INFORMATION SYSTEMS RESEARCH
Vol. 30, No. 2, June 2019, pp. 540–562
ISSN 1047-7047 (print), ISSN 1526-5536 (online)
http://pubsonline.informs.org/journal/isre/

Storm Clouds on the Horizon? New Entry Threats and R&D Investments in the U.S. IT Industry

Yang Pan,[a] Peng Huang,[b] Anandasivam Gopal[b]

[a] E.J. Ourso College of Business, Louisiana State University, Baton Rouge, Louisiana 70803; [b] Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742

Contact: yangpan@lsu.edu, https://orcid.org/0000-0002-1540-1806 (YP); phuang1@rhsmith.umd.edu, https://orcid.org/0000-0002-7872-0396 (PH); agopal@rhsmith.umd.edu, https://orcid.org/0000-0001-9270-5961 (AG)

Received: February 6, 2018
Revised: June 4, 2018
Accepted: July 26, 2018
Published Online in Articles in Advance: June 6, 2019

https://doi.org/10.1287/isre.2018.0816

Abstract. The threats of new entry by startups in the fast-moving information technology (IT) industry have important implications for firm decision making. Although analytical work on the strategic responses to new entry threats (NETs) through preemptive research and development (R&D) has produced contradictory predictions, empirical analysis of this relationship is limited, largely because of the absence of a reasonable measure. In this work, we make two contributions. First, we develop and validate a measure of these threats through text mining using product descriptions provided by incumbent firm 10-K filings and business descriptions provided by startups. This novel measure of NET differs significantly from *observed entry* and *competition*. Second, we study the R&D investment strategies of IT firms facing new entry threats. Using a sample of U.S. IT firms over the period 1997–2013, we show that incumbents on average reduce R&D spending when there are greater threats from the startup space. More importantly, we show that the effect is not uniform—firms that operate in industries with strong network effects or high levels of technological cumulativeness invest relatively more in R&D when they face greater NET. Our work adds to the literature on the relationship between product market threats and firm decision making by expanding the scope of this line of work to include the role of threats of new entry—a central construct in the field of strategy and industrial organization—and by highlighting conditions that influence the effectiveness of preemptive R&D investments as a response to NET in the IT industry.

History: Anindya Ghose, Senior Editor; Prasanna Tambe, Associate Editor.
Supplemental Material: The online appendices are available at https://doi.org/10.1287/isre.2018.0816.

Keywords: new entry threats • text mining • R&D investment • uncertainty • innovation • disruptive technology

1. Introduction

The information technology (IT) industry has played a central and critical role in driving economic growth over the last two decades, especially in the United States (Jorgenson and Stiroh 1999, Jorgenson et al. 2000). This contribution to economic growth is partly a product of the constant technological changes and fast clock speed (Fine 1998) that define the industry. The rate of change in the industry's external environment, including the development of new technologies, shifts in consumer preferences, and fast-moving market dynamics, far exceeds that seen in other industries (Mendelson and Pillai 1999, Brynjolfsson and McAfee 2011). The result is a short product lifecycle (Mendelson and Pillai 1998), great volatility in the market structure, and a hypercompetitive context where advantages, if any, tend to be short lived (Wiggins and Ruefli 2005).

This environmental volatility is, in part, related to the potential disruption brought by *new entry threats* (NETs) often emerging from the entrepreneurial ecosystem, a phenomenon prevalent in most high-tech contexts but of particular importance in the IT industry. Entrepreneurship within the IT sector is particularly prominent, with over 70% of all venture capital (VC)-funded startups associated with the IT industry (Gompers and Lerner 2001). The presence of intense entrepreneurial activity in the product market of an incumbent firm backed by influential venture capitalists can lead to significant turbulence in the future and therefore, warrants strategic responses by incumbents. Indeed, the threats that emerge from entrepreneurial ecosystems are often deemed a significant source of risk and volatility by managers of IT product and services firms.[1] One important strategic response is through preemptive investments in research and development (R&D), because startups usually lead the assault on incumbents with novel and superior technologies. The importance of R&D investments, as an input to the innovation process and a key enabler of competitive advantage within the IT industry (Sambamurthy et al. 2003), has been highlighted in the information systems (IS) literature (King et al. 1994). A long line of analytical work in industrial organization literature has attempted to shed light on how incumbents respond to potential entry through their R&D investments (e.g., Reinganum 1983, Fudenberg and Tirole 1984, Lukach et al. 2007), often providing contradictory predictions. For example, earlier work by Gilbert and Newbery (1982) argues that an incumbent with monopoly power is likely to engage in preemptive invention in its effort to deter entry. In contrast, Reinganum (1983) maintains that, because of the probabilistic nature of the inventive process, an incumbent has a lower incentive to invest in R&D than a potential entrant. Similarly, Fudenberg and Tirole (1984), in their study of preentry strategic investments, conclude that an incumbent has an incentive to underinvest in R&D. More recent work by Lukach et al. (2007) suggests that whether entry threat stimulates or discourages R&D investments is contingent on if the entry can be deterred or is inevitable.

Surprisingly, on the empirical side, there is a scarcity of studies that investigate this relationship, forming the central research question that we aim to address in this paper. Specifically, we ask the following question: In the presence of threats of new entry, how do incumbents in the IT industry respond in terms of their investments in R&D? Addressing this important question will provide insight into how firms in the IT sector choose among a set of strategic countermeasures available to them in the face of potential new entry while also unveiling the conditions under which creative destruction is likely to occur. The lack of prior empirical evidence is primarily owing to the difficulty in measuring new entry threats. As one of the five competitive forces articulated by Porter (2008), NET has been a key theoretical construct in the strategy and industrial organization literatures, but a valid measurement has remained elusive for several reasons. By definition, threats from startups represent *forward-looking* estimations of the extent to which the potential entry of new competition may influence cash flows or product market performance. Therefore, these threats have not manifested as yet but may or may not materialize at some point in the future, leading to their probabilistic nature. Although it is tempting to use measures of *competition* (Hoberg and Phillips 2016) or *observed entry* (Aghion et al. 2009) as surrogates for new entry threats, such approaches are potentially flawed, because these represent fundamentally different constructs pertaining to realized competitive dynamics, whereas the threats of new entry represent forecasts about future competition (Goolsbee and Syverson 2008), and because the source of NET in the IT industry often comes from startups rather than mature competitors. Methodologically, existing measures of competition or realized entry rely heavily on static industry classification codes (Aghion and Howitt 1992, Becker-Blease 2011); however, there is no clear analog of the Standard Industrial Classification/North American Industry Classification System (SIC/NAICS) codes extending into the entrepreneurial ecosystem, where startups tend to straddle existing industry classifications (Sarasvathy 2001). Thus, notwithstanding the importance of new entry threats in theoretical work (Spence 1977), the absence of an established measure represents a significant gap in the literature.

We take on this challenge by developing and validating a novel measure of new entry threats using text analytics. Machine learning techniques are increasingly being used to develop new measurement schemes for strategic and economic concepts that have been difficult to assess and quantify in the past. Intuitively, the measure that we develop is based on the extent to which an incumbent firm's description of its product markets in its 10-K filings overlaps with the business descriptions of new entrepreneurial firms that receive first-round funding from VCs during the same period. To capture the ebbs and flows of venture funding and the startups that receive them, we use business descriptions of startups receiving VC funding from VentureXpert. The focus on very early-stage startups is appropriate here, because they collectively represent movement in the incumbent's product space toward possible future competition. The presence of VC funding makes them credible threats while also signaling quality. Contemporaneously, these startups do not constitute significant competition at this stage, because they lack mature products or an established customer base. Because many of them are still in the phases of product or service design, they are years away from becoming realized entrants into the incumbent's product market. We also conduct a series of validation tests, showing that variations in the measure do relate to increased competition as well as lower profitability in the future. Furthermore, we show that firms facing high NET are likely to experience greater turbulence in the forms of layoffs and bankruptcies in the future, dynamics consistent with prior theory (Becker-Blease 2011).

Theoretically, IT firms facing the threats of new entry may choose to respond in different ways. On the one hand, some may choose to increase their investments in R&D in the hope of establishing advantage in technological and process capabilities to successfully compete in the future if and when the perceived threats do materialize (Reinganum 1983, Lukach et al. 2007). On the other hand, incumbents may choose to avoid R&D spending, where the payoffs are generally uncertain, and choose instead to either conserve their cash (Klepper and Simons 1997, Hoberg et al. 2014) in anticipation of acquiring technology licensing later from pioneers or invest in complementary capabilities, like marketing, distribution, and manufacturing, that can help the incumbent compete more effectively should the anticipated threats materialize (O'Connor and Rafferty 2012).

Finally, incumbents may choose not to respond through their R&D spending, essentially deferring any action until the uncertainty associated with new entry threats is resolved (Hoberg et al. 2014).

These potential variations in strategic moves on the part of incumbents also suggest that the specific actions may depend on idiosyncratic features of firms within the IT industry. We consider two factors that may determine the effectiveness of preemptive R&D as a response to NET. Particularly, prior literature suggests that the timing and valuation of R&D are partly influenced by (1) the strength of network effects (NEs) (Kristiansen 1998), capturing demand-side economies of scale, and (2) the degree of technological cumulativeness (TC) (Oriani and Sobrero 2008) defined by the extent to which control over an earlier stream of innovations is needed for exploiting later ones within an industry segment. The presence of network effects within a subindustry can affect market dynamics by creating winner-take-all scenarios rapidly (Zhu et al. 2006), thereby significantly influencing the payoffs of preemptive responses to NET. Similarly, technological cumulativeness captures the extent to which a firm's technological assets build on prior investments, capturing elements of path dependence (Breschi and Malerba 1997). Here again, incumbent responses to NET will be determined by how important current investments in R&D are in ensuring future competitiveness and averting being locked out. Thus, we examine a second related research question: to what extent is the relationship between new entry threats and R&D spending moderated by idiosyncratic IT industry characteristics, such as the presence of network effects and technological cumulativeness?

We use firm-level data on the U.S. IT industry over the period 1997–2013 to conduct our analyses. Our results show, interestingly, that incumbent IT firms on average reduce their R&D spending when they face high NET, consistent with the reasoning of financial conservatism noted in Hoberg et al. (2014). Furthermore, the results indicate that firms operating in contexts with strong network effects and where technological cumulativeness is important tend to respond to high NETs by investing relatively more in R&D, all else being equal. We show that these results hold even after controlling for competition and realized entry, both derived from the Text-based Network Industry Classification (TNIC) database (Hoberg and Phillips 2016, Kim et al. 2016), thus revealing the incremental effects that new entry threats and the implied future volatility have on incumbent decision making about innovation spending.

Our work provides several contributions to extant IS research. First, we address the gap in the IS literature pertaining to how forecasts of volatility in product markets influence the R&D decisions of incumbent IT firms. Although the broader economics literature has provided much insight on the roles of competition and observed entry, we evaluate the role that new entry threats, emerging from the entrepreneurial ecosystem, play within the IT industry. The rapid rate of technological change and the extent to which entrepreneurship feeds this change within the IT industry (Brynjolfsson and McAfee 2011) make the study of new entry threats from startups particularly relevant relative to other industry sectors. A societal focus on IT-based entrepreneurship and the development of entrepreneurial ecosystems, in the forms of incubators and accelerators, have radically improved the abilities of new ventures to threaten incumbents. In these contexts, it is important to understand how NET may be accurately measured and how incumbents react to these threats. In addition, by studying the moderating role of network effects and technology cumulativeness, we highlight the relevance of idiosyncratic characteristics of the IT industry in shaping the effectiveness of preemptive R&D as a response to NET. Our work here directly contributes to these gaps in the literature and to the broader literature on the drivers of R&D investments within the IT sector (Bardhan et al. 2013), with a particular focus on environmental volatility and competitive dynamics.

We also make a contribution in methodology by devising and validating a new measure of NET using text analysis, with potential domains of application that may extend beyond the IS field. We build on prior work in finance and marketing that has used text analysis to construct measurement schemes that are subsequently applied to studying competitive dynamics (Tetlock et al. 2008, Hoberg and Phillips 2010). Our measure of new entry threat is particularly suitable for capturing emerging threats from the overall entrepreneurial space rather than focusing on individual entrepreneurs who are easy to ignore or discount. We believe that many empirical questions in the industrial organization literature can be addressed through the use of the measure that we invent. In the spirit of encouraging research in related fields, we intend to provide open access to the NET data to the academic community.

2. Theoretical Background

2.1. Product Market Threats and the Threat of New Entry

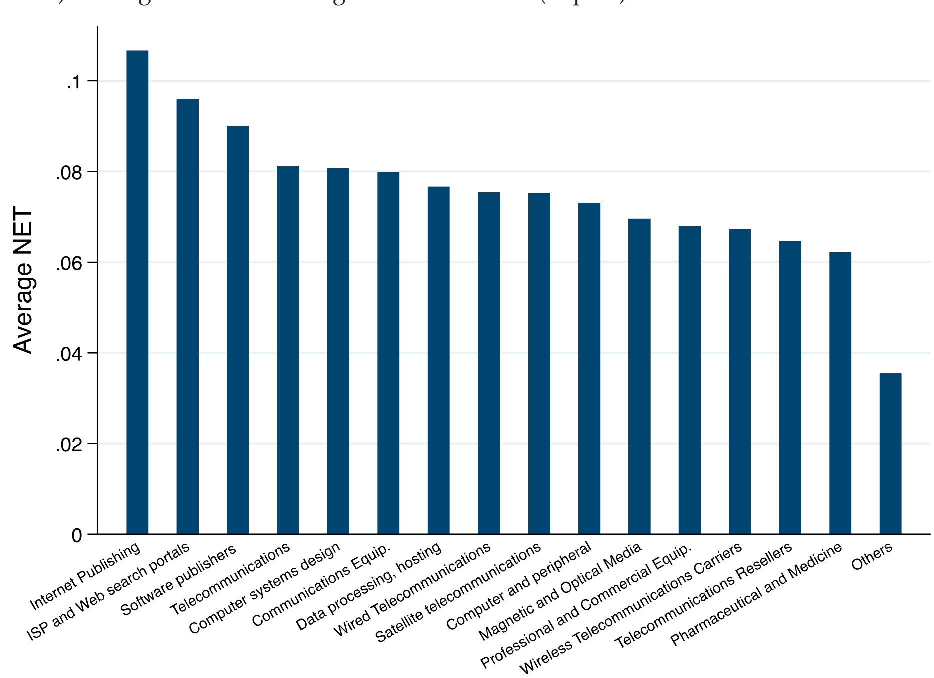
All firms face *product market threats*, defined as sources of instability and uncertainty in a firm's product market, which threaten the sustainability of the firm's earnings as well as the viability of its current product portfolio (Hoberg et al. 2014). Within this broad notion, literature has suggested three important factors—competition, observed entry, and the *threat of new entry*—that shape the competitive dynamics and firms' survival (Porter 2008). Competition typically addresses the extent to which other incumbents exert pressure on the focal firm (Aghion 2003) and is typically determined by the number and relative sizes of other players within the focal firm's product or factor markets, such as those within the same SIC/NAICS codes (Blundell et al. 1999). Observed entry, by contrast, is the extent to which new firms enter the focal firm's market either as a newly created entity or through lateral diversification (Khessina and Carroll 2008).

Interestingly, barring some analytical studies, extant empirical literature sheds little light on the strategic implications of the third source of instability—the threats emerging from the entrepreneurial ecosystem in the form of startups. Technology entrepreneurship has become a prominent phenomenon in the United States (Hsu 2008) owing to the introduction of economic policies to foster such activities (Kuratko and Menter 2017) as well as the development of organizational forms to help such entrepreneurs gain traction, such as incubators (Colombo and Delmastro 2002) and seed accelerators (Cohen and Hochberg 2014). IT startups are at the center of the entrepreneurial ecosystem: the majority of the startups funded by venture capitalists are associated with information technology (Gompers and Lerner 2001), and most ventures in accelerators and incubators also tend to be technology based (Pauwels et al. 2016). Entrepreneurial activities drive much of the disruption for incumbent firms, thus giving rise to significant product market threats in the future. In contrast to competition and realized entry, the new entry threats are, by definition, forward looking, because they imperil the "stability and sustainability of future earnings" (Brav et al. 2005).

Viewed differently, although competition and realized entry describe equilibrium market outcomes, new entry threats emerging from startups are inherently about disequilibrium, often leading to changes in competitive dynamics (Cockburn and MacGarvie 2011) and creative destruction (Aghion and Howitt 1992). This is consistent with a technology lifecycles perspective wherein the volatility emerging from new entry threats is typically associated with the *introduction* stage of a technology's development, when fundamental technological problems are being tackled through radical innovation and design (Utterback 1994). In contrast, realized entry and competition tend to be associated with the *growth* and *maturity* stages in the lifecycle, when a broad range of market applications based on the technology has been developed and uncertainty about the technology is reduced (Haupt et al. 2007). Thus, although competition and realized entry are deterministic and can be observed by the focal firm with relative ease, new entry threats are probabilistic perceptions of future entry (Goolsbee and Syverson 2008), because not all perceived threats will materialize. Furthermore, such threats may emerge from unlikely quarters—developments in seemingly unrelated fields can lead to important applications in the focal market, such as the threats posed by smartphone-based ride-sharing app developers to the providers of taxi service. However, if these threats do materialize, the new products or services introduced by startups can upend existing market structures and disrupt business models significantly (Gompers and Lerner 2001).

Some evidence indicating the relevance of NET in the IT sector is manifested in the statistics based on the NET measure that we develop (described in detail later). For example, in Figure 1, we show the 15 four-digit NAICS industries with the highest levels of new entry threats.

Figure 1. (Color online) Average NET Across High-Tech Industries (Top 15)



Interestingly, with the exception of pharmaceuticals, all other sectors depicted are associated with the IT industries,[2] an observation that is consistent with faster clock speed and hypercompetition within the IT industries in general (Mendelson and Pillai 1998). As prior research suggests, a key strategic response on the part of IT incumbents is through innovation (Sambamurthy et al. 2003, Kim et al. 2016), because new entrants typically lead the attack on incumbents with new and innovative products. However, given the difficulty associated with assessing new entry threats, there is a surprising gap in prior research on how incumbents respond to NET through their innovation spending. Indeed, "the analysis of R&D investments is surely one of the most difficult problems of investment under uncertainty" (Schwartz and Moon 2000).

2.2. New Entry Threats and R&D Investments

The relationship between product market threats and firm spending on innovation has been studied extensively under the context of competition and realized entry, with little empirical work devoted to the role of new entry threats. We first briefly review the literature with respect to competition and realized entry before providing arguments for the effects of new entry threats on innovation investments.

When firms operate in competitive environments, they face uncertainty with respect to how investments in R&D will help improve their market position (Dixit and Stiglitz 1977). Prior literature has presented contrasting views on the link between competition and innovation spending. On the one hand, firms may choose to reduce R&D investments, because the presence of competition reduces any postinnovation rents for the innovator (Grossman and Helpman 1991). Consistent this view, monopoly power is generally conducive to innovation investments given uncertain payoffs (Schumpeter 1942). Firms with monopoly power face lower market uncertainty, are able to secure postinnovation rents and preempt entry, and face less financial constraint (Dasgupta and Stiglitz 1980, Gilbert and Newbery 1982). On the other hand, some firms choose to increase their R&D spending in a competitive environment so as to "escape competition" (Aghion et al. 2005). Under this view, innovation offers the possibility of creating new products that supersede existing ones, thereby allowing the innovator to differentiate themselves (Arrow 1962). The empirical evidence regarding the effects of competition on innovation is mixed. Nickell (1996) and Blundell et al. (1999) show a positive association between competition and innovation, whereas Dixit and Stiglitz (1977) and Tang (2006) report the opposite. In comparison, Aghion et al. (2005) find an inverted U-shaped relationship, which is conditional on a firm's distance to the technological frontier. More recently, Kim et al. (2016) show that, within the IT sector specifically, the relationship is positive because of the critical role of innovation and the limited effectiveness of other countermeasures, such as through marketing and price competition. A short summary of the empirical work is shown in Table A1 in the online appendices outlining the mixed results in general.

With respect to the effects of realized entry, the literature extends the reasoning from above, because realized entry usually leads to greater competitive intensity. In addition, innovation investment decisions are moderated by the heterogeneity in the focal firm's industry in terms of existing competitive intensity before entry, the extent to which intellectual property is critical, and the tradeoff between radical and incremental innovations (Mitchell and Singh 1992, Chandy and Tellis 2000). The empirical evidence here is again mixed. Aghion et al. (2009) show that firms in technologically advanced industries respond to realized entry by increasing their innovation spending, whereas Kuester et al. (1999) show that firms tend to reduce their innovation spending. Geroski (1989) finds that entry and innovation spending seem to go hand in hand with increased firm productivity, implying a positive effect. Table A2 in the online appendices presents a summary of empirical work on the relationship between realized entry and innovation; whereas the direct effect of realized entry is ambiguous, in technologically advanced and competitive industries, firms tend to respond by increasing their innovation investments (Aghion et al. 2009, Kim et al. 2016). These results mirror those observed for the effect of competition.

In the case of new entry threats emerging from the entrepreneurial ecosystem, there are several sources of uncertainty that incumbents face with their investment decisions. First, there is uncertainty regarding whether the observed threats will eventually materialize in later years or peter out without entering the mainstream product markets (Gompers and Lerner 2001). Second, during the introduction stage of a technology lifecycle, there is often a multitude of competing designs and standards, and a dominant design is yet to be established (Jovanovic and MacDonald 1994). At this crucial juncture, endorsing the wrong standard by committing significant R&D investments prematurely can turn out to be fatal (Shapiro and Varian 1999). Third, similar to the contexts of competition and realized entry, there is uncertainty about whether preemptive investments in R&D will be successful in enhancing the firm's competitive position through the introduction of new products and services (Huchzermeier and Loch 2001, Gunther McGrath and Nerkar 2004). In fact, the probability of failure in R&D investments at this point is even greater, because the nature of innovation tends to be radical rather than incremental. In summary, a firm's response to new entry threat faces much greater uncertainty than when it responds to competition and realized entry, leading to a more complex decision process on the part of incumbents.

On the one hand, when IT incumbents sense new entry threats, some may choose to aggressively counter the possibility of entrepreneurs eventually entering their product markets by preemption through R&D spending (Gilbert and Newbery 1982, Lukach et al. 2007) in the spirit similar to preemptive pricing changes (Goolsbee and Syverson 2008). With this approach, the incumbent's strategy entails ramping up on R&D investments, with the belief that any resulting innovations from this spending will help deter potential entrants by making the markets unattractive to them while also helping to contribute new products and services to the market on the margin. In the event that the expected threats do not materialize, these investments in R&D can still help the incumbent defend itself from competition and lateral entry by other incumbents (Aghion et al. 2009, Kim et al. 2016) given the presence of hypercompetition and rapid technological changes in this industry (McAfee and Brynjolfsson 2008).

On the other hand, investments in R&D are generally considered highly risky and may not always deliver value (Huchzermeier and Loch 2001). Therefore, as the threat of new entry from startups increases, risk-averse managers may prefer to invest in countermeasures that have more assured payoffs, such as advertising and marketing (Gatignon et al. 1989), strategic alliances (Gimeno 2004), or specialized manufacturing capabilities (Cohen et al. 2000). These complementary capabilities have been shown to be crucial to the commercialization of new innovations (Teece 1986). Alternatively, they may choose to simply hold on to cash (Hoberg et al. 2014) in preparation for acquiring technology licensing through markets for technology (Arora et al. 2001) should the threats materialize, thereby securing a hedge against the failure of R&D efforts or endorsing the wrong standard. Indeed, because responses to new entry threats have to account for the three sources of uncertainties in R&D as we highlighted above, a sensible strategy may well be forbearance (i.e., to resist taking any obvious action) (Gatignon et al. 1989). Therefore, this reasoning would predict a negative relationship between new entry threats and the firm's R&D spending.

Given these diverging arguments and the lack of prior systematic theoretical treatment in the context of the IT industry, we allow the empirical analyses to provide suitable guidance. Through the introduction of a new measure for new entry threat, we can capture the direct effect of NET on innovation spending while controlling for both competition as well as realized entry.

2.3. Moderating Factors

Prior research examining the effects of product market threats on innovation has attested to the underlying heterogeneity both within and across industries (Cornaggia et al. 2015). The presence of factors, such as the stages of technology evolution, the specific form of uncertainty faced, and the costs associated with deferring investment decisions, can moderate the relationship between such threats and innovation investments (Klepper 1996, Lukach et al. 2007, Czarnitzki and Toole 2011). Even within the context of a single industry, there are variations in the degree to which these threats can disrupt an incumbent's product markets and the extent to which investments in preemptive R&D may be effective. We identify two such factors that may be viewed as boundary conditions (i.e., they shape the extent to which an IT incumbent may choose to respond to NET through elevated R&D spending) as argued below.

2.3.1. Network Effects. Many IT product markets are associated with varying degrees of direct or indirect network externalities, where the utility that a user derives from a product or service depends on the number of other users who are in the same network (Katz and Shapiro 1985). In addition, multisided platforms that serve two or more different groups of customers who are subject to indirect network effects are common in the IT industries (Evans 2003). In industries with high network effects, there is a significant first-mover advantage, and being late to the game may foreclose the opportunity to invest or enter the marketplace in the future. In the extreme case, the presence of network effects may simply lock out late entrants (McGrath 1997).

Therefore, when new technology opportunities arise and VC-backed startups are particularly active, firms operating in high-network effects markets run the risk of facing very high potential costs if they choose to defer investments in R&D in the current period but then need to catch up later because of the winner-takes-all nature of competition (Weeds 2002). If the cost of deferring such investments is high, especially if and when the perceived new entry threats do materialize, the IT incumbent operating in market where network effects are particularly strong should choose to increase its R&D spending as soon as NET emerge. Furthermore, in such markets, immediate investments in R&D may also create valuable options for expansion later on. Thus, we expect a firm that operates in an industry with strong network effects to make relatively higher and earlier R&D investments in response to high new entry threats.

2.3.2. Technological Cumulativeness. Innovations in the IT industries often depend on technological cumulativeness, which is defined as "the degree of serial correlation among innovations and innovative activities" (Breschi and Malerba 1997). Cumulativeness represents the extent to which earlier innovations are needed for building the next generation of innovative products, thereby capturing an element of path dependence by which technology progresses sequentially. Prior research shows that, in industries such as semiconductors and computers, technological advances often build on and interact with elements of existing technologies (Merges and Nelson 1990, Ham Ziedonis and Hall 2001). In technological fields that display high levels of cumulativeness, the success of R&D effort hinges on the relevant continuities with prior innovative activities. Firms that do not have control over earlier innovations are often unable to exploit subsequent ones (Green and Scotchmer 1995), potentially because of a failure to develop relevant absorptive capacity (Cohen and Levinthal 1990) or being locked out of the market (Hill and Rothaermel 2003).

The lockout effect associated with high technological cumulativeness is likely to influence the value of using R&D as a preemptive response to new entry threats, because it is important for the incumbent to secure a foothold of the relevant technological competence in the early stage of the technology's lifecycle (Oriani and Sobrero 2008). Here too, as in the case with network effects, deferring investments in R&D to a later date becomes an expensive option. Additionally, prior research notes that a high level of technological cumulativeness is often associated with high appropriability of innovations (Breschi and Malerba 1997), which is necessary to ensure future success if and when new entry threats materialize. Thus, in sectors where technological cumulativeness is important, incumbents will be more willing to respond aggressively in terms of innovation spending in the face of new entry threats. The analyses presented next describe how we test for the direct effect as well as the moderating effects of new entry threats on R&D spending in the IT industry. We start by describing the development of a new text-based measure for new entry threats.

3. A Text-Based Measure of New Entry Threats

As discussed above, a measurement of new entry threats has remained elusive in the literature given the forward-looking nature of this construct. Most existing work has used current competitive intensity or observed (actual) entry as surrogates (Chen et al. 1992, Aghion et al. 2009), but these are approximations at best and do not reflect the construct as discussed in earlier theoretical work (Caves and Porter 1977, Porter 2008). The challenge for the incumbent is to spot the disruptive entrepreneurial firms that constitute such threats early on and respond adequately (Rigby et al. 2002). Although detecting a specific disruptor is imprecise and uncertain (Markides 2006), broad movements within the entrepreneurial space that intersect with the incumbent's product markets can still be recognized as significant threats. Prior work studying technology fads/cascades has discussed these broader trends as being useful predictors of firm and individual behavior (Abrahamson 1991, Bikhchandani et al. 1998). We argue that large-scale new venture formation within a certain product market is a valid representative of new entry threats, leading potentially to impending competition in the future.

3.1. Methodology: From Words to New Entry Threats

The use of text analysis requires descriptive text corpora from firms to construct appropriate measures. A considerable body of work has used public filings provided by the firms, specifically annual reports (10-Ks) in the United States, to create measures of firm fundamentals, such as competitive intensity, industry classes, and firm strategy (Tetlock et al. 2008, Hoberg and Phillips 2010, Tetlock 2011). These documents are useful as sources of data for two reasons. First, business descriptions provided by public firms must be representative and accurate as required by financial market regulations. Thus, product descriptions of public firms contain timely information about their products, markets, and competitors. Second, as firms evolve, these descriptions are modified and updated to reflect the changing nature of their businesses, thereby providing longitudinal variation. Thus, for incumbents, we utilize their annual 10-K filings as informative and credible sources of text.

We also need a source of text describing the entrepreneurial ecosystem that can be used to characterize the extent to which they represent credible threats. For this purpose, we use the VentureXpert data set and focus on startups backed by venture capital funding. Using VentureXpert data allows us to focus on IT entrepreneurs who have received venture capital funding, and their ventures are, therefore, of baseline quality and represent credible threats to incumbents. It is important to note that individual new ventures included in VentureXpert are typically too small and early stage to count as competitors or threats to incumbent firms. Therefore, we refine our definition of new entry threats from startups in two ways. First, we observe that the threats of new entry rarely originate from a single entrepreneur, but instead, they are from broad collective movements in the startup space (i.e., evidence of systematic entrepreneurial movements into a specific subindustry is more representative of NET for an incumbent). Following Hoberg et al. (2014), we identify new entry threats at the level of an "industry"; for the purposes of our analysis, we treat the whole set of entrepreneurial ventures in VentureXpert that receive VC funding as the relevant "industry." Second, it is unlikely that all entrepreneurial firms represent emerging new entry threats to the incumbent. Therefore, we consider those firms that receive *first-round funding* in a given year as posing new entry threats to incumbents. If the entrepreneurial ecosystem observes value in a specific industry or technology space and systematically invests in new ventures at early funding stages, there is likely to be a groundswell of new ventures associated with this industry segment entering the VentureXpert data set in a given year, which could then potentially lead to significant realized entry in subsequent years, thereby representing NET for the incumbent. Thus, the new entry threats measure here is based on (1) new ventures that receive *first-round* venture capital funding and (2) the collective body of all entrepreneurs who receive first-round funding rather than the individual entrepreneurs. The new entry threats captured here are not meant to be exhaustive in scope; for example, other sources of potential new entry include startups receiving crowd funding as well as those operating in foreign markets. However, in the interest of tractability and parsimony, we adopt this conservative approach to capture the set of potential new entrants that are most relevant in the IT sector.

The degree of new entry threat to an incumbent depends on how closely the startups are related to the primary market of the incumbent. We, therefore, require a measure of similarity between the product portfolios observed in the VC-funded startup space and the incumbent's product market. More to the point, we need to measure the similarity between the text from the incumbent describing its product market on the one hand and text from the entrepreneurial space on the other hand. We use the cosine text similarity approach to capture this similarity (Sebastiani 2002). For publicly traded incumbents, the source of business descriptions is section 1 of their 10-K annual filings. For startups, we use their business descriptions from the VentureXpert database. We extract all detailed business descriptions from startups that received first-stage funding and aggregate these descriptions for each year t; the individual business descriptions are short, with the typical description consisting of four to five sentences. Aggregating these for a given year provides a more representative and useful document of entrepreneurial space. Cosine similarity between this collective entrepreneurial document and an incumbent's business description forms the basis for measuring new entry threat, effectively by calculating their overlap in word usage.

Specifically, after the respective text documents are available, we parse semantics at the sentence level with the Natural Language Processing Toolkit (Bird et al. 2009) and retain the nouns and proper nouns, which are the most meaningful elements in product descriptions. We remove commonly used English stop words. We also omit geographical words, such as country, state, and city names, as well as the words describing time periods, such as months and dates, following Hoberg and Phillips (2016). Our results are robust to the inclusion of these stop words. Figure 2 presents a histogram of frequencies of the number of unique words used in the product descriptions of the incumbent firms, showing that the typical firm uses roughly 700 unique words. Figure 3 displays the number of startups that received first-round funding by year,

Figure 2. (Color online) Histogram of the Number of Unique Words Used in Incumbents' 10-K Product Descriptions

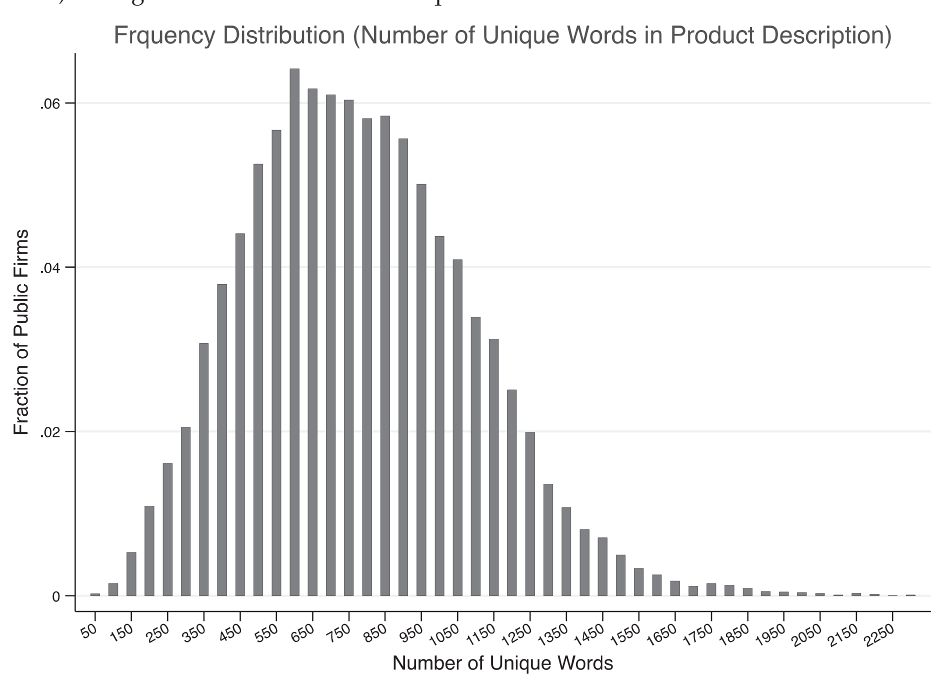
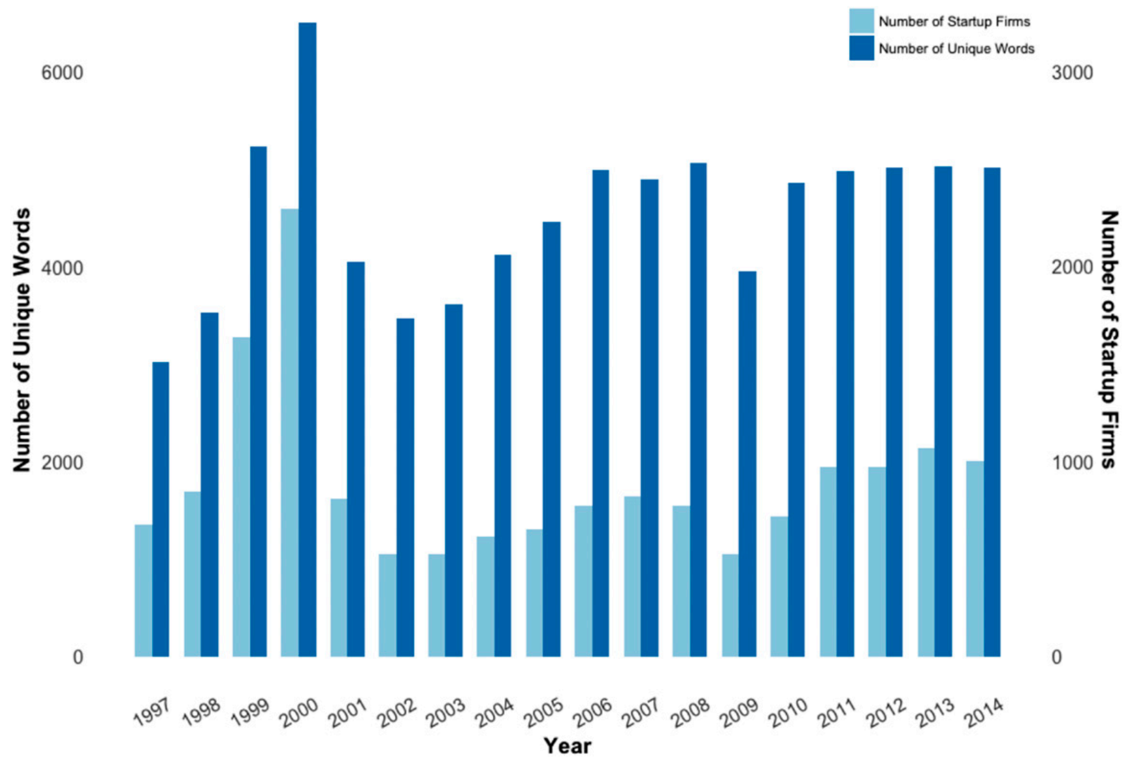


Figure 3. (Color online) The Number of Startups and the Number of Unique Words Used in the Startups' Product Descriptions by Year



which ranges from 522 to 2,299, and the number of unique words used in the collective startups' product descriptions across the years, which ranges from 3,020 to 6,508 words.

Next, we define all incumbents' business descriptions and the aggregated startup document as a cumulative document corpus for each year t (that is, the corpus includes $n + 1$ documents in total, with n being the number of incumbents in year t). Subsequently, we build document vectors for each incumbent's text and the aggregated startup text in year t. Let J_t denote a scalar equal to the length of the words dictionary, which includes all unique words used in the document corpus of year t. Let W_{it} represent an ordered vector of length J_t describing the pattern in which the J_t words are used in document i ($i = 0$ represents the aggregated file from the startup ecosystem) in year t. We use *term frequency times inverse document frequency* (*TF-IDF*) (Tata and Patel 2007) as the weight for each word in the document vector. In this case, each element j in W_{it} captures the relative importance of word (or term) j in document i given its within-document and cross-document frequencies. *Term frequency* (f_{it}) is defined as the number of occurrences of words j in document i. The normalized term frequency is defined as

$$TF_{ji} = \frac{f_{ji}}{\max_k f_{ki}}. \tag{1}$$

That is, the term frequency of term j in document i is f_{ji} normalized by the maximum number of occurrences of any term in document i. This normalization process helps correct for biases caused by the length of a document (i.e., term frequency gets inflated in longer documents). *Inverse document frequency* (IDF_{ji}) for a term is defined in the following fashion. Suppose that term j appears in n_j of the N documents in the collection; then, $IDF_j = \log(N/n_j)$. Naturally, a term that appears in many documents, such as "service," gets a lower *IDF* weight (and therefore, is treated as less important), whereas a term that occurs in only a few documents, such as "encryption," gets a higher *IDF* weight. The weighting score of *TF-IDF* for term j in document i is defined as $TF_{ji} \times IDF_j$. Intuitively, words with high within-document frequency obtain higher weighting, and those with high cross-document frequency are weighted less. Lastly, because our main interest is in the similarity between the document representing an incumbent and the aggregate document representing startups in a year, we operationalize the measure of new entry threats for incumbent firm i in year t as

$$NET\_TFIDF_{i,t} = SIMc\left(\overrightarrow{W_{i,t}}, \overrightarrow{W_{0,t}}\right) = \left(\frac{\overrightarrow{W_{i,t}} \cdot \overrightarrow{W_{0,t}}}{\overrightarrow{|W_{i,t}|} \times \overrightarrow{|W_{0,t}|}}\right), \tag{2}$$

where W_{it} denotes incumbent firm i's document vector in year t ($i = 1,2,3\ldots n$) and W_{0t} represents the aggregated startups' business descriptions document vector. By construction, the cosine measure of new entry threat NET_TFIDF_{it} is bounded within range [0, 1], with higher values representing greater threats of new entry for the incumbent (because the two word vectors are closer in unit vector space).

There are several reasons why the cosine similarity score calculated using TF-IDF weighted words vector is an appropriate measure of new entry threats. First, the properties of TF-IDF are well understood given its wide application in the studies of information processing and text analysis (Hiemstra 2000, Aizawa 2003, Aral and Van Alstyne 2011). Second, the measure is intuitive given its consideration of words frequency within as well as between documents. Third, the cosine similarity's normalization builds in a natural control for document length, because it measures the angle between two word vectors on a unit sphere.

3.2. Validation of the NET Measure

On the basis of the methodology defined above, we calculate NET for all public firms found in Compustat within the high-tech industries (including IT industries) from 1997 to 2014. Our sample of high-tech industries is defined using the 46 four-digit NAICS codes in Hecker (2005). Because NET is a new measure, in this section, we illustrate the validity of our measure through a series of tests. These include assessing whether our measure captures changing trends in the startup space over time; examining how NET is associated with the changes in the competitive dynamics of the incumbent's industry in subsequent years; comparing the turbulence experienced by high-NET firms with those experienced by low-NET firms in subsequent years; and examining how NET in a selected number of industries varies according to industry-level demand shocks caused by major, well-known exogenous events. We believe that these tests collectively provide sufficient validation for the measure and describe these in some details below.

3.2.1. Capturing Changing Trends in the Entrepreneurial Space. As a first step to validating our NET measure, we examine whether our approach of using text indeed accounts for shifting technology trends within the startup space as reflected in the "hottest" words used in product descriptions of startup firms across years in our sample. Table 1 shows the list of 20 words with the highest *TF-IDF* weights in the collective startup product descriptions document in three selected years—2000, when the dot com bubble was at its peak; 2003, when the stock market reached the lowest point after the dot com collapse; and 2006, when the economy had recovered from the dot com bubble.

Table 1. Top 20 Words in Entrepreneurial Firms' Documents in Selected Years

Year	Words list
2000	onlin, Internet, softwar, wireless, softwareprovid, servicesprovid, solut, web, content, broadband, ecommerc, softwaredevelop, servicesdevelop, platform, media, drug, email, network, enterpris, patient, portal
2003	diseas, patient, drug, therapeut, cancer, softwar, therapi, treatment, protein, wireless, discord, video, molecul, inhibitor, tissu, healthcar, biotechnolog, antibodi, pharmaceut, cell
2006	onlin, diseas, patient, cancer, drug, therapeut, search, publish, media, video, therapi, cloud, blog, antibodi, tissu, treatment, game, softwar, advertis, platform

Note. All documents are stemmed and preprocessed—therefore, the words presented here are in their root form.

We observe several interesting patterns. First, there is significant longitudinal variation in the most influential words that venture-funded startups use to describe their products and services. Second, the changes in the vocabulary reflect systematic shifts in technology trends that are consistent with observations in high tech. For example, in 2000, the VC-funded entrepreneurial space was dominated by firms related to the internet or software industries: words, such as "online." "internet," "software," "web," "email," "broadband," "ecommerce," and "portal," were among the most frequently used in their product descriptions. In fact, all but 2 among the top 20 words were related to information and communication technology industries in 2000. However, in 2003, we observe a significant change in the vocabulary used to describe funded startups. The use of internet-related words was dramatically reduced, replaced by words such as "disease," "patient," "drug," "treatment," "therapy," "protein," "biotech," and "antibody." These changes show that the VC-funded startup space had shifted systematically from internet/software to pharmaceutical and biotech industries after the dot com bubble. Interestingly, in 2006, we see the word list reflecting a balance between IT and biotech industries. Although some software and internet-related terms resurface, they do so with a completely different emphasis. Terms, such as "search," "cloud," "blog," "advertising," "video," and "game," become more influential, reflecting a trend toward cloud computing, social media, online advertising, and video games in the IT industry. Overall, Table 1 provides evidence for the critical longitudinal variation in words used to describe the entrepreneurial firms in different years, showing that our measure captures underlying trends within the startup space with fidelity.

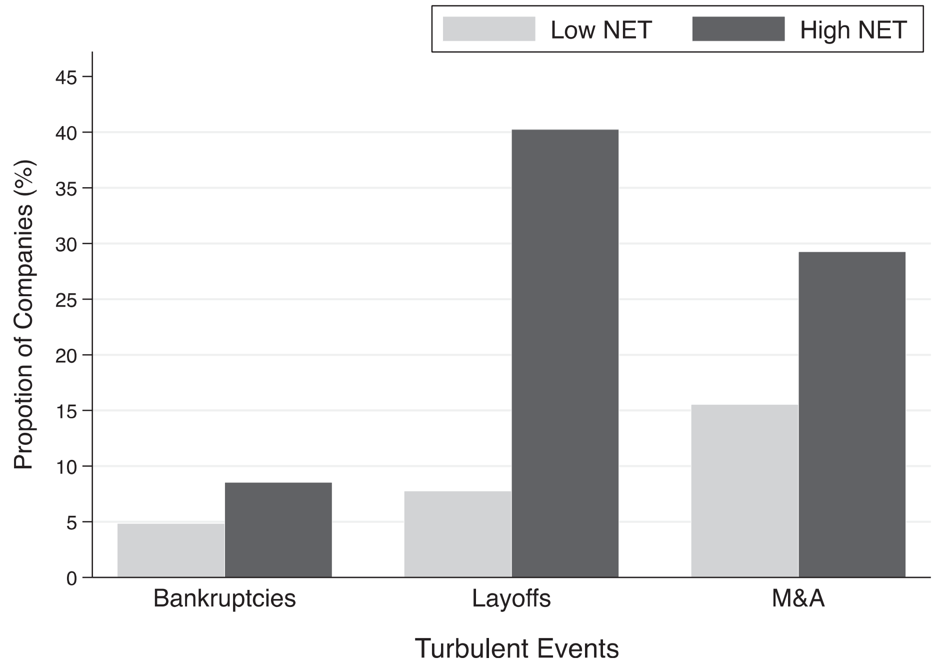
3.2.2. New Entry Threat and Future Competitive Dynamics. If the NET measure indeed captures the extent to which incumbents face potential new entry, one way to validate the measure is to examine how NET is associated with the changes in the competitive landscape of subsequent periods. This would imply that firms with higher values of NET are likely to face, in the subsequent two to three years, an increase in the number of direct competitors, all else being equal. This may happen through a number of mechanisms. To start with, some fraction of the new startups may eventually go public and become a rival for the incumbent. In other cases, competing firms may form alliances or joint ventures with the startups, or acquire technology licensing from them, and invade the product space of the focal incumbent (Mitchell and Singh 1992). Relatedly, prior literature has shown that one of the direct consequences of new entry is the reduction of profitability in the industry as a whole because of increased competitive intensity (Audretsch and Mata 1995). To test these expectations, in Online Appendix A.1, we present details of several model-free illustrations that confirm a positive association between NET and the number of future rivals faced by an incumbent as well as a negative association between NET and future profitability for the same incumbent firm.

3.2.3. Turbulence Experienced by High-NET Vs. Low-NET Firms. One of the defining characteristics of new entry threats is that they threaten the sustainability of the firm's future earnings and viability of its product portfolio (Hoberg et al. 2014). As a result, if some of the threats do indeed materialize, firms with high levels of NET are more likely to suffer from deteriorating operational performance, and therefore, they are more likely to experience turbulent events, such as liquidation or downsizing, during the difficult time period that follows. We test these eventualities here. We contrast high-NET firms with those having low NET to study the likelihood of turbulent events, such as filing for bankruptcy, announcing significant downsizing or layoffs, and being acquired by or merged with other companies in the five years after the measurement of NET.

Specifically, for each year in our sample frame (1997–2014), we select 10 firms with the highest NET scores and 10 with the lowest NET scores, and we put them into a *high group* and a *low group*, respectively. For each observation in the high group and the low group, we conduct a search in the Lexis/Nexis database to identify news releases that are related to turbulent events associated with the firm in the subsequent five years using a Boolean query combining the company name and keywords, such as "bankruptcy," "liquidation," "layoff," "cut jobs," "merger," and "acquisition." Sources are limited to four types: newspapers, business and industry news, U.S. newspapers, and web news. We also search through news releases from the firm to confirm the dates and details of events thus identified, and we report the frequency of such events across the two sets of firms, representing high and low rates of new entry threats. As an illustration, in Online Appendix 2, we present a summary of these events for the two groups of firms in 2009. Figure 4 reports the comparison of the rates of incidents between the two groups for three types of events: bankruptcies, layoffs, and acquisitions/mergers. As expected, companies facing high levels of NET have much greater likelihood of experiencing turbulent events—8.5% (7 companies) of the 82 companies filed bankruptcy, 40.2% (33 companies) announced significant layoffs, and 29.3% (24 companies) were acquired by or merged with other firms in the subsequent five years. In comparison, among the companies in the low group, only 4.8% (5 companies) of 103 companies filed bankruptcy, 7.8% (8 companies) announced layoffs, and 15.5% (16 companies) were acquired or merged. Two-sided t tests show that the probabilities of "layoffs" and "M&A" are significantly different between the two groups ($p < 0.05$ and $p < 0.01$, respectively). Companies facing low NET fared much better than those facing high NET in subsequent years, providing evidence that the measure is forward looking and correlates with firms' future competitive dynamics.

3.2.4. NET and Industry-level Exogenous Demand Shocks. Another source of validation for NET arises from the expectation of observing higher values in NET when an exogenous event creates the potential for new entry into that industry. The underlying argument here is that a sudden increase (or decrease) in industry-level demand because of an exogenous event disrupts the relationship between supply and demand and therefore, encourages (or discourages) entry. As a result, incumbents in the industry will likely face increasing (or decreasing if demand shock is negative) levels of NET after the events. We consider four selected high-tech industries and their associated exogenous events as exemplars. The first is the military armored vehicle manufacturing industry[3] after the terrorist attacks on September 11, 2001. The second event pertains to the dot com collapse in March 2000 and the resulting changes in the internet-related industries. Third, we consider changes in the software publishing industry after Apple's announcement of a major and critical software development kit release for iOS, which drove hundreds of app developers into the mobile apps market. Fourth, we trace changes in the biotechnology industry, which experienced a demand boost after the complete sequencing of the human genome. We choose these events because of their importance in shaping the trajectory of these high-tech industries, thereby potentially offering (or extinguishing in the case of a negative shock) new opportunities for entrepreneurial firms. We present the details of these analyses in Online Appendix 3.

Figure 4. Turbulent Events in the Subsequent Five Years



Notes. Contrasts between high NET and low NET firms. M&A, mergers and acquisitions.

To summarize the results, in all four cases, we find that the longitudinal variations in the NET measure in selected industries are consistent with demand shocks after well-known industry-wide exogenous events, adding to the validity of this measure. We now turn to addressing the primary research question of interest pertaining to the influence of NET on the incumbent's R&D investments.

4. Data and Empirical Analyses

In this section, we start by describing the sample and other variables that are used in our empirical investigations. We then present a panel data model to test the association between NET and R&D investments in the specific industrial context that we study. Furthermore, we examine the robustness of the estimates to endogeneity concerns by using a number of techniques, such as dynamic panel generalized method of moments (GMM) estimations and instrumental variables regressions. We then proceed to test the effects of the proposed boundary conditions (network effects and technological cumulativeness) that may moderate the relationship between NET and firm R&D investments.

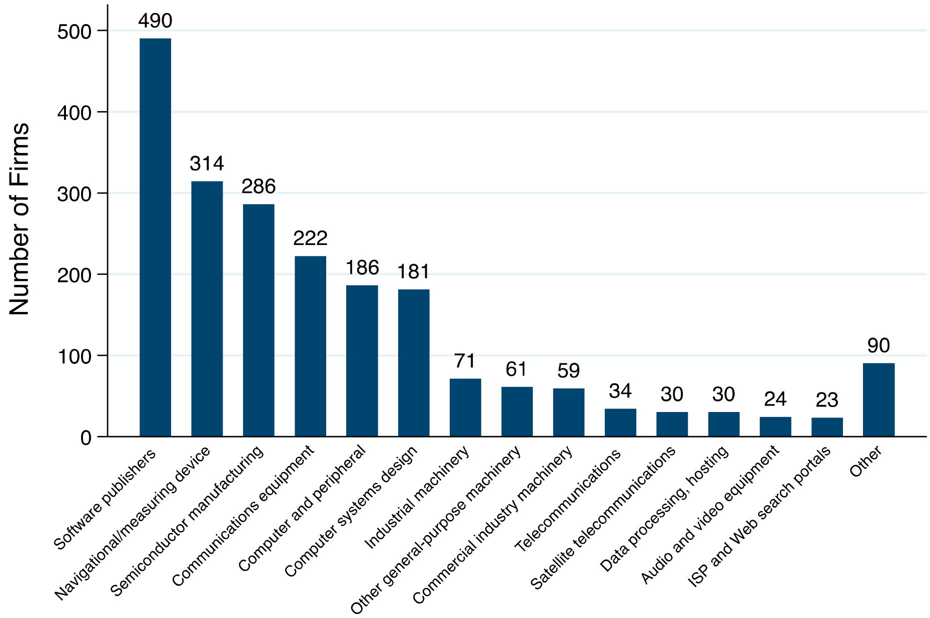
4.1. Data and Variables

We restrict the analyses to the set of firms in the IT industries using the 24 four-digit NAICS industry codes that include IT software, hardware, and services industries over the period 1997–2013 (Kim et al. 2016). We list the NAICS codes as well as their text descriptions in Table A3 in the online appendices. Figure 5 shows the distribution of firms among the subsectors in the IT industry. We obtain financial data and other firm characteristics from Compustat. Our primary data set consists of 2,101 publicly traded firms over a 17-year period with 14,410 firm-year observations, representing an unbalanced panel. The sample period includes years when there was considerable turbulence in the IT industry (e.g., during the internet boom) as well as the less volatile years. The long panel also includes the period of the global financial crisis in 2008 and the period of recovery afterward, which significantly affected IT-related venture capital funding and entrepreneurial activities in general. Together, the data set provides considerable longitudinal variation in our measure of new entry threats that allows us to use firm-level fixed effects models to control for many unobserved firm heterogeneities. We describe the variables in our main analyses below.

4.1.1. R&D Investments. Following prior literature, we measure R&D investments using R&D intensity, which is defined as R&D expenditures over a firm's total asset (Blonigen and Taylor 2000, Hall 1988). R&D expenditures reflect contemporaneous managerial decisions that are closely associated with a firm's investment strategy. The mean value of R&D intensity in our sample is 12.88%.

4.1.2. Network Effects. We adopt the measure of network effects invented by Srinivasan et al. (2004) and Wang et al. (2010). Specifically, the authors in these studies identified 45 product categories that are characterized

Figure 5. (Color online) Sample Distribution Across Four-Digit NAICS Industries



by varying degrees of network effects as well as the pioneers (leading firms) in each of the product categories. These products range from computer hardware (e.g., workstations, mainframe computers, and personal computers), computer software (e.g., database, personal finance software, word processing software, and spreadsheet software), and consumer entertainment electronics devices (e.g., home VCRs, DVD players, videogame, and color television) to telecom equipment (telephones, fax machines, and wireless telephones) and other office suppliers (such as printers and scanners). Two groups of raters—academic experts and MBA students with background in high-tech marketing strategies—were asked to rate separately the degrees of direct and indirect network effects associated with each product category on a one (no network externality) to seven scale (very high externality) (Srinivasan et al. 2004). The strength of network effects is then computed by adding scores for both direct and indirect network externalities (with a range of 2–14).

We match the 45 product categories to four-digit NAICS industries by referencing the industry classifications of the pioneers identified in each product category (Wang et al. 2010). As a result, we successfully identify network effects for 15 four-digit NAICS industries.[4] Based on this matching, we constructed two different forms of the measure of network effects: a continuous score associated with each four-digit NAICS industry[5] and a binary measure of high-network effects industries, with value that is set to one for industries in which the score of network effects is higher than the median value of 8.4 among the 45 product categories (Wang et al. 2010) and zero otherwise.

4.1.3. Technological Cumulativeness. Patent self-citation or citation referring to previous patents owned by the same patentee has been used as an indicator of the degree of cumulative, sequential invention by a firm in prior research (Lanjouw and Schankerman 2001). Following this interpretation, Oriani and Sobrero (2008) defined the degree of technological cumulativeness by referring to the average percentage of patent self-citations at the industry level. We adopt a similar definition and use the National Bureau of Economic Research (NBER) patent data to calculate the backward self-citation rate for each patent, which is the number of backward citations made to a patent with the same assignee code (self-citations) divided by the total number of backward citations (see Hall et al. 2001 for full details). Using the patent self-citation data, we first compute a continuous measure of *technological cumulativeness* as the average percentage of self-citations at firm level using the firm's patent portfolio during the period of 1996–2006 when the patent data are available. Then, following Oriani and Sobrero (2008), we also construct an industry-level binary indicator of technological cumulativeness. In particular, we calculate the average percentages of patent self-citation at the industry level using all of the patents filed by the patentees in the focal industry and then identify a set of high-technological cumulativeness industries, for which the variable *high-tech cumulativeness industries* is set to one when the average percentage of self-citations of the industry is in the top quartile among all of the industries in our sample and zero otherwise. We use four-digit NAICS-based industry classification to define high-tech cumulativeness industries. Because of the static industry membership used by

NAICS, the variable high-tech cumulativeness industries is constant over time.[6]

4.1.4. Control Variables. Following prior literature (Kaplan and Zingales 1997, Becker-Blease 2011, Atanassov 2013), we control for various firm characteristics that may affect a firm's R&D investment decision, including firm size (sales), age, profitability, asset tangibility, leverage, capital expenditure, growth opportunity, and financial constraints. We control for competition for the incumbent firm by the Herfindahl–Hirschman index based on the TNIC scheme created by Hoberg and Phillips (2010). We choose TNIC over NAICS, because TNIC classifications are updated every year as firms file 10-K reports, allowing for a more accurate measure of competition (Kim et al. 2016). We also control for *realized entry* into the incumbent's product market by taking advantage of the dynamic nature of the TNIC scheme: the TNIC captures the set of rivals for each focal firm in a given year based on the text included in the firms' 10-K filings. We, therefore, define realized (observed) entry for a firm as the number of the firm's rivals in year t less the number of rivals in year $t-1$. Finally, we control for the firm's product diversification using the entropy measure of sale shares in different lines of business of Jacquemin and Berry (1979). Entropy is computed using data on firm sales in each six-digit NAICS business as reported by the Compustat Segment database. Diversification is thus measured as the within-firm mean of $Entropy_{it}$ over the sample period.[7] We summarize variable definitions in Table 2 and provide descriptive statistics and correlations in Table 3.

4.2. Baseline Analysis of NET on R&D Investments

We start with a baseline model that builds on prior IS work explaining firm innovation spending (Kim et al. 2016). Particularly, we model the effect of NET on an incumbent firm's R&D investment using a two-way fixed effects panel data specification below:

$$R\&D\_Intensity_{it} = \eta_t + \lambda_i + \beta \times NET_{it} + \gamma X_{it} + \mu_{it}. \quad (3)$$

Here, i indexes firms, and t indexes time periods. The variable NET_{it} is our measure of new entry threats. X_{it} is a set of firm characteristics that affects a firms' investment decisions. We control for time-invariant unobservable firm characteristics by including firm fixed effects λ_i. We also include year fixed effects η_t to control for economy-wide shocks. Heteroskedasticity robust standard errors are clustered at the firm level to control for serial correlation (Wooldridge 2010); μ_{it} represents the idiosyncratic errors.

We report the results from the fixed effects model in column (1) of Table 4. For comparison, we also present a random effects panel data model in column (2) of Table 4. In the random effects models, the unobserved individual heterogeneity λ_i is assumed to be uncorrelated with the included regressors (Greene 2003). We find that the coefficient of *NET* is negative in the fixed effects model and significant at the 5% level, indicating that greater new entry threats are associated with lower level of R&D investments, all else being equal. A Hausman's test comparing the fixed effects and random effects estimates rejects the orthogonality between the random effects and the regressors ($p < 0.01$). Therefore, without the strong assumption that the firm heterogeneity is uncorrelated with the regressors, random effect estimates are likely to be biased. We, therefore, interpret our results using the fixed effects estimates.

To illustrate the magnitude of the effect of NET on R&D investments, consider a one standard deviation (SD) increase in new entry threat (before standardization, NET has a mean of 0.07 and an SD of 0.05). The coefficient of NET implies a 0.35 percentage point reduction in R&D intensity (recall that the sample average R&D intensity is 12.88 percentage points), which translates to a reduction of $6.5 million in R&D investments based on the mean level of total asset in the sample. This finding—firms in the IT industries invest less in R&D as uncertainty increases in the face of potential new entry—is interestingly consistent with the conclusions of research studying firms' R&D investment decisions in manufacturing when facing market uncertainty (Czarnitzki and Toole 2011, 2013) as well as capital investment decisions of firms facing profit uncertainty (Ghosal and Loungani 2000).

Although a fixed effects model controls for many sources of unobserved firm heterogeneities, a particular concern here is that the presence of preemptive R&D might deter entry or that other unobserved industry-wide shocks (such as technological opportunities) may influence entrepreneurial entry, VC funding decisions, and incumbent R&D investments simultaneously, causing our NET measure to be endogenous. As a way to address these issues, we relax the assumption that NET is strictly exogenous and test a model using the GMM-based dynamic panel data models estimator (Arellano and Bond 1991, Blundell and Bond 1998). Taking advantage of our long panel and the large number of firms in our data set, we construct internal instruments within the data following the conventions of the dynamic panel data methods. Specifically, we use the *differences GMM* estimator, using lag terms of our endogenous variables, *NET* and *L.R&D Intensity*, and all differences of other exogenous variables, including year dummies, as our instrument variables for the differenced equation. We use the second lag and onward of endogenous variables for the difference GMM specifications.[8] We check the validity of the moment conditions required by the differences GMM estimator using the Hansen test, which does not reject the assumption that our instruments are exogenous (Arellano and

Table 2. Variable Definitions and Data Sources

Variable and data source	Definition
Innovation variable (data source: Compustat)	
R&D_intensity$_{it}$	Research and development expenditure to total assets ratio of firm *i* in year *t*
New entry threat variable (data source: VentureXpert and 10-K files)	
NET$_{it}$	A text-based measure of threat from new entry by term frequency–inverse document frequency-weighted cosine similarity between business description of startups and established firms
Boundary conditions (data source: Compustat and NBER Patent Database)	
Network effects$_i$	Network effects index constructed by examining network externality of 45 categories, including computer hardware, computer software, consumer electronics, etc., from 1950 to 2007 (Srinivasan et al. 2004, Wang et al. 2010)
StrongNE_industry$_i$	Equal to one if network effect is greater than median of network effect in 45 categories defined by Wang et al. (2010), zero otherwise
Tech Cumulativeness$_{it}$	The average patent self-citation rate of all of the applied patents for firm *i* in year *t*
HighTC_industry$_i$	Equal to one if the four-digit NAICS industry average patent self-citation rate is at top quartile, zero otherwise
Firm characteristics (data source: Compustat and TNIC)	
Sale$_{it}$	Total sale of firm *i* in year *t* (in $ billion)
Age$_{it}$	Number of years since listing of firm *i* in year *t*
ROA$_{it}$	Operating income before depreciation to total assets ratio of firm *i* in year *t*
Asset_tangibility$_{it}$	Net property, plants, and equipment to total assets ratio of firm *i* in year *t*
Leverage$_{it}$	Total debt of firm *i* in year *t* divided by its total assets
CapExp/Assets$_{it}$	Capital expenditure to total assets ratio of firm *i* in year *t*
Tobin's Q$_{it}$	Market to book ratio of firm *i* in year *t* as defined in Brown and Caylor (2006)
KZ Index$_{it}$	Kaplan–Zingales index (Kaplan and Zingales 1997) is a relative measurement of reliance on external financing; companies with higher Kaplan–Zingales index scores are more likely to experience difficulties when financial conditions tighten, because they may have difficulty financing their ongoing operations[a]
TNIC_HHI$_{it}$	Herfindahl–Hirschman index of firm *i* in year *t* based on TNIC (Hoberg et al. 2014)
Observed Entries$_{it}$	Number of firm *i*'s rivals in year *t* less number of firm *i*'s in year $t-1$
Product Diversification$_i$	Product diversity measured by over sample within-firm mean of entropy measure of sale shares in different lines of business firm *i*

[a]Following Chemmanur and Tian (2018), we use the regression coefficients from Kaplan and Zingales (1997) to compute the Kaplan–Zingales index as $-1.002 \times \mathit{Cash\,flow} - 39.368 \times \mathit{Dividends} - 1.315 \times \ \mathit{Cash\,flow} + 0.28 \times Q + 3.18 \times \mathit{Leverage}$.

Bond 1991, Roodman 2009). We also test the validity of the GMM assumptions in our model. The test results are reported at the bottom of Table 4, which indicates that our model specification shows no significant serial correlation in the first-differenced disturbances.

We report the results from the Arellano–Bond estimator of dynamic panel data models, treating *NET* as exogenous first in column (3) of Table 4 and then as endogenous in column (4) of Table 4. We observe that the coefficient estimate of *NET* in column (3) of Table 4 is similar to that of the fixed effects model, consistent with our main finding that new entry threats reduce R&D investments. Moreover, the coefficient of *NET* in column (4) of Table 4 ($\beta = -1.334$) is significantly higher than that from the fixed effects model. The larger estimate in the dynamic panel data model, accounting for the endogeneity of *NET*, suggests that the presence of endogeneity, if any, likely causes a downward bias in the fixed effects model, whereas the baseline model generates more conservative estimates of the effect of *NET*. Overall, the Arellano–Bond estimates provide additional support for the finding that the uncertainty associated with potential entry reduces the inclination of incumbents to invest in R&D. Our results are consistent with existing work showing that firms facing market turbulence are more likely to respond conservatively (Brav et al. 2005, Hoberg et al. 2014).

To further alleviate endogeneity concerns, particularly the reverse causality issues related to R&D investments as an entry barrier, in Online Appendix 4, we present a two-stage least squares analysis using an alternative identification strategy involving the use of two instrumental variables for NET that represent industry-level incentives and barriers to potential entry. Here again, we find that the coefficient estimate of *NET* remains negative and is larger than that from the fixed effects model in Table 4, showing that the results reported here are robust to the endogeneity of NET.

Table 3. Summary Statistics and Correlation Coefficients

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. *R&D Intensity* (%)	12.88	15.53	1.000													
2. *NET*	0.07	0.05	0.115*	1.000												
3. *Network Effects*	7.71	1.97	0.078*	0.233*	1.000											
4. *Tech Cumulativeness*	0.06	0.07	−0.076*	−0.194*	−0.154*	1.000										
5. *Sale* ($ billion)	1.36	7.03	−0.086*	0.086*	0.030*	0.159*	1.000									
6. *Age*	15.52	12.37	−0.219*	−0.212*	−0.180*	0.333*	0.285*	1.000								
7. *ROA* (%)	−1.26	35.99	−0.667*	−0.050*	−0.048*	0.063*	0.086*	0.187*	1.000							
8. *Asset Tangibility* (%)	13.34	11.91	−0.052*	−0.160*	−0.053*	0.131*	0.038*	0.103*	0.022*	1.000						
9. *Leverage* (%)	9.34	17.21	−0.084*	−0.069*	−0.041*	0.042*	0.056*	0.109*	0.027*	0.230*	1.000					
10. *CapExp/Assets* (%)	3.93	4.28	0.063*	−0.004	0.065*	0.045*	0.009	−0.080*	−0.055*	0.613*	0.070*	1.000				
11. *Tobin's Q*	2.48	2.96	0.194*	0.072*	0.084*	0.043*	−0.017*	−0.140*	−0.207*	−0.049*	−0.045*	0.103*	1.000			
12. *KZ Index*	0.25	2.69	0.046*	0.016	−0.017*	0.015	−0.008	−0.033*	−0.060*	0.056*	0.232*	0.066*	0.205*	1.000		
13. *TINC-HHI*	0.23	0.21	−0.117*	−0.219*	−0.043*	0.059*	−0.030*	0.158*	−0.013	−0.000	0.049*	−0.057*	−0.042*	−0.030*	1.000	
14. *Observed Entries*	−6.33	37.25	−0.085*	−0.074*	−0.078*	0.059*	0.029*	0.100*	0.074*	0.040*	0.045*	−0.008	0.032*	0.014	−0.007	1.000
15. *Product Diversification*	0.94	0.51	−0.161*	−0.107*	−0.105*	0.139*	0.255*	0.363*	0.148*	0.130*	0.099*	−0.001	−0.066*	0.001	0.062*	0.060*

Notes. This table reports the summary statistics for primary variables constructed based on the sample of U.S. public firms in the IT industries from 1997 to 2013. (We include all IT Industries, such as hardware, software, and IT services industries, which are defined by four-digit NAICS code: 2211, 3332, 3333, 3336, 3339, 3341, 3342, 3343, 3344, 3345, 3346, 5112, 5161, 5171, 5172, 5173, 5174, 5179, 5112, 5181, 5182, 5413, 5415, 5416, and 5417.) Please see Table 2 for the description of the variables. Pearson correlation coefficients are reported for our sample of 14,410 firm-year observations.

*$p < 0.05$.

4.3. Testing for Moderation

4.3.1. Network Effects. We test for the moderating effect of NEs using the index adopted from Srinivasan et al. (2004) and Wang et al. (2010). The results are reported in Table 5. We first use the continuous NE measure and present the fixed effects model estimates in column (1) of Table 5. The coefficient of the interaction term *NET* × *Network Effects* (β = 0.113) is positive and significant ($p < 0.10$). To understand the economic significance of the boundary condition of network effects, we use the binary measure of *strong network effect industries*. This variable is set to one for nine four-digit NAICS industries where their network effects scores are higher than the median.[9] The fixed effects model estimates that incorporate this variable are reported in column (2) of Table 5. We observe that the estimated coefficient of the interaction *NET* × *Strong Network Effects Industries* is positive (β = 0.578) and statistically significant ($p < 0.05$). We also present split-sample analyses comparing firms with strong NE with the ones with weak NE (in columns (3) and (4), respectively, of Table 5) using the binary variable of *Strong Network Effects Industries* to divide the sample. The results from column (3) of Table 5 show that a one-SD increase in NET is associated with an approximately 0.61-percentage point decrease in R&D intensity (relative to an average R&D intensity of 11.35 percentage points of firms in weak NE industries), translating to a reduction of $11.69 million in R&D investment based on the mean level of total asset in the weak NE sample. We also find that only firms in weak NE industries significantly reduce their R&D investments in response to NET, whereas firms operating in high-NE industries seem to be insensitive to NET. These results support the notion that firms operating in industries characterized by strong network effects face significant first-mover benefits and high costs of catching up ex post. Therefore, they are more likely to invest in R&D than firms in weak NE markets when facing new entry threats (Weeds 2002). The presence of network effects is thus a significant moderating factor of the relationship between NET and R&D spending.

4.3.2. Technological Cumulativeness. To test the moderating effect of TC on the relationship between NET and R&D investment, we use the rate of backward self-citations of patents as a measure of technological cumulativeness (Oriani and Sobrero 2008). We first use the continuous firm-level measure of technological cumulativeness and interact this variable with the NET measure. The results from the fixed effects panel data model are presented in column (1) of Table 6. In column (2) of Table 6, we report the results using the binary industry-level measure of technological cumulativeness, where the binary variable, *High-Tech Cumulativeness Industries*, is

Table 4. New Entry Threats and R&D Investments

			Dynamic panel models	
Dependent variable: *R&D Intensity* (%)	Fixed effects (1)	Random effects (2)	NET as exogenous (3)	NET as endogenous (4)
NET	−0.350**	−0.078	−0.335*	−1.334***
	(0.144)	(0.114)	(0.173)	(0.335)
L.R&D Intensity			0.119***	0.122***
			(0.040)	(0.039)
ln*(Sales)*	0.153	0.034	0.864*	1.012**
	(0.365)	(0.113)	(0.515)	(0.504)
ln*(Age)*	1.084*	0.279	1.304*	0.881
	(0.554)	(0.303)	(0.767)	(0.768)
ROA	−26.454***	−27.007***	−28.648***	−28.720***
	(2.424)	(0.294)	(2.956)	(2.952)
PPE/Assets	16.152***	9.609***	14.880***	14.853***
	(2.426)	(1.285)	(3.321)	(3.328)
Leverage	−2.042	−2.874***	−3.046	−2.987
	(1.324)	(0.559)	(1.904)	(1.908)
Capx/Assets	8.150*	7.904***	12.941*	13.032*
	(4.561)	(2.423)	(7.013)	(7.049)
Tobin's Q	0.179**	0.158***	0.195	0.194
	(0.087)	(0.029)	(0.121)	(0.121)
KZ Index	0.096	0.093***	0.006	0.003
	(0.086)	(0.029)	(0.057)	(0.057)
Competition	0.053	−1.588***	−0.578	−0.827
	(0.535)	(0.476)	(0.552)	(0.561)
Observed Entries	−0.001	−0.003*	−0.002	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Product Diversification	−0.046	−0.737**	−0.276	−0.284
	(0.443)	(0.360)	(0.436)	(0.436)
Firm fixed effects	Yes	—		
Year fixed effects	Yes	Yes	Yes	Yes
Observations	14,410	14,410	10,490	10,490
No. of firms	2,101	2,101	1,557	1,557
Adjusted R^2	0.761		—	—

Notes. This table reports the estimates for *R&D Intensity* as dependent variables. The sample was constructed based on the sample of U.S. public firms in the IT industries from 1997 to 2013. Robust standard errors are in parentheses. New entry threats are standardized with mean of 0 and SD of 1. The dynamic model in column (3) treats *L.R&D Intensity* as an endogenous variable. The model in column (4) assumes that both *NET* and *L.R&D Intensity* are endogenous variables. Instruments for differenced equation: GMM type, *L(2/.).NET*, *L(2/.).L.R&D Intensity* (i.e., all available lags from lag2 onward and all differences of exogenous variables, including year dummies). The Arellano–Bond test results for zero autocorrelation in the first differences errors are as follows: In the model where NET is treated as exogenous, the z-value is −3.69 in AR(1) ($Pr > z = 0.00$) and the z-value is −1.85 for AR(2) ($Pr > z = 0.065$). In the model where NET is endogenous, the z-value is −3.79 for AR(1) ($Pr > z = 0.00$) and the z-value is −1.89 for AR(2) ($Pr > z = 0.058$).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

constructed based on four-digit NAICS codes as described earlier.

Here again, we find that firms that operate in fields with high technological cumulativeness are less likely to cut their R&D investments when facing a high level of NET, in contrast to those that operate in fields with low technological cumulativeness. Furthermore, effect size calculations show that, for firms with technological cumulativeness at the first quartile of the sample (low TC with 0% patent self-citation), a one-SD increase in NET is associated with a decrease of $19.63 million ($p < 0.01$) in R&D investment based on the mean value of total asset in the sample. However, the effect size for firms with third quartile level of technological cumulativeness (high TC with 9.2% patent self-citation) is only a $6.62 million ($p < 0.1$) reduction in R&D investment, indicating that firms experiencing high levels of technological cumulativeness are less likely to reduce their R&D investment as a result. Note that all of these marginal effects are over and above those associated with competition and realized entry, the other two sources of product market threats.

Table 5. Boundary Condition—Network Effects

Dependent variable: *R&D Intensity* (%)	Network effects[a] continuous measure (1)	Network effects[b] binary measure (2)	Weak NE industries (3)	Strong NE industries (4)
NET	−1.191**	−0.648***	−0.606***	−0.011
	(0.500)	(0.160)	(0.177)	(0.231)
NET × *Network Effects*	0.113*	—	—	—
	(0.063)	—	—	—
NET × *Strong NE Industries*	—	0.578**	—	—
	—	(0.274)	—	—
ln(*Sales*)	−0.072	0.167	0.678	−0.214
	(0.383)	(0.366)	(0.508)	(0.568)
ln(*Age*)	0.890	1.067*	0.425	3.161***
	(0.606)	(0.554)	(0.688)	(0.954)
ROA	−25.675***	−26.456***	−31.340***	−21.837***
	(2.512)	(2.424)	(3.835)	(2.836)
PPE/Assets	18.286***	16.064***	11.321***	30.210***
	(2.757)	(2.427)	(2.138)	(6.951)
Leverage	−2.030	−2.031	−3.242	−0.553
	(1.461)	(1.323)	(2.027)	(1.340)
Capx/Assets	8.850*	8.239*	5.244	12.070*
	(4.973)	(4.562)	(5.517)	(6.809)
Tobin's Q	0.163*	0.181**	0.430***	0.057
	(0.088)	(0.087)	(0.117)	(0.097)
KZ Index	0.103	0.096	0.076	0.107
	(0.092)	(0.086)	(0.082)	(0.149)
Competition	0.012	0.059	0.618	−0.379
	(0.592)	(0.535)	(0.601)	(1.020)
Observed Entries	−0.002	−0.001	0.002	−0.002
	(0.002)	(0.002)	(0.003)	(0.003)
Product Diversification	0.364	−0.027	0.388	−0.470
	(0.429)	(0.443)	(0.570)	(0.590)
Year and firm fixed effects	Yes	Yes	Yes	Yes
Observations	13,228	14,410	8,797	5,613
No. of firms	1,942	2,101	1,135	966
Adjusted R^2	0.753	0.761	0.762	0.763

Notes. This table reports the estimates for *R&D Intensity* as dependent variables. The sample was constructed based on the sample of U.S. public firms in the IT industries from 1997 to 2013. Robust standard errors are in parentheses. New entry threats are standardized with mean of 0 and SD of 1.

[a]Continuous measures of network effects are collected from Wang et al. (2010).

[b]We identify products with strong network effects following Srinivasan et al. (2004) and Wang et al. (2010) as telecommunication devices and personal computer, operating system and software, internet services provider, personal data assistant services, etc., including nine four-digit NAICS: 5112, 5181, 5182, 5173, 5413, 5415, 5416, and 5417.

$*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

The use of the industry-level measure of technological cumulativeness also confirms our finding. In column (2) of Table 6, the coefficient of the interaction *NET* × *High-Tech Cumulativeness Industries* is positive and significant ($p < 0.1$), showing that firms in high-TC industries make greater R&D investments than those operating in low-TC industries. Moreover, the estimates in column (2) of Table 6 indicate that firms in high-TC industry do not seem to reduce their R&D investments when facing high NET; indeed, the effect of NET on R&D for this group is statistically insignificant ($p = 0.643$). To compare the differences in the marginal effects of NET between the two groups of firms, the subsample analyses, divided by the binary variable of *High-Tech Cumulativeness Industries*, are reported in columns (3) and (4) of Table 6. The results show that, for firms in low-TC industries (column (3) of Table 6), a one-SD increase in NET is associated with a 0.45-percentage point decrease in R&D intensity (relative to the average R&D intensity of 13.00 percentage points for firms in this subsample), translating to a reduction of \$8.01 million in R&D investment based on the mean level of total asset in the low-TC sample. By contrast, firms in high-TC industries (column (4) of Table 6) seem to be less sensitive and do not significantly reduce their R&D investments in response to new entry threats. These results provide empirical evidence for the role of technology cumulativeness as a boundary condition that shapes the relationship between NET and R&D spending.

5. Discussion and Conclusion

It is well known that innovation is one of the building blocks of competitive advantage in the IT industry (Giarratana 2004), and R&D investments, as an input of the innovation process, represent particularly important managerial decisions (Schwartz and Moon 2000). It is also established that the IT industry tends to be volatile, where creative destruction often emerges from entry by new entrepreneurial ventures backed by venture capital (McAfee and Brynjolfsson 2008). Although product market threats, in the form of competition and observed entry, have been studied extensively in the literature, there is little work that studies the role of new entry threats per se, a notable gap given the relevance of this construct in the IT industry. Our work here thus addresses an important question: how do IT incumbents adjust their R&D investments in response to increasing threats of new entry? This question remains understudied in IS research for several reasons. First, there is a lack of established measures of new entry threats, which presents a significant challenge to the empirical studies in this area. Second, the effect of new entry threat is easily conflated with other forms of product market threats, such as competition or observed entry. Third, there are significant heterogeneities in terms of the effectiveness of preemptive R&D as a response to NET among firms in the IT industries, adding to the empirical complexity.

In this paper, we focus our attention on the relationship between NET and R&D investments by overcoming some of these challenges. First, we develop a text-based measure of new entry threats by analyzing the product descriptions of both incumbent firms and startups. We conduct a series of validation tests and show that the NET measure indeed captures impending threats from the startup space. Second, we specifically control for

Table 6. Boundary Condition—Technological Cumulativeness

Dependent variable: *R&D Intensity* (%)	Firm-level TC measure (1)	TC by four-digit NAICS (2)	Low-TC industries (3)	High-TC industries (4)
NET	−0.830***	−0.466***	−0.454***	0.056
	(0.193)	(0.166)	(0.170)	(0.264)
NET × *Tech Cumulativeness*	5.887***	—	—	—
	(1.970)	—	—	—
NET × *High-TC Industries*	—	0.586*	—	—
	—	(0.302)	—	—
ln(*Sales*)	0.547	0.157	0.138	−0.010
	(0.444)	(0.366)	(0.430)	(0.557)
ln(*Age*)	0.794	1.112**	1.001	2.166**
	(0.677)	(0.553)	(0.638)	(1.072)
ROA	−30.678***	−26.457***	−27.318***	−22.797***
	(3.227)	(2.425)	(2.908)	(2.995)
PPE/Assets	14.061***	16.260***	14.970***	22.967***
	(2.735)	(2.423)	(2.593)	(6.507)
Leverage	−3.633**	−2.095	−3.502**	2.444
	(1.776)	(1.327)	(1.651)	(1.671)
Capx/Assets	5.619	7.882*	3.680	20.679***
	(5.470)	(4.562)	(5.293)	(7.524)
Tobin's Q	0.166*	0.176**	0.163	0.169
	(0.095)	(0.087)	(0.104)	(0.138)
KZ Index	0.190	0.105	0.143	0.046
	(0.143)	(0.088)	(0.129)	(0.090)
Competition	−0.300	0.080	0.073	0.592
	(0.605)	(0.535)	(0.581)	(1.132)
Observed Entries	−0.001	−0.001	0.002	−0.011**
	(0.002)	(0.002)	(0.002)	(0.005)
Product Diversification	0.095	−0.043	−0.106	0.159
	(0.547)	(0.443)	(0.529)	(0.607)
Year and firm fixed effects	Yes	Yes	Yes	Yes
Observations	10,678	14,400	11,075	3,335
No. of firms	1,313	2,100	1,578	523
Adjusted R^2	0.781	0.761	0.763	0.760

Notes. This table reports the estimates for *R&D Intensity* as dependent variables. The sample was constructed based on the sample of U.S. public firms in the IT industries from 1997 to 2013. Robust standard errors are in parentheses. New entry threats are standardized with mean of 0 and SD of 1.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

competition and realized entry in our analyses, thereby allowing the incremental effects of new entry threats to be estimated separately. Third, we study the role of a couple of boundary conditions that are particularly salient in the IT industry, thereby revealing some sources of the heterogeneities observed in this relationship. Using data on a panel of 2,101 firms in the U.S. IT industry over the period 1997–2013, we find that, on average, higher levels of new entry threats are associated with lower R&D spending within the IT industry. Our finding is robust to a number of alternative regression specifications addressing the endogeneity of NET. We also find that firms operating in industries with strong network effects or facing high levels of technological cumulativeness do not reduce their R&D investments as much when facing NET; indeed, in some contexts, they may even show net increases in R&D spending as a result.

We attribute the finding of a negative average association between NET and R&D investments to the "cautionary effects" of uncertainty that have been well documented in prior literature (Bulan 2005, Bloom et al. 2007). Most IT firms are likely to take a "wait-and-see" approach to investment decisions in response to the heightened uncertainties associated with new entry threats. From an industry lifecycle standpoint (Klepper 1996), time periods associated with high threats of new entry into the market are often associated with impending technological shifts, wherein a multitude of competing designs and standards emerges and a dominant design has yet to be established (Jovanovic and MacDonald 1994). During these times, delaying R&D investments can be of particular value, because they allow the firm to avail of flexibility while avoiding premature investments that may not pay off in the future. However, we also show a number of exceptions to this observation: IT firms operating in markets with strong network effects, with the associated winner-takes-all dynamics, may choose to respond more aggressively in the face of NET. Similarly, to the extent that innovation has path dependence, a more proactive stance may be warranted, which we see in the case of technological cumulativeness.

Our research here makes a significant theoretical contribution by paying greater attention to the innovation-related investments made by firms within the IT industry, especially given the fast-paced, dynamic environment within this industry sector. We extend the study of product market threats, which has hitherto placed its emphasis mostly on realized entry and competition, to evaluating the role of new entry threats on firm-level R&D decision making. The overriding emphasis in this research on actual entry (Aghion et al. 2009) or competition (Aghion et al. 2005, Kim et al. 2016) is not altogether surprising given the empirical challenges in measuring new entry threats. A few researchers have tried to address the role of NET more directly: for example, Goolsbee and Syverson (2008) and Seamans (2013) represent rare and notable exceptions, but both are limited to specialized contexts where incumbent responses are observed mostly through pricing. However, in the IT industry, an arguably more effective response may be through the preemption of innovation, because the market dynamics are primarily driven by the introduction of new products and services, where pricing strategies play a secondary role. Furthermore, given the prominent role played by the startup ecosystem within the IT industry, our work brings together the threats that startups represent to incumbents as well as the strategic responses that such incumbents may choose, opening up new avenues for additional theory building.

Our work also draws linkages with research studying competition and R&D spending, recognizing that some new entry threats will become realized entries and eventually part of the competition. The relationship between competition and R&D spending is not without ambiguity as discussed earlier. Within the IT industry specifically, recent work studying the impact of competition provides results that are complementary to those that we present here. For example, Kim et al. (2016) show that, when faced with competition, IT firms are likely to invest more in R&D per se, even if some of these additional investments shift to more flexible options, like corporate venture capital. We show that, when competition has not materialized but appears merely as a threat, firms are more inclined, on average, to wait and see what the future may bring rather than commit investments prematurely into R&D. A number of mechanisms may be behind these differences, including the nature of R&D in different stages of technology lifecycles (Abernathy and Utterback 1978), the tradeoffs between preemptive R&D and investments in complementary assets (Teece 1986), or the existence of markets for technology (Arora et al. 2001). Nevertheless, we see some commonalities between the responses by IT firms to competition as observed in the literature and those to new entry threats as we show here, such as their preference for more flexible options as environmental uncertainty increases.

In addition to theoretical contributions, we make a methodological contribution by creating and validating a new measurement of new entry threat from the startup ecosystem. Our text-mining approach, in contrast to earlier measurement of market threats based on industry classifications or market shares, not only captures forward-looking threats in a firm's competitive environment but also, changes over time as new ventures are funded and incumbents change their product and service offerings. In this spirit, our NET measure is similar to the TNIC industry classification developed by Hoberg and Phillips (2016). There are several empirical and theoretical contexts where new entry threats faced by incumbents play a central role; the availability of a standard and accepted measure will help by allowing for comparability across models and theories.

Our work here also paves the way for future research in related areas. As mentioned earlier, new entry threats as a construct have seen significant theoretical development (Porter 2008), but empirical research has been relatively sparse. We hope that the availability of such a measure will lead to growing interests in empirical work addressing the role of NET on various outcomes, such as mergers and acquisitions, product pricing strategies, corporate governance, and IT investment choices. Furthermore, innovation-related decisions regarding patent applications, technological alliances, and licensing agreements are all made under varying degrees of new entry threats within the IT industry. We hope that our work will kick start additional work that empirically examines these related topics.

Acknowledgments

The authors would like to thank the senior editor and the associate editor for their support of this paper. Earlier versions of the paper were presented at the University of Minnesota, University of Washington, Georgia Tech, Louisiana State University, and Boston College; the authors thank the participants for their feedback. The new entry threats data set that is described in this paper is available for download for academic research only at: https://sites.google.com/view/newentrythreats/home.

Endnotes

[1] See https://www.imd.org/publications/articles/the-battle-for-digital-disruption-startups-vs-incumbents.

[2] A full set of NET values across all high-tech industry sectors is available on request from the authors.

[3] Military armored vehicle manufacturing industry is the only one that is closely related to the military goods and services industry in the high-tech sector (Hecker 2005).

[4] The 15 four-digit NAICS industries include machinery manufacturing, communication and audio/video equipment manufacturing, semiconductor equipment manufacturing, software publishers, computer and peripheral equipment manufacturing, internet service providers and web search portals, data processing, hosting and related services, etc. The highest three network effect industries are data processing, computer design, and scientific-related services (with a score of 10.7); telecommunications resellers (10); and software publishers (9.05). The lowest three network effect industries are commercial and service industry machinery manufacturing (3.9); industrial machinery manufacturing (4.1); and navigational, measuring, electromedical, and control instruments manufacturing (5.8).

[5] We use the median value if multiple network effects indexes from different products are mapped to same industry.

[6] In our sample, the top three NAICS4-based tech cumulativeness industries are data processing, hosting, and related services; computer and peripheral equipment manufacturing; and industrial machinery manufacturing.

[7] Because all Compustat firms do not report sales by lines of business, we are able to calculate this variable only for 954 firms.

[8] We experimented with different lags and their combinations; results are fully robust to varying lag structures.

[9] Specially, the nine high network industries are defined with four-digit NAICS codes: 5112, 3341, 5181, 5182, 5173, 5413, 5415, 5416, and 5417, including software publishers; computer and peripheral equipment; internet service providers and web search portals; data processing, hosting, and related services; telecommunications resellers; etc.

References

Abernathy WJ, Utterback JM (1978) Patterns of industrial innovation. *Tech. Rev.* 80(7):40–47.

Abrahamson E (1991) Managerial fads and fashions: The diffusion and rejection of innovations. *Acad. Management Rev.* 16(3):586–612.

Aghion P (2003) *Empirical Estimates of the Relationship Between Product Market Competition and Innovation. Institutions, Innovation and Growth: Selected Economic Papers* (Edward Elgar Publishing, Paris), 142–169.

Aghion P, Howitt P (1992) A model of growth through creative destruction. *Econometrica* 60(2):323–351.

Aghion P, Bloom N, Blundell R, Griffith R, Howitt P (2005) Competition and innovation: An inverted-u relationship. *Quart. J. Econom.* 120(2):701–728.

Aghion P, Blundell R, Griffith R, Howitt P, Prantl S (2009) The effects of entry on incumbent innovation and productivity. *Rev. Econom. Statist.* 91(1):20–32.

Aizawa A (2003) An information-theoretic perspective of tf–idf measures. *Inform. Processing Management* 39(1):45–65.

Aral S, Van Alstyne M (2011) The diversity-bandwidth trade-off. *Amer. J. Sociol.* 117(1):90–171.

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Arora A, Fosfuri A, Gambardella A (2001) Markets for technology and their implications for corporate strategy. *Indust. Corporate Change* 10(2):419–451.

Arrow K (1962) Economic welfare and the allocation of resources for invention. Universities-National Bureau Committee for Economic Research, Committee on Economic Growth of the Social Science Research Council, eds. *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton University Press, Princeton, NJ), 609–626.

Atanassov J (2013) Do hostile takeovers stifle innovation? Evidence from antitakeover legislation and corporate patenting. *J. Finance* 68(3):1097–1131.

Audretsch DB, Mata J (1995) The post-entry performance of firms: Introduction. *Internat. J. Indust. Organ.* 13(4):413–419.

Bardhan I, Krishnan V, Lin S (2013) Research note—business value of information technology: Testing the interaction effect of it and R&D on Tobin's q. *Inform. Systems Res.* 24(4):1147–1161.

Becker-Blease JR (2011) Governance and innovation. *J. Corporate Finance* 17(4):947–958.

Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascades. *J. Econom. Perspect.* 2(3):151–170.

Bird S, Klein E, Loper E (2009) *Natural Language Processing with Python* (O'Reilly Media, Inc., Sebastopol, CA).

Blonigen BA, Taylor CT (2000) R&D intensity and acquisitions in high-technology industries: Evidence from the U.S. electronic and electrical equipment industries. *J. Indust. Econom.* 48(1):47–70.

Bloom N, Bond S, Van Reenen J (2007) Uncertainty and investment dynamics. *Rev. Econom. Stud.* 74(2):391–415.

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics* 87(1):115–143.

Blundell R, Griffith R, Van Reenen J (1999) Market share, market value and innovation in a panel of British manufacturing firms. *Rev. Econom. Stud.* 66(3):529–554.

Brav A, Graham JR, Harvey CR, Michaely R (2005) Payout policy in the 21st century. *J. Financial Econom*. 77(3):483–527.

Breschi S, Malerba F (1997) *Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics, and Spatial Boundaries* (Frances Pinter, London).

Brown LD, Caylor ML (2006) Corporate governance and firm valuation. *J. Accounting Public Policy* 25(4):409–434.

Brynjolfsson E, McAfee A (2011) *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy* (Digital Frontier Press, Lexington, MA).

Bulan LT (2005) Real options, irreversible investment and firm uncertainty: New evidence from us firms. *Rev. Financial Econom*. 14(3):255–279.

Caves RE, Porter ME (1977) From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *Quart. J. Econom*. 91(2):241–261.

Chandy RK, Tellis GJ (2000) The incumbent's curse? Incumbency, size, and radical product innovation. *J. Marketing* 64(3):1–17.

Chemmanur TJ, Tian X (2018) Do antitakeover provisions spur corporate innovation? A regression discontinuity analysis. *J. Financial Quant. Anal*. 53(3):1163–1194.

Chen M-J, Smith KG, Grimm CM (1992) Action characteristics as predictors of competitive responses. *Management Sci*. 38(3):439–455.

Cockburn IM, MacGarvie MJ (2011) Entry and patenting in the software industry. *Management Sci*. 57(5):915–933.

Cohen S, Hochberg Y (2014) Accelerating startups: The seed accelerator phenomenon. Working paper, University of Georgia, Athens.

Cohen WM, Levinthal DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart*. 35(1): 128–152.

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552, National Bureau of Economic Research, Cambridge, MA.

Colombo MG, Delmastro M (2002) How effective are technology incubators?: Evidence from Italy. *Res. Policy* 31(7):1103–1122.

Cornaggia J, Mao Y, Tian X, Wolfe B (2015) Does banking competition affect innovation? *J. Financial Econom*. 115(1):189–209.

Czarnitzki D, Toole AA (2011) Patent protection, market uncertainty, and R&D investment. *Rev. Econom. Statist*. 93(1):147–159.

Czarnitzki D, Toole AA (2013) The R&D investment–uncertainty relationship: Do strategic rivalry and firm size matter? *Managerial Decision Econom*. 34(1):15–28.

Dasgupta P, Stiglitz J (1980) Industrial structure and the nature of innovative activity. *Econom. J*. 90(358):266–293.

Dixit AK, Stiglitz JE (1977) Monopolistic competition and optimum product diversity. *Amer. Econom. Rev*. 67(3):297–308.

Evans DS (2003) Some empirical aspects of multi-sided platform industries. *Rev. Network Econom*. 2(3):1–19.

Fine CH (1998) *Clockspeed: Winning Industry Control in the Age of Temporary Advantage* (Perseus Books, Reading, MA).

Fudenberg D, Tirole J (1984) The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *Amer. Econom. Rev*. 74(2):361–366.

Gatignon H, Anderson E, Helsen K (1989) Competitive reactions to market entry: Explaining interfirm differences. *J. Marketing Res*. 26(1):44–55.

Geroski PA (1989) Entry, innovation and productivity growth. *Rev. Econom. Statist*. 71(4):572–578.

Ghosal V, Loungani P (2000) The differential impact of uncertainty on investment in small and large businesses. *Rev. Econom. Statist*. 82(2):338–343.

Giarratana MS (2004) The birth of a new industry: Entry by start-ups and the drivers of firm growth: The case of encryption software. *Res. Policy* 33(5):787–806.

Gilbert RJ, Newbery DM (1982) Preemptive patenting and the persistence of monopoly. *Amer. Econom. Rev*. 72(3):514–526.

Gimeno J (2004) Competition within and between networks: The contingent effect of competitive embeddedness on alliance formation. *Acad. Management J*. 47(6):820–842.

Gompers P, Lerner J (2001) The venture capital revolution. *J. Econom. Perspect*. 15(2):145–168.

Goolsbee A, Syverson C (2008) How do incumbents respond to the threat of entry? Evidence from the major airlines. *Quart. J. Econom*. 123(4):1611–1633.

Green JR, Scotchmer S (1995) On the division of profit in sequential innovation. *RAND J. Econom*. 26(1):20–33.

Greene WH (2003) *Econometric Analysis* (Prentice Hall, Upper Saddle River, NJ).

Grossman GM, Helpman E (1991) Quality ladders in the theory of growth. *Rev. Econom. Stud*. 58(1):43–61.

Gunther McGrath R, Nerkar A (2004) Real options reasoning and a new look at the R&D investment strategies of pharmaceutical firms. *Strategic Management J*. 25(1):1–21.

Hall BH (1988) The effect of takeover activity on corporate research and development. Auerbach AJ, ed. *Corporate Takeovers: Causes and Consequences* (University of Chicago Press, Chicago), 69–100.

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498, National Bureau of Economic Research, Cambridge, MA.

Ham Ziedonis R, Hall BH (2001) The effects of strengthening patent rights on firms engaged in cumulative innovation: Insights from the semiconductor industry. Libecap GD, ed. *Entrepreneurial Inputs and Outcomes: New Studies of Entrepreneurship in the United States*, Advances in the Study of Entrepreneurship, Innovation & Economic Growth, vol. 13 (Emerald Group Publishing Limited, Bingley, UK), 133–187.

Haupt R, Kloyer M, Lange M (2007) Patent indicators for the technology life cycle development. *Res. Policy* 36(3):387–398.

Hecker DE (2005) High-technology employment: A NAICS-based update. *Monthly Labor Rev*. 128(7):57–72.

Hiemstra D (2000) A probabilistic justification for using tf×idf term weighting in information retrieval. *Internat. J. Digital Libraries* 3(2):131–139.

Hill CW, Rothaermel FT (2003) The performance of incumbent firms in the face of radical technological innovation. *Acad. Management Rev*. 28(2):257–274.

Hoberg G, Phillips G (2010) Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Rev. Financial Stud*. 23(10):3773–3811.

Hoberg G, Phillips GM (2016) Text-based network industries and endogenous product differentiation. *J. Political Econom*. 124(5): 1423–1465.

Hoberg G, Phillips G, Prabhala N (2014) Product market threats, payouts, and financial flexibility. *J. Finance* 69(1):293–324.

Hsu DH (2008) Technology-based entrepreneurship. Shane S, ed. *Handbook of Technology and Innovation Management* (Wiley, West Sussex, UK), 367–388.

Huchzermeier A, Loch CH (2001) Project management under risk: Using the real options approach to evaluate flexibility in R&D. *Management Sci*. 47(1):85–101.

Jacquemin AP, Berry CH (1979) Entropy measure of diversification and corporate growth. *J. Indust. Econom*. 27(4):359–369.

Jorgenson DW, Stiroh KJ (1999) Information technology and growth. *Amer. Econom. Rev*. 89(2):109–115.

Jorgenson DW, Stiroh KJ, Gordon RJ, Sichel DE (2000) Raising the speed limit: U. S. economic growth in the information age. *Brookings Papers Econom. Activity* 31(1):125–235.

Jovanovic B, MacDonald G (1994) The life-cycle of a competitive industry. *J. Political Econom*. 102(2):322–347.

Kaplan SN, Zingales L (1997) Do investment-cash flow sensitivities provide useful measures of financing constraints? *Quart. J. Econom*. 112(1):169–215.

Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. *Amer. Econom. Rev*. 75(3):424–440.

Khessina OM, Carroll GR (2008) Product demography of de novo and de alio firms in the optical disk drive industry, 1983-1999. *Organ. Sci*. 19(1):25–38.

Kim K, Gopal A, Hoberg G (2016) Does product market competition drive cvc investment? Evidence from the U.S. IT industry. *Inform. Systems Res*. 27(2):259–281.

King JL, Gurbaxani V, Kraemer KL, McFarlan FW, Raman K, Yap C-S (1994) Institutional factors in information technology innovation. *Inform. Systems Res*. 5(2):139–169.

Klepper S (1996) Entry, exit, growth, and innovation over the product life cycle. *Amer. Econom. Rev*. 86(3):562–583.

Klepper S, Simons KL (1997) Technological extinctions of industrial firms: An inquiry into their nature and causes. *Indust. Corporate Change* 6(2):379–460.

Kristiansen EG (1998) R&D in the presence of network externalities: Timing and compatibility. *RAND J. Econom*. 29(3):531–547.

Kuester S, Homburg C, Robertson TS (1999) Retaliatory behavior to new product entry. *J. Marketing* 63(4):90–106.

Kuratko DF, Menter M (2017) The role of public policy in fostering technology-based nascent entrepreneurship. Cunningham J, O'Kane C, eds. *Technology-Based Nascent Entrepreneurship*, Palgrave Advances in the Economics of Innovation and Technology (Palgrave Macmillan, New York), 19–52.

Lanjouw JO, Schankerman M (2001) Characteristics of patent litigation: A window on competition. *RAND J. Econom*. 32(1):129–151.

Lukach R, Kort PM, Plasmans J (2007) Optimal R&D investment strategies under the threat of new technology entry. *Internat. J. Indust. Organ*. 25(1):103–119.

Markides C (2006) Disruptive innovation: In need of better theory. *J. Product Innovation Management* 23(1):19–25.

McAfee A, Brynjolfsson E (2008) Investing in the it that makes a competitive difference. *Harvard Bus. Rev*. 86(7/8):98–107.

McGrath RG (1997) A real options logic for initiating technology positioning investments. *Acad. Management Rev*. 22(4):974–996.

Mendelson H, Pillai RR (1998) Clockspeed and informational response: Evidence from the information technology industry. *Inform. Systems Res*. 9(4):415–433.

Mendelson H, Pillai RR (1999) Industry clockspeed: Measurement and operational implications. *Manufacturing Service Oper. Management* 1(1):1–20.

Merges RP, Nelson RR (1990) On the complex economics of patent scope. *Columbia Law Rev*. 90(4):839–916.

Mitchell W, Singh K (1992) Incumbents' use of pre-entry alliances before expansion into new technical subfields of an industry. *J. Econom. Behav. Organ*. 18(3):347–372.

Nickell SJ (1996) Competition and corporate performance. *J. Political Econom*. 104(4):724–746.

O'Connor M, Rafferty M (2012) Corporate governance and innovation. *J. Financial Quant. Anal*. 47(2):397–413.

Oriani R, Sobrero M (2008) Uncertainty and the market valuation of R&D within a real options logic. *Strategic Management J*. 29(4):343–361.

Pauwels C, Clarysse B, Wright M, Van Hove J (2016) Understanding a new generation incubation model: The accelerator. *Technovation* 50:13–24.

Porter ME (2008) The five competitive forces that shape strategy. *Harvard Bus. Rev*. 86(1):25–40.

Reinganum JF (1983) Uncertain innovation and the persistence of monopoly. *Amer. Econom. Rev*. 73(4):741–748.

Rigby DK, Christensen CM, Johnson M (2002) Foundations for growth: How to identify and build disruptive new businesses. *MIT Sloan Management Rev*. 43(3):22–32.

Roodman D (2009) How to do xtabond2: An introduction to difference and system gmm in stata. *Stata J*. 9(1):86–136.

Sambamurthy V, Bharadwaj A, Grover V (2003) Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS Quart*. 27(2):237–263.

Sarasvathy SD (2001) Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Acad. Management Rev*. 26(2):243–263.

Schumpeter JA (1942) *Capitalism, Socialism and Democracy* (Harper & Row, New York).

Schwartz ES, Moon MA (2000) *Evaluating Research and Development Investments* (Oxford University Press, New York).

Seamans RC (2013) Threat of entry, asymmetric information, and pricing. *Strategic Management J*. 34(4):426–444.

Sebastiani F (2002) Machine learning in automated text categorization. *ACM Comput. Surveys* 34(1):1–47.

Shapiro C, Varian HR (1999) The art of standards wars. *Calif. Management Rev*. 41(2):8–32.

Spence AM (1977) Entry, capacity, investment and oligopolistic pricing. *Bell J. Econom*. 8(2):534–544.

Srinivasan R, Lilien GL, Rangaswamy A (2004) First in, first out? The effects of network externalities on pioneer survival. *J. Marketing* 68(1):41–58.

Tang J (2006) Competition and innovation behaviour. *Res. Policy* 35(1):68–82.

Tata S, Patel JM (2007) Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM SIGMOD Record* 36(2):7–12.

Teece DJ (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Res. Policy* 15(6):285–305.

Tetlock PC (2011) All the news that's fit to reprint: Do investors react to stale information? *Rev. Financial Stud*. 24(5):1481–1512.

Tetlock PC, Saar-Tsechansky M, Macskassy S (2008) More than words: Quantifying language to measure firms' fundamentals. *J. Finance* 63(3):1437–1467.

Utterback J (1994) *Mastering the Dynamics of Innovation* (Harvard Business School Press, Boston).

Wang Q, Chen Y, Xie J (2010) Survival in markets with network effects: Product compatibility and order-of-entry effects. *J. Marketing* 74(4):1–14.

Weeds H (2002) Strategic delay in a real options model of R&D competition. *Rev. Econom. Stud*. 69(3):729–747.

Wiggins RR, Ruefli TW (2005) Schumpeter's ghost: Is hypercompetition making the best of times shorter? *Strategic Management J*. 26(10):887–911.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

Zhu K, Kraemer KL, Gurbaxani V, Xu SX (2006) Migration to open-standard interorganizational systems: Network effects, switching costs, and path dependency. *MIS Quart*. 30(Special Issue):515–539.