STRATEGIC IMPLICATIONS OF POWER-LAW DISTRIBUTIONS IN THE CREATION AND EMERGENCE OF NEW VENTURES: POWER-LAW ANALYSES IN THREE PANEL STUDIES

G. Christopher Crawford
University of Louisville, Christopher.Crawford@Louisville.edu

Bill McKelvey
University of California

Recommended Citation
Frontiers of Entrepreneurship Research: Vol. 32: Iss. 12, Article 1.
Available at: http://digitalknowledge.babson.edu/fer/vol32/iss12/1
STRATEGIC IMPLICATIONS OF POWER-LAW DISTRIBUTIONS IN THE CREATION AND EMERGENCE OF NEW VENTURES: POWER-LAW ANALYSES IN THREE PANEL STUDIES

G. Christopher Crawford, University of Louisville, USA
Bill McKelvey, University of California, Los Angeles, USA

ABSTRACT

Whereas the creation and emergence of high-growth firms are central topics in entrepreneurship research, senior scholars lament the absence of a comprehensive theory to explain and predict this rare phenomenon. Counter to the assumptions of normal distributions and independent observations, we develop hypotheses that entrepreneurial outcomes are distributed according to a power-law, where interdependence and extreme outliers have a disproportionate and co-evolutionary effect on the environment. Tests on six different outcome measures within the Panel Study of Entrepreneurial Dynamics II, Kauffman Firm Survey, and Inc. 5000 find that power-law distributions pervade the domain. We use a complexity science perspective to explain these results, which provide an empirically validated foundation for future theory-building efforts.

INTRODUCTION

The study of high performance firms is of primary importance to entrepreneurship scholars, yet sampling and understanding these organizations is elusive (Acs & Szerb, 2007; Aldrich & Ruef, 2006; Davidsson, 2005). On one hand, high potential firms backed by venture capital (VC) funds are very rare. While there is a sizable estimated population of 7.4 million nascent enterprises in the United States (Reynolds & Curtain, 2008), VCs fund about 820 companies per year – only about 0.03% of all new ventures. In the last 25 years, VC-backed firms like Microsoft and Federal Express and Google have generated 12% of the sales and employment in the U.S. (Shane, 2008) – these firms influence Schumpeterian change in the economy. Thus, a VC-backed firm would be one example of an “extreme” outcome, an outlier for both its rarity and its potential scale of influence on the environment.

On the other hand, however, little is known about how these firms scale up into extreme outcomes. First, the typical venture is much smaller in size, capital, and scope. Based on representative samples, 93% of all new firms are founded by one person or a family-based team (Ruef, Aldrich, & Carter, 2003); the median new firm only needs $15,000 to effectively begin operations; and 90% of founders expect their new business to have little influence the market – in fact, only 19% of new venture founders would like their organization to grow as large as possible (Reynolds, 2007). The outcomes of new ventures are equally dreary, where 60% of start-ups cease to exist after five years and 90% of new firms never grow beyond their first year (Headd, 2003). Thus, it appears that healthy, growing entrepreneurial firms are also outliers.

Whereas Axtell (1999) suggests that the “average” firm does not exist, and Shane (2008) adds that the “average” new firm fails, scholars continue to build organizational theory focusing on the average, using Gaussian statistical techniques, and assuming normal distributions of outcomes.
As one example, for nascent firms, Reynolds (2007:141) finds that “Both the average requirement to start a business ($1.4 million), and that expected to be provided by the nascent entrepreneur ($800,000) are dramatically skewed by extremely high values.” The average, then, does not help scholars understand the true dynamics of the environments in which these firms operate, nor does it provide any instructive relevance for practitioners. We develop hypotheses to suggest that these skewed outcomes are consistent with the characteristics of power-law distributions, where the size and scope of a venture are inversely proportional to their probability of occurrence.

Using three theoretically relevant datasets and rigorous maximum-likelihood methods from physics, we find that power-law distributions are consistent across time, at multiple units and levels of analysis. Thus, we propose that these findings provide an empirically validated basis for a comprehensive theory of venture growth. Our study makes a significant contribution to the entrepreneurship literature by providing a highly generalizable framework that explains both established and conflicting findings, successfully predicts aggregate regularities, and explicates new macro-level causal relationships; our power-law analyses make a methodological contribution to organizational science and provide normative guidance to research and practice.

THEORETICAL AND EMPIRICAL CONTEXT

Recent summaries of the literature on entrepreneurial growth exhibit appear exasperated at the domain’s lack of a theory about firm growth. As evidence:

“Attempts to provide solutions to how explained variance can be improved, unfortunately, are largely futile (McKelvie & Wiklund, 2010: 227);”

“Consequently, there has been disquiet concerning a perceived lack of well-founded knowledge about the causes, effects, and process of growth (Leitch, Hill, & Neergaard, 2010: 250);

“We have observed that theoretical predictions have been of limited use in understanding the growth of firms, if not downright misleading…Empirical research into firm growth has nonetheless come up against some major obstacles. The main message that seems to emerge is that growth is largely a random process. (Coad, 2007:59).”

Indeed, some scholars suggest that the field is “tilting at windmills” – referencing Cervantes’ misguided and hapless Don Quixote – in an attempt to develop a comprehensive theory of growth (Leitch et al., 2010). From the beginning, entrepreneurship scholars have identified the difficulty in building theory for myriad reasons. The survival biases, problems with data collection, small sample case studies, unobserved heterogeneity in the phenomena, ontological and epistemological differences about the ephemeral nature of opportunities, interrelated complexities of social networks, and endogenous individual motivations - all inherent in almost all aspect of entrepreneurship research – provide numerous potential obstacles for scholars (Bull, Thomas, & Willard, 1995; Davidsson, 2005; Katz & Gartner, 1988). These obstacles become exacerbated without robust empirical validation and subsequent lack of replication studies. As well, traditional statistical analyses for hypothesis testing assume normal distributions of data and additive, linear relationships, which may not be appropriate to describe and understand the potentially non-linear phenomena of entrepreneurship (Delmar & Shane, 2006). If we – as researchers, entrepreneurs, policy makers, or teachers – accept that growth is random, then all of our efforts may be better spent elsewhere.
Instead of giving up, however, Coad (2007), in his conclusion, draws from Herbert Simon’s (1968) seminal theory-development instruction, and recommends that scholars adopt a Simonian methodology. Here, facts 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 the “striking empirical regularities” in the data, regularities where “standard statistical tests of hypotheses [are] inappropriate (443).” Simon was referring to the ubiquity of power-law distributions – the type about which he wrote his article, “On a Class of Skew Distribution Functions” (1955). Power-law relationships link two variables in the form of $F \sim N^{-\alpha}$, where $F$ is frequency, $N$ is rank (the variable) and $\alpha$, the exponent, is constant (Clauset, Shalizi, & Newman, 2009; Sornette, 2006). In contrast to the traditionally assumed normal (Gaussian) distribution of new venture outcomes (e.g., firm revenue, number of employees), where events are completely independent and identically distributed, a power-law (Paretian) distribution acknowledges the fundamental interconnectedness and interdependence of events (Andriani & McKelvey, 2009). When graphed on regular scales, the power-law distribution looks like Figure 1a; as stylized in Figure 1b, a well-formed rank/frequency distribution appears as a downward sloping straight line when plotted on log-log axes.

Though rare, outcomes in the long tail of a power-law distribution – most easily recognized as the largest circle at the bottom right of Figure 1b – is of disproportionate influence on the entire system. In entrepreneurship, for example, the non-linear outcomes beyond the dotted threshold line might include a new firm that generates 1000 new jobs, $100M in revenue, or a 40,000% increase in growth. Instead of viewing outliers as anomalies that reduce the statistical significance of future analyses, a power-law distribution emphasizes the potential co-evolutionary influence of these data. These unique events represent something fundamentally new and qualitatively different; the distinctiveness of the outcome and its complex effects make it relevant to both practitioners and scholars (Benbya & McKelvey, 2011; Rivkin & Siggelkow, 2007). Studies show that Gaussian techniques are not applicable to firms in all regions of the distribution (Boisot & McKelvey, 2010). Specifically, linear analyses cannot explain, the outcomes of both self-employed founders of Mom & Pop retail stores ($N=17$ million) and the non-linear, extreme event of Wal-Mart ($N=1$). If data are distributed in such a manner, it would seem logical to conceive that theory building (and subsequent theory testing efforts with linear methods) would prove to be difficult. Indeed, studies show that Gaussian assumptions and methods to explain these highly skewed distributions can lead to incorrect conclusions, misspecified theories, and misleading normative recommendations – all of which reduce the credibility of scholarly research (O’Boyle & Aguinis, 2011).

Power-law distributions have been found in U.S. firm size (Axtell, 1999), overall industry growth (Stanley et al., 1996), industry sectors and firm growth rate (Zanini, 2008), network structure (Barabási, et al., 2002), market share prices (Kohli & Sah, 2006), firm innovations (Poole, Van de Ven, Dooley, & Holmes, 2000), stock returns (Linn & Tay, 2007), and competitive performance advantages (Powell, 2003), among many others. However, with the exception of one conceptual conference abstract, entrepreneurship scholars have not investigated power-law distributions with enough depth to identify or explain their existence, propose how they emerge, and study their overall effects. Power-law distributions, then, appear important to similar domains of interest to those in the study of entrepreneurship. Explaining and predicting these distributions in manner that is relevant for both practitioners and scholars, however, will require selecting the appropriate outcomes of interest.
HYPOTHESIS DEVELOPMENT

Following Simon’s (1968) framework, we start somewhat atheoretically in an attempt to understand the data and answer the question, “Are outcomes in the domain of entrepreneurship power-law distributed?” If they are, explaining how and why they emerge will provide a more solid foundation for building theory on new venture performance. We first identify outcomes relevant to a comprehensive theory of growth; then, after testing, we use a complexity science perspective to explain the emergence of these distributions.

Outcomes

Entrepreneurial outcomes have been measured by revenue, employees, growth, profit, performance, success, economic well-being, survival, market-share, amount of venture capital funding, IPO underpricing, return on investment, and return on equity, among others (Van de Ven & Engleman, 2004; Wiklund, Davidsson, Audretsch, & Karlsson, 2011). Some of these are more relevant and measurable than others. If the entire population of entrepreneurs is considered, not all outcomes will be equally important to theory or practice.

As multiple studies have shown, most businesses are founded with the intention to live a certain lifestyle – not to grow (Carter, Gartner, Shaver, & Gatewood, 2003). Thus, measures based on market-share, venture capital, or public offerings outcomes are not relevant to all practitioners. One of the perks of a lifestyle-based firm is the ability to write off many business-related activities when filing a tax return. Though academic research on the topic is limited, anecdotal evidence suggests that many returns are creatively configured to avoid showing profit as a means of paying less tax. Thus, measures of profit return on equity, and return on investment may be of questionable reliability and validity. And, while profit may be stringently measured by auditors of larger firm, research shows that high growth is not related to profitability (Markman & Gartner, 2002). Whereas firm survival has been used as a valid outcome measure, it is not within the scope of this study’s primary interest – the extreme, non-linear outcomes in the tail of the distribution that may cause co-evolutionary effects. It may be important to note, however, that new firm exit from the market has been empirically shown to follow a power-distribution (DiGuilmi, Gallegati, & Ormerod, 2004). Finally, completely subjective measures like “success” or “quality of life” are difficult to conceptualize and assess in larger companies or on new venture teams with multiple members.

Both revenue and employees, then, have the potential to be theoretically and practically relevant. Each measure can be reliably assessed in all firms, regardless of size or ownership structure, and each has the potential – if large enough – to influence co-evolutionary change in the environment. Growth in revenue and employees is also measurable, but may not be immediately relevant to the majority of entrepreneurs who don’t begin with the expectation to grow. However, as we elucidate in the following section, we implicitly assume that all new ventures have the potential to grow, even if the founder begins the process with limited expectations and scarce resources. Thus, we propose that growth measures for both revenue and employees are generalizable to the entire domain. In addition, how growth is measured can implicitly bias one organizational form over another. For example, relative growth (measured as a percent) favors firms that start with smaller revenue or employee figures, while absolute growth (measured as difference in amount) favors larger firms (Delmar et al., 2003). For ease of exposition and to be consistent with practitioner-based
understanding, we use the term *growth* to indicate relative increase percent and *gain* to indicate absolute increase in amount. Subsequently, six outcomes appear of interest to the domain and to practice: revenue, employees, revenue growth, revenue gain, employee growth, and employee gain. Based on 1) the difficulties in building theory as suggested at the beginning of this paper, 2) the influence and effect of outliers in the domain, and 3) the ubiquity of power-law distributed phenomena in similar domains, we put forth the following hypotheses:

- **Hypothesis 1:** Revenue of emerging firms is power-law distributed.
- **Hypothesis 2:** Employees of emerging firms are power-law distributed.
- **Hypothesis 3a+b:** Revenue growth and gain of emerging firms are power-law distributed.
- **Hypothesis 4a+b:** Employee growth and gain of emerging firms are power-law distributed.

**Method**

The research focuses on understanding venture outcomes in the domain – from the initial stage of venture organizing to the potential co-evolutionary effects on the larger environment. To answer the research question, we chose an exploratory, quantitative design with three theoretically relevant samples (including two longitudinal panel studies), each at a different stage of venture emergence. Consistent with complexity science epistemology and its foundational tenets of self-organized systems and bottom-up emergence (Cilliers, 1998; Lichtenstein, 2011), all firms in the study are privately owned, so that any initial expectations for future growth originate from the founder, not from a corporate mandate.

**Samples and Procedures**

We analyze three samples. All three samples include the entire spectrum of business types and industries, and all data are collected from similar time periods to mitigate potential cohort affects in our analysis (Aldrich & Ruef, 2006; McKelvey & Aldrich, 1983). As well, each sample is large enough to allow our theory to be more generalizable to the entire domain of entrepreneurship.

First, data collected in the Panel Study of Entrepreneurial Dynamics II (PSED) was a replication of the PSED I, upon which more than 100 peer-reviewed journal articles have been published (Davidsson, 2006). The PSED II’s baseline data collection started in 2005 as a representative sample of 31,845 adults in the contingent United States. Information was collected from 1,214 subjects who affirmatively answered the question, “Are you, alone or with others, currently trying to start a new business, including any self-employment or selling any goods or services to others?” to assess the level of embryonic entrepreneurial activity (Reynolds, 2007). Whereas the PSED is a representative sample of the nascent entrepreneurial population (where every adult in the United States had a non-zero probability of being contacted), the next two samples would be considered theoretically relevant samples (Reynolds & Curtain, 2008; Shadish, Cook, & Campbell, 2002) of ongoing businesses and extreme outcomes, respectively.

Second, the Kauffman Firm Survey (KFS) started with a random sample of 32,469 businesses from a Dun & Bradstreet list which identified almost 250,000 firms that started operations in 2004. A start-up includes any independent business that was established by a single person or a team, or purchased as an existing business or new franchise. Businesses were excluded if they had a federal identification number, income on Schedule C, or paid federal Social Security or state
unemployment insurance or taxes prior to or after 2004. The overall sample includes 4,928 firms. Using a stratified sampling methodology that was weighted toward high-technology firms, there are 2,034 high-technology and 2,894 non-high-tech businesses (DesRoches, Robb, & Mulcahy, 2009). We perform analyses within the National Opinion Research Center (NORC) enclave, a data repository housing the restricted-access KFS microdata that provides a higher level of refinement compared to the data available to the general public.

Third, the Inc. 500® (INC) self-selected sample annually collects and publishes revenue, employee, and three-year growth rate data on the 5000 fastest-growing, privately held, for-profit companies in the United States (Markman & Gartner, 2002). The 500 companies with the highest growth rate are written up in Inc. Magazine, and the publisher’s website provides information on the top 5000 companies. Whereas there is a nominal $100 application fee, the exposure and publicity that come from being on the list – including more than 25 million views on the publisher’s website and in print editions – suggest that these firms represent the U.S. population of private, high-growth firms. As well, since a data certification from an internal auditor/controller must accompany each firm’s application, and since certified public accountants check and verify the data prior to publication, this sample possesses a high degree of both reliability and validity.

To be included, firms must have generated some revenue by March 31, 2007, at least $100,000 in revenue in 2007, and at least $2M in revenue in 2010 (Inc.Magazine, 2011). Based on these criteria, the firms in this 2011 cross-sectional sample represent an oversampling of the very highest performing firms in the KFS. The high face validity and theoretical generalizability of the INC complement our efforts to build a comprehensive theory of growth new ventures: companies from all industries and sectors are included; organizations are privately owned; and, the large sample size provides relevant data on extreme outcomes at the macro-level.

Outcome Variables

Consistent with recommendations from entrepreneurship and complexity science scholars, we use multiple measures to understand the heterogeneous contexts of organizations and the individuals who found them (Cilliers, 1998; Delmar, Davidsson, & Gartner, 2003). Similarly, we measure data at the venture’s initial conditions (PSED Year 1), as well as over time (Year 5).

The first measure of outcome with the potential to create co-evolutionary effects on the environment is the number of Employees (Flier, Van Den Bosch, & Volberda, 2003). We are agnostic as to whether the employees are full- or part-time. Since there is intuitive reason to believe that an unemployed individual takes a job at an entrepreneurial venture, the new influx of income could influence that person’s environment, regardless of whether that job is full- or part-time. Thus, we use the question, “Right now, how many people, not counting the owners but including exclusive subcontractors, are working either full or part-time for this new business?” for PSED Employees, from Wave B (Yr1) and Wave F (Yr5), the new organization’s fifth year of operation. Similarly, KFS Employees uses data from the fourth follow-up (covering the business’s fifth year), where the respondent answered the question, “Not counting owner(s), on December 31, 2008, how many people worked for your company?” Finally, we measure INC Employees as the reported number of employees.

PSED Revenue is data from Wave B (Yr1) and Wave F (Yr5); KFS Revenue uses data from the fifth year of collection; INC Revenue is the self-reported amount for 2010. Unless otherwise
specified, we delete cases piece-wise if any of the responses were either “No response” or “I don’t know.” Consistent with Markman & Gartner (2002) and Inc. Magazine, we calculate Revenue Growth and Employee Growth using formula (1) and Revenue Gain and Employee Gain using formula (2) for all three studies.

\[
\frac{((\text{Year 5} / \text{Year 2}) - 1) \times 100}{(\text{Yr 5} - \text{Year 2})}
\]  

(1)  

(2)

**Data Analysis**

Empirically, few phenomena adhere to a power law over all values; instead, the power-law most often applies for values greater than some minimum. Here, only the datapoints in the tail of the distribution follow a power law; if viewed through the lens of traditional statistics, these points are considered outliers, freaks, anomalies, mistakes, black swans, or non-linear outcomes (Taleb, 2007) - the sciences of complexity, however, view these outliers as the ones that matter, the ones that influence all the surrounding points (Rivkin & Siggelkow, 2007). We use MATLAB (R2010a) software and follow the protocol and techniques for calculating power-law model fit, as described by Clauset, Shalizi, & Newman (2009). The authors suggest a three-step method: 1) estimate the parameters for the scaling exponent \(\alpha\) and the minimum value in the distribution that exhibits power-law behavior \(x_{\text{min}}\); 2) calculate the standard errors of the estimates in step one using a semi-parametric bootstrap analysis; and, 3) calculate the goodness-of-fit between the data and the power-law with a non-parametric bootstrap analysis. Below, we briefly describe these procedures, and encourage the reader to explore the appendices in Clauset, et al. (2009) for the mathematical computations which underlie each technique.

Referring back to figure 1b: the parameter \(\alpha\) is the slope of the power-law tail; \(x_{\text{min}}\) is the dotted line that separates the Gaussian world from the Paretian world; and, the Kolmogorov-Smirnov (KS) goodness-of-fit test assesses how far away the circles are from the slope’s line. The \(\alpha\) and \(x_{\text{min}}\) estimates are calculated with the `plfit.m` MATLAB script found at [www.santafe.edu/~aaronc/powerlaws](http://www.santafe.edu/~aaronc/powerlaws). Once the scaling parameter and lower-bound threshold of the distribution are estimated, we calculate the uncertainties of each using the `plvar.m` script. This step requires a semi-parametric approach that first generates 10,000 random numbers, distributed according to a power-law with the estimated \(\alpha\) and \(x_{\text{min}}\) from the previous step, and selects the value that generates the minimum difference between the observed data and the estimated model. Finally, the script `parplpva2.m` assesses the relative fit of the model. The model’s goodness of fit, calculated with the K-S test, indicates how different the maximum likelihood estimates for all of the model parameters are in comparison to the actual data. Values below 0.10 are considered very good fit (White, Enquist, & Green, 2008). We do not compare the power-law fit of our data with a fit to every competing distribution (of which there is an infinite number). Of primary interest is to show that the data are not normally distributed, and that a power-law distribution provides a more accurate representation of actual outcomes.

**Results**

In Table 1, we show results from the fitting of a power-law form to each of the outcome variables. It is interesting to note that a normal distribution has a skewness of 0, and the data is considered
skewed if that number is above 3 (Greene, 2007). Only one of the 24 models we ran (PSED Employees Yr 2) was within the normal range; and, since power-laws are a result of underlying connectivity dynamics and take time to emerge (Bak & Chen, 1991), this is not a surprising result. As well, this variable is the only one with a non-significant K-S test, overwhelmingly supporting our power-law hypotheses. Most importantly, this validates the foundational premise that the data are not distributed normally and that non-linear methods like those presented here may provide a theory-based alternative means of analysis.

We also point out that PSED Revenue Yr 2 is significant. How could this be? Though these distributions take time to emerge, the fractal nature of these distributions suggests that when power-laws exist at one level, they also exist at preceding levels (Boisot & McKelvey, 2010). Revenue, in contrast to Employees, could be more easily captured by a nascent venture where, for example, a wage employee could be enticed to form her own company when a potential stakeholder promises payment for services not provided by the existing firm. More interestingly, though, is the idea that venture inputs – like human, social, intellectual, and financial capital – might be power-law distributed. This could be an intriguing question for future research.

Analyses and Assessments: Tipping Points and Potential Scalability

Our analysis diverges somewhat from Andriani & McKelvey’s (2009) suggestion that that distribution assessment should only focus on the power-law tail, not the entire distribution. We assess the entire distribution of each variable for theoretical and practical reasons. Theoretically, complexity science scholars use the concepts of phase transition, threshold, bifurcation point, critical value, or tipping point to identify the point where a system changes from an additive, linear state into a multiplicative, non-linear state; this state is qualitatively different than the previous (Prigogine, 1980) – think of water reaching a boiling point at 212°F. Once this threshold has been crossed, the system cannot return to its previous state without some form of dissonance – what McKelvey (2004b:87) calls “irreversible order-creation conditions for lower scale dissipative structures”. Within figure 1b, this point is the dotted line that separates the Gaussian and Paretian regions. Practically, firms beyond this point are operating in a much more interdependent, highly scalable, non-linear environment and, thus, have the potential to influence outcomes at a higher level; firms below this point have little ability to cause non-linear outcomes within their existing environment.

We propose the region within the standard error of the tipping point is what complexity science scholars have termed the region of emergent complexity. Here, organisms are more likely to survive because they have a solid enough foundation of resources, yet maintain enough flexibility to change when environmental perturbations dictate (Andriani & McKelvey, 2009). In the fifth year of the PSED, the critical threshold (x_min) for number of employees is 2, with a standard error of 1. How could this have non-linear effects, and how might this facilitate a venture’s survival? Even with one employee, a founder’s efforts can become more focused and more scalable. If it is an administrative or labor employee, the founder has the potential to put more attention on activities – like sales or networking – that generate immediate revenue or procure future sales contracts. Present and future cash flows can pay bills, fund expansions, upgrade facilities, among a plethora of other things that could increase the legitimacy and probability of survival for a new venture. Additionally, in small work environments, a founder’s relationship with the employees become interdependent (Wiklund, Davidsson, & Delmar, 2003), where any decision to expand...
or cease operations has the potential to influence everyone in company. Where there are between one and three employees (the standard error of the \( x_{\text{min}} \)), it also seems more likely that a founder would know the relatives of those under her employ and, thus, important decisions may have a cascading influence on the local environment. When a business has no employees, however, there is little opportunity for non-linearity and little possibility for adaptive efficacy. Without a single employee, either to pick up the slack (e.g., continue administrative work if the founder is sick for a few days) or to provide some additional slack (e.g., answer a company phone call while the founder greets a new client at the door), all of a founder's attention must be addressed at proximate tasks and her ability to efficaciously adapt is reduced. While having two or three employees may not appear to be an 'extreme' event, it is important to note that the median and the mode of the distribution are zero.

Consistent with seminal and recent complexity science research (Bak & Chen, 1991; Benbya & McKelvey, 2011), outcomes beyond the tipping point also present the possibility of an unexpected negative extreme event. For number of employees, the area beyond the critical threshold and beyond the region of emergent complexity is four. Though these employees can help scale the business, additional venture resources must dissipated to sustain them (Davis, Eisenhardt, & Bingham, 2009). If a venture has a non-linear number of employees and only linear revenue (the mean revenue in the fifth year of the PSED is only $35,000), for example, it may be financially impossible to support those employees – a “decreasing share of the accessible energy differentials” for the founder (Levie & Lichtenstein, 2010:334), which could result in lay-offs or firm closure; that is, unless there is a sufficient base of alternate venture resource endowments (e.g., external funding, a cache of cash), an extreme negative event is likely. A more appropriate benchmark for revenue might be the PSED's fifth year's non-linear threshold of $600,000, an amount where both the entrepreneur and the employees could conceivably have the opportunity for adequate compensation.

As a venture emerges to a conceptually more established, dynamic state (Levie & Lichtenstein, 2010) in the KFS, there is an increased potential for non-linear cascades of environmental influence. In the fifth year of operations, the critical threshold for number of employees is 35, and the standard error is ± 11. Here, there appears to be a different underlying dynamic (as evidenced by a significantly different alpha), and the influence of new business operations goes beyond the family relationships described above. This would seem to confirm the results of the Wiklund, et al (2003) study of Swedish companies with more than 20 employees, which finds that the primary component of a manager’s willingness to grow is his concern about the potential reduction of a “family-like” relationship with employees.

In the region of emergent complexity between 24 and 46 employees, those relationships may still come into play. However, if a new business creates more than 46 jobs, though the strength of the tie between a founder and each employee may be reduced, the influence of the firm's outcome cascades to a higher environment. New job creation of this magnitude could affect a local economy and potentially revitalize a small town; as well, a firm of this size may extract premium offers from corporations looking to acquire it – all of these could be considered positive extreme event. It is also important to note that outcomes beyond the threshold also present the possibility of an unexpected negative extreme event, depending on the perspective of the subject under study. Thus, as tie strength and interdependence are decreased between a founder and each employee, there is less relational embeddedness (Hite, 2005) and an increased incentive for the founder to actually sell the firm, thereby indirectly affecting the employees.
Similar numbers and increasing cascades of influence for both positive and negative extreme events exist for firms at an advanced state of emergence in the INC. Here, the critical threshold begins at 240 employees – enough to effectively influence a city’s tax infrastructure and attract employees from outside the proximate geographical region. It is interesting to note that Walmart – a company that has survived numerous lawsuits about gender discrimination and fair wages, has withstood attempts at unionization with nearly 2,100,000 employees and 8,970 locations, and has been an exemplar case-study for both the revitalization and the destruction of many small towns – has an average of 234 employees per store (en.wikipedia.org/Walmart, accessed 10/24/2012).

**Implications and Conclusion**

The norm of normality appears to be an antiquated myth. In all three studies, we find statistically significant power-law distributions and identify tipping points in the data, where outcomes transition from additive/linear to multiplicative/non-linear. Entrepreneurial firms with outcomes beyond thresholds produce co-evolutionary effects on the environment, like those first envisioned by Schumpeter. Though power-law distributions have been shown to be ubiquitous in management and strategy research, this study is the first to approach these distributions in entrepreneurship. Most important, however, the significance at multiple levels of outcomes suggest, like Simon (1968), that the findings, “even if approximate, cannot be accidental, but must reveal underlying lawfulness (443).” Indeed, the power-law dynamics of growth pervade the domain, and the findings have significant implications for theory, practice, pedagogy, and policy.

**Implications**

Our findings provide a starting point for a more cohesive theory of entrepreneurial growth that can lend insight to practice and research. Practitioners can use tipping points as benchmarks to either enhance survival or to facilitate growth. Future theoretical development will need to account for power-laws in the data and examine the mechanisms that drive their emergence. Alvarez and Barney (2007) call for a framework that that subsumes discovery and creation theories by encompassing the nature of three components: the decision-making context, opportunities, and entrepreneurs. Power-law distributions exhibit linear and non-linear data, suggesting that decisions are made in environments of both Gaussian risk and Paretian (i.e., Knightian) uncertainty. Similarly, new venture opportunities are plentiful and easily discovered (e.g., a “For Sale” sign in an existing pizzeria) in the body of the distribution and can be created according to the idiosyncratic resources and interactions in the tail of the distribution. Finally, since outcomes in entrepreneurship are power-law distributed, complexity science proffers that venture inputs are likely to be similarly skewed, suggesting that there are both ex ante differences between entrepreneurs and non-entrepreneurs and that variation emerges ex post, after founders interact with the environment.

Given that practicing entrepreneurs found their businesses in environments that are power-law distributed, our theories should reflect a similar context and engage both regions of the distribution. For practitioners, understanding the new venture’s position in relation to the tipping points of the distribution of outcomes will be vital. Most businesses don’t want to grow, so counseling for the venture’s sustainability is more appropriate. Those with linear expectations will improve their probability of survival by enhancing the efficiency of their existing processes and by systematically searching for ways to further embed their venture within existing strong-tie
relationship networks (i.e., “doing things right”); however, to grow, these ventures will need to improve the effectiveness of their processes by interacting with an expanded weak-tie network and learning from environmental feedback (i.e., “doing the right things”). Similarly, founders with non-linear growth expectations will benefit most by doing things that enhance the legitimacy and resource endowments of the new venture: forming a diverse team, developing proprietary processes, and importing enough financial capital to provide a flexible foundation and the ability to capture new opportunities as they emerge.

For pedagogy, most MBA instruction is already implicitly focused around extreme (both positive and negative) outcomes. The findings presented here suggest the need for a strategically holistic curriculum – one that is focused on understanding a venture’s internal scalability relative to the complexity and potential extreme outcomes at all levels and units of analysis. In power-law environments niche opportunities may be discovered to avoid competing head-on with the most powerful competitors; in environments that are not power-law distributed, companies with sufficient resources may most successfully compete by creating opportunities with potential stakeholders that are path-dependent and inimitable.

Policy interventions, in the form of tax breaks, business incubators, or grants, are primarily instituted as a means of generating positive externalities (e.g., spurring innovation or creating new jobs). However, only those few new firms beyond critical thresholds will influence the market in such a way. Thus, the efficacy of interventions may be enhanced by providing incentives for firms 1) to begin the venture with non-linear inputs, like a diverse team or a substantial equity investment, and/or 2) to achieve outcomes beyond the tipping point of the distribution.

Conclusion

The significant power-law findings in this study lend support for the development of a scale-free theory of entrepreneurial performance. A system is scale-free when there is no single node that can represent the “average” of the entire distribution (Boisot & McKelvey, 2010) – this would be especially true in entrepreneurship. When power-laws exist, a scale-free theory suggests that they emerge as a result of one generative mechanism; here, one explanation can reliably characterize and predict all outcomes in a domain (Andriani & McKelvey, 2009; Simon, 1968). What could this mechanism be in entrepreneurship? Future studies exploring this topic will be interesting, indeed.

Power-law distributions are found in earthquakes, where tens of thousands of small and insignificant quakes occur regularly, and very few quakes of significant magnitude occur every 150 years. If building codes – normative rules to improve the stability and longevity of the structures – were based on averages and normal distributions, they would be foolhardy and detrimental to long-term performance; therefore, codes are designed around the extreme. In a similar way, this study finds that entrepreneurial outcomes are most accurately described with power-laws; as scholars, the extremes should also guide our theory-building efforts and insightful suggestions for practice. In sum, the field’s reliance on the “average” may hinder theory development to all stakeholders. Without a wholesale a paradigm shift – acknowledging that the norm of normality does not exist – the domain will continue to tilt at windmills.

CONTACT: Christopher Crawford; Christopher.Crawford@Louisville.edu; (T) 513-295-8717.
ACKNOWLEDGEMENTS

We would like to thank the Ewing Marion Kauffman Foundation for its financial assistance with some portions of this research.

REFERENCES


---

**Figure 1:**

a) Rank/Frequency Distribution on Linear Scales;

b) Distribution Plotted on Log-Log Scales
<table>
<thead>
<tr>
<th>Outcomes</th>
<th>n</th>
<th>med</th>
<th>mean</th>
<th>skew</th>
<th>s.d.</th>
<th>max</th>
<th>$x_{\text{min}}$</th>
<th>$\alpha$</th>
<th>K-S</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employees</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED - Yr2</td>
<td>54</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>7 ± 1</td>
<td>3.50 ± 0.48</td>
<td>0.43</td>
</tr>
<tr>
<td>PSED - Yr5</td>
<td>68</td>
<td>0</td>
<td>30</td>
<td>12</td>
<td>122</td>
<td>1,500</td>
<td>2 ± 1</td>
<td>1.73 ± 0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>KFS</td>
<td>1,462</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>14</td>
<td>320</td>
<td>22 ± 7</td>
<td>2.56 ± 0.21</td>
<td>0.07</td>
</tr>
<tr>
<td>INC</td>
<td>5,000</td>
<td>51</td>
<td>364</td>
<td>40</td>
<td>3,769</td>
<td>194,000</td>
<td>244 ± 69</td>
<td>1.95 ± 0.04</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Employee Gain (#)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED</td>
<td>59</td>
<td>0</td>
<td>33</td>
<td>33</td>
<td>38</td>
<td>1,495</td>
<td>2 ± 0</td>
<td>1.83 ± 0.17</td>
<td>0.08</td>
</tr>
<tr>
<td>KFS</td>
<td>643</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>8</td>
<td>175</td>
<td>9 ± 5</td>
<td>2.31 ± 0.23</td>
<td>0.08</td>
</tr>
<tr>
<td>INC</td>
<td>4,998</td>
<td>15</td>
<td>74</td>
<td>22</td>
<td>399</td>
<td>14,680</td>
<td>114 ± 32</td>
<td>2.12 ± 0.06</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Employee Growth (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>3,011</td>
<td>29,900</td>
<td>400 ± 320</td>
<td>2.04 ± 0.72</td>
<td>0.09</td>
</tr>
<tr>
<td>KFS</td>
<td>367</td>
<td>67</td>
<td>143</td>
<td>5</td>
<td>257</td>
<td>2,600</td>
<td>233 ± 164</td>
<td>2.52 ± 0.52</td>
<td>0.09</td>
</tr>
<tr>
<td>INC</td>
<td>4,198</td>
<td>90</td>
<td>202</td>
<td>8</td>
<td>427</td>
<td>8,050</td>
<td>566 ± 138</td>
<td>2.66 ± 0.14</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Revenue ($000)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED - Yr2</td>
<td>119</td>
<td>60</td>
<td>40</td>
<td>6</td>
<td>1,600</td>
<td>12,000</td>
<td>500 ± 560</td>
<td>1.76 ± 0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>PSED - Yr5</td>
<td>145</td>
<td>42</td>
<td>35</td>
<td>7</td>
<td>1,400</td>
<td>12,000</td>
<td>600 ± 36</td>
<td>1.78 ± 0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>KFS</td>
<td>1,901</td>
<td>125</td>
<td>84</td>
<td>25</td>
<td>4,500</td>
<td>160,000</td>
<td>1,900 ± 134</td>
<td>2.19 ± 0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>INC</td>
<td>5,000</td>
<td>10,700</td>
<td>73,135</td>
<td>36</td>
<td>631,131</td>
<td>30,700,000</td>
<td>158,000 ± 444,000</td>
<td>1.79 ± 0.09</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Revenue Gain ($000)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED</td>
<td>111</td>
<td>45</td>
<td>22</td>
<td>14</td>
<td>2,182</td>
<td>4,500</td>
<td>70 ± 56</td>
<td>1.82 ± 0.19</td>
<td>0.09</td>
</tr>
<tr>
<td>KFS</td>
<td>1,630</td>
<td>115</td>
<td>844</td>
<td>25</td>
<td>47,280</td>
<td>160,000</td>
<td>1,700 ± 1,060</td>
<td>2.14 ± 0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>INC</td>
<td>4,998</td>
<td>4,991</td>
<td>21,764</td>
<td>19</td>
<td>99,350</td>
<td>3,770,175</td>
<td>16,200 ± 10,300</td>
<td>2.00 ± 0.09</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Revenue Growth (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSED</td>
<td>58</td>
<td>140</td>
<td>63</td>
<td>4</td>
<td>1,753</td>
<td>95,900</td>
<td>1,100 ± 1,608</td>
<td>1.76 ± 0.23</td>
<td>0.08</td>
</tr>
<tr>
<td>KFS</td>
<td>921</td>
<td>100</td>
<td>70</td>
<td>29</td>
<td>4,052</td>
<td>79,900</td>
<td>180 ± 155</td>
<td>1.80 ± 0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>INC</td>
<td>4,998</td>
<td>94</td>
<td>331</td>
<td>16</td>
<td>1,107</td>
<td>40,882</td>
<td>114 ± 399</td>
<td>1.86 ± 0.24</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*Gain* is absolute difference between year two and year five for PSED and KFS, between 2007 and 2010 for INC;  
*growth* is relative percentage;  
*Max* is largest value in distribution;  
*$x_{\text{min}}$* is the tipping point, where data change from linear to non-linear;  
*$\alpha$* is slope of the power-law tail;  
Kolmogorov-Smirnov (K-S) goodness-of-fit statistic is considered robust when ≤ 0.10 – significant K-S statistics are bolded.