Power law distributions in entrepreneurship: Implications for theory and research ☆

G. Christopher Crawford [a,*], Herman Aguinis [b,1], Benyamin Lichtenstein [c,2], Per Davidsson [d,e,3], Bill McKelvey [f,4]

[a] Ohio University, College of Business, Athens, OH 45701, United States
[b] Indiana University, Kelley School of Business, Bloomington, IN 47405, United States
[c] University of Massachusetts-Boston, College of Management, Boston, MA 02125, United States
[d] Queensland University of Technology, Business School, Brisbane Qld 4001, Australia
[e] Jönköping International Business School, Sweden
[f] Kedge Business School, Management Department, 13288 Marseille Cedex 9, France

ARTICLE INFO

ABSTRACT

A long-held assumption in entrepreneurship research is that normal (i.e., Gaussian) distributions characterize variables of interest for both theory and practice. We challenge this assumption by examining more than 12,000 nascent, young, and hyper-growth firms. Results reveal that variables which play central roles in resource-, cognition-, action-, and environment-based entrepreneurship theories exhibit highly skewed power law distributions, where a few outliers account for a disproportionate amount of the distribution's total output. Our results call for the development of new theory to explain and predict the mechanisms that generate these distributions and the outliers therein. We offer a research agenda, including a description of non-traditional methodological approaches, to answer this call.

© 2015 Elsevier Inc. All rights reserved.

1. Executive summary

A long-held assumption in entrepreneurship research is that normal (i.e., Gaussian) distributions characterize variables of interest for both theory and practice. In other words, scores on variables such as firm resources (e.g., human capital and financial resources) and firm performance and outcomes (e.g., revenue, revenue growth) are assumed to aggregate around the mean, which is stable and

* Corresponding author. Tel.: +1 513 295 8717.
E-mail addresses: Crawforg@Ohio.edu (G.C. Crawford), HAguinis@Indiana.edu (H. Aguinis), Benyamin.bml@Gmail.com (B. Lichtenstein), Per.Davidsson@Qut.edu.au (P. Davidsson), Mckelveybill1@Gmail.com (B. McKelvey).
[1] Tel.: +1 812 856 0780.
[2] Tel.: +1 617 287 7887.
[3] Tel.: +61 7 3138 2051.
[4] Tel.: +1 310 825 7796.

meaningful, suggesting that observations can be accurately characterized by some combination of the mean and standard deviation. Our study challenges the normality assumption by examining more than 12,000 nascent, young, and hyper-growth firms. Results reveal that 48 out of 49 variables that play central roles in resource-, cognition-, action-, and environment-based entrepreneurship theories exhibit highly skewed power law distributions. In sharp contrast to normal distributions, in power law distributions the majority of observations are far to the left of the mean, a few outliers account for a disproportionate amount of the entire distribution's output and, consequently, the distribution's average is undefined and relatively meaningless in many cases. In a nutshell, results offer empirical evidence for the conclusion that *variables of interest in entrepreneurship should be assumed to follow a power law distribution unless proven otherwise.*

The discovery regarding the pervasive presence of power laws across many types of variables central to most theories in entrepreneurship suggests that more attention needs to be given to those outliers that make a disproportionate contribution. For example, 95% of all U.S. businesses are small (employing 20 people or fewer) and more than 60% of all new jobs are created by a mere .03% of all entrepreneurial start-ups. These high-influence firms drive innovation in whole sectors of the economy; they are the ones that change the competitive landscape of an industry, spur continued global innovation, and are the ones that are of most interest from a practice perspective. If entrepreneurship research continues to assume normality and focus on the mean, as the most frequently used data-analytic tools such as ordinary least squares regression and ANOVA do, it may continue to achieve statistically significant results, but the domain is unlikely to make important theoretical progress. Moreover, relying on the normality assumption, our results will likely have little value for policy makers and practitioners, who are not so much interested in a hypothetical average, but primarily in the very successful cases. Our results point to the need to examine the entire distribution of a phenomenon; in particular, to focus on extreme cases, rather than explaining them away as anomalies that must be fixed via data transformations that squeeze distributions under a Gaussian curve. We offer a research agenda that emphasizes the need to first learn whether a particular distribution is normal or not and then understand the emergence mechanisms of power law distributions. We hope the implementation of such an agenda will lead to results that will help advance entrepreneurship theory and practice in important ways.

2. Introduction

Entrepreneurship researchers, like scholars in all scientific fields, make assumptions about the phenomena under investigation. However, we should be aware of these assumptions and, perhaps more importantly, question them when necessary (Alvesson and Sandberg, 2011). One assumption that is rarely questioned, but which has significant ramifications for our view of entrepreneurship, is whether the phenomena we study follow a normal (i.e., Gaussian) curve. If the distributions of the variables we study are normal, it makes perfect sense to measure the "average" of a particular range of scores because most of them are clustered around the distribution's mean. But, what if underlying distributions are not Gaussian? If an underlying distribution deviates from normality and, instead, follows a power law (i.e., where the majority of scores are far to the left of the mean and a few outliers account for a disproportionate amount of the entire distribution's output), the mean is meaningless in many cases. Thus, one consequence of violating the assumption of normality is that results focusing on the mean may inadvertently misrepresent the nature of the phenomenon under investigation (Abbott, 1988). Moreover, not only might descriptive statistics be misleading, inferential results based on the most frequently used techniques in entrepreneurship, such as ordinary least squares (OLS) regression, structural equation modeling, hierarchical linear modeling, and meta-analysis (Dean et al., 2007) may similarly misrepresent entrepreneurship phenomena. For example, in the case of OLS regression, a regression coefficient is interpreted as the *mean* increase in an outcome given a one-point increase in a predictor. But, again, meaningful interpretation of such results relies on the validity of the normality assumption.

The normality assumption is deeply embedded in the quantitative tools of entrepreneurship and social science research in general. As noted above, the vast majority of the statistical techniques used in the domain rely on the assumption of normality as the foundation of hypothesis testing. The normality assumption also explains why outliers (i.e. cases that are more than three standard deviations from the mean) are usually seen as statistical anomalies that must be cleansed from the data (Aguinis et al., 2013; Andriani and McKelvey, 2007). However, in contrast to that view, some of the most important companies of our time—Apple, Google, Facebook, Walmart—are extreme outliers. Far from being anomalies that must be "fixed" or deleted to facilitate subsequent analysis (Aguinis and Joo, 2015), these are highly impactful companies that have major effects on all firms in the environment—these and other fast-growth companies change the competitive landscape of an industry and spur continued global innovation.

Extant research has examined whether social phenomena are more accurately described by power law distributions than normal distributions (Aguinis et al., in press; Andriani and McKelvey, 2009; Axtell, 1999; Boisot and McKelvey, 2010; Meyer et al., 2005; Zanini, 2008). Empirical analyses have discovered non-normal distributions of many phenomena, including the size of industries (Zanini, 2008) and world economies (Buldyrev et al., 2003); the individual and team actions driving technological breakthroughs (Fleming, 2007; Fleming and Sorenson, 2004); the structure of networks (Barabási et al., 2002); corporate competitive advantage (Powell, 2003); and the performance of individual workers, ranging from entertainers to politicians and researchers (Aguinis et al., in press; O'Boyle and Aguinis, 2012). Scholars have only recently approached the topic of power law distributions in entrepreneurship (c.f., Crawford and McKelvey, 2012; Crawford et al., 2014). However, there is a dearth of research on the pervasiveness of these distributions throughout the domain. Moreover, and perhaps more importantly, if distributions of key variables are indeed non-normal, how would this discovery change the way we theorize and study entrepreneurship phenomena? The present study addresses these knowledge gaps.