Journal of Business Venturing 30 (2015) 696–713

Contents lists available at ScienceDirect

Journal of Business Venturing

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

Article history:
Received 4 September 2013
Received in revised form 11 December 2014
Accepted 2 January 2015
Available online 27 January 2015

Field Editor: Dimo Dimov

Keywords:
Entrepreneurship
Generative mechanisms
Growth
Outliers
Power law distribution

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.

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

☆ We thank Field Editor, Dimo Dimov, and three anonymous *Journal of Business Venturing* reviewers, who provided us with detailed and highly constructive feedback which allowed us to substantially improve our manuscript. Also, we would like to thank the Ewing Marion Kauffman Foundation for its financial assistance with some portions of this research, and express sincere appreciation to Alicia Robb and E.J. Reedy for their support and encouragement of continued entrepreneurship research using Kauffman Firm Survey data. We gratefully acknowledge the significant financial support for CAUSEE data collection from Australian Research Council grants DP0666616 and LP0776845, the service provider BDO, and National Australia Bank.

* 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.

http://dx.doi.org/10.1016/j.jbusvent.2015.01.001
0883-9026/

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.

Our approach is to test four dozen common input and outcome variables[5] in entrepreneurship, to empirically determine whether they are better characterized by a power law (PL) or a normal distribution. As a form of replication, we select generalizable variables that are central to seminal theories used to explain and predict entrepreneurship, including resource-, cognition-, action-, and environment-based perspectives. We analyze over 12,000 entrepreneurial firms[6] across four data sets at different stages of development (i.e., nascent, young, and hyper-growing): The Panel Study of Entrepreneurial Dynamics II (PSED)—a five-year tracking of the initial conditions and behaviors leading to success or failure of nascent entrepreneurs; the Comprehensive Australian Study of Entrepreneurial Emergence (CAUSEE)—a longitudinal project following nascent and young firms over a four year period; the Kauffman Firm Survey (KFS)—a five-year longitudinal tracking of new businesses in the U.S.; and the Inc. 5000 (INC)—a dataset of hyper-growth firms that includes three-year growth outcomes. We utilize a novel nonparametric data-analytic approach developed in the field of physics for assessing the extent to which a variable's distribution conforms to a PL (i.e., Clauset et al., 2009). Our findings show that virtually no variables exhibit normal distributions: of the 49 variables we test, 48 can be more accurately characterized as PLs compared to normal distributions. Stated differently, we find strong evidence that the norm of normality in entrepreneurship research is not empirically justified.

Our manuscript makes a unique, value-added contribution because it goes beyond the empirical discovery of the pervasiveness of non-normal distributions; most importantly, we discuss how these PL findings have the potential to change the conversation in entrepreneurship research. For example, we identify theoretical and methodological implications from our study as a springboard for what Cornelissen and Durand (2013, p. 154) called “a coherent and sustainable program of research” by formulating a research agenda for future study. Next, we elucidate the assumptions of normality that permeate organization science, and then identify how these assumptions have played a similarly central role in the development of theory in entrepreneurship.

3. Theory: transitioning from Gaussian to power law distributions

Theories that assume Gaussian distributions appear to be the norm in organization science (Delbridge and Fiss, 2013; Meyer et al., 2005). In some cases, the normality assumption is actually made explicit, as in Wiklund and Shepherd's (2011, p. 927) statement that “In any sample of firms, it can reasonably be assumed that performance will vary normally around a mean.” Most often, though, the assumption is implicit. For example, Aguinis and Lawal (2012) content analyzed methodological challenges reported by authors of 75 empirical articles published in the *Journal of Business Venturing* (JBV) from January 2005 to November 2010; they found the least frequently mentioned problem (2%) was violation of assumptions such as normality. However, none of the authors of those JBV empirical articles reported tests to assess the shape of the underlying distribution. Recognizing that normality permeates most aspects of entrepreneurship research, we begin with a critical analysis of this assumption, arguing instead that it may not fit with the reality of modern social systems in general, and entrepreneurship in particular.

3.1. Gaussian distributions, data adjustments, and the intractable “outlier problem”

A key characteristic of a normal distribution is that data aggregate around the mean, which is stable and meaningful, suggesting that observations can be accurately characterized by some combination of the mean and standard deviation (Greene, 2011). As such, within any given sample most observations are bunched around the mean, whereas only a small number of cases are far away from the average. As mentioned earlier, the simplicity of this arrangement allows for the use popular data analytic techniques based on the general linear model (e.g., Greene, 2011).

But, what happens when non-normality is observed empirically? For researchers in entrepreneurship and other fields, the most common “solution” for addressing the “problem” that the data do not fit a Gaussian worldview is to make data adjustments, with techniques such as mathematically transforming skewed distributions or simply dropping the outliers in order to make the sample better reflect the “true” underlying normal curve (Aguinis and Joo, 2015; Aguinis et al., 2013; Greene, 2011). As illustrated in Godfrey et al. (2009), “we ran regression diagnostics to look for outliers and removed seven observations that substantially skewed regression results, consistent with normal practice;” similar practices are common in entrepreneurship (e.g. van Stel et al., 2007).

Although these adjustments often improve the likelihood of gaining statistical significance, they may unwittingly reduce the validity and the accuracy of the conclusions, for these techniques mask an ontological reality that may not conform to normality. For example, removing the statistical outliers in a distribution of businesses does not reduce the influence of those outliers in practice. Thus, findings are likely to lack internal and external validity and, therefore, be of little value to inform either theory or practice. As evidence, consider the impact of a new Walmart on existing mom-and-pop retailers in small town, or the way that Amazon.com has led to the closing of so many independent local book stores. In both cases, squeezing the data into a normal curve by eliminating the outliers—Walmart or Amazon—is equivalent to removing the most important drivers of the system. In our manuscript, we propose *changing our conceptualization* of the distribution of key variables from normal to non-normal (if such change is warranted based on empirical observations), rather than *changing our data* to fit our existing, and possibly incorrect, conceptualization.

[5] We use the general term “input variables” to encompass items that precede and purportedly explain the outcomes of interest in entrepreneurship; these variables have also been called “antecedent,” “determinant,” “explanatory,” “explanantia,” “independent,” and “predictor” variables in the literature. Similarly, we use “outcome variables” as a general term to encompass “consequent,” “criterion,” “dependent,” and “explananda” variables.

[6] The understanding of “entrepreneurial” varies in the literature (cf. Gartner, 1990; Mitchell, 2011). Our conceptual and empirical use of the term captures a broad variety of criteria such as “new entrant” and “founder-managed” as well as “high growth,” which is likely to reflect underlying “innovation” along some dimension.

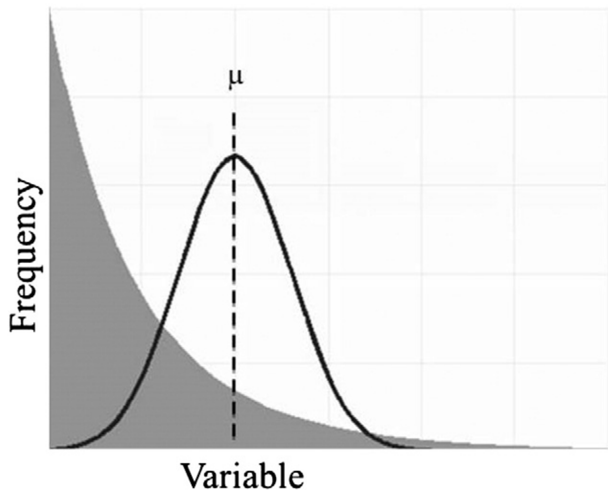

Fig. 1. A normal distribution (black) overlaying a power law distribution (gray) (modified from O'Boyle and Aguinis, 2012).

3.2. An alternative: power law distributions

Herbert Simon's (1955, 1968) seminal work provides guidance for disciplines where skewed distributions and outliers disproportionately influence theory building and theory testing. He proposed an inductive method, where variables of interest are first pursued through empirical investigation and, subsequently, theories are formulated as attempts to explain the "stylized facts" that emerge. Simon's writing in 1968 was an attempt to guide theory-building efforts and explain some "striking empirical regularities [in data, where] standard statistical tests of hypotheses [are] inappropriate" (p. 443). Simon was referring to the ubiquity of PL distributions, the topic of his classic article, "On a class of skew distribution functions" (1955). In entrepreneurial terms, the outliers are the high-impact firms (Acs, 2008; Bhide, 2000)—new ventures with radically new technology or novel business models that often transform entire industries. Because these firms have the potential to drive outcomes regionally and nationally, they are a very important part of any sample of entrepreneurial firms, and thus, cannot simply be excluded from the scope of our theorizing and dropped from our empirical analysis. Though "normal practice" may advocate removing outliers (Aguinis et al., 2013), PL distributions explicitly assume the existence and influence of observations that are many orders of magnitude greater than the average. In fact, in PL distributions these extreme observations are expected.

Fig. 1 includes a normal (i.e., Gaussian) curve and an alternative generic PL distribution, illustrating key differences between them. The black line of the familiar bell-shaped normal distribution suggests that most of the data are clustered around a mean (μ) and then fan out into disappearing, symmetrical tails, where the probability of a positive or negative extreme event is near zero. In contrast, the gray PL identifies that the majority (the highest frequency) of outcomes (e.g., company employees) appear at the top of the Y-axis, whereas a minority (i.e., the extreme outcomes) are out at the end of the long-tailed X-axis, and the largest company (the extreme outcome) is out at the end of the long-tailed X-axis. A PL distribution has unstable means, near-infinite variance, and a greater proportion of extreme events (Sornette, 2006). Here "unstable" suggests that the mean can change significantly if a more "extreme sized" observation is added or deleted from the sample; and, since it does not characterize these more "extreme" entities, the mean becomes meaningless and even misleading.

Consider the following example, especially in terms of its theoretical implications. Using 2008 United States data, the average number of employees in all firms is $\mu = 4$ and, if we only count the firms that report at least one employee, it is $\mu = 20$. However, these mean values have little relevance in relation to the 22 million mom-and-pop firms with zero employees and no intention to hire any[7]; nor do they have any relevance to Walmart, with 2.2 million employees.[8] At best, this "average" is useful for describing only a small portion of companies; moreover, computing correlations and other statistics based on the average and variance (i.e., squared average deviation from the mean) will not lead to meaningful inferences for a very large percentage of firms.

To define a PL more formally, the probability of an event sized x is inversely proportional to its size raised to some exponent, as $1/x^{-\alpha}$ (Clauset et al., 2009). Because α is expressed as an exponent, as α decreases to one, the tail of the distribution becomes longer (i.e., heavier) and a greater proportion of the entire distribution lies in the most extreme values (i.e., a larger proportion of extreme values). In statistical terms, we use the term power law to refer to those heavy-tailed distributions where a disproportionate amount of the total outcomes is captured by a small group of extreme scores, while most observations (raw count) are far to the left of the mean. Also, although other heavy-tailed distributions exist (e.g., exponential or log-normal), we focus on PLs because of the cross-disciplinary findings on the phenomena and because of the interesting, underlying processes that are purported to generate these power laws (Andriani and McKelvey, 2009; Clauset et al., 2009).

7 https://www.census.gov/econ/smallbus.html, accessed Aug 7, 2014; 121 M paid employees/27.8 M firms.

8 2012 Form 10-K, Wal-Mart Stores, Inc.

So far, we have used the phrase "input and outcome variables" because it is the nature of theory to suggest that a particular antecedent "causes" a certain consequence (Bacharach, 1989; Shadish et al., 2002). The next main section provides a rationale for the variables we included in our study. We first identify and describe input variables from four general theoretical frameworks used in entrepreneurship research at multiple levels of analysis. These are the role of individual and team-level resources, the qualities of founder cognition, aspects of entrepreneurial action and agency, and the influence of the environment. Our analysis offers a rationale for pursuing a formal examination of the shape of key variables in each of these theories. Second, we address key outcome variables that are generalizable and relevant to the entire domain of entrepreneurship.

3.3. Theoretically relevant variables: resources, cognitions, action, and environments

3.3.1. Resource-based theories: human, social, and financial capital

Many theories of entrepreneurship note that the success of new ventures depends on the resource endowments of the firm, including human capital, social capital, and financial capital. Overall, greater amounts of these types of capital—the farther beyond the "average" compared to other firms—increase the probability that a venture will start-up and survive. However, there have been studies providing evidence that resource-based variables have negative or null effect on outcomes (cf. Kim et al., 2006; Parker, 2009). Indeed, our analysis posits that the significant criticisms regarding the impracticality of the resource-based view (see Arend and Levesque, 2010) stem from the field's definition of competitive advantage: "profits relative to the *average* competitor in an industry" (Peteraf and Barney, 2003, italics added). If the mean is not as meaningful as it is believed to be, due to the presence of non-normal distributions, then assessment of resource-based advantages would be severely biased. Based on work conducted mainly outside of the field of entrepreneurship, we posit that all three types of capital (resources) are likely to be PL distributed.

Human capital describes the cumulative amount of individuals' knowledge, education, and professional skill. It reflects the capacity to make skillful decisions, act effectively, and interact with others through school, industry, and professional networks (Allen et al., 2007; Shane and Stuart, 2002). A study by O'Boyle and Aguinis (2012) examined more than 600,000 individuals; their results show that many aspects of human performance that require knowledge, education, skills, and experience are PL distributed, suggesting that there is a minority of individuals who perform at a level that is many orders of magnitude above the rest—thereby providing the human capital foundation for competitive advantages in the market.

Social capital is comprised of individuals' networks and the assets that may be mobilized through them (Nahapiet and Ghoshal, 1998); these networks of relationships are a resource for social action (Burt, 1997; Kreiser et al., 2013), and entrepreneurial success (Aldrich and Kim, 2007). Theoretically, the broader the social networks of the entrepreneur and her team, the more likely they will create or discover an opportunity. However, extant research suggests that networks are PL distributed (e.g., Barabási, 2009), where most nodes have very few connections but a few outliers are connected to a very large percentage of the entire population.

Finally, *financial capital* has been shown to be PL distributed in terms of individual wealth (Heinsalu and Patriarca, 2014; Pareto, 1897) and investors' wealth (Solomon and Richmond, 2001). This distribution of financial variables seems to hold across both time and space (Abul-Magd, 2002; Ning and You-Gui, 2007). A PL distribution for financial capital would mean that those entrepreneurs with financial capital at the extreme end of the long tail possess a valuable and scarce resource, which allows them to purchase, hire, or partner with alternative sources of capital without putting in the requisite activities necessary to acquire them organically.

3.3.2. Cognition-based theory: expectations founded on previous experience

A cognitive perspective of entrepreneurship is often used to understand why an entrepreneur choses one decision over another (Baron, 2004). Following Chiles et al.'s research on Radical Austrian Economics (2007, 2010a, 2010b), we argue that cognition is endogenously influenced by previous experience and expectations for future outcomes. This idea is related to Bandura's (1989) concept of self-efficacy, which describes one's confidence in achieving a specific task, and which has been studied extensively in entrepreneurship (Chen et al., 1998), new venture performance (Hmieleski and Corbett, 2008), and expectations for future outcomes (Cassar, 2014; Gatewood et al., 2002). Because the most important source of confidence is experience (Bandura, 1989) which, as we mentioned in the previous section is likely PL distributed, then the "average" amount of experience or self-efficacy on an entrepreneurial team may not accurately explain action or expectations, but may follow a PL distribution.

3.3.3. Action-based theories: behaviors and path-dependence

Entrepreneurship requires action (McMullen and Shepherd, 2006), and extant research on start-up activities suggests that new ventures are more likely to emerge (i.e., officially begin operations) when nascent entrepreneurs conduct organizing activities at high rates, spread out over time, and such activities occur later in the start-up process (Hopp and Sondereggger, 2014; Lichtenstein et al., 2007). Research in other fields has shown that many types of human activity (e.g., distance traveled from home, duration of travel), and also the number of personal interactions, are PL distributed (González et al., 2008; Gulati et al., 2012; Song et al, 2010). Based on the work by Lichtenstein et al. (2007) and the general PL patterns of human activities, we suspect the same may be true of entrepreneurial action in both number of activities conducted and the total time working on the venture. As Arthur (1988) explained, human activity becomes self-reinforcing and recursive, so the more feedback an entrepreneur receives—either from potential stakeholders or her own experimentation—the more likely similar behaviors will be repeated, which could potentially turn the distribution of activities non-normal. In fact, these differences in activity could lead to path-dependence, which has been suggested as a source of firm variation within both discovery and creation theories of entrepreneurial action (Alvarez and Barney, 2007).

3.3.4. Environment-based theories

Entrepreneurship scholars have long studied the effect of the environment on new venture creation, in terms of variations in the number of new ventures founded over time (Aldrich, 1990), as well as the relative munificence of different environments (Edelman and Yli-Renko, 2010). Munificence may be generated by increased availability of capital and other resources, as well as through changing consumer preferences that generates new opportunities for the entrepreneur and for existing businesses (McMullen et al., 2007; Plummer et al., 2007). Previous research has modeled environments (e.g., industry sectors) as being Gaussian (e.g., Ganco and Agarwal, 2009). However, if the distribution of firms within industries follows a power law, the influence of outlier firms on start-ups would be disproportionately high. For example, Farjoun and Levin (2011) examined this possibility by modeling industry dynamism through the use of fractals, a complexity science technique that reveals the power law nature of these contexts. Fractal mathematics show how activity at the most micro-level of a system aggregates to higher-order activity; due to self-similarity, these dynamics are mostly likely to be PL distributed (Gell-Mann, 1988; Mandelbrot, 2009). Extant research has shown that most environments—like global economies, geographic industrial clusters, and the size of all registered U.S. firms—exhibit PL distributions (Andriani and McKelvey, 2009; Flier et al., 2003; Zanini, 2008).

3.4. Outcomes in entrepreneurship

Consistent with our inclusion of input variables that can apply to all ventures regardless of size, we are interested in generalizable outcome measures that can be applicable to ventures at all stages of emergence and early development. Entrepreneurial outcomes discussed in the literature include, but are not limited to, revenue, employees, growth, profit, economic well-being, survival, market-share, amount of venture capital funding, IPO (under)pricing, wealth creation, and other or combined indicators of “performance” or “success” (Davidsson, 2004; Delmar et al., 2003; Garnsey, et al., 2006; Leitch, et al., 2010; Shepherd and Wiklund, 2009; Steffens, et al., 2009; Van de Ven and Engleman, 2004). Some of these indicators are dichotomous (e.g., survival); subjective and/or hard to measure (well-being; value creation); known to apply only to a small share of the business population (VC funding; IPOs) or cannot be meaningfully compared across industries (asset value; market share). Albeit general and highly relevant, profitability is difficult to assess reliably and also less suitable at very early stages. Hence we do not include profit indicators in our analysis.

Instead, we investigate growth in revenue and employees. These are the most commonly used growth/size indicators used in the literature (Delmar, 1997) and there is accumulating consensus that they are the most generally applicable as well as the most theoretically and practically relevant ones in cross-industry studies (Davidsson et al., 2010). All firms need people to produce outputs and sales to be sustainable; thanks to the longitudinal nature of our early-stage samples, both employees and revenue are relevant to (and measured by) every firm in our analyses. Further, these measures are neutral to whether the growth is organic or via acquisition (Lockett et al., 2011), and they complement each other, as sales and employment growth do not always correlate highly (Chandler et al., 2009; Shepherd and Wiklund, 2009). To avoid biasing influence of initial firm size, we employ both relative (percentage) and absolute measures of sales and employment growth (cf. Delmar et al., 2003).

4. Method

Our investigation leverages four separate, yet complementary, datasets of entrepreneurial input and outcome variables at different stages of venture emergence: two representative longitudinal samples from different countries that begin at the nascent stage, one longitudinal study of newly formed firms, and one cross-sectional sample of hyper-growth firms. Next we identify the characteristics of the individuals and ventures providing data to ensure that our analysis is consist with the level(s) of analysis of each study, and consistent with the primary unit(s) of analysis and boundary conditions of the theories from which we draw.

4.1. Databases: PSED II, CAUSEE, KFS, and INC

4.1.1. PSED II

The second Panel Study of Entrepreneurial Dynamics (PSED II) was a replication and extension of the PSED I from 1998–2001. The two data sets and their international counterpart studies have generated more than 100 peer-reviewed journal articles as well as multiple dissertations, books, and book chapters (Davidsson and Gordon, 2012; Frid, 2013; Gartner et al., 2004). The PSED II's baseline data started in 2005 as a representative sample of 31,845 adults in the contiguous United States, contacted via random digit dialing. Using a well-developed protocol of screening questions, a team of interviewers identified 1214 subjects as nascent entrepreneurs, meaning they were engaged in the process of founding a new venture, but had not yet achieved full-fledged start-up status, nor positive income for six months or more (Reynolds, 2007). We analyzed all resource- and cognition-based variables at the initial point of data collection, Wave A, and all action-based inputs at Wave B for both the founder and the founding team. Similarly, we analyzed reported venture revenue and employee outcome variables at Wave A (Year 0) and Wave D (Year 3) to remain consistent with the INC 5000 data, described below.

4.1.2. CAUSEE

The Comprehensive Australian Study of Entrepreneurial Emergence offers a longitudinal survey of nascent and young firms on a yearly basis over a four-year period. Using essentially the same sampling approach as PSED II, CAUSEE used more than 30,000 screening interviews to identify samples of 625 nascent firms and 559 young firms. This data set complements our study because it permits a more culturally generalizable assessment of our findings for both PSED II and KFS data. From the CAUSEE data set, we analyzed the founder's expected number of employees and total revenue after 12 months of operations from Wave 1 (the initial collection of data), as well as the venture's actual number of employees and revenue at Wave 1 (Year 0) and Wave 4 (Year 3).

4.1.3. KFS

The Kauffman Firm Survey (KFS) started with a random sample of 32,469 U.S. businesses from a Dun & Bradstreet list which identified almost 250,000 firms that started operations in 2004. A start-up is defined as any independent business that was established by a single person or a team, or purchased as an existing business or new franchise. Businesses are excluded if they have a federal identification number, income on Schedule C, or paid federal Social Security or state unemployment insurance or taxes prior to or after 2004. Using a stratified sampling methodology that is weighted toward high-technology firms, the sample includes 2034 high-technology firms and 2894 non-high-tech businesses (DesRoches et al., 2009). We perform analyses within the National Opinion Research Center (NORC) enclave, a data repository housing the restricted-access KFS micro-data that provides a higher level of refinement (e.g., continuous variables instead of binned or artificially polycotomized scores, as identified in Aguinis et al., 2009) compared to the data available to the general public. From this data set, we analyzed revenue and employee variables at the venture level from 'Year 0' (the initial collection of data) and Year 3 (the fourth year of data collection).

4.1.4. INC 5000

Every year, *Inc. Magazine* collects revenue, employee, and three-year growth rate data from ongoing businesses at the extreme high end of performance, and then publishes small case vignettes—in both print and online—on the fastest-growing, privately held, for-profit companies in the United States (Markman and Gartner, 2002). The 500 companies with the highest revenue growth rate are written up in *Inc.*, and the publisher's website provides information on the top 5000 companies. In contrast to the PSED II, CAUSEE, and KFS data sets, INC 5000 is a self-selected, pay-to-enter sample. However, given the fact that the fee is only $100, and the Inc. website generates 25 million views for the companies on the Inc. 5000 list (Inc. Magazine, 2011), the cost is a nominal marketing expense for such guaranteed global publicity. Data certification and internal accounting checks insure high degrees of reliability and validity in the data (Shadish et al., 2002). The sample is more skewed toward technology firms than the three other samples, suggesting INC could be considered an oversampling of the highest performing firms in the KFS. From this data set, we analyzed the 3-year firm revenue growth rate percentage published online; and, with the help of a research assistant, constructed a data set of Year 0 and Year 3 revenue, employee, and relative growth figures from links on the Inc. website to almost all 5000 firms. Within our constructed data set, we sorted revenue figures into self-reported industry sectors, which are reported as environmental-level variables (industry sectors) in Table 1.

By using these four independently created databases, we offer some assurance that our findings are not the idiosyncratic outcomes of how and why a particular database was created. The following sections describe the input and outcome variables, respectively, and identify the datasets from which they were collected.

4.2. Input variables

4.2.1. Resource-based variables

We identify five different human capital resources in the PSED II. *Employees Supervised* is the total number of individuals who reported to the respondent in her previous position. *Number of Owners* is the answer to question "How many total people or businesses or financial institutions will share ownership of the business?" For *Team Education* we re-scaled the education level of each owner as described in Appendix 1; the sum for all owners is *Team Education*. For *Team Industry Experience* and *Previous Ventures Founded*, we sum the responses for all owners.

Social capital relationships include strong and weak tie networks (Granovetter, 1983). We assess *Weak Ties* with the response to "How many other people, who will not have an ownership share, have made a distinctive contribution to the founding of this new business, such as planning, development, financial resources, materials, training, or business services?" *Strong Ties* is measured by response to "How many other people, who will not have an ownership share, have provided significant support, advice, or guidance on a regular basis to this new business?" Other entrepreneurship scholars have used this question to conceptually represent social capital (Bruderl and Preisendörfer, 1998).

Financial capital variables are assessed at the individual and team/venture level. We use the survey's summation of a founder's assets and debts to measure *Individual Net Worth*, and the summation of all source funds invested in the company for *Individual Investment*. Similarly, *Team Funding* is the summation of all funds invested for up to five owners; we did not calculate any additional funds for more than five. *Venture Debt* is the total amount of all debts to all owners of the company, as recorded in the PSED II in Wave A.

4.2.2. Cognition-based variables

We analyze expectations about the amount of revenue and the number of future employees in the PSED II and in CAUSEE to investigate whether the distributions of long- and short-term expectations are PL distributed. For *PSED II Expected Yr5 Employees*, in Wave A, the founder was asked, "During the fifth year of operation, how many managers or employees, including exclusive subcontractors, will be working for this (new) business, not counting owners?" In *CAUSEE*, Wave 1, *Expected 12 m (month) Employees* asks for number of employees expected at the end of the following 12 months. Similar questions were posed for revenue. *PSED II Expected Yr5 Revenue* asks "What annual revenue is expected when the business is in its fifth year of operation?" Similar questions were asked for *CAUSEE Expected 12 m Revenue*. Less than 7% of firms responded with "No response" or "I don't know;" these were not included in our analysis.

Table 1
Descriptors of distributions of input and outcome variables used in theories of entrepreneurship.

	n	Mean	Med	Skew	sd	Min	Max	α	K-S
Input variables									
Resources: human capital									
Employees supervised[a]	1179	20	4	12	83	0	1500	2.12	0.07
Number of owners	1213	2	1	8	1	1	30	3.50	0.04
Owner(s) Education	1214	25	18	7	21	8	377	4.38	0.05
Team industry experience	1214	14	8	2	16	0	149	5.50	0.09
Previous ventures founded	1214	2	1	4	3	0	30	3.06	0.04
Resources: social capital									
Strong ties	1214	1	0	10	2	0	50	2.61	0.02
Weak ties	1190	2	1	5	2	0	25	2.86	0.05
Resources: financial capital ($000)									
Individual net worth	892	667	152	23	5810	0	153,000	2.44	0.04
Individual investment	1191	23	2	18	110	0	3000	2.11	0.07
Team investment[b]	1213	42	3	29	398	0	13,000	2.04	0.06
Venture debt	121	212	8	5	766	0	5000	1.55	0.09
Cognitions									
PSED expected Yr5 employees	1173	34	3	19	335	0	8500	1.75	0.07
CAUSEE expected 12 m employees[c]	700	9	1	26	133	0	3500	2.12	0.08
PSED expected Yr5 revenue ($000)	1105	31	10	14	2530	0.2	500,000	1.71	0.08
CAUSEE expected 12 m rev ($000)	1232	4141	100	14	35,124	0.06	700,000	1.53	0.06
Actions									
Total team activities	1214	2	7	1	4	0	51	3.50	0.09
Total team hours[d]	1211	1931	400	7	5681	0	73,000	2.16	0.07
Environment: industry sector ($000)									
Construction	84	20	8	3	34	6	1137	2.03	0.07
Retail	204	86	10	8	416	3	4000	2.20	0.05
Manufacturing[e]	233	96	11	7	403	3	3800	1.79	0.06
Consumer products & services	246	66	13	8	277	2	2900	1.74	0.04
Software	308	22	128	17	54	2	748	2.27	0.03
Business products & services	592	113	13	7	318	2	4600	1.49	0.06
Franchises	126	47	11	4	100	2	742	1.60	0.08
IPO	81	901	161	8	3466	2	30,700	1.69	0.07
Outcome variables									
Revenue ($000)									
PSED II Yr0	119	400	60	6	1600	1	500	1.76	0.05
PSED II Yr3	126	360	42	7	1753	1	15,000	1.80	0.08
CAUSEE Yr0	587	1311	80	18	12,921	0.15	270,000	1.92	0.06
CAUSEE Yr3	150	339	160	5	788	0.50	6000	1.94	0.08
KFS Yr0	2602	393	40	49	7500	0	375,000	2.00	0.04
KFS Yr3	2209	1153	120	21	9866	0	250,000	1.97	0.04
INC	4990	44,056	10,000	17	223,984	2000	62,000,000	2.09	0.02
Revenue growth									
PSED ($000)	111	45	22	14	2182	1	4500	1.82	0.09
CAUSEE ($000)	202	959	100	6	3392	0.3	31,800	1.67	0.05
KFS ($000)	2085	698	60	24	6129	1	200,400	1.99	0.04
INC ($000)	4990	21,764	4991	19	99,350	1000	30,770,175	2.00	0.01
INC (%)	4990	298	127	11	763	2	19,812	2.57	0.02
Employees									
PSED II Yr0	125	4	0	6	3	0	16	3.50	0.43
PSED II Yr3	57	11	4	5	27	0	170	2.02	0.08
CAUSEE Yr0	309	9	1	3	7	0	45	3.21	0.08
CAUSEE Yr3	137	16	4	9	77	0	900	1.99	0.07
KFS Yr0	4823	2	0	11	7	0	165	2.20	0.06
KFS Yr3	2944	4	1	30	21	0	900	2.55	0.07
INC[f]	4990	202	50	26	1173	1	52,152	2.09	0.02
Employee growth									
PSED	59	33	1	33	38	0	1495	1.83	0.08
CAUSEE	61	25	4	9	114	0	894	1.87	0.09
KFS	643	1	1	16	8	0	175	2.31	0.08
INC	4198	74	15	22	399	0	14,680	2.12	0.02
INC (%)	4198	202	90	8	427	0	8050	2.66	0.03

n = number of observations with values of zero or larger; med = median; sd = standard deviation; min = score with the smallest value (minimum); max = score with the largest value (maximum); α = scaling exponent (i.e., slope) of the power law curve [the lower the value, the more of the total distribution resides in the tail]; K-S = Kolmogorov–Smirnov goodness-of-fit statistic, which compares hypothesized PL to Gaussian [the lower the value, the higher the probability of an underlying power law distribution]. PSED = Panel Study of Entrepreneurial Dynamics. CAUSEE = Comprehensive Australian Study of Entrepreneurial Emergence All K-S values are statistically significant (i.e., $p < 0.10$) except for the value of PSED II Employees Yr0, suggesting better fit with an underlying PL distribution compared to a normal distribution. Variables with superscripted letters are plotted as histograms on linear scales in Fig. 2.

4.2.3. Action-based variables

We measure *Total Team Activities* in the PSED II as the sum of all entrepreneurial activities (out of a list of such activities identified in Appendix 1) completed by owners of the venture over the first two years of organizing. Other studies have used PSED-type data to measure activity, calling it terms like "gestation activities," (Davidsson and Gordon, 2012), "efforts to create a venture" (Edelman and Yli-Renko, 2010), and "venture organizing activity" (Delmar and Shane, 2004). We use all 31 activities included in the PSED II. Similarly, we examine *Total Team Hours* as the total number of hours devoted to the formation of the new business, as reported by the primary respondent for all owners, to the question, "How many hours in total have you [and other owners] devoted to this new business?

4.2.4. Environment-based variables

We review the aggregated revenue of industry sectors, which can represent the environmental resources commonly measured in institutional theory and population ecology studies of density dependence (Aldrich and Ruef, 2006; Delmar and Shane, 2004). We separate all INC 5000 firms from the 2010 survey into self-identified industry sectors, analyzing those that have the closest match to the majority of nascent firm NAICS codes. We include sectors that have more than 50 firms, since results are more prone to bias for data sets of smaller size (Clauset et al., 2009). We list those sectors according to the number of firms in each: *Construction*, *Retail*, *Manufacturing*, *Consumer Products & Service*, *Software*, and *Business Products and Services*. Since research on entrepreneurship and entrepreneurial performance have been studied in the context of franchises and initial public offerings (cf., Gulati and Higgins, 2003; Kaufmann and Dant, 1999), we also analyze the distribution of revenue of firms purchased under a franchise agreement as *Franchises* and those firms that have filed for an Initial Public Offering as *IPO*. The latter two variables provide a relatively diverse range of industries, similar to the composition of sectors above.

4.3. Outcome variables

We analyze the outcome variables from all four data sets, focusing on the level and the growth of revenue and number of employees as our key metrics, for as mentioned above these are the only outcome measures that are unambiguous and valid in all types of firms regardless of size or ownership structure. Since many companies in the PSED II, CAUSEE, and KFS databases start small—where a large portion have either zero employees or zero revenue—measuring growth as a relative percentage would make the growth rate go to infinity, and disproportionately skew the analysis. To avoid this problem, we report growth in revenue and number of employees in absolute terms, from the initial collection of data (Year 0) and from the fourth wave of data collection (Year 3), as shown below in Eq. (1); and, for the more established firms in the INC 5000 dataset, we also measure growth as a relative percentage increase, calculated using Eq. (2) as follows:

$$(\text{Year 3}–\text{Year 0}) \tag{1}$$

$$[((\text{Year 3}/\text{Year 0})–1) * 100]. \tag{2}$$

When calculating revenue growth for both the PSED II and CAUSEE, there were about 20 cases from each dataset with missing values for Year 0. We analyzed these distributions with two missing data handling techniques: one by inserting a '0' for Year 0 (thus recognizing the achievement of a venture within a small sample), and another by removing the entire observation (thus reducing sample size and the potential richness of the data). In the analysis, described next, we found no substantive difference in any single parameter between the two missing data handling techniques and, since we use the reported revenue and employee outcomes for Yr3, we report growth results with the observations included.

4.4. Data analysis

To assess the presence of a PL in the data we used MATLAB (R2013b) software and followed the protocol for calculating PL model fit as described in Clauset et al. (2009). First, using the *plfit.m* MATLAB script found at www.santafe.edu/~aaronc/powerlaws/, we estimated the parameters for the scaling exponent of a power probability density function, $p(x) \sim x^{-\alpha}$, via maximum likelihood estimation (MLE). Running a semi-parametric Monte Carlo bootstrap calculation 1000 times, the script computes the Komolgorov–Smirnov (K-S) goodness of fit statistic. The K-S test is a non-parametric goodness of fit index similar to chi-square—like the chi-square statistic, smaller K-S values indicate better conformity to a power law because the null hypothesis is that there would be no absolute deviation between the observed and a perfectly formed power law distribution (Clauset et al., 2009). In addition, α is a scaling parameter that represents the overall dynamics of the distribution: the closer the number to 1.0, the longer the tail, and the greater proportion of the total distribution is in the tail (i.e., greater proportion of extreme scores). So, a distribution with $\alpha = 1.2$ has a greater proportion of extreme scores compared to a distribution with $\alpha = 2.2$.

Researchers in many social sciences have relaxed the definition of a statistically exact "normal distribution" (i.e., skew is exactly 0, and the mean, median, and mode are equal) to a more general approximation. Similarly, and following Aguinis et al. (in press), we refer to a "power law" as a heavy-tailed distribution where observations are clearly dominated by a small proportion of units, and where the majority of the units are to the left of the mean. As we discuss in the results below, PLs show a superior fit to the data compared to a normal distribution (based on the K-S statistic), but we acknowledge that distributions may not meet the traditional exactitude of a power law, where the distribution has a never-ending tail and (nearly) infinite variance. As identified in Clauset et al. (2009), there are alternative heavy-tailed distributions such as log-normal and exponential. Our primary goal is not to describe the precise shape of the distribution for each of the variables we examine but, rather, to challenge the assumption of normality in order to potentially capture entrepreneurial phenomena more accurately.

5. Results

Table 1 includes distribution descriptors for input and outcome variables. In addition to K-S and α values, we also report traditional descriptive statistics—including the mean, median (med), skewness (skew), standard deviation (sd), and the minimum (min) and maximum (max) value. Information in Table 1 shows that the majority of distributions are heavily skewed and the means and median values are dissimilar. Thus, this information, in the form of the familiar skew, mean, and median metrics, offers an initial glimpse into the non-normal nature of the distributions. Table 1 also includes the K-S value for each distribution. K-S values ≤ 0.10 provide consistent and unbiased estimates supporting a hypothesis that the distribution is more accurately characterized by a PL than a normal distribution (Clauset et al., 2009).

Results in Table 1 show that virtually all variables are more accurately described as a PL than a normal distribution. Specifically, of the 49 variables we analyzed, only one—PSED II Employees Yr0—is better approximated with a normal (Gaussian) distribution than a PL distribution, with a K-S statistic larger than 0.10.[9] In short, results offer strong evidence regarding the pervasive presence of PL distributions in entrepreneurial inputs and outcomes.

In addition to the analytic results, Fig. 2 shows six histograms that serve as exemplars of the shape of these distributions. Each histogram is very similar to the stylized PL curve shown in Fig. 1, where the majority of observations are at the low end, to the left of the mean. As shown in Fig. 2, and consistent with the high percentage of α values below 3 (80% of the variables) in Table 1, results show that the proportion of extreme scores in each of these distributions is much larger than would be expected if there were an underlying normal distribution. In fact, if there were an underlying normal distribution, about 20% of the firms have scores that are so high (approximately 3 standard deviations above the mean or higher), the probability of their occurrence is effectively zero.

6. Discussion

Our results suggest that a normal distribution is not an accurate depiction of empirical reality for entrepreneurial firms for 48 out of the 49 variables we examined. Instead, PL distributions are pervasive in entrepreneurial inputs and outcomes; thus, assumptions of normality may only be applicable in special cases. Accordingly, we find support for the following conclusion regarding entrepreneurship research: *Variables of interest should be assumed as PL distributed unless proven otherwise*. In so doing, our study extends a growing recognition that the social world seems to be organized according to power law distributions (Aguinis and O'Boyle, 2014; Aguinis et al., in press; Andriani and McKelvey, 2007, 2009; Barabási et al., 2002; Boisot and McKelvey, 2010; Powell, 2003). As we describe next, the discovery of this empirical result has significant implications for our theorizing about entrepreneurship.

6.1. Implications for theory and research

The most obvious question for future research raised by our findings is: What drives the performance of those new ventures at the positive extreme of the distribution—the high potential companies that have a significant impact on entire industries, places, and patterns of human behavior? Early entrepreneurship research found that the drivers of "marginal survival" and "high performance" are, in part, qualitatively different (Cooper et al., 1994; Dahlqvist et al., 2000). However in that context, "high performance" was modest compared to extreme elite firms. There is reason to suspect that the processes shaping performance at the extremely high end may again be qualitatively different from "high performance." To gain insights about these important extreme cases requires longitudinal research that captures a sufficient number of them at an early stage (i.e., before they have proven what they are about to become), and comparing their rapid growth to the processes that lead to more modest outcomes. This is a very challenging task, but might be accomplished in one of three ways: (a) pooling several extant "nascent entrepreneurship" data sets, so as to arrive at an analyzable group of cases with extreme performance outcomes; (b) undertaking new and very large, representative studies for the same purpose, and (c) conducting studies guided by a sampling logic aiming at theoretical representation of the relevant outcome distribution, rather than statistical representation of present-day empirical populations, which will always be dominated by the "modest majority" (Davidsson and Gordon, 2012). Note that these proposed research designs should not aim to include solely what researchers think *a priori* these "high(er) potential" cases might look like, for there may not exist identifiable seeds of eventual greatness at the early stages of a new venture.

Another relevant question is whether the main task for entrepreneurship research is to explain variance in financial performance (e.g., sales; profits) among firms after they have been established, or to explain how they come into existence in the first place (Gartner, 1988; Venkataraman, 1997). The latter focus would ask the following: Compared to more modest start-ups, what characterizes the emergence journey of those ventures that already have very significant levels of resource investments and numbers of employees at the point where they can be regarded "operational" rather than "nascent" ventures? A challenge here is that the

[9] A likely explanation for this result is that power laws take time to emerge and are less likely when there are constraints on a system (Aguinis et al., in press). Nascent firms, with limited legitimacy in the marketplace (i.e., questionable ability to provide competitive pay, benefits, and job stability at the inception of the firm), are somewhat constrained in their abilities to hire employees.

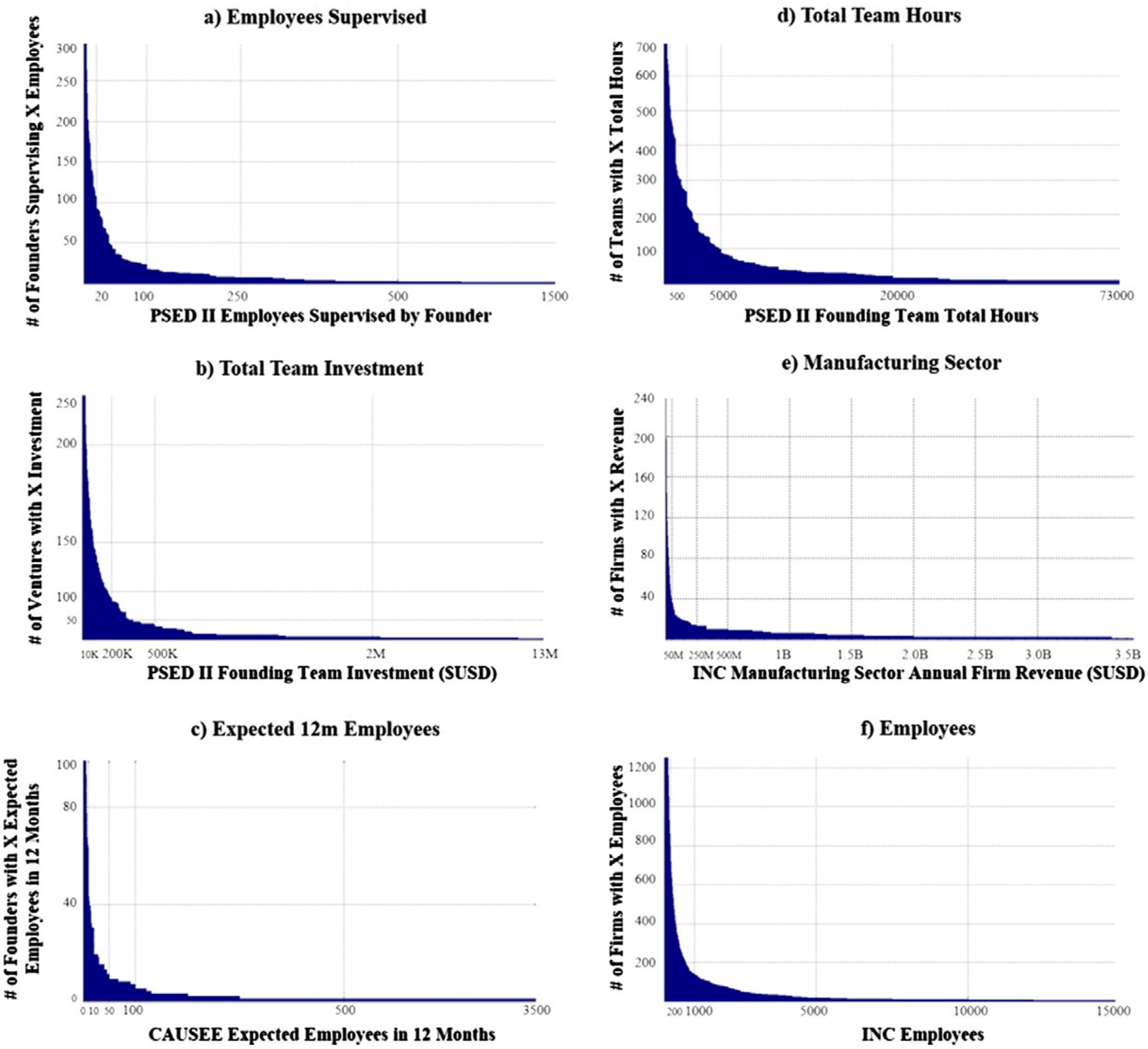

Fig. 2. Frequency distributions of illustrative input and outcome variables in entrepreneurship research.

"gestation period" of start-ups is often quite long (Reynolds, 2007) and likely even more so for ambitious and innovative start-ups (Samuelsson and Davidsson, 2009). Therefore, if a performance criterion like "positive cash flow" is used as a marker of the transition into an operational firm (Lichtenstein et al., 2007; Newbert et al., 2013), there is considerable risk that many of the rare ventures that will eventually show outstanding performance will not do so within the time window of the study. One such example would be Amazon.com, which went seven full years of operations without a quarterly profit (Spector, 2002). Therefore, in the pursuit of insights into the drivers of emergence of extreme new ventures, a way around the problem of long delays in financial performance might be to examine the development of configurations of extreme values on what are commonly regarded as "antecedents" or "independent" variables, which our analysis demonstrated are also power law distributed.

In a Gaussian world, differences in inputs explain different outcomes that remain relatively close to the mean of the distribution. Most of these explanations are based on models of linear causality, which dominate entrepreneurship research (Dean et al., 2007; Delbridge and Fiss, 2013; McKelvey, 2004a). Unfortunately, these theoretical frameworks have difficulty explaining outcomes where the highest performers are not just six standard deviations away from the lowest, but can be 20 times or even 1000 times larger. What types of theories could allow entrepreneurship to reach the goal of improving our understanding of the full range of observations—not just those close to the mean—without the need to squeeze observed scores into a normal distribution? In particular, how can we understand those extremely influential ventures that are frequently discarded as anomalies, outliers, or errors when assuming that the underlying distribution is normal? How should we reconceptualize both inputs and outcomes in entrepreneurship with the goal of improving our understanding of substantive phenomena? We address these questions next.

6.1.1. Generative causal processes of PL distributions

Complexity science researchers have identified a series of generative mechanisms (i.e., causal processes) that yield power law distributions. Andriani and McKelvey (2009) summarized this literature by describing 15 specific causes of PLs which they organized into four categories: positive feedback, contextual effects, ratio imbalances, and multiple scale-free causes. These generative mechanisms provide plausible theoretical explanations for PLs. Based on our results regarding resources, cognitions, actions, and environments, we elucidate mechanisms within each category that may be directly applicable to entrepreneurship: *preferential attachment* in positive feedback; *self-organized criticality* and *phase transitions* in contextual effects; *hierarchical modularity* in ratio imbalances; and *combination theory* in multiple scale-free causes. We review each with an eye toward new theory development in entrepreneurship.

6.1.1.1. Positive feedback mechanisms. Barabási's (2009) *preferential attachment* model focuses on networks, explaining how the larger nodes exhibit a Matthew Effect, where the rich get richer, based on differences in initial conditions. According to this model, when new agents enter a system, they prefer to connect to the node that is most easily recognized, which usually is the largest node in a system. Here, the causal mechanism that drives the generation of a power law is the node's initial endowment—its size. For example, Google's initial development of its search engine attracted customers, which increased its visibility. This attracted talented engineers, who invented new methods and products that attracted additional customers. This led to increased profits and attracted merger partners, which then attracted even more customers. Thus, when new customers are looking for a search engine, Google is most likely to be more prominently listed since it is currently the most popular, and it is more likely to be chosen. These are recursive processes inherent in cumulative advantage: over time, small advantages can lead to extreme differences in outcomes. Aguinis et al. (in press) proposed that a meta-theoretical principle of cumulative advantage explains the emergence of PLs within the context of individual performance. This conclusion was reached based on 633,876 productivity observations collected from approximately 625,000 individuals in occupations including research, entertainment, politics, sports, sales, and manufacturing, among others. Thus, given that some new ventures begin the creation process with more resources than others, cumulative advantage has the potential to explain PLs in entrepreneurship.

6.1.1.2. Contextual effects mechanisms. Another theoretical explanation for the presence of PL distributions is Bak and Chen's (1991) *self-organized criticality* (SOC), which describes a dynamic system that has built up to a critical point of stable disequilibrium. This model is best represented by a sandpile: once the pile has built up, similar grains of sand, when dropped one-by-one, have differential effects—most grains that are dropped only move one or two grains on the pile, while a few grains have extreme cascading effects, moving hundreds of grains. Measured over time, the sizes of all these effects are distributed according to a power law. Thus, when a system is positioned at a critical point, the addition of a single new input can cause dramatic change. In entrepreneurial terms, for a company poised in SOC, the addition of one input (e.g., a company adding a star employee, the presentation of a business plan in Silicon Valley instead of in a university classroom) can sometimes cause a nonlinear avalanche of outcomes. Going further, Newman (2005) showed that in these SOC systems, there is an increased likelihood of these extreme "black swan" events.

A related stream of research focuses on the threshold to SOC, and how it reflects a phase transition in the system. As Boisot and McKelvey (2010, p. 422) explained, "Beyond certain thresholds, complexity can lead to phase transitions toward either emergent order—that is, dissipative structures that maintain themselves in existence by continuously importing free energy from their environment and exporting bound energy back into it—or greater chaos." At this threshold or bifurcation point, the system changes from an additive, linear state into a multiplicative, nonlinear state that is qualitatively different from its previous state (Dooley and Van de Ven, 1999; McKelvey, 2004b). Dissipative structures theory (Prigogine, 1955; Prigogine and Stengers, 1984) has been used by entrepreneurship scholars to explore the idea of phase transitions, as a way of explaining how new regimes of order can emerge in organizations and systems pushed into disequilibrium (Chiles et al., 2004; Lichtenstein, 2000, 2014; Slevin and Covin, 1997). Chiles et al. (2010a) used the term *kaleidic* to express this world of perpetual disequilibrium which leads to abrupt shifts from one phase to another. In each of these PL generating theories—SOC and dissipative structures—the non-linear PL outcomes are caused by the dynamics of the disequilibrium state on its context (i.e., contextual effects).

6.1.1.3. Ratio imbalance mechanisms. A third category of generative mechanisms that drive the emergence of PL distributions shows how some cost-driven efficiency in growth results in internal hierarchies that follow a PL. One exemplar of this mechanism was discovered by Carniero (1987), who found that native villages will grow to a certain size, but then – due to inefficiences and communication breakdowns – will split into two self-contained villages. Over time the entire set of villages follows a PL.

In management, the classic example is Simon's (1962) theory of hierarchical modularity, which claims that "nearly decomposable systems" are the most adaptable in a dynamic environment. In practical terms, when organizational sub-systems (e.g. work units, product lines) are loosely coupled, each one has more flexibility to adopt the most efficient method of achieving its goals; in aggregate, this confers a high degree of effectiveness to the firm as a whole. In theory (McKelvey, 2012), if all firms in a sector pursue a modular approach, the entire landscape will become fractal, i.e., PL distributed. (See also Andriani and McKelvey, 2007).

6.1.1.4. Multiple scale-free mechanisms. The fourth category refers to systems with multiplicative effects; that is, where the aggregate interactions that lead to a phenomenon are not additive but multiplicative. A good example occurs in food webs, in which a fractal structure of predators and of niche resources will multiply to generate a fractal structure of a given species (Pimm, 1982; Preston, 1948). Likewise, combination theory shows that even when each element of a process is not fully non-normal, their interaction often results in PL distributions (West and Deering, 1995). Such combinations are likely in entrepreneurship, heightening the need to examine the underlying structure of each input and output variable in a theory.

6.1.2. New theories of entrepreneurship

Whereas current entrepreneurship theories of resources, cognitions, actions, and environmental influences are focused on the differences in inputs and outcomes across individual ventures, a PL conceptualization requires explanations about the entire set of ventures, through causal process that yield the whole range of scores. Thus, for example, new theories will have to explain outcomes that range from the failure of a first-time entrepreneur to the success of an INC 5000 company. By attempting to also account for the extremes in the distribution, our theories can reflect empirical reality more closely.

An intriguing approach draws on Drazin and Sandeland's (1992) autogenesis theory, which proposes that expectations about future outcomes are what drive interactions among interdependent agents. These expectations create tensions *within* the agent (as she compares her current state with a projected future state) and *between* agents (where differences can be amplified through positive feedback loops). In combination, outcomes are PL distributed. In entrepreneurship, research has similarly demonstrated that aspirations or expectations for growth are predictive of the actual growth outcomes that ensue (Delmar and Wiklund, 2008; Mok and van den Tillaart, 1990; Wiklund and Shepherd, 2003). It seems conceivable, then, to propose that expectations for growth in entrepreneurship—from the bottom-up (from founders) and from the top-down (from a VC or other potential stakeholders)—could be a driving causal process in the domain of entrepreneurship.

A focus on extremes also emphasizes the conditions that distinguish the small percentage of ventures that emerge in the tail of a power law distribution. According to power law theory, these extremes are differentiated from the rest of the data at a minimal threshold point in the distribution—what McKelvey (2004b, p. 319) calls "the first critical value." Thus, new theories and methods are needed to understand what is happening at either side of this critical inflection point. In other words, how is the system different before the threshold versus afterwards? Research into threshold points has pointed to a variety of pathways worth exploring in this regard (Granovetter, 1978; Granovetter and Soong, 1986; Macy, 1991; Pierce and Aguinis, 2013). One approach to examining this issue was identified by Aguinis and O'Boyle (2014), who compared differences between distributions that are Gaussian versus those that are PL. Specifically, in PL distributions, only minor differences in an input (e.g., firm revenue) can generate dramatic differences in outcomes to the system (e.g., firm market capitalization). Thus, new theories could focus on the points of leverage that drive fast-growth ventures, and compare these across samples to find which inputs have the strongest causal relation to extreme outcomes. Next we describe methodological approaches that will facilitate future empirical research.

6.1.3. Non-traditional empirical methods for PL data

Our results suggest that entrepreneurship research focusing on inputs or outcomes of individual, team, or venture constructs cannot use methods that assume normal distributions, including techniques like OLS regression, structural equation modeling, ANOVA, hierarchical linear modeling, or meta-analysis. Though virtually all statistical approaches in entrepreneurship are based on the assumption of Gaussian distributions (Dean et al., 2007), other research domains have made methodological progress with techniques that do not rely on this assumption.

One technique, used in our study, draws from Clauset et al.'s (2009) approach for validating PL distributions. At a minimum, this method should be utilized in conjunction with any statistical analysis to confirm whether the data are in fact normally distributed. This should be seen as a precondition for using traditional data analytic tools. Moreover, future research can aim at understanding reasons for the presence of PL distributions when they are found—as discussed in the previous section. For example, to test a preferential attachment hypothesis, future research could identify critical values within a distribution of initial firm resources, then use nonlinear correlations (e.g., Kendall's Tau or Spearman's Rho) on longitudinal data to describe how resources above and below the threshold influence a firm's outcomes over time. A test of self-organized criticality could identify minimal thresholds in environmental resources (e.g., the distribution of city size or available venture capital per city) and propose how founders' organizing processes might be different in areas above and below the critical point. For autogenesis, scholars could study how initial expectations for growth may have a nonlinear association with outcomes, by finding parallels between power law slopes (i.e., alpha) of nascent expectations and subsequent outcomes, and by conducting an analysis of long-range correlations which, according to Sornette (2006, p. 223), "are the signatures that inform us about the underlying mechanisms" of a nonlinear system.

Another complementary approach, identified by Kruschke et al. (2012), is the use of Bayesian statistics, which allow the researcher to "…specify the functional form of the likelihood function and prior distribution" (p. 728) in advance of the analysis. The PLs we find can be the empirical grounding for future Bayesian models to specify prior and posterior distributions (what we label as inputs and outcomes, respectively). In so doing, researchers can compare the likelihood of outcomes under Gaussian conditions and PL conditions, allowing a far greater range of possible explanations. Although Bayesian models are rare in organization science—Kruschke et al. (2012) found them in only 0.5% of 10,000 articles across 10 years in top journals—they would be particularly useful at identifying the conditions that differentiate normal versus PL outcomes.

The discovery of PLs in entrepreneurship makes the case for a renewed interest in building thick descriptions of the processes in which entrepreneurs engage (Poole, et al., 2000; Van de Ven and Engleman, 2004). Recall that PLs are generated by recursive interactions among agents in a system (Andriani and McKelvey, 2009). Thus, research designs that can study these interactions in depth (e.g., ethnographies and comparative longitudinal case studies) can reveal the underlying dynamics of interaction loops between entrepreneurs and their customers, suppliers, and other stakeholders. Overall, rich longitudinal and qualitative research of organizations at the extremes provides behavioral insights that are unavailable in purely quantitative designs.

Finally, propositions about which mechanisms drive the emergence of power laws could be examined within agent-based simulation models. These models are important to entrepreneurship theory because each agent within the model can be heterogeneously endowed with resources and decision rules for interacting with other agents and for capturing resources in the environment (McMullen and Dimov, 2013). In contrast to deterministic simulations like systems dynamics (cf., Lomi et al., 2010), agent-based computational models are stochastic and probabilistic: even if the same exact input values are used for each agent in a model, the probabilistic interactions among the agents make the same exact outcome values highly unlikely in subsequent runs of the simulation. Hence, Monte Carlo experiments that keep all inputs the same are run multiple times (e.g., 30–100) to estimate the most probable outcome. Agent-based models like these are particularly well-suited to build what Davis et al. (2007) termed "simple theory"—where theoretical constructs are only partially developed within a domain. Proposing a mechanism like preferential attachment as the driver of power laws, for example, agents that start with slightly greater endowments (e.g., human, social, or financial capital) than others could be programmed to have a slightly higher probability of successfully capturing resources (e.g., bank or venture capital funding) in the environment; over time, as network research by Barabási (2009) has demonstrated, the cumulative outcomes are likely to be power law distributed.

Most relevant to our power law arguments in this paper, agent-based computational models can be programmed with the distribution of both agent and environmental resources specified *a priori*. NetLogo is one agent-based modeling toolkit (Wilensky, 1999), available online as freeware, which scholars could use to enhance the verisimilitude of theories about entrepreneurship that more accurately reflect empirical reality. Within NetLogo, agent-based simulations of preferential attachment and self-organized criticality already exist in the model library—these models can provide a validated baseline for the development of future theory, where scholars can change the settings of existing parameters, add parameters, and compare how well model outcomes correspond to empirical—i.e., power law—results (Rand and Rust, 2011). More creative studies could mix the methods we describe above to build more accurate models of entrepreneurship. One ambitious research project could combine an ethnographic study to understand how entrepreneurs interact, then create a computational model in NetLogo to demonstrate how micro-level interactions among agents can generate macro-level outcomes, and then validate those outcomes with one of the representative samples we analyzed in our study.

6.2. Implications for policy and practice

For policy, current interventions in the form of tax breaks, business incubators, or grants, are instituted because they are supposed to increase environmental munificence and, hypothetically, stimulate new entrepreneurial activity—to spur innovation and create new jobs. However, these approaches are most often of limited impact, because the majority of founders do not expect to grow (Amezcua et al., 2013; Arshed et al., 2014). Thus, they avoid doing things that lead to growth, like forming teams with diverse knowledge, asking for external funding, or developing proprietary processes. From that perspective, the efficacy of interventions may be enhanced by providing incentives for firms to: (1) begin the venture with qualitatively different inputs, like a diverse team or a substantial equity investment; and/or (2) to achieve outcomes in the tail of the distribution. In these ways new firms procure enough resources as a foundation for potential growth, and all firms are given aggressive, yet cognitively achievable performance goals.

Like the PL distributed reward systems outlined in Aguinis and O'Boyle (2014), entrepreneurs with the intention of changing the global landscape would be best served by finding employees in the tail of the distribution—"stars"—and putting systems in place that reward outlier performance. Similarly, as mentioned in the previous section, our data-analysis method can estimate where the tail of the distribution begins. Emerging companies can use these estimates as empirical benchmarks to encourage employees to push the envelope on performance.

7. Conclusion

Our results offer empirical evidence that the assumption of normality in entrepreneurship research is untenable in most cases. Accordingly, these empirical results call for an important shift in terms of future entrepreneurship theory and research. PL distributions suggest that more attention needs to be given to those outliers that make a disproportionate contribution. For example, according to Shane (2008), 95% of all U.S. businesses are small (employing 20 people or fewer), 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 focus on the mean, it may continue to achieve statistically significant results, but the domain is unlikely to make important theoretical progress. Moreover, 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. Instead, 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. Our proposed agenda suggests that researchers should first determine whether a particular distribution is normal or not, and then understand the emergence mechanisms of power law distributions. Our approach offers directions for future research whose results we hope will help advance entrepreneurship theory and practice in important ways.

Appendix 1. Constructs, variables, and items

Human capital

Team education

What is the highest level of education you have completed: up to the eighth grade, some high school, high school degree, technical or vocational degree, some college, community college degree, a bachelor's degree, some graduate training, a master's degree, or a law degree, medical degree, or Doctorate? Each was coded as follows:

☐ up to the eighth grade = 8	☐ a bachelor's degree = 16
☐ some high school = 10	☐ some graduate training = 17
☐ high school degree = 12	☐ a master's degree = 18
☐ technical or vocational degree = 13	☐ a law degree, medical degree, or Doctorate = 22
☐ some college = 14	
☐ community college degree = 15	

When there were more than five owners, we added 12 years for each additional member. The sum for the entire ownership team is recorded as Team Education.

Previous ventures founded

How many other businesses have you helped to start as an owner or a part-owner? Besides the new business discussed in this interview, how many other businesses do you own? We sum the occurrences for the entire team.

Actions

For this study, we sum all 31 the activities into one variable, *Total Team Activities*. Activities are based on all 31 entrepreneurial activities measured in the PSED II data, which are:

1. Established credit with supplier
2. Started marketing efforts
3. Received income from first sale
4. Retained accountant for business
5. Discussions with potential customers
6. Retained lawyer for business
7. Purchased liability insurance
8. Started development of prototype
9. Prepared financial projections
10. Determined regulatory requirements
11. Opened bank account
12. Became member of trade association
13. Listed business in phone book
14. Established internet communication
15. Applied for federal EIN
16. Filed DBA with government
17. Paid unemployment tax
18. Paid social security tax
19. Filed federal income tax
20. Registered with D&B
21. Received external funding
22. Hired employees
23. Purchased major equipment
24. Bought raw materials
25. Collected competitive information
26. Defined market opportunity
27. Asked for funds
28. Focused full-time on new venture
29. Developed proprietary technology or process
30. Applied for patent
31. Business plan development

References

Abbott, A., 1988. Transcending general linear reality. Sociol. Theory 6 (Fall), 169–186.
Abul-Magd, A.Y., 2002. Wealth distribution in an ancient Egyptian society. Phys. Rev. E. 66 (5), 057104.
Acs, Z., 2008. Foundations of High-impact Entrepreneurship. Now Publishers, Boston.

Aguinis, H., Joo, H., 2015. Debunking myths and urban legends about how to identify influential outliers. In: Lance, C.E., Vandenberg, R.J. (Eds.), More Statistical and Methodological Myths and Urban Legends. Routledge, New York, pp. 206–223.

Aguinis, H., Lawal, S.O., 2012. Conducting field experiments using eLancing's natural environment. J. Bus. Ventur. 27 (4), 493–505.

Aguinis, H., O'Boyle, E.H., 2014. Star performers in twenty-first-century organizations. Pers. Psychol. 67, 313–350.

Aguinis, H., Pierce, C.A., Culpepper, S.A., 2009. Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. Organ. Res. Methods 12 (2), 623–652.

Aguinis, H., Gottfredson, R.K., Joo, H., 2013. Best-practice recommendations for defining, identifying, and handling outliers. Organ. Res. Methods 16 (2), 270–301.

Aguinis, H., O'Boyle, E., Gonzalez-Mule, E., Joo, H., 2015. Cumulative advantage: Conductors and insulators of heavy-tailed productivity distributions and productivity stars. Pers. Psychol. http://dx.doi.org/10.1111/peps.12095 (in press).

Aldrich, H.E., 1990. Using an eological perspective to study organizational founding rates. Enterp. Theory Pract. 14 (3), 7–24.

Aldrich, H.E., Kim, P., 2007. Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search. Strateg. Entrep. J. 1 (1–2), 147–166.

Aldrich, H.E., Ruef, M., 2006. Organizations Evolving. 2nd ed. Sage Publications, London.

Allen, S.D., Link, A.N., Rosenbaum, D.T., 2007. Entrepreneurship and human capital: Evidence of patenting activity from the academic sector. Enterp. Theory Pract. 31 (6), 937–951.

Alvarez, S.A., Barney, J.B., 2007. Discovery and creation: alternative theories of entrepreneurial action. Strateg. Entrep. J. 1 (1–2), 11–27.

Alvesson, M., Sandberg, M., 2011. Generating research questions through problematization. Academy of Management Review 36 (2), 247–271.

Amezcua, A., Grimes, M., Bradley, S., Wiklund, J., 2013. Organizational sponsorship and founding environments: A contingency view on the survival of business incubated firms, 1994–2007. Acad. Manag. J. 56 (6), 1628–1654.

Andriani, P., McKelvey, B., 2007. Beyond gaussian averages: Redirecting international business and management research toward extreme events and power laws. J. Int. Bus. Stud. 38 (7), 1212–1230.

Andriani, P., McKelvey, B., 2009. From Gaussian to Paretian thinking: causes and implications of power laws in organizations. Organ. Sci. 20 (6), 1053–1071.

Arend, R.J., Levesque, M., 2010. Is the resource-based view a practical organizational theory? Organ. Sci. 21 (4), 913–930.

Arshed, N., Carter, S., Mason, C., 2014. The ineffectiveness of entrepreneurship policy: Is policy to blame? Small Bus. Econ. http://dx.doi.org/10.1007/s11187-014-9554-8.

Arthur, B., 1988. Self-reinforcing mechanisms in economics. In: Anderson, P., Arrow, K., Pines, D. (Eds.), The Economy as an Evolving Complex System. Wiley, New York, pp. 9–33.

Axtell, R.J., 1999. The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distribution Center on Social and Economic Dynamics Working. Brookings Institution, Washington, D.C.

Bacharach, S.B., 1989. Organizational theories: Some criteria for evaluation. Acad. Manag. Rev. 14 (4), 496–515.

Bak, P., Chen, K., 1991. Self-organized criticality. Sci. Am. 264 (1), 46–54.

Bandura, A., 1989. Human agency in social cognitive theory. Am. Psychol. 44 (9), 1175–1184.

Barabási, A.-L., 2009. Scale-free networks: A decade and beyond. Science 325 (5939), 412–413.

Barabási, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A 311 (3), 590–614.

Baron, R., 2004. The cognitive perspective: A valuable tool for answering entrepreneurship's basic “why?” questions. J. Bus. Ventur. 19 (2), 221–240.

Bhide, A., 2000. The Origin and Evolution of New Businesses. Oxford University Press.

Boisot, M., McKelvey, B., 2010. Integrating modernist and postmodernist perspectives on organizations: A complexity science bridge. Acad. Manag. Rev. 35 (3), 415–433.

Bruderl, J., Preisendörfer, P., 1998. Network support and the success of newly founded business. Small Bus. Econ. 10 (3), 213–225.

Buldyrev, S.V., Dokholyan, N.V., Erramilli, S., Hong, M., Kim, J.Y., Malescio, G., Stanley, H.E., 2003. Hierarchy in social organization. Physica A 330 (3), 653–659.

Burt, R.S., 1997. The contingent value of social capital. Adm. Sci. Q. 42 (2), 339–361.

Carniero, R., 1987. The evolution of complexity in human societies and its mathematical expression. Int. J. Comp. Sociol. 28 (1), 111–128.

Cassar, G., 2014. Industry and startup experience on entrepreneur forecast performance in new firms. J. Bus. Ventur. 29 (1), 137–151.

Chandler, G., McKelvie, A., Davidsson, P., 2009. Asset specificity and behavioral uncertainty as moderators of the sales growth: Employment growth relationship in emerging ventures. J. Bus. Ventur. 24 (4), 373–387.

Chen, C.C., Greene, P.G., Crick, A., 1998. Does entrepreneurial self-efficacy distinguish entrepreneurs from managers? J. Bus. Ventur. 13 (4), 295–316.

Chiles, T., Meyer, A., Hench, T., 2004. Organizational emergence: The origin and transformation of Branson, Missouri's musical theaters. Organ. Sci. 15 (5), 499–520.

Chiles, T., Bluedorn, A., Gupta, V.K., 2007. Beyond creative destruction and entrepreneurial discovery: A radical Austrian approach to entrepreneurship. Organ. Stud. 28 (4), 467–493.

Chiles, T., Tuggle, C.S., McMullen, J., Bierman, L., Greening, D., 2010a. Dynamic creation: elaborating a radical Austrian approach to entrepreneurship. Organ. Stud. 31 (1), 7–46.

Chiles, T.H., Vultee, D.M., Gupta, V.K., Greening, D.W., Tuggle, C.S., 2010b. The philosophical foundations of a radical Austrian approach to entrepreneurship. J. Manag. Inq. 19 (2), 138–164.

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.

Cooper, A., Gimeno-Gascon, F.J., Woo, C., 1994. Initial human and financial capital as predictors of new venture performance. J. Bus. Ventur. 9 (5), 371–396.

Cornelissen, J.P., Durand, R., 2013. More than just novelty: Conceptual blending and causality. Acad. Manag. Rev. 37 (1), 152–154.

Crawford, G.C., McKelvey, B., 2012. Strategic implications of power law distributions in the creation and emergence of new ventures: Power law analyses in three panel studies. Front. Entrep. Res. 32 (12), 1.

Crawford, G.C., McKelvey, B., Lichtenstein, B., 2014. The empirical reality of entrepreneurship: How power law distributed outcomes call for new theory and method. J. Bus. Vent. Insights 1 (1–2), 3–7.

Dahlqvist, J., Davidsson, P., Wiklund, J., 2000. Initial conditions as predictors of new venture performance: a replication and extension of the Cooper et al. study. Enterp. Innov. Manag. Stud. 1 (1), 1–17.

Davidsson, P., 2004. Researching Entrepreneurship. Springer, New York.

Davidsson, P., Gordon, S.R., 2012. Panel studies of new venture creation: a methods-focused review and suggestions for future research. Small Bus. Econ. 39 (4), 853–876.

Davidsson, P., Achtenhagen, L., Naldi, L., 2010. Small firm growth. Found. Trends Entrep. 6 (2), 69–166.

Davis, J.P., Eisenhardt, K.M., Bingham, C.B., 2007. Developing theory through simulation methods. Acad. Manag. Rev. 32 (2), 480–499.

Dean, M.A., Shook, C.L., Payne, G.T., 2007. The past, present, and future of entrepreneurship research: Data analytic trends and training. Enterp. Theory Pract. 31 (4), 601–618.

Delbridge, R., Fiss, P., 2013. Styles of theorizing and the social organization of knowledge. Acad. Manag. Rev. 38 (3), 325–331.

Delmar, F., 1997. Measuring growth: methodological considerations and empirical results. In: Donckels, R., Miettinen, A. (Eds.), Entrepreneurship and SME Research: On its Way to the Next Millennium. Ashgate, Aldershot, UK and Brookfield, VA, pp. 199–216.

Delmar, F., Shane, S., 2004. Legitimating first: organizing activities and the survival of new ventures. J. Bus. Ventur. 19 (3), 385–410.

Delmar, F., Wiklund, J., 2008. The effect of small business managers' growth motivation on firm growth: A longitudinal study. Enterp. Theory Pract. 32 (3), 437–457.

Delmar, F., Davidsson, P., Gartner, W.B., 2003. Arriving at the high-growth firm. J. Bus. Ventur. 18 (2), 189–216.

DesRoches, D., Robb, A., Mulcahy, T., 2009. Kauffman Firm Survey (KFS) — Baseline/first/second Follow-ups: Study Metadata Documentation. SSRN eLibrary.

Dooley, K.J., Van de Ven, A.H., 1999. Explaining complex organizational dynamics. Organ. Sci. 10 (3), 358–372.

Drazin, R., Sandelands, L., 1992. Autogenesis: A perspective on the process of organizing. Organ. Sci. 3 (2), 230–249.

Edelman, L.F., Yli-Renko, H., 2010. The impact of environment and entrepreneurial perceptions on venture-creation efforts: Bridging the discovery and creation views of entrepreneurship. Enterp. Theory Pract. 34 (5), 833–856.

Farjoun, M., Levin, M., 2011. A fractal approach to industry dynamism. Organization Studies 32 (6), 825–851.

Fleming, L., 2007. Breakthroughs and the 'long tail' of innovation. MIT Sloan Manag. Rev. 49 (1), 68–74.
Fleming, L., Sorenson, O., 2004. Science as a map in technological search. Strateg. Manag. J. 25 (8–9), 909–928.
Flier, B., Van Den Bosch, F., Volberda, H., 2003. Co-evolution in strategic renewal behaviour of British, Dutch and French financial incumbents: Interaction of environmental selection, institutional effects and managerial intentionality. J. Manag. Stud. 40 (8), 2163–2187.
Frid, C., 2013. Publications based on the panel study of entrepreneurial dynamics. http://www.psed.isr.umich.edu/psed/documentation.
Ganco, M., Agarwal, R., 2009. Performance differentials between diversifying entrants and entrepreneurial start-ups: A complexity approach. Acad. Manag. Rev. 34 (2), 228–253.
Garnsey, E., Stam, E., Heffernan, P., 2006. New firm growth: Exploring processes and paths. Ind. Innov. 13 (1), 1–20.
Gartner, W.B., 1988. "Who is an entrepreneur?" is the wrong question. Am. Small Bus. J. 12 (4), 11–31.
Gartner, W.B., 1990. What are we talking about when we are talking about entrepreneurship? J. Bus. Ventur. 5 (1), 15–28.
Gartner, W.B., Shaver, K.G., Carter, N.M., Reynolds, P.D. (Eds.), 2004. Handbook of Entrepreneurial Dynamics: The Process of Business Creation. Sage Publications, Thousand Oaks, CA.
Gatewood, E.J., Shaver, K.G., Powers, J.B., Gartner, W.B., 2002. Entrepreneurial expectancy, task effort, and performance. Enterp. Theory Pract. 27 (2), 187–206.
Gell-Mann, M., 1988. The concept of the institute. In: Pines, D. (Ed.), Emerging Synthesis in Science. Addison-Wesley, Boston, MA, pp. 1–15.
Godfrey, P.C., Merrill, C.B., Hansen, J.M., 2009. The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. Strateg. Manag. J. 30 (4), 425–445.
González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782.
Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83, 1420–1443.
Granovetter, M., 1983. The strength of weak ties: A network theory revisited. Sociol. Theory 1 (1), 201–233.
Granovetter, M., Soong, R., 1986. Threshold models of interpersonal effects in consumer demand. J. Econ. Behav. Organ. 7, 83–100.
Greene, W.H., 2011. Econometric Analysis. 7th ed. Prentice-Hall, Englewood Cliffs, NJ.
Gulati, R., Higgins, M.C., 2003. Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. Strateg. Manag. J. 24 (2), 127–144.
Gulati, R., Sytch, M., Tatarynowicz, A., 2012. The rise and fall of small worlds: Exploring the dynamics of social structure. Organ. Sci. 23 (2), 449–471.
Heinsalu, E., Patriarca, M., 2014. Kinetic models of immediate exchange.The. Eur. Phys. J. B 87 (8), 1–10.
Hmieleski, K.M., Corbett, A.C., 2008. The contrasting interaction effects of improvisational behavior with entrepreneurial self-efficacy on new venture performance and entrepreneur work satisfaction. J. Bus. Ventur. 23 (4), 482–496.
Hopp, C., Sondereggger, R., 2014. Understanding the dynamics of nascent entrepreneurship—prestart experience, Intentions, and Entrepreneurial success. J. Small Bus. Manag. http://dx.doi.org/10.1111/isbm.12107.
Inc.Magazine, 2011. http://www.inc.com/inc5000apply/2011/faqs.html (December 14, 2011).
Kaufmann, P.J., Dant, R.P., 1999. Franchising and the discipline of entrepreneurship research. J. Bus. Ventur. 14 (1), 5–16.
Kim, P.H., Aldrich, H.E., Keister, L.A., 2006. Access (not) denied: The impact of financial, human, and cultural capital on entrepreneurial entry in the United States. Small Bus. Econ. 27 (1), 5–22.
Kreiser, P.M., Patel, P.C., Fiet, J., 2013. The influence of changes in social capital on firm-founding activities. Enterp. Theory Pract. 37 (3), 539–568. http://dx.doi.org/10.1111/etap.12039.
Kruschke, J., Aguinis, H., Joo, H., 2012. The time has come: Bayesian methods for data analysis in the organization sciences. Organ. Res. Methods 15 (4), 722–752.
Leitch, C., Hill, F., Neergaard, H., 2010. Entrepreneurial business growth and the quest for a "comprehensive theory": Tilting at windmills? Enterp. Theory Pract. 34 (2), 249–260.
Lichtenstein, B., 2000. Emergence as a process of self-organizing: New assumptions and insights from the study of non-linear dynamic systems. J. Organ. Chang. Manag. 13 (6), 526–544.
Lichtenstein, B., 2014. Generative Emergence: A New Discipline of Organizational, Entrepreneurial, and Social Innovation. Oxford University Press, New York.
Lichtenstein, B., Carter, N., Dooley, K., Gartner, W.B., 2007. Complexity dynamics of nascent entrepreneurship. J. Bus. Ventur. 22 (2), 236–261.
Lockett, A., Wiklund, J., Davidsson, P., Girma, S., 2011. Organic and acquisitive growth: Re-examining, testing and extending Penrose's growth theory. J. Manag. Stud. 48 (1), 48–74.
Lomi, A., Larsen, E., Wezel, F., 2010. Getting there: Exploring the role of expectations and preproduction delays in processes of organizational founding. Organ. Sci. 21 (1), 132–149.
Macy, M., 1991. Chains of cooperation: Threshold effects in collective action. Am. Sociol. Rev. 56, 730–747.
Mandelbrot, B.B., 2009. "New methods of statistical economics," revisited: Short versus long tails and Gaussian versus power-law distributions. Complexity 14 (3), 55–65.
Markman, G.D., Gartner, W.B., 2002. Is extraordinary growth profitable? A study of Inc. 500 high-growth companies. Enterp. Theory Pract. 27 (1), 65.
MATLAB and Statistics Toolbox Release R2013b, The MathWorks, Inc., Natick, Massachusetts, United States.
McKelvey, B., 2004a. Toward a complexity science of entrepreneurship. J. Bus. Ventur. 19 (3), 313–341.
McKelvey, B., 2004b. Toward a 0th law of thermodynamics: Order-creation complexity dynamics from physics and biology to bioeconomics. J. Bioecon. 6 (1), 65–96.
McKelvey, B., 2012. Fixing the UK's economy. In: McGlade, J., Strathern, M., Richardson, K. (Eds.), Complexity in Human and Natural Systems. ISCE Publishing, Litchfield Park.
McMullen, J.S., Dimov, D., 2013. Time and the entrepreneurial journey: The problems and promise of studying entrepreneurship as a process. J. Manag. Stud. 50 (8), 1481–1512.
McMullen, J.S., Shepherd, D.A., 2006. Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. Acad. Manag. Rev. 31 (1), 132–152.
McMullen, J.S., Plummer, L.A., Acs, Z.J., 2007. What is an entrepreneurial opportunity? Small Bus. Econ. 28 (4), 273–283.
Meyer, A.D., Gaba, V., Colwell, K.A., 2005. Organizing far from equilibrium: Nonlinear change in organizational fields. Organ. Sci. 16 (5), 456.
Mitchell, R.K., 2011. Increasing returns and the domain of entrepreneurship research. Enterp. Theory Pract. 35 (4), 615–629.
Mok, A.L., van den Tillaart, H., 1990. Farmers and small businessmen: A comparative analysis of their careers and occupational orientation. In: Donckels, R., Miettinen, A. (Eds.), New Findings and Perspectives in Entrepreneurship. Avebury, Aldershot, UK, pp. 203–230.
Nahapiet, J., Ghoshal, S., 1998. Social capital, intellectual capital, and the organizational advantage. Acad. Manag. Rev. 23 (2), 242–266.
Newbert, S.L., Tornikoski, E.T., Quigley, N.R., 2013. Exploring the evolution of supporter networks in the creation of new organizations. J. Bus. Ventur. 28 (2), 281–298.
Newman, M.E.J., 2005. Power laws, Pareto distributions, and Zipf's law. Contemp. Phys. 46 (5), 323–351.
Ning, D., You-Gui, W., 2007. Power-law tail in the Chinese wealth distribution. Chin. Phys. Lett. 24 (8), 2434–2436.
O'Boyle, E.H., Aguinis, H., 2012. The best and the rest: Revisiting the norm of normality of individual performance. Pers. Psychol. 65, 79–119.
Pareto, V., 1897. Cours d'Economie Politique. Rouge, Paris.
Parker, S.C., 2009. The Economics of Entrepreneurship. Cambridge University Press, New York, NY.
Peteraf, M.A., Barney, J.B., 2003. Unraveling the resource-based tangle. Manag. Decis. Econ. 24 (4), 309–323.
Pierce, J.R., Aguinis, H., 2013. The too-much-of-a-good-thing effect in management. J. Manag. 39 (2), 313–338.
Pimm, S., 1982. Food Webs. 2nd edition. University of Chicago Press.
Plummer, L.A., Haynie, J.M., Godesiabois, J., 2007. An essay on the origins of entrepreneurial opportunity. Small Bus. Econ. 28 (4), 363–379.
Poole, M.S., Van de Ven, A.H., Dooley, K., Holmes, M.M., 2000. Organizational change and innovation processes: Theory and methods for research. Oxford University Press, New York.
Powell, T.C., 2003. Varieties of competitive parity. Strateg. Manag. J. 24 (1), 61–86.
Preston, F., 1948. The commonness, and rarity, of species. Ecology 29, 254–283.
Prigogine, I., 1955. Thermodynamics of Irreversible Processes. Thomas, Springfield.
Prigogine, I., Stengers, I., 1984. Order Out of Chaos. Bantam, New York.
Rand, W.M., Rust, R.T., 2011. Agent-based modeling in marketing: Guidelines for rigor. Int. J. Res. Mark. 28 (3), 181–193.

Reynolds, P.D., 2007. Entrepreneurship in the United States. Kluwer, Boston.
Samuelsson, M., Davidsson, P., 2009. Does venture opportunity variation matter? Investigating systematic process differences between innovative and imitative new ventures. Small Bus. Econ. 33 (2), 229–255.
Shadish, W., Cook, T., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. 2nd ed. Wadsworth Publishing.
Shane, S., 2008. The Illusions of Entrepreneurship: The Costly Myths that Entrepreneurs, Investors, and Policy Makers Live By. Yale University Press, New Haven and London.
Shane, S., Stuart, T., 2002. Organizational endowments and the performance of university start-ups. Manag. Sci. 48 (1), 154–170.
Shepherd, D.A., Wiklund, J., 2009. Are we comparing apples with apples or apples with oranges? Appropriateness of knowledge sccumulation across growth studies. Enterp. Theory Pract. 33 (1), 105–123.
Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 42 (3/4), 425–440.
Simon, H.A., 1962. New developments in the theory of the firm. Am. Econ. Rev. 1–15 (May).
Simon, H.A., 1968. On judging the plausibility of theories. In: Staal, v. Rootselaar (Ed.), Logic, methodology and philosophy of sciences III. North-Holland, Amsterdam.
Slevin, D.P., Covin, J.G., 1997. Strategy formation patterns, performance, and the significance of context. J. Manag. 23 (2), 189–209.
Solomon, S., Richmond, P., 2001. Power laws of wealth, market order volumes and market returns. Physica A 299 (1), 188–197.
Song, C., Qu, Z., Blumm, N., Barabási, A.-L., 2010. Limits of predictability in human mobility. Science 327 (5968), 1018–1021.
Sornette, D., 2006. Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization, and Disorder: Concepts and Tools. Springer-Verlag, Berlin.
Spector, R., 2002. Amazon.com: Get Big Fast. Harper Business Publishing, New York.
Steffens, P., Davidsson, P., Fitzsimmons, J., 2009. Performance configurations over time: Implications for growth- and profit-oriented strategies. Enterp. Theory Pract. 33 (1), 125–148.
Van de Ven, A.H., Engleman, R.M., 2004. Event-and outcome-driven explanations of entrepreneurship. J. Bus. Ventur. 19 (3), 343–358.
van Stel, A., Storey, D., Thurik, R., 2007. The effect of business regulations on nascent and young business entrepreneurship. Small Bus. Econ. 28, 171–186.
Venkataraman, S., 1997. The distinctive domain of entrepreneurship research: An editor's perspective. In: Katz, J., Brockhaus, J. (Eds.), Advances in Entrepreneurship, Firm Emergence, and Growth vol. 3. JAI Press, Greenwich, pp. 119–138.
West, B., Deering, B., 1995. The Lure of Modern Science: Fractal Thinking. World Scientific, Singapore.
Wiklund, J., Shepherd, D.A., 2003. Aspiring for, and achieving growth: The moderating role of resources and opportunities. J. Manag. Stud. 40 (8), 1911–1941.
Wiklund, J., Shepherd, D.A., 2011. Where to from here? EO-as-experimentation, failure, and distribution of outcomes. Enterp. Theory Pract. 35 (5), 925–946.
Wilensky, U., 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL (http://ccl.northwestern.edu/netlogo/).
Zanini, M., 2008. Using power curves to assess industry dynamics. McKinsey Q. 1–5 (November).