Journal of Business Venturing Insights 22 (2024) e00492

Contents lists available at ScienceDirect

Journal of Business Venturing Insights

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

Stars everywhere: Revealing the prevalence of star performers using empirical data published in entrepreneurship research

Kaushik Gala *, Andreas Schwab

Department of Management and Entrepreneurship, Ivy College of Business, ISU, USA

A R T I C L E I N F O

Keywords:
Star performers
Star entrepreneurs
Performance distributions and outliers
Skewness
Simulations and heuristics

A B S T R A C T

Scholars have long called for moving beyond a narrow focus on average performance toward a more direct investigation of the variance in performance. While a few studies have evaluated *star entrepreneurs*, most empirical research continues to focus on average performers. This lacuna has constrained not only the development of theories but also the accumulation of data on the distribution of performance. In response, this study uses simulations and heuristics to extract distributional information from descriptive statistics commonly reported in published research (i.e., mean, standard deviation, and sample size). Applying this approach to studies recently published in high-impact entrepreneurship journals shows that (a) the suggested methodology can provide rough estimates of the skew and shape of performance distributions, and (b) right-skewed, heavy-tailed distributions featuring star performers are ubiquitous in entrepreneurship, thus reinforcing calls for more direct studies of performance distributions in entrepreneurship.

1. Introduction

Management research has increasingly recognized an important pattern across organizational settings – a majority of poor and modest performers and a few *star* performers (Aguinis and O'Boyle Jr, 2014; Bradley and Aguinis, 2023; Joo et al., 2017). In response, scholars have called for explicitly investigating these 'outliers' to enrich the theoretical understanding of the antecedents, mechanisms, and boundary conditions of star performance in organizational contexts (Gibbert et al., 2021; Hymer and Smith, 2024). However, empirical studies of this important phenomenon remain rare despite calls for the comprehensive examination and reporting of performance variance (Aguinis et al., 2013; Maula and Stam, 2020). Thus, management research faces the substantial risks that arise from adopting a Gaussian perspective focused on central tendencies and ignoring right-skewed, heavy-tailed distributions of performance (Andriani and McKelvey, 2009; McKelvey and Andriani, 2005).

Entrepreneurship research, too, tends to focus on *average* performance and largely avoids a fine-grained investigation of the observed *spectrum* of performance (Crawford et al., 2014, 2015; Dean et al., 2007). This myopic tendency persists despite decades of research and an accumulated body of literature with thousands of peer-reviewed publications seeking to understand the drivers of entrepreneurial performance (Landström et al., 2012). Indeed, studies capturing the tails of performance distributions are especially critical in entrepreneurial contexts because star entrepreneurs and ventures often contribute the majority of socio-economic value (Clark et al., 2023; Crawford et al., 2024a; Gala et al., 2024).

Pioneering studies of star performance in entrepreneurship used publicly available datasets that report performance for participants, such as the Kauffman Firm Survey and the Panel Study of Entrepreneurial Dynamics (e.g., Crawford, 2013; Crawford et al.,

* Corresponding author. 3219 Gerdin, 2167 Union Dr, Ames, IA, 50011, USA.
E-mail address: kgala@iastate.edu (K. Gala).

https://doi.org/10.1016/j.jbvi.2024.e00492
Received 30 May 2024; Received in revised form 20 August 2024; Accepted 23 August 2024
Available online 21 September 2024

2015). However, replicating and extending their findings to a broader array of entrepreneurial contexts has been challenged by (a) the lack of reporting of distributional characteristics and (b) the lack of sharing of raw data in published studies (Crawford et al., 2022). Therefore, this study investigates the following question: *What can we infer about the skewness and shapes of performance distributions from empirical studies published in high-impact entrepreneurship journals?*

Notably, strategic management research often involves publicly listed firms, with many papers focused on US-based firms that constitute market indices such as the S&P 500 (Richard et al., 2009). This commonality enabled Certo et al. (2024) to use public data to investigate the tails of performance distributions for public firms. They reported estimates of variance (i.e., skewness and kurtosis) for measures such as net income and stock returns and evidence of right-skewed, heavy-tailed distributions. In contrast to strategic management research, investigations into entrepreneurial performance primarily focus on private firms, which are not required to report performance data publicly. Indeed, most empirical studies of entrepreneurship use proprietary data, do not share such data, and rarely report skewness of performance. To address this gap, we explore the possibility of *'reverse engineering'* the skewness and shapes of distributions using commonly reported descriptive statistics.

2. Empirical studies of entrepreneurial performance (2021–2023)

To explore the feasibility of 'reverse engineering' skewness and distribution shapes, we identified recently published papers in three high-impact entrepreneurship journals: Entrepreneurship Theory and Practice (ETP), Journal of Business Venturing (JBV), and Strategic Entrepreneurship Journal (SEJ). We focused on the 2021–2023 period to capture (a) recently reported performance data and (b) current practices of measuring and reporting entrepreneurial performance. This focus also enabled us to examine the extent to which researchers have heeded editorial calls to move beyond a simplistic focus on averages by explicitly investigating the distribution of performance. We started with an initial pool of all papers published during 2021–2023 in the above three journals, as shown in Table A1.1.

Notably, the distributional perspective we adopt and promote in this study relies upon the outcome of interest being measured on a continuous scale. Thus, performance distributions are more relevant for quantitative measures, such as revenue, but less so for (a) binary measures, such as whether a business survived or not, or (b) categorical measures, such as whether the performance was 'poor,' 'good,' or 'excellent.'

2.1. *Continuous and positive measures of performance*

We identified 123 studies that reported descriptive statistics for continuous measures of performance (see Table A1.2). We included revenue, revenue growth, profitability, number of employees, and assets as measures of venture performance (Delmar, 2019; Murphy et al., 1996). Given the nascency of ventures typically studied in entrepreneurial research, we also included funds raised from angel investors, venture capital firms, and crowdfunding platforms as alternative performance measures.

However, we excluded crowdfunding goals and intentions because they did not reflect actual or realized performance. We further excluded distant performance measures such as personal wealth, new product introductions, and patent counts. We also excluded studies reporting the performance of angel and venture capital investors and the performance of sports teams, artists, and chess players to remain consistent with this study's focus on *venture* performance. Of the 123 studies that reported venture performance, 109 reported *continuous* measures of performance. Thus, fourteen studies were excluded because they reported either binary measures (e.g., firm survival) or ordinal measures (e.g., managerial perceptions of firm performance measured using Likert scales).

For conceptual and analytical simplicity, we focus on continuous performance measures that only take on positive values. For example, we include revenue and number of employees but exclude revenue growth and employee growth because the latter can be both positive and negative. Thus, of the 109 studies shortlisted above, we exclude six studies that report only growth outcomes. Another ten studies are excluded because they do *not* report the mean, standard deviation (SD), or both; we need these as inputs for our heuristic- and simulation-based analyses. Thus, the final set consists of 93 studies, listed in **Appendix 1**. Because some of these studies report multiple performance measures, the final sample consists of 122 reported measures of performance. These studies encompass a wide variety of entrepreneurial actors (e.g., digital platform entrepreneurs, family businesses, and venture-funded startups), industries (e.g., health sciences, high-tech, and manufacturing), and geographic regions (e.g., Asia, Europe, and North America).

Some of the studies in the sample also report the minimum value (41%), the maximum value (40%), and the median value (14%). Moreover, 66% of studies transform the performance variable, with all but one of these studies using log transformations. Only a third of these studies justify transforming the performance measure, with the most common rationale being that the untransformed performance data was "highly" or "strongly" skewed. Only 5% of the studies in the sample report how they handled outliers. Notably, only one paper out of 93 – Hunt et al. (2022) – reports the value of skewness. For the two continuous measures in their study, they report skewness values of 4.01 and 8.66, indicating severe non-normality based on a commonly used threshold of skewness greater than 2.0 (Blanca et al., 2013; Micceri, 1989; West et al., 1995).

2.2. *Coefficient of variation*

Given the mean and standard deviation of performance, we can compute the *actual* values for the coefficient of variation (CoV), a standardized measure of inequality for continuous measures (Allison, 1978). For the 122 performance measures across the 93 studies, the CoV varies from 0.1 to 11.6, as shown in Fig. 1. More than half of these values exceed 1.0, indicating relatively high variance (Allison, 1980). This high volatility around the reported mean makes the latter less informative as an estimate of performance. This high volatility could be symmetric around the mean (Kokko et al., 1999); a lack of such symmetry, however, would imply relatively high levels of skewness (Doane and Seward, 2011).

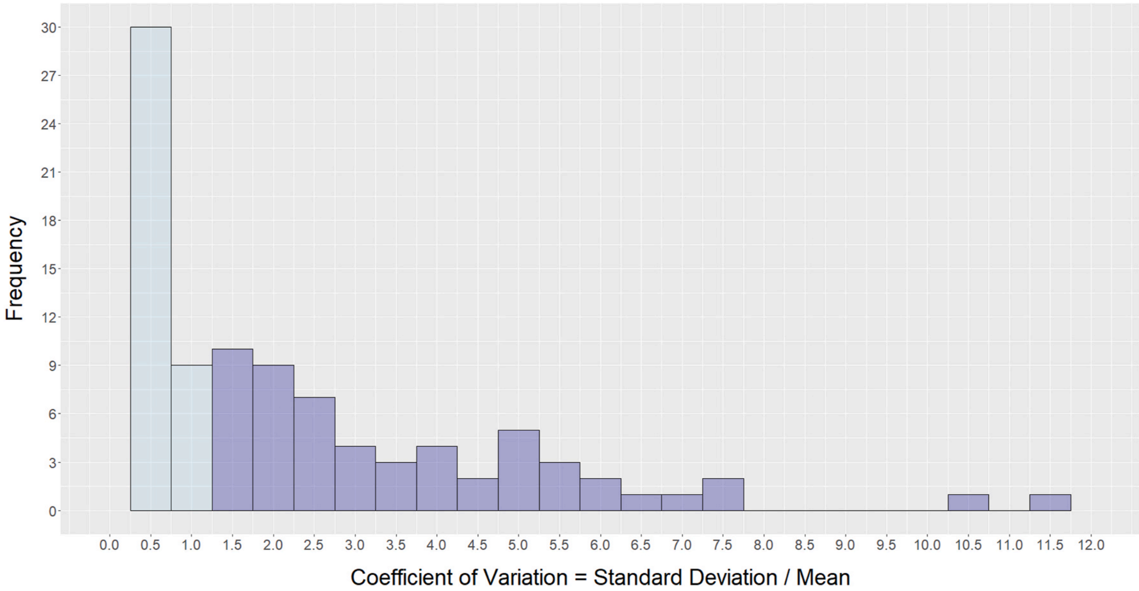

Fig. 1. Distribution of Coefficient of Variation based on reported mean and standard deviation.
Note: Bars in darker shade indicate CoV >1.25 (i.e., relatively high variance around the mean).
(122 measures from 93 papers published in JBV, ETP, and SEJ during 2021–2023).

2.3. Preliminary findings

Our analysis of this sample of studies reveals two critical issues. First, we observe the absence of descriptive information about the distribution of performance, including the omission of skewness an absence of explicit investigations of distributional shapes. Second, we note a lacuna in sharing raw data via online supplements or open-source data repositories (Grégoire et al., 2024). Had the underlying performance data been publicly available for the papers in the sample, we could have computed skewness for each reported measure. Moreover, we could have used distribution pitting techniques such as those employed by Crawford et al. (2024a), Joo et al. (2017), and Gala et al. (2024) to identify the best fitting distribution shapes (e.g., lognormal, power law).

3. Inferring the skewness of performance using simulations

Given the absence of reported values of skewness of performance or sharing of raw data, this study explores whether simulations can help estimate skewness using commonly reported descriptive statistics, specifically mean, standard deviation, and sample size.

3.1. Simulation approach

To estimate skewness, we first need to identify the family of distributions that best 'fits' the observed data. However, none of the studies in the sample explicitly identifies the shape of performance distributions. Therefore, consistent with prior research on star performers in the organizational literature, we consider four unimodal families of distributions: normal (i.e., Gaussian), exponential (i.e., gamma family), lognormal, and power law (i.e., Pareto family). These distribution families *"likely explain the majority of natural phenomena"* (Joo et al., 2017, p. 1025) and have been found to adequately characterize many distributions of performance in organizational contexts (Bradley and Aguinis, 2023; Crawford et al., 2024a). Notably, the exponential distribution involves the sum of two or more distributions, while the lognormal distribution involves a multiplicative combination of two or more distributions (Kleiber and Kotz, 2003; Limpert et al., 2001). Fig. 2 illustrates these four families.

For each reported performance measure, we generated random numbers from four distributions, one for each unimodal family of distributions, as shown in Fig. 3. We ensured that the simulated datasets had the same sample size and *approximately* the same mean and standard deviation as reported in the corresponding paper. We repeated this process for each distribution family fifty times to generate fifty performance distributions. We then computed an estimate of skewness as the average of the respective skewness values across the fifty distributions generated for that family. Thus, for each performance measure, we generated four averaged estimates of skewness, one for each distribution family.

3.2. Validation of the simulation approach

Before applying this simulation approach to the 93 entrepreneurship studies in the sample, we tested the accuracy of simulation-based estimates of skewness using statistical data published in studies that explicitly investigated performance distributions, specifically Aguinis et al. (2016) and Crawford et al. (2015). These studies include numerous *continuous* and *positive* measures of performance with a wealth of statistics such as minimum, median, maximum, and skewness, a practice largely absent in entrepreneurship research.

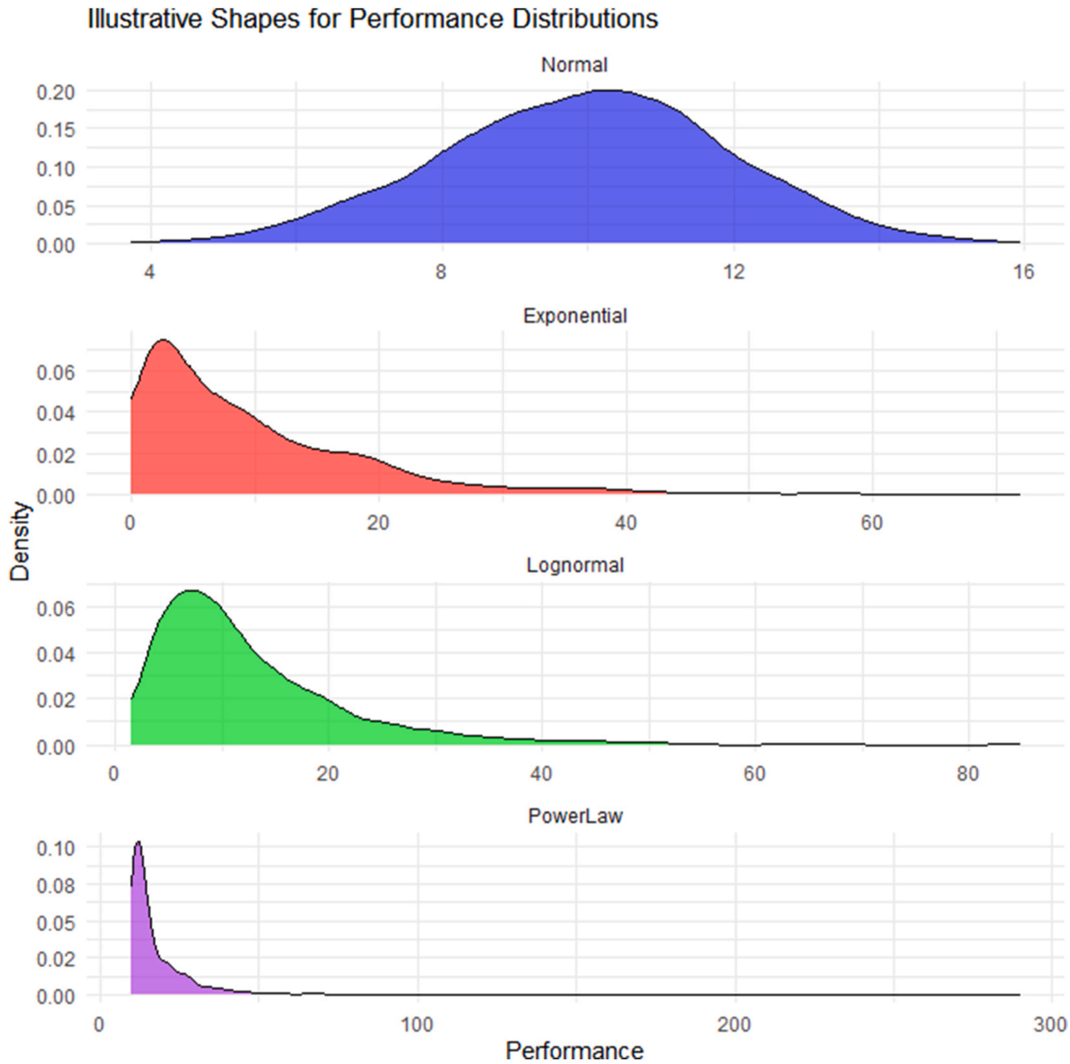

Fig. 2. Depiction of Generic distributions.
Note: This figure use different scales for both axes.

Of the four chosen distribution families, lognormal distributions provide relatively more *accurate* (average error of 14%)*,* whereas gamma distributions provide relatively more *conservative* (average error of −33%) estimates of skewness for the performance measures reported by Aguinis et al. (2016). Similarly, for the performance measures reported by Crawford et al. (2015), lognormal distributions provide relatively more *accurate* (average error of 26%)*,* whereas gamma distributions provide relatively more *conservative* estimates of skewness (average error of −28%).

Importantly, our simulations failed to generate random numbers fitting normal distributions in 71%–94% of the cases, despite 10,000 attempts in each iteration. This finding reflects the severe non-normality of performance in most entrepreneurial contexts. Moreover, the Pareto distribution provided the least accurate estimates. In theory, Pareto distributions such as the pure power law have infinite mean and, depending on the shape parameter, potentially infinite variance (Clauset et al., 2009; Stumpf and Porter, 2012). In the context of entrepreneurship, Pareto distributions may be less likely candidates than the lognormal or gamma distributions due to constraints on venture performance imposed by factors such as specialization, market size, dis-economies of scale, limited resources, and governmental regulation (Canback et al., 2006; Fan et al., 2021).

In contrast to our findings, Crawford et al. (2024) found that pure power laws best characterized the distribution of performance – when measured as the number of employees – for US-based nascent ventures in Waves 4 and 5 of the Panel Study of Entrepreneurial Dynamics (PSED II). Their findings, however, indicate increasing inequality in venture performance over time, manifesting as a shift from power law with exponential cutoff distributions to pure power law distributions over time. In line with our findings, Gala et al. (2024) find support for lognormal distributions as a suitable fit for distributions of entrepreneurial performance on a digital platform; similar findings for lognormal distributions were also reported by Crawford et al. (2024).

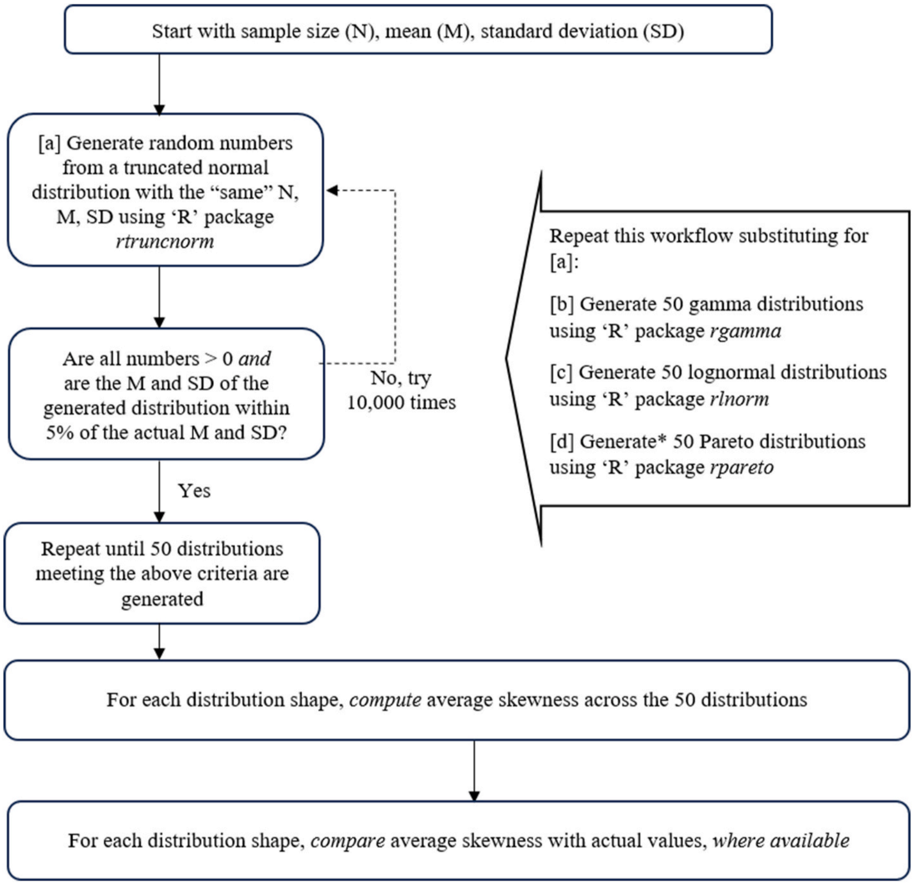

Fig. 3. Simulation workflow for each continuous and positive measure of performance.[2]
*Note: *rpareto* requires X_{min} and α as inputs. We vary X_{min} from 0.1 * M up to 5.0 * M in 1000 increments, and α from 4.0 down to 1.0 in 1000 increments. Thus, we make one million attempts to generate one Pareto distribution with the same N and "similar" M and SD as the actual.

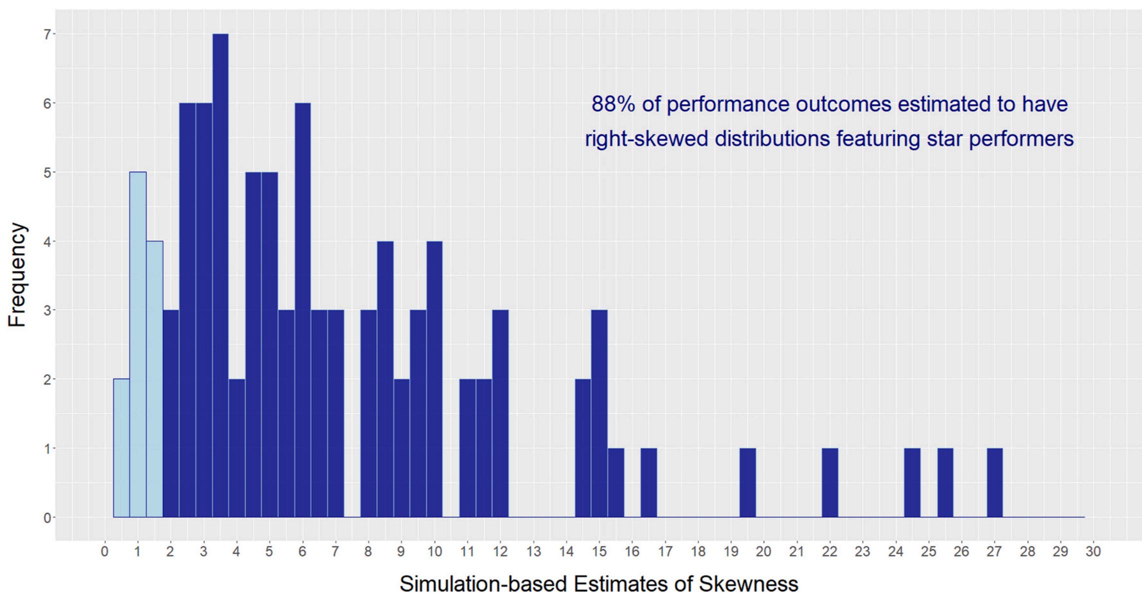

Fig. 4. Distribution of simulation-based estimates of skewness.
Note: n=95; Bars in darker shade indicate skewness > 2.0; Using data reported in ETP, JBV, SEJ (2021-2023)

Appendix 2 reports detailed findings related to the accuracy of skewness estimated using simulations. In sum, these results suggest that skewness can be estimated – albeit only to a rough approximation – by 'reverse engineering' the commonly reported descriptive statistics (i.e., mean, standard deviation, and sample size) for continuous and positive measures of performance.

3.3. *Simulation-based estimation of skewness*

The preceding validation of the simulation approach identified the gamma distribution as the most suitable for generating relatively *conservative* estimates of skewness with accuracy approaching that of the lognormal distribution. Therefore, we used this distribution to estimate skewness for continuous and positive performance measures reported in the 93 empirical studies in the sample. Fig. 4 illustrates the range of skewness values generated using simulations. The majority of estimates of skewness (87%) are greater than 2.0, indicating high variance (Micceri, 1989; West et al., 1995). To the extent that these estimates reflect the *actual* skewness of performance, they highlight the presence and prevalence of star entrepreneurs.

These simulation-based estimates of skewness are subject to a crucial limitation. An infinite number of distributions can theoretically have the same mean and standard deviation given a specific sample size. Anscombe's (1973) quartet is a compelling illustration of this statistical reality. For example, Fig. 5 shows four datasets with different scatterplots yet the same sample size and nearly identical descriptive statistics (i.e., mean and standard deviation).

Even though studies have shown exponential (i.e., gamma-family) distributions as a better fit than normal distributions for *some* performance outcomes (Crawford et al., 2024a; Gala et al., 2024), the assumption of the gamma distribution for *all* performance measures in this study may limit the overall accuracy of the simulation results. Nevertheless, in the absence of entrepreneurship studies explicitly reporting the skewness of performance, simulation-based estimates can serve as rough approximations, while the limitations of simulation-based methodologies will reinforce calls for studies to explicitly report distributional information (Crawford et al., 2022; Clark et al., 2023).

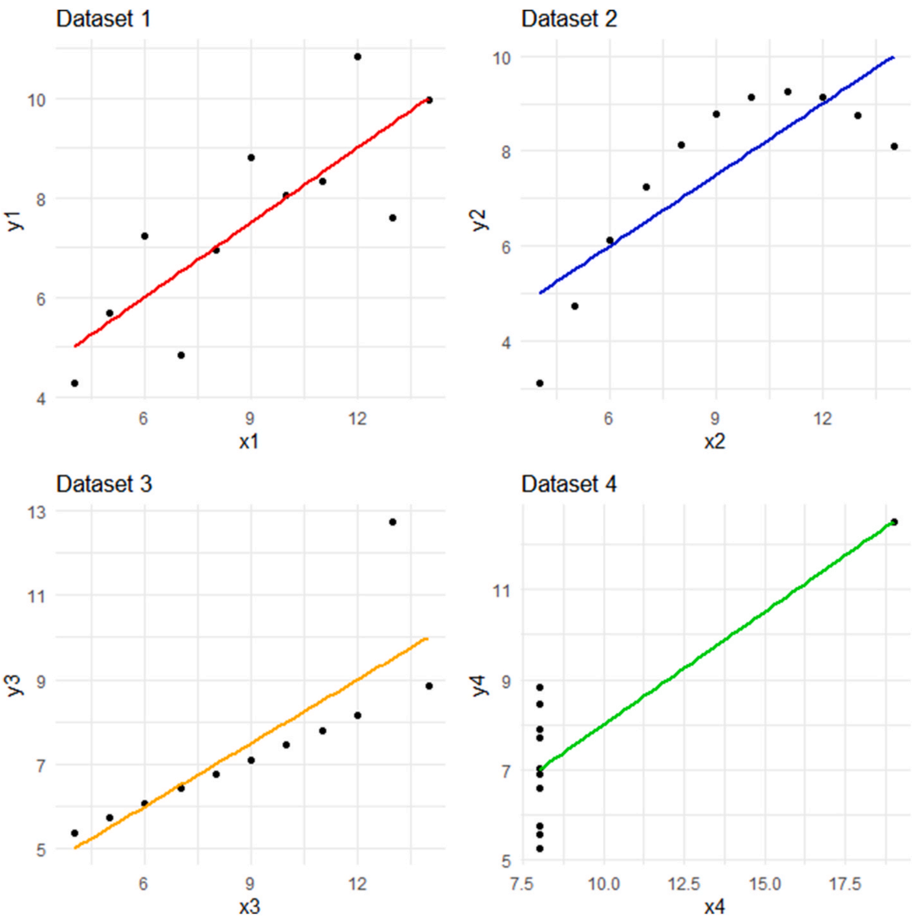

Fig. 5. Anscombe's quartet.
Note: All four datasets above have nearly identical means (9, 7.5) and variance (11, 4.125) for the two variables (x, y).

4. Inferring the shapes of performance distributions

In a second step, we move beyond estimating skewness and introduce a methodology developed by statisticians and used in other fields (e.g., econophysics) to infer distributional shapes using skewness, mean, and standard deviation. Specifically, we use a heuristic tool called the moment-ratio (MR) plot, which plots the skewness of a variable on the Y-axis and the corresponding Coefficient of Variation (CoV) on the X-axis (Cirillo, 2013; Vargo et al., 2010; Vogel and Fennessey, 1993). The reference chart is split into zones, each associated with a specific family of distributions, as illustrated in Cirillo (2013, p. 5954, Fig. 5). If the observed performance falls in the Paretian zone, its distribution *likely* belongs to the Pareto family (Cirillo, 2013). Thus, the moment-ratio plot serves as a simple yet useful heuristic to identify the distributional shape that 'best' represents the observed performance in a particular dataset[1]. **Appendix 3** provides further details on how these moment-ratio plots are created.

4.1. Validation of the moment-ratio plot

To assess the accuracy of MR plots in identifying the most likely distribution shape, we applied this heuristic to the performance measures reported by Aguinis et al. (2016). The shapes inferred from the MR plots were compared to those reported by Joo et al. (2017), who applied a distribution-pitting methodology to these same datasets. Distribution pitting seeks to identify which one of several distribution shapes best fits a given dataset. For example, Joo et al. (2017) developed a methodology that considers seven shapes (i.e., normal, Poisson, Weibull, exponential, lognormal, power law, and power law with exponential cutoff).

Findings indicate that MR plots are reasonably accurate in identifying the exponential (i.e., gamma family) and lognormal distributions as the 'best fits' for the measures reported by Aguinis et al. (2016). Similarly, MR plots are reasonably accurate in identifying the power law with exponential cutoff (i.e., gamma family) and lognormal distributions as the 'best fits' for 24 *outcome variables* reported in Crawford et al. (2015; 2024a). See **Appendix 3** for further details on these validation results.

4.2. Application of the moment-ratio plot

To illustrate the value of moment-ratio plots, we created one for each of the three continuous performance measures reported by Hunt et al. (2022), the only paper in our set of 93 studies that reports skewness. To demonstrate generalizability, we searched for additional studies that (a) reported mean, SD, and skewness for performance and (b) were published during the 2000–2020 period in the same three journals (see Table A3.2 in Appendix 3 for details). Fig. 6 shows the shapes inferred using reported values of skewness, mean, and SD of performance. For the Hunt et al. (2022) study, one measure (i.e., number of businesses failed) falls in the exponential zone while the other two (i.e., jobs created and number of businesses founded or acquired) fall in the lognormal zone.

Thus, MR plots enable the identification of likely distribution shapes when only three descriptive statistics (i.e., mean, standard deviation, and skewness) are available. This evaluation of performance variability goes beyond the normal versus power-law dichotomy that dominated early research. Instead, the MR plot allows a more nuanced view of performance distributions and, consequently, opportunities to capture the multifaceted nature of entrepreneurship (Kuckertz et al., 2023).

Such fine-grained analyses are valuable for three reasons. First, research has highlighted the limitations of methods based on p-values in differentiating between different types of heavy-tailed distributions (Broido and Clauset, 2019; Mitzenmacher, 2004; Nair et al., 2022). Second, *pure* power law distributions have infinite means and often infinite variances, which limits their ability to accurately describe entrepreneurial performance, which typically faces many substantial constraints (Crawford et al., 2024a; Stumpf and Porter, 2012). Third, identifying specific distribution shapes – possibly using distribution pitting methods like those developed by Joo et al. (2017) – is an important step toward identifying potential mechanisms that engender star performers in entrepreneurship. Notably, only a single study in the sample reported skewness of performance. Thus, we reiterate calls for empirical studies to report, at the minimum, the skewness of performance outcomes to provide insights into the asymmetry of distributions and enable the use of MR plots for inferring the likely shape of such distributions.

5. Discussion

Entrepreneurship research has increasingly identified the substantial variance in performance and the disproportionate influence of star entrepreneurs (Crawford et al., 2024a; Gala et al., 2024). While scholars have begun to explore the antecedents and implications of non-normal distributions of entrepreneurial performance (Booyavi and Crawford, 2023; Crawford et al., 2024b; Khurana et al., 2023), most studies of venture performance fail to evaluate and report performance distributions. Extending the analyses of published empirical studies to reveal the impact of star performers is inhibited by studies failing to publish their underlying datasets and the performance data for private firms being confidential (Crawford et al., 2022).

Notably, Crawford et al. (2024a) provide a comprehensive list of methodologies used by extant empirical research to determine distribution shapes. These methods include simple goodness-of-fit statistics (e.g., chi-square), distribution pitting, and maximum likelihood estimation (Crawford et al., 2024; Table 1, p. 5). All such methods, however, require access to the raw data, which severely limits their applicability, given the current state of reporting practices in entrepreneurship research. In contrast, the moment-ratio plots and simulations introduced in this paper require only standard descriptive statistics for the variables of interest to estimate

[1] Given that we selected the *gamma* distribution to arrive at simulation-based estimates of skewness for the studies in our sample, it would be tautological to use these estimates in a moment-ratio plot to infer distributional shapes. Therefore, validation of the MR plots focuses on studies which reported not only mean and standard deviation but also skewness.

[2] 'R' code for the simulation methodology is available upon request.

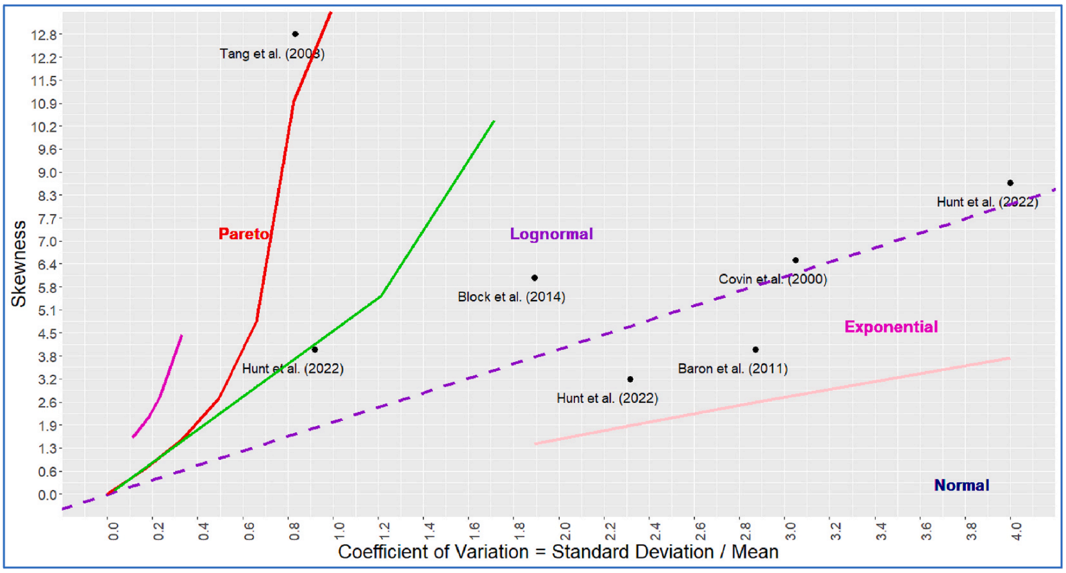

Fig. 6. Moment-ratio plot for illustrative studies.

distributional features of interest. Thus, the methods introduced in this paper help extract additional valuable insights into performance distributions using only three commonly reported descriptive statistics (i.e., mean, standard deviation, and sample size).

5.1. Implications and contributions

This study draws scholarly attention to the substantial variance in entrepreneurial performance by re-analyzing the descriptive statistics reported in studies published in high-impact journals through the lens of skewness and distributional shapes. Evidence for non-normal distribution of performance outcomes has been reported for over a decade in management research, with pioneering empirical studies by Aguinis and O'Boyle Jr, 2014 and Crawford et al. (2014, 2015). However, current empirical research in entrepreneurship rarely adopts a distributional perspective and instead continues to focus on average performance. Therefore, this study seeks to reinvigorate the discussion on performance distributions and outliers – it does so by extracting additional evidence of star performance from a broad range of average-focused studies.

Our findings of heavily right-skewed and non-normal distributions reinforce calls for an explicit investigation of performance distributions and outliers (Crawford et al., 2022; Clark et al., 2023; Gala et al., 2024). The consistency of (a) high skewness and high CoV values and (b) exponential and lognormal-like distributions when describing entrepreneurial performance highlights the strong influence of star performers on cumulative outcomes.

Furthermore, this study underscores the need for (a) future empirical studies to report statistics such as median, maximum, skewness, and kurtosis for more fine-grained insights into performance and (b) journals to demand such information (Maula and Stam, 2020; Wennberg and Anderson, 2020). By adopting such simple practices, scholars and editors can contribute to research on star entrepreneurs. Of even greater value would be efforts to include plots of performance distributions in published papers (Schwab, 2018) and to disclose raw performance data while addressing confidentiality issues (Quigley et al., 2023). Finally, this study highlights the opportunity to enrich entrepreneurship research by adopting methodological techniques from other disciplines, such as econophysics, hydrology, and quality technology, that regularly examine extreme outcomes (Cirillo, 2013; Vargo et al., 2010; Vogel and Fennessey, 1993).

5.2. Limitations and future research

Inferring the skewness and shapes of distributions using only a few descriptive statistics engenders some limitations. For example, the choice of four unimodal families of distributions may introduce inaccuracies – other families or a combination of two or more families may enhance the accuracy of the simulation methodology. Similarly, the choice of the gamma distribution to generate conservative estimates of skewness may severely underestimate the skewness of performance outcomes for some data sets. Thus, inferences about the prevalence of star performance may indicate lower-than-actual variance and extremity in the distribution of performance.

Accordingly, we encourage future studies to explore the simulation accuracy for newly introduced families of distributions, such as the exponentiated exponential and the double Pareto-lognormal, which may be better suited for specific entrepreneurial contexts (Gupta and Kundu, 2001; Reed and Jorgensen, 2004). Future research may also expand the scope to entrepreneurship studies published in journals such as the Academy of Management Journal, Journal of Applied Psychology, Journal of Management, and Strategic Management Journal to identify studies that not only report skewness and other distributional information and thus, further validate heuristic- and simulation-based approaches. Additionally, scholars may employ the heuristics and simulations introduced in this study for thought experiments that may lead to novel conceptual insights into and hypotheses about substantial variance in venture performance (Breig et al., 2018; Davis et al., 2007; Lu and Dimov, 2023).

This study examined only output variables, even though significant skewness has also been reported for input variables (e.g., number of owners, human capital, and social capital) in entrepreneurial contexts. Crawford et al. (2015), for example, report CoV values ranging from 0.5 to 81.6 and skewness values ranging from 1 to 29 for input variables. Such substantial variance and skew in inputs in representative samples of entrepreneurship calls for simulation-based investigations to examine the prevalence of skewness in various antecedents previously reported in entrepreneurship studies. Future research may also investigate whether and how non-normal distributions in antecedents transmute into non-normal distributions in outcomes (Shim, 2016).

Finally, this study's findings are limited in their ability to help infer the underlying generative mechanisms of right-skewed, heavy-tailed performance distributions. For example, Andriani and McKelvey (2009, Table 2, p. 1058) posit 15 mechanisms that often engender power law distributions. Thus, identifying the power law as the 'best fitting' shape using a distribution-pitting method does not necessarily translate into identifying a specific generative mechanism (Joo et al., 2017).

To establish clearer links between generative mechanisms and distribution shapes, future studies may leverage computational methods, such as agent-based modeling or system dynamics, that account for the influence of critical thresholds, feedback loops, initial conditions, and non-linear change on star performance (Bort et al., 2024; Crawford, 2009; Kearney and Lichtenstein, 2023; Lichtenstein, 2018). Such studies can draw upon the power law perspective conceived by organizational researchers (Andriani and McKelvey, 2009) and further developed by entrepreneurship scholars (Crawford, 2013; Crawford et al., 2014, 2015). This perspective involves a framework of constructs (i.e., resource endowments, expectations for future growth, and engagement) and draws upon complexity science to explicate the emergence of influential outliers.

5.3. Conclusion

This study introduces a heuristic- and simulation-based approach to the methodological toolbox for researchers interested in studying performance distributions. Moreover, it provides evidence for the prevalence of star performers across a wide variety of contexts by re-analyzing the statistics reported in recent entrepreneurship studies. In doing so, it highlights the need for researchers to adopt a distributional perspective and to directly examine the full range of performance, especially star performers.

Research support

There is no research support to disclose.

Relationships

There are no additional relationships to disclose.

Patents and intellectual property

There are no patents to disclose.

Other activities

There are no other activities to disclose.

CRediT authorship contribution statement

Kaushik Gala: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Investigation, Conceptualization. **Andreas Schwab:** Writing – review & editing, Supervision, Methodology, Conceptualization.

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.

Acknowledgements

We thank the editor Philip T. Roundy and an anonymous reviewer for their constructive comments throughout the review process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbvi.2024.e00492.

References

Aguinis, H., Gottfredson, R.K., Joo, H., 2013. Best-practice recommendations for defining, identifying, and handling outliers. Organ. Res. Methods 16 (2), 270–301.
Aguinis, H., O'Boyle Jr, E., 2014. Star performers in twenty-first century organizations. Person. Psychol. 67 (2), 313–350.
Aguinis, H., O'Boyle Jr, E., Gonzalez-Mulé, E., Joo, H., 2016. Cumulative advantage: conductors and insulators of heavy-tailed productivity distributions and productivity stars. Person. Psychol. 69 (1), 3–66.
Allison, P.D., 1978. Measures of inequality. Am. Socio. Rev. 865–880.
Allison, P.D., 1980. Inequality and scientific productivity. Soc. Stud. Sci. 10 (2), 163–179.
Andriani, P., McKelvey, B., 2009. From Gaussian to Paretian thinking: causes and implications of power laws in organizations. Organ. Sci. 20 (6).
Anscombe, F.J., 1973. Graphs in statistical analysis. Am. Statistician 27 (1), 17–21.
Blanca, M.J., Arnau, J., López-Montiel, D., Bono, R., Bendayan, R., 2013. Skewness and kurtosis in real data samples. Methodology 9, 78–84.
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., 2024. The strategic advantage of impulsivity in entrepreneurial action: an agent-based modeling approach. Entrep. Theory Pract. 48 (2), 547–580.
Bradley, K.J., Aguinis, H., 2023. Team performance: nature and antecedents of nonnormal distributions. Organ. Sci. 34 (3), 987–1352.
Breig, R., Coblenz, M., Pelz, M., 2018. Enhancing simulation-based theory development in entrepreneurship through statistical validation. J. Bus. Ventur. Insights 9, 53–59.
Broido, A.D., Clauset, A., 2019. Scale-free networks are rare. Nat. Commun. 10 (1), 1017.
Canback, S., Samouel, P., Price, D., 2006. Do diseconomies of scale impact firm size and performance? A theoretical and empirical overview. ICFAI Journal of Managerial Economics 4 (1), 27–70.
Certo, S.T., Raney, K., Albader, L., Busenbark, J.R., 2024. Out of shape: the implications of (extremely) nonnormal dependent variables. Organ. Res. Methods 27 (2), 195–222.
Cirillo, P., 2013. Are your data really Pareto distributed? Phys. Stat. Mech. Appl. 392 (23), 5947–5962.
Clark, D.R., Crawford, G.C., Pidduck, R.J., 2023. Exceptionality in entrepreneurship: systematically investigating outlier outcomes. J. Bus. Ventur. Insights 20, e00422.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.
Crawford, G.C., 2009. A review and recommendation of simulation methodologies for entrepreneurship research. Available at SSRN 1472113.
Crawford, G.C., 2013. Causes of Extreme Outcomes in Entrepreneurship. University of Louisville. Doctoral dissertation.
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 (5), 696–713.
Crawford, G.C., Joo, H., Aguinis, H., 2024a. Under the weight of heavy tails: a power law perspective on the emergence of outliers in entrepreneurship. J. Bus. Ventur. Insights 21, e00447.
Crawford, G.C., Linder, C., Lechner, C., Villani, E., 2024b. Outlier entrepreneurs: nonlinear paths and novel ventures. J. Bus. Ventur. Insights 21, e00437.
Crawford, G.C., McKelvey, B., Lichtenstein, B.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., Skorodziyevskiy, V., Frid, C.J., Nelson, T.E., Booyavi, Z., Hechavarria, D.M., et al., 2022. Advancing entrepreneurship theory through replication: a case study on contemporary methodological challenges, future best practices, and an entreaty for communality. Entrep. Theory Pract. 46 (3), 779–799.
Davis, J.P., Eisenhardt, K.M., Bingham, C.B., 2007. Developing theory through simulation methods. Acad. Manag. Rev. 32 (2), 480–499.
Dean, M.A., Shook, C.L., Payne, G.T., 2007. The past, present, and future of entrepreneurship research: data analytic trends and training. Entrep. Theory Pract. 31 (4), 601–618.
Delmar, F., 2019. Measuring growth: methodological considerations and empirical results. In: Entrepreneurship and SME Research. Routledge, pp. 199–215.
Doane, D.P., Seward, L.E., 2011. Measuring skewness: a forgotten statistic? J. Stat. Educ. 19 (2).
Fan, T., Schwab, A., Geng, X., 2021. Habitual entrepreneurship in digital platform ecosystems: A time-contingent model of learning from prior software project experiences. Journal of Business Venturing 36 (5), 106140.
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.
Gibbert, M., Nair, L.B., Weiss, M., Hoegl, M., 2021. Using outliers for theory building. Organizational Research Methods 24 (1), 172–181.
Grégoire, D.A., Ter Wal, A.L., Little, L.M., Bermiss, S., Kotha, R., Gruber, M., 2024. Mobilizing new sources of data: opportunities and recommendations. Acad. Manag. J. 67 (2), 289–298.
Gupta, R.D., Kundu, D., 2001. Exponentiated exponential family: an alternative to gamma and Weibull distributions. Biom. J.: Journal of Mathematical Methods in Biosciences 43 (1), 117–130.
Hunt, R.A., Lerner, D.A., Johnson, S.L., Badal, S., Freeman, M.A., 2022. Cracks in the wall: entrepreneurial action theory and the weakening presumption of intended rationality. J. Bus. Ventur. 37 (3), 106190.
Hymer, C.B., Smith, A.D., 2024. Making Exceptions Exceptional: A Cross-Methodological Review and Future Research Agenda. Journal of Management 01492063241237225.
Joo, H., Aguinis, H., Bradley, K.J., 2017. Not all nonnormal distributions are created equal: improved theoretical and measurement precision. J. Appl. Psychol. 102 (7), 1022.
Kearney, C., Lichtenstein, B., 2023. Generative emergence: exploring the dynamics of innovation and change in high-potential start-up ventures. Br. J. Manag. 34 (2), 898–914.
Khurana, I., Tamvada, J.P., Audretsch, D.B., 2023. The weaker sex? A tale of means and tails. J. Bus. Ventur. Insights 20, e00407.
Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. John Wiley & Sons.
Kokko, H., Mackenzie, A., Reynolds, J.D., Lindström, J., Sutherland, W.J., 1999. Measures of inequality are not equal. Am. Nat. 154 (3), 358–382.
Kuckertz, A., Scheu, M., Davidson, P., 2023. Chasing mythical creatures - a (not-so-sympathetic) critique of entrepreneurship's obsession with unicorn startups. J. Bus. Ventur. Insights 19, e00365.
Landström, H., Harirchi, G., Åström, F., 2012. Entrepreneurship: exploring the knowledge base. Res. Pol. 41 (7), 1154–1181.
Lichtenstein, B., 2018. Applying the 15 complexity sciences: methods for studying emergence in organizations. In: Handbook of Research Methods in Complexity Science. Edward Elgar Publishing, pp. 525–544.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. Bioscience 51 (5), 341–352.
Lu, J., Dimov, D., 2023. A system dynamics modelling of entrepreneurship and growth within firms. J. Bus. Ventur. 38 (3), 106285.
Maula, M., Stam, W., 2020. Enhancing rigor in quantitative entrepreneurship research. Entrep. Theory Pract. 44 (6), 1059–1090.
McKelvey, B., Andriani, P., 2005. Why Gaussian statistics are mostly wrong for strategic organization. Strat. Organ. 3 (2), 219–228.
Micceri, T., 1989. The unicorn, the normal curve, and other improbable creatures. Psychol. Bull. 105 (1), 156.
Mitzenmacher, M., 2004. A brief history of generative models for power law and lognormal distributions. Internet Math. 1 (2), 226–251.
Murphy, G.B., Trailer, J.W., Hill, R.C., 1996. Measuring performance in entrepreneurship research. J. Bus. Res. 36 (1), 15–23.
Nair, J., Wierman, A., Zwart, B., 2022. The fundamentals of heavy tails: Properties, emergence, and estimation, 53. Cambridge University Press.
Quigley, T.J., Hill, A.D., Blake, A., Petrenko, O., 2023. Improving our field through code and data sharing. J. Manag. 49 (3), 875–880.

Reed, W.J., Jorgensen, M., 2004. The double Pareto-lognormal distribution—a new parametric model for size distributions. Commun. Stat. Theor. Methods 33 (8), 1733–1753.
Richard, P.J., Devinney, T.M., Yip, G.S., Johnson, G., 2009. Measuring organizational performance: towards methodological best practice. J. Manag. 35 (3), 718–804.
Schwab, A., 2018. Investigating and communicating the uncertainty of effects: the power of graphs. Entrep. Theory Pract. 42 (6), 823–834.
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.
Stumpf, M.P., Porter, M.A., 2012. Critical truths about power laws. Science 335 (6069), 665–666.
Vargo, E., Pasupathy, R., Leemis, L., 2010. Moment-ratio diagrams for univariate distributions. J. Qual. Technol. 42 (3), 276–286.
Vogel, R.M., Fennessey, N.M., 1993. L moment diagrams should replace product moment diagrams. Water Resour. Res. 29 (6), 1745–1752.
Wennberg, K., Anderson, B.S., 2020. Enhancing the exploration and communication of quantitative entrepreneurship research. J. Bus. Ventur. 35 (3), 105938.
West, S.G., Finch, J.F., Curran, P.J., 1995. Structural equation models with nonnormal variables: problems and remedies. In: Hoyle, R.H. (Ed.), Structural Equation Modeling: Concepts, Issues, and Applications. Sage Publications, Inc, pp. 56–75.