# Measuring Growth Entrepreneurship: Business Registration, Digital Presence and the Growth Dynamics of Young Firms

Jorge Guzman, MIT Sloan School
Scott Stern, MIT Sloan School and NBER

Growth entrepreneurs – new companies that are founded with the intent of building and sustaining significant commercial enterprises – are only a small subset of entrepreneurial activity but are often believed to be critical to the role that entrepreneurship plays in economic growth. Despite the potential importance of this subset of entrepreneurs to economic performance, the ability to identify growth enterpreneurs (particularly at the earliest stages of their business) remains elusive. A significant share of prior research in entrepreneurship focuses on self-employment or self-reported entrepreneurial intentions (such as the Global Entrepreneurship Monitor). Within the domain of innovation-based entrepreneurship, the preponderance of the literature focuses on firms that achieve significant entrepreneurial milestones (such as the receipt of venture funding, or perhaps inclusion within a sample of innovation-oriented firms (e.g., Hsu, 2004). A recent literature has begun to exploit business registration and the establishment of stand-alone business enterprises as a measure of entrepreneurship (Haltiwanger, 2011; Levine, and Rubinstein, 2013), often with a focus on specific types of firms (e.g., manufacturing enterprises).

This paper offers a new approach to estimate the population of growth-oriented businesses within particular regions and over time. Our approach combines three related insights. First, we exploit the fact that incorporation is essential for the establishment of relationships that are required for growth businesses, such as raising formal capital, building certain types of business relationships, establishing certain types of employee relationships (e.g., providing health insurance), and establishing certain types of physical presences. Second, incorporation records have become increasingly available electronically, and we have been able to obtain incorporation records (sometimes going back in time for many years) from a number of US states, as well as several countries and regions throughout the world. Our preliminary assessment is that most states now make incorporation data available online, and that we will be able to exploit data from a wide range of states and countries. A significant advantage of incorporation data is that it provides (essentially) a population measure of potential growth entrepreneurs and measures a significant activity for each of these entrepreneurs at a very early stage. Finally, and perhaps most importantly, we have developed an approach that allows us to gather the "online record" of individually incorporated companies from the time of their incorporation onwards. Key elements of a firm's online presence include, among other elements, the establishment of a website, the use of email addresses (that are listed publicly on the web), presence in social network sites such as LinkedIn, Faceboook, or Twitter, and observed "news" related to the growth of a

business (such as the grant of a patent, the opening of a facility, or the receipt of venture capital).

Our empirical approach will be to combine incorporation data with online presence measures to develop metrics for the level and nature of entrepreneurship in a given region at a point in time. To be specific, our approach will allow us to measure, at the earliest stages, whether incorporating firms within a given region are also engaged in complementary activities that are indicative of growth entrepreneurship (e.g., have they established a website? Does any employee list them on LinkedIn? Do they use Monster.com to hire? Do they have a Twitter account?). Our longitudinal approach will then allow us to estimate the relationship between more traditional "outcomes" of growth entrepreneurship (such as VC funding) with these early indicators. In the ideal case, we will be able to use these "online breadcrumbs" to predict the level of growth entrepreneurs who achieve particular milestones within 5 years by examining the initial online presence indicators of incorporations in the first few years after incorporation.

Our research has already involved experimentation with preliminary algorithms that combine backward-looking web scraping with incorporation data to establish whether these data might be useful for developing a systematic index and database of growth entrepreneurs within a particular region at a point in time. For example, using a sample of 2006 Massachusetts incorporations, we find a surprisingly high rate of active online presence as of 2013. For example, more than 20% of firms incorporated in 2006 maintain a Facebook presence (as of 2013) and our initial algorithm suggests that approximately 60% continue to maintain a website. While our current algorithm may have a modest upward-bias (and we have plans for how to eliminate such biases), this simple calculation provides some insight into the ability to link incorporations and online presence. In addition to these concrete linkages between incorporated enterprises and online presence, we also plan to use textual and semantic analysis to create refined filters and indices for linking "early" online activity to the potential for growth and milestones. Importantly, the project benefits from the significant "big data" experience of one of the authors (Guzman), in his work as the Data Engineering Director on the Billion Prices Project.

This paper offers three potential contributions. First, our approach offers a systematic and scalable approach for measuring growth entrepreneurship at its earliest stages. Essentially, our argument is that incorporation is a necessary but not sufficient condition for growth entrepreneurship, and that we can employ a predictive algorithm that will allow us to filter growth entrepreneurs without having to make assumptions about founder backgrounds, firm location, or industrial focus. Second, our dynamic algorithmic approach offers the potential to offer a predictive and potentially real-time tool for estimating the number and nature of growth entrepreneurs in a given location at a point in time. This tool could potentially be used by policymakers or others to capture the intensity of growth entrepreneurship over time, and allow for the development of policies to address particular regions and environments. Finally, our approach offers direct insight into

the nature of growth entrepreneurship. By establishing the empirical relationship between early measures of online presence and traditional milestones of growth entrepreneurship, our approach provides insight into the relationship between the earliest stages of firm formation (e.g., hiring patterns, statements of firm identity, marketing efforts), and firm growth over time.