Journal of Business Venturing Insights 21 (2024) e00447

Contents lists available at ScienceDirect

Journal of Business Venturing Insights

journal homepage: www.elsevier.com/locate/jbvi

Under the weight of heavy tails: A power law perspective on the emergence of outliers in entrepreneurship

G. Christopher Crawford [a,*], Harry Joo [b], Herman Aguinis [c]

[a] The University of Tampa, Sykes College of Business, 401 W. Kennedy Blvd, Tampa, FL, 33606, USA
[b] University of Dayton, 300 College Park, Dayton, OH, 45469-2271, USA
[c] The George Washington University, 2201 G St NW, Washington, DC, 20052, USA

ARTICLE INFO

JEL classification:
C14
D39
M13

Keywords:
Generalizable outcomes
Generative mechanisms
Heavy-tailed distributions
Outliers
Power law perspective

ABSTRACT

A fundamental discovery in entrepreneurship is that firm outcomes do not follow a symmetrical Gaussian curve. Instead, most are heavily right-skewed distributions in which a few extreme outliers (e.g., rock star firms like Airbnb, Tesla, and Uber) account for a disproportionate amount of the output. Although past research usually described outcome distributions as shaped following the power law, our study asks the following question: *What other less extreme distributions of generalizable firm outcomes exist in entrepreneurship?* Our investigation leverages four representative datasets from the U.S., Europe, and Australia, comprising 32 samples with about 22,000 ventures. We implemented a precise data-analytic approach that compares each sample (i.e., empirical distribution) against multiple theoretical distribution shapes to identify the best fit. Results showed that, across nearly all samples, the pure power law was not the dominant distribution. Instead, the annual revenue distribution is shaped as a power law with an exponential cutoff, and the number of employees distribution is shaped lognormally. Combined, these suggest the existence of top-down limitations on the highest performing firms. Accordingly, we offer an agenda for future research focused on (a) identifying and releasing systemic constraints, (b) examining and falsifying the underlying generative mechanisms that cause the emergence of heavy-tailed distributions and the outliers therein, and (c) conducting multi-level, mixed-method studies to investigate how micro-level interactions aggregate into macro-level heavy-tailed distributions. Our paper makes significant contributions to the power law perspective and future efforts to explain and predict the emergence of rock star firms in entrepreneurship.

1. Introduction

Challenging a long-held and often-implicit assumption that firm outcomes follow normal (i.e., Gaussian) distributions where outliers are inconsequential or exceedingly rare, studies in entrepreneurship have uncovered that firm outcomes are power law-distributed, where a minority of outlier-sized firms explain most of the distribution's variability (Aguinis et al., 2018a; Crawford et al., 2015; Crawford and McKelvey, 2018; Shim, 2016; Su et al., 2019). Power law distributions are said to be so pervasive that they are the "empirical reality of entrepreneurship," where the surprisingly abundant outliers within can create new marketspaces, redefine the nature of competition, and transform the world's expectations of what is possible (Crawford et al., 2014: 3). High-profile examples Airbnb, Amazon, Tesla, and Uber are firms that have creatively destroyed industries and changed the dynamics of human interactions.

* Corresponding author.
E-mail addresses: GCrawford@UT.edu (G.C. Crawford), yjoo01@udayton.edu (H. Joo), haguinis@gwu.edu (H. Aguinis).

https://doi.org/10.1016/j.jbvi.2023.e00447
Received 4 September 2023; Received in revised form 14 December 2023; Accepted 26 December 2023

Outlier firms are not just "statistical freaks" or "markedly different" than others; they are *better*, primarily because they have improved processes and achieved performance heights that others have not. Thus, understanding how these outliers and the distributions within which they reside has become an important focus of entrepreneurship research.

Recent work by Clark et al. (2023) espouses a power law perspective framework to explain and predict the emergence of outliers (i.e., exceptional outcomes) and skewed distributions at multiple levels of analysis. This framework (Andriani and McKelvey, 2009; Boisot and McKelvey, 2010; Booyavi and Crawford, 2023; Crawford et al., 2015, 2023) proposes that the emergence of outliers is predicted by four input constructs: *endowments* (e.g., human, social, intellectual, financial capital); *expectations* (e.g., envisioned future outcomes); *engagement* (e.g., actions and interactions with potential stakeholders); and *environments* (e.g., resource munificence, opportunities). The framework assumes that the variables within these constructs will be power law distributed and that the generative mechanism—a concept we define in the following section—driving both input and outcome distributions is self-organized criticality (SOC). SOC, conceptualized initially as an explanation for the size and frequency of extreme events like landslides and earthquakes (Bak et al., 1987), is used in the power law perspective to suggest that once an interdependent variable exceeds some minimal threshold of performance, it has the potential to have nonlinear effects on other variables. This threshold becomes the critical point beyond which a "normal" observation becomes an outlier. Subsequently, SOC as a mechanism with the power law perspective suggests that founder interactions with potential stakeholders (e.g., engaging with investors, customers, and suppliers) can lead to explosive, nonlinear jumps in generalizable firm outcomes (Crawford, 2015). In turn, across the entire domain of new venture creation and growth, outliers emerge, and these firm outcomes become power-law distributed. Thus, as Boisot and McKelvey's (2010) foundation of complexity science assumptions suggest, a power law perspective provides an overarching ontological understanding of the domain, with the utility to plausibly explain and predict the full scope of outcomes in entrepreneurship—from the tens of millions Mom & Pop retail businesses with zero employees to the highest performing rock star firms with valuations in the trillions of dollars (Crawford et al., 2023).

However, previous power law studies suggest that these distributions may not be statistically precise (i.e., "pure") power laws (c.f., O'Boyle and Aguinis, 2012), and this idea has received little empirical attention. More specifically, in their conceptual paper, Boisot and McKelvey (2010, p. 416) note regarding organizational outcomes, "other (less extreme) skew distributions, reflecting the different ways that phenomena interact, are also possible." Though this is consistent with seminal work by Simon (1955), identifying "less extreme" is an important caveat, because other research on heavy-tailed distributions proposes that if a distribution is not a pure power law, different underlying processes may drive its emergence (c.f., Bradley and Aguinis, 2023; Joo et al., 2017; Stumpf and Porter, 2012). Thus, the shape of the distribution is an important first step for investigating the causal mechanisms that drive the emergence of outliers. In turn, this could lead to significant insights for theory, policy, and practice (Leitch et al., 2010). This, therefore, focuses our attention on the distribution of generalizable entrepreneurial outcomes, leading to our primary research question, *What other less extreme distributions of generalizable firm outcomes exist in entrepreneurship*?

To answer this question, we leverage a method initially used in the study of individual and team performance, distribution pitting (Bradley and Aguinis, 2023; Joo et al., 2017), which compares empirical data with conceptual distributions (both heavy-tailed and normal) to determine the best fit. We use this method on annual revenue and number of employees data from four representative samples of entrepreneurial activity. Specifically, we included the nascent and emerging firms in Panel Study of Entrepreneurial Dynamics I & II and the Comprehensive Australian Study of Entrepreneurial Emergence, as well as hypergrowth firms in the INC 5000, both in the U.S. and in Europe (N~22,000). We pitted each of the empirical distributions against four different, heavy-tailed distributions: (a) pure power law, (b) lognormal, (c) power law with an exponential cutoff, and (d) exponential (Fig. 1 includes technical details and a visual representation of each type).

Results uncovered very few pure power law distributions. Instead, across all samples, we find that annual revenue is distributed according to the power law with an exponential cutoff and number of employees is shaped as a lognormal distribution. These findings support the power law perspective that "other, less extreme' heavy-tailed distributions exist. Our improved empirical precision in the specific shape of heavy-tailed distributions now serve as a vital stepping stone for further investigations into discovering the generative mechanisms that cause these distributions at multiple levels. Moreover, the lack of pure power law distributions suggests it is possible that self-organized criticality may not be the causal mechanism driving generalizable entrepreneurial outcomes such as revenue and number of employees. In turn, this places urgency on revisiting the search for mechanisms (i.e., "the different ways that phenomena interact") in the domain, as first suggested by McKelvey (2004) and later by Crawford et al. (2014) in the first issue of *Journal of Business Venturing Insights*. Next, we offer a brief review of the evolving research stream on heavy-tailed distributions and the underlying drivers of them. Then, we delve into distribution pitting methodology and describe our results. Finally, we discuss implications for theory and future research as well as policy and practice, including suggestions for a research agenda that tie together the four power law perspective constructs with the four classifications of mechanisms proposed by Andriani and McKelvey (2009) and Crawford et al. (2015), aimed at identifying the underlying mechanisms generating the emergence of non-normal distributions and outliers in entrepreneurship.

2. Skewed research

2.1. Domains of distributions, variables, mechanisms, and methods

Over the last decade, a significant stream of research has focused on the emergence of power law (i.e., skewed, heavy-tailed, non-normal) distributions. As shown in Table 1, the studies investigated a range of distribution shapes, including the "pure" power law, lognormal, power law with an exponential cutoff, and exponential tail. This diversity in distribution shapes (and subsequent discussions of the mechanisms causing them) indicate a complex landscape in entrepreneurship and management, where different internal aspects and external forces can drive the emergence of outliers and lead to highly skewed aggregate outcomes in terms of firm size, performance, and growth. Moreover, Table 1 reflects studies conducted at various levels of analysis, including individual, firm, community, country, and even specific platforms like Udemy and Instagram. The variables studied are equally diverse, ranging from student reviews, annual household income, and posts on social media to more traditional business metrics like annual revenue, market capitalization, and scientific publications. This diversity underscores the multifaceted nature of the domain, where a broad spectrum of factors, from micro-level individual actions to macro-level conditions, can influence the emergence of outliers and highly skewed outcome distributions. It is important to note Andriani and McKelvey (2009) identify 15 potential underlying mechanisms of these distributions and call them "scale-free" because each applies at all units and levels of analysis of the phenomena under study; later, for future research, we use the authors' four classifications of these mechanisms (i.e., *positive feedback*, *multiple distributions*, *ratio*

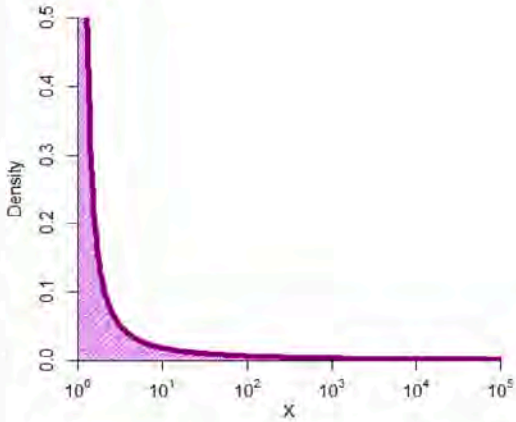
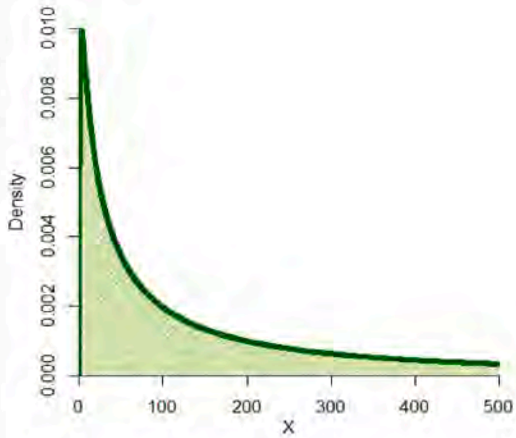
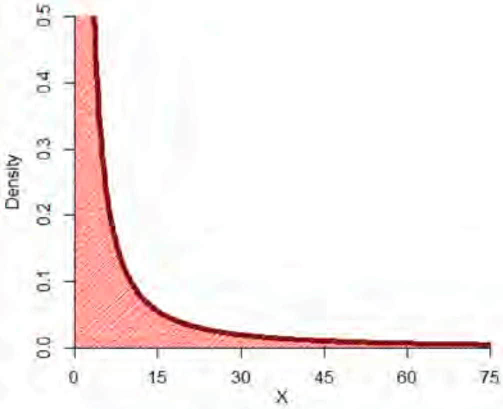
Description of distribution shape	Visual representation of each specific shape
1. Pure power law ($\alpha = 1.5$): $$p(x) \propto x^{-\alpha},$$ where x is a set of values from a variable. Alpha (α) (> 1) is the rate of decay (i.e., how quickly the distribution's right-tail "falls"). So, the lower the value of α (closer to 1), the heavier the distribution's right tail. The pure power law lacks any constraint to how large outliers can be—a unique feature that no other distribution shape has. When α is close to 2, it is The pure power law consists of itself and generally has the heaviest right tail, with the largest outliers, among all distribution shapes in our study.	
2. Lognormal ($\mu = 5, \sigma = 2$): $$p(x) \propto e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}},$$ where x is a set of values from a variable. Euler's number $e \approx 2.718$. $\ln(x)$ is the natural log of x and is normally distributed. Mu (μ) (> 0) is the mean. Sigma (σ) (> 0) is the standard deviation. μ does not affect the heaviness of the distribution's right-tail. In contrast, σ does. The higher the value of σ (further away from 0), the heavier the distribution's right tail. Higher values in the lognormal tend to increasingly differ from each other (i.e., increasingly larger distances between the higher-value cases). Unlike the pure power law, the lognormal decays rapidly at the extreme end of its right tail (Taleb, 2007: 326). The lognormal consists of itself and generally has the second-heaviest right tail among the four shown here.	

Fig. 1. Description and visual representation of the four shapes of heavy-tailed distributions.

3. Power law with exponential cutoff ($\alpha = 1.5, \lambda = 0.01$):

$$p(x) \propto x^{-\alpha} e^{-\lambda x},$$

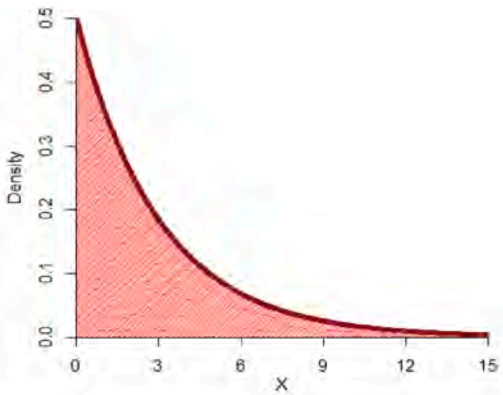
where x is a set of values from a variable. Euler's number $e \approx 2.718$. Both alpha α (> 1) and lambda λ (> 0) are rates of decay, or how quickly the distribution's right-tail falls. So, the lower the values of α (closer to 1) and λ (closer to 0), the heavier the distribution's right tail. The rate of decay λ is stronger in terms of making the distribution's right tail fall and thus increasingly influences the distribution's shape along higher values of x. The tail is the third heaviest among the four in our analysis.



4. Exponential distribution ($\lambda = 0.5$):

$$p(x) \propto e^{-\lambda x},$$

where x is a set of values from a variable. Euler's number $e \approx 2.718$. Lambda (λ) (> 0) is the rate of decay, or how quickly the distribution's right-tail falls. So, the lower the value of λ (closer to 0), the heavier the tail. This has the lightest tail among the four heavy-tailed distributions we use in our analysis.



Note. In each of the visuals, the x-axis represents values of a continuous variable, whereas the y-axis ("Density") represents the likelihood of the continuous variable taking on a given value or range of values. Some text and images were originally published in Joo, Aguinis, and Bradley (2017).

Fig. 1. (*continued*).

imbalances, and *contextual effects*) and associate them with the four power law perspective constructs.

As background, a generative mechanism in the context of social sciences, particularly in complex systems and organizational theory, refers to the underlying process or set of processes that give rise to observed patterns, structures, or behaviors within a system (c.f., Baron, 1998)—these reflect, calling back to Boisot and McKelvey (2010), "the different ways that phenomena interact." It explains how specific outcomes or patterns emerge from the interactions of simpler elements within that system. They are called "generative' because they generate the observed phenomena, often through the interaction of basic elements following certain rules or principles. Generative mechanisms are often not directly observable; instead, they are inferred from the patterns or behaviors they produce. Understanding a generative mechanism involves identifying the rules or principles governing the interactions of the system's components. We discuss mechanisms in more depth at the end of this paper.

Finally, the methodologies used to determine distribution shapes in Table 1 vary across the studies, including the chi-square statistic, maximum likelihood estimation (MLE) semi-parametric bootstrap analysis, Gabaix-Ibragimov Zipf Plots, coupled Simon models, and distribution pitting. These different methodological approaches reflect the evolving nature of entrepreneurship and management research in this field, showcasing both traditional and innovative data-analytic techniques to understand complex phenomena. Our study is positioned to solely focus on the shape of generalizable outcome distributions as a foundation for future

Table 1
Summary of distribution shape studies in entrepreneurship (listed chronologically from most recent to oldest).

Authors	Publication	Level of Analysis	Sample Type	Sample Size	Variable	Data-analytic Approach to Determine Shape	Distribution Shapes Found	Proposed Generative Mechanisms
Gala et al. (2024)	*JBV*	Individual	Udemy Teaching Platform with 52 Domains	~12,000	Number of Student Reviews	Distribution Pitting	Lognormal Power Law w/ Exponential Cutoff	Proportional Differentiation
Crawford et al. (2023)	*JBVI*	Individual	PSED II	~1200	Annual Household Income	MLE Semi-parametric Bootstrap Analysis	Power Law	Endowments, Expectations for Future Growth, and Engagement
Thietart and Malaurent (2023)	*AMD*	Community	Instagram and Twitter (now known as X)	N/A	Posts & Tweets per day	Two-stage: autoregressive model and BDS statistics	Power Law	Black noise: Multiplicative interconnectivity
Booyavi and Crawford (2023)	*JBVI*	Individual Nascent Venture	PSED II	~1200	Annual Revenue	MLE Semi-parametric Bootstrap Analysis	Power Law	Endowments, Expectations for Future Growth
Bradley and Aguinis (2023)	*OS*	Team Distribution	274 Performance Distributions	~700,000	Team Performance	Distribution Pitting	Power Law w/ Exponential Cutoff	Incremental Differentiation Authority Differentiation Temporal Stability
Podobnik et al. (2020)	*CSF*	Individual Firm Country	Forbes Wealthiest S&P 500 SCImago OCED Member Countries	~1000	Net Worth Market Cap Scientific Publications per Country GDP Growth Rate	Gabaix-Ibragimov Zipf Plots Coupled Simon Model	Power Law	STEM Education Preferential Attachment
Su et al. (2019)	*PHH*	Firm	American Family Business Survey SMDC	~3200	Annual Revenue	Chi-Square Statistic	Power Law	N/A
Aguinis et al. (2018b)	*MR*	Individual Firm	CEO Execucomp Database Compustat Database	~4500	CEO Compensation Firm Tobin's Q and Return on Assets	MLE Semi-parametric Bootstrap Analysis	Power Law	Job Autonomy Job Complexity
Aguinis et al. (2018a)	*JAP*	Individual	Researchers and Publications in Four Domains	~59,000	# Published Articles	Distribution Pitting	Power Law w/ Exponential Cutoff	Incremental Differentiation
Crawford and McKelvey (2018)	*EE*	Individual Nascent Venture Young Firm Hypergrowth Firm	PSED II KFS INC 5000	~6000	Net Worth Strong Ties Employees Supervised Total Hours Annual Revenue Employees (#)	MLE Semi-parametric Bootstrap Analysis	Power Law	Self-organized Criticality Interacting Fractals Preferential Attachment Meta-Constructs: Endowments, Expectations, Engagement, Environments
Joo et al. (2017)	*JAP*	Individual	229 Occupational Domains	~625,000	Individual Output	Distribution Pitting	Exponential Tail	Incremental Differentiation
Aguinis et al. (2016)	*PP*	Individual	229 Occupational Domains	~625,000	Individual Output	MLE Semi-parametric Bootstrap Analysis	Power Law	Cumulative Advantage Job Autonomy Job Complexity
Shim (2016)	*JBVI*	Individual Nascent Ventures	PSED II	~1200	Expectations for Growth Revenue, Employees	MLE Semi-parametric Bootstrap Analysis	Initial Lognormal Emerges to Power Law	Multiplicative Process
Thietart (2016)	*SMJ*	Firm	Danone	1	Strategic events	Test for Colored Noise & Residuals Statistical Summary	Power Law	Pink Noise: Self-organized Criticality
Crawford et al. (2015)	*JBV*	Individual Nascent	CAUSEE PSED II	~12,000	25 input variables: Endowments, Expectations, Engagement,	MLE Semi-parametric Bootstrap Analysis	Power Law	Future Research on: Contextual Effects

(*continued on next page*)

Table 1 (*continued*)

Authors	Publication	Level of Analysis	Sample Type	Sample Size	Variable	Data-analytic Approach to Determine Shape	Distribution Shapes Found	Proposed Generative Mechanisms
		Venture Young Firm Hypergrowth Firm	KFS INC 5000		Environments Annual Revenue ($), Growth (%), & Gain (+), Employees (#), Growth (%), & Gain (+)			Positive Feedback Multiplicative Processes Ratio Imbalances
Crawford et al. (2014)	*JBVI*	Nascent Ventures Young Firm Hypergrowth Firm	PSED II KFS INC 5000	~11,000	Annual Revenue Number of Employees	MLE Semi-parametric Bootstrap Analysis	Power Law	Future Research on: Self-organized Criticality Phase Transitions Preferential Attachment
O'Boyle and Aguinis (2012)	*PPsych*	Individual	Researchers Entertainers Politicians Amateur Athletes Professional Athletes	~633,000	Individual Output	Chi-Square Statistic	Power Law	Future Research on: Matthew Effect

Notes. Publication: JBVI: *Journal of Business Venturing Insights,* AMD*: Academy of Management Discoveries,* JBV: *Journal of Business Venturing,* OS: *Organization Science,* CSF: *Chaos, Solitons, and Fractals,* PHH: *The Palgrave handbook of heterogeneity among family firms,* MR: *Management Research,* JAP: *Journal of Applied Psychology,* EE: *Edward Elgar Handbook of Research Methods in Complexity Science: Theory and Applications,* PPsych: *Personnel Psychology,* SMJ: *Strategic Management Journal.* Sample: CAUSEE (Comprehensive Australian Study of Entrepreneurial Emergence), PSED II (Panel Study of Entrepreneurial Dynamics II), KFS (Kauffman Firm Survey), INC 5000 (Inc. Magazine 500 Fastest Growing Privately Owned Firms in the United States), S&P 500 (Standard & Poor's 500 Largest Publicly Owned Companies in the United States), Scimago (Scimago Journal & Country Rank), OCED (Organization for Cooperation and Development), SMDC (Small Business Development Center). Sample Size: is approximate to account for specific samples in the study and reflect the sum of total observations. Data Analytic Approach to Determine Shape: the maximum-likelihood estimation bootstrap analysis compares empirical data with a specific heavy-tailed distribution and calculates both a Kolmogorov-Smirnov (KS) statistic and a *p*-value, with plausible hypothesis support indicated by a KS below 0.1 and *p*-value above 0.1 (Clauset et al., 2009); distribution pitting is a falsification-based method for comparing distributions to assess how well each one fits a given data set, assessed by loglikelihood ratio and *p*-value (Joo et al., 2017); Gabaix-Ibragimov Zipf Plots calculates a Zipf exponent (ξ), which correlates to a specific alpha (α) where a power law is indicated when $\alpha \in (0,2)$ (Gabaix and Ibragimov, 2011) and the coupled Simon model uses a stationary cumulative distribution that exhibits power law scaling, then fits that to an empirical distribution (1955); two-stage method of autoregressive model is a linear estimation of the time series that filters all nonlinear causal effects and BDS statistics assess whether series residuals are independently, identically, distributed (Poole et al., 2000); test for colored noise and residuals statistical summary use *p*-value above 0.1 and alpha (α) between 0 and 2 from Clauset et al. (2009), respectively, to support hypothesized power law distribution; and the chi-square statistic forces empirical data to conform with both a Gaussian distribution and a Paretian (i.e., power law) distribution and determines a better fit, indicated by the lower χ^2 value (Aguinis and Harden, 2009).

research on the mechanisms driving these distributions. Next, we explore the semi-parametric distribution pitting method that we used to identify the shape of firm outcomes more precisely.

3. Method

We measured the generalizable outcomes of each firm according to annual revenue (i.e., sales) and employment (i.e., number of employees), consistent with Crawford et al.'s (2014, 2015) empirical investigation of power law distributions in entrepreneurship. We did not use other indicators of firm size, such as total asset value or market capitalization, because their relevance is limited to only certain types of firms or industries (i.e., publicly traded, capital-intensive) (c.f., Delmar et al., 2003).

3.1. Revenue

For annual firm revenue, we used four datasets consisting of 16 samples, spanning about 22,000 firms that are consistent with definitions of entrepreneurship, whether a fast-growing venture (Markman and Gartner, 2002) or the creation of a new venture (Davidsson and Wiklund, 2006). The first datasets was collected from Inc. 5000, consisting of cross-sectional data on privately held, for-profit firms varying in age and size. Many firms were just a few years old, while some were 10 years or older; several firms had as low as $2 million USD in annual revenue, whereas a few had more than $4 billion USD. A key aspect of Inc. 5000 firms is that they are deemed hypergrowth, which is consistent with the fact that two of the four distributions we investigate allow for sudden or rapid growth (i.e., pure power law and lognormal, as shown in Fig. 1). From Inc. 5000, we specifically collected data on firms in the United States (US) in 2015 and 2016 and Europe (EU) in the same two years. As a result, we derived four Inc. 5000 datasets, with four samples of annual revenue, all of which represent 20,000 firms.

The other three datasets were collected from Panel Study of Entrepreneurial Dynamics (PSED) I & II and Comprehensive Australian Study of Entrepreneurial Emergence (CAUSEE). We used these three because, unlike the Inc. 5000, they consist of longitudinal data tracking cohorts of nascent entrepreneurial activity. PSED I & II follow a cohort of nascent firms in the US, whereas CAUSEE tracks a combination of nascent and young Australian firms over time. We collected three waves (i.e., samples) of annual firm revenue from PSED I (encompassing years 1998–2000); five waves from PSED II (2005/2006–2010/2011); and four waves from CAUSEE (2007–2011)—adding up to 12 samples and representing about 2000 firms. Thus, we obtained 16 samples of annual revenue across all five datasets spanning about 22,000 firms.

3.2. Employment

We collected data on number of employees (i.e., employment or headcount). Our data from PSED I and II and CAUSEE includes full-time and part-time employees, whereas, for Inc. 5000, firms only reported full-time employees. Combined, we collected 16 samples of annual firm employment. In sum, we have 32 observed firm size distributions to analyze.

3.3. Data-analytic approach: distribution pitting

We implemented the distribution-pitting method with the Dpit package in R (Joo et al., 2017) to pit four different shapes of heavy-tailed distributions (i.e., pure power law, lognormal, power law with exponential cutoff, and exponential). Distribution pitting is

Table 2
Summary of results for revenue and employment samples, based on PSED I, PSED II, CAUSEE, and INC 5000 US & EU.

	Revenue: Best-fitting distribution shape	Employment: Best-fitting distribution shape
Longitudinal		
PSED I	**($USD)**	
Wave 1, $N = 152$	Power law with exponential cutoff	Power law with exponential cutoff
Wave 2, $N = 198$	Power law with exponential cutoff	Power law with exponential cutoff
Wave 3, $N = 78$	Power law with exponential cutoff	Power law with exponential cutoff
PSED II		
Wave 1, $N = 119$	Lognormal	Power law with exponential cutoff
Wave 2, $N = 132$	Power law with exponential cutoff	Power law with exponential cutoff
Wave 3, $N = 126$	Power law with exponential cutoff	Undetermined
Wave 4, $N = 145$	Power law with exponential cutoff	Pure power law
Wave 5, $N = 137$	Power law with exponential cutoff	Pure power law
CAUSEE	**($AUD)**	
Wave 1, $N = 891$	Lognormal	Lognormal
Wave 2, $N = 722$	Power law with exponential cutoff	Lognormal
Wave 3, $N = 613$	Power law with exponential cutoff	Lognormal
Wave 4, $N = 448$	Power law with exponential cutoff	Undetermined
Cross-Sectional		
INC 5000	**($USD)**	
US 2015	Power law with exponential cutoff	Lognormal
US 2016	Power law with exponential cutoff	Lognormal
	(€EU)	
EU 2015	Power law with exponential cutoff	Lognormal
EU 2016	Power law with exponential cutoff	Lognormal

Note. See Fig. 1 for technical details and a visual representation of each distribution shape. "Undetermined" refers to a result in which none of the distribution shapes was identifying as the best-fitting.

applied to each of the 32 samples to determine the best-fitting shape. Besides understanding the precise shape of the distribution this pitting technique also allows us to see the frequency and magnitude of outliers. To control for the possibility that the best-fitting shape for a sample is not a heavy-tailed distribution and, instead, a (fairly) symmetrical one, the method also pits the four heavy-tailed shapes against a normality-based category. Because the heavy-tailed shapes consist of four specific distributions (as shown in Fig. 1) and the normality-based shape category consists of three specific distributions (i.e., Normal, Poisson, and Weibull, as described in Joo et al., 2017), we consider a total of seven specific distributions and thus conducted 21 pairwise comparisons per sample. This distribution pitting method adopts a philosophy of science approach based on falsifying models as a means of advancing scientific knowledge. Specifically, it employs three decision rules that help actively rule out distribution shapes until only one is remaining, which is deemed the best-fitting one (or more accurately, the least unlikely). Alternatively, if multiple shapes remain even after all comparisons and decision rules per sample, the conclusion for the sample is undetermined. Full details on the distribution-pitting method are in Appendix 1.

4. Results

4.1. Revenue

A summarized in Table 2, 14 out of the 16 annual revenue samples are shaped according to a power law with an exponential cutoff distribution. The only exceptions were two lognormal distributions, and they were only for the first year of data collection on nascent firms in the PSED II and CAUSEE. None were shaped according to a pure power law distribution.

As a robustness check, we examined whether our conclusions based on the four Inc. 5000 samples still hold after splitting the samples into 100 random industry-specific subsets for 2016 and 2015. Among those, 84 were distributed as a power law with exponential cutoff (with an equal number from the US and EU), five were pure power-law distributed (all from the EU, including Engineering, Software, Health Care, IT Services, and Travel & Hospitality), ten were exponentially distributed, and one was undecided. Combined, these checks suggest that our initial findings were robust.

4.2. Employment

Table 2 shows that the employment distributions were mostly lognormal (7 out of 16) and power law with an exponential cutoff (5 out of 16). Two were undetermined and two were pure power-law distributed. All four INC 5000 samples of established hyper-growth companies were lognormally shaped. Also, the first three years of CAUSEE show lognormal distributions for number of employees (with an undetermined result in year four). In somewhat of a contrast, for nascent firms in the PSED I and II, the number of employees was shaped as a power law with exponential cutoff for the first few years; however, in the final two years of the PSED II, the distributions emerge into pure power laws. These results mirror Shim's (2016) findings that distributions become more skewed over time (i.e., outliers increase their advantages; tails get heavier), with the only two pure power law-distributed samples among the 32 we studied.

As an additional robustness check, we split the employment samples into 100 random industry-specific subsets, with 50 from the US and 50 from the EU. Among those, 56 were lognormal, nine (9) were distributed as a power law with exponential cutoff, 19 were shaped as exponential, 15 were undetermined, with only one as a pure power law (Consumer Products & Services in the EU). Combined, these checks support our initial findings, albeit less so than the results for revenue. We also note that these results are consistent with Certo et al.'s (2023: p.1) different skewness varying "substantially across measures, samples, and years."

Looking at an overview of both generalizable outcome variable results, we see the lognormal distribution of employment in contrast to the power law with an exponential cutoff distribution of annual revenue. We know from Fig. 1 that lognormal distributions have heavier tails with more influential outliers vis-à-vis power law with exponential cutoff, which, according to complexity science (McKelvey, 2004) and the power law perspective (c.f., Crawford and McKelvey, 2018), suggests that there is likely some sort of environmental constraint on annual revenue. For example, this constraint could be some kind of top-down rule (possibly a tax regulation) that does not influence the number of employees in a firm. It is interesting to note that this constraint is consistent across venture stages, multiple countries, and firm age. Next, we transition into a discussion on the implications of our heavy-tailed findings.

5. Discussion

5.1. Implications for theory and a research agenda on distributions and outlier emergence

We find that pure power law distributions are not nearly as dominant as they have been assumed to be based on using more traditional methodological approaches (e.g., Aguinis et al., 2018b; Crawford et al., 2015; Crawford and McKelvey, 2018; Shim, 2016; Su et al., 2019). Although none of the 32 distributions of revenue and employment are normally distributed, only two were power pure power law distributed. The majority are distributed following a power law with an exponential cutoff and a lognormal shape. Whereas we found little evidence of pure power law distributions in the domain's most generalizable outcome variables, we did confirm one primary empirical fact: literally no outcome variables are normally distributed—all are heavy-tailed. We emphasize that in these alternative, less-extreme distributions, outliers still dominate; outliers still drive aggregate system outcomes; outliers are still the primary instigators of innovation, creative disruption, and the emergence of new order; outliers still disproportionately influence the statistical and behavioral properties of the population. However, in nearly all prior empirical studies, many of the most important observations have been deleted to obtain statistical significance. Outliers, then, are still, as Certo et al. (2023: p.7) describe, "empirically problematic but conceptually salient." This also suggests future studies that investigate the overall influence of outliers in the distribution (which could be calculated as alpha (α) with the maximum-likelihood technique in Table 1) would be of benefit to the

Table 3
Future research agenda: generative mechanisms for distribution shapes and outliers.

Generative mechanism classification (*and associated power law perspective construct*)	Description
1. Contextual effects (*Environments*)	• Definition: These mechanisms focus on exogenous effects that trigger scale-free (SF) dynamics (i.e., influences at multiple levels of analyses). Different kinds of external impacts initiate different SF causal processes, with the common factor being the contextual influence on and from the local and global environment. A mechanism where trivial events (e.g., winning a pitch competition) can trigger sudden increases in size ranging from small to extremely large for a few firms (Andriani and McKelvey, 2009; Boisot and McKelvey, 2010, 2011). • Example mechanisms: *self-organized criticality*, *niche proliferation*, *contagion bursts*, *phase transition*, *incremental differentiation, pink noise*. • Aspects of these mechanisms can be found in Aislabie (1992), Derbyshire and Garnsey (2014), Lichtenstein (2000), Nicholls-Nixon (2005), Thietart (2016), and Thietart and Malaurent (2023). • Implications: Given the potential inherent bursts in these mechanisms, outlier-sized firms are inherently unpredictable or nondeterministic (Taleb, 2007, 2020). So, rather than trying to predict outlier-sized firms individually, researchers should identify and study aspects of the broader *environment* that allow more or fewer outliers to exist (Aguinis et al., 2016). In other words, entrepreneurship theory must focus on “plausible anticipation rather than prediction” of outlier-sized firms (Crawford, 2012: 79). • Power law perspective associations: *Environments* refer to the local and global munificence surrounding and influencing an entity or system, where outcomes have high long-range correlations with initial conditions. When there are limited constraints on a system (e.g., a capitalistic market), exogenous factors and contextual effects play a vital role in setting off scale-free causal processes once they reach a point of criticality. However, top-down constraints like excessive organizational rules or governmental regulations impede the emergence and performance of outliers, leading to incremental differentiation and truncated tails. In contrast, with self-imposed governance and no rules, there is the potential for extreme, beyond-outlier ‘Dragon-King’ outcomes (c.f., Sornette and Ouillon, 2012). • Potential research questions: What organizational rules constrain the highest performing individuals within a firm? | What regulations constrain the highest performing firms in an industry? | In what ways do phase transitions influence the engagement strategies of tech startups versus traditional businesses? | How do deregulated environments facilitate nonlinear increases in the endowments and outcomes of emerging firms? | What role does the environment play in elevating startups’ expectations after winning events like pitch competitions? | How do varying socio-economic environments shape the engagement approaches of social entrepreneurship ventures? | In what ways do local and global environmental factors impact the expectations of firms, leading to ‘Dragon-King’ outcomes? | How do exogenous environmental factors like political shifts alter the engagement strategies of small and medium enterprises? | What impact do environmental constraints, such as governmental regulations, have on the endowments of firms in developing economies? | How do environmental shifts drive the engagement tactics of green technology startups amidst scale-free causal processes? | What influence do contextually driven environmental events have on the expectations and engagement strategies of startups during market disruptions?
2. Positive feedback (*Engagement*)	• Definition: These mechanisms propose that large differences in firm size are a function of positive feedback loops, which are based on the interaction (i.e., multiplication, positive feedback loops) between two factors: initial size and initial growth rate (Gibrat, 1931; Mitzenmacher, 2004). • Example mechanisms: *preferential attachment*, *least effort*, *spontaneous order creation*, *irregularly generated gradients*, *proportionate differentiation*.• Aspects of these mechanisms can be found in Barabási (2005), Chiles et al. (2004), Garnsey et al. (2006), and Simon (1993). • Implications: Joo et al. (2017) suggest that proportionate differentiation means that outliers are predictable, and the predictors should relate to the *initial* characteristics of firms (i.e., initial size and initial growth rate) and their interactions over time. • Power law perspective associations: *Engagement* involves an individual’s or groups’ active participation and involvement in a particular process or system. In scenarios where positive feedback is present, engagement can intensify this effect. Sustained interactions with varying intensity over time can reinforce loops where success breeds further success, amplifying initial advantages, and leading to power law distributions in outcomes. This mapping underscores the role of continuous and active involvement in magnifying positive feedback loops in various contexts. • Potential research questions: How does preferential attachment influence the engagement strategies of e-commerce platforms in customer acquisition? | In what ways do least effort and spontaneous order creation enhance the engagement and productivity of startup teams? | How do irregularly generated gradients shape the engagement and decision-making processes in high-growth firms? | What role do initial size and growth rate, as aspects of engagement, play in predicting the success of family businesses? | How does early market engagement influence the innovation cycles in biotech firms through positive feedback loops? | Can sustained user engagement, as a form of positive feedback, explain the growth of online learning platforms? | How does initial engagement with the market impact the development of social media platforms? | In what ways do early success and positive feedback loops shape the engagement strategies of crowd-funded startups? | How do positive feedback mechanisms from early market engagement influence the scaling of influencer-driven marketing startups? | Can the theory of preferential attachment be applied to understand the engagement tactics of rapidly growing marketing startups?
3. Multiple distributions (*Endowments*)	• Definition: Mechanisms here indicate that SF dynamics are due to a combination of somewhat skewed distributions and outliers across various variables. These contribute multiplicatively to form heavy-, fat-,

(*continued on next page*)

Table 3 (*continued*)

Generative mechanism classification (*and associated power law perspective construct*)	Description
	and long-tailed power law distributions. • Example mechanisms: *combination theory*, *interactive breakage theory*, *interacting fractals* •Implications: It is important to understanding the shape of distributions. Crawford et al. (2015) found power law distributions in 25 theoretically relevant entrepreneurship input variables; combinations of these variables will have multiplicative effects, suggesting that the importance of omitted variable bias is much greater than proposed. • Power law perspective associations: *Endowments* refer to an entity's initial resources or capabilities. Endowments include human, social, intellectual, and financial capital, each characterized by heavy-tailed variables. In this context, the multiplicative effect of these inputs can lead to outlier outcomes and heavy-tailed distributions at higher levels of analysis. For example, a company with exceptional human capital (talented employees) and substantial financial capital might see exceptional team and company outcomes relative to others with fewer endowments. This multiplicative effect of outlier-tainted variables across different types of capital results in heavy-tailed distributions in organizational performance, innovation, or growth. • Potential research questions: How do combinations of different types of capital (human, social, financial) as endowments influence the scaling of fintech startups? | In what ways do endowments, in the form of interactive breakage theory and interacting fractals, explain growth patterns in multinational corporations? | How do varying distributions of resources as endowments impact the performance of non-profits versus for-profit organizations? | Can endowment combinations, such as intellectual and financial capital, explain the success of interdisciplinary R&D teams in tech firms? | How do endowments in the form of global networks and local insights drive the success of multinational consulting firms? | In what ways do endowments like diverse team skills and robust financial backing impact AI startups? | How do endowment combinations affect the resilience and adaptability of small businesses during economic downturns? | Can the combination theory, as an aspect of endowments, explain the success rate of technology startups in innovation hubs? | How do the endowments of intellectual and financial capital influence the innovation output in pharmaceutical companies? | In what ways do endowments, including human, social, intellectual, and financial capital, result in heavy-tailed distributions in organizational performance?
4. Ratio imbalances (*Expectations*)	• Definition: This category includes mechanisms where the scale-free cause is some "cost-driven efficiency" requiring constant or periodic adjustment. • Example mechanisms: *event bursts*, *hierarchical modularity*, *random walk*, *surface-volume law*, *autogenesis, black noise* • Key feature: comparisons of a current situation vis-à-vis envisioned outcomes motivate extra engagement; when individual or venture goals are not met, tensions instigate constant adjustments in staying on course where individuals will prioritize activities and show bursts of communication, entertainment, work, and travel activities followed by long delays (Amitrano, 2012; Drazin and Sandelands, 1992; Song et al., 2010). • Implications: Aspirations matter. Without understanding subjective individual or company goals, a scholarly, objective understanding of outcomes may only exist at a surface level. • Power law perspective associations: *Expectations* involve anticipations or beliefs about future performance or outcomes. In the context of ratio imbalances, these expectations can significantly impact allocating resources and efforts. For instance, high expectations for an organization's particular project or department might lead to disproportionate resource allocation, creating efficiency imbalances. This imbalance can manifest in heavy-tailed distributions as certain areas outperform others based on the weight of expectations placed upon them internally (self-regulated goals) or externally (existing or potential stakeholder objectives). The expectation-driven allocation of resources, attention, and efforts can create self-reinforcing successes or failures, exemplifying how expectations can shape power law dynamics. • Potential research questions: How do expectations for future economies of scale and economies of score influence the outcomes of social enterprises? | How do event bursts influence the expectations and goal-setting processes in the product development cycles of consumer electronics startups? | In what ways do expectations drive adjustments in organizational structures, as seen in hierarchical modularity, impacting efficiency in large corporations? | What role do expectations play in driving resource allocation and strategic adjustments in startups facing aspiration-driven challenges? | How do varying stakeholder expectations influence the strategic pivots and resource allocation in startups during market adversity? | In what ways do expectations shape the engagement and strategic decisions of tech startups striving for outlier growth targets? | How do expectations influence the management of ratio imbalances and efficiency in rapidly scaling tech startups? | What impact do individual and organizational expectations have on the strategic planning and resource allocation in entrepreneurial ecosystems? | How do expectations drive the allocation of resources and efforts, creating efficiency imbalances in different areas of an organization? | In what ways do expectations for novelty impact the engagement and resource allocation strategies of organizations seeking to innovate and grow?

domain—Crawford and McKelvey (2018) provide extensive guidance for using the parameters in this technique to test potential mechanisms.

For theory, our study makes significant contributions to the power law perspective's framework for understanding the emergence of outliers (Booyavi and Crawford, 2023; Clark et al., 2023; Crawford, 2015; Crawford et al., 2014) in several meaningful ways. First, this perspective hypothesizes that outcomes in social systems, when measured on a continuous scale without a pre-imposed limit, would be power law distributed. Our study, while finding pure power laws in only a small percentage of outcomes, provides a more nuanced understanding that suggests these outcomes would be more accurately characterized as heavy tail distributions. Second, the methodological precision we introduce with Dpit while investigating entrepreneurial *outcomes* might suggest that the power law perspective's *input* constructs—*endowments*, *expectations*, *engagement*, and *environments*—would be similarly heavy-tailed. This will be important for future theory development, considering the power law perspective proposes that outlier outcomes are primarily a result of outlier inputs (i.e., one or more construct variables with a value beyond some critical threshold in the distribution).

In all, our findings pave the way for a future research agenda regarding the mechanism(s) driving the emergence of heavy-tailed distributions and outlier firms in entrepreneurship. As proffered by Andriani and McKelvey (2009) and Boisot and McKelvey (2011), these heavy-tailed, "less extreme" power law distributions and the outliers therein are "caused" by some underlying generative mechanism(s). In the following sections, we identify what those mechanisms could be, classify some of their primary characteristics related to the study of entrepreneurship, associate those classifications directly with the input constructs of the power law perspective, and identify methods of testing and or ruling out individual mechanisms.

5.2. Classifying mechanisms and power law perspective associations

In Table 3, we identify how the four classifications of causal mechanisms proposed by Andriani and McKelvey (2009)—contextual effects, positive feedback, multiple distributions, and ratio imbalances—encompass the mechanisms known to generate heavy-tailed outcomes. For future theory development, Table 3 also (1) identifies how these four classifications can be directly associated with the power law perspective's input constructs and (2) suggests research questions that could be addressed.

In pure power law distributions, outliers disproportionately influence the statistical and behavioral characteristics of the entire system (e.g., Jeff Bezos's $100B+ net worth or Apple's $2T+ market capitalization). In less extreme distributions like the ones documented in our study, however, the outliers are much less influential on the larger system, suggesting instead that entrepreneurs must possess unique combinations of not-quite-outlier input variables—what Clark et al. (2023) call "critical configurations"—that could still generate outlier outcomes (c.f., Leppänen et al., 2023). Our results also support calls for using agent-based models to develop theory about which mechanism(s) drive individual agent (i.e., entrepreneur) behaviors at multiple levels that can demonstrate how micro-level interactions emerge into macro-level outcomes (Dimov and Pistrui, 2020).

For example, Bort et al. (2023) presented a model exploring the competitive edge of chronic impulsivity by simulating agents in competition for resources, demonstrating a skewed distribution of amassed resources over time. This study could serve as a blueprint for future explorations in entrepreneurship to discern among the generative mechanisms of power law distributions as described by Andriani and McKelvey (2009) and Boisot and McKelvey (2010). This model's validation and open availability can enable scholars to delve into diverse research questions, many of those in Table 3, by modifying the model's parameters to align with real-world data on new ventures, their initial conditions, potential emergence, and growth. This iterative process of validation and refinement through comparing simulated and empirical outcomes assists in pinpointing the mechanism(s) most coherent with entrepreneurial dynamics, thereby contributing significantly towards the enhancement of entrepreneurship theory. Models like this—mixed with nonlinear, non- or semi-parametric or quantile regression, and nonlinear correlation techniques—can build and test theories that explain and predict outlier-based phenomena at multiple levels with a high degree of utility and plausibility.

5.3. Implications for policy and practice

For policy—the "governance principles that guide courses of action and behavior in organizations and societies" (Aguinis et al., 2022, p. 858)—our findings suggest that community and scholastic programs to discover and increase the participation of outlier (i.e., star performer, Aguinis and O'Boyle, 2014) individuals in new venture creation at multiple stages of development could build opportunity recognition skills, early self-efficacy, resilience, and an entrepreneurial mindset, the combination of which increases the probability of both startup survival and individual satisfaction (White and Hertz, 2022). Similarly, as part of a corporate entrepreneurship strategy, executives and managers should encourage and support high-potential and high-performing employees to explore intrapreneurial opportunities that facilitate continual firm growth (Crawford and Kreiser, 2015; Kuratko et al., 2015). Thus, institutional programs dedicated to supporting innovation can serve as catalysts for transformative change, unleashing the potential of exceptional individuals to embark on groundbreaking endeavors. By providing essential resources and removing constraints, these programs not only empower visionary thinkers and creators but also set the stage for extraordinary leaps in progress and development. As such, the investment in such initiatives promises not just incremental advancements, but the possibility of explosive growth and disruptive breakthroughs, heralding a new era of societal advancement.

For practice, our findings suggest that the importance placed on initial conditions and interactions inherent in the power law perspective should focus on building knowledge resources (i.e., endowments) and interacting with potential stakeholders (i.e., engagement) as early as feasible in the nascent stage of organizing—even *before* building a minimum viable product, what Savoia (2019) calls a "pretotype," to understand whether customers are actually interested in the venture idea. Actions like this can increase the volume and frequency of feedback loops, leading to more opportunities to adjust product/service offerings to proven customer needs, improving product-market fit, and decreasing the time to profitability, the combination of which is vital for growth (Davidsson et al., 2009).

In conclusion, addressing these research areas will enhance our theoretical understanding of distribution shapes and outliers in entrepreneurship and provide practical insights for identifying, nurturing, and leveraging outlier talents in the entrepreneurial landscape. In particular, studies should acknowledge the outliers and build them into the development of new theory, rather than explaining them away as anomalies. Our proposed research agenda on generative mechanisms—which emphasizes the importance of identifying the distribution's shape and explaining the mechanisms that drive the emergence of outliers in heavy-tailed distributions—offers clear directions for future study. This agenda emphasizes the emergence of non-normal distributions and not only "tolerates," but explicitly acknowledges the presence and outsized influence of outliers.

CRediT authorship contribution statement

G. Christopher Crawford: Conceptualization, Data curation, Investigation, Project administration, Writing – original draft, Writing – review & editing. **Harry Joo:** Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Herman Aguinis:** Writing – original draft, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4.0 in order to proofread and enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge editor Philip Roundy and one anonymous reviewer for their highly developmental insights that substantially improved the content and positioning of the manuscript. We recognize the significant financial support for the CAUSEE study that made part of this research possible. The CAUSEE data collection, documentation, and release were funded by the Australian Research Council (grants DP0666616 and LP0776845); the QUT Business School; industry partners BDO Australia and National Australia Bank; and the Australian Government Department of Industry, Innovation and Science. A portion of the research described in this manuscript was presented at the Babson College Entrepreneurship Research Conference (Waterford, Ireland, June 2018). Crawford deeply appreciates the generous funding from The National Science Foundation (Science of Organizations Program), grant number 1734567 titled "Modeling the Emergence of Outliers in Entrepreneurship" to G. Christopher Crawford (PI); the Ewing Marrion Kauffman Foundation's Dissertation Fellowship for "Causes of Extreme Outcomes in Entrepreneurship: Endowments, Expectations, Engagement, and Environments;" as well as The University of Tampa's International Professional Development Grant, Provost David Stern, and Dean Frank Ghannadian in the Sykes College of Business. Crawford also expresses gratitude for the constructive feedback and encouragement on the development of the power law perspective from Ted Baker, Zahra Booyavi, Chao Chen, Daniel Clark, Per Davidsson, Dimo Dimov, Matthew Ferrell, Hannibal Fleming, Casey Frid, Diana Hechavarria, Christian Lechner, Xuanye Li, Benyamin Lichtenstein, Christian Linder, †Bill McKelvey, Thomas Nelson, Robert Pidduck, Maija Renko, Jaehu Shim, Vitaliy Skorodziyevskiy, Didier Sornette, Ehsan Teymourian, Elisa Villani, Chris Willis, and co-authors Herman Aguinis and Harry Joo. Correspondence concerning this article should be addressed to Christopher Crawford, The University of Tampa, 401 W. Kennedy Blvd, Tampa, FL 33606 USA. Email: GCrawford@UT.edu.

Appendix 1

Distribution pitting method description

We used a falsification-based methodology called distribution pitting (Joo et al., 2017), which compares distinct theoretical distributions against one another regarding how well each fits a given sample (i.e., observed distribution). We implemented distribution pitting using the R package Dpit, available on the CRAN.

Distribution pitting uses three decision rules to identify the likely dominant theoretical distribution and its associated generative mechanism for each sample. The first decision rule is to generate and interpret distribution pitting statistics for each pairwise comparison of theoretical distributions. Given that Dpit considers seven theoretical distributions (i.e., pure power law, lognormal, exponential, power law with an exponential cutoff, normal, Poisson, and Weibull), there are 21 pairwise comparisons involved in the first decision rule. In turn, per pairwise comparison, the Dpit package provides two types of distribution pitting statistics: a loglikelihood ratio (LR) and its associated *p*-value. A positive LR value indicates greater empirical support for the theoretical distribution mentioned first in the focal pairwise comparison. In contrast, a negative LR value indicates greater empirical support for the secondly mentioned theoretical distribution. The *p*-value associated with each LR value indicates the extent to which the non-zero LR value is likely due to random fluctuations alone. Because the null hypothesis is set to $LR = 0$, the lower the *p*-value, the less likely that the LR value is simply due to chance. As recommended, we used a *p*-value cutoff 0.10 (Clauset et al., 2009). Based on the LR and associated *p*-values per sample, we ruled out any theoretical distribution found to have a significantly inferior fit to the sample compared to another theoretical distribution, even just once. If only one theoretical distribution was never identified as having a significantly inferior fit, we concluded that the particular theoretical distribution (i.e., the sole surviving theoretical distribution) is the likely dominant distribution for the sample.

If the first decision rule led to more than one surviving theoretical distribution for the focal sample, we used the second decision rule to exclude additional theoretical distributions. Rooted in the principle of parsimony, the second decision rule is to identify the theoretical distribution with more parameters as being the worse match to the sample. Although theoretical distributions with more parameters have an equivalent or superior fit to the sample, they are associated with reduced parsimony and, therefore, a higher risk that the fitted model will be sample-specific (i.e., not generalizable). Out of the 21 pairwise comparisons between theoretical distributions, three comparisons involve nested distributions: (1) pure power law (one parameter) is nested within power law with exponential cutoff (two parameters), (2) exponential distribution (one parameter) is nested within power law with exponential cutoff (two parameters), and (3) exponential distribution (one parameter) is nested within the Weibull distribution (two parameters). For example, if the exponential and Weibull distributions equally fit a sample, we identified the latter as having the worse fit and, thus, ruled it out.

If the first and second decision rules still led to more than one surviving theoretical distribution for the focal sample, we used the third decision rule, which is also based on the principle of parsimony. For a pairwise comparison between two theoretical distributions, the third decision rule is to identify the theoretical distribution with a greater range of possible distribution shapes as having the worse fit. Among the seven theoretical distributions considered, some theoretical distributions (i.e., lognormal, Poisson, and Weibull) are more "flexible" in that they can assume a broader range of distribution shapes encompassing the shapes of other theoretical distributions (i.e., pure power law, exponential, power law with an exponential cutoff, and normal), which in contrast are "inflexible" in that they have a narrower range of distribution shapes. So, for a given sample, if a flexible distribution and an inflexible distribution remained after using the second decision rule, we identified the flexible distribution (i.e., the theoretical distribution with a greater range of possible distribution shapes) as having the worse fit and, therefore, ruled it out.

References

Aguinis, H., Harden, E.E., 2009. Cautionary note on conveniently dismissing χ^2 goodness-of-fit test results: implications for strategic research. In: Bergh, D.D., Ketchen, D.J. (Eds.), Research Methodology in Strategy and Management, vol. 5. Emerald Group Publishing, UK, pp. 111–120. Howard House.

Aguinis, H., O'Boyle Jr., E., 2014. Star performers in twenty-first century organizations. Personnel Psychology 67 (2), 313–350.

Aguinis, H., O'Boyle, E., Gonzalez-Mulé, E., Joo, H., 2016. Cumulative advantage: conductors and insulators of heavy-tailed productivity distributions and productivity stars. Person. Psychol. 69, 3–66.

Aguinis, H., Ji, Y.H., Joo, H., 2018a. Gender productivity gap among star performers in STEM and other scientific fields. J. Appl. Psychol. 103, 1283–1306.

Aguinis, H., Martin, G.P., Gomez-Mejia, L.R., O'Boyle, E.H., Joo, H., 2018b. The two sides of CEO pay injustice: a power law conceptualization of CEO over and underpayment. Manag. Res. J. Iberoam. Acad. Manag. 16 (1), 3–30.

Aguinis, H., Jensen, S.H., Kraus, S., 2022. Policy implications of organizational behavior and human resource management research. Acad. Manag. Perspect. 36 (3), 857–878.

Aislabie, C., 1992. Sudden change in a model of small firm growth. Small Bus. Econ. 4, 307–314.

Amitrano, D., 2012. Variability in the power-law distributions of rupture events. Eur. Phys. J. Spec. Top. 205, 199–215.

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

Bak, P., Tang, C., Wiesenfeld, K., 1987. Self-organized criticality: an explanation of $1/f$ noise. Phys. Rev. Lett. 38 (4), 381–384.

Barabási, A.L., 2005. The origin of bursts and heavy tails in human dynamics. Nature 435 (7039), 207–211.

Baron, R.A., 1998. Cognitive mechanisms in entrepreneurship: why and when entrepreneurs think differently than other people. J. Bus. Ventur. 13 (4), 275–294.

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

Boisot, M., McKelvey, B., 2011. Connectivity, extremes, and adaptation: a power-law perspective of organizational effectiveness. J. Manag. Inq. 20, 119–133.

Booyavi, Z., Crawford, G.C., 2023. Different, but same: a power law perspective on how rock star female entrepreneurs reconceptualize "gender equality". J. Bus. Ventur. Insights 19, e00374.

Bort, J., Wiklund, J., Crawford, G.C., Lerner, D.A., Hunt, R.A., 2023. The strategic advantage of impulsivity in entrepreneurial action: an agent-based modeling approach. Entrep. Theory Pract. 10422587231178882.

Bradley, K.J., Aguinis, H., 2023. Team performance: nature and antecedents of nonnormal distributions. Organ. Sci. 34, 1266–1286.

Certo, S.T., Raney, K., Albader, L., Busenbark, J.R., 2023. Out of shape: the implications of (extremely) nonnormal dependent variables. Organ. Res. Methods, 10944281231167839.

Chiles, T.H., Meyer, A.D., Hench, T.J., 2004. Organizational emergence: the origin and transformation of Branson, Missouri's musical theaters. Organ. Sci. 15, 499–519.

Clark, D.R., Crawford, G.C., Pidduck, R.J., 2023. Exceptionality in entrepreneurship: systematically investigating outlier outcomes. J. Bus. Ventur. Insights. *JBVI_e0422*.

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

Crawford, G.C., 2012. Disobeying power laws: perils for theory and method. J. Organ Dysfunct. 1, 75–81.

Crawford, G.C., 2015. The emergence of outliers in entrepreneurship: a self-organized criticality framework. Front. Entrepr. Res. 35 (16). Article 1.

Crawford, G.C., Kreiser, P.M., 2015. Corporate entrepreneurship strategy: extending the integrative framework through the lens of complexity science. Small Bus. Econ. 45, 403–423.

Crawford, G.C., McKelvey, B., 2018. Using maximum likelihood estimation methods and complexity science concepts to research power law-distributed phenomena. In: Handbook of Research Methods in Complexity Science. Edward Elgar Publishing, pp. 227–253.

Crawford, G.C., Linder, C., Lechner, C., Villani, E., 2023. Outlier entrepreneurs: nonlinear paths and novel ventures. J. Bus. Ventur. Insights. JBVI_e00437.

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. Ventur. Insights 1, 3–7.

Crawford, G.C., Aguinis, H., Lichtenstein, B., Davidsson, P., McKelvey, B., 2015. Power law distributions in entrepreneurship: implications for theory and research. J. Bus. Ventur. 30, 696–713.

Davidsson, P., Wiklund, J., 2006. Conceptual and empirical challenges in the study of firm growth. In: Sexton, D., Landström, H. (Eds.), The Blackwell Handbook of Entrepreneurship, pp. 179–199. Blackwell.

Davidsson, P., Steffens, P., Fitzsimmons, J., 2009. Growing profitable or growing from profits: putting the horse in front of the cart? J. Bus. Ventur. 24, 388–406.

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

Derbyshire, J., Garnsey, E., 2014. Firm growth and the illusion of randomness. J. Bus. Ventur. Insights 1, 8–11.

Dimov, D., Pistrui, J., 2020. Recursive and discursive model of and for entrepreneurial action. Eur. Manag. Rev. 17, 267–277.
Drazin, R., Sandelands, L., 1992. Autogenesis: a perspective on the process of organizing. Organ. Sci. 3 (2), 230–249.
Gabaix, X., Ibragimov, R., 2011. Rank—1/2: a simple way to improve the OLS estimation of tail exponents. J. Bus. Econ. Stat. 29, 24–39.
Gala, K., Schwab, A., Mueller, B.A., 2024. Star entrepreneurs on digital platforms: Heavy-tailed performance distributions and their generative mechanisms. J. Bus. Ventur. 39 (1), 106347.
Garnsey, E., Stam, E., Heffernan, P., 2006. New firm growth: exploring processes and paths. Ind. Innovat. 13, 1–20.
Gibrat, R., 1931. Les inégalités économiques. Recueil Sirey.
Joo, H., Aguinis, H., Bradley, K.J., 2017. Not all nonnormal distributions are created equal: improved theoretical and measurement precision. J. Appl. Psychol. 102, 1022–1033.
Kuratko, D.F., Hornsby, J.S., Hayton, J., 2015. Corporate entrepreneurship: the innovative challenge for a new global economic reality. Small Bus. Econ. 45, 245–253.
Leitch, C., Hill, F., Neergaard, H., 2010. Entrepreneurial and business growth and the quest for a "comprehensive theory": tilting at windmills? Entrep. Theory Pract. 34, 249–260.
Leppänen, P., George, G., Alexy, O., 2023. When do novel business models lead to high performance? A configurational approach to value drivers, competitive strategy, and firm environment. Acad. Manag. J. 66 (1), 164–194.
Lichtenstein, B.B., 2000. Self-organized transitions: a pattern amid the chaos of transformative change. Acad. Manag. Perspect. 14, 128–141.
Markman, G.D., Gartner, W.B., 2002. Is extraordinary growth profitable? A study of Inc. 500 high–growth companies. Entrep. Theory Pract. 27, 65–75.
McKelvey, B., 2004. Toward a complexity science of entrepreneurship. J. Bus. Ventur. 19 (3), 313–341.
Mitzenmacher, M., 2004. A brief history of generative models for power law and lognormal distributions. Internet Math. 1, 226–251.
Nicholls-Nixon, C.L., 2005. Rapid growth and high performance: the entrepreneur's "impossible dream?". Acad. Manag. Perspect. 19, 77–89.
O'Boyle Jr., E., Aguinis, H., 2012. The best and the rest: Revisiting the norm of normality of individual performance. Person. Psychol. 65 (1), 79–119.
Podobnik, B., Crawford, G.C., Lichtenstein, B., Lipic, T., Wild, D., Zhang, X., Stanley, H.E., 2020. The new wealth of nations: How STEM fields generate the prosperity and inequality of individuals, companies, and countries. Chaos, Solit. Fractals 141, 110323.
Poole, M.S., Van de Ven, A.H., Dooley, K., Holmes, M.E., 2000. Organizational Change and Innovation Processes: Theory and Methods for Research. Oxford University Press.
Savoia, A., 2019. The Right It: Why So Many Ideas Fail and How to Make Sure Yours Succeed. In: HarperOne.
Shim, J., 2016. Toward a more nuanced understanding of long-tail distributions and their generative process in entrepreneurship. J. Bus. Ventur. Insights 6, 21–27.
Simon, H., 1955. On a class of skew distribution functions. Biometrika 44, 425–440.
Simon, H.A., 1993. Strategy and organizational evolution. Strat. Manag. J. 14, 131–142.
Song, C., Koren, T., Wang, P., Barabási, A.L., 2010. Modelling the scaling properties of human mobility. Nat. Phys. 6 (10), 818–823.
Sornette, D., Ouillon, G., 2012. Dragon-kings: mechanisms, statistical methods and empirical evidence. Eur. Phys. J. Spec. Top. 205 (1), 1–26.
Stumpf, M.P.H., Porter, M.A., 2012. Mathematics: critical truths about power laws. Science 335, 665–666.
Su, E., Holt, D.T., Pollack, J.M., 2019. The distribution of family firm performance heterogeneity: understanding power law distributions. In: Memili, E., Dibrell, C. (Eds.), The Palgrave Handbook of Heterogeneity Among Family Firms: 407–429. Palgrave Macmillan.
Taleb, N.N., 2007. The Black Swan: the Impact of the Highly Improbable. Random House.
Taleb, N.N., 2020. Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications. *arXiv preprint arXiv:2001.10488*.
Thietart, R.A., 2016. Strategy dynamics: agency, path dependency, and self-organized emergence. Strat. Manag. J. 37, 774–792.
Thietart, R.A., Malaurent, J., 2023. How do social media contribute to the emergence of "extreme" events? The rise of Tour Paris 13: The "Sistine Chapel" of street art. Acad. Manag. Discov. https://doi.org/10.5465/amd.2022.0229.
White, R.J., Hertz, G.T., 2022. Teaching opportunity recognition: Meeting the challenges of message, method and measurement. J. Entrepreneursh. 31 (2S), S52–S78. https://doi.org/10.1177/09713557221107436.