Identifying High-Growth Firms

Daniel Halvarsson*
Identifying High-Growth Firms

Daniel Halvarsson*

Abstract

This paper investigates the role(s) of high-growth firms (HGFs) in the robust growth-rate distribution. HGFs are identified as firms for which the growth-rate distribution exhibits power-law decay. In contrast to the traditional means of identifying HGFs, a distributional approach eliminates the need to specify an arbitrary growth rate or percentage share. The latter approach is illustrated by the growth-rate distribution for Swedish data on incorporated firms at the aggregate level and at the 2-digit industry level. The empirical results indicate that a power law is sometimes present in the growth-rate distribution and suggest that HGFs are rarer than previously thought.

Keywords: High-growth firms · Gazelles · Firm growth-rate distribution · Laplace distribution · Power law

JEL classification: L11 · L25 · D22

*The Royal Institute of Technology, Division of Economics, SE-100 44 Stockholm, Sweden; and The Ratio Institute, P.O Box 3203, SE-103 61 Stockholm, Sweden, tel: +46760184541, e-mail: daniel.halvarsson@ratio.se. I thank seminar participants at Ratio for providing valuable comments and suggestions on a previous version of this paper.
1 Introduction

Rapid firm growth is observed in only a handful of firms (Henrekson and Johansson, 2010). These high-growth firms (HGFs) or “Gazelles” are believed to be important for the creation of new jobs and are the subject of a growing field of research (Birch and Medoff, 1994; Brüderl and Preisendörfer, 2000; Davidsson and Henrekson, 2002; Delmar et al, 2003; Littunen and Tohmo, 2003; Halabisky et al, 2006; Acs and Mueller, 2008; Acs, 2011). Empirical evidence suggests that these firms are more prone to hiring groups of people that traditionally have had difficulty entering the labor market (Coad et al, 2011). These attributes have made HGFs an interesting phenomena from a policy perspective, where they are considered conducive to economic growth and seen as potential instruments for combating unemployment (European-Commission, 2010; Höflzl, 2011; Daunfeldt and Halvarsson, 2011; Daunfeldt et al, 2012a).

The purpose of this paper is to determine ways of differentiating HGFs from other firms that have lower growth rates. In the previously published literature, HGFs are often considered to be a certain percentage of firms with the highest growth rates (Henrekson and Johansson, 2010), or in other words, firms with growth rates higher than the e.g. 99th, 95th or 90th percentile (Daunfeldt and Halvarsson, 2012b).

The percentage shares that are designated for HGFs differ somewhat between studies but usually range between 1 and 10 percent fastest growing firms. In an attempt to standardize the definition, OECD/EUROSTAT (henceforth OECD) adopted a measure based on a minimum growth boundary (Ahmad, 2008). What distinguishes a HGF from other firms, according to OECD, is that the HGFs have experienced an annualized employment growth rate of at least 20 percent over a 3-year period and had an initial size of no less than 10 employees. Thus, the current means by which HGFs are selected are not based on theoretical arguments, begging the questions of why high growth should occur at precisely 20 percent or why high growth should pertain to the fastest growing “X” percent of firms?

Heuristics can, of course, be important, because they facilitate comparisons between studies. However, there remains a great need to develop a more theoretical understanding of HGFs (beyond a list of empirical descriptives). By investigating the definition of HGFs with regard to the properties of the stylized firm growth-rate distribution, this paper takes a potential first small step towards a theoretical understanding of HGFs.
Within the literature on firm growth, there exists a parallel discussion to HGFs about the statistical properties of firm growth-rate distributions (Stanley et al., 1996; Lee et al., 1998; Axtell, 2001; Bottazzi and Secchi, 2003, 2006; Fagiolo and Luzzi, 2006). In contrast to the common Gaussian distribution, the firm growth-rate distribution is distinctly tent-shaped, with heavy tails and a high singular spike at growth rates of approximately zero. Evidence across countries and at different levels of industry aggregation suggests that the growth-rate distribution is at least Laplace distributed (see e.g. Stanley et al., 1996; Bottazzi and Secchi, 2005; Erlingsson et al., 2012), with tails that possibly follow a power-law distribution above some growth rate (Fu et al., 2005; Gabaix, 2011; Schwarzkopf et al., 2010). Its ubiquitous shape has been regarded as “one of the most important pieces of evidence able to throw some light on the underlying drivers of corporate growth[...].” (Dosi and Nelson, 2010, p.44).

The emergence of a Laplace distribution in firm growth rates has been related to various suggestive generating mechanisms, such as classical competition in Alfarano and Milaković (2008) and Alfarano et al. (2012), increasing returns in Bottazzi and Secchi (2006) and financial fragility in Gatti et al. (2005). In addition, a Laplace distribution follows from generalized versions of the central limit theorem, where the Laplace distribution constitutes a limit to schemes of random sums of random variables (e.g. Klebanov et al., 2006). While these explanations may account for part of its ubiquity, they do not explain the presence of power-law tails observed in Fu et al. (2005) for international pharmaceutical firms. These researchers show that the heavy tails in their model arise from bursts of growth experienced by smaller firms having only a few products. If the commercialization of additional products is successful, the resulting burst could generate a power-law tail in the growth-rate distribution.

However, for the purposes of this paper, irrespective of the true generating mechanism behind the shape of the growth-rate distribution, it offers a formalism that can be used to distinguish high-growth firms from the bulk of other firms with zero or low growth rates. In particular, I exploit the possibility of a probabilistic break between the Laplace and power-law distribution in the growth-rate distribution to identify HGFs as firms for which the growth-rate distribution decays with a power law. Furthermore, this break can be estimated directly from the data, which circumvents the problem of imposing some arbitrary percentage share or growth threshold. In contrast to previous definitions, a distributional definition of HGFs allows the share and growth requirement to vary across regions, over different levels of industry aggregations, and between
countries. Moreover, by tying HGFs to power-law signatures in the growth-rate distribution, the unique statistical properties of power laws may shed new light on the composition of these firms.

The above approach is illustrated by examining the aggregate and 2-digit industry growth-rate distributions for incorporated firms in Sweden. The analysis considers three consecutive 3-year periods, in which the event of a power law is estimated by a novel technique developed by Clauset et al (2007; 2009). The results confirm the presence of a power law in some industries, but a power-law presence could be rejected for the aggregate growth-rate distribution in two out of three periods. The results also indicate that the number of HGFs could be considerably lower, if they exist at all, when this definition of HGF is used rather than previous definitions. Except for three industries, the presence of a power law in the growth-rate distribution could be rejected in at least one period, suggesting that the growth-rate distribution and thus HGFs might be sensitive to external conditions that fluctuate over time.

The contributions of this paper are threefold. First, it provides an alternative approach for studying HGFs, an approach that eliminates the need to choose some appropriate percentage share or growth threshold. Second, this paper goes one step further than many previous studies that attempt to fit a parametric function to the empirical distribution of firm growth. A formal test is provided to test for the presence of a power law in the right tail of the growth-rate distribution. Third, new evidence is provided about the minimum growth requirements and frequencies of HGFs at the aggregate and at the 2-digit industry level.

The remainder of the paper is structured as follows. The following section briefly reviews literature related to HGFs and the growth-rate distribution. Then, section 3 introduces a new definition for HGFs based on the properties of the growth-rate distribution, and section 4 briefly covers the estimation strategy. In section 5, data are presented with descriptive statistics. Section 6 presents the results. The final section summarizes the main contributions of this paper and concludes with a discussion of limitations and suggestions for future research.
2 Related literature

2.1 High-growth firms

The increasing interest in HGFs is centered on the idea that some firms are more important than others. Compared to other categories of firms, HGFs are found to be responsible for a large share of net job creation in the economy (Birch and Medoff, 1994; Delmar et al, 2003; Littunen and Tohmo, 2003; Acs and Mueller, 2008). Furthermore, Coad et al (2011) argue that these firms are also more likely to hire groups of people who are known to have difficulties entering the labor market (such as immigrants and young persons). In a survey of the empirical literature, Henrekson and Johansson (2010) conclude that the typical HGF is a young and small-sized firm that grows organically. While larger and older firms are represented among HGFs, they are often found to be much fewer in number. In their survey, these researchers have also found that HGFs exist in all industries, with no discernible over-representation in high-technology industries.

The empirical findings in the HGF literature can be interpreted in light of Schumpeter’s two technological regimes (Capasso et al, 2009). Early on, Schumpeter (1911) focused on the roles of new-entry and small entrepreneurial firms as important drivers of innovation and economic growth. Later, Schumpeter (1942) instead emphasized the role of large firms as the motors of innovation. Among these large firms, the innovative process is routinized and propelled by increasing returns to scale. The skewness among HGFs towards small firms seems to favor Schumpeter’s earlier view, which is supported by evidence of their innovativeness. Looking at HGFs in 16 different countries, Hözl (2009) finds evidence that HGFs are more innovative than other firms if the country is near the technological frontier. Furthermore, Stam and Wennberg (2009) also find evidence that R&D matters for HGFs, especially in the earlier stages of business.

Most previously published studies of HGFs are empirical and use a number of different definitions to differentiate HGFs from the general population of firms (Delmar et al, 2003; Daunfeldt et al, 2010). Perhaps the most common approach is to select the top 1, 3, 5 or 10 percent fastest growing firms in a population, where growth is measured either by percentage change or by first difference (Henrekson and Johansson, 2010). Other definitions include the OECD definition, which proposes to define HGFs as firms with an annualized
employment growth rate of at least 20 percent over a 3-year period, provided
the firm has at least ten employees at the beginning of that period (Ahmad,
2008).

The above definitions rely on the researcher to choose the relevant minimum
growth boundary or the percentage share. For instance, consider the definition
based on the top 10-percent of highest growing firms. In this case, HGFs are
likely to include firms with hardly any growth (Bjuggren et al, 2010), which by
any relevant criteria should not be regarded as having “high” growth.

Moreover, regardless of which percentage share is used, some firms will al-
ways be classified as HGFs, irrespective of their growth characteristics. More
importantly, however, there is no reason to expect that high growth should oc-
cur for a certain percentage of the firms, or why 20 percent is the most relevant
benchmark. The OECD definition also leaves little room for variation in the
growth requirement for different industries.\(^1\)

Conditions for high growth may e.g. differ substantially for services and man-
facturing industries. For instance, hospitality industries (restaurants, cafes,
etc.), where capital intensity and economies of scale are less pronounced, typi-
cally experience small sunk costs compared to manufacturing firms, where young
firms must rapidly attain a minimum efficient scale in order to survive (Mans-
field, 1962; Audretsch et al, 2004). Therefore, it is certainly plausible that
different growth requirements should apply to different industries.\(^2\)

The current void in theoretical modeling is perhaps the biggest challenge for
future HGF research. Even if a formal economic foundation is far from being
realized, the statistical properties of the stylized growth-rate distribution offer
a convenient way to examine and compare these firms.

\subsection{2.2 The firm growth-rate distribution}

The distribution of (log) firm growth is characterized by heavy tails and a high
singular peak. The tent-shaped distribution was first discovered by Stanley
et al (1996) using U.S. firm data. Since then, it has generated a growing body
of literature. The evidence is striking and points to a growth-rate distribution
for which the distribution is at least Laplace (double-exponential), which is
\footnote{Daunfeldt et al (2012a) argue that the OECD definition also systematically discriminates against small firms, which are known to be important job creators (Birch, 1979). For a comprehensive study of the existing definitions of HGFs, I refer to the study by Daunfeldt et al (2010).}

\footnote{The minimum efficient scale is a standard concept in industrial organization and refers to the minimum firm size required to produce at a long term average cost minimum.}
symptomatic of firm dynamics in which most firms do not grow and where only a small fraction experience rapid growth rates (e.g., Lee et al, 1998; Bottazzi and Secchi, 2003; Reichstein and Jensen, 2005; Fagiolo and Luzzi, 2006; Coad and Planck, 2011; Schwarzkopf et al, 2010; Alfarano et al, 2012).

This property of growth rates is found to be robust and applicable to different growth indicators, different countries, and for different levels of industry aggregations. Alongside empirical regularities, such as Zipf’s law and Gibrat’s law, the tent-shape of the growth-rate distribution is now regarded as a robust stylized fact in empirical industrial organization and evolutionary economics (Dosi and Nelson, 2010).

The discovery of a Laplace-type growth-rate distribution was surprising, because it contradicts one of the basic building blocks of firm dynamics, namely Gibrat’s law of proportionate effect (see, for example, Geroski, 1995; Caves, 1998; Lotti et al, 2003; Audretsch et al, 2004) for an exhaustive survey of Gibrat’s law). This law stipulates that firms grow at the same proportional rate, independent of their size, which implies that the firm growth-rate distribution is i.i.d. and Gaussian. Dosi and Nelson (2010) remarks that the presence of non-Gaussian tails may have important implications for understanding the firm-growth process and suggests the existence of an underlying correlating mechanism. This would not be the case, however, under Gibrat’s law, in which growth rates are purely random.

The Laplace distribution can be expressed by its density function,

\[ p(g) = \frac{1}{2\sigma} \exp \left( -\frac{|x - \mu|}{\sigma} \right), \tag{1} \]

where, the parameter \( \sigma > 0 \) is a scaling parameter that determines the width of the distribution.\(^3\)

The field has yet to reach a consensus explanation for the emergence of this shape of the firm growth-rate distributions, but a number of mechanisms have been suggested. Bottazzi and Secchi (2006) suggests that the distribution results from the interdependence of competing firms, where increasing returns from growth generate the heavy tails.

In another possible explanation, provided by Coad and Planck (2011), the Laplace shape results from the hierarchical structure of a firm. In their model, growth opportunities (for the firm) lead to the potential hiring of additional

\(^3\)Taking the log of \( p(g) \) produces the well-known shape of two straight lines emanating down from the center of the distribution.
workers at the lowest level. In need of supervision, these new employees may then induce the firm to hire additional personnel higher up in the organization. On the other hand, Alfarano and Milaković (2008) and Alfarano et al (2012) show that the Subbotin distribution with the nested Laplace distribution constitutes a type of statistical equilibrium (as conceived by Foley, 1994) that emerges under conditions alluding to classical competition. Interestingly, Alfarano et al (2012) demonstrate that the dispersion parameter $\sigma$ in (1) is determined by two interdependent competitive forces. While firms systematically drift towards profit/growth equalization, this drift is intimately connected to a stochastic component, specific to the firm, that drives the dynamics of growth. However, only in the Laplace equilibrium is the competitive pressure equal for all firms, independent of their profit levels/firm size (Alfarano et al, 2012).

The above models shed some light on the possible generating mechanism for the Laplacian. However, further studies present evidence that the growth-rate distribution is endowed with even heavier tails (Reichstein and Jensen, 2005; Bottazzi et al, 2011). Examining French manufacturing firms, Bottazzi et al (2011) discovered that high growth rates are characterized by substantially heavier tails than described by equation (1), and remarks that “the Laplace distribution of growth rates cannot be considered as a universal property valid for all sectors” (Bottazzi et al, 2011, p.2). Moreover, Fu et al (2005), Schwarzkopf et al (2010) and Gabaix (2011) find that the tails of the growth-rate distribution is better described by a power law.

Fu et al (2005) refine previous findings and argue that the Laplace distribution do constitute a good fit for the central part of the growth-rate distribution but that a power-law distribution is a better description for the tails. As with the Laplace distribution, a power law can be defined via its distribution function, here by its counter cumulative distribution,

$$P(g > x) = \left( \frac{x}{g_{min}} \right)^{-\gamma}, \text{for } x > g_{min} > 0,$$

(2)

where $g$ is the growth rate of some firm, and $\gamma$ is the characteristic exponent that determines the frequency of high-growth events. The crucial parameter for the study of HGFs is $g_{min}$, which constitutes a minimum growth boundary around which the probabilistic behavior changes. If (2) is indeed an accurate description of the tails of the growth-rate distribution, this growth boundary determines the growth rate where the probability for high growth suddenly increases.

To account for the excess variance observed in empirical growth rates, Fu
et al (2005) construct a proportional growth model based on the constituent units of the firm, where growth can occur by either expanding the number of products sold (scope) or by increasing the size of already existing products (scale). The mechanism presented in Fu et al (2005) predicts that the Laplace center in the growth-rate distribution is mainly due to the presence of large firms with a lot of existing products in place. If a large multi-product firm grows, either in scope or in scale, the resulting growth-rate distribution is likely Laplace. On the other hand, for smaller firms with more limited product lines, the commercialization of an additional product, if successful, can result in extreme bursts of growth that account for the power-law signatures in the growth-rate distribution. Thus, they find that the resulting growth-rate distribution is Laplace for $|g| \to 0$ and decays with a cubic power law for $|g| \to \infty$ (Fu et al, 2005).\footnote{The resulting probability distribution in Fu et al (2005) can be approximated by $P(g) \approx \frac{2V}{\sqrt{g^2 + 2V}} \left( \left( |g| + \sqrt{g^2 + 2V} \right)^2 \right)$. For $|g| \to 0$, the cusp is Laplace, $P(g) \sim \exp(-|g|)$ and for $|g| \to \infty$, the tails decay with a cubic power law, $P(g) \sim |g|^{-3}$.}

The rapid growth among small and young firms agree with the findings in the literature regarding Gibrat’s law, where small firms are found to grow at a faster pace than large firms (Evans, 1987a; Hart and Oulton, 1996; Dunne et al, 1989; Dunne and Hughes, 1994; Calvo, 2006; Evans, 1987b). According to Audretsch et al (2004), this is especially true in the manufacturing industry.

3 Arriving at a distributional definition of high-growth firms

Investigations of firm growth-rate distributions suggest that the tails might be different. While most firms are contained under a Laplace cusp, there are some firms with rapid growth rates, where the growth-rate distribution has a possible power-law distribution (Fu et al, 2005). The existence of such a property suggests a non-arbitrary way to separate firms with rapid growth rates (HGFs) from firms with lower or marginal growth. If such a break exists in the empirical growth-rate distribution, it will occur for some minimum growth boundary $g_{\min}$, which motivates the following distributional definition of HGFs:

**Definition 1.** HGFs are defined as the set of all firms $i = 1, ..., n$ with growth rates higher than $g_{\min}$, above which the growth-rate distribution decays with a power law.
HGFs := \left\{ \text{i.s.t. } P(g_i > x) = \left( \frac{x}{g_{\text{min}}} \right)^{-\gamma} \right\}, \text{ for } x > g_{\text{min}} > 0. \quad (3)

As before, the parameter $\gamma > 0$ determines the shape of the distribution, where a small value indicates a higher frequency of HGFs (while a large value indicates a lower frequency of HGFs). Definition 1 suggests an alternative approach for selecting HGFs. Because the existence of $g_{\text{min}}$ can be determined and its value estimated directly from the data, the researcher is not required to choose a suitable percentage share or growth requirement (as is customary in the HGF literature today).

The definition also allows the number of HGFs to vary between samples. The same considerations apply to the minimum required growth rate $g_{\text{min}}$, which, in contrast to the OECD definition, is not fixed at a rate of 20 percent. Even if a distributional definition of HGFs would solve some of the problems associated with previous definitions, the question would still remain as to why a power law in the growth-rate distribution is the signature feature of HGFs.

Although power laws have been observed in the right tail of the growth-rate distribution, power laws have, to my knowledge, never previously been associated with HGFs outright. Power laws go far beyond considerations of firm growth-rate distributions and have been documented for a variety of other economic phenomena. Well-known cases include the distribution of firm sizes (Axtell, 2001), job vacancies (Gunz et al, 2002), entrepreneurship and innovations (Poole, 2000), CEO pay (Gabaix and Landier, 2008) and ‘Fordist’ organization structures (Stanley et al, 1996). The observation that similar empirical regularities, such as power laws, arise in diverse phenomena suggests that some universal traits may govern basic growth process (Schwarzkopf et al, 2010). In fact, there exist a number of causal mechanisms capable of generating power laws. These are sometimes referred to as scale-free (invariant) theories.\footnote{See Andriani and McKelvey (2009) for a comprehensive survey of possible mechanisms.}

As argued by Fu et al (2005) and Gabaix (2011), power laws in firm growth can stem from the distribution of the constituent units of firms. Power-law bursts in growth rates have also been associated with different types of organizational structures. For the typical ‘Fordist’ organization structure, where
orders are carried out from top to bottom, Stanley et al (1996) give an example where such top-down management can give rise to episodes of power-law growth. The same applies to self-organized bottom-up management (Andriani and McKelvey, 2009). They conclude that the process governing firm growth is likely to be scale invariant (Andriani and McKelvey, 2009).

Finally, scale-free (power-law) dynamics by their nature may hold attractive properties from the perspective of an entrepreneur or intrapreneur, a consideration that is well captured by the following suggestive quote:

“Who more than entrepreneurs wouldn’t like to let loose SF (scale free) dynamics in their firms? Think of how many small entrepreneurial ventures stay that way simply because the emergent growth dynamics they had at the one- or two-level size failed to scale up as levels increased. Think how many large organizations show failing intrapreneurship for the same reason—the hundreds of “butterfly ideas” never become meaningful butterfly events, never produce butterfly effects, and never spiral into multilevel SF causal dynamics producing power-law signatures.” (Andriani and McKelvey, 2009, p.1056, parenthesis added).

However, identifying the exact dynamic that generates the power-law signatures observed for HGFs is outside the scope of this paper and therefore remains speculative. It is sufficient, for the purposes of this paper, to impose some distinguishing feature that separates HGFs from other firms, regardless of whether a power law in the growth-rate distribution is the most suitable feature or not. The following section describes the empirical strategy used to test for a power law in the firm growth-rate distribution and to estimate the minimum growth boundary $g_{min}$.

4 Empirical strategy for identifying high-growth firms

Now that HGFs have been defined, the remaining challenge lies in identifying these firms, which translates into the problem of testing for the existence of a

6While empirically examining aspects behind power laws in the cross-sectional growth-rate distribution may prove challenging, testing the effect from various industry specific variables on e.g. the minimum growth boundary $g_{min}$ is entirely plausible. Given a large enough sample size of estimated values on $g_{min}$, a second stage regression analysis, containing industry variables, could in principle be conducted. An analysis of such sorts, however, would likely require more data points on $g_{min}$ than available at the 2-digit industry level and has not been attempted in this paper.
power law somewhere in the right tail of the growth-rate distribution. One common way to ‘test’ for a power law is to inspect the logarithm of the distribution function in equation (2),

$$\log P(g) = \gamma \log g_{\text{min}} - \gamma \log x,$$

(4)

which is linear in $\log x$ with a negative slope of $-\gamma$, resulting in a so-called log-log plot. However, a graphical approach introduces a host of biases and possible Type I and Type II errors (Goldstein et al., 2004; Clauset et al., 2009; Eeckhout, 2009). For example, Eeckhout (2009) notes that a log-log plot severely distorts the tails of the data, making it difficult to assess the correct values of the parameters. To identify HGFs without falsely overestimating their presence, a more systematic method is needed.

To overcome the bias associated with graphical methods, Clauset et al. (2009) developed an appropriate method for estimating the parameters $g_{\text{min}}$ and $\gamma$. The following two-step procedure is implemented here:

1. Estimation To find an unbiased estimate of $\gamma$, a correct minimum growth boundary $g_{\text{min}}$ must be determined. This is because a power law diverges for values of $g \to 0$. To find $g_{\text{min}}$, one has to find the value of the parameter $g_{\text{min}}$ that minimizes the Kolmogorov-Smirnov distance $D$,

$$D = \max_{g > g_{\text{min}}} |P_e(g) - P(g)|,$$

(5)

evaluated over an array of possible $g_{\text{min}}$ values. The distance function $D$ measures the maximum absolute vertical distance between the empirical distribution $P_e(g)$ and the best fit power-law distribution $P(g)$, computed for $g > g_{\text{min}}$. Should the growth-rate distribution follow a power law somewhere in the right tail, the estimate $g_{\text{min}}$ then becomes the most probable minimum growth boundary. For each $g_{\text{min}}$, the parameter $\gamma$ is estimated via maximum likelihood,

$$\hat{\gamma} = m \left[ \sum_{i=1}^{m} \log \frac{g_i}{g_{\text{min}}} \right]^{-1},$$

(6)

Clauset et al. (2009) also developed a third step, in which a statistically significant power law is examined with regards to other distributions. Thus, the two-step procedure undertaken here only allows for a test of a power law. It is still possible that other distributions may provide a better fit.
where \( g_i > g_{\text{min}} \) is the growth rate for firms \( i = 1, \ldots, m \leq n \). Plugging \( \hat{g}_{\text{min}} \) into (6) then gives the appropriate estimate of \( \gamma \).\(^8\)

2. **Testing** Because it is possible to identify a boundary and to estimate the parameters even if the growth-rate distribution does not follow a power law, a formal test is needed. One way to accomplish this is to generate a large number of synthetic data sets with properties similar to the empirical data. Then, p-values can be determined by comparing the Kolmogorov-Smirnov distance in (5) for each synthetic data set to the empirical data. Importantly, when interpreting the p-values, a high p-value is considered evidence for the existence of a power law (in contrast to the usual approach) because, in this case, the null hypothesis is the existence of a power law. Clauset et al (2009) suggest using a p-value of 0.1 to judge the significance. Thus, for p-values higher than 0.1, the null hypothesis of a power law cannot be rejected. (For a detailed description of the method, see Clauset et al, 2009)\(^9\)

Finally, to find the variance of \( \hat{g}_{\text{min}} \) and \( \hat{\gamma} \), 50 new data sets are bootstrapped from the empirical data and estimated separately, using step 1.

### 5 Data and descriptive statistics

The distributional approach used to identify HGFs is undertaken using Swedish data for incorporated firms covering the period 1995-2010. The data set comes from the Swedish Patent and registration office (PRV) and contains information on a number of accounting variables, such as the number of employees, registration dates, sales, R&D expenditures, and profits. It is compiled by PAR-AB that specializes in collecting detailed market information often used by decision makers in business.

The unit of analysis here is the firm growth-rate distribution, composed of growth rates for individual firms. I have chosen to study the growth-rate distribution of growth rates measured over 3-year periods instead of, for instance,

\[^8\]Here \( P^e(g) = \frac{1}{k} \sum_{i=1}^{k} I_{g_i \geq g_k} \) is the empirical distribution of \( g \), and \( I \) is the indicator function taking the value 1 if \( g_i \geq g_k \), and 0 otherwise. The best-fitted distribution is given by the CDF, \( P(g) = \left( \frac{g}{g_k} \right)^{-\gamma_k} \). For each possible \( g_k \) in the data set, the exponent \( \gamma_k \) is computed by maximum likelihood. Then, whichever \( g_k \) minimizes the Kolmogorov-Smirnov distance in (5) gives the desired estimate \( \hat{g}_{\text{min}} \) and the corresponding \( \hat{\gamma} \).

\[^9\]To locate the most probable boundary \( \hat{g}_{\text{min}} \), I use the program ‘pfit.m’; and, to calculate p-values, I use ‘plval.m’ described in Clauset et al (2009).
annual growth rates, which are frequently used in previously published studies of the firm growth-rate distribution. Even if annual growth rates figure in the empirical analysis of HGFs, longer growth rates are more common, as made apparent by OECD’s definition. Longer growth rates should also contain less noise than annual growth rates, a favorable condition when studying empirical distributions.

Due to limitations in the data, the first and last time periods were dropped. To improve the quality of the dataset the following additional treatments have been employed. Because, according to Henrekson and Johansson (2010), HGFs are mainly found among organically growing firms, merging or acquiring firms have been excluded (to the extent that such information is available). Firms that belong to a business group have also been disregarded, because subsidiaries can sometimes grow at the expense of other firms in the group. Moreover, from the 3-year measure of growth, firms that either entered or exited sometime during this period are not included. Furthermore, in addition to the aggregate firm growth-rate distribution, the study also examines the growth-rate distribution for firms in a number of 2-digit industries. Firms with invalid or missing NACE (rev. 2) codes have also been excluded from the sample. Finally, a minimum number of constituent firms is required to study its growth-rate distribution. I look at industries with at least 1000 registered firms over the three periods (2000-2003, 2003-2006 and 2006-2009), which results in a wide range of 21 industries including manufacturing, services, retail and wholesale. The industries are listed in Table A.1.1 in Appendix A.1.

To measure growth, there are a number of different indicators, such as the number of employees, sales or market shares, to choose from. These should be distinguished from, for instance, value added and profits that are more suitable for measuring firm performance than firm growth (Coad, 2009). Surveying HGFs, Henrekson and Johansson (2010) finds that employment is the most common growth indicator. Because employment growth is easily related to job-creation statistics and unemployment figures, the number of employees is often used for policy relevance. Thus, keeping with tradition, this study uses the number of employees to measure firm size.

Herein, firm growth $g$ is measured by the log difference of the number of employees.
employees $S$ over a 3-year period. To compare growth-rate distributions across industries, each growth rate is normalized by the average industry growth $\bar{g}_{I,t}$ and its standard deviation $\sigma_{g_{I,t}}$. Thus, the growth from time $t-3$ to $t$ for firm $i$ in industry $I$ is measured as follows:

$$g_{i \in I,t} = \left( \log \frac{S_{i \in I,t}}{S_{i \in I,t-3}} - \bar{g}_{I,t} \right) / \sigma_{g_{I,t}}.$$  \hspace{1cm} (7)

Table 1 presents descriptive statistics for $g$ computed for the period 2006-2009. The final sample contains 81,028 firms registered in 21 different industries. As a result of the normalizing scheme, mean growth and zero standard deviations are equal to zero and one.

Figure 1 displays the shape of the growth-rate distribution once all industries have been combined. The kernel density, with a logarithmic scaled vertical axis, displays the typical tent-shape. Focusing on high firm growth, the interesting segment in Figure 1 is the right tail, where an upward bend suggests the presence...
of a power law (above some minimum growth boundary).

However, as pointed out above, graphical inspection can be extremely difficult, which is why I go on to present the results from the empirical test.

6 Results

Results are presented for the aggregate growth-rate distribution, where all 21 industries have been combined, and then for each respective industry. Because only the cross-sectional growth-rate distribution is considered, results from three periods (2000-2003; 2003-2006; 2006-2009) are provided to reflect possible business cycle effects. A statistically significant power law is taken as confirmation that HGFs exist.

6.1 The aggregate growth-rate distribution

All the results from estimating (2) are presented in Tables 2 to 4. The aggregate results, presented at the bottom row of each table, provide mixed evidence for the presence of a power law. Significance is only observed in the latest period of 2006-2009, where a highly significant p-value of 0.98 is reported. Observe, once more, that a high p-value means that the null hypothesis of a power law is less likely to be rejected.

This result can also be illustrated by graphing the log-log plot for the right tail of the growth-rate distribution described above.\textsuperscript{11} If the tail follows a power

\textsuperscript{11}To plot the empirical data, I use the program ‘plplot.m’ described in Clauset et al (2009).
law above some minimum growth boundary, the plot should exhibit a straight line. The estimated power law is here demarcated with a straight line, beginning at the estimated boundary that constitutes the event of high growth. In the earlier periods, where a power law can be rejected, the growth-rate distribution decays faster, seen in Figure 2 (b) and (c) as a downward bend from the line in the growth-rate distribution.

For the significant period, high growth is estimated to begin at a growth rate of $\hat{g}_{min} = 5.09$, which corresponds to 75 HGFs that comprise 0.09 percent of the sample. The estimated exponent $\hat{\gamma}$ is found to be 7 and is more than twice the magnitude found by Fu et al (2005) for international pharmaceutical firms.

This result indicates that HGFs should occur less frequently even if they follow the same type of power-law dynamic. An important difference, however, is that Fu et al (2005) study involves annual growth rates, which are known to be more volatile than longer growth rates. For the two earlier periods, HGFs are absent, meaning that a power law is rejected. In both cases, the numbers of proposed HGFs are substantially larger, reflecting the relatively low estimates

6.2 The industry-level growth-rate distribution

The results for the 2-digit industry level suggest that, even when it is rejected at the aggregate level, a power law can still be present in the growth-rate distribution at the industry level. For industries like metals manufacturing (NACE-25), wholesale trade (NACE-45) and retail trade (NACE-46), a power law is statistically significant throughout all periods. In the latest period, for instance, 20 metals-manufacturing HGFs have growth rates higher than 3.1, which translates to an increase of more than 20 times the initial size over three years. As a percentage share of all firms within this industry, HGFs constitute 0.72 percent, which is quite close to the common 1-percent definition of HGFs. This result can be interpreted in light of Audretsch et al (2004), where young and small firms in manufacturing are expected to grow at a faster rate than their counterparts in the service industry.

For each of the other industries, a power law could be rejected in at least one period. Two exceptions to this observation are legal and accounting services (NACE-69) and management consultancy activities (NACE-70) that never registered HGFs in any of the periods. This result suggests that, although power laws are sometimes present at the industry level, there is a substantial variation over time, and little systematic differences can be discerned between industries.

In considering the scale-free properties of power-law distributions, it is interesting to assess whether the sum of HGFs found at the industry level is also reflected in the number of HGFs found in the aggregate growth-rate distribution. In the 2006-2009 period, when considering only statistically significant results, the total number of industry HGFs amounts to 83 firms, compared to 75 at the aggregate level. These values are surprisingly close and warrant further, more detailed investigations to better understand which industries create the power-law tails and the HGFs observed for higher levels.

Finally, Figure 3 presents the corresponding log-log plot for growth-rate distributions at the industry level. Compared to the aggregate plots in Figure 2, the industry plots display a more lumpy growth profile, with clustering in some industries around certain growth rates. Furthermore, as demonstrated by the right tails, the industry graphs also suggest that high firm growth is characterized by substantial inter-industry heterogeneity.

\[ \exp(3.1)-1 \approx 20 \]

\[ ^{12}\text{Thus, } \exp(3.1)-1 \approx 20 \]
Figure 3. Estimated power laws in the right tail for 2-digit industry growth-rate distributions in 2006-2009.
### Table 2. Estimation results for 2-digit growth-rate distributions during period 2006-2009.

<table>
<thead>
<tr>
<th>Industry (NA CE rev.2)</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\sigma}_\gamma$</th>
<th>$\hat{g}_{min}$</th>
<th>$\hat{\sigma}<em>{g</em>{min}}$</th>
<th>p-value</th>
<th>$D^b$</th>
<th>Share$^c$</th>
<th>HGFs$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crop and animal prod.</td>
<td>4.515</td>
<td>0.772</td>
<td>2.217</td>
<td>0.660</td>
<td>0.000</td>
<td>0.215</td>
<td>1.46%</td>
<td>36</td>
</tr>
<tr>
<td>2. Forestry and logging</td>
<td>6.727</td>
<td>1.263</td>
<td>2.684</td>
<td>0.634</td>
<td>0.253</td>
<td>0.153</td>
<td>1.18%</td>
<td>15</td>
</tr>
<tr>
<td>16. Manufacture of wood</td>
<td>2.733</td>
<td>0.012</td>
<td>0.689</td>
<td>0.000</td>
<td>0.000</td>
<td>0.185</td>
<td>17.19%</td>
<td>169</td>
</tr>
<tr>
<td>25. Manufacture of metal</td>
<td>5.457</td>
<td>0.324</td>
<td>3.100</td>
<td>0.324</td>
<td>0.102</td>
<td>0.155</td>
<td>0.72%</td>
<td>20</td>
</tr>
<tr>
<td>41. Construction of buildings</td>
<td>3.956</td>
<td>0.442</td>
<td>1.771</td>
<td>0.354</td>
<td>0.000</td>
<td>0.188</td>
<td>3.98%</td>
<td>128</td>
</tr>
<tr>
<td>43. Specialized construction</td>
<td>6.143</td>
<td>0.502</td>
<td>4.864</td>
<td>0.559</td>
<td>0.370</td>
<td>0.157</td>
<td>0.08%</td>
<td>9</td>
</tr>
<tr>
<td>45. Wholesale and retail, motor</td>
<td>2.661</td>
<td>0.010</td>
<td>0.656</td>
<td>0.000</td>
<td>0.000</td>
<td>0.217</td>
<td>18.93%</td>
<td>668</td>
</tr>
<tr>
<td>46. Wholesale trade, ex motor</td>
<td>5.458</td>
<td>0.304</td>
<td>4.838</td>
<td>0.456</td>
<td>0.839</td>
<td>0.098</td>
<td>0.16%</td>
<td>13</td>
</tr>
<tr>
<td>47. Retail trade, ex motor</td>
<td>5.197</td>
<td>0.354</td>
<td>4.598</td>
<td>0.666</td>
<td>0.188</td>
<td>0.138</td>
<td>0.17%</td>
<td>18</td>
</tr>
<tr>
<td>49. Land transport</td>
<td>5.837</td>
<td>0.853</td>
<td>2.742</td>
<td>0.450</td>
<td>0.000</td>
<td>0.158</td>
<td>1.07%</td>
<td>69</td>
</tr>
<tr>
<td>56. Food services</td>
<td>9.448</td>
<td>1.780</td>
<td>3.991</td>
<td>0.791</td>
<td>0.429</td>
<td>0.158</td>
<td>0.26%</td>
<td>8</td>
</tr>
<tr>
<td>62. Computer programming</td>
<td>5.942</td>
<td>0.308</td>
<td>3.460</td>
<td>0.203</td>
<td>0.005</td>
<td>0.179</td>
<td>0.08%</td>
<td>25</td>
</tr>
<tr>
<td>68. Real estate activities</td>
<td>5.772</td>
<td>0.142</td>
<td>3.364</td>
<td>0.000</td>
<td>0.003</td>
<td>0.206</td>
<td>0.56%</td>
<td>16</td>
</tr>
<tr>
<td>69. Legal and accounting</td>
<td>2.132</td>
<td>1.882</td>
<td>0.598</td>
<td>0.725</td>
<td>0.000</td>
<td>0.280</td>
<td>14.94%</td>
<td>611</td>
</tr>
<tr>
<td>70. Management consultancy</td>
<td>5.600</td>
<td>0.145</td>
<td>3.237</td>
<td>0.067</td>
<td>0.000</td>
<td>0.178</td>
<td>0.84%</td>
<td>48</td>
</tr>
<tr>
<td>73. Advertising research</td>
<td>4.161</td>
<td>0.146</td>
<td>2.218</td>
<td>0.100</td>
<td>0.000</td>
<td>0.202</td>
<td>2.38%</td>
<td>50</td>
</tr>
<tr>
<td>74. Other professional, science</td>
<td>5.232</td>
<td>0.121</td>
<td>3.106</td>
<td>0.000</td>
<td>0.006</td>
<td>0.199</td>
<td>1.01%</td>
<td>20</td>
</tr>
<tr>
<td>81. Services to buildings</td>
<td>4.450</td>
<td>0.112</td>
<td>2.203</td>
<td>0.189</td>
<td>0.004</td>
<td>0.144</td>
<td>2.65%</td>
<td>37</td>
</tr>
<tr>
<td>85. Education</td>
<td>6.167</td>
<td>0.658</td>
<td>3.748</td>
<td>0.572</td>
<td>0.059</td>
<td>0.182</td>
<td>0.74%</td>
<td>12</td>
</tr>
<tr>
<td>86. Health activities</td>
<td>4.012</td>
<td>0.127</td>
<td>3.408</td>
<td>0.065</td>
<td>0.057</td>
<td>0.163</td>
<td>0.60%</td>
<td>20</td>
</tr>
<tr>
<td>90. Creative, arts</td>
<td>6.115</td>
<td>0.046</td>
<td>3.768</td>
<td>0.203</td>
<td>0.556</td>
<td>0.239</td>
<td>0.00%</td>
<td>3</td>
</tr>
<tr>
<td>∑ All industries</td>
<td>7.052</td>
<td>0.7982</td>
<td>5.088</td>
<td>0.7319</td>
<td>0.984</td>
<td>0.035</td>
<td>0.00%</td>
<td>75</td>
</tr>
</tbody>
</table>

Note: The bold text indicates that a power law could not be rejected with a p-value > 0.1. Because the power-law distribution is the null hypothesis, large p-values should indicate that there is little chance that the alternative hypothesis is acceptable.

$^a$ Standard errors are the computed bootstrap values.

$^b$ $D$ is the vertical distance between the empirical distribution and the fitted power law for the estimates $\hat{\gamma}$ and $\hat{g}_{min}$.

$^c$ Share is the percentage share of firms with growth rates higher than $\hat{g}_{min}$ with respect to the full distribution of firms.

$^d$ HGFs is the number of firms with growth rates higher than $\hat{g}_{min}$. 
<table>
<thead>
<tr>
<th>Industry (NACE rev.2)</th>
<th>$\hat{\gamma}$</th>
<th>$\sigma^a$</th>
<th>$\hat{\gamma}_{\text{min}}$</th>
<th>$\sigma_{\hat{\gamma}_{\text{min}}}^a$</th>
<th>p-value</th>
<th>$D^b$</th>
<th>Share$^c$</th>
<th>HGFs$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crop and animal prod.</td>
<td>2.978</td>
<td>0.013</td>
<td>0.853</td>
<td>0.000</td>
<td>0.000</td>
<td>0.225</td>
<td>13.88%</td>
<td>351</td>
</tr>
<tr>
<td>2. Forestry and logging</td>
<td>2.661</td>
<td>0.241</td>
<td>0.716</td>
<td>0.135</td>
<td>0.000</td>
<td>0.227</td>
<td>18.01%</td>
<td>212</td>
</tr>
<tr>
<td>16. Manufacture of wood</td>
<td>7.296</td>
<td>2.062</td>
<td>4.125</td>
<td>1.503</td>
<td>0.073</td>
<td>0.138</td>
<td>0.80%</td>
<td>8</td>
</tr>
<tr>
<td>25. Manufacture of metal</td>
<td>6.667</td>
<td>1.101</td>
<td>4.374</td>
<td>0.781</td>
<td>0.124</td>
<td>0.164</td>
<td>0.45%</td>
<td>13</td>
</tr>
<tr>
<td>41. Construction of buildings</td>
<td>7.551</td>
<td>0.250</td>
<td>3.670</td>
<td>0.000</td>
<td>0.464</td>
<td>0.119</td>
<td>0.59%</td>
<td>18</td>
</tr>
<tr>
<td>45. Specialized construction</td>
<td>5.202</td>
<td>0.899</td>
<td>3.339</td>
<td>0.434</td>
<td>0.000</td>
<td>0.127</td>
<td>0.82%</td>
<td>89</td>
</tr>
<tr>
<td>46. Wholesale and ret., motor</td>
<td>2.686</td>
<td>0.009</td>
<td>0.778</td>
<td>0.000</td>
<td>0.000</td>
<td>0.219</td>
<td>16.22%</td>
<td>554</td>
</tr>
<tr>
<td>47. Retail trade, ex motor</td>
<td>7.222</td>
<td>0.933</td>
<td>5.117</td>
<td>0.592</td>
<td>0.511</td>
<td>0.144</td>
<td>0.11%</td>
<td>9</td>
</tr>
<tr>
<td>49. Land transport</td>
<td>11.268</td>
<td>1.809</td>
<td>5.480</td>
<td>0.536</td>
<td>0.785</td>
<td>0.122</td>
<td>0.09%</td>
<td>7</td>
</tr>
<tr>
<td>56. Food services</td>
<td>6.281</td>
<td>0.385</td>
<td>3.953</td>
<td>0.173</td>
<td>0.319</td>
<td>0.132</td>
<td>0.27%</td>
<td>17</td>
</tr>
<tr>
<td>62. Computer programming</td>
<td>9.788</td>
<td>0.917</td>
<td>4.500</td>
<td>0.315</td>
<td>0.285</td>
<td>0.188</td>
<td>0.21%</td>
<td>7</td>
</tr>
<tr>
<td>68. Real estate activities</td>
<td>5.752</td>
<td>0.158</td>
<td>2.558</td>
<td>0.025</td>
<td>0.032</td>
<td>0.185</td>
<td>0.90%</td>
<td>25</td>
</tr>
<tr>
<td>69. Legal and accounting</td>
<td>10.669</td>
<td>1.481</td>
<td>4.515</td>
<td>1.438</td>
<td>0.056</td>
<td>0.284</td>
<td>0.29%</td>
<td>11</td>
</tr>
<tr>
<td>70. Management consultancy</td>
<td>4.769</td>
<td>0.566</td>
<td>3.134</td>
<td>0.216</td>
<td>0.000</td>
<td>0.195</td>
<td>0.90%</td>
<td>45</td>
</tr>
<tr>
<td>73. Advertising research</td>
<td>3.928</td>
<td>0.038</td>
<td>2.062</td>
<td>0.000</td>
<td>0.000</td>
<td>0.190</td>
<td>2.60%</td>
<td>55</td>
</tr>
<tr>
<td>74. Other professional, science</td>
<td>6.241</td>
<td>0.127</td>
<td>3.197</td>
<td>0.000</td>
<td>0.120</td>
<td>0.169</td>
<td>0.77%</td>
<td>14</td>
</tr>
<tr>
<td>81. Services to buildings</td>
<td>3.220</td>
<td>0.915</td>
<td>1.806</td>
<td>0.840</td>
<td>0.001</td>
<td>0.170</td>
<td>2.58%</td>
<td>34</td>
</tr>
<tr>
<td>85. Education</td>
<td>5.724</td>
<td>0.835</td>
<td>3.226</td>
<td>0.694</td>
<td>0.474</td>
<td>0.146</td>
<td>0.60%</td>
<td>9</td>
</tr>
<tr>
<td>86. Health activities</td>
<td>5.257</td>
<td>0.209</td>
<td>3.442</td>
<td>0.000</td>
<td>0.246</td>
<td>0.214</td>
<td>0.26%</td>
<td>8</td>
</tr>
<tr>
<td>90. Creative, arts</td>
<td>4.155</td>
<td>0.131</td>
<td>4.100</td>
<td>0.449</td>
<td>0.439</td>
<td>0.322</td>
<td>0.63%</td>
<td>3</td>
</tr>
<tr>
<td>$\sum$ All industries</td>
<td>6.900</td>
<td>1.1446</td>
<td>4.325</td>
<td>0.7982</td>
<td>0.064</td>
<td>0.064</td>
<td>0.25%</td>
<td>199</td>
</tr>
</tbody>
</table>

Note: The bold text indicates that a power law could not be rejected with a p-value > 0.1. Because the power-law distribution is the null hypothesis, large p-values should indicate that there is little chance that the alternative hypothesis is acceptable.

$^a$ Standard errors are the computed bootstrap values.

$^b$ $D$ is the vertical distance between the empirical distribution and the fitted power law for the estimates $\hat{\gamma}$ and $\hat{\gamma}_{\text{min}}$.

$^c$ Share is the percentage share of firms with growth rates higher than $\hat{\gamma}_{\text{min}}$ with respect to the full distribution of firms.

$^d$ HGFs is the number of firms with growth rates higher than $\hat{\gamma}_{\text{min}}$. 

21
Table 4. Estimation results for firm growth-rate distribution during the period 2000-2003.

<table>
<thead>
<tr>
<th>Industry (NA CE rev.2)</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\sigma}$</th>
<th>$\hat{\gamma}_{min}$</th>
<th>$\hat{\sigma}_{min}$</th>
<th>$p$-value</th>
<th>Share</th>
<th>$\Sigma HFG$s</th>
<th>$\Sigma$ firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crop and animal prod.</td>
<td>0.278</td>
<td>0.008</td>
<td>0.001</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Forestry and logging</td>
<td>2.564</td>
<td>0.318</td>
<td>0.246</td>
<td>0.488</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Manufacture of wood</td>
<td>2.399</td>
<td>0.009</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Construction of buildings</td>
<td>5.424</td>
<td>0.128</td>
<td>0.638</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Wholesale and ret., motor</td>
<td>2.397</td>
<td>0.008</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Retail trade, ex motor</td>
<td>7.046</td>
<td>0.316</td>
<td>0.246</td>
<td>0.488</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Land transport</td>
<td>6.365</td>
<td>0.156</td>
<td>0.246</td>
<td>0.488</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Food services</td>
<td>5.334</td>
<td>0.007</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Computer programming</td>
<td>6.321</td>
<td>0.076</td>
<td>0.879</td>
<td>0.879</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Real estate activities</td>
<td>5.413</td>
<td>0.007</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Management consultancy</td>
<td>5.124</td>
<td>0.044</td>
<td>0.389</td>
<td>0.389</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Advertising consulting</td>
<td>5.413</td>
<td>0.007</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Other professional, sci</td>
<td>2.433</td>
<td>0.076</td>
<td>0.879</td>
<td>0.879</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Education</td>
<td>5.413</td>
<td>0.007</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Health activities</td>
<td>3.718</td>
<td>0.006</td>
<td>0.389</td>
<td>0.389</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Total industries</td>
<td>5.709</td>
<td>0.316</td>
<td>0.246</td>
<td>0.488</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: The bold text indicates that a power-law could not be rejected with a $p$-value > 0.1. Because the power-law distribution is the null hypothesis, large $p$-values should indicate that there is little chance that the alternative hypothesis is tenable.

a Standard errors are the computed bootstrap values.

b $D$ is the vertical distance between the empirical distribution and the fitted power-law for the estimates $\hat{\gamma}$ and $\hat{\sigma}$.

c Share is the percentage share of firms with growth rates higher than $\hat{\gamma}_{min}$ with respect to the full distribution of firms.

d $\Sigma HFG$s is the number of firms with growth rates higher than $\hat{\gamma}_{min}$.

22
7 Summary and concluding remarks

The lack of theoretical considerations in the definition of so-called high-growth firms (HGFs) has lead to a variety of definitions in the literature. The common requirement is a high growth rate; but other than that, the literature offers little guidance for determining at what rate high growth is likely to occur. According to the definition adopted by the OECD (Ahmad, 2008), high growth allegedly occurs at an annualized employment growth rate of 20 percent over a 3-year period. Other definitions prescribe that HGFs comprise some percentage of all firms having the highest growth rate, regardless of the rate at which they are growing. This has lead to a rather vague notion of HGFs that involves inclusive definitions (such as the highest growing 10 percent of all firms).

To address this problem, this paper takes a broader perspective on the process of growth. By looking at the properties of growth itself, specifically at the statistical properties of the growth-rate distribution, I argue that it is possible to characterize high growth without the need for determining relevant percentage shares or required growth rates beforehand.

It is generally understood that the growth-rate distribution, with its high peak and heavy tails, describes a dynamic where most firms do not grow but where a few firms experience bursts of very high growth (Henrekson and Johansson, 2010). In addition, the growth-rate distribution has been found to be at least Laplace shaped, sometimes with tails similar to a power law (Fu et al, 2005; Schwarzkopf et al, 2010; Gabaix, 2011). In this type of distribution, the high bursts of growth experienced by a few firms are thus captured in the power-law tails. This is essentially the same high-growth episode often associated with HGFs. By combining the determination of a power law among the highest growing firms in the economy with the notion of HGFs, the challenge of identifying HGFs is reduced to estimating the right tail of the growth-rate distribution.

With a distributional definition, it is possible to test whether HGFs are present in the sample and, if they are, to estimate at what rate high growth is likely to commence. Furthermore, in contrast to the one-size-fits-all definition proposed by the OECD, a distributional definition also allows the requirements for high growth to vary with context.

Take, for example, firms in a market economy, where the prospects for high growth are likely better than for firms in economies more hostile to competition. If this is true, it should be reflected by a growth-rate distribution with thicker
tails for economies more conducive to free markets.

An additional feature of a distributional definition is the robustness of the growth-rate distribution itself, which is considered an extremely stable pattern (Dosi and Nelson, 2010). According to Smilor and Feeser (1991), the identifying patterns among successful growing firms may also have important implications for creating successful economic policy.

Using data for Swedish incorporated firms, the distributional definition was tested for the aggregate growth-rate distribution as well as for the 2-digit industry-level growth-rate distribution. Overall, the evidence suggests that a power law is sometimes present in the aggregate growth-rate distribution and is accepted for one out of three periods.

Adopting the HGF definition, the results indicate that HGFs comprise a smaller share than was previously thought, often as small as a fraction of a percent. However, for some industries, the shares of HGFs (in relation to the full samples of firms) were estimated to be close to the 1 percent definition that is often used in the literature (Daunfeldt et al, 2010).

These results are in stark contrast to the growth requirements proscribed by the OECD. In general, the present results suggest that the boundary to high growth occurs at a (log) growth rate of at least 3 to 5, which entails more than a 20-fold increase in size over a 3-year period. Results for 2-digit industries are mixed; only three industries display power-law signatures in their growth-rate distributions over the complete period.

Investigating HGFs using the growth-rate distribution approach does, however, have some limitations, some of which are discussed in the next section.

7.1 Limitations and suggestions for future research

The most obvious limitation of the approach is the simple intuition behind using a heuristic, such as the 20 percent mark used by OECD, compared to a definition based on an aggregate property in the growth-rate distribution. While a simple percentage growth rate can be used as a yardstick for firm-by-firm comparison, the testing of a power law requires a large amount of data. Future studies could very well establish such a heuristic for the growth requirements associated with the growth-rate distribution; however, the method is computationally more involved and may require constant fine-tuning as conditions continuously change. Nonetheless, the results presented here could be used to navigate betweenalready existing definitions, in that the suggested definition seems to favor the
1-percent definition over more inclusive ones. A distributional definition could also be considered complementary, in the sense that it reflects a subcategory of HGFs with even more exceptional growth rates, i.e. super-HGFs.

In addition, the present study involves the potential problem of model uncertainty, because, for some industries, the power law is estimated using only a few observations. Even after correcting for small sample bias, the certainty of some estimates can still be questioned. In the present study, the power-law hypothesis has been tested on industries at the 2-digit aggregation level. When applied to finer industry cross-sections, problems associated with small samples become increasingly difficult to navigate. Hence, the estimator used in this paper is arguably more efficient for higher levels of industry aggregation and perhaps more suited for cross-country comparisons or aggregate comparisons over time.

There is also a potential problem regarding the interpretation of statistically insignificant results. The interpretation adopted in the present study follows from Definition 1 of HGFs, where a rejection of the null hypothesis (of a power law) implies the absence of HGFs in the sample. To my knowledge, this question has yet to be discussed in the literature. While the OECD definition allows for situations where no firms meet the growth requirement, it too is very inclusive. These caveats aside, the idea of defining HGFs in terms of an aggregate property, such as the firm growth-rate distribution, introduces a new dimension to the literature. Future research may benefit from inspecting more closely, possibly with the aid of individual case studies, the characteristics and compositions of the few firms identified as HGFs. When it comes to the role played by various institutional settings, for example, such a definition may prove to be more applicable to different conditions than do individual firm growth rates, which are inherently stochastic.

Finally, and more generally, promising new research have been advanced on the relationship between the growth-rate distribution and macroeconomic fluctuations. For instance, Holly et al (2012) find evidence suggesting that asymmetries in the growth-rate distribution function as propagators of business cycles, and where the right tail of the growth-rate distribution shows particular resilience with respect to macroeconomic shocks.
References


Daunfeldt, S., Halvænson (2012b). Are high-growth firms one-hit wonders? evidence from sweden, the Swedish Retail Institute working papers, No. 73.


## A Appendix

### A.1 Tables

#### Table A.1.1. NACE rev 2. Industries

<table>
<thead>
<tr>
<th>NACE rev. 2</th>
<th>Industry description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop and animal production, hunting and related service activities</td>
</tr>
<tr>
<td>2</td>
<td>Forestry and logging</td>
</tr>
<tr>
<td>16</td>
<td>Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials</td>
</tr>
<tr>
<td>25</td>
<td>Manufacture of fabricated metal products, except machinery and equipment</td>
</tr>
<tr>
<td>41</td>
<td>Construction of buildings</td>
</tr>
<tr>
<td>43</td>
<td>Specialized construction activities</td>
</tr>
<tr>
<td>45</td>
<td>Wholesale and retail trade; repair of motor vehicles and motorcycles</td>
</tr>
<tr>
<td>46</td>
<td>Wholesale trade, except of motor vehicles and motorcycles</td>
</tr>
<tr>
<td>47</td>
<td>Retail trade, except of motor vehicles and motorcycles</td>
</tr>
<tr>
<td>49</td>
<td>Land transport and transport via pipelines</td>
</tr>
<tr>
<td>56</td>
<td>Food and beverage service activities</td>
</tr>
<tr>
<td>62</td>
<td>Computer programming, consultancy and related activities</td>
</tr>
<tr>
<td>68</td>
<td>Real estate activities</td>
</tr>
<tr>
<td>69</td>
<td>Legal and accounting activities</td>
</tr>
<tr>
<td>70</td>
<td>Activities of head offices; management consultancy activities</td>
</tr>
<tr>
<td>73</td>
<td>Advertising and market research</td>
</tr>
<tr>
<td>74</td>
<td>Other professional, scientific and technical activities</td>
</tr>
<tr>
<td>81</td>
<td>Services to buildings and landscape activities</td>
</tr>
<tr>
<td>85</td>
<td>Education</td>
</tr>
<tr>
<td>86</td>
<td>Human health activities</td>
</tr>
<tr>
<td>90</td>
<td>Creative, arts and entertainment activities</td>
</tr>
</tbody>
</table>

32