Journal of Business Venturing 39 (2024) 106347

Contents lists available at ScienceDirect

Journal of Business Venturing

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

Star entrepreneurs on digital platforms: Heavy-tailed performance distributions and their generative mechanisms

Kaushik Gala *, Andreas Schwab, Brandon A. Mueller

Department of Management and Entrepreneurship, Ivy College of Business, Iowa State University, 3219 Gerdin, 2167 Union Dr, Ames, IA 50011, USA

ARTICLE INFO

JEL classifications:
L26
M13
D31

Keywords:
Digital platforms
Entrepreneurial performance
Star entrepreneurs
Lognormal performance distributions
Proportional differentiation

ABSTRACT

This study extends emerging theories of star performers to digital platforms, an increasingly prevalent entrepreneurial context. It hypothesizes that the unique characteristics of many digital platforms (e.g., low marginal costs, feedback loops, and network effects) produce heavy-tailed performance distributions, indicating the existence of *star entrepreneurs*. Using longitudinal data from an online learning platform, proportional differentiation is identified as the most likely generative mechanism and lognormal distribution as the most likely shape for distributions of entrepreneurial performance in digital contexts. This study contributes theory and empirical evidence for non-normal entrepreneurial performance with implications for scholars and practitioners of digital entrepreneurship.

Executive summary

The performance of 'star' entrepreneurs on digital platforms can be 100- or 1000-fold that of their average competitors. When performance is plotted as a distribution, star performers reside in the tails of these distributions. The assumption of a normal distribution of performance in the bulk of entrepreneurship research implies that most performance observations are clustered around the average. Instead, most entrepreneurs on digital platforms exhibit sub-par performance, while a minority captures a major fraction of the generated value. This paper argues that the unique characteristics of digital contexts - nearly zero marginal costs, feedback loops, and network effects - drive such extreme performance. Using data from Udemy, a digital platform where independent producers (entrepreneurs) offer educational videos (digital products) to a large pool of potential customers, we provide evidence that entrepreneurial performance is lognormally rather than normally distributed. We further identify proportional differentiation as the underlying generative mechanism. Thus, star performance on digital platforms is not driven only by the rich-get-richer effect. Instead, both the initial value of performance and the rate at which it is accumulated play important roles in explaining extreme performance outcomes. This discovery has important implications for entrepreneurship theory and practice. Our findings, for example, signal that some late entrants who successfully pursue high customer accumulation rates in domains with high knowledge intensity can become star entrepreneurs.

* Corresponding author.
E-mail address: kgala@iastate.edu (K. Gala).

https://doi.org/10.1016/j.jbusvent.2023.106347
Received 16 November 2022; Received in revised form 26 August 2023; Accepted 28 August 2023

1. Introduction

Management researchers are increasingly recognizing the importance of non-linear, non-normal, and asymmetric distributions for accurately capturing and understanding the performance of individuals, teams, and organizations (Andriani and McKelvey, 2007, 2009; Meyer et al., 2005; Beamish and Hasse, 2022). The performance of *star performers*, by definition, dramatically exceeds that of their 'average' peers. Empirical evidence for such non-normally distributed performance has been reported across occupations, professions, and industries (Aguinis and O'Boyle, 2014; Aguinis et al., 2018; Asgari et al., 2021; Bradley and Aguinis, 2023; Joo et al., 2017).

Entrepreneurship researchers, too, have begun to pay more attention to non-normal performance distributions (Crawford et al., 2015; Hunt et al., 2022; Wiltbank et al., 2015), with power-law distributions and complexity science serving as suitable starting points and theoretical perspectives (Crawford and Kreiser, 2015; McKelvey, 2004; Schindehutte and Morris, 2009; Selden and Fletcher, 2015). Most of this research, however, has focused on supporting the general relevance of non-normal performance distributions, leaving more focused explorations of specific entrepreneurial contexts – whose unique characteristics may substantively influence performance distributions – untouched.

Digital platforms, defined as "*a shared, common set of services and architecture that serves to host complementary offerings, including digital artifacts*" (Nambisan, 2017, p. 1032), represent an increasingly prevalent context for impactful entrepreneurial activities (Cusumano, 2022; Cutolo and Kenney, 2021; Kenney et al., 2021). The digital economy powered by such platforms accounts for over 10 % of the U.S. GDP (Nicholson, 2020) and engages an increasing fraction of the global workforce (Autor et al., 2020; World Bank, 2022). Many of the most valuable global companies (e.g., Alphabet, Amazon, Apple) operate digital platforms and marketplaces (Forbes, 2022) on which millions of individuals and firms engage in entrepreneurial activity (Nambisan et al., 2019). The creator economy alone, constituted by individuals targeting online audiences with branded digital content, is expected to grow to half a trillion dollars by 2027 (Goldman Sachs, 2023).

Scholars have theorized how the unique attributes of digital platforms – such as their broad access, low distribution costs, high levels of transparency, and inbuilt feedback loops – may influence entrepreneurs' and organizations' performance (Donaker et al., 2019; Kordrostami and Rahmani, 2020; Rifkin, 2014). Winner-take-all effects and positive network externalities enabled by digital technologies may amplify performance differentials among entrepreneurs operating on such platforms (Benzell and Brynjolfsson, 2019; Menz et al., 2021; Nambisan et al., 2019; von Briel et al., 2018). More broadly, scholars have suggested that the scalability of digital firms' resource bundles – arising from significant economies of scale in production and distribution – combined with strong network effects can result in a "*disproportionately higher boost to firm size from specialization*" (Giustiziero et al., 2023, p. 1410). In developing a theory of the digital firm, these scholars have argued that greater and more persistent scalability favors hyper-specialization and rapid, global growth with increasing returns to scale (Arthur, 2009).

To examine extreme performance in this increasingly prevalent context for entrepreneurship, this study seeks to develop and test theory centered on the mechanisms that generate heavy-tailed distributions of entrepreneurial performance on digital platforms. These generative mechanisms can help explain when and how extremely high-performing entrepreneurs emerge. Such theory development centered on non-normality and positive skew stands in contrast to common assumptions of normality of performance in entrepreneurship research (Dean et al., 2007; Delmar et al., 2022). To the extent that the distribution of actual performance differs materially from a normal distribution, the propositions and findings from extant research may have to be revisited to account for heavy-tailed entrepreneurial performance (Crawford et al., 2022).

We test related hypotheses on longitudinal data for Udemy.com, an online platform with over 49 million global learners and over 50,000 online instructors (Udemy, 2021). While the average annual income of an instructor is about $2800, dozens of instructors earn more than $1 million annually (Udemy, 2021). An example of a *star entrepreneur* (analogous to a star performer) on Udemy is the instructor Dr. Angela Yu with a cumulative base of over 1.9 million students since 2015 (The App Brewery, 2023; Udemy, 2023b). Similarly, nearly 950,000 students have enrolled across courses offered by Chris Dutton, the founder of Maven Analytics (Udemy, 2023c). Such instructors are, in essence, entrepreneurs as they discover and exploit opportunities to meet the market demand for educational content (Shepherd et al., 2019). In the platform context, scholars have increasingly recognized the dispersion, redefinition, and reorganization of work wherein 'producers' such as Udemy instructors act as mini- or proto-entrepreneurs, albeit with highly unequal outcomes (Calvo et al., 2022; Johnson et al., 2022; Kenney and Zysman, 2016). Indeed, digital entrepreneurs such as Udemy instructors are expected to dynamically "*opt in and out based on their own individual goals, motivations, capabilities, constraints, and contributions*" (Nambisan, 2017, p. 1035).

This study seeks to make multiple contributions to the entrepreneurship literature. First, it theorizes that the inherent nature of digital platforms can drive heavy-tailed performance distributions for entrepreneurs who operate on such platforms (Nambisan et al., 2019). In doing so, it not only explicates how digital contexts for entrepreneurship differ substantively from physical contexts and why these differences may amplify performance differentials, but also seeks to enrich emerging research on the theory of the digital firm by drawing attention to 'hyperscaling' outcomes for entrepreneurs operating on digital platforms (Giustiziero et al., 2023).

Second, it provides arguments and empirical evidence supporting the proposition that such distributions are likely lognormal in shape for entrepreneurs on digital platforms, extending prior research that has identified exponential tail or power law with exponential cutoff distributions as the most likely characterizations in broader contexts of individual or team performance (Bradley and Aguinis, 2023; Joo et al., 2017). Here, this study contributes to the literature on digital entrepreneurship by predicting a specific distribution shape that more accurately characterizes non-normal performance on digital platforms. Moreover, it contributes to emerging empirical research on pervasive non-normality in entrepreneurship variables by illustrating how a falsification approach can be used to theorize and identify the most suitable distributional shapes and appropriate analytical methods can be adopted for hypothesis testing where distributions and their parameters are the focal dependent variables.

Third, this study introduces and provides empirical support for a specific underlying generative mechanism of *proportional differentiation* to explain the observed lognormal distribution of performance for entrepreneurs involved in creating and distributing digital goods via digital platforms. This mechanism involves the combined effect of initial performance and the rate at which performance accumulates over time (Andriani and McKelvey, 2009; Banerjee and Yakovenko, 2010). The current study is likely the first in entrepreneurship research that conducts empirical tests of generative mechanisms. It shows the need for explicitly accounting for star entrepreneurs so that more variance can be explained (Makino and Chan, 2017), and the underlying success factors that generate such heavy-tailed performance distributions can be identified (Gibbert et al., 2021).

Finally, the evidence found in this study for lognormal performance distributions and their generative mechanism extends emerging theories, which have primarily investigated star performers in more traditional salaried or professional settings (Aguinis and O'Boyle Jr., 2014; Aguinis et al., 2018; Asgari et al., 2021; Bradley and Aguinis, 2023), to those in digital and entrepreneurial contexts. The reported insights into extreme entrepreneurial performance on digital platforms – drawing upon concepts and theories from complexity science, information economics, and work design – have important implications for entrepreneurship theory and practice; these are elaborated upon in the Discussion section of this paper.

2. Theory and hypotheses

2.1. Performance distributions

Emerging evidence suggests that individual performance frequently adheres to non-normal distributions (Asgari and Hunt, 2016; Asgari et al., 2021; O'Boyle Jr and Aguinis, 2012). When captured as firm size or growth, organizational performance also has been found to follow non-normal distributions and resemble Pareto or Zipf distributions with noticeably heavy tails (e.g., Axtell, 2001; Simon and Bonini, 1958). These observations have made scholars of individual, team, and firm performance increasingly concerned about violations of normality, a ubiquitous assumption in management and organizational research (Aguinis et al., 2021; Andriani and McKelvey, 2011; Meyer et al., 2005). Past empirical research on non-normal distributions of individual performance has noted their resemblance to power laws; scholars have also started to identify factors that may influence such heavy-tailed distributions (Aguinis et al., 2016). Recent investigations have shifted from focusing only on power law distributions toward exploring a broader variety of non-normal distributions, including exponential tail and lognormal distributions (Joo et al., 2017). Because each class of distributions is associated with a different generative mechanism, identifying these mechanisms is essential to predict the presence and prevalence of star performers and the shape of performance distributions (Andriani and McKelvey, 2009).

Entrepreneurship research, too, has started to investigate performance distributions and found strong evidence of heavy tails, with far more star entrepreneurs than predicted by assumptions of normality (Crawford et al., 2015). Following their counterparts in management science (Boisot and McKelvey, 2010), entrepreneurship scholars have conceptualized these distributions as power laws, calling attention to the widespread prevalence of tail extremity not only for output variables, such as employee count and revenue, but also input variables, such as industry experience and investment (Crawford et al., 2014; Crawford et al., 2022). These findings have led some researchers to call for a reconceptualization of entrepreneurship research and a move beyond long-held assumptions of normality (Crawford, 2018). Instead, they recommend the explicit analysis of valid and influential outliers (Aguinis et al., 2013) to engage in comprehensive examinations of heavy-tailed performance distributions and their underlying mechanisms.

2.1.1. Entrepreneurs on digital platforms

Research into performance variance has drawn sharp distinctions between 'average' ventures and high-growth startups, as well as between 'average' individuals and highly skilled professionals (Huberman and Adamic, 1999; Mankiw, 2013; Shane, 2009). Moreover, scholars have theorized that extreme performance will likely accelerate across industries and sectors as digital technology rapidly penetrates the global economy (Birkinshaw, 2018; Menz et al., 2021). Today, the digital economy represents nearly 10 % of the U.S. GDP (Nicholson, 2020), digital platforms engage an increasing fraction of the global workforce (World Bank, 2022), and high-performing digital ventures capture increasing economic returns (Arthur, 1994).

What characteristics make digital platforms and the entrepreneurs who operate on them a particularly relevant context for scholarly inquiry into extreme performance? First, digital platforms have disrupted traditional industries and created new ones (Menz et al., 2021). They have become a highly relevant entrepreneurial context in record time (Evans and Gawer, 2016; Cusumano, 2022). Some digital platforms have evolved into cross-industry ecosystems whose constituents range from millions of software developers to billions of users (Cozzolino et al., 2021). Second, digital platforms seem to augment the likelihood of winner-take-all effects for digital entrepreneurs (Cusumano et al., 2019) while simultaneously subjecting digital entrepreneurs to the mercy of a platform's policies and priorities (Nambisan and Baron, 2021; Rietveld and Schilling, 2021). For instance, the approximately 500,000 active sellers on Amazon's U.S. e-commerce platform have average annual revenues of approximately $200,000, yet several hundred sellers have annual revenues of over $10 million (Hahnbeck, 2021). At Poshmark.com, a digital platform for social commerce focused on fashion, the average annual revenue across nearly 4.5 million sellers was approximately $310, yet some entrepreneurs have earned over $100,000 in a year (Business Insider, 2018; Business Insider, 2022; Poshmark, 2020). Similarly, the approximately 5.4 million sellers on the Etsy.com online marketplace have average annual revenues of approximately $2200, yet dozens of them have annual revenues of over $100,000 (Etsy, 2023; MarketPlacePulse, 2023).

Therefore, building upon extant research that has found non-normal performance distribution for (a) high-performing individuals across a variety of professional contexts (Aguinis and O'Boyle Jr., 2014) and (b) entrepreneurial firms (Crawford et al., 2015), we propose that the distribution of performance on digital platforms is likely to be heavy-tailed and contain a greater proportion of – and more extreme – star entrepreneurs. In the following sections, we develop novel theory and related hypotheses to: (1) predict specific distributions most likely to occur in the context of digital platform entrepreneurship, (2) explicate generative mechanisms that may give rise to these non-normal distributions, and (3) examine factors that may amplify or diminish the extremity of non-normal performance.

2.2. Non-normal performance distributions for entrepreneurs on digital platforms

Emerging research into the digital transformation of organizations (Vial, 2019) has argued that the advent of digital artifacts, infrastructure, and platforms (Nambisan, 2017) is likely to fundamentally alter the entrepreneurial process (Fossen and Sorgner, 2021). Given that the traditional limits of physicality do not apply in the digital world (Nambisan, 2017), the likelihood and tail extremity of non-normal distributions in entrepreneurial performance may be greater in the digital context.

Once created, digital offerings such as software products have low or nearly zero marginal costs of production and distribution (Adner et al., 2019; Bakos and Brynjolfsson, 1999; Rifkin, 2014). Their commercialization is limited primarily by market size, which often surpasses that of traditional markets due to substantially lower geographic constraints. In general, digital technologies have been theorized as external enablers of venture creation (von Briel et al., 2018) because they give entrepreneurs a broader reach not only to end customers but also to labor markets, suppliers, and partners (Stallkamp et al., 2022). Therefore, we extend prior research on non-normal performance (Aguinis et al., 2016; Joo et al., 2017) to performance distributions on digital platforms by hypothesizing the presence and, importantly, a high propensity of star entrepreneurs in this context. The non-normality and extremity observed in input variables (e.g., number of owners, previous ventures founded) and output variables (e.g., number of employees, revenue) in physical contexts for entrepreneurship (Crawford et al., 2015) are likely to not only replicate but, arguably, exacerbate in digital contexts for entrepreneurship (Menz et al., 2021). In sum, when viewed from a distributional perspective, the performance for a pool of entrepreneurs on a digital platform is likely to be asymmetric and heavy-tailed, in stark contrast to a symmetric, normal shape.

Hypothesis 1. Performance distributions for entrepreneurs on digital platforms are more likely to resemble non-normal distributions than normal distributions.

2.3. Types of non-normal performance distributions on digital platforms

If entrepreneurial performance on digital platforms is not normally distributed (H1), we face an equally important but more challenging question: Which specific non-normal distribution shape best characterizes entrepreneurial performance in the digital context? We explore this question by considering seven distribution shapes commonly considered in recent studies of non-normal performance distributions (see Fig. 1; Bradley and Aguinis, 2023; Joo et al., 2017). For exponential tail (λ), Poisson (μ), and pure power law (α) distributions, one parameter is sufficient to characterize the distribution. However, two parameters are needed to characterize the lognormal (μ, σ), normal (μ, σ), power law with exponential cutoff (α, λ), and Weibull (β, λ) distributions. All distributions except the pure power law have finite means and variances. Poisson and Weibull distributions are relatively similar in their low degree of asymmetry and skew. They also share the same underlying generative mechanism (i.e., homogenization). Therefore, we lumped them with the normal distribution into a broader category of symmetric or potentially-symmetric distributions, following Joo et al. (2017).

2.3.1. Lognormal performance distributions and digital platform entrepreneurs

We adopt the preceding taxonomy of performance distributions to argue that lognormal distributions most suitably capture entrepreneurial performance on digital platforms. Lognormal distributions have 'heavy' but finite right tails; in other words, their tails are more 'fat' than 'long'. Compared with normal or exponential tail distributions, this indicates a greater prevalence of extreme performance. Compared with pure power laws, this indicates range limits to extreme performance. Notably, for firm size as a measure of performance, prior research has reported some evidence for lognormal distributions (Chesher, 1979), particularly for small and midsize firms (Aitchison and Brown, 1957; Bee et al., 2017), including entrepreneurial firms (Hernandez, 2019). Similarly, recent research into the distributions of team performance in physical contexts (e.g., engineering, politics, sports) has reported some evidence for lognormal distributions (Bradley and Aguinis, 2023).

For digital platforms, we argue that winner-take-all effects amplified by low marginal costs are likely to drive performance to greater extremes than in physical production contexts. However, limits imposed by market size and access – combined with rising costs of user attrition and customer support – may eventually limit the extent of extreme success. Therefore, we posit that the performance of entrepreneurs operating on digital platforms is better characterized by a lognormal distributional shape than the other shapes illustrated in Fig. 1. In the following section, we introduce theory-based arguments for why the lognormal distribution is likely to better characterize entrepreneurial performance on digital platforms than (a) pure power law distributions, (b) exponential tail and power law with exponential cutoff distributions, or (c) symmetric distributions.

2.3.2. Lognormal versus pure power law distributions

The performance of entrepreneurs operating on digital platforms may not follow power law distributions, which have been explored in the context of organizational performance of new ventures (Crawford et al., 2015) and large, multiproduct firms (Bee et al., 2017). Power law distributions are theorized to arise from a variety of underlying processes (Barabási and Albert, 1999) – such as, for

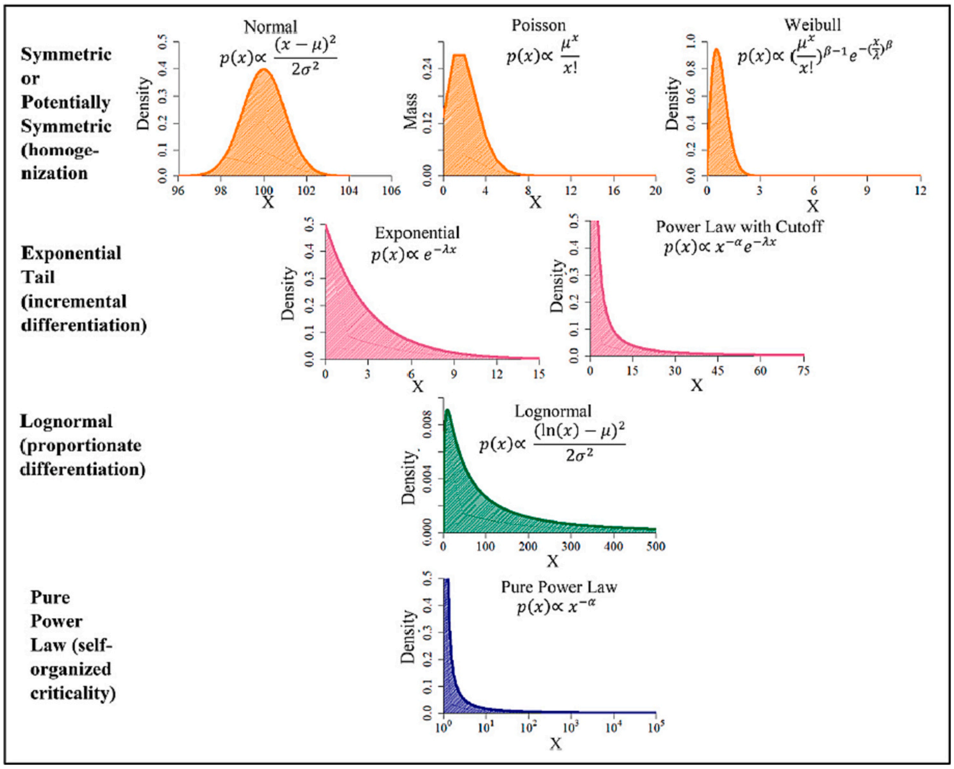

Fig. 1. Depiction of generic distributions.
Note: This figure is adapted from Bradley and Aguinis (2023), p. 1268, Figure 1, which is based on Joo et al. (2017), p. 1024, Figure 1. The graphs are not equally scaled.

example, preferential attachment or self-organized criticality (Andriani and McKelvey, 2009; McKelvey et al., 2012). Preferential attachment involves positive feedback loops, where the larger nodes in a system attract more new incoming agents (e.g., popular Amazon sellers attracting more first-time customers), leading to what is frequently referred to as 'rich-get-richer' growth patterns associated with power law distributions of performance.

Self-organized criticality involves a highly dynamic and self-reinforcing system with built-in sensitivities such that unpredictable and seemingly small exogenous events can disrupt its stability due to the complex interdependencies between elements in the system. Such 'shocks' to a system in a critical state tend to trigger broader non-deterministic changes with extremely high variability in outcomes (Andriani and McKelvey, 2009). For example, complex decentralized production systems can be highly sensitive to disruptions of logistical processes. The impact of these disruptions can prove difficult to contain and suddenly unravel an entire production system, as automotive firms discovered during the recent pandemic (Ramani et al., 2022). This pandemic event also induced spontaneous adaptations by entrepreneurs, employees and consumers, with home confinement and the shift to remote work resulting in unpredictable, power law outcomes – exemplified by Zoom, which by April 2020 attracted 300 million meeting participants, 30 times the amount just four months earlier (NPR, 2021).

Theoretically, 'pure' power law outcome distributions have infinite mean and variance (Newman, 2005), neither of which empirically holds for digital entrepreneurship, where performance is heavy-tailed but within a measurable range. Not only are the most extreme performance outliers on such platforms limited by the network size, specifically, and the market size, broadly, but they are also likely to face competition for market share from peers and for revenue share from the platform owner-operator (Srinivasan and Venkatraman, 2018; Tavalaei and Cennamo, 2021). Similarly, there may be demand-side limits to extreme popularity, wherein new customers do not necessarily follow the purchasing decisions of past customers by focusing exclusively on the top-rated entrepreneurs. While positive feedback loops involving customer and user reviews play an important role in enabling star entrepreneurs through preferential attachment, potential customers may not blindly choose the most popular entrepreneur. Instead, they often rely on multiple factors as they evaluate various entrepreneurs and their offerings, thus mixing talent with popularity (Franck and Nüesch, 2012).

Digital platforms are unlikely to be in a state of self-organized criticality vulnerable to unpredictable shocks because (a) the decentralized and distributed nature of entrepreneurial activity in such ecosystems may offer greater flexibility and resilience amidst exogenous shocks (Floetgen et al., 2021), (b) digital entrepreneurs and platforms may adapt faster amidst dynamic environments (Fan et al., 2021), and (c) the owner-operators of digital platforms spanning geographic boundaries may adopt loose-coupling strategies to improve responsiveness amidst external shocks (Nambisan and Luo, 2021). In sum, underlying causes such as self-organized criticality may not be the primary driver and pure power laws not the most suitable characterizations of non-normal performance distributions for entrepreneurs on digital platforms.

2.3.3. *Lognormal versus exponential tail and power law with exponential cutoff distributions*

In their endeavor to fit specific shapes to different output distributions, scholars have suggested that (a) exponential tail and (b) power law with exponential cutoff distributions may best fit many professional employment contexts (Joo et al., 2017). Both these distributions have been linked to the generative mechanism of *incremental differentiation*, which are primarily driven by the rate at which performance accumulates over time (Andriani and McKelvey, 2009). For individual performance in the physical context, rapidly diminishing returns and increasing costs are likely to enforce productivity limits on high-performing individuals. Therefore, incremental differentiation is arguably more relevant to physical contexts than proportional differentiation, wherein both initial performance levels and the rate at which performance accumulates over time make a difference However, entrepreneurs who operate on digital platforms are likely to face much lower marginal costs of production, scaling, and distribution given the digital nature of their products (Bakos and Brynjolfsson, 1999; Goldfarb and Tucker, 2019; Rifkin, 2014). Zero or nearly zero marginal costs facilitate substantial increases in production volume. In contrast, individual entrepreneurs producing physical products or services face relatively higher costs, limiting their ability to increase output.

The limitations on the output generated in the physical world are substantively different from those in the digital world. For example, a high-performing doctor may face diminishing returns because of time and energy limitations in treating patients, but an entrepreneur who commercializes data, algorithms, or software on platforms is less likely to face such limitations (Nambisan, 2017). Moreover, the assumption that output loops involving positive feedback may be short-lived because of diminishing returns (Van de Rijt et al., 2014) may not hold for digital platforms where customer reviews – when prominently displayed, publicly available, and persistently stored – play a central role in how some entrepreneurs benefit disproportionately (Bolton et al., 2013; Kordrostami and Rahmani, 2020). Here, entrepreneurial performance may instead benefit from a greater number of customers whose reviews accelerate the acquisition of new customers, who often rely on heuristics when faced with voluminous information (Aljukhadar et al., 2012). Moreover, online customer reviews – by virtue of being easily accessible and searchable – are likely to have greater reach and, therefore, greater influence than physical word-of-mouth, which may have limited reach. Similarly, initial performance in the digital context may not arise entirely from dumb luck (Barabási, 2012). Instead, it may arise from first-mover advantages (Varadarajan et al., 2008) and hence be salient to the generative mechanism for heavy-tailed entrepreneurial performance on digital platforms. In sum, incremental differentiation may not be the primary generative mechanism – and exponential tail or power law with exponential cutoff distributions may not offer the most suitable characterizations – of non-normal performance distributions for entrepreneurs on digital platforms.

2.3.4. *Lognormal versus symmetric distributions*

Symmetric distributions include the normal, Poisson, and Weibull distributions (Joo et al., 2017). Extending the arguments for Hypothesis 1, wherein we posit non-normality, we argue that the Poisson and Weibull distributions – with their slight skew and lack of extreme data points – are less likely to characterize entrepreneurial performance on digital platforms. Homogenization, the mechanism theorized to underlie the symmetric class of distributions, involves floor and ceiling constraints on output. Such constraints are often relevant to physical contexts such as laborious jobs or managerial work (Joo et al., 2017) but may not apply to digital contexts, particularly for digital goods distributed via digital platforms. Based on these preceding pairwise arguments and conjectures about the mechanisms that generate alternative performance distributions, we suggest:

Hypothesis 2. Performance distributions for entrepreneurs on digital platforms are more likely to resemble lognormal distributions than (a) pure power law, (b) exponential tail, (c) power law with exponential cutoff, or (d) symmetric distributions.

2.4. *Influence of domain characteristics on entrepreneurial performance on digital platforms*

In addition to the generative mechanisms of star entrepreneurs, specific characteristics of the domains in which entrepreneurs operate may influence the performance distributions of entrepreneurs on digital platforms. In this study, an entrepreneurial *domain* represents the category where entrepreneurs offer their products and services on a digital platform, such as ‘Books’ or ‘Electronics’ on Amazon.com. Thus, *domains* in the entrepreneurship context are analogous to *occupations* in the job context, which scholars have used to study the outcome distributions of star performers (Aguinis et al., 2016).

2.4.1. *Knowledge intensity of an entrepreneurial domain*

The knowledge intensity of an entrepreneurial domain is characterized by “*the kinds of knowledge, skill, and ability demands that are placed on*” entrepreneurs and their customers on digital platforms as reflected in complexity, information processing, problem-solving, skill variety, and specialization (Morgeson and Humphrey, 2006, p. 1323). Drawing upon previous work-design literature (Hackman and Oldham, 1976; Parker et al., 2017), we suggest that these five characteristics may influence the distribution of entrepreneurial performance in a domain.

Complexity implies tasks that are difficult to perform and require high levels of mental effort (Campbell, 1988; Hunter et al., 1990). Task complexity has been linked to creativity and relatively higher levels of performance differences between individuals (Coelho and Augusto, 2010). Information processing, wherein large amounts of data create higher cognitive demands, can also augment the likelihood and extent of extreme performance in a domain by conferring advantages to entrepreneurs better at handling voluminous information (Lord and Maher, 1990). Similarly, problem-solving has been theorized to be positively associated with entrepreneurial opportunity discovery (Shane, 2003), reflecting the advantages that may accrue to entrepreneurs capable of *"generating unique or innovative ideas or solutions, diagnosing and solving nonroutine problems, and preventing or recovering from errors"* (Morgeson and Humphrey, 2006, p. 1323). In our study, skill variety (wherein multiple skills are required) and specialization (wherein specialized skills are required) respectively represent the breadth and depth aspects of entrepreneurial 'work' (Morgeson and Humphrey, 2006). High-performing entrepreneurs in a domain that demands many unique skills may better retain and compound their advantage over peers compared with those in a domain where fewer or generic skills may suffice.

2.4.2. Knowledge intensity in the digital context

Knowledge characteristics are particularly relevant to digital platforms, where information and communication play a focal role in the creation, delivery, and consumption of goods and services (Nambisan, 2017). The digital nature of interactions between entrepreneurs and their stakeholders on these platforms provides the former with vast amounts of data they can use to better understand customers, competitors, or potential partners. Entrepreneurs with the requisite cognitive and intellective resources may better exploit this high volume of digitized information (Frenkel et al., 1995). Moreover, marketplaces such as Amazon.com, Airbnb.com, or Etsy.com lack synchronicity; buyers can explore and purchase products or services at a time and place different than when they were listed or offered by the sellers (Derave et al., 2021). Therefore, the greater the knowledge intensity of the entrepreneurial domain, the more information processing is required on the part of entrepreneurs to enable quick product searches, seamless purchasing and delivery, and prompt feedback from and timely support for customers.

Consequently, high-performing entrepreneurs in a digital domain with higher knowledge intensity may compound their advantage faster than entrepreneurs in less knowledge-intensive domains. For instance, entrepreneurs operating in the artificial intelligence or gaming industries are more likely to outperform their peers, for whom imitation may be onerous if not impossible. When compared with entrepreneurs operating in domains such as home improvement consulting or real estate advisory, we expect higher knowledge intensity to lead to greater tail extremity in performance. In sum, the knowledge intensity of a domain may act as a conductor of cumulative advantage (Aguinis et al., 2016; DiPrete and Eirich, 2006), thus increasing the probability of star digital entrepreneurs with sustained extreme levels of performance.

Hypothesis 3. Knowledge intensity of a domain is positively associated with the tail extremity of the performance distribution for entrepreneurs on digital platforms.

2.5. Generative mechanism for lognormal distribution of entrepreneurial performance

To our knowledge, prior entrepreneurship research has not empirically investigated the relevance and impact of different generative mechanisms on non-normal performance distributions. Given the established links between specific non-normal distributions (e.g., lognormal) and their generative mechanisms (e.g., proportional differentiation; Andriani and McKelvey, 2009), this is a critical gap in the literature we seek to address. Extending the earlier arguments that predict lognormal distributions for entrepreneurial performance on digital platforms, we hypothesize that proportional differentiation – a multiplicative effect between 'initial value' and 'accumulation rate' (Gumbel, 1958; Mitzenmacher, 2004) – will act as the primary generative mechanism of star entrepreneurs on digital platforms. For entrepreneurial activity, the *initial value* signifies the initial number of customers, funds, or resources available to or acquired by an entrepreneur, while the *accumulation rate* indicates the corresponding increase in customers, funds, or resources by the entrepreneur. Scholars have applied this notion of proportional differentiation as a generative mechanism in theorizing lognormal distributions for nascent firms (Hernandez, 2019; Shim, 2016). We extend this notion to entrepreneurs on digital platforms.

2.5.1. Initial value: the 'rich get richer'

On digital platforms, higher initial performance may be driven by early entry (Varadarajan et al., 2008), better IT-enabled resources (Nevo and Wade, 2010), network externalities (Shapiro et al., 1998), or sheer luck (Barabási, 2012; Denrell and Liu, 2012). Early star entrepreneurs may be able to amortize their costs across a larger number of customers, introduce product updates more frequently, or benefit from greater cost efficiencies. These competitive advantages may enable them to enter a self-reinforcing cycle of accelerating growth. In addition, the nearly zero marginal cost of digital goods may facilitate rapid scaling (Adner et al., 2019; Bakos and Brynjolfsson, 1999; Rifkin, 2014). The tendency of digital platform owner-operators to promote high-performing entrepreneurs (Hagiu, 2006; Oh et al., 2015) may further compound these initial advantages of star entrepreneurs, thus resulting in non-normal performance distributions.

The centrality of public feedback loops on digital platforms – for example, customer reviews (Cusumano et al., 2019; Donaker et al., 2019) – implies that the *accumulation rate* for entrepreneurs is affected by reviews, ratings, or recommendations from their customers (Belleflamme and Peitz, 2018). Positive reviews from past customers are likely to draw new customers with relatively lower effort from such entrepreneurs (Le Mens et al., 2018; Yadav and Pavlou, 2014). While the cost of supporting a rapidly growing customer base and negative reviews can dampen growth momentum, high-performing entrepreneurs may still benefit from this Matthew effect (Rigney, 2010; Wan, 2015), wherein initial success breeds more success (Van de Rijt et al., 2014).

Hypothesis 4a. Higher initial performance has a positive effect on the future performance of entrepreneurs on digital platforms.

2.5.2. Accumulation rate: overcoming initial disadvantages

Proportional differentiation allows entrepreneurs to overcome a relative deficit in *initial value* by outperforming their competitors over a sustained period, which allows the multiplicative effect to transmute into a cumulative advantage, analogous to compound interest (Aguinis et al., 2016). Thus, fast followers or late entrants might overcome their initial disadvantage vis-à-vis early winners by accumulating new customers and users relatively faster. Relatively lower marginal costs of production and distribution, relatively higher productivity ceilings, and relatively weaker growth-diminishing effects on digital platforms (Nambisan, 2017) may enable entrepreneurs to outperform their peers over time, despite a relatively low level of initial performance. Often, platform policies attempt to nudge users toward emerging star performers (Gokkaya and Wai, 2017), which at times has backfired when these policies unintentionally disrupt – rather than augment – the monopolistic advantages of incumbents (Borghans and Groot, 1998). In other words, greater effort, better strategy, positive externalities, or sheer luck have, in theory, the potential to enable some entrepreneurs to overcome their initial performance disadvantage.

Hypothesis 4b. Relative rate of performance accumulation has a positive effect on the future performance of entrepreneurs on digital platforms.

3. Methods

3.1. Data collection

Data was collected from Udemy.com, a digital platform that brings together producers and consumers of online educational courses created by entrepreneurial ventures, often involving one or several individuals. Angela Yu (Founder, The App Brewery), Chris Dutton (Founder, Maven Analytics), and Jose Portilla (Founder, Pierian Training) exemplify Udemy instructors with extreme performance; each is an education technology ('EdTech') entrepreneur with a million or more students on Udemy alone (Udemy, 2023b, 2023c, 2023d). High-performing Udemy instructors may also include employees of EdTech ventures such as John Pauler (Udemy, 2023e). Notably, of the 109 Udemy instructors in our sample with more than 500,000 students each, none is an established company or an employee of a large company; this has also been observed in various rankings of top instructors (e.g., Online Courses Galore, 2022).

Regardless of their employment status, we treat each instructor who operates on Udemy as an entrepreneur, following the well-established, opportunity-centric definition of entrepreneurs as those who *discover*, *evaluate*, *and exploit* opportunities (Shane and Venkataraman, 2000). Udemy instructors *discover* topics that learners are interested in, *evaluate* topics suitable for online instruction, and *exploit* learners' interest by offering paid versions of recorded courses through the Udemy.com digital platform. Arguably, this empirical context exemplifies the five key elements of entrepreneurial agency, i.e., "*ability, motivation, opportunity, institutions, and process skill*" (McMullen et al., 2021a, p. 1199). Indeed, a forward-looking view by leading entrepreneurship scholars has embraced digitalization and the gig economy, wherein seeing and exploiting opportunities – the essence of entrepreneurship – will be an *everyday-everyone* phenomenon (van Gelderen et al., 2021).

Moreover, we draw upon the emerging literature on digital entrepreneurship that has highlighted the changing nature of work in the digital economy and the replacement of traditional notions of 'job', 'employment', and 'labor' by 'income generation' (Calvo et al., 2022). A decrease in pre-defined loci of entrepreneurial agency has been emphasized as a distinct element of digital entrepreneurship, unlike the traditional, static notion of self-employed individuals or full-time founders (Nambisan, 2017). In evaluating the nature of work on digital platforms, Bearson et al. (2021) suggested a typology based on value creation, which categorizes the empirical context of our study as *platform-dependent consignment content creation* because the products sold on Udemy are purely digital or virtual in nature. Other scholars have coined the term *contentpreneurs* to describe entrepreneurs such as Udemy instructors who "*utilize online content as a key pillar of their business strategy in order to directly monetize the content for resource acquisition*" (Johnson et al., 2022, p.1). Such entrepreneurs abound on digital platforms such as Instagram (Mardon et al., 2018) and YouTube (Ashman et al., 2018; Guinez-Cabrera and Aqueveque, 2022), which are primarily business-to-consumer and involve purely digital products. Regardless of the terminology used, we suggest that Udemy instructors typify digital entrepreneurs who often pursue opportunities with fluid boundaries and enact them through diffuse and distributed agency (Nambisan, 2017).

Finally, we contrast Udemy instructors with the sellers on Coursera, another digital platform focused on online learning (Coursera, 2023). Coursera sellers are typically universities and multi-national companies that offer paid – often certified – course content and degrees. Arguably, Coursera exemplifies corporate entrepreneurship wherein large, established companies such as Google and leading universities such as Stanford pursue initiatives aimed at "*creating and adding new business, or at fostering innovation, change and renewal*" (Urbano et al., 2022, p. 1545).

3.1.1. Courses, instructors, and domains on Udemy

Udemy.com (Udemy, 2021; Udemy, 2023a) has over 50,000 instructors and 100,000 courses in 13 main categories (e.g., Lifestyle, Marketing, and Office Productivity). Each category has multiple, narrower domains (e.g., Branding, Digital Marketing, and Social Media). Each of these 130 domains lists numerous courses, which platform users can search using filters for language, topic, or learner level. Our study focuses on paid courses offered in English, which account for about 63 % of all courses (Udemy, 2021). We randomly sampled four domains for each of the 13 categories. For each of these 52 sampled domains (Appendix 1), we collected data for all courses and corresponding instructors. After excluding courses that were discontinued (2.4 %), lacked data for the number of students (4.1 %), or lacked data for the number of reviews (8.2 %), the final stratified random sample contained 27,779 courses. For these courses, we collected data for corresponding instructors; excluding duplicate (1.7 %) and private (5.2 %) profiles resulted in data for

12,391 instructors.

The tests for Hypothesis 4a and 4b required longitudinal data. Therefore, we collected course information for the same 12,391 instructors in the 52 domains in a second round, *16 weeks* after the first round. This duration was chosen because the median course length is 3 h, and 95 % of courses are less than 15 h long. Based on an estimated completion rate of 1 h per week (ClassCentral, 2021), students enrolled in 95 % of all courses are likely to have completed the cycle from course sign-up to course completion and posting of customer reviews within the 16-week window. Notably, students need not wait for course completion to provide the course a rating and review; they can do so at any time after starting the course (Udemy, 2023f). The Spearman correlation[1] between course length (in hours) and the cumulative number of students for that course is 0.21 [$CI_{95\,\%}$: 0.20, 0.22]. Therefore, we expect the results to hold even if a small fraction of the sampled courses has a duration significantly longer than 15 h and, therefore, a cycle time from signup-to-review that exceeds 16 weeks. The resulting panel dataset has an attrition rate of 3.7 % and contains data for 11,937 instructors.

3.2. Dependent variables

3.2.1. Entrepreneurial performance

For each instructor within each domain, performance was measured as the cumulative number of students enrolled across all the paid courses offered by a Udemy instructor. Performance was also measured, alternatively, as the cumulative number of reviews received across all the paid courses offered by a Udemy instructor.

3.2.2. Tail extremity of performance distribution

Following extant research, which uses distribution parameters (Bradley and Aguinis, 2023; Joo et al., 2017) as measures of tail extremity, we used the lognormal scale (μ) and shape (σ) parameters simultaneously to test Hypothesis 3 (Limpert et al., 2001; Montoya et al., 2017). Fig. 2 illustrates our exploration of this alternative test of the heaviness of the tails. Solid lines represent the same μ but a higher σ than dashed lines of the same color. The higher the σ for a given μ, the faster the decay of the right tail and the diminished likelihood of extreme performers. Conversely, the higher the μ for a given σ, the taller the peak and the greater the proportion of high-but-not-extreme performers. Following the arguments for relatively fatter than longer tails, we measure the tail extremity of the lognormal performance distribution using μ instead of σ.

This choice of μ over σ as the focal dependent variable aligns with the theorizing for Hypothesis 3, which centered on the role of digital technologies in skewing performance away from a normal distribution, driven by low or zero marginal cost of production, network effects, and the positive feedback loops generated by reviews. However, the performance for a given entrepreneurial domain is not unlimited in the extremity. Not only is the number of customers interested in a given product finite, but alternatives may also be available on the same or other platforms. Though performance is likely to skew right and be characterized by tail extremity, i.e., star entrepreneurs, the distribution of performance is unlikely to manifest as an extreme lognormal – or a pure power law – with near-infinite performance outliers. In sum, because this study posits heavily right-skewed, but not unlimited, performance and focuses on *fat* – not *long* – right tails, μ was chosen as the DV for testing Hypothesis 3, with σ accounted for as a control variable.

3.3. Independent variables

3.3.1. Knowledge intensity

The knowledge intensity of each domain was measured using five knowledge characteristics: (a) complexity, (b) information processing, (c) problem-solving, (d) skill variety, and (e) specialization. These knowledge characteristics are part of the work-design questionnaire self-report measure theorized, developed, and validated by Morgeson and Humphrey (2006) and subsequently used in empirical research (e.g., Peiró et al., 2020). In our study, we used the course information publicly available on Udemy.com to construct the following proxy measures to capture each of the five knowledge characteristics (Morgeson and Humphrey, 2006). First, 'complexity' was computed as the average number of lectures across courses in a domain based on the assumption that instructors will use a greater number of lectures for more complex courses for ease of consumption by students. Second, 'information processing' was measured as the average video length across courses in a domain based on the assumption that longer courses require students to consume more information.

Third, the 'problem-solving' component was measured using the average number of downloadable resources across courses in a domain, based on the observation that most courses that involve a higher degree of problem-solving, such as accounting, product design, or programming, require students to analyze and refine spreadsheets, design templates, or develop software code (Udemy, 2023a). Fourth, 'skill variety' was measured as the average number of instructors per course across courses in a domain. A larger number of instructors enables a greater variety of instructor skills. Finally, a domain's degree of 'specialization' was measured using the average number of informational articles across courses in a domain, based on the observation that courses with a relatively larger number of articles were often related to highly specialized domains such as business law, network security, and software engineering. For each sampled domain, these five component scores were normalized and averaged into an overall measure of knowledge intensity, following Morgeson and Humphrey (2006).

[1] The Spearman rank correlation provides a more accurate measure of correlation between non-normally distributed variables with high kurtosis than the Pearson product-moment correlation coefficient (De Winter et al., 2016).

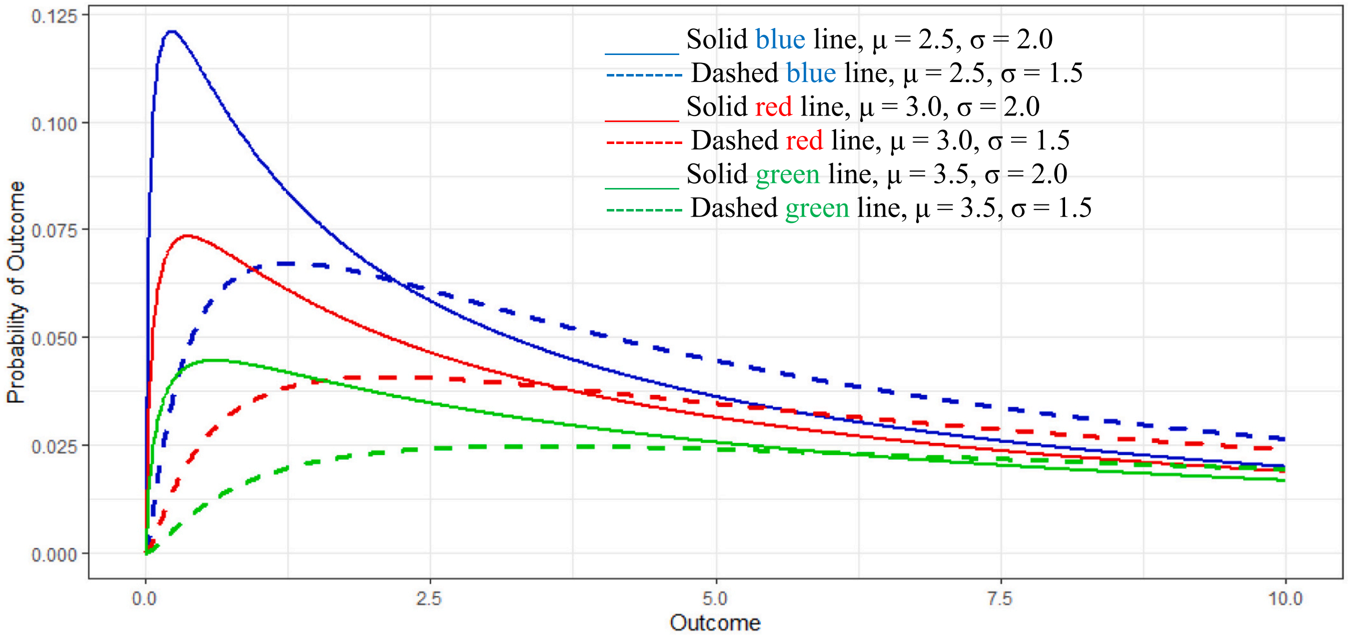

Fig. 2. Illustrative lognormal distributions.

3.3.2. *Initial performance and performance accumulation rate*

For each instructor, performance data were collected at two times spread 16 weeks apart. The initial value of instructor performance was measured as the cumulative number of students in mid-May 2022. The accumulation rate was measured as the number of new students added over the following 16 weeks, i.e., by mid-September 2022.

3.4. *Control variables*

Following prior research (Aguinis et al., 2016), three instructor-level variables were used to control for differences in educational level, professional commitment, and use of social media. The binary variable '*Degree*' was set to 1 if the instructor profile on Udemy explicitly mentioned an educational degree. The binary variable '*Professional*' was set to 1 if the instructor profile on Udemy listed a website. '*Social*' counts the number of social media platforms for each instructor, specifically Facebook, LinkedIn, Twitter, and YouTube. The number of instructors varied across domains ('*n*' range: 41 to 2622 for the 52 domains in the sample). End users on digital platforms may change their search and selection heuristics when faced with information overload due to an overwhelming number of alternatives (Aljukhadar et al., 2012). To account for this potential effect on the assessment of distributional shapes, we controlled for the number of available instructors in a given domain in the analyses for Hypothesis 3.

4. Results

4.1. *Hypothesis 1: non-normal distribution*

Hypothesis 1 (H1) posited the non-normality of performance, measured as each instructor's number of students, and alternatively as the number of reviews. The Pearson correlation between these two performance measures was 0.65 [$CI_{95\ \%}$: 0.64, 0.66], and the Spearman correlation was 0.84 [0.84, 0.85]. Table 1 shows the descriptive statistics across all sampled 52 domains; there is greater tail extremity in the number of reviews than in the number of students, as evidenced by the respective kurtosis values.

Fig. 3 shows a histogram with the number of instructors plotted against the number of students per instructor. The largest panel shows the distribution for the entire range of performance (number of students from 1 to 2,225,509). The middle panel shows the distribution for the range from 1 to 100,000; the smallest panel shows the distribution for the range from 1 to 5000. The distributions are non-normal, skewed, and heavy-tailed.

The non-normality of the distribution shown in Fig. 3 was statistically tested using the Kolmogorov-Smirnov (K-S) test ($D = 0.43$; $p = 2.2e\text{-}16$) and the Anderson-Darling (A-D) test ($A = 3616.8$; $p < 2.2e\text{-}16$); the latter is a stronger test for H1 because it is more sensitive in the detection of heavy right tails (D'Agostino and Stephens, 1986; Engmann and Cousineau, 2011). The null hypothesis of a normal distribution was rejected in each case.

In addition, we plotted histograms for instructor performance in each domain in our sample ('*n*' range: 41 to 2622, where '*n*' is the number of instructors in each domain). In visual evaluations, each histogram displayed a non-normal, heavy-tailed distribution, consistent with Hypothesis 1 (H1). Moreover, the *p*-values for the K-S and A-D tests for each of the 52 domains indicated evidence against the null hypothesis of a normal distribution of performance (p-value range: 8.3e-07 to 2.2e-16).

Fig. 4 shows histograms for domains with the minimum and the maximum number of instructors. In sum, these histograms and statistical tests are consistent with not only non-normality but also heavy-tailed distribution of entrepreneurial performance and star entrepreneurs on Udemy. For the domain featured in the left panel, just 12 (29 %) of 41 instructors account for about 80 % of total students across all instructors; for the one on the right, just 311 (12 %) of 2622 instructors account for about 80 % of total students across all instructors.

4.2. *Hypothesis 2: lognormal distribution*

Hypothesis 2 (H2) was tested in two stages. First, we investigated the overall functional form of the performance distribution for all 12,391 sampled instructors using the 'R' package *poweRlaw* (Clauset et al., 2009; Gillespie, 2014). In a second step, we conducted fine-grained, pairwise comparisons of 'fit' between seven alternative types of distributions using the '*R' package Dpit*, developed by Joo et al. (2017). Consistent with prior research, this investigation of the overall shape of performance distributions did not employ any statistical control variables (Aguinis et al., 2016; Bradley and Aguinis, 2023; Crawford et al., 2015; Joo et al., 2017).

For the pooled performance distribution of all 12,391 instructors, we fitted a lognormal distribution with the following parameters: scale ($\mu = 6.95$), shape ($\sigma = 2.31$), and x_{min} (369) using the 'R' package *poweRlaw* (Clauset et al., 2009; Gillespie, 2014), which employs maximum likelihood estimation. The K-S goodness-of-fit statistic (0.009) was below a conservative critical value (0.011) computed as $1.22/\sqrt{12{,}391}$ for $\propto = 0.1$ (Wicklin, 2019), which indicated general support for lognormality (H2). For more fine-grained evaluation,

Table 1
Descriptive statistics for performance across domains.

Performance	Mean	Median	Skew	Kurtosis	SD	Min	Max
No. of students	10,906	551	16.1	358.5	62,989	1	2,225,509
No. of reviews	874	41	28.0	1164.9	7,101	1	408,356

Note: n = 12,391.

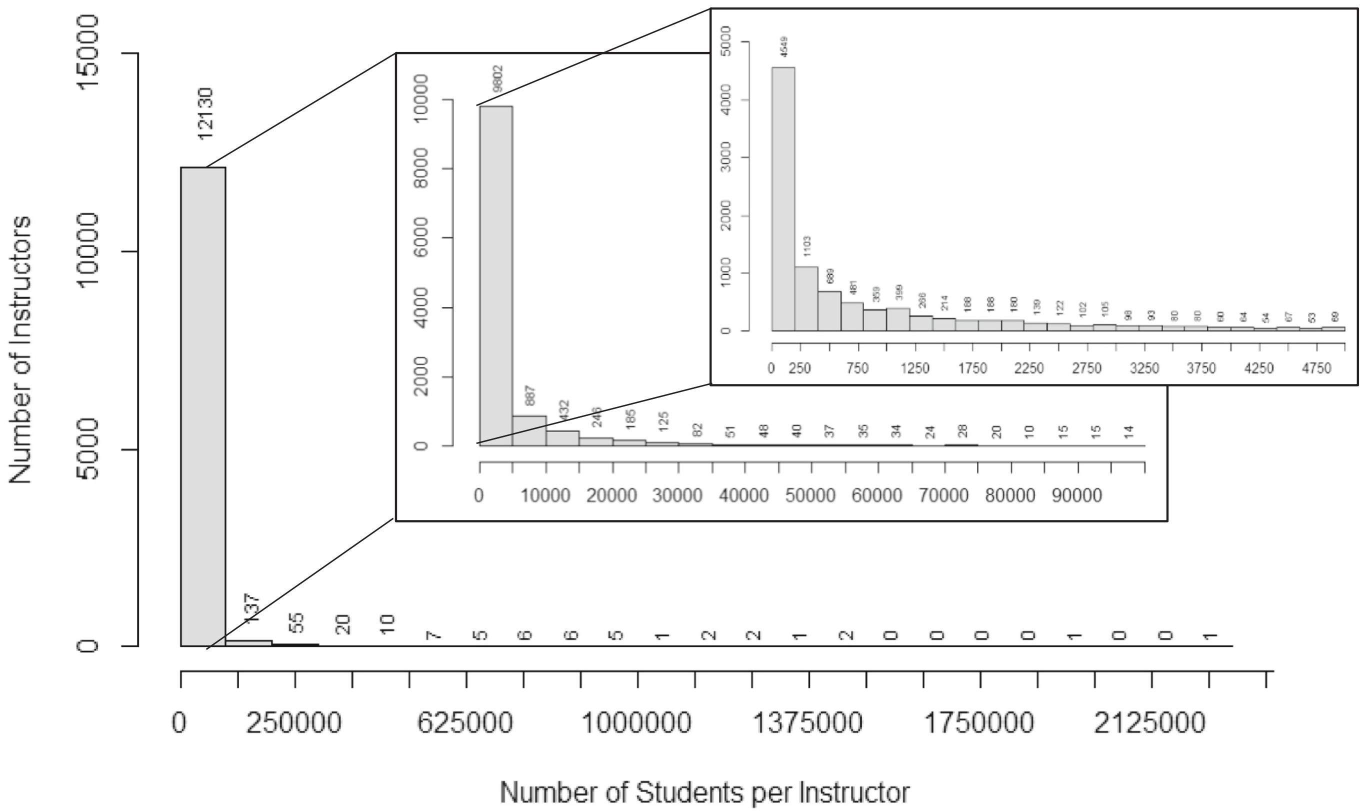

Fig. 3. Histogram of performance across domains.

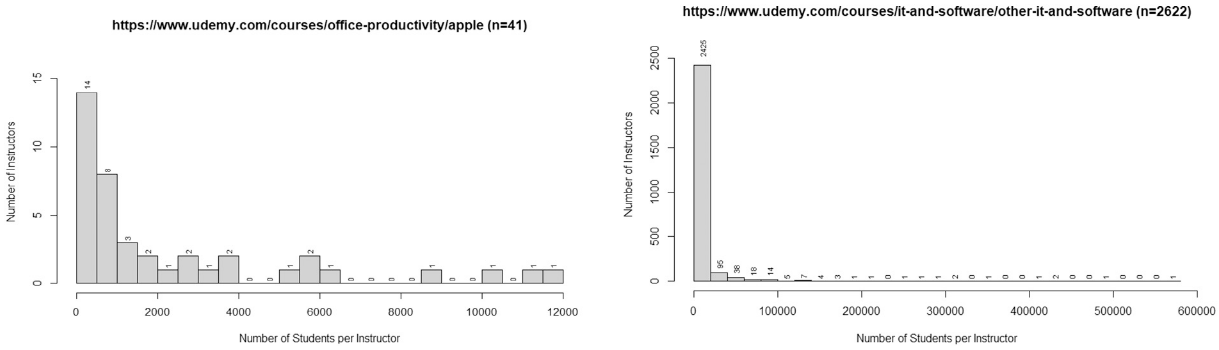

Fig. 4. Histogram of performance for two specific domains.

Table 2
Frequency of performance distributions by domain.

		Domain count	Percentage
Dominant distribution	Exponential tail	0	0
	Lognormal	34	65 %
	Normal	0	0
	Power Law (PL)	0	0
	PL with cutoff	0	0
	Poisson	0	0
	Weibull	0	0
Co-dominant distributions	Lognormal and PL with cutoff	16	31 %
	Lognormal, PL with cutoff, and Weibull	1	2 %
	Lognormal and Poisson	1	2 %
No dominant or co-dominant distributions		0	0
Total		52	100 %

we compared the observed performance distribution in each of the 52 domains across seven alternative distribution shapes using the distribution pitting methodology proposed by Joo et al. (2017). Their *Dpit* package estimates log-likelihood ratios (LR) and p-values for each pairwise comparison, resulting in 21 comparisons for each domain.

With the null hypothesis of no difference (LR = 0) between the two distributions and a p-value cutoff of 0.1 (Clauset et al., 2009), the dominant distribution for each domain was the distribution that was never identified as being a worse fit than any of the other six distributions across the 21 pairwise comparisons. If two distributions were equally dominant, their direct pairwise comparison was used to identify the dominant distribution. In case of an inconclusive direct comparison, the principle of parsimony was used, following Joo et al. (2017), to mark the distribution with fewer parameters as the *dominant* distribution or, if they had the same number of parameters, mark both distributions as *co-dominant*. Fig. 1 shows the number of parameters for each distribution. Table 2 summarizes the findings for domain-wise tests of distribution fit.

The lognormal distribution was the dominant fit for 34 (65 %) of the 52 domains and the co-dominant fit for the remaining 18 domains. In 16 of these 18 cases, the power law with exponential cutoff and the lognormal distributions were co-dominant. In one domain, the lognormal, power law with exponential cutoff, and Weibull distributions were co-dominant. Finally, in one domain, the lognormal and the Poisson distributions were co-dominant. Thus, the lognormal distribution was never dominated by any other type of distribution. These results provide strong support for H2, that the performance of entrepreneurs on digital platforms follows a lognormal distribution with heavy tails, indicating the prevalence of star entrepreneurs. The observation that, after the Poisson distribution, the normal distribution was the second-worst fit for 51 of the 52 domains offers further support for H1.

For each of the 52 domains, we also evaluated the lognormality of its distribution using the Kolmogorov-Smirnov goodness-of-fit statistic. We applied a conservative critical value, computed as $1.22/\sqrt{n_i}$ for $\propto = 0.1$, where '*n*' is the number of instructors in domain '*i*' (Wicklin, 2019). This statistical test provided further evidence at the domain level that was consistent with Hypothesis 2, i.e., lognormal distribution of entrepreneurial performance within domains. Fig. 5 illustrates how the observed distribution for a specific domain where lognormal was dominant (green) differs from the observed distribution in another domain where lognormal was co-dominant with the power law with exponential cutoff (red) and from the observed distribution in a third domain where lognormal was co-dominant with the power law with exponential cutoff and Weibull (blue). Performance is log-transformed for ease of visual interpretation and to highlight the variance in tail extremity across distributions (Wennberg and Anderson, 2020). As expected, the lognormal distribution has *fatter* and *longer* tails than the Weibull and power-law with exponential cutoff distributions.

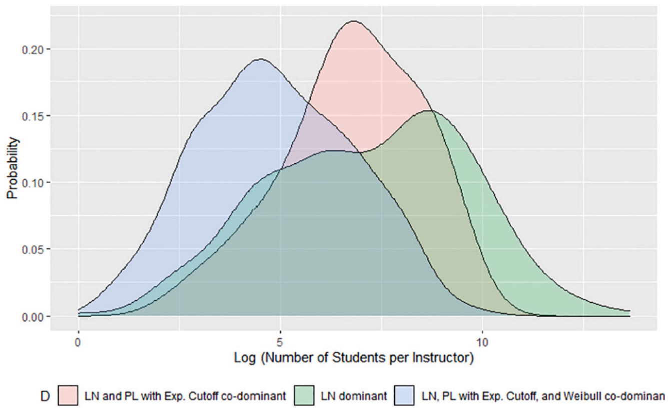

Fig. 5. Log-transformed performance distributions for domains where lognormal (LN) was either dominant or co-dominant with Power Law (PL) with Exponential Cutoff and Weibull Distributions.

Table 3
Correlation analysis for domain knowledge intensity (Hypothesis 3).

	Lognormal shape (σ) parameter	Number of instructors	Knowledge intensity
Lognormal scale (μ) parameter	−0.254† [−0.512, 0.046]	−0.024 [−0.319, 0.275]	0.446** [0. 172, 0.656]
Lognormal shape (σ) parameter		0.388** [0.102, 0.613]	−0.132 [−0.412, 0.172]
Number of instructors			0.209 [−0.093, 0.476]

Note: $n = 44$; †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Numbers in brackets indicate 95 % confidence intervals.

4.3. Hypothesis 3: knowledge intensity

Hypothesis 3 (H3) proposed a moderating effect of a domain's knowledge intensity on the tail extremity of its performance distribution. First, we identified the domains where the lognormal distribution was dominant or co-dominant (Table 2), and the K-S statistic was below the conservative cutoff value specified above. Next, for each domain, three aforementioned instructor-level control variables (*Degree*, *Professional*, and *Social*) were factored in using the residual procedure for variance partitioning (Hambrick and Quigley, 2014), following the supplemental analyses conducted by Aguinis et al. (2016, p. 45). The residuals were then tested for their 'fit' to a lognormal distribution.

Of the 52 domains for which residuals were computed, the lognormal distribution was the dominant fit for 5 domains. Moreover, it was identified as the co-dominant fit for another 39 domains. For these 44 domains with a lognormal distribution of residuals, the scale (μ range: 6.73 to 10.61) and shape (σ range: 0.54 to 1.88) parameters were estimated. Table 3 shows the pairwise Spearman correlations and corresponding confidence intervals based on the lognormal parameters that 'fit' the residuals. We found evidence consistent with the hypothesized positive association ($r = 0.45$ [0.17, 0.66], $p = 0.002$) between a domain's knowledge intensity and its lognormal scale parameter (μ).

Fig. 6 depicts the performance distribution for three domains to illustrate how the tail extremity of the lognormal distribution is greater for domains with higher knowledge intensity. Performance is log-transformed for ease of visual interpretation and to highlight the variance in tail extremity across distributions (Wennberg and Anderson, 2020). The three distributions of performance represent domains with low (green, $\mu = 5.7$, $\sigma = 2.7$), medium (blue, $\mu = 7.0$, $\sigma = 1.8$), and high (red, $\mu = 7.8$, $\sigma = 2.1$) levels of knowledge intensity.

Table 4 shows the results from additive nonparametric multiple regression for testing Hypothesis 3. Unlike OLS regression, nonparametric regression relaxes the assumption of linear relationships between the independent variables and the dependent variable; it allows for heavy-tailed or skewed distributions of one or more variables that constitute the regression analysis (Stone, 1985; Fox, 2005). The non-normal distributions observed in the focal and control variables make additive nonparametric regression more suitable than OLS regression for testing Hypothesis 3.[2] We implemented this method using the 'R' package *mgcv*, which provides nonparametric estimates for generalized additive models (R: MGCV, 2022). In the first column of Table 4, s(Variable) indicates the use of smoothing splines for that variable; we retained the default smoothing parameters of the *s* function in the *mgcv* package (Fox, 2002). Based on an ANOVA test of two models, one with only the control variables and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.14$; $p = 0.004$) of an association between the knowledge intensity of a domain and the lognormal scale

[2] We thank an anonymous reviewer for suggesting this regression technique.

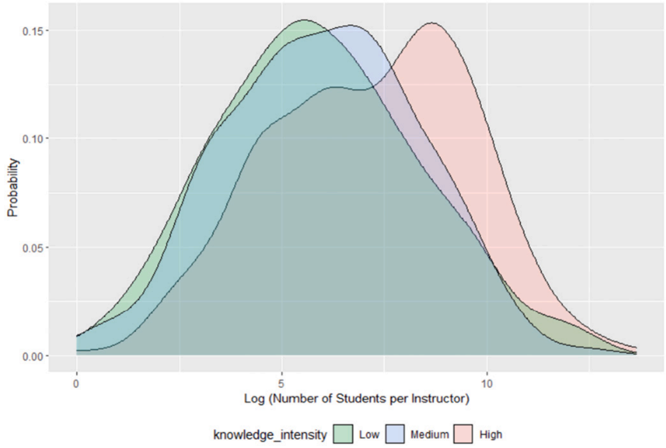

Fig. 6. Log-transformed performance distributions for domains with low, medium, and high knowledge intensity.

Table 4
Additive nonparametric regression of lognormal scale parameter (μ).

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1 s(Lognormal sigma) + s(Number of instructors)	38.840	26.932					0.15
Model 2 s(Lognormal sigma) + s(Number of instructors) + s(Knowledge intensity)	37.986	21.923	0.854	5.097	10.321	0.004**	0.29

Note: n = 44; $^{\dagger}p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

parameter (μ) for the fitted distribution.

Fig. 7 shows the component plot (created using the 'R' package *ggeffects*) for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005). The right panel illustrates the positive relationship between knowledge intensity and the lognormal scale parameter (μ) as a measure of tail extremity in performance. These results are consistent with Hypothesis 3.

4.4. *Hypotheses 4a and 4b: proportional differentiation*

Hypotheses 4a and 4b provide a direct empirical test for the underlying generative mechanism that created the observed lognormal distribution of entrepreneurial performance.

As illustrated in Table 5, support for both H4a and H4b would represent evidence of proportional differentiation as the generative mechanism (Joo et al., 2017). However, support for H4b without support for H4a would indicate evidence of incremental differentiation (Joo et al., 2017). Conversely, support for H4a but not for H4b would signify preferential attachment as the generative mechanism of non-normal distributions (Mitzenmacher, 2004). Table 5 excludes the generative mechanism of pure power laws, which could be a combination of self-organized criticality and exogeneous shocks, because (a) the empirical test of Hypothesis 4a and 4b does not measure these constituents, (b) the findings for Hypothesis 2 indicate that pure power laws were not identified as the dominant fit for any domain, and (c) pure power laws theoretically imply infinite mean and variance, whereas the data for Udemy shows extreme but finite performance and therefore finite mean and variance.

To the best of our knowledge, generative mechanisms have not been tested in management and entrepreneurship research. We used the suggestions on non-traditional empirical methods by Crawford et al. (2015) as the starting point for this analysis. Accordingly, we tested H4a and H4b using Somers' Delta (Somers, 1962), a directional estimate of the association between ordinal-scaled "*predictor variable X and outcome variable Y*" (Metsämuuronen, 2021; Newson, 2006, p. 1). To capture changes in entrepreneurial performance over time, we used the longitudinal data collected at time points T1 and T2, 16 weeks apart. First, the *initial value* of performance for each domain, measured as the number of students at time T1, was treated as predictor (X), while performance at time T2 was treated as the outcome variable (Y). Somers' D for this pair was computed using the R package *DescTools*, which converts continuous data into ordinal data per the Somers' D measure (R: Somers' Delta, 2022). Such ordinal alternatives to Pearson correlations offer more accurate measures of association between non-normally distributed variables with high kurtosis (De Winter et al., 2016). Second, the *accumulation rate* for each instructor, computed as the increase in students from T1 to T2, was treated as predictor (X), while performance at

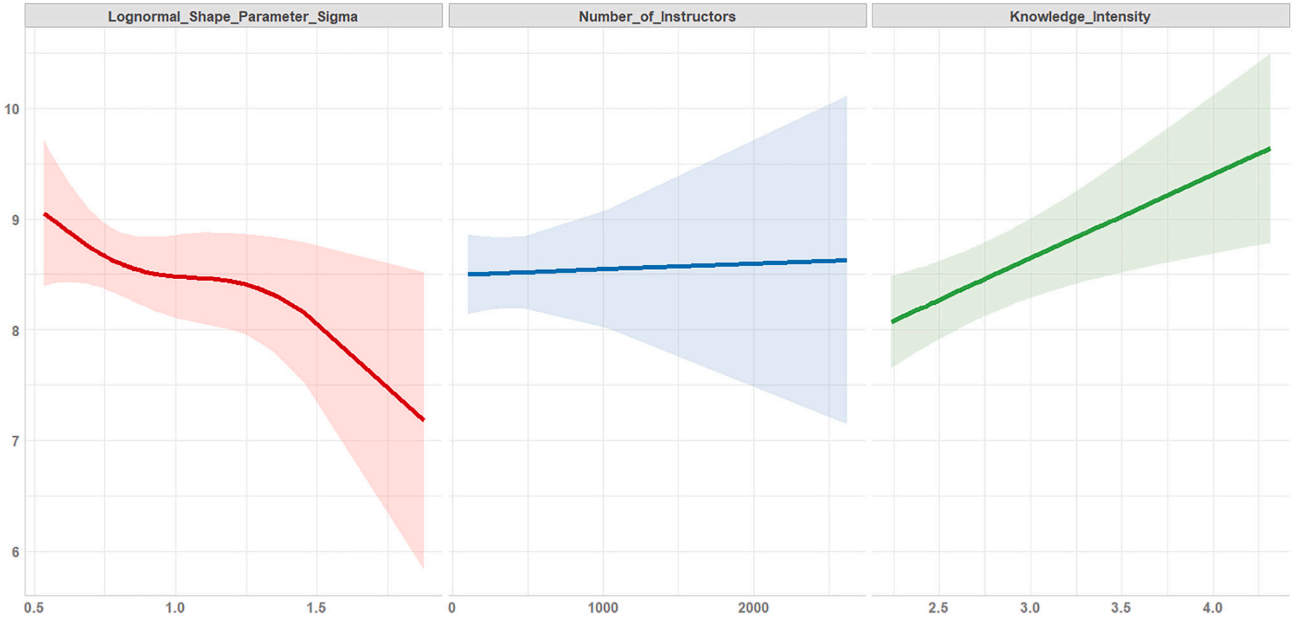

Fig. 7. Component plot of (a) lognormal shape parameter, (b) number of instructors, and (c) knowledge intensity with smoothing functions from additive nonparametric regression.

Table 5
Generative mechanisms and their constituents.

Mechanism that generates heavy-tailed distributions	Initial value of performance	Accumulation rate of performance
Proportional differentiation	✓	✓
Preferential attachment (also called Matthew effect, rich-get-richer effect)	✓	✖
Incremental differentiation	✖	✓

Table 6
Somers' D measure of association across domains.

Somers' D	All domains (n = 52)	Domains where lognormal dominant (n = 30)	Domains where lognormal co-dominant (n = 22)
Initial Value, Outcome pair			
Average	0.96	0.96	0.96
[Lowest LCI, Highest UCI]	[0.86, 0.99]	[0.86, 0.99]	[0.88, 0.99]
Accumulation Rate, Outcome pair			
Average	0.36	0.34	0.37
[Lowest LCI, Highest UCI]	[0.01, 0.71]	[0.01, 0.67]	[0.08, 0.71]

Note: LCI – Lower Boundary of 95 % Confidence Interval; UCI – Upper Boundary of 95 % Confidence Interval.

time T2 was treated as the outcome variable (Y). Again, the corresponding Somers' D was computed. The Spearman rank correlation between initial value and accumulation rate ranged from 0.22 to 0.75 across the 52 domains, indicating a moderate degree of correlation between the two predictors.

As summarized in Table 6, Somers' D estimates indicated a positive association between the initial value (performance at time T1) and the outcome (performance at time T2) for each of the 52 domains; the corresponding 95 % confidence intervals did not include zero. Moreover, the average Somers' D was 0.96, providing evidence against incremental differentiation. This finding is consistent with H4a. Somers' D estimates for the relationship between the accumulation rate (increase in performance from time T1 to T2) and outcome (performance at time T2) were also positive for each of the 52 domains, and the corresponding 95 % confidence intervals did not include zero. This finding is consistent with H4b. Table 6 also shows that the Somers' D estimates remain the same whether the lognormal distribution is solely dominant ($n = 30$) or co-dominant ($n = 22$) with other distributions.

Together, these Somers' D estimates, by supporting both H4a and H4b, indicate evidence of proportional differentiation as the generative mechanism for the lognormal distribution of entrepreneurial performance on Udemy. Notably, the average effect (Somers' D) of the initial value on the performance outcome at T2 was greater than the effect of the accumulation rate. Further investigations showed the same effect size differences for each of the 52 domains, and the confidence intervals of the two effect size estimates did not overlap for any domain. In other words, while entrepreneurs with a lower initial value of performance may be able to surpass those with a higher initial value via a superior accumulation rate over time, early performance advantages have a relatively stronger influence on the eventual performance and its distribution. Thus, the observed data are consistent with H4a and H4b, suggesting proportional differentiation as an important generative mechanism of heavy-tailed distributions of performance on digital platforms.

4.5. Robustness tests

For Hypotheses 1 and 2, the number of reviews was used as an alternate measure of entrepreneur performance to test whether the lognormal distribution was a better 'fit' for the performance distribution of instructors across the sampled domains. Reviews are a relevant measure of performance because they are more likely to be provided by students who not only paid for and consumed the entire course but also had definitive – positive or negative – opinions about the instructor and the content (Kordrostami and Rahmani, 2020). Histograms and tables for this measure are available in Appendix 5.

The statistical tests for Hypotheses 1 and 2 were repeated using the number of reviews, and the performance distribution across all domains was found to be non-normal. Moreover, using the *Dpit* package, the lognormal distribution was found to be a better 'fit' than the other six types of distributions specified in the package. Also, the K—S goodness-of-fit statistic (0.007) was below the conservative critical value (0.011) computed as $1.22/\sqrt{12{,}391}$ for $\propto = 0.1$ (Wicklin, 2019), thus indicating lognormality. The *Dpit* package identified the lognormal distribution as the dominant fit for 44 (85 %) of the 52 domains and a co-dominant fit for the remaining 8 domains. Thus, using the alternate performance measure, we found evidence consistent with Hypotheses 1 and 2.

The robustness of findings for Hypothesis 3 was tested using an alternative measure, ***kurtosis,*** to capture the tail extremity of performance distributions. Kurtosis uses sampled observations to estimate the propensity of extreme data points in the population (Westfall, 2014). For a sample of size n with mean μ and standard deviation σ, kurtosis is computed as

$$K = \frac{1}{\sigma^4}\left[\frac{1}{n}\sum\nolimits_{i=1}^{n}(x_i - \mu)^4\right].$$

Scholars have recently explicated the relevance of kurtosis to non-normal distributions with sufficient sample size, highlighting it as a measure of the contribution of extreme data points vis-à-vis that of other points in the distribution (Celikoglu and Tirnakli, 2018). See Appendix 2 for results consistent with Hypothesis 3 for this alternate measure of tail extremity. We also tested the robustness of this study's findings using OLS regression instead of additive unrestricted nonparametric multiple regression. Results again showed a positive association between the knowledge intensity of a domain and the tail extremity of entrepreneurial performance for that domain, consistent with Hypothesis 3. Results also did not change when we dropped the *Social* control variable, which accounted for 66 % of the instructors who had social media accounts.

Reported findings for H4a and H4b were robust when using Kendall's Tau-A – instead of Somers' D – to estimate ordinal association (Liebetrau, 1983); corresponding results are reported in Appendix 3. Hypotheses 4a and 4b were also tested using the number of reviews as an alternate measure of entrepreneurial performance. The initial value at T1 and accumulation rate from T1 to T2 were both positively associated with the number of reviews at T2, suggesting proportional differentiation as the generative mechanism using the alternate performance measure. See Table A5.6 in Appendix 5 for the results. Furthermore, we established the robustness of results supporting H1, H2, and H3 by testing these hypotheses on the performance data collected in the second round for both performance measures. See Appendix 4 and 5 for details.

5. Discussion

As the digitization of goods and services continues to reduce the marginal costs of production and distribution (Bakos and Brynjolfsson, 1999; Goldfarb and Tucker, 2019; Shapiro et al., 1998), digital platforms are becoming increasingly central to related entrepreneurial activity (van Gelderen et al., 2021), with feedback loops and network effects acting as tailwinds for star entrepreneurs (Cusumano, 2022; Nambisan et al., 2018; Belleflamme and Peitz, 2018). With trillions of dollars flowing through the global digital economy, digital technologies and platforms have drawn increasing scholarly attention (e.g., Fan et al., 2021; Johnson et al., 2022; McMullen et al., 2021b; Paul et al., 2023; Zaheer et al., 2019).

We seek to contribute to this literature by exploring the performance distributions of entrepreneurs on the digital platform Udemy.com. We find support for lognormal performance distributions, indicating the prevalence of star entrepreneurs and their generative mechanism of proportional differentiation. These findings supplement prior research on heavy-tailed performance distributions in organizational settings that suggested power-law or exponential tail distributions of performance (Bradley and Aguinis, 2023; Crawford et al., 2015; Joo et al., 2017).

5.1. Implications for theory

Considering the unique characteristics of digital technologies and platforms (Cutolo and Kenney, 2021; Nambisan, 2017), this study provides substantive empirical evidence that the distribution of entrepreneurial performance on such platforms is far from normal. Instead, it is dominated by heavy tails created by a handful of star entrepreneurs. This finding alerts scholars to the risks of not accounting for heavy-tailed distributions of entrepreneurial performance. On Udemy (within the sampled 52 domains), the top 7 % of instructors accounted for about 80 % of students. Thus, focusing on the 'average' entrepreneurs while ignoring star entrepreneurs may severely limit our understanding, explanation, and prediction of the nature of success in digital-platform entrepreneurship. Explicitly accounting for star entrepreneurs is critical not only for explaining more variance in empirical data (Makino and Chan, 2017) but also for identifying the underlying success factors that generate such heavy-tailed performance distributions (Gibbert et al., 2021). Thus, this study answers recent calls to revisit long-held assumptions of normal distribution of performance in organizational research (Beamish and Hasse, 2022; Crawford et al., 2022; Dean et al., 2007; Delmar et al., 2022).

Reported findings extend non-normal performance research to entrepreneurship on digital platforms. We contribute the insight that performance is better characterized by the lognormal distribution, which differs from the exponential tail and power law with exponential cutoff distributions typically found in other employment and work contexts (Bradley and Aguinis, 2023; Joo et al., 2017). Thus, this study identifies the need to adapt prevailing statistical methodologies to account for lognormal distributions by drawing upon Paretian paradigms (McKelvey and Andriani, 2005) and scholarly guidelines for incorporating non-normality in statistical analyses (Becker et al., 2019; Rönkkö et al., 2022). Specifically, pervasive non-normality in entrepreneurship variables calls for (a) identification of the most suitable distributional shapes using a pitting or falsification approach, (b) examination of the proportion or weight that outliers carry in terms of the overall pool, (c) adoption of appropriate analytical methods (e.g., rank correlations, nonparametric regression) for hypothesis testing, and (d) theoretically informed expansion of the taxonomy of distribution shapes to account for novel generative mechanisms of extreme performance.

Notably, the longitudinal nature of this study's data enabled the identification of proportional differentiation – where *both* initial performance and the rate at which performance accumulates over time matter – as the generative mechanism of star entrepreneurs on digital platforms. This finding paves the way for a more systematic and deeper understanding of how extreme performance emerges in digital contexts. It extends the broader star performance literature, which has primarily investigated traditional employment and work environments, and identifies incremental differentiation as the likely generative mechanism (Bradley and Aguinis, 2023; Joo et al., 2017). Thus, this study also contributes to the literature on generative mechanisms of star performers (Asgari et al., 2021; Crawford et al., 2015; Crawford et al., 2022), which has theorized but, to the best of our knowledge, not directly tested and verified such mechanisms.

Furthermore, this study identifies knowledge intensity (Morgeson and Humphrey, 2006) as a domain-level contextual factor that influences the observed non-normal entrepreneurial performance on digital platforms. Scholars should interpret this as a signal to examine the potentially important role of other contingency factors that may alter the tail extremity of performance distributions. Together, these findings of the underlying generative mechanisms and moderating factors of extreme performance may help scholars progress toward a coherent and parsimonious theory of digital platform entrepreneurship. More broadly, this study contributes an interdisciplinary perspective to entrepreneurship theory by combining concepts from scholarly research on complexity science (e.g., self-organized criticality), information economics (e.g., zero marginal costs), and work design (e.g., knowledge intensity).

5.2. Implications for practice

The reported heavy-tailed performance on a specific digital platform offers practical insights for active and future digital-platform entrepreneurs (Cutolo and Kenney, 2021). First, it draws attention to the high variance in performance that may often occur in such pursuits, wherein star performers may operate orders of magnitude better than their ‘average’ peers. On Udemy, the highest-performing instructor had 204 times as many students as the average instructor. This insight may inform entrepreneurial entry, product development, or scaling decisions. It can also help entrepreneurs map expected performance distributions to their odds of success in such winner-take-all environments.

Our exploration and explication of positive skew in performance distributions are particularly relevant for contexts where (a) the goods or services being offered are purely digital in nature as is their distribution, (b) the reviews and ratings provided by customers to entrepreneurs are publicly visible, and (c) the digital platform has enough critical mass in terms of offerings, engagement, and visibility to enable strong network effects. In addition, the reported findings alert practitioners to consider the impact of specific domain characteristics, such as knowledge intensity, on the tail extremity of performance and the prevalence of star entrepreneurs. Indirectly, this study's findings reveal the possibility that other, ill-understood, contingency factors may shape the odds for extreme performance in digital entrepreneurship.

This study's empirical finding of proportional differentiation as an important generative mechanism of star entrepreneurs on digital platforms offers insights to novice or average entrepreneurs who aspire to become star performers. While initial out-performance may arise from dumb luck (Barabási, 2012; Denrell and Liu, 2012) or first-mover advantages (Varadarajan et al., 2008), the finding that the accumulation rate – not only the initial value – of performance influences eventual outcomes implies that entrepreneurs can and should pursue competitive strategies and tactics targeted at overcoming disadvantages against entrenched star entrepreneurs (Srinivasan and Venkatraman, 2018). In other words, an initial performance advantage helps star entrepreneurs but does not guarantee their long-term success. New and fast-growing later entrants have the potential to achieve star status. This may be particularly salient for entrepreneurs operating on B2C platforms such as Udemy, wherein creating and distributing purely digital goods removes many scaling constraints associated with physical products.

Moreover, entrepreneurs may overcome initial disadvantages by proactively identifying and selling through platforms that have reached inflection points, characterized by a critical mass and concomitant network effects but a lack of entrenched star entrepreneurs (Hagiu and Rothman, 2016; Stummer et al., 2018). This study's empirical support for proportional differentiation as the likely generative mechanism of extreme performance is consistent with the theorized centrality of ratings, reviews, and recommendations to star performance on digital platforms (Belleflamme and Peitz, 2018; Le Mens et al., 2018; Luca and Zervas, 2016). Accordingly, those competing with star entrepreneurs ought to promptly evaluate and respond to customer reviews (Archak et al., 2011) and provide incentives for new reviews (Zhang et al., 2020) to leverage the feedback loops built into transparent digital platforms for a higher accumulation rate of performance.

5.3. Limitations and future research

Empirically, this study focuses on a single digital platform, Udemy, to test its hypotheses. In this, it differs from prior research (e.g., Bradley and Aguinis, 2023; Crawford et al., 2015; Joo et al., 2017), which has used a compilation of publicly available, cross-sectional data to investigate the performance of hundreds of thousands of individuals and teams using a variety of performance measures, largely in physical contexts. While our study's longitudinal data and focus on a single platform offers some benefits with respect to internal validity and enables the empirical investigation of the mechanism that underlies star performance, we recommend the constructive replication of this study across a variety of digital platforms, entrepreneurs, and offerings to evaluate the generalizability of the reported findings (Anderson et al., 2019). Thus, we encourage future research involving digital platforms that differ in (a) the nature and scale of goods sold, (b) the level of transparency of ratings and reviews, and (c) the extent to which network effects and externalities influence performance. Scholars should, for example, consider investigating star entrepreneurs on platforms that connect enterprises to consumers (e.g., Coursera.org), involve the distribution of physical goods such as handmade crafts (e.g., Etsy.com), or focus on peer-to-peer business models such as local commerce (e.g., Offerup.com).

On business-to-business (B2B) platforms, sellers and buyers may be inclined toward repeat business and, therefore, rely less on reviews and ratings than those on business-to-consumer (B2C) platforms, limiting the tail extremity of performance for entrepreneurs on B2B platforms. Similarly, the combination of physical and digital offerings on digital platforms, with the concomitant higher marginal costs of production and distribution, should constrain the tail extremity in entrepreneurial performance. Clearly, future research is needed not only to substantiate these arguments but also to quantify the influence of these differences in product and platform characteristics on the tail extremity of performance distributions. Moreover, an ecosystem perspective may be well suited to the examination of entrepreneurship on digital platforms (Eckhardt et al., 2018), wherein platform policies drive value creation, with platform owner-operators competing with platform-based entrepreneurs for value appropriation (Chen et al., 2022).

Following extant research (Bradley and Aguinis, 2023; Joo et al., 2017), this study focused on identifying a dominating generative mechanism of extreme performance. In doing so, we did not consider combinations of generative mechanisms (Andriani and McKelvey, 2009), partially in recognition of the substantial statistical power needed for these, more challenging, investigations. Future research may, therefore, explore combinations of exponential tail, lognormal, and power law shapes (Andriani and McKelvey, 2009) or novel distributions such as the double Pareto-lognormal proposed by Reed and Jorgensen (2004), thereby enhancing the explanatory power of investigations into extreme performance.

For the generative mechanism of proportional differentiation, future research can examine the relative influence of initial value and accumulation rate on final performance. For novice entrepreneurs to become or compete with star entrepreneurs, an increase in accumulation rate of performance is crucial to overcome initial disadvantages. Therefore, scholars should study the factors (e.g., platform size and policies, industry or market factors, and strength of network effects) that determine whether such outperformance in accumulation rate can be compounded over time to overcome initial luck or first-mover advantages of entrenched star entrepreneurs. While this study lacked access to the 'true' initial value of performance for Udemy instructors, future longitudinal studies with access to initial and several subsequent performance measures could not only trace the evolution of star performers but also identify when and how late entrants can outperform entrenched entrepreneurs.

This study did not investigate the role of exogenous shocks, such as fundamental changes in digital platform regulations and practices, which may influence the likelihood and strength of non-normal outcome distributions. Digital platforms are often owned and operated by large technology corporations with substantial discretion to set, change, and enforce changes in platform policy (Cutolo and Kenney, 2021). For instance, an entrepreneur may experience an exogenous shock by being banned from or prominently featured on a digital platform. Alternatively, a sudden substantial change in the revenue-sharing policies of a digital platform may alter the extremity of entrepreneurial outcomes. For example, Etsy's 'Star Seller' program (Etsy, 2022) and associated revisions in advertising policies and listing fees drew criticism from new and 'average' entrepreneurs because the program implicitly granted greater cumulative advantages to entrenched star entrepreneurs (EtsyStrike, 2022).

Such disruptions to entrepreneurial performance on digital platforms may increase in conjunction with the pervasiveness and power of platforms in the digital economy (Cioffi et al., 2022). For example, a high 'take rate' by the platform owner-operator may increase tail extremity by making economies of scale crucial for the survival of 'average' entrepreneurs while, at the same time, it may decrease tail extremity by motivating incumbent star entrepreneurs to pursue other distribution channels. Therefore, we encourage scholars to empirically explore how extreme outperformance by entrepreneurs is potentially amplified or dampened by exogenous shocks such as (a) changes in revenue-sharing arrangements, (b) negative network externalities such as being delisted from the platform, or (c) positive network externalities such as being featured prominently on the platform (Gokkaya and Wai, 2017). Such investigations promise to enrich not only digital entrepreneurship theory and practice but may also inform policy decisions by owner-operators of digital platforms and their governmental and industry regulators (Cioffi et al., 2022).

Another future research direction is the focused investigation of extreme values applying, for example, Extreme Value Theory, which suggests a different set of distributions to capture extremely rare but impactful events (Charras-Garrido and Lezaud, 2013). Similarly, scholars may consider Laplace distributions for more complex cases when entrepreneurial performance can take on both positive and negative outcomes (e.g., return on investment, profitability), which implies the relevance of two tails for extreme phenomena in natural and social sciences (Gel, 2010). Finally, we call for future exploration at the intersection of outcome distributions and (a) entrepreneurial demographics such as race, age, or gender, (b) platform geography, for example, developed economies versus developing economies, and (c) entrepreneurial expertise, for instance, breadth or depth of experience in platform-based entrepreneurship, to isolate the influence of these characteristics and, in doing so, identify boundary conditions for proportional differentiation and lognormal performance distributions in digital platform entrepreneurship.

5.4. Conclusion

This study investigates heavy-tailed performance distributions and star entrepreneurs on digital platforms. We argue that such performance outliers are not anomalies to be discarded but the very focus and goal of entrepreneurs and, therefore, deserving of theoretical and empirical investigation. Studies that overlook these deviations from normality in performance distributions limit their explanatory power, predictive accuracy, and understanding of the mechanisms that drive extreme performance. We theorize and find empirical support for lognormal distributions as a better 'fit' for entrepreneurial performance in digital contexts. Importantly, we introduce novel theory for the prevalence of star entrepreneurs on digital platforms; we provide empirical evidence for a specific generative mechanism – proportional differentiation – to explain the observed, heavy-tailed distribution of entrepreneurial performance. Reported findings motivate and guide future research that delves into the increasingly prevalent phenomenon of digital entrepreneurship.

Declaration of funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Kaushik Gala: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Andreas Schwab:** Formal analysis, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. **Brandon A. Mueller:** Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

The authors do not have permission to share data.

Acknowledgements

We thank field editor Satish Nambisan and three anonymous reviewers for their constructive feedback throughout the review process. We thank our colleagues Annaleena Parhankangas and Scott G. Johnson for their valuable comments on earlier versions of this manuscript.

Appendix 1. List of 52 (of a total of 130) Udemy domains included in the sample

ID Website
16 https://www.udemy.com/courses/development/databases
18 https://www.udemy.com/courses/development/software-testing
20 https://www.udemy.com/courses/development/software-engineering
50 https://www.udemy.com/courses/business/industry
52 https://www.udemy.com/courses/business/media
58 https://www.udemy.com/courses/business/real-estate
60 https://www.udemy.com/courses/business/other-business
64 https://www.udemy.com/courses/marketing/search-engine-optimization
70 https://www.udemy.com/courses/marketing/marketing-fundamentals
76 https://www.udemy.com/courses/marketing/advertising
80 https://www.udemy.com/courses/marketing/content-marketing
98 https://www.udemy.com/courses/office-productivity/apple
100 https://www.udemy.com/courses/office-productivity/google
102 https://www.udemy.com/courses/office-productivity/sap
108 https://www.udemy.com/courses/office-productivity/other-productivity
114 https://www.udemy.com/courses/design/user-experience
116 https://www.udemy.com/courses/design/game-design
122 https://www.udemy.com/courses/design/fashion
124 https://www.udemy.com/courses/design/architectural-design
132 https://www.udemy.com/courses/it-and-software/it-certification
134 https://www.udemy.com/courses/it-and-software/network-and-security
136 https://www.udemy.com/courses/it-and-software/hardware
140 https://www.udemy.com/courses/it-and-software/other-it-and-software
156 https://www.udemy.com/courses/personal-development/happiness
168 https://www.udemy.com/courses/personal-development/self-esteem-and-confidence
170 https://www.udemy.com/courses/personal-development/stress-management
172 https://www.udemy.com/courses/personal-development/memory
186 https://www.udemy.com/courses/lifestyle/travel
188 https://www.udemy.com/courses/lifestyle/gaming
190 https://www.udemy.com/courses/lifestyle/home-improvement
194 https://www.udemy.com/courses/lifestyle/other-lifestyle
196 https://www.udemy.com/courses/photography-and-video/photography-fundamentals
198 https://www.udemy.com/courses/photography-and-video/photography-tools
220 https://www.udemy.com/courses/photography-and-video/other-photography-and-video

226 https://www.udemy.com/courses/health-and-fitness/sports
236 https://www.udemy.com/courses/health-and-fitness/self-defense
242 https://www.udemy.com/courses/health-and-fitness/meditation
244 https://www.udemy.com/courses/health-and-fitness/other-health-and-fitness
300 https://www.udemy.com/courses/music/music-fundamentals
302 https://www.udemy.com/courses/music/vocal
304 https://www.udemy.com/courses/music/music-techniques
306 https://www.udemy.com/courses/music/music-software
362 https://www.udemy.com/courses/development/development-tools
370 https://www.udemy.com/courses/photography-and-video/digital-photography
380 https://www.udemy.com/courses/teaching-and-academics/humanities
523 https://www.udemy.com/courses/teaching-and-academics/online-education
525 https://www.udemy.com/courses/teaching-and-academics/other-teaching-academics
527 https://www.udemy.com/courses/teaching-and-academics/teacher-training
542 https://www.udemy.com/courses/finance-and-accounting/finance-certification-and-exam-prep
544 https://www.udemy.com/courses/finance-and-accounting/financial-modeling-and-analysis
548 https://www.udemy.com/courses/finance-and-accounting/money-management-tools
552 https://www.udemy.com/courses/finance-and-accounting/other-finance-and-accounting

Appendix 2. Results for alternate measure of tail extremity for Hypothesis 3

Hypothesis 2 was tested using kurtosis as an alternate measure of tail extremity. This involved the 44 domains for whose residuals the lognormal was identified as the dominant or co-dominant fit, following the procedure in Section 4.3. Table A2.1 shows the pairwise Spearman correlations and corresponding confidence intervals indicating a positive correlation ($r = 0.29$ [−0.01, 0.54], $p = 0.056$) between a domain's knowledge intensity and the kurtosis of its residuals.

Table A2.1
Correlation analysis for Hypothesis 3.

	Number of instructors	Knowledge intensity
Kurtosis of distribution of residuals	0.372* [0.085, 0.602]	0.290† [−0.007, 0.541]
Number of instructors		0.209 [−0.094, 0.477]

Note: n = 44; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001. Numbers in brackets indicate 95 % confidence intervals.

Table A2.2 shows the results from additive nonparametric multiple regression with kurtosis as the dependent variable. Based on an ANOVA test of two models, one with only the control variable, number of instructors, and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.14$, $p = 0.006$) of an association between the knowledge intensity of a domain and its kurtosis, i.e., the tail extremity of its performance distribution.

Table A2.2
Additive nonparametric regression of kurtosis.

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1 *s*(Number of instructors)	39.774	8693					0.14
Model 2 *s*(Number of instructors) + *s*(Knowledge intensity)	38.745	11,020	1.029	11,307	8.283	0.006**	0.28

Note: n = 44; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.

Fig. A2.1 shows the component plot for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005); the right panel illustrates the positive relationship between knowledge intensity and kurtosis. These results are consistent with Hypothesis 3.

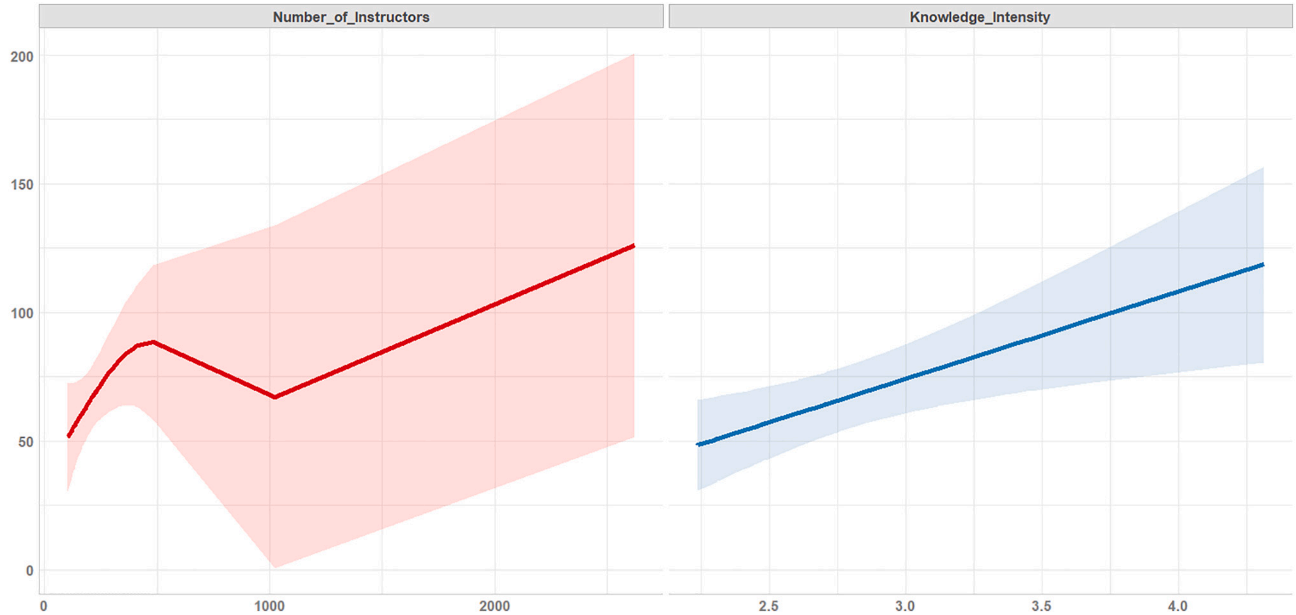

Fig. A2.1. Component plot of (a) number of instructors, and (b) knowledge intensity with smoothing functions from additive nonparametric regression.

Appendix 3. Analysis and results for Hypothesis 4a and 4b with Kendall's Tau-A instead of Somers' D as measure of association

Table A3.1
Kendall's Tau-A measure of association (performance = number of students).

Kendall's Tau-A	All domains (n = 52)	Domains where lognormal dominant (n = 30)	Domains where lognormal co-dominant (n = 22)
Initial Value, Outcome pair			
Average	0.96	0.96	0.96
[Lowest LCI, Highest UCI]	[0.85, 1.00]	[0.85, 0.99]	[0.88, 1.00]
Accumulation Rate, Outcome pair			
Average	0.36	0.34	0.37
[Lowest LCI, Highest UCI]	[0.01, 0.71]	[0.01, 0.67]	[0.08, 0.71]

Note: CI – 95 % Confidence Interval. LCI – Lower Confidence Interval. UCI – Upper Confidence Interval.

Appendix 4. Analysis and results of hypothesis testing for time T2 data (performance = cumulative number of students per instructor)

Table A4.1 shows the descriptive statistics for performance across all 52 domains listed in Appendix 1.

Table A4.1
Descriptive statistics for performance across domains.

Performance	Mean	Median	Skew	Kurtosis	SD	Min	Max
No. of students	11,874	642	16.1	365.9	68,398	2	2,535,147
No. of reviews	985	47	28.5	1213.7	7926	2	459,609

Note: n = 11,937.

4.1. Hypothesis 1

The non-normality of the performance distribution across all domains was statistically tested using the Kolmogorov-Smirnov (K-S) test (D = 0.43; *p* = 2.2e-16) and the Anderson-Darling (A-D) test (A = 3485.7; p = 2.2e-16) with the result that the null hypothesis of a normal distribution was rejected in each case. Moreover, the *p*-values for the K-S and A-D tests for each of the 52 domains indicated evidence (p-value range: 3.4e-07 to 2.2e-16) against the null hypothesis of a normal distribution of performance within each domain.

Thus, Hypothesis 1 was supported across and within domains for Time T2 data.

4.2. Hypothesis 2

For entrepreneurial performance across all 52 domains in the sample, the lognormal distribution was found to be dominant (n = 11,937). The K-S goodness-of-fit statistic (0.009) was below a conservative critical value (0.011) computed as $1.22/\sqrt{11{,}397}$ for $\propto = 0.1$ (Wicklin, 2019), thus indicating lognormality. The scale ($\mu = 6.90$), shape ($\sigma = 2.35$), and x_{min} (235) parameters of the fitted lognormal distribution were estimated for the whole sample using the *poweRlaw* (Clauset et al., 2009; Gillespie, 2014).

Furthermore, the *Dpit* package was used to identify the dominant or co-dominant distribution for performance, measured as the number of students per instructor. Table A4.2 summarizes the findings for domain-wise tests of distribution fit. The lognormal distribution was identified as the dominant fit for 30 (58 %) of the 52 domains. Moreover, the lognormal distribution was identified as the co-dominant fit for the remaining 22 domains. In 20 of these 22 cases, the power law with exponential cutoff was co-dominant with the lognormal. Hypothesis 2 was supported across and within domains for Time T2 data.

Table A4.2
Frequency of performance distributions by domain (time T2).

		Domain count	Percentage
Dominant distribution	Exponential tail	0	0
	Lognormal	30	58
	Normal	0	0
	Power Law (PL)	0	0
	PL with cutoff	0	0
	Poisson	0	0
	Weibull	0	0
Co-dominant distributions	Lognormal and PL with cutoff	20	38
	Lognormal, PL with cutoff, and Weibull	1	2
	Lognormal and Poisson	1	2
No dominant or co-dominant distributions		0	0
Total		52	100

4.3. Hypothesis 3 (lognormal parameters as measure of tail extremity)

For each domain where the lognormal distribution was dominant or co-dominant, performance was regressed on three instructor-level control variables (i.e., *Degree*, *Professional*, and *Social*) and the residuals were then tested for their 'fit' to a lognormal distribution. Of the 52 domains for which residuals were computed, the lognormal distribution was the dominant fit for five domains. Moreover, it was identified as the co-dominant fit for another 40 domains. For these 45 domains with a lognormal distribution of residuals, the scale (μ range: 6.41 to 10.71) and shape (σ range: 0.54 to 1.81) parameters were estimated.

Table A4.3 shows the pairwise Spearman correlations and corresponding confidence intervals based on the lognormal parameters that fit the residuals. We found a positive correlation ($r = 0.42$ [0.14, 0.64], $p = 0.004$) between the knowledge intensity of a domain and the lognormal scale parameter (μ) as a measure of tail extremity for that domain, indicating support for Hypothesis 3 for data collected in the second round.

Table A4.3
Correlation analysis for Hypothesis 3 (time T2).

	Lognormal shape (σ) parameter	Number of instructors	Knowledge intensity
Lognormal scale (μ) parameter	−0.270† [−0.522, −0.026]	0.012 [−0.282, 0.305]	0.419** [0.144, 0.635]
Lognormal shape (σ) parameter		0.340 [0.051, 0.576]	−0.112 [−0.393, 0.188]
Number of instructors			0.257* [−0.039, 0.512]

Note: n = 45; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001. Numbers in brackets indicate 95 % confidence intervals.

Table A4.4 shows the results from additive nonparametric multiple regression using data for time T2 with the lognormal scale parameter as the dependent variable. Based on an ANOVA test of two models, the first with only the control variables and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.17$; $p = 0.002$) of an association between the knowledge intensity of a domain and the lognormal scale parameter (μ) for the fitted distribution.

Table A4.4
Additive nonparametric regression of lognormal scale parameter (μ).

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1	42	31.821					0.12
s(Lognormal sigma) + *s*(Number of instructors)							
Model 2	41	25.186	1	6.635	10.801	0.002**	0.29
s(Lognormal sigma) + *s*(Number of instructors) + *s*(Knowledge intensity)							

Note: n = 45; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.

Fig. A4.1 shows the component plot for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005); the right-most (third) panel illustrates the positive relationship between knowledge intensity and the lognormal scale parameter. These results are consistent with Hypothesis 3, for data collected at time T2 and performance measured using the cumulative number of students per instructor.

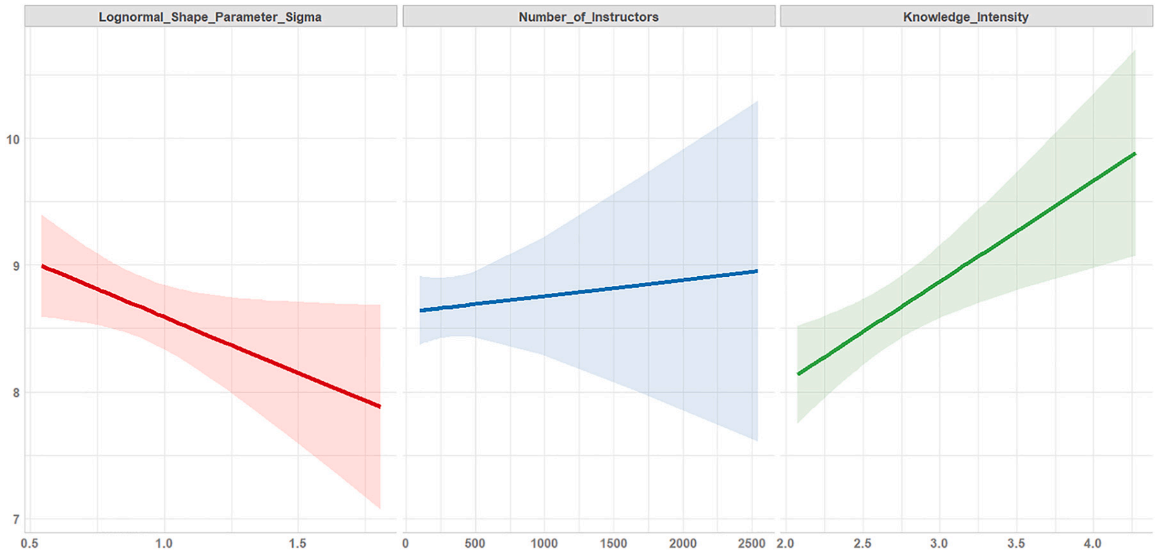

Fig. A4.1. Component plot of (a) lognormal shape parameter, (b) number of instructors, and (c) knowledge intensity with smoothing functions from additive nonparametric regression.
(DV = Lognormal parameter; Time T2).

4.4. *Hypothesis 3 (kurtosis as measure of tail extremity)*

Using pairwise Spearman correlations and corresponding confidence intervals based on kurtosis as the alternate measure of tail extremity (Table A4.5), we found a positive correlation ($r = 0.56$ [$CI_{95\ \%}$: 0.32, 0.74], $p = 5.8e\text{-}5$) between the knowledge intensity of a domain and the kurtosis of the residuals for that domain. These results for data at time T2 are consistent with Hypothesis 3.

Table A4.5
Correlation analysis for Hypothesis 3 (Time T2).

	Number of instructors	Knowledge intensity
Kurtosis	0.512***	0.562***
	[0.257, 0.700]	[0.322, 0.735]
Number of instructors		0.304*
		[0.012, 0.549]

Note: n = 45; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001. Numbers in brackets indicate 95 % confidence intervals.

Table A4.6 shows the results from additive nonparametric multiple regression for testing Hypothesis 3 using data for time T2 and kurtosis as the alternate measure of tail extremity. Based on an ANOVA test of two models, the first with only the control variable and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.07$; $p = 0.004$) of a statistically significant association between the knowledge intensity of a domain and the kurtosis for the fitted distribution.

Table A4.6
Additive nonparametric regression of kurtosis.

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1	36.011	51,770					0.67
s(Number of instructors)							
Model 2	34.425	38,689	1.586	13,082	7.476	0.004**	0.74
s(Number of instructors) + *s*(Knowledge intensity)							

Note: n = 45; $^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Fig. A4.2 shows the component plot for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005); the right panel illustrates the positive relationship between knowledge intensity and kurtosis. These results are consistent with Hypothesis 3 for data collected at time T2 and performance measured using the cumulative number of students per instructor.

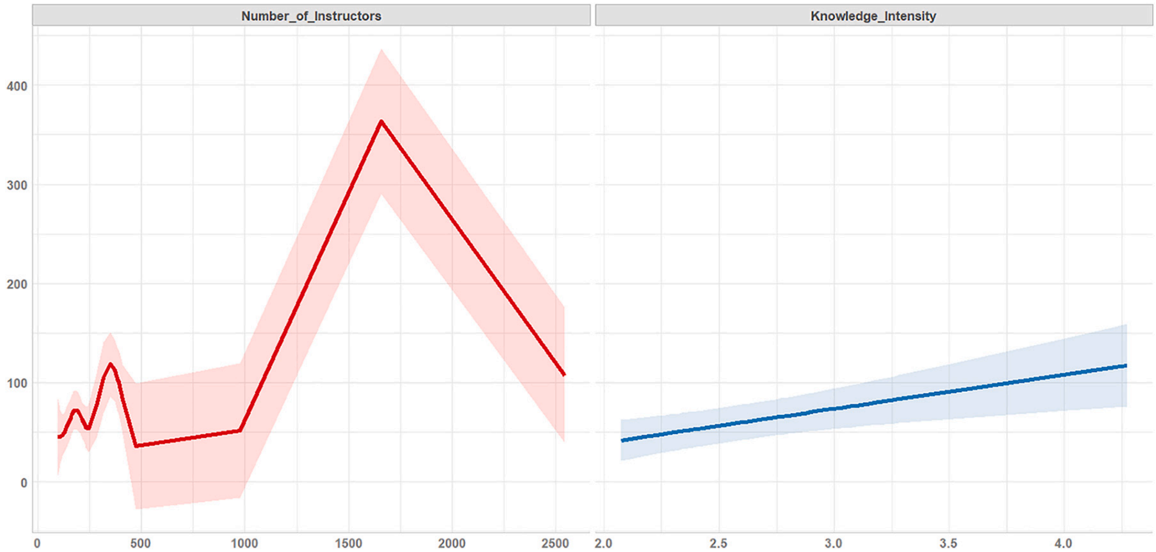

Fig. A4.2. Component plot of (a) number of instructors, and (b) knowledge intensity with smoothing functions from additive nonparametric regression.
(DV = Kurtosis; Time T2).

Appendix 5. Analysis and results of hypothesis testing with reviews as measure of performance (using data for time T1)

5.1. Hypothesis 1

Using the cumulative number of reviews per instructor as an alternate measure of performance, the non-normality of the performance distribution across all domains was statistically tested using the Kolmogorov-Smirnov (K-S) test (D = 0.45; p = 2.2e-16) and the Anderson-Darling (A-D) test (A = 3950.5; p = 2.2e-16) with the result that the null hypothesis of a normal distribution was rejected in each case. Moreover, the p-values for the K-S and A-D tests for each of the 52 domains indicated evidence (p-value = 2.2e-16) against the null hypothesis of a normal distribution of performance within each domain. Fig. A5.1 shows a histogram with the number of instructors plotted against the number of reviews per instructor.

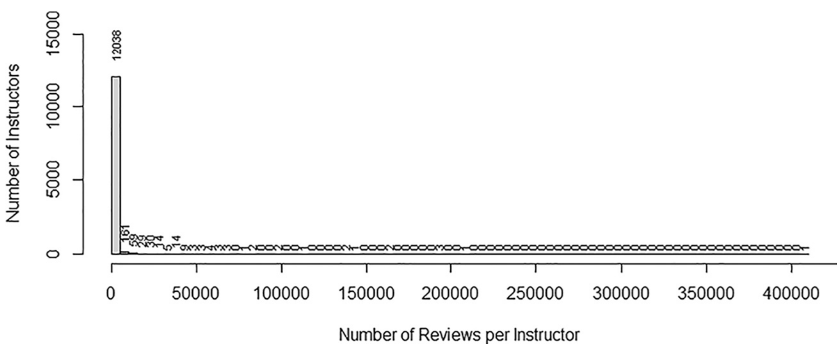

Fig. A5.1. Histogram of performance (number of reviews) across domains.

In addition, we plotted histograms for instructor performance in each domain in our sample (‘n’ range: 41 to 2622, where ‘n_i’ is the number of instructors in domain ‘i’). In visual evaluations, each histogram displayed a non-normal, heavy-tailed distribution. Fig. A5.2 shows histograms for domains with the minimum and the maximum number of instructors. In sum, these histograms and statistical tests indicate support for H1 using the alternate measure of performance.

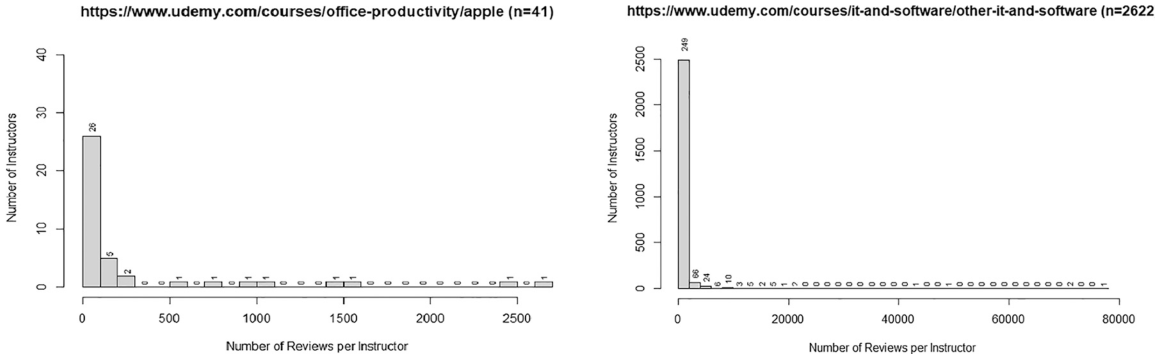

Fig. A5.2. Histogram of performance (number of reviews) for two specific domains.

5.2. *Hypothesis 2*

For the alternate performance measure, H2 was tested as described in Section 4.2. Table A5.1 summarizes the findings for domain-wise tests of distribution fit. The lognormal distribution was the dominant fit for 44 (84 %) of the 52 domains and the co-dominant fit for the remaining 8 domains; it was never dominated by any other type of distribution. These results provide further support for H2 and our observation that, after the Poisson distribution, the normal distribution was the second-worst fit for 47 of the 52 domains offers further support for H1.

Table A5.1
Frequency of performance distributions by domain.

		Domain count	Percentage
Dominant distribution	Exponential tail	0	0
	Lognormal	44	84
	Normal	0	0
	Power Law (PL)	0	0
	PL with cutoff	0	0
	Poisson	0	0
	Weibull	0	0
Co-dominant distributions	Lognormal and Poisson	5	10
	Lognormal, PL with cutoff, and Weibull	2	4
	Lognormal and Weibull	1	2
No dominant or co-dominant distributions		0	0
Total		52	100

5.3. *Hypothesis 3 (lognormal scale parameter as measure of tail extremity)*

Hypothesis 3 was tested using the number of reviews as an alternate measure of entrepreneurial performance in a manner similar to that described in Section 4.3. After regressing the outcome variable on corresponding instructor-level control variables, the residuals were found to fit the lognormal distribution for 47 of 52 domains.

Using pairwise Spearman correlations and corresponding confidence intervals based on the lognormal parameters that ‘fit’ the residuals (Table A5.2), we found a positive association ($r = 0.54$ [$CI_{95\ \%}$: 0.30, 0.72], $p = 0.0001$) between the knowledge intensity of a domain and the lognormal scale parameter (μ) for that domain, when performance was measured using the cumulative number of reviews per instructor. This finding was consistent with Hypothesis 3.

Table A5.2
Correlation analysis for Hypothesis 3.

	Lognormal shape (σ) parameter	Number of instructors	Knowledge intensity
Lognormal scale (μ) parameter	−0.432** [−0.640, −0.166]	0.055 [−0.236, 0.337]	0.544*** [0.304, 0.719]

(*continued on next page*)

Table A5.2 (*continued*)

	Lognormal shape (σ) parameter	Number of instructors	Knowledge intensity
Lognormal shape (σ) parameter		−0.083 [−0.361, 0.210]	−0.044 [−0.327, 0.247]
Number of instructors			0.059 [−0.232, 0.340]

Note: n = 47; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001. Numbers in brackets indicate 95 % confidence intervals.

Table A5.3 shows the results from additive nonparametric multiple regression for testing Hypothesis 3 using reviews as the measure of performance and the lognormal scale parameter as the measure of tail extremity. Based on an ANOVA test of two models, the first with only the control variables and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.287$; $p = 5.6e\text{-}5$) of an association between the knowledge intensity of a domain and the lognormal scale parameter (μ) for the fitted distribution.

Table A5.3
Additive nonparametric regression of lognormal scale parameter (μ).

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1 *s*(Lognormal sigma) + *s*(Number of instructors)	43.761	46.164					0.170
Model 2 *s*(Lognormal sigma) + *s*(Number of instructors) + *s*(Knowledge intensity)	41.806	29.100	1.955	17.063	12.668	5.6e-5***	0.457

Note: n = 47; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.

Fig. A5.3 shows the component plot for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005); the right-most (third) panel illustrates the positive relationship between knowledge intensity and the lognormal scale parameter. These results are consistent with Hypothesis 3 with performance measured as the cumulative number of reviews per instructor.

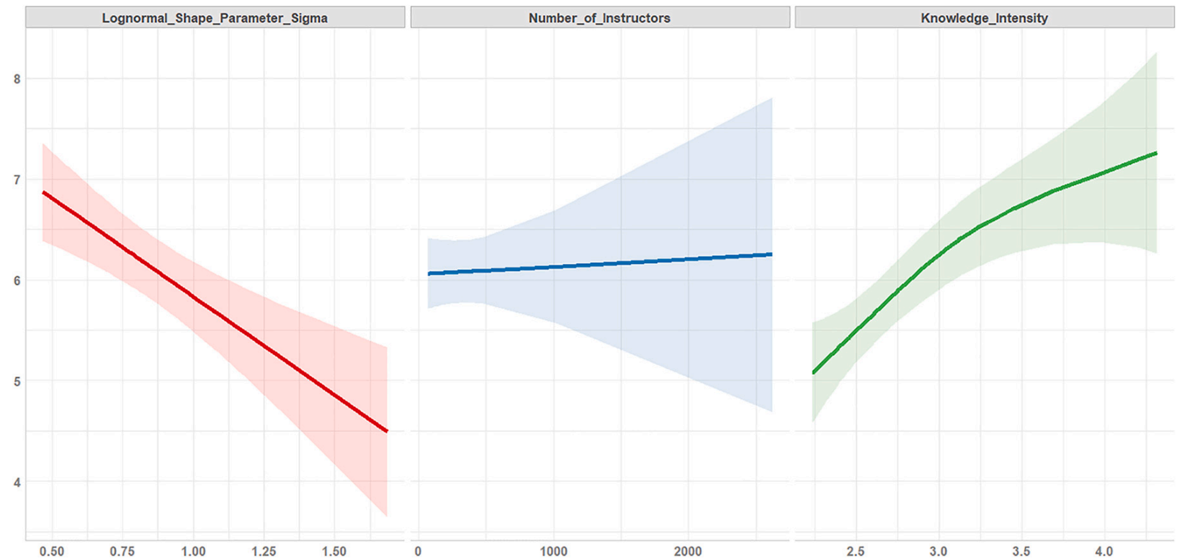

Fig. A5.3. Component plot of (a) lognormal shape parameter, (b) number of instructors, and (c) knowledge intensity with smoothing functions from additive nonparametric regression.
(DV = Lognormal scale parameter; Performance = Reviews).

5.4. *Hypothesis 3 (kurtosis as measure of tail extremity)*

Using pairwise Spearman correlations and corresponding confidence intervals based on kurtosis as the measure of tail extremity (Table A5.4), the correlation of 0.19 ([$CI_{95\,\%}$: −0.11, 0.45], $p = 0.211$) between the knowledge intensity of a domain and the kurtosis of the residuals as the *alternate* measure of tail extremity for that domain was *not* statistically significant.

Table A5.4
Correlation analysis for Hypothesis 3.

	Number of instructors	Knowledge intensity
Kurtosis	0.630*** [0.418, 0.777]	0.186 [−0.107, 0.449]
Number of instructors		0.059 [−0.232, 0.340]

Note: n = 47; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001. Numbers in brackets indicate 95 % confidence intervals.

Table A5.5 shows the results from additive nonparametric multiple regression for testing Hypothesis 3 using reviews as the measure of performance and kurtosis as the alternate measure of tail extremity. Based on an ANOVA test of two models, the first with only the control variable and the second with the addition of knowledge intensity, we find evidence ($\Delta R^2 = 0.03$; $p = 0.065$) of a *marginally* significant association between the knowledge intensity of a domain and the lognormal scale parameter (μ) for the fitted distribution.

Table A5.5
Additive nonparametric regression of kurtosis.

Model	Resid. Df	Resid. dev	Df	Deviance	F	Pr (>F)	Adj. R^2
Model 1 s(Number of instructors)	41.806	108,975					0.50
Model 2 s(Number of instructors) + s(Knowledge intensity)	40.767	100,008	1.040	8967	3.561	0.065†	0.53

Note: n = 47; †p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.

Fig. A5.4 shows the component plot for each independent variable with the corresponding partial regression line and a 95 % confidence envelope (Fox, 2005); the right panel illustrates the positive relationship between knowledge intensity and kurtosis.

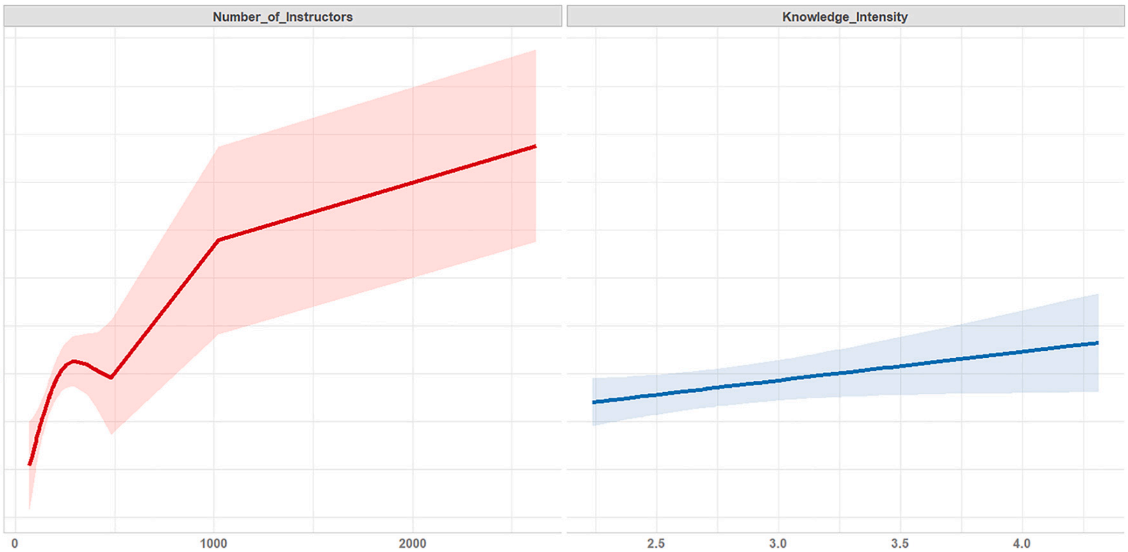

Fig. A5.4. Component plot of (a) number of instructors, and (b) knowledge intensity with smoothing functions from additive nonparametric regression.
(DV = Kurtosis; Performance = Number of Reviews).

5.5. *Hypothesis 4a and 4b*

Hypothesis 4a and 4b were tested using the number of reviews as an alternate measure of performance in a manner similar to that described in Section 4.4. First, the initial value of performance for each domain, measured as the number of reviews at time T1, was treated as predictor (X), while performance at time T2 was treated as the outcome variable (Y). Somers' D for this pair was computed using the R package *DescTools*. Second, the accumulation rate for each instructor, computed as the increase in reviews from T1 to T2, was treated as predictor (X), while performance at time T2 was treated as the outcome variable (Y). Again, the corresponding Somers' D was computed. The Spearman rank correlation between *initial value* and *accumulation rate* ranged from 0.47 to 0.84 across the 52 domains. As summarized in Table A5.6, Somers' D estimates indicated a positive association between the *initial value* (performance at time T1) and the *outcome* (performance at time T2) for each of the 52 domains; the corresponding 95 % confidence intervals did not include zero.

Table A5.6
Somers' D measure of association (performance = number of reviews).

Somers' D	All domains (n = 52)	Domains where lognormal dominant (n = 44)	Domains where lognormal co-dominant (n = 8)
Initial Value, Outcome pair			
Average	0.95	0.95	0.94
[Lowest LCI, Highest UCI]	[0.86, 0.99]	[0.87, 0.99]	[0.86, 0.99]
Accumulation Rate, Outcome pair			
Average	0.57	0.56	0.60
[Lowest LCI, Highest UCI]	[0.34, 0.81]	[0.34, 0.71]	[0.38, 0.81]

Note: CI – 95 % Confidence Interval LCI – Lower Confidence Interval UCI – Upper Confidence Interval.

Moreover, the average Somers' D was 0.95, providing evidence against incremental differentiation (Joo et al., 2017; Bradley and Aguinis, 2023). This finding is consistent with H4a. Somers' D estimates for the relationship between the *accumulation rate* (increase in performance from time T1 to T2) and *outcome* (performance at time T2) were also positive for each of the 52 domains, and the corresponding 95 % confidence intervals did not include zero. This finding is consistent with H4b. Table A5.6 also shows that the Somers' D estimates remain the same whether the lognormal distribution is solely dominant (n = 44) or co-dominant (n = 8) with other distributions.

Together, these Somers' D estimates, by supporting both H4a and H4b, indicate evidence of *proportional differentiation* as the generative mechanism for the lognormal distribution of entrepreneurial performance, when measured as the cumulative number of reviews per instructor.

References

Adner, R., Puranam, P., Zhu, F., 2019. What is different about digital strategy? From quantitative to qualitative change. Strat. Sci. 4 (4), 253–261.
Aguinis, H., O'Boyle Jr., E., 2014. Star performers in twenty-first century organizations. Pers. Psychol. 67 (2), 313–350.
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., Gonzalez-Mulé, E., Joo, H., 2016. Cumulative advantage: conductors and insulators of heavy-tailed productivity distributions and productivity stars. Pers. Psychol. 69 (1), 3–66.
Aguinis, H., Ji, Y.H., Joo, H., 2018. Gender productivity gap among star performers in STEM and other scientific fields. J. Appl. Psychol. 103 (12), 1283.
Aguinis, H., Hill, N.S., Bailey, J.R., 2021. Best practices in data collection and preparation: recommendations for reviewers, editors, and authors. Organ. Res. Methods 24 (4), 678–693.
Aitchison, J., Brown, J., 1957. The Lognormal Distribution: With Special Reference to its Uses in Economics. Cambridge University Press.
Aljukhadar, M., Senecal, S., Daoust, C.E., 2012. Using recommendation agents to cope with information overload. Int. J. Electron. Commer. 17 (2), 41–70.
Anderson, B.S., Wennberg, K., McMullen, J.S., 2019. Enhancing quantitative theory-testing entrepreneurship research. J. Bus. Ventur. 34 (5), 105928.
Andriani, P., McKelvey, B., 2007. Beyond Gaussian averages: redirecting international business and management research toward extreme events and power laws. J. Int. Bus. Stud. 38 (7), 1212–1230.
Andriani, P., McKelvey, B., 2009. Perspective—from Gaussian to Paretian thinking: causes and implications of power laws in organizations. Organ. Sci. 20 (6), 1053–1071.
Andriani, P., McKelvey, B., 2011. From skew distributions to power-law science. In: The SAGE Handbook of Complexity and Management, pp. 254–273.
Archak, N., Ghose, A., Ipeirotis, P.G., 2011. Deriving the pricing power of product features by mining consumer reviews. Manag. Sci. 57 (8), 1485–1509.
Arthur, W.B., 1994. Increasing Returns and Path Dependence in the Economy. University of Michigan Press.
Arthur, W.B., 2009. The Nature of Technology: What it Is and How it Evolves. Simon and Schuster.
Asgari, E., Hunt, R.A., 2016. Crowding out effects of star performers in corporate entrepreneurship. Front. Entrepr. Res. 36 (13), 3.
Asgari, E., Hunt, R.A., Lerner, D.A., Townsend, D.M., Hayward, M.L., Kiefer, K., 2021. Red giants or black holes? The antecedent conditions and multilevel impacts of star performers. Acad. Manag. Ann. 15 (1), 223–265.
Ashman, R., Patterson, A., Brown, S., 2018. 'Don't forget to like, share and subscribe': digital autopreneurs in a neoliberal world. J. Bus. Res. 92, 474–483.
Autor, D., Dorn, D., Katz, L.F., Patterson, C., Van Reenen, J., 2020. The fall of the labor share and the rise of superstar firms. Q. J. Econ. 135 (2), 645–709.
Axtell, R.L., 2001. Zipf distribution of U.S. firm sizes. Science 293 (5536), 1818–1820.
Bakos, Y., Brynjolfsson, E., 1999. Bundling information goods: pricing, profits, and efficiency. Manag. Sci. 45 (12), 1613–1630.
Banerjee, A., Yakovenko, V.M., 2010. Universal patterns of inequality. New J. Phys. 12 (7), 075032.
Barabási, A.L., 2012. Luck or reason. Nature 489 (7417), 507–508.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.
Beamish, P.W., Hasse, V.C., 2022. The importance of rare events and other outliers in global strategy research. Glob. Strateg. J. 12 (4), 697–713.
Bearson, D., Kenney, M., Zysman, J., 2021. Measuring the impacts of labor in the platform economy: new work created, old work reorganized, and value creation reconfigured. Ind. Corp. Chang. 30 (3), 536–563.
Becker, T.E., Robertson, M.M., Vandenberg, R.J., 2019. Nonlinear transformations in organizational research: possible problems and potential solutions. Organ. Res. Methods 22 (4), 831–866.
Bee, M., Riccaboni, M., Schiavo, S., 2017. Where Gibrat meets Zipf: scale and scope of French firms. Physica A 481, 265–275.
Belleflamme, P., Peitz, M., 2018. Inside the engine room of digital platforms: reviews, ratings, and recommendations. In: Working Papers, Aix-Marseille School of Economics, WP 2018 – Nr. 06.
Benzell, S.G., Brynjolfsson, E., 2019. Digital Abundance and Scarce Genius: Implications for Wages, Interest Rates, and Growth (No. w25585). National Bureau of Economic Research.
Birkinshaw, J., 2018. How is technological change affecting the nature of the corporation? J. Br. Acad. 6(s1, 185–214.
Boisot, M., McKelvey, B., 2010. Integrating modernist and postmodernist perspectives on organizations: a complexity science bridge. Acad. Manag. Rev. 35 (3), 415–433.
Bolton, G., Greiner, B., Ockenfels, A., 2013. Engineering trust: reciprocity in the production of reputation information. Manag. Sci. 59 (2), 265–285.
Borghans, L., Groot, L., 1998. Superstardom and monopolistic power: why media stars earn more than their marginal contribution to welfare. J. Inst. Theor. Econ. 546–571.

Bradley, K.J., Aguinis, H., 2023. Team performance: nature and antecedents of nonnormal distributions. Organ. Sci. 34 (3), 987–1352.

Business Insider, 2018. Meet the entrepreneur who has made $1 million by selling clothes with a popular app. https://www.businessinsider.com/poshmark-seller-makes-1-million-dollars-2018-6.

Business Insider, 2022. My husband and I made $200,000 in revenue last year on Poshmark. https://www.businessinsider.com/i-made-200000-last-year-selling-shoes-poshmark-with-husband-2022-7.

Calvo, A.G., Kenney, M., Zysman, J., 2022. Understanding Work in the Online Platform Economy: A Critical Review. University of California, Berkeley Roundtable on the International Economy, Berkeley.

Campbell, D.J., 1988. Task complexity: a review and analysis. Acad. Manag. Rev. 13 (1), 40–52.

Celikoglu, A., Tirnakli, U., 2018. Skewness and kurtosis analysis for non-Gaussian distributions. Physica A 499, 325–334.

Charras-Garrido, M., Lezaud, P., 2013. Extreme value analysis: an introduction. J. Soc. Fr. Stat. 154 (2), 66–97.

Chen, L., Tong, T.W., Tang, S., Han, N., 2022. Governance and design of digital platforms: a review and future research directions on a meta-organization. J. Manag. 48 (1), 147–184.

Chesher, A., 1979. Testing the law of proportionate effect. J. Ind. Econ. 403–411.

Cioffi, J.W., Kenney, M.F., Zysman, J., 2022. Platform power and regulatory politics: Polanyi for the twenty-first century. New Polit. Econ. 27 (5), 820–836.

ClassCentral, 2021. 157K Courses, 425M enrollments: Breaking down Udemy's massive catalog. https://www.classcentral.com/report/udemy-by-the-numbers/.

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

Coelho, F., Augusto, M., 2010. Job characteristics and the creativity of frontline service employees. J. Serv. Res. 13 (4), 426–438.

Coursera, 2023. Coursera|degrees, certificates, & free online courses. https://www.coursera.org/.

Cozzolino, A., Corbo, L., Aversa, P., 2021. Digital platform-based ecosystems: the evolution of collaboration and competition between incumbent producers and entrant platforms. J. Bus. Res. 126, 385–400.

Crawford, G.C., 2018. Skewed Opportunities: How the Distribution of Entrepreneurial Inputs and Outcomes Reconceptualizes a Research Domain. Proceedings-Academy of Management.

Crawford, G.C., Kreiser, P.M., 2015. Corporate entrepreneurship strategy: extending the integrative framework through the lens of complexity science. Small Bus. Econ. 45 (2), 403–423.

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., 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., Skorodziyevskiy, V., Frid, C.J., Nelson, T.E., Booyavi, Z., Hechavarria, D.M., Teymourian, E., 2022. Advancing entrepreneurship theory through replication: a case study on contemporary methodological challenges, future best practices, and an entreaty for communality. Enterp. Theory Pract. 46 (3), 779–799.

Cusumano, M., 2022. The evolution of research on industry platforms. Acad. Manag. Discov,. 8 (1), 7–14.

Cusumano, M.A., Gawer, A., Yoffie, D.B., 2019. The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power. Harper Business.

Cutolo, D., Kenney, M., 2021. Platform-dependent entrepreneurs: power asymmetries, risks, and strategies in the platform economy. Acad. Manag. Perspect. 35 (4), 584–605.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-Fit Techniques. Marcel Dekker.

De Winter, J.C., Gosling, S.D., Potter, J., 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol. Methods 21 (3), 273.

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

Delmar, F., Wallin, J., Nofal, A.M., 2022. Modeling new-firm growth and survival with panel data using event magnitude regression. J. Bus. Ventur. 37 (5), 106245.

Denrell, J., Liu, C., 2012. Top performers are not the most impressive when extreme performance indicates unreliability. Proc. Natl. Acad. Sci. 109 (24), 9331–9336.

Derave, T., Prince Sales, T., Gailly, F., Poels, G., 2021. Comparing digital platform types in the platform economy. In: International Conference on Advanced Information Systems Engineering. Springer, Cham, pp. 417–431.

DiPrete, T.A., Eirich, G.M., 2006. Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments. Annu. Rev. Sociol. 271–297.

Donaker, G., Kim, H., Luca, M., Weber, M., 2019. Designing better online review systems. Harv. Bus. Rev. 97 (6), 122–129.

Eckhardt, J.T., Ciuchta, M.P., Carpenter, M., 2018. Open innovation, information, and entrepreneurship within platform ecosystems. Strateg. Entrep. J. 12 (3), 369–391.

Engmann, S., Cousineau, D., 2011. Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. J. Appl. Quant. Methods 6 (3), 1–17.

Etsy, 2022. Star seller. https://www.etsy.com/starseller.

Etsy, 2023. Investor presentation March 2023. https://s22.q4cdn.com/941741262/files/doc_financials/2023/_etsy-elevator-pitch-march-2023-(1).pdf.

EtsyStrike, 2022. Etsy strike. https://etsystrike.org/.

Evans, P.C., Gawer, A., 2016. The Rise of the Platform Enterprise: A Global Survey. University of Surrey.

Fan, T., Schwab, A., Geng, X., 2021. Habitual entrepreneurship in digital platform ecosystems: a time-contingent model of learning from prior software project experiences. J. Bus. Ventur. 36 (5), 106140.

Floetgen, R.J., Strauss, J., Weking, J., Hein, A., Urmetzer, F., Böhm, M., Krcmar, H., 2021. Introducing platform ecosystem resilience: leveraging mobility platforms and their ecosystems for the new normal during COVID-19. Eur. J. Inf. Syst. 30 (3), 304–321.

Forbes, 2022. The global 2000. https://www.forbes.com/global2000.

Fossen, F.M., Sorgner, A., 2021. Digitalization of work and entry into entrepreneurship. J. Bus. Res. 125, 548–563.

Fox, J., 2002. Nonparametric regression. In: Appendix to: An R and S-PLUS Companion to Applied Regression, pp. 1–7.

Fox, J., 2005. Nonparametric Regression. In: Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons.

Franck, E., Nüesch, S., 2012. Talent and/or popularity: what does it take to be a superstar? Econ. Inq. 50 (1), 202–216.

Frenkel, S., Korczynski, M., Donoghue, L., Shire, K., 1995. Re-constituting work: trends towards knowledge work and info-normative control. Work Employ. Soc. 9 (4), 773–796.

Gel, Y.R., 2010. Test of fit for a Laplace distribution against heavier tailed alternatives. Comput. Stat. Data Anal. 54 (4), 958–965.

Gibbert, M., Nair, L.B., Weiss, M., Hoegl, M., 2021. Using outliers for theory building. Organ. Res. Methods 24 (1), 172–181.

Gillespie, C.S., 2014. A Complete Data Frame Work for Fitting Power Law Distributions. arXiv Preprint arXiv:1408.1554.

Giustiziero, G., Kretschmer, T., Somaya, D., Wu, B., 2023. Hyperspecialization and hyperscaling: a resource-based theory of the digital firm. Strateg. Manag. 44 (6), 1391–1424.

Gokkaya, B.U., Wai, L., 2017. Using probabilistic tag modeling to improve recommendations. In: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Halifax, Nova Scotia – Canada, August 13–17, 2017, 7.

Goldfarb, A., Tucker, C., 2019. Digital economics. J. Econ. Lit. 57 (1), 3–43.

Goldman Sachs, 2023. The creator economy could approach half-a-trillion dollars by 2027. https://www.goldmansachs.com/intelligence/pages/the-creator-economy-could-approach-half-a-trillion-dollars-by-2027.html.

Guinez-Cabrera, N., Aqueveque, C., 2022. Entrepreneurial influencers and influential entrepreneurs: two sides of the same coin. Int. J. Entrep. Behav. Res. 28 (1), 231–254.

Gumbel, E.J., 1958. Statistics of Extremes. Columbia University Press.

Hackman, J.R., Oldham, G.R., 1976. Motivation through the design of work: test of a theory. Organ. Behav. Hum. Perform. 16 (2), 250–279.
Hagiu, A., 2006. Pricing and commitment by two-sided platforms. RAND J. Econ. 37 (3), 720–737.
Hagiu, A., Rothman, S., 2016. Network effects aren't enough. Harv. Bus. Rev. 94 (4), 17.
Hahnbeck, 2021. Number of $10M+ US Amazon sellers increased by 40% in the last 12 months. https://www.hahnbeck.com/blog/2021/10/20/number-of-10-million-amazon-sellers.
Hambrick, D.C., Quigley, T.J., 2014. Toward more accurate contextualization of the CEO effect on firm performance. Strateg. Manag. J. 35 (4), 473–491.
Hernandez, I.F.R., 2019. Exploring Log-Normal Distributions in Nascent Entrepreneurship Outcomes: International Comparisons and Agent-Based Modelling (Doctoral dissertation). University of Bath.
Huberman, B.A., Adamic, L.A., 1999. Growth dynamics of the world-wide web. Nature 401 (6749), 131.
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.
Hunter, J.E., Schmidt, F.L., Judiesch, M.K., 1990. Individual differences in output variability as a function of job complexity. J. Appl. Psychol. 75 (1), 28.
Johnson, N.E., Short, J.C., Chandler, J.A., Jordan, S.L., 2022. Introducing the contentpreneur: making the case for research on content creation-based online platforms. J. Bus. Ventur. Insights 18, e00328.
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), 1022–1053.
Kenney, M., Zysman, J., 2016. The rise of the platform economy. Issues Sci. Technol. 32 (3), 61.
Kenney, M., Bearson, D., Zysman, J., 2021. The platform economy matures: measuring pervasiveness and exploring power. Soc. Econ. Rev. 19 (4), 1451–1483.
Kordrostami, E., Rahmani, V., 2020. Investigating conflicting online review information: evidence from Amazon.com. J. Retail. Consum. Serv. 55, 102–125.
Le Mens, G., Kovács, B., Avrahami, J., Kareev, Y., 2018. How endogenous crowd formation undermines the wisdom of the crowd in online ratings. Psychol. Sci. 29 (9), 1475–1490.
Liebetrau, A.M., 1983. Measures of Association, Vol. 32. Sage.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. BioScience 51 (5), 341–352.
Lord, R.G., Maher, K.J., 1990. Alternative information-processing models and their implications for theory, research, and practice. Acad. Manag. Rev. 15 (1), 9–28.
Luca, M., Zervas, G., 2016. Fake it till you make it: reputation, competition, and yelp review fraud. Manag. Sci. 62 (12), 3412–3427.
Makino, S., Chan, C.M., 2017. Skew and heavy-tail effects on firm performance. Strateg. Manag. J. 38 (8), 1721–1740.
Mankiw, N.G., 2013. Defending the 1 percent. J. Econ. Perspect. 27 (3), 21–34.
Mardon, R., Molesworth, M., Grigore, G., 2018. YouTube beauty gurus and the emotional labour of tribal entrepreneurship. J. Bus. Res. 92, 443–454.
MarketPlacePulse, 2023. Top 100 Etsy sellers. https://www.marketplacepulse.com/top-etsy-sellers.
McKelvey, B., 2004. Toward a complexity science of entrepreneurship. J. Bus. Ventur. 19 (3), 313–341.
McKelvey, B., Andriani, P., 2005. Why Gaussian statistics are mostly wrong for strategic organization. Strateg. Organ. 3 (2), 219–228.
McKelvey, B., Lichtenstein, B.B., Andriani, P., 2012. When organisations and ecosystems interact: Toward a law of requisite fractality in firms. Int. J. Complex. Leadersh. Manag. 2 (1), 104–136.
McMullen, J.S., Brownell, K.M., Adams, J., 2021a. What makes an entrepreneurship study entrepreneurial? Toward a unified theory of entrepreneurial agency. Enterp. Theory Pract. 45 (5), 1197–1238.
McMullen, J.S., Ding, A.W., Li, S., 2021b. From cultural entrepreneurship to economic entrepreneurship in cultural industries: the role of digital serialization. J. Bus. Ventur. 36 (6), 106157.
Menz, M., Kunisch, S., Birkinshaw, J., Collis, D.J., Foss, N.J., Hoskisson, R.E., Prescott, J.E., 2021. Corporate strategy and the theory of the firm in the digital age. J. Manag. Stud. 58 (7), 1695–1720.
Metsämuuronen, J., 2021. Directional nature of Goodman–Kruskal gamma and some consequences: identity of Goodman–Kruskal gamma and Somers delta, and their connection to Jonckheere–Terpstra test statistic. Behaviormetrika 48 (2), 283–307.
Meyer, A.D., Gaba, V., Colwell, K.A., 2005. Organizing far from equilibrium: nonlinear change in organizational fields. Organ. Sci. 16 (5), 456–473.
Mitzenmacher, M., 2004. A brief history of generative models for power law and lognormal distributions. Internet Math. 1 (2), 226–251.
Montoya, J.A., Figueroa, G., Navarro, G., 2017. On the problem of comparing the means and medians of two independent lognormal distributions. J. Data Sci. 16, 631–656.
Morgeson, F.P., Humphrey, S.E., 2006. The work design questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. J. Appl. Psychol. 91 (6), 1321.
Nambisan, S., 2017. Digital entrepreneurship: toward a digital technology perspective of entrepreneurship. Enterp. Theory Pract. 41 (6), 1029–1055.
Nambisan, S., Baron, R.A., 2021. On the costs of digital entrepreneurship: role conflict, stress, and venture performance in digital platform-based ecosystems. J. Bus. Res. 125, 520–532.
Nambisan, S., Luo, Y., 2021. Toward a loose coupling view of digital globalization. J. Int. Bus. Stud. 52 (8), 1646–1663.
Nambisan, S., Siegel, D., Kenney, M., 2018. On open innovation, platforms, and entrepreneurship. Strateg. Entrep. J. 12 (3), 354–368.
Nambisan, S., Wright, M., Feldman, M., 2019. The digital transformation of innovation and entrepreneurship: progress, challenges and key themes. Res. Policy 48 (8), 103773.
Nevo, S., Wade, M.R., 2010. The formation and value of IT-enabled resources: antecedents and consequences of synergistic relationships. MIS Q. 34 (1), 163–183.
Newman, M.E., 2005. Power laws, Pareto distributions and Zipf's law. Contemp. Phys. 46 (5), 323–351.
Newson, R., 2006. Confidence intervals for rank statistics: Somers' D and extensions. Stata J. 6 (3), 309–334.
Nicholson, J.R., 2020. New Digital Economy Estimates. Bureau of Economic Analysis.
NPR, 2021. A pandemic winner: how zoom beat tech giants to dominate video chat. https://www.npr.org/2021/03/19/978393310/a-pandemic-winner-how-zoom-beat-tech-giants-to-dominate-video-chat.
O'Boyle Jr., E., Aguinis, H., 2012. The best and the rest: revisiting the norm of normality of individual performance. Pers. Psychol. 65 (1), 79–119.
Oh, J., Koh, B., Raghunathan, S., 2015. Value appropriation between the platform provider and app developers in mobile platform mediated networks. J. Inf. Technol. 30 (3), 245–259.
Online Courses Galore, 2022. Top Udemy instructors. https://onlinecoursesgalore.com/top-udemy-instructors.
Parker, S.K., Morgeson, F.P., Johns, G., 2017. One hundred years of work design research: looking back and looking forward. J. Appl. Psychol. 102 (3), 403.
Paul, J., Alhassan, I., Binsaif, N., Singh, P., 2023. Digital entrepreneurship research: a systematic review. J. Bus. Res. 156, 113507.
Peiró, J.M., Bayona, J.A., Caballer, A., Di Fabio, A., 2020. Importance of work characteristics affects job performance: the mediating role of individual dispositions on the work design-performance relationships. Personal. Individ. Differ. 157 (109808), 1–11.
Poshmark, 2020. Form S-1 registration statement. https://www.sec.gov/Archives/edgar/data/1825480/000119312520320132/d66583ds1.htm.
R: MGCV, 2022. R documentation for mixed GAM computation vehicle with automatic smoothness estimation package. https://cran.r-project.org/web/packages/mgcv/mgcv.pdf.
R: Somers'' Delta, 2022. R documentation for DescTools package. https://search.r-project.org/CRAN/refmans/DescTools/html/SomersDelta.html.
Ramani, V., Ghosh, D., Sodhi, M.S., 2022. Understanding systemic disruption from the Covid-19-induced semiconductor shortage for the auto industry. Omega 113, 102720.
Reed, W.J., Jorgensen, M., 2004. The double Pareto-lognormal distribution—a new parametric model for size distributions. Commun. Stat. Theory Methods 33 (8), 1733–1753.
Rietveld, J., Schilling, M.A., 2021. Platform competition: a systematic and interdisciplinary review of the literature. J. Manag. 47 (6), 1528–1563.
Rifkin, J., 2014. The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism. St. Martin's Press.
Rigney, D., 2010. The Matthew Effect: How Advantage Begets Further Advantage. Columbia University Press.

Rönkkö, M., Aalto, E., Tenhunen, H., Aguirre-Urreta, M.I., 2022. Eight simple guidelines for improved understanding of transformations and nonlinear effects. Organ. Res. Methods 25 (1), 48–87.
Schindehutte, M., Morris, M.H., 2009. Advancing strategic entrepreneurship research: the role of complexity science in shifting the paradigm. Enterp. Theory Pract. 33 (1), 241–276.
Selden, P.D., Fletcher, D.E., 2015. The entrepreneurial journey as an emergent hierarchical system of artifact-creating processes. J. Bus. Ventur. 30 (4), 603–615.
Shane, S., 2003. A General Theory of Entrepreneurship: The Individual-Opportunity Nexus. Edward Elgar.
Shane, S., 2009. Why encouraging more people to become entrepreneurs is bad public policy. Small Bus. Econ. 33 (2), 141–149.
Shane, S., Venkataraman, S., 2000. The promise of entrepreneurship as a field of research. Acad. Manag. Rev. 25 (1), 217–226.
Shapiro, C., Varian, H.R., Carl, S., 1998. Information Rules: A Strategic Guide to the Network Economy. Harvard Business Press.
Shepherd, D.A., Wennberg, K., Suddaby, R., Wiklund, J., 2019. What are we explaining? A review and agenda on initiating, engaging, performing, and contextualizing entrepreneurship. J. Manag. 45 (1), 159–196.
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.A., Bonini, C.P., 1958. The size distribution of business firms. Am. Econ. Rev. 48 (4), 607–617.
Somers, R.H., 1962. A new asymmetric measure of association for ordinal variables. Am. Sociol. Rev. (27), 799–811.
Srinivasan, A., Venkatraman, N., 2018. Entrepreneurship in digital platforms: a network-centric view. Strateg. Entrep. J. 12 (1), 54–71.
Stallkamp, M., Hunt, R.A., Schotter, A.P., 2022. Scaling, fast and slow: the internationalization of digital ventures. J. Bus. Res. 146, 95–106.
Stone, C.J., 1985. Additive regression and other nonparametric models. Ann. Stat. 13 (2), 689–705.
Stummer, C., Kundisch, D., Decker, R., 2018. Platform launch strategies. Bus. Inf. Syst. Eng. 60, 167–173.
Tavalaei, M.M., Cennamo, C., 2021. In search of complementarities within and across platform ecosystems: complementors' relative standing and performance in mobile apps ecosystems. Long Range Plan. 54 (5), 101994.
The App Brewery, 2023. Programming bootcamp and courses. https://www.appbrewery.co/.
Udemy, 2021. Investor presentation Q4 2021. https://investors.udemy.com/static-files/070b3028-2703-4240-9ab7-83440266c8a6.
Udemy, 2023a. Online courses – learn anything, on your schedule | Udemy. https://udemy.com.
Udemy, 2023b. Dr. Angela Yu, developer and lead instructor, The App Brewery. https://www.udemy.com/user/4b4368a3-b5c8-4529-aa65-2056ec31f37e/.
Udemy, 2023c. Chris Dutton, Founder, Maven Analytics. https://www.udemy.com/user/chrisdutton3/.
Udemy, 2023d. Jose Portilla, Head of Data Science, Pierian Training. https://www.udemy.com/user/joseportilla/.
Udemy, 2023e. John Pauler, Analytics Expert. https://www.udemy.com/user/john-pauler/.
Udemy, 2023f. Leaving and editing a course review. https://business-support.udemy.com/hc/en-us/articles/115005537968-Leaving-and-Editing-a-Course-Review.
Urbano, D., Turro, A., Wright, M., Zahra, S., 2022. Corporate entrepreneurship: a systematic literature review and future research agenda. Small Bus. Econ. 1–25.
Van de Rijt, A., Kang, S.M., Restivo, M., Patil, A., 2014. Field experiments of success-breeds-success dynamics. Proc. Natl. Acad. Sci. 111 (19), 6934–6939.
van Gelderen, M., Wiklund, J., McMullen, J.S., 2021. Entrepreneurship in the future: a Delphi study of ETP and JBV editorial board members. Enterp. Theory Pract. 45 (5), 1239–1275.
Varadarajan, R., Yadav, M.S., Shankar, V., 2008. First-mover advantage in an internet-enabled market environment: conceptual framework and propositions. J. Acad. Mark. Sci. 36 (3), 293–308.
Vial, G., 2019. Understanding digital transformation: a review and a research agenda. J. Strateg. Inf. Syst. 28 (2), 118–144.
Von Briel, F., Davidsson, P., Recker, J., 2018. Digital technologies as external enablers of new venture creation in the IT hardware sector. Enterp. Theory Pract. 42 (1), 47–69.
Wan, Y., 2015. The Matthew effect in social commerce. Electron. Mark. 25 (4), 313–324.
Wennberg, K., Anderson, B.S., 2020. Enhancing the exploration and communication of quantitative entrepreneurship research. J. Bus. Ventur. 35 (3), 105938.
Westfall, P.H., 2014. Kurtosis as peakedness, 1905–2014. RIP. Am. Stat. 68 (3), 191–195.
Wicklin, R., 2019. Critical values of the Kolmogorov-Smirnov test. https://blogs.sas.com/content/iml/2019/05/20/critical-values-kolmogorov-test.html.
Wiltbank, R., Dew, N., Read, S., 2015. Investment and returns in successful entrepreneurial sell-outs. J. Bus. Ventur. Insights 3, 16–23.
World Bank, 2022. Digital platforms multiply market size and employment in services, but only when disciplined by competition. https://blogs.worldbank.org/voices/digital-platforms-multiply-market-size-and-employment-services-only-when-disciplined.
Yadav, M.S., Pavlou, P.A., 2014. Marketing in computer-mediated environments: research synthesis and new directions. J. Mark. 78 (1), 20–40.
Zaheer, H., Breyer, Y., Dumay, J., 2019. Digital entrepreneurship: an interdisciplinary structured literature review and research agenda. Technol. Forecast. Soc. Chang. 148, 119735.
Zhang, M., Wei, X., Zeng, D.D., 2020. A matter of reevaluation: incentivizing users to contribute reviews in online platforms. Decis. Support. Syst. 128, 113158.