Studying Firm Growth Distributions with a Large Administrative Employment Database

Jay Dixon†  Robert Petrunia‡  Anne-Marie Rollin§

January 13, 2017

Abstract

This paper uses business administrative data from Statistics Canada’s Longitudinal Employment Analysis Program database to describe the annual firm growth rate distribution in Canada over the 2000-to-2009 period. The use of the administrative tax database is essential to study firm growth as it allows us to look at the universe of Canadian employer firms and capture the whole firm growth distribution. The results show that the distributions of employment growth rates in Canada have more density both in the centre and in the tails than a normal distribution. The evidence paints a picture of firm growth dynamics whereby a large number of firms change very little each year, and a small number markedly grow or decline. These results hold across years, industries, as well as firm size and age classes.

JEL classification: L11.

Key Words: Administrative Data; Firm Growth.

*The views stated herein are those of the authors and do not reflect the views of Innovation, Science and Economic Development Canada.
†Innovation, Science and Economic Development Canada, 235 Queen Street, Ottawa, Ontario, Canada K1A 0H5. Email: jay.dixon@canada.ca. Phone: +1(343) 291 2626. Fax: +1 (613) 952 1936
‡Department of Economics, Lakehead University, 955 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1. Email: rpetruni@lakeheadu.ca. Phone: +1 (807) 343 8384. Fax: +1 (807) 346-7936.
§Statistics Canada, 100 Tunney’s Pasture Driveway, Ottawa, Ontario, Canada K1A 0T6. Email: anne-marie.rollin@canada.ca. Phone: +1 (613) 951 3116
1 Introduction

Governments collect large and highly detailed data in the course of their operations. One such task, administering the business tax system, generates large amounts of information on firms’ activities and has nearly universal coverage. Although research is not their main purpose, these administrative business data give researchers the potential to study the behaviour of whole populations of firms. Researchers are increasingly exploiting this potential with research that often involves answering policy-related questions. The increase in policy-oriented research using administrative data is a function of: (i) the complete or partial replacement of business surveys with administrative data in order to reduce the response burden of firms; (ii) increasing openness from governments and national statistical offices in providing researchers access to business microdata; and (iii) policymakers’ interest in facilitating access to administrative databases in order to study their own policies. The use of big administrative data indeed allows researchers to assess policy impacts and help them to improve policy designs.

Wider access to administrative data has already transformed the study of firm dynamics. The main finding to emerge thus far is that firms are heterogeneous, both in terms of (observable and unobservable) characteristics and of performance.¹ For policymakers interested in encouraging productivity growth or innovation, this heterogeneity has profound implications for policy design. Understanding why some firms grow and others do not, or knowing what firms are likely to grow, and which are prone to shrink can lead to more efficient and better targeted policies to promote economic growth.

This article provides an example of using business administrative data from a specific country to build on empirical work done elsewhere. We exploit Statistics Canada’s Longitudinal Employment Analysis Program (LEAP) dataset to study firm growth distributions in Canada over the period 2000 to 2009. The LEAP database covers close to the entire population of employer firms, including both incorporated and unincorporated businesses. We describe how the study of firm growth rate distributions for all industries sheds new light on results obtained in the past in the field of business economics. This more detailed treatment, such as looking at the growth distribution for

an industry or firm type, is only possible using administrative data.

The results from LEAP show that the growth distributions of Canadian firms are essentially centered at zero, and have a non-Gaussian shape with more density at the center and in the tails of the distribution. These findings hold for various slices of the data, including: (i) yearly observations; (ii) industrial sectors; (iii) firm age categories; and (iv) firm size categories. These results demonstrate that the majority of firms change very little in terms of size, year over year. However, a nontrivial number of firms experience extremely positive growth, while another group experiences extremely negative growth.

These extreme outcomes have drawn more attention in the study of firm dynamics. Studies on high-growth firms (HGFs) find that the majority of them are young.\textsuperscript{2} Schreyer (2000) shows that high-growth firms create a substantial portion of jobs. Pakes and Ericson (1995) demonstrate that a substantial amount of research and development investment leads to firm growth. Hall (2002) points out that small firm’s R&D investment is subject to greater financial constraints, as innovation is speculative in nature and does not create tangible assets like capital investment. This theoretical framework suggests a benefit of R&D subsidies, especially when there is a high degree of uncertainty associated with innovation. At the other end, Jovanovic (1982) suggests a significant amount of job destruction is due to the rapid shrinking of young firms. Young firm are the innovators and growers, but may also require support to overcome any hurdles.

Traditional empirical research on firms has typically come in two forms. First, there are many studies of economy-wide or industry-level economic aggregates.\textsuperscript{3} However, because of firm heterogeneity, there are difficulties in mapping these studies’ conclusions back into the micro-economic behaviour of firms. At the other extreme, there are studies of specific firm behaviour from purpose-made surveys or case studies. The data for this research tend to be selective on the industries or firm characteristic. Large, already successful publicly-traded firms often receive better coverage than the smaller, younger and riskier entrepreneurial firms popularly thought to drive growth. The data often feature short time-series and can suffer a large amount of attrition of unknown origin. Selectivity, non-representation and other prob-

\textsuperscript{2}For example, see Delmar, Davidsson, and Gartner (2003), Contini and Revelli (1989), Birch, Haggerty, and Parsons (1995) and Henrekson and Johansson (2010).

\textsuperscript{3}For example, Fortin (1996) and Freedman and Macklem (1998) look at aggregate Canadian economy in the early 1990s as suggestive for activities of firms.
lems mean that the results from studies using survey data are difficult to generalize for policy purposes.

Firm-level administrative data, by contrast, are extensive and comprehensive. Business administrative data have recently shifted interest in economics research from describing the “average” or “typical” firm to describing the behavior of different subpopulations of businesses found in the tails of the distribution (start-ups, gazelles and exits to name a few).

The economy consists of many interacting groups of individuals with varied and evolving legal and real relationships between them. Administrative data comes with enough information to flexibly combine individual observations and link them across time according to the needs of individual researchers. Due to near universal coverage, administrative databases can also be used to merge information from multiple data sets, including surveys. Unlike attrition problems in surveys, changes in the administrative data’s population over time have definite interpretations. For example, entry and exit (either permanent or temporary) is not a data problem, but an important part of firm dynamics.

The nature of administrative databases means researchers find working with them difficult and challenging for reasons that go beyond the usual difficulties of working with “big data”. The source of these administrative databases, such as LEAP, are micro-level tax records, which contain sensitive firm or individual tax information. The sensitive nature of tax information leads to confidentiality requirements that go beyond the typical confidentiality requirements affecting micro-level survey data. Research is often held back in part because of the difficulty for researchers outside the government to access the relevant administrative data. As this study shows, administrative firm databases continue to add to the study of firm dynamics as governments and researchers improve access and learn better ways to maintain confidentiality when working with such databases.

---

4 Einav and Levin (2014) examine the use of administrative databases in economics.

5 A firm leaving the database may do so temporarily, returning later. Thus, distinguishing permanent exits from temporary shut-downs is often difficult.

6 Statistics Canada established the Canadian Centre for Data Development and Economic Research (CDER) in 2012 to allow secure access to business microdata for analytical purposes. Results released by researchers are vetted by an analyst from Statistics Canada in order to preserve the confidentiality of the data. The LEAP database discussed in this paper is available at CDER. For a list of all available datasets, or to learn more about data access, visit the CDER website, [http://www.statcan.gc.ca/eng/cder/index](http://www.statcan.gc.ca/eng/cder/index).
The paper is organized as follows. Section 2 discusses the LEAP database and measurement issues. Section 3 presents firm growth distributions. Finally, Section 4 concludes.

2 LEAP Database - Turning Administrative Data into a Firm Level Database

Statistics Canada’s Longitudinal Employment Analysis Program (LEAP) is an administrative database that contains payroll and employment information for all employer businesses in the Canadian economy. The source of information in LEAP are T4 “Statement of Remuneration Paid” slips. In a given year, an employer must issue a T4 slip to any employee earning $500 or more. Each T4 slip contains income paid by the firm to a worker along with any payroll deductions, such as employment insurance, Canadian Pension Plan or union dues, paid by the worker. An individual receives multiple T4 slips if he or she works at multiple firms during a year.

Canadian employers must provide T4 slips to their employees before the last day of February in the following calendar year to which the slips apply. If they fail to do so, a daily penalty of $25 can be imposed, up to a maximum of $2,500. Employees need these T4 slips to prepare their own personal tax income returns before the end of April. Since both employers and employees have a fiscal incentive to fill T4 slips, compliance is very high, which means that the coverage of LEAP is very close to the complete universe of employer firms.

The income information, taken from the T4 slips, is aggregated to create a total payroll for the firm. This aggregation to the firm using Statistics Canada’s Business Registry results in the creation of the LEAP database.\textsuperscript{7} LEAP is a longitudinal file and tracks annual firm employment back to 1983. During the construction of LEAP, significant efforts ensure the consistency of the data. Exceptional care is given to guarantee that firms tracked over time are truly the same entities.

The LEAP removes “spurious” births and deaths of firms that can arise from corporate restructuring: for example, mergers and acquisitions (M&A),

\textsuperscript{7}The administrative data in LEAP are structured at the level of the “statistical enterprise,” which is the level associated with a complete set of financial statements. For simplicity sake, this statistical unit is referred to as “the firm” in this study.
spinoffs, and divestitures. These transactions represent jobs changing firms rather than the creation of new jobs or the destruction of existing jobs. False births and deaths are removed through “labour-tracking”. Clusters of employees of appearing and disappearing firms are compared to the clusters of employees of other firms in the previous year (for appearing firms) or the following year (for disappearing firms). If a substantial portion of employees is found previously or subsequently in another firm, a connection is made between the two firms involved, and the structure of the firm in year \( t \) is applied to the data of both firms in year \( t - 1 \) and all preceding years. This sort of cleaning, which is done using flows of employees, is feasible in LEAP because the administrative T4 data simultaneously cover all employers and all employees.

Statistics Canada was a leader in the development of labour-tracking methods used to create longitudinal business databases, such as LEAP, in the 1970s (Baldwin, Dupuy, and Penner (1992)).\(^8\) Recently, Geurts (2016) uses administrative data from Belgium to show that labour-tracking removes more bias in employment dynamics measures than other traditional methods used by statistical agencies.

Construction of the LEAP database results in the formation of annual vintages. When an additional year of data becomes available, a new longitudinal file, called a “vintage”, is created. LEAP vintage 2009 tracks firm employment from 1983 to 2009; vintage 2008 tracks employment from 1983 to 2008; and so on. Firm structure changes over time due to activities such as mergers, acquisitions, spinoffs and divestitures. Each vintage holds the structure of the firm constant using information for the most recent year available. For example, LEAP vintage 2009 aggregates all earlier annual data based on the firm’s structure in 2009; vintage 2008 aggregates earlier data based on its structure in 2008.\(^9\) Thus, two firms that merge in 2009 are considered to be a single firm from 1983 to 2009 in the 2009 vintage, but are recorded as separate firms from 1983 through 2008 in the 2008 vintage.

---

\(^8\)Statistics Canada is still using the labour-tracking technique in its new longitudinal business database for research purpose called the National Account Longitudinal Microdata File (NALMF), which contains T4 data but also other administrative data (see Rollin (2014)).

\(^9\)If Firms A and B merge in 2005 to create Firm C then all 2005 and later vintages retroactively impose Firm C (merged firm) and remove Firms A and B in all years. For years prior to the merger, a synthetic history of Firm C results from the aggregation of Firms A and Bs’ activities prior to 2005.
This treatment abstracts almost completely from establishments changing hands. Firm growth is only “organic growth”, due to the addition of new establishments to the firm or new employees at existing establishments.

2.1 Measuring Firm Size and Firm Age

The LEAP database’s main advantage is near universal coverage of Canadian employer firms. However, this advantage comes with the cost that limited information is available on the firm, as is typical with other administrative databases. Administrative data are quite different from surveys which allow, by design, the collection of any desired information about the unit of interest. LEAP limits annual firm information to employment, payroll and industry. The firm growth literature points to simultaneous dependence of firm growth on both firm age and size (see Evans (1987) and Huynh and Petrunia (2010)).

Unfortunately, LEAP does not contain a direct measure of firm age or size. Firm-size can be constructed using information on firm payroll. This information allows an estimation of the firm’s employment through a constructed variable called Average Labour Units (ALUs). A firm’s ALUs equals the firm’s total payroll divided by the average annual earnings for a worker\(^{10}\) \((\bar{w}_{jkt})\) in the relevant industry (NAICS) \((j)\), province \((k)\), firm size class \((l)\) and time \((t)\) or:

\[
ALU_{it} = \frac{\text{total payroll}_{it}}{\bar{w}_{jkt}}
\]

The LEAP database also contains an indirect measure of firm age, whose construction is in the following manner. Firm age can be inferred can be inferred from earlier vintages of the data. The first year the firm ID appears in all LEAP vintage defines the firm’s birth year. Current year minus birth year becomes the measure of firm age.

2.2 Use of LEAP to Study Firm Growth Distributions

Although not designed to study firm growth, the LEAP database is flexible enough to study the distribution of annual firm growth vigilantly. Quasi-universal coverage of employer firms and careful creation of firm histories in

---

\(^{10}\)This average annual earnings comes from Statistics Canada’s Survey of Employment, Payroll and Hours (SEPH)
its various vintages provide this flexibility. Moreover, the LEAP allows researcher to study the whole business sector, as opposed to just manufacturing firms that is a feature in most studies. LEAP allows us to draw conclusions across different industries, firm size classes, and firm age classes. However, some careful decisions are needed in order to create the best analytical file to draw those growth distributions. Using, as is, one vintage of LEAP, is not fully satisfactory.

The fact that each vintage pushes the firm’s last year structure back in time risks misclassifying the sources of growth for every year except the last two. For example, the structure in the t-1 vintage for a firm that was involved in M&A activity in $t-1$ is clearly different from the structure in the $t+1$ vintage. The merged firm, as it exists in year t, cannot be held responsible for the consolidated operations in year $t-2$ of the separate firms from which it was formed.\footnote{This misclassification issue for the historical data contained in each LEAP vintage could be measured only when every single vintage of LEAP started being kept in the early 2000 (the storage cost of such a large dataset had lowered sufficiently.). The new longitudinal database created by Statistics Canada, the National Accounts Longitudinal Microdata File (NALMF), will eventually allow data to be assembled longitudinally in various ways, for different research projects (Rollin (2014)).}

Statistics Canada does keep earlier vintages of the LEAP. Therefore, this paper uses multiple vintages of the data to address the misclassification problem in earlier years of each vintage. It uses only data from the last two years of each LEAP vintage from 2000 to 2009, which results in a series of ten cross-sections, one for each vintage. The ten cross-sectional vintages are pooled to study the distribution of annual growth rates over the 2000-2009 period. For each cross-section, the calculation of year-on-year firm growth rates means the firm structure is held constant for only the last two years.

Another aspect of the data that researchers must keep in mind is that when firms are active for only part of the year, their employment is underestimated by the ALU measure. This underestimation is more important for start-ups created at the end of the calendar year, and for exits that close down in the first months of the year. Underestimation of the size of entries and exits has implications for growth rates in the year after entry, and in the year before exit. Growth rates of start-ups that survive into the next year are overestimated, and growth rates between the year before exit and the exit year are more negative than they actually were. Due to this data limitation, the analysis in Section 3 looks only at incumbent firms that are
at least two years old.\textsuperscript{12}

\subsection*{2.3 Measurement of Firm Growth}

We follow the approach of Davis, Haltiwanger, and Schuh. (1996) and Haltiwanger, Jarmin, and Miranda (2013) to measure firm employment growth. In this approach, firm size in year $t$, $X_{it}$, is the average employment size of the firm in years $t$ and $t-1$ or

$$X_{it} = \frac{ALU_{it} + ALU_{i,t-1}}{2}$$

This firm size is then used to calculate firm growth in year $t$ as “average-year” growth rate or

$$g_{it} = \frac{ALU_{it} - ALU_{i,t-1}}{X_{it}}$$

The log growth rate and this “average-year” growth rate measure produce similar values over the range from a contraction of 50 percent to an expansion of doubling from the base year size ($ALU_{i,t-1}$) of the firm. This growth rate measure has advantages. The measure falls between values of $-2$ (exit) and 2 (entry) without extreme values. The “average-year” growth rate measure limits the effects of short-term growth and regression to the mean. Regression to the mean occurs if a firm tends to a certain specific target size. A period of positive growth causes the firm to exceed its target size, which is then followed by a period of negative growth as the firm returns to its target size.

We exclude firm observations with $X_{it}$ below a value of one ALU. This exclusion of observations prevents the small-base problem. Small changes in employment can result in extremely high growth rates with small values of $X_{it}$.

\section*{3 Studying Firm Growth Distributions}

Existing research supports the proposition that firms are heterogeneous and grow for many different reasons. The degree of heterogeneity in firms’ characteristics and the seeming randomness of their performance outcomes creates

\textsuperscript{12}Excluding exiting incumbents a year before exit would require using three consecutive years of data, rather than the two year cross-section.
problems both for research and for policy. Existing research generates few lessons to guide policy-makers interested in promoting growth. Many researchers conclude that the best way to gain insight into why firms grow is to treat it as a random process, with each firm’s outcome being a realization of that process. Administrative data allows us to draw the raw distributions of firm characteristics and outcomes.

Leptokurtic, tent-shaped distributions have been recognized as a robust feature of firm-level data generally (see Bottazzi and Secchi (2003) for a discussion of firm growth distributions). Little (1962) noted unexplained but persistent (and, in his sample, growing) leptokurtosis in his data on 441 large and medium-sized U.K. firms obtained from Moodies Services, Ltd. Stanley, Amaral, Buldyrev, Havlin, Leschhorn, Maass, Salinger, and Stanley (1996) established these distributions for 2,000 publicly traded manufacturing firms using Compustat data. The nature of these distributions has also been taken up using administrative data on French and Italian manufacturing firms in Bottazzi and Secchi (2003) and Bottazzi and Secchi (2006).

Many models implicitly or explicitly assume that individual firm growth is the consequence of a variety of independent random shocks. If this assumption were correct then the central limit theorem (CLT) suggests that the resulting firm growth rate distribution is approximately normal. Instead, they resemble exponential distributions of the sort generally produced by the interaction of heterogeneous agents. A few papers, such as Coad (2006), Coad (2010) and Buldyrev, Amaral, Havlin, Leschhorn, Maass, Salinger, Stanley, and Stanley (1997), propose simple abstract models that generate these distributions. More work needs to be done on working out the causes of these regularities, as well as their policy implications.

The above-mentioned research on firm growth relies on privately collected data available on large public firms from Moodies or Compustat, or sector-specific administrative data available from Italian and French statistical agencies. The advantage of the LEAP database is that it covers all sectors of the economy and essentially all firms, irrespective of size and ownership characteristics. The comprehensive nature of the dataset allows us to precisely establish the growth distribution of all firms, but also to slice the data in a variety of ways. We are thus able to compare the distribution of firms’ outcomes across a variety of characteristics. We look at firm growth in aggregate, by years, by broad sector, and finally by size and age.
3.1 Firm Growth Distributions in LEAP

Figure 1 presents the distribution of Canadian firms’ employment growth rates over the 2000-2009 period, alongside a normal distribution with the same mean and variance. The firm growth distribution has higher central peak and fatter tails than the comparable normal distribution. Table 1 presents summary statistics corresponding to the firm growth distribution presented in figure 1. The central tendency of the distribution, as measured by the mean or the median, is close to zero. This fact, and the tight concentration of density around the centre, imply that most firms do not change their employment much year-on-year. Half of all firms (between the 25th and 75th percentiles) grow or shrink by less than 15 percent. The distribution is skewed to the left, with a skewness statistic of -0.75 and a lower mean than median growth (-0.02 versus 0.00). Kurtosis, which measures the peakedness in the centre and the tail thickness, is 4.86. This kurtosis measure is greater than the normal (3.0), but lower than the Laplace distribution (6.0). Skewness and kurtosis suggested that there is greater density in the left tail of the distribution.

The characteristics of the overall distribution have both methodological implications for researchers and practical implications for policymakers. First, a large mass of firms change very little in size each year. However, a small but significant number of firms grow rapidly and another small group of firms decline markedly. From a researcher’s point-of-view, this result means that methods aimed at parsing average effects in a representative sample is unlikely to yield significant insights. Instead, studying firm growth requires methods that focus on the outliers. For policy-makers, the results suggest that their programs may be more effective if they can identify firms and target firms with the potential for high growth, or those at particular risk of shrinking rapidly.

3.2 Yearly Firm Growth Distributions

Firm outcomes may be affected by aggregate (economy-wide) shocks. For example, positive shocks may open new opportunities for more firms to grow, but not all firms will perceive or seize them. Of those that do, many will not succeed in translating them into growth. The impact of aggregate shocks may not be noticeable from the performance average or median firm, but it may be detectable in the tails. Negative shocks, on the other hand, may cause more
firms to shrink rapidly, especially if the shocks impair access to credit markets for vulnerable firms. Other firms may not be able to exploit opportunities for want of funding. Again, the full impact of these sorts of aggregate shocks may be difficult to disentangle from observing a representative sample of firms.

The incidence and intensity of aggregate shocks vary over time. Figure 2 shows yearly firm growth distributions for the 2000-2009 period. There is very little difference in these annual distributions: they essentially overlap with each other and have the same shape with a large peak occurring close to zero and thick tails. The 2000-2009 period was generally a time of expansion for the Canadian economy. However, the period does include two global economic downturns: (i) 2001-2002; and (ii) 2007-2009 Global financial crisis. The 2001-02 downturn was relatively minor for the Canadian economy, which avoided a recession. Thus, the business cycle effects are likely small in this year. As a result of the Global financial crisis, Canada experienced a short recession which started in October 2008 with a fast and pronounced decline in GDP growth and employment. This recession ended in May 2009. The 2008 and 2009 yearly distributions are remarkably similar to the previous years. This finding suggests that aggregate shocks, even very large ones, have little impact on the spectrum of firms' growth outcomes observed on an annual basis.

3.3 Firm Growth across Sectors

Many studies on firm dynamics are limited to the manufacturing sector. Some studies do look at “high growth firms”, but they often consider them to be synonymous with the high technology sector of the economy. Administrative data allows us to look at the growth distributions by industry for the entire Canadian economy. This study is restricted to the business sector. We exclude all firms classified in the North American Industry Classification System (NAICS) sectors 91 (public administration), 61 (education services) and 62 (health care and social assistance).13

Figures 3 and 4 depict firm growth distributions conditional on their primary two digit NAICs classification. Figure 3 compares agriculture (NAICS 11), manufacturing (NAICS 31-33), and mining and oil and gas (NAICS 22).

13In Canada, the education and health sectors are heavily financed by the public sector. Firms in the excluded sectors have different motives than firms in the business sector - they are not necessarily profit maximizing.
Figure 4 depicts retail trade (NAICS 44-45), wholesale trade (NAICS 42), finance and real estate (NAICS 51 and 52), transportation, construction and utilities (NAICS 48,49, 23 and 22) and other services (NAICS 54,56,71,72 and 81).

There are differences in the firm growth distributions across sectors. In figure 3, the firm growth distribution for the agriculture sector has a substantial peak at zero, while the firm growth distribution for the mining, oil and gas extraction sector is relatively flat with more density over the -0.7 to -0.2 and 0.2 to 1 range. There is also more of a positive skew for the mining, oil and gas extraction sector firm growth distribution. Since the 2000-2009 period witnessed a significant global commodity boom, the positive skew denotes more favourable opportunities for firms in the sector generally. The greater density in the right tail of the distribution suggests implies greater opportunities for high growth in resources.

The firm growth distributions for the other five sectors are similar in shape to the overall firm growth distribution in Figure 1. A popular conception is that high-growth firms are concentrated in high-tech manufacturing industries. If true, it would imply that the distribution for that industry should have a thicker right tail than the others. In fact, the thickness of the tails, where the candidates for high-growth firms and rapidly shrinking firms reside, is similar among all industries. That is, although rapid growth and rapid decline are popularly associated with the high-tech industry, the outcomes of most of the other industries have similar distributions. This finding is in line with the high-growth literature, which reports the presence of HGFs across most industries (Henrekson and Johansson (2010)).

3.4 Firm Growth by Firm Size and Firm Age

Much of the literature on firm dynamics centers on the relationship between size and growth, sometimes accounting for firm age. The results of the present analysis suggest that the differences in conditional means and variances by firm size are driven not by the mass of firms in the centre of the distribution, but by the firms in the tails. In particular, the tails of the distribution are thicker for smaller size classes, especially for firms with fewer than 10 employees (figure 5). But the influence of size seems to be asymmetrical: the negative tail gets thicker faster than the positive tail as firm size decreases.

The growth distributions of younger firms have thicker tails than do older firms (figure 6). But the age distributions exhibit the opposite asymmetry to
the size distributions. The positive tail is especially thick for younger firms. The positive skew of these distributions suggests that, unlike their older counterparts, younger firms either tend to be high growing, or to have already exited (in which case, they are not in the data).\textsuperscript{14} The size and age results presented here suggest that small firms are particularly prone to negative outcomes, while the outcomes for younger firms are skewed positively: small firms have a greater chance of shrinking rapidly, while young firms have a greater chance of experiencing high growth.

These findings raise a number of issues. First, the non-Gaussian nature of the distributions of firm growth rates requires an explanation, but this is left for future research. Second, how much do firms in the tails of the distribution contribute to overall employment dynamics? Third, although much of the literature on firm dynamics looks at firms’ expected growth based on characteristics such as size and age, the differences between various types of firms are most pronounced in one or both tails.

4 Conclusions

Large administrative databases provide researchers the opportunity to investigate the universe of businesses. The LEAP database is one such administrative database, which contains employment and payroll information for Canadian firms taken from T4 “Statement of Renumeration Paid” slips issued by an employer to all of its employees. We use the LEAP database to show firm growth distributions do not have a standard Gaussian shape. The firm growth distribution has more density in center and tails than does a Normal distribution with the same mean and variance. This finding holds across all years of analysis, in various sectors of the economy, and across all firm size and age classes.

These results demonstrate that there is a significant number of high growth firms and rapidly shrinking firms. These types of firms are of interest to policy makers. The high growth firms are the creators of employment growth and they also often engender innovation and productivity improvements. Alternatively, the rapidly shrinking firms are a significant source of job destruction. A better understanding of the underlying factors determining high growth firms and rapidly shrinking firms enables better designed

\textsuperscript{14}Figures 5 and 6 were also done with narrower size and age classes. The same tendency was observed. The tails are heavier for smaller and younger groups.
policies to promote firm growth. However, a better understanding of the heterogeneity across firms requires further study using the near universal population available with administrative data, such as LEAP, and fully exploiting the extensive information available from the tax record sources.

Survey data are useful when examining central tendencies. Standard regression analysis focuses on these central tendencies by comparing relationships between variables at the average. Administrative firm level databases, such as LEAP, allows us to examine the whole firm growth distribution. These databases allow and require researchers to examine the full firm distribution with non-parametric or quantile regression methods, as done by Huynh, Jacho-Chávez, Petrunia, and Voia (2011).

Problems do occur when using large administrative databases. As with other “big data”, computing and software restrictions have made working with administrative data difficult. This problem is less substantial with the advent of modern computing. A more serious problem is that highly confidential firm or individual tax information provides the source of administrative data. The sensitive nature of tax information leads to confidentiality requirements that goes beyond the typical confidentiality requirements of micro-level survey data. A challenge going forward when using business administrative data is for researchers and statistical agencies to work together to allow researchers better and easier access to these administrative records, while ensuring confidentiality.
References


Table 1: Summary statistics for distribution of employment growth rates, Canada, 2000 to 2009

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>5.3 million</td>
</tr>
<tr>
<td>Mean:</td>
<td>-0.02</td>
</tr>
<tr>
<td>Standard deviation:</td>
<td>0.4</td>
</tr>
<tr>
<td>Skewness:</td>
<td>-0.75</td>
</tr>
<tr>
<td>Kurtosis:</td>
<td>4.86</td>
</tr>
<tr>
<td>10th percentile:</td>
<td>-0.43</td>
</tr>
<tr>
<td>25th percentile:</td>
<td>-0.15</td>
</tr>
<tr>
<td>Median:</td>
<td>0</td>
</tr>
<tr>
<td>75th percentile:</td>
<td>0.15</td>
</tr>
<tr>
<td>90th percentile:</td>
<td>0.36</td>
</tr>
<tr>
<td>Interquartile range:</td>
<td>0.3</td>
</tr>
<tr>
<td>Interdecile range:</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Note: All observations from 2000 to 2009 are pooled to obtain the distribution of employment growth rates.
Figure 1: Distribution of employment growth rates (Canada, 2000-2009), compared with normal distribution

![Distribution of employment growth rates](image1)

**Note:** All observations from 2000 to 2009 are pooled to obtain the distribution of employment growth rates.

**Source:** Statistics Canada, Longitudinal Employment Analysis Program, 2000 to 2009.

Figure 2: Distribution of employment growth rates, by year, Canada, 2000-2009

![Distribution of employment growth rates by year](image2)

**Source:** Statistics Canada, Longitudinal Employment Analysis Program, 2000 to 2009.
Figure 3: Distribution of employment growth rates, by industry, goods sector, Canada, 2000 to 2009

Figure 4: Distribution of employment growth rates, by industry, service sectors, Canada, 2000-2000
Figure 5: Distribution of employment growth rates, by firm size classes, Canada, 2000-2009

![Graph showing distribution of employment growth rates, by firm size classes.]

**Note:** ALUs signifies Average Labour Units.

**Source:** Statistics Canada, Longitudinal Employment Analysis Program, 2000 to 2009.

Figure 6: Distribution of employment growth rates, by firm age classes, Canada, 2000-2009

![Graph showing distribution of employment growth rates, by firm age classes.]

**Source:** Statistics Canada, Longitudinal Employment Analysis Program, 2000 to 2009.