Chasing Entrepreneurial Firms

Elsie Echeverri-Carroll
University of Texas at Austin

Maryann Feldman
University of North Carolina at Chapel Hill

May 17, 2017

Version Prepared for the 2017 Industry Studies Association Annual Meetings

Abstract: The search for a reliable dataset of entrepreneurial firms is ongoing. We analyze and assess longitudinal data on startups from two data sources—the National Establishment Time-Series (NETS) database and the Secretary of State (SOS) business registry data. Our primary purpose in this paper is to assess the usefulness and reliability of these two databases in measuring startup activity along a number of dimensions. Using the data to measure software startups, we conclude that data need to be carefully cleaned and adjusted from several biases before they become reliable. We carefully document our methodology and make suggestions for others to increase the utility of these sources. In particular, we find the NETS data enables identification of startup trends in high-tech industries at the regional level and the contribution of startups to local employment.

Acknowledgement— This study was funded by the Ewing Marion Kauffman Foundation. The authors wish to thank the Kauffman Foundation for their generous support. Evan Johnston, Jayson Varkey, and Alyse Polly, the research assistants who worked in preparing complex data for analysis as well as figures and tables, deserve special recognition. They displayed great intelligence in analyzing, organizing, and interpreting the detailed data. We thank Scott Stern and Jorge Guzman for sharing data from the North Carolina Secretary of State. Comments and suggestions appreciated.

Introduction

Defining a robust and reliable set of entrepreneurial firms has become a holy grail for researcher interested in industrial economics, regional technological change and economic growth. We observe that firms such as Apple, Microsoft, Google, and Facebook were once struggling startups that have grown to have significant impact on job creation, innovation, and productivity. Yet the existing statistical infrastructure is in many ways inadequate to investigate questions around the birth new entrepreneurial startup firms and their development over time (Goetz et al. 2015). High quality data on entrepreneurial activity are necessary for empirical research. Moreover, significant public resources are devoted to promoting entrepreneurship in an effort to create robust regional economies, with little information to help guide or evaluate these efforts.

Haltiwanger et al. (2016) note that, U.S. statistical system was historically designed to inform national economic policy, with an emphasis on collecting data for large, firms in traditional industries that produced tangible goods and employed large numbers of workers. These data are not entirely adequate to study new technologies that produce services, new business models that rely on contractual employment relationships and a new realization that the locus of economic activity has shifted from large firms to local geographies that are defined by the arbitrary location of firms. To fill this void, there are many efforts to bring new data to the study of entrepreneurship. Significant efforts are underway at the U.S. Census Bureau (Davis et al. 1996, Goetz et al. 2015, Haltiwanger et al. 2013, Kane 2010, Haltiwanger et al. 2012). In the absence of access to Census data, researchers have used the National Establishment Time Series (NETS) database (Neumark et al, 2011, Bell-Masterson and Stangler 2015) Most recently, Guzman and Stern (2016) demonstrate the utility of using of new business filings from Secretary of State (SOS) offices across the U.S. While the NETS data are proprietary, Secretary of State data are available from administrative records. Both data offer individual address information.

The purpose of this paper is to evaluate the usefulness and reliability of longitudinal databases of entrepreneurial activity. Following the work of Guzman and Stern (2015a, 2015b, 2016), we used business registration records from the Texas and North Carolina Secretaries of State to study entrepreneurial trends in Austin and the Research Triangle region. We focus on for-profit business establishments that are registered under the legal form of corporations, limited- and limited-liability partnerships, and limited liability companies. We compare the Secretary of State

data with the NETS data and find similar trends. We note discrepancies that future researchers might consider not only in identifying startups but also when considering indicators of high-growth and high quality startups. Our results show the potential to use NETS data to qualify entrepreneurs along several proxies for quality of startups at the regional level.

The next section begins with a description exiting efforts to provide longitudinal data on new firm formation. We focus specifically on replicating Stern and Guzman's use of filings with the Secretary of States (SOS) and then make comparisons with the National Establishment Time Series (NETS). Section Two considers the adjustments needed to measure startup activity longitudinally in NETS so as to be comparable to the startup trends in SOS. Section three presents an example of how NETS could be used to measure entrepreneurial quality in a region by studying high-tech entrepreneurial trends. Section four questions whether fuzzy matching of company names can be used to integrate the SOS and the NETS data, as these databases provide complementary information about startups. Section five makes suggestions for future research and summarizes the analysis on the usefulness and reliability of these two databases for measuring entrepreneurial trends.

Available Data on Entrepreneurial Firms

Guzman and Stern (2016) noted that a practical requirement for any growth-oriented entrepreneur is business registration. While it is possible to found a new business without registration, companies must register with a secretary of state as a corporation, limited- or limited-liability partnership, or limited liability company. The act of registering the firm triggers the legal creation of the company and signal an intention to move beyond the project or idea stage. Business registration records reflect the population of businesses that adopt a legal form in order to take advantage of the benefits of limiting personal liability, preferential tax status and the ability to issue and trade ownership shares. Guzman and Stern (2016) note that registration is a practical prerequisite for growth and one proxy for entrepreneurship quality. The concept of registration is key to all sources of entrepreneurial data yet there is no universal identifier, which can be linked across data sets, except for firm name.

The Census Bureau maintains a Business Registry, which is the universe of employers and is updated continuously (DeSalvo et al 2016). The Census Bureau's Business Register captures firms and establishments with paid employees. The date of formation for a new firm is

attributed to the first year of positive employment. The *Longitudinal Business Database* (LBD) is constructed from annual snapshots from the Census Bureau's Business Register, and is a confidential database available to qualified researchers through secure Federal Statistical Data Centers. The LBD relies on employment data from the Quarterly Census of Employment and Wages (QCEW), which is restricted to employers who are required to pay unemployment insurance (UI). As a result, the BDS only includes data on workers covered by either state unemployment insurance laws or by the Unemployment Compensation for Federal Employees Program (UI collectively). However, two important components of the workforce are not eligible for UI: self-employed and those firms with unpaid workers or contractual employees (Donegan 2017). Unfortunately, we do not have access to the LBD to make comparisons or evaluate trends.

The *Business Dynamics Statistics* (BDS) is the first publicly available dataset that incorporates the age of firms and therefore allows the researcher to define startups as age-zero firms (Haltiwanger et al. 2008). The BDS is based on the LBD and provides annual measures of business dynamics including the number of startups in the U.S. economy by state and MSA for 1976–2014. Researchers and policymakers can use BDS data to measure entrepreneurial trends in a metropolitan area. However, publicly available data on the Census Bureau website restricts the capacity to qualify startups. For instance, it does not allow identification of high-tech and non-high-tech startups since the website does not provide a cross-tabulation of firm age, industry sector, and MSA: firm age is reported by sector only at the two-digit SIC code. Once again, it is difficult to use these data for comparisons against the other data sources we examine.

Another source used to study entrepreneurship is the National Establishment Time Series (NETS), produced by Walls and Associates and based on a compilation and reconciliation of annual Dun and Bradstreet (D&B) establishment data (Neumark et al. 2005). D&B is a credit rating service and has a profit motive to identify and assemble information on the population of business establishments. They have adopted a massive data-collection procedure, with particular efforts devoted to identifying the birth and death of establishments. Every establishment identified is assigned a Data Universal Numbering System (DUNS), which has become a standard means of tracking businesses, and has been adopted by many government agencies in the United States and internationally. In contrast to the QCEW, D&B asks companies how many people work at an establishment, and includes workers not covered by unemployment insurance.

As a result, the NETS data seemingly includes a larger set of establishments. Moreover, the data provides the establishment address and tracks changes in location and ownership.

Neumark et al. (2005) examine the reliability of NETS employment data and find a correlation of 0.994 with the Quarterly Census of Employment and Wages (QCEW) over four-years and at the county level. To further consider founding date on individual companies, Neumark et al. (2005) find a correlation of 0.87 between start dates reported in NETS and the start dates reported by the company websites (75 percent corresponded exactly, 88 percent within one year, 92 percent within two years). Neumark et al. (2005, 2011) conclude that NETS tracks establishment births accurately, adding to the overall evidence of the reliability of the NETS data.¹ The NETS data, we should note, however is proprietary, and must be purchased by researchers.

Guzman and Stern (2015) introduce a new approach that examines for-profit business registrations from the administrative records from Secretary of States (SOS). This approach has wide appeal as every state tracks new business registrations and the data are available, although with varying procedures, and costs. One interesting wrinkle is that Delaware registration is preferred by business establishments seeking a business advantages and is required by many venture capitalists (Guzman and Stern 2015b). In a series of papers, Guzman and Stern focus on focus is on entrepreneurial quality, defined by a specific set of characteristics. The next section replicates their methodology for Austin, Texas and the Research Triangle region of North Carolina.

Figure 1. Number of For-Profit Registration Filings in Austin and the Research Triangle, 1990 to 2015

Measuring Quality (Growth-Oriented) Startup Trends with SOS Data

Following the work of Guzman and Stern (2015a, 2015b, 2016), we used business registration records from the Texas and North Carolina Secretaries of State to study entrepreneurial trends in Austin and the Research Triangle.² We focus on for-profit business establishments that are registered under the legal form of corporations, limited- and limited-liability partnerships, and

¹ However, they note that the NETS data makes use of employment data imputation, especially for young firms. As a consequence, they find it is best to consider average changes in employment over three or more years when using NETS data, as short-term employment changes do not correlate well with other employment data sources.

² Guzman and Stern (2015a, 2015b, and 2016) provide a rich and detailed overview of these data in the data appendix of their publications.

limited liability companies. It is important to note that establishments that are sole proprietorships or general partnerships are not required to file with a secretary of state but with the office of the county clerk in the county where they maintain their business premise; therefore, the SOS data does not include most sole proprietorships or general partnerships. We included businesses founded in Austin and the Research Triangle as well as those founded in the two regions but registered in Delaware or any other state. In a manner, similar to Guzman and Stern (2015a, 2015b), we restrict our sample of startups to those satisfying one of the following criteria:

- (i) A for-profit establishment registered in Texas or North Carolina with its principal office in an Austin–Round Rock MSA ZIP code³ or a Research Triangle county;
- (ii) A for-profit establishment registered in Delaware or any other state with principal office in an Austin–Round Rock MSA ZIP code or a Research Triangle county.

Our sample of startups in the Austin metro includes 223,921 and 168,648 unique filings⁴ with the Texas and North Carolina SOSs (including registrations in other states such as Delaware) between January 1, 1990, and December 31, 2015, with a primary address in an Austin ZIP code or Research Triangle county and a for-profit legal organization (e.g., corporation, limited liability partnership, limited partnership, limited liability company). As Figure 1 demonstrates the SOS data show very similar growth trends in entrepreneurial activity in both regions between 1990 and 2008; however, after the Great Recession (2008-2009), entrepreneurial activity grew rapidly in Austin while slowed down significantly in the Research Triangle region.

Figure 2. Number of For-Profit Austin Establishment
Registration Filings: 1990 to 2015

There were 8,599 Austin startups registered in Delaware from 1990 to 2015. Data on registrations obtained from the North Carolina Secretary of State did not delineate the state of registration and are omitted from the analysis of growth-oriented startups in this section. Figure 2 shows startups registered in Delaware with principal address in an Austin MSA ZIP code between

³ We use zip codes to define the Austin MSA and the Research Triangle Region. The ZIP code to MSA (or CBSA) crosswalk is taken from the U.S. Department of Housing and Urban Development (HUD: https://www.huduser.gov/portal/datasets/usps_crosswalk.html). These data may be provided upon request to the authors. The county composition of the two regions has held constant.

⁴ Each entity that registers with the SOS is assigned a unique filing number that represents the establishment. If there are multiple filings for a given entity (i.e., a legal amendment to the entity's file such as a change of legal structure, addition of an officer, name change, etc.), the same unique filing number is associated with each amendment.

1990 and 2015. Particularly striking is the rapid increase of registrations in Delaware in the period after the dot-com bust in 2000 and before the Great Recession in 2008. There is also a steep upward trend for this proxy of venture-capital-backed startups from 2010 to 2015 in line with the Kauffman Foundation recent rankings of Austin as the number-one entrepreneurial city in the United States (Morelix et al. 2015, Reedy et al. 2016).

Figure 2 also shows the filings of corporations, another proxy for quality of startups as they usually grow faster than startups organized under other legal business structures. New startups filing as corporations (outside Delaware) showed similar growth patterns than those filing in Delaware, growing rapidly between 1990 and 2000 and again after the Great Recession from 2010 to 2015. However, possibly venture capital-backed firms that registered in Delaware slowed down for a short period after the 2000 dot-com bust (2000–2002) and after the Great Recession (2008–2010), while those registered as corporations (but not in Delaware) slowed down for a long period (2001–2010).

Comparing Entrepreneurial Trends Using NETS and SOS Data

For establishments that existed before 1990, NETS records 1989 as the establishments' first year of existence. Additionally, NETS data for the previous year are reported as data for the following year. For instance, the 2014 NETS data include establishments created through the end of 2013, not those created in 2014. We restrict our sample to establishments that reported a first year between 1990 and 2013.⁵ We also need to remove not-for-profit (e.g., religious or charitable organizations, etc.) and government establishments. The legal status variable (LegalStat: G = proprietors, H = partnership, I = corporation, J = non-profit) in NETS would have potentially allowed us to identify non-profit establishments. However, this variable has 276,250 missing cases (69.7% of establishments) in the Austin-Round Rock NETS data and 241,770 (68.5%) missing in the Research Triangle NETS data. We delete the relatively few establishments whose variable legal status identifies them as non-profit establishments. We then delete establishments with NAICS codes that Choi et al. (2013) suggest identify types of businesses that are not-for-profit

⁵ Neumark et al. (2005) removed NETS observations for 1990 and 1991, as D&B drastically improved its methodology for data collection in 1992. We chose to retain these years, as they do not show a divergence from the trends seen in the SOS data.

and government entities to compensate for the large number of missing observations for the legal status variable.⁶

A preliminary definition of a startup in NETS is an establishment whose first year in this database is reported between 1990 and 2013, with a first address location in a ZIP code included in the Austin-Round Rock MSA or in a Research Triangle county, that is not a non-profit or government entity and reported itself as its own headquarters in the first year. A first examination of the raw data shows the total number of entrepreneurial firms in the NETS is significantly different from that of the SOS in both regions (Figure 3A and 3B). For example, for the Austin metro, NETS showed about 309,377 startups during the 1990–2013 period, while the SOS showed about 177,009 during the same period. Similarly, for the Research Triangle, NETS showed 281,646 startups between 1990 and 2013, while the SOS only reported 146,844 in the same time interval. While we anticipate some differences due to measurement errors and to the methods of data collection to explain the gaps of 132,368 and 134,802 startups between NETS and the SOS data in the Austin and Research Triangle regions, respectively, the unusual large gap over time between the NETS and the SOS datasets calls for an adjustment of the data to make them more compatible.

Figure 3. Comparison of SOS and NETS, 1990 to 2015: Raw numbers

This disparity between NETS and SOS startups counts in both regions may be due to several reasons. One reason is that the SOS data does not include sole proprietors: these establishments file with the county clerk, not the SOS, so we need to exclude sole proprietors from NETS in order to compare it with SOS data. Once again, the legal status variable would have potentially allowed us to identify sole proprietors and exclude them from NETS, but the large number of missing values for this variable requires that we find alternative methods of identification. Choi et al. (2013) addressed this issue by excluding establishments with only one employee in their first year in NETS, as this database counts the owner of the business as an employee.⁷ After removing the relatively few establishments whose variable legal status identifies

⁶ NAICS: 92 (government and armed forces), 8131 (religious and charitable organizations), 4821 (railroad employment), 6111 (private and public elementary and secondary school), 1141 (commercial fish and shellfish related sectors), 8141 (domestic workers), and 11 (agricultural workers on small farms).

⁷ Choi et al. (2013) also excluded establishments with two employees. However, we decided to remove only establishments with one employee. Startup trends from NETS shift significantly below the SOS startup trends when we exclude establishments with two employees. This trend seems to indicate that removing establishments with two employees will remove a large portion of LLP, LLC, LP, and corporations with one owner and one employee.

them as sole proprietors and establishments that only had one employee in their first year as an approximation of sole proprietors, the adjusted NETS data show 181,111 startups in Austin and 160,005 in the Research Triangle region for the 1990–2013 period fulfilling the following criteria:

- (i) First year is between 1990 and 2013 with a first address ZIP code included in the Austin-Round Rock MSA or the Research Triangle region;
- (ii) Establishment reported itself as its own headquarters in its first year;
- (iii) Establishment does not have a non-profit-associated NAICS code in any year or a non-profit legal status;
- (iv) Establishment has more than one employee in its first year and does not have a sole proprietor legal status.

Figures 4A and 4B show startup trends in the Austin metropolitan area and the Research Triangle region from 1990 to 2013 using raw and adjusted (excluding startups with one employee) NETS data and for-profit establishments from the SOS database. It is important to note the gap becomes much smaller in both regions, with only 4,102 and 13,161 more startups in the adjusted NETS (after correcting for sole proprietor's bias) than in the SOS data for Austin and the Research Triangle, respectively, between 1990 and 2013.

Figure 4. Austin: SOS vs NETS Raw vs NETS Adjusted, 1990 to 2013

Raw and adjusted NETS and SOS data show upward trends in entrepreneurial activity for both Austin and the Research Triangle region for the period 1990 to 2007 and a slow down during the Great Recession between 2008 and 2009. However; after this period, the SOS and NET trends differed and as already noted, the SOS data shows different entrepreneurial trends in both regions (upward entrepreneurial activity in Austin and stagnation in new startups in the Research Triangle).

Why are trends different across NETS and SOS in the last four years (2010-2013)? A possible explanation for the spike in entrepreneurial activity in 2010 in both regions that is not picked up by the SOS data is a temporal effect of the Great Recession on sole proprietors and general partnerships as it could have pushed laid-off workers to pursue entrepreneurship in record numbers.⁸ The larger number of startups in the adjusted NETS than in SOS for Austin and the

⁸ According to Don Wall's email on March 29, 2017, the spike in startup activity in 2010 is an explosion of businesses organized as sole proprietors after the 2008–2009 recession. He finds a similar trend in the U.S.

Research Triangle region could be an indication that we are not able to efficiently control for the sole proprietors' bias by removing new firms with one employee.

NETS show a drop in startup counts in the last three years of our samples (2011-2013) while SOS data show an upward count in Austin and a constant count in the Research Triangle. A possible explanation for the divergent patterns between SOS and NETS could be that most companies filing with the SOS are captured quite rapidly in the database available to researchers. In contrast, new firms, particularly those born in the last four years, take time to get into the D&B database. We started working on this project in 2016, and at that time Don Walls provided establishment (all establishments including startups) NETS data through 2012 which included 359,018 establishments in the Austin MSA and 323,288 establishments in the Research Triangle. More recently, in 2017 we received an update through 2013 which includes 396,271 establishments in the Austin MSA and 353,023 establishments in the Research Triangle. However, figure 5 shows that the distribution of the additional 37,262 establishments in Austin and 28,108 in the Research Triangle in the updated NETS data are not solely concentrated in 2013.⁹ In fact, in the case of Austin, only 12,949 (34.75%) of the new establishments started in 2013, 19,440 (52.17%) started between 2010 and 2012, and the remaining 4,873 (13.08%) started before 2010. Similarly, for the Research Triangle region, 10,537 (37.49%) of the new establishments started in 2013, 14,689 (52.26%) started between 2010 and 2012, and 2,882 (10.25%) started before 2010 in the updated NETS data for this region. In spite of these biases in the data for recent years, we conclude NETS and SOS are good sources to measure entrepreneurial activity between 1990 and 2010, particularly qualify entrepreneurial trends, such as those of high-tech firms.

Figure 5. Comparing the 2013 and 2014 NETS

High-Tech Startup Trends in Austin and the Research Triangle Using NETS

The SOS data does not allow us to study establishment births by industry. In contrast, NETS data allows us not only to study establishments' births but also to differentiate them by the type of industry to which they belong. High-tech industries pay above average wages and have

⁹ Of the 359,018 establishments in the first Austin dataset, only 359,009 were found in the updated data. This accounts for the difference of nine resulting in 37,262 new establishments between the two Austin datasets. Similarly, of the 324,915 establishments in the first Research Triangle dataset, only 324,510 were found in the updated data. This accounts for the difference of 405 establishments and results in 28,108 new establishments for the Research Triangle.

important employment multiplier effects in the local economy, but the question is how to define high-tech industries. Hecker (2005) points out there is no single definition of high-technology industries (or establishments); however, there is wide agreement on their general characteristics. In particular, he cites a report from the Office of Technology Policy (1982) describing high-technology firms as those engaged in the design, development, and introduction of new products or innovative manufacturing processes through the systematic application of scientific and technical knowledge. To classify industries by their relative innovativeness, studies have used a large variety of proxies for innovation (Chapple et al. 2004). However, in most academic studies, high-tech industries are those with a large proportion of workers in scientific, technical, or technology-oriented occupations (Hadlock et al. 1991; Hecker 1999, 2005; Luker and Lyons 1997; Markusen et al. 1986; Yu 2004; Haltiwanger et al., 2014).

Hecker (2005) defines four Standard Occupational Classification (SOC) categories as technology-oriented: engineers; life and physical scientists; computer professionals and mathematicians (except actuaries); and engineering, computer, and scientific managers. Workers in these occupations need in-depth knowledge of theories and principles of science, engineering, and mathematics (Hecker 1999, 2005). Such knowledge is generally acquired through specialized post-high school education, ranging from an associate degree to a doctorate, in some field of technology. Using data on employment by occupation, Hecker (1999, 2005) finds which NAICS categories are relatively intensive in technology-oriented workers and classifies them as high-tech. As noted by Decker et al. (2015), Hecker's definition of high-tech industries has become standard in the literature. In a recent study, Goldschlag and Miranda (2016) found that Hecker's (2005) methodology of defining the high-tech sector through the relative concentration of technology-oriented workers in an industry to be remarkably stable over time when applied to more recent 2012 and 2014 industry-occupation employment data (85% of Hecker's original high-tech industries satisfy the same criteria in 2014).

The NETS data provides the annual industrial classification of an establishment by its primary 2012 NAICS code at the six-digit level of specificity. To implement Hecker's (2005) definition of high-tech industries, we constructed a crosswalk of his 46 four-digit, 2002 high-tech NAICS codes to 198 six-digit, 2012 high-tech NAICS codes.¹⁰ We classify an establishment as

¹⁰ A complete list of high-tech NAICS codes can be provided by the authors.

high-tech based on its first reported NAICS code in NETS (variable name: NAICS##) and whether it fits Hecker's list of high-tech NAICS.

Our focus in this section is on total high-tech entrepreneurial trends and those in the top high-tech sectors in Austin and the Research Triangle. Figure 6 shows the number of new high-tech entrepreneurial establishments in both regions from 1990 to 2010. Both regions depict a steady growth of high-tech startups in the 1990s, a slowdown after the dot-com bust in 2000 and during the 2008–2009 Great Recession. These trends match what we already observed for the total entrepreneurial activity in both regions for the same period.

Figure 6. High-Tech Startup Births in Austin and the Research Triangle, 1990 to 2010

High-Tech Startups by Industry Sectors

As described by Echeverri-Carroll and Oden (2016), Austin specializes in four high-tech sectors: computer manufacturing, semiconductor manufacturing, software, and high-tech business services (e.g., architecture, engineering, management, etc.). Semiconductor and computer manufacturing are traded sectors defined mainly by large global high-tech firms that experience high international competition. Creating a new computer or semiconductor manufacturing facility is very expensive relative to a software startup. It explains why most of the high-tech entrepreneurial activity in the city is dominated by software and high-tech services. The Research Triangle region has specialized mainly in the biotechnology and pharmaceutical sector, software, and high-tech business services (Echeverri-Carroll et al., 2016; Feldman and Lowe, 2011).

We use industry-level data from NETS to study entrepreneurial trends in 3 high-tech sectors: software, high-tech business services, and biotechnology and pharmaceutical manufacturing.¹¹ Figures 7A and 7B show, total startup births in the software industry in Austin have ranged from 100 in 1990 to as many as 300 in 2001 and 310 in 2010, while software startup births in the Research Triangle ranged from 30 in 1990 to 260 in 2001 and 240 in 2010. It is also important to note the significant increase in software startups between 2009 and 2010: an increase

¹¹ The NAIC codes that define startups in the high-tech business services and biotechnology sectors is based on Osman (2015) and Echeverri-Carroll et al. (2016) while those that define the software sector come from definitions of this sector in Osman (2015), Spigel (2013), Bessen and Hunt (2007), Rosenthal and Strange (2006), and Saxenian (1994).

of 174.34 percent in Austin and 206.33 percent in the Research Triangle. Particularly striking is the significant growth of startups in the high-tech business service sector.

Figure 7. Software, High-Tech Business Services, and Biotechnology Startup Births, 1990 to 2010

The example in this section for the case of the Austin and the Research Triangle regions illustrate the potential to use the adjusted NETS data to develop different proxies for quality entrepreneurs including those in the high-tech sectors. However, integrating SOS and NETS data would expand possibilities for selecting quality entrepreneurs. SOS and NETS offer complementary benefits for measuring different dimensions of entrepreneurial activity, and it will be helpful to match the two databases based on the only common variable, the name of the new firms. For example, the SOS data offer the opportunity to get up-to-date startup counts for establishments organized as corporations, limited- or limited-liability partnerships, or limited liability companies in a geographical area. The NETS offer the opportunity to analyze these entrepreneurial firms on several longitudinal quality indicators such as employment, revenue, geographical movements, and industry by eight-digit SIC and six-digit NAICS codes. Integrating these two databases by the startup name would offer a plethora of research opportunities for quality entrepreneurs' economic activity, but we will discuss in the following section that this is not an easy task.

Matching Entrepreneurial Firm Names at SOS and NETS

The SOS and NETS databases are the two primary sources of data on new entrepreneurial firms at the national and regional levels in this study. The significance of these databases is that they have the potential to capture the population of startups rather than just a sample. To attempt to integrate the two databases we must use the only overlapping variable, the name of the entrepreneurial business, as the basis for matching registrations in the SOS to establishments in NETS. Although there are statistical packages that facilitate name-matching processes, their usefulness is limited in cases where there is a large deviation of names between the two databases. The exercise of matching names in NETS and SOS is complex because many entrepreneurial firm

names in the NETS do not exactly match startup names in the SOS database. These are some examples: Accelrted Dgital Solution Inc. (NETS) versus Accelerated Digital Solutions Inc. (SOS); Doner John & Assoc. (NETS) versus John Doner & Associates Inc. (SOS). Differences could be due to a variety of reasons such as the use of abbreviations, misspellings, registration under different names in each database, etc. An exploratory analysis of the NETS data shows that the abbreviation of names and the differences in spellings do not follow a recognizable pattern. For example, the shorthand notation of “software” as “sftwr” in the NETS record for Transportation Sftwr Specialists is not used consistently as shorthand in the rest of the NETS database, evidenced by company name records such as Software Devices, Lamonica Software, etc. Since there are not identifiable patterns behind the differences in company names in the NETS and SOS datasets, this paper uses both manual and approximated matching methods (by means of statistical software) in the name-matching exercise. The following matching exercise is conducted only for the Austin NETS and SOS datasets as the process is very labor-intensive as noted below.

Approximate matching is conducted using fuzzy matching techniques. These techniques are used to identify pairs of words that have a high probability of being the same. This paper uses the COMPGED and INDEX functions in SAS as fuzzy matching tools. We will provide some intuition for the COMPGED function here and a short introduction to the INDEX function in stage three. The COMPGED function compares two string variables¹² and returns a value (COMPGED score) based on the measure of dissimilarity¹³ between the two strings. The COMPGED function gives a high value if the strings are highly dissimilar and a low value if the strings are almost alike (e.g., an exact match will return a value of 0). The paper identifies string pairs with the highest probability of being actual matches by choosing low COMPGED scores and then manually verify whether they are valid matches. However, it should also be noted that when comparing two lists of string variables, the COMPGED function will return values for every possible pair of variables that can be formed between the two lists (e.g., if there are 100 names in list A and 200 names in list B, the COMPGED function will score all 20,000 possible pairs).

The scoring of every possible pair of names for big datasets such as NETS and SOS is computationally time-consuming and can also result in a huge number of invalid pairs with low

¹² A string variable is a variable that can contain letters, numbers, and other characters.

¹³ The measure of dissimilarity is obtained by computing the number of deletions, insertions, or swaps needed to transform one of the strings being compared into the other string (also referred to as the Levenshtein distance).

COMPGED scores. Since each pair with a low COMPGED score might not be an actual match (e.g., the pair Technologies Unlimited Inc. and Technology Unlimited Inc. has a low COMPGED score since the company names are almost alike yet represent different companies), the resulting dataset of pairs from the COMPGED function needs to be verified manually, making the process very complex. The magnitude of this problem demonstrates the need for smaller samples that make the manual process of verifying the potential matches more manageable.

Matching Startups Names in the Software Industry

To make the name-matching process manageable, we focus only on startups in the software sector. We selected the software-startup subsample as there are not obvious reasons to anticipate different challenges in the name-matching process for this group compared to those arising from matching the full set of NETS startups, other than a considerable decrease in the magnitude of the task. Section three already discusses the NAICS codes that define this industry as well as the industry's increasing importance to Austin's high-tech economy. As previously mentioned, we started working on this project in 2016 and procured NETS data from Don Walls for the 1990–2012 period. Similarly, we procured early data from the Texas Secretary of State for the period 1990–2015. We obtained more current data for both NETS and the Texas SOS in 2017.

While the previous sections in this paper present results on startup counts using the most recent NETS and SOS databases, the labor-intensive name-matching exercise discussed in this section uses data from the first NETS and SOS databases for the 1990–2010 period because most of the name-matching exercise was conducted in the fall of 2016. From the first databases, this paper uses the subset of 3,533 entrepreneurial software firms in NETS to be matched against 112,931 SOS records in Austin between 1990 and 2010. This paper performs the matching process in five stages, which will be described next.

Stage 1

A prerequisite for the matching process is cleaning the entrepreneurial firm name variable in NETS and the entity name variable in SOS of extraneous spaces, special characters (e.g., #,!,\$, etc.), and words that do not add value to the matching process (e.g., “Inc.,” “Ltd.,” “LLC,” “Corporation,” etc.). In the first matching stage, we use a SAS query command (Proc SQL) to find exact matches between the entrepreneurial firm name in NETS and the entity name in the SOS

data set. As Guzman and Stern (2015b) note, entities within the same state cannot have the same name (due to Secretary of State restrictions); therefore, any exact matches we obtain for companies located in Austin (therefore in Texas) will certainly be the same company in the two databases. We find 1,257 (36%) exact name matches, leaving us with 2,276 (64%) entrepreneurial firm names remaining to be matched.

Stage 2

In the second stage, we again utilize the SAS command Proc SQL, which allows us to select a set of entrepreneurial firm pairs fulfilling the following two requirements: the COMPGED scores are below an arbitrarily low threshold value, and the ZIP code of each new entrepreneurial firm pair is the same in NETS and SOS databases. We manually verify the quality of matching in the resulting pairs in order to identify valid matches. We then repeat the process with a higher threshold value for the COMPGED score and observe that in most cases of valid matches, the new entrepreneurial firms names differed by extraneous spaces between words, letters missing in words (e.g., Intelligent Automation Systems Inc. versus Intellgent Automtn Systems Inc., Leosoft Inc. versus Leo Soft Inc., etc.), or because the company name in one database was a shorter version of the corresponding name in the other database (e.g., Isquare-R Inc. versus Isquare Inc., Shokwave Software Inc. versus Shokwave Inc.). We identified 58 new entrepreneurial firms that are cross-listed with the same name in the two databases using this method. We are then left with 2,218 (62%) unmatched entrepreneurial firms.

Stage 3

In the third stage, we observe that of the 1,257 exact matches we obtained in the first stage, 698 (56%) entrepreneurial firms did not have matching ZIP codes in the SOS dataset. We decided to continue the matching process based only on the entrepreneurial firm name without imposing the restriction of an exact match on entrepreneurial firm ZIP code. We then repeat the stage-two process to obtain pairs with the highest probability of being valid matches and were able to find 70 additional valid matches between the two datasets. We are now left with 2,148 (60%) unmatched entrepreneurial firms.

Stage 4

Certain entrepreneurial firms in NETS have registered with SOS under a slightly different name—e.g., Zunke Network Solutions (SOS) versus Zunke Associates (NETS). Therefore, they could not be captured using the COMPGED function. In such cases, even though the entrepreneurial firm names share a portion that is exactly the same, COMPGED calculates the score for the entire string and gives high scores in these cases, which is misleading. To account for this problem, we use the INDEX function in SAS in the fourth stage of the matching process. The INDEX function finds a substring of characters (e.g., zunke) within character strings of entrepreneurial firm names (e.g., zunke associates). The INDEX function maps the substrings in one database with the strings in the other database. It searches strings from left to right, looking for the occurrence of the specified substring. When a pair of matching names occurs, it gives a non-zero value. If the substring is not found within any letter string (entrepreneurial firm name), it gives a zero value. The resulting list of pairs with non-zero values is verified manually to ascertain whether they are valid matches or not. We find 69 valid name matches between the SOS and NETS listings and are left with 2,079 (58%) entrepreneurial firms that remain unmatched at the end of stage four.

Stage 5

We ran out of possibilities to continue finding matches using software and needed to rely on manual matching. We cross-checked the remaining 2,079 entrepreneurial firms using publicly available information from the Texas Comptroller of Public Accounts¹⁴ and other websites.¹⁵ The Comptroller's website allowed us to identify entrepreneurial firms in NETS that were registered with the SOS but were missing from our sample. We used the other websites to determine whether these establishments had undergone a name change. From these processes, we were able to identify seven establishments that had changed names in the past and were therefore recorded under different names in our SOS and NETS databases (e.g., Anxdea Software in NETS data is registered

¹⁴ Because companies registered with the Texas SOS pay Franchise taxes collected by the Texas Comptroller of Public Accounts, this organization provide data on registration with the Texas Secretary of States in the following Website: <https://mycpa.cpa.state.tx.us/coa/coaSearch.do>.

¹⁵ Websites like www.manta.com , www.smallbusinessdb.com and www.buzzfile.com provides information on small businesses within an area using a name based search. Information provided includes company industry sector, address, owner's name and its aliases. We were able to identify previous names of a few companies due to helpful information on these sites. However, these sites do not contain an exhaustive list of all businesses.

as Timerewards Software in the SOS database). We are left with 2,072 (58%) unmatched establishments at the end of stage five.

We then question why we could not find matching names for 2,072 of our NETS establishments in the SOS. We look manually at the 2,072 unmatched establishment names in NETS and found the following reasons in order of importance:

- 1,413 records (68.19 percent of cross-checked records) in the NETS data set are establishments for which no corresponding SOS record could be found using both our SOS data set and the Texas Comptroller records. One possible explanation for this divergence is that many of the 1,413 NETS startups we could not find in the SOS are sole proprietors or general partnerships.
- 306 (14.76 percent of cross-checked records) in the NETS data set are establishments that were born in the Austin MSA (according to NETS) but registered with SOS under an address outside Austin. For example, Perficient had a first address within the Austin MSA and a first year of 1997 according to NETS; however, according to Texas Comptroller records, it is registered in St. Louis, Missouri, in 1999. This divergence could be explained by differences in data collection in both the SOS and NETS. The NETS firstyear variable is a proxy for a startup founding date, while the SOS data is an indication of when the startup filed with the Texas SOS as an LP, LLP, LLC, or corporation. Our sample of 1,461 matched startups in NETS and SOS shows only 50 percent with the same founding year (NETS) and filing year (SOS). However, 75.08 percent matched within ± 2 years and 80.42 percent matched within ± 3 years. In spite of the close temporal proximity between the founding year and the registration year for most of the software startups, our estimates indicate that about 15 percent of the bias is explained by these divergences.
- 284 (13.70 percent of cross-checked records) in the NETS data set are establishments that have an address in Austin and are registered with the Texas Comptroller Office under the same address; however, these establishment filings are missing from our SOS dataset. It seems that it takes some time for the records to be updated in the publically available database, so the sample researchers receive may vary depending on the time of the year they request data from the SOS.

- 36 (1.73 percent of cross-checked records) in the NETS dataset are establishments that NETS recorded as being their own headquarters in their first year; however, according to the company websites and the previously mentioned online resources, these establishments were branches of existing firms since their inception. For example, Computer Sciences Corporation (CSC), with headquarters in Falls Church, Virginia, opened a branch office in Austin in 2006. NETS captured the opening of this new establishment in Austin accurately in 2006, it but did not capture the Virginia-based headquarters. Instead, it reported that the CSC Austin location was the headquarters of the business.
- 33 records (1.59 percent of cross-checked records) in the NETS dataset are establishments that have not yet been registered with any SOS, according to the Comptroller Office website. They are most likely sole proprietors or general partnerships.

In sum, the most important explanations for unmatched names in the two databases are that it is difficult to remove all sole proprietors or general partnerships from NET (68% of cases), the founding year and address in NETS and SOS capture different points in the life of the establishment (15%), and missing cases in the SOS (14%).

Conclusions

The search for a reliable dataset of entrepreneurial firms is ongoing. The LBD is the “gold standard” of data on entrepreneurial activity at both the national and regional level. However, it is a confidential database available only to qualified researchers through secure Federal Statistical Data Centers. Researchers have in practice many restrictions to accessing this fine-grained, government-produced data. Researchers unable to easily access government-produced data are increasingly considering privately developed alternative data. We analyze and assess longitudinal data on startups from two data sources—the National Establishment Time-Series (NETS) database and the Secretary of State (SOS) business registry data.

This paper finds that entrepreneurial trends using NETS and the SOS data are similar for the period 1990 to 2010 after adjusting NETS by removing establishments organized as sole proprietors, removing government and not-for-profit entities, and proxy sole proprietors by

removing startups with one employee. These adjustments to NETS data are carefully discussed in this paper. In contrast, the data diverge for the 2010–2013 period, indicating both a spike in sole proprietors in 2010 only in NETS and significantly divergent growth dynamics for the last three years (2010–2013), when NETS drops significantly in startup counts. We find that two issues make it difficult to argue in favor of using the NETS data to measure entrepreneurial activity in the recent years. First, with the D&B data collection method there is lag in capturing new entrepreneurial establishments particularly in the last three or four years when the data become available. Second, the large number of missing values (70 percent) in the NETS variable that identifies the legal status make it difficult to remove all sole proprietors and general partnerships from NETS. Furthermore, dropping startups with only one employee as a proxy for sole-proprietors may not be a good strategy when macroeconomic conditions affect entrepreneurial activity.

The paper recognizes that integrating the SOS and the NETS using startups' names would offer many opportunities to analyze startups on a number of quality indicators. We conducted a matching process for 3,533 entrepreneurial establishments in NETs with 112,931 startup names in the SOS data. To our knowledge this is a much larger database of entrepreneurial firms than samples used in previous studies (usually databases of less than 200 startups) to test the validity of NETS to measure new firm births. We were able to match names for only 41.15 percent (1,454) of startups in NETS and SOS, using only SAS matching codes. To try to understand why we could not match 58.85 percent (2,079) of these establishments, we check them manually.

We found that the NETS data reports more startups than SOS mainly for three reasons. First, most of NETS establishments (68 percent) were probably sole proprietors or general partnerships that could not be removed from NETS because 70 percent of the values are missing from the variable that identifies the legal structure and probably the standard approximation of removing firms with one employee does not in effect remove all sole-proprietors. Second, about 15 percent of the cases are explained because NETS reports year of founding and SOS reports year of incorporation which may also capture different addresses. Third, about 14 percent of the cases are missing from the SOS which seems to provide slightly different data samples depending on the time of the year in which the data are requested.

Following Guzman and Stern, our results maintains that it is important to study quality entrepreneurs and focuses on high-tech industries associated with good-paying jobs and local employment multiplier effects. The report uses NETS and the Hecker (2005) definition of high-

tech industries to present data on human-capital intensive or high-tech startups as an indicator of entrepreneurship quality in Austin and the Research Triangle. These regional examples show the potential to use NETS data to qualify entrepreneurs along several proxies for quality of startups at the regional level, including not only the possibility of identifying high-tech startups but also indicators of high-growth startups by using employment and sales data available in NETS.

Bibliography

- Bessen, J., and Hunt, R. M. (2007) "An Empirical Look at Software Patents," *Journal of Economics and Management Strategy*, 16: 157–189.
- Chapple, K., Markusen, A., Schrock, G., Yamamoto, D., and Yu, P. (2004) "Gauging Metropolitan 'High-Tech' and 'i-Yech' Activity," *Economic Development Quarterly*, 18: 10-29.
- Choi, T., Robertson, J., and Rupasingha, A. (2013) "High-Growth Firms in Georgia," Working Paper 2013-20, Federal Reserve Bank of Atlanta.
- Davis, S., Haltiwanger, J., and Schuh, S. (1996) *Job Creation and Destruction*, Cambridge, MA: MIT Press.
- Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2015) "Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S," Working Paper 21776, National Bureau of Economic Research, Cambridge, MA.
- DeSalvo, B., F. Limehouse, and S.D. Klimek (2016) "Documenting the Business Register and Related Economic Business Data " US Census Bureau Center for Economic Studies Paper No. CES-WP- 16-17.
- Donegan, M. (2017) *Measuring Entrepreneurship: How does Database Selection Influence our Understanding of Entrepreneurial Regions?* Working Paper, University of North Carolina, Chapel Hill.
- Echeverri-Carroll, E., and Oden, M. (2016) "High-Tech Entrepreneurial Trends in the Austin MSA," Report 3, Kauffman Foundation.
- Echeverri-Carroll, E., Feldman, M., Gibson, D., Lowe, N., and Oden, M. (2016) "A Table of Two Innovative Entrepreneurial Regions: The Research Triangle and Austin," Report 1, Kauffman Foundation.
- Echeverri-Carroll, E., and Ayala, S.G. (2009) "Wage differentials and the spatial concentration of high-technology industries," *Papers in Regional Science*, 88(3):623-641.
- Feldman, M. P., and Lowe, N. (2011) "Restructuring for Resilience," *Innovations* 6:129-146.
- Feldman, M. P., and Lowe, N. (2015) "Triangulating Regional Economies: The Promise of Digital Data," *Research Policy*, 44(9): 1785-1793.
- Geffen, J. V. (2003) "Metropolitan Statistical Area Changes," *Texas Labor Market Review*, September. http://www.tracer2.com/admin/uploadedPublications/951_tlmr0309.pdf.
- Gordian, C. (2016) Personal communication, 27 September.
- Goetz, Christopher, Henry Hyatt, Erika McEntarfer, and Kristin Sandusky. (2015) "The Promise and Potential of Linked Employer-Employee Data for Entrepreneurship Research," Working Paper 21639, Cambridge, MA: National Bureau of Economic Research.
- Goldschlag, N. and Miranda, J. (2016) "Business Dynamics Statistics of High Tech Industries," U.S. Census Bureau, Center for Economic Studies.
- Guzman, J., and Stern, S. (2015a) "Where Is the Silicon Valley?" *Science* 347(6222).
- Guzman, J., and Stern, S. (2015b) "Nowcasting and Placecasting Entrepreneurial Quality and Performance," Working Paper 20954, Cambridge, MA: National Bureau of Economic Research.
- Guzman, J., and Stern, S. (2016) "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988–2014," Working Paper 22095, Cambridge, MA: National Bureau of Economic Research.
- Hadlock, P., Hecker, D., and Gannon, J. (1991) "High-Technology Employment: Another View," *Monthly Labor Review*, July, 26-30.

- Haltiwanger, J., Hurst, E., Miranda, J., and Schoar, A. (2016) *Measuring entrepreneurial businesses: Current knowledge and challenges*. Chicago Press: Chicago.
- Haltiwanger, J., Hathaway, I., Miranda, J., (2014) “Declining Business Dynamism in the U.S. High-Technology Sector,” Kansas City: Kauffman Foundation.
- Haltiwanger, J., Jarmin, J., Miranda, J. (2009a) “Jobs Created from Business Startups in the United States. Business Dynamics Statistics Briefing,” Kansas City: Kauffman Foundation.
- Haltiwanger, J., Jarmin, J., Miranda, J. (2009b) “Entrepreneurship across States. Business Dynamics Statics Briefing,” Kansas City: Kauffman Foundation.
- Haltiwanger, J., Jarmin, J., Miranda, J. (2013) “Who Creates Jobs? Small Versus Large Versus Young,” *The Review of Economics and Statistics*, May, 95(2) 347-361.
- Haltiwanger, J. (2012) “Job Creation and firm Dynamics in the United States,” In J Lerner and S Stern (Eds.) *Innovation Policy and the Economy* (12)17-38, University of Chicago Press.
- Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2008) “Business formation and dynamics by business age: Results from the new business dynamics statistics,” Center for Economic Studies, preliminary paper.
- Hecker, D. E. (1999) “High-Tech Employment: A Broader View,” *Monthly Labor Review*, June, 18-28.
- Hecker, D. E. (2005) “High-Technology Employment: A NAICS-Based Update,” *Monthly Labor Review*, 128, 57.
- Kane, T. (2010) “The Importance of Startups in Job Creation and Job Destruction. Kauffman Foundation Research Series: Firm formation and Economic Growth,” Kansas City: Kauffman Foundation.
- Labor Market and Career Information Department (LMCI) of the Texas Workforce Commission, “2000 Census vs. 2010 Census MSA Definitions.”
- Luker, W., and Lyons, D. (1997) “Employment Shifts in High-Technology Industries, 1988-1996,” *Monthly Labor Review*, June, 12-25.
- Markusen, A., Hall, P., and Glasmeier, A. K. (1986) *High Tech America: The What, How, Where, and Why of Sunrise Industries*. Boston: Allen and Unwin Publishers.
- Morelix, A., Fairlie, R., Russell, J., Reedy, E. (2015) The Kauffman Index: Startup Activity, Metropolitan Area and City Trends. The Kauffman Foundation.
- Moretti, E. (2013) *The New Geography of Jobs*. New York: Frist Mariner Books.
- Neumark, D., Wall, B., and Zhang, J. (2011) “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series,” *Review of Economics and Statistics*, February 93: 16-29.
- Neumark, D., Zhang, J., Wall, B. (2005) “Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time Series,” IZA Discussion Papers, No. 1774.
- Osman, T. (2015) “The Shadow of Silicon Valley: The Dispersion of the Information Technology Industry throughout the San Francisco Bay Area, 1990-2010,” Doctoral dissertation, University of California–Los Angeles.
- Reedy, E., Fairlie, R., Russell, J., Morelix, A. (2016). “The Kauffman Index: Startup Activity, Metropolitan Area and City Trends,” The Kauffman Foundation.
- Rosenthal, S.S., and Strange, W.C. (2006) “The Micro-Empirics of Agglomeration Economies,” In RJ Arnott and DP McMillen (Eds.) *A Companion to Urban Economics*, Blackwell Publishing Ltd, Oxford, UK.

- Saxenian, A. (1994). *Regional Advantage—Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
- Spigel, B. (2013) “The Emergence of Regional Cultures and Practices: A Comparative Study of Canadian Software Entrepreneurship,” Doctoral dissertation, University of Toronto.
- Walls and Associates (2013) “National Establishment Time-Series (NETS) Database: 2013 Database Description.”
- Yu, P. D. (2004) “Focus on High-Tech—What’s in a Name? Gauging High-Tech Activity,” *Regional Review*, Federal Reserve Bank of Boston 14:6-9.

Figure 1. Number of For-Profit Registration Filings in Austin and the Research Triangle, 1990 to 2015

Source: Texas and North Carolina Secretaries of State (SOS)

Notes: For-profit registration filings include LPs, LLPs, LLCs, and corporations

**Figure 2. Number of For-Profit Austin Establishment
Registration Filings: 1990 to 2015**

Source: Texas Secretary of State (SOS)

Notes: Delaware (DE) filings include LPs, LLPs, LLCs, and corporations

Figure 3. Comparison of SOS and NETS, 1990 to 2015: Raw numbers

Source: Texas and North Carolina Secretaries of State (SOS) and the National Establishment Time Series (NETS)

Notes: SOS startups include for-profit registrations (corporations, LPs, LLPs, and LLCs). NETS startups in the raw data include for-profit establishments that started in the Austin MSA or Research Triangle region, respectively, between 1990 and 2013 and reported themselves as their own headquarters in their first year.

Figure 4. Austin: SOS vs NETS Raw vs NETS Adjusted, 1990 to 2013

Source: Texas and North Carolina Secretaries of State (SOS) and the National Establishment Time Series (NETS)

Notes: SOS startups include for-profit registrations (corporations, LPs, LLPs, and LLCs). NETS startups in the raw data include for-profit establishments that started in the Austin MSA or Research Triangle region, respectively, between 1990 and 2013 and reported themselves as their own headquarters in their first year. NETS startups in the adjusted data additionally remove sole proprietors.

Source: National Establishment Time Series (NETS)

Notes: NETS startups in the raw data include for-profit establishments that started in the Austin MSA or Research Triangle region, respectively, between 1990 and 2013, reported themselves as their own headquarters in their first year, and are not sole proprietors. High-tech industries are those with high concentrations of technology-oriented workers. High-tech subsectors are discussed in section three.