New Entry Threats, Firm Governance, and Innovation in the U.S. IT Industry

Completed Research Paper

Yang Pan
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742-1815 U.S.A.
ypan@rhsmith.umd.edu

Peng Huang
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742-1815 U.S.A.
huang@umd.edu

Anandasivam Gopal
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742-1815 U.S.A.
agopal@rhsmith.umd.edu

Abstract

We examine how firms in the U.S.-based IT industry adjust their innovation strategy based on two factors that have been under-studied in the literature: the potential new entry threats the firm faces in its product market, and the firm's governance structure. We argue that: 1) turbulence in the product markets caused by threat from entrepreneurial startups affects the focal firm's investments in long-term risky activities such as innovation, and 2) this relationship is moderated by the extent to which management of the firm is protected through corporate governance. While extant research has discussed the potential disruptions that incumbent firms may experience from start-ups and their technologies, the absence of acceptable measures of industry classification for startups as well as the inability to accurately gauge when they represent a credible threat has limited empirical research into this question. We contribute a new measure to identify these threats through text analyses based on product descriptions provided by incumbent firm 10-K filings and product blurbs provided by start-ups. This new measure differs significantly from approaches that use static industry classifications, which are backward-looking and do not fully account for industry evolution over time, thereby not capturing the threat of new entry. We measure the degree to which managers are protected through corporate governance using the G-Index (Gompers et al. 2003). Our analyses, performed on a sample of IT firms during the period 1997-2007, show that incumbent firms react to new entry threats by systematically reducing innovation. Moreover, more governance provisions protecting managers amplify this effect: protected managers reduce innovation to a greater extent when the firm faces intensive new entry threats. We discuss the implications for research and practice.

Keywords: Innovation, corporate governance, entrenched management, new entry threats, disruptive technology, text mining
Introduction

The Information Technology (IT) industry plays a critical role in the economic growth in the past two decades (Jorgenson and Stiroh 1999; Jorgenson et al. 2000). This industry is characterized by constant technological changes and fast clockspeed (Fine 1998). The rate of changes in the industry’s external environment, including the development of new technologies, shifts in consumer preferences, and fast-moving market dynamics far exceeds those seen in other industries (Brynjolfsson and McAfee 2011, Mendelson and Pillai 1999). The result is a compression of product life-cycles (Mendelson and Pillai 1998), volatility in market structure and a competitive context where advantages, if any, tend to be short-lived (Wiggins and Rueffli 2005). Thus, the ability to develop new product and process innovations is a key driver of a firm’s competitive advantages (Kleis et al. 2012). A significant part of this fast-moving dynamic is fed by the concurrent high rate of firm entry in the form of entrepreneurial ventures; technology startups are constantly bringing new products and business models to the market (Giarratana 2004). Indeed, over 70% of all venture-capital funded startups tend to be associated with the IT industry (Gompers and Lerner 2001), indicating that they contribute significantly to the threat of new entry experienced by incumbent firms.

Intense entrepreneurial activity in the product market of an incumbent can cause turbulence in the market – entrepreneurs incorporating new and innovative technologies, backed by influential venture capitalists and their deep networks, can significantly affect the future market potential of incumbents. In the presence of such threats of new entry, how do incumbents respond in terms of their investments in innovation? This forms the first research question we address. Extant literature suggests that, on the one hand, firms may increase their investments in their innovative activities, thereby building technological and process capabilities to successfully compete in the future as the perceived nascent threats materialize (Lukach et al. 2007; Reinganum 1983). On the other hand, it is also possible that firms eschew risky investments in innovation that, at the best of times, have uncertain payoff and instead choose to either conserve cash (Hoberg et al. 2014; Klepper and Simons 1997) or invest in complementary capabilities vital to the commercialization of innovation, such as manufacturing or marketing (Teece 1986), as a differentiation strategy, should the anticipated threats materialize (O’Connor and Rafferty 2012). In this paper, we address this underlying tension in the literature regarding how firms respond to new entry threat, in terms of their innovation strategy within the IT industry.

While we provide arguments about firm responses, we recognize that firms do not make decisions; people do (March and Shapira 1987). Within firms, managers are responsible to setting directions regarding the firm’s investments. Managers vary in their abilities or inclinations to invest in risky activities with uncertain long-term payoffs, such as innovation (Galasso and Simcoe 2011, Jensen and Meckling 1976). The variations in managerial characteristics lead to agency problems, especially in public firms where control and ownership are divided between shareholders and managers. One mechanism through which these agency problems are managed is the set of customs, rules and institutions collectively referred to as corporate governance (Gompers et al. 2003). The specific form of governance instituted within the firm creates a balance of power between the shareholders and the managers in terms of authority and control, which is likely to have significant implications for innovation strategies. For example, Google, among other recent Information Technology companies such as Facebook, Groupon, and LinkedIn, has adopted a dual-class share structure that gives super-voting stocks to its founders. Such dual-class share structure allegedly allows the founders/managers to be protected so that they can “concentrate on the long term” performance of the firm.1 However, there is a broad debate about whether such protection afforded to managers is beneficial to the firm (Bechuk and Cohen 2005) under the viewpoint that protected managers tend to be insulated from the discipline of the market, leading to inferior firm performance (Bertrand and Mullainathan 2003). Thus, our second research question is: does the presence of protected management moderate the firm’s response to new entry threats in the form of the firm’s innovation strategy?

In this paper, we combine perspectives from new entry threats and corporate governance to understand the implications for IT firm’s innovation strategy. There is likely to be significant interplay between new entry threat and governance for two reasons. First, any theorizing of a firm’s response to threats in its

---

1 http://dealbook.nytimes.com/2012/04/13/new-share-class-gives-google-founders-tighter-control/?_r=0
market by definition has to reflect the incentives of the firm's management to make risky decisions. Therefore, as a first step, the firm's governance regime becomes an important intermediate variable. Second, firm governance effectively represents a boundary condition in understanding the relationship between new entry threats and innovative activity, i.e. how a firm's response to new entry threats may only be relevant when managers have protection, and therefore are long-term thinking. Thus, we posit that it is important to consider the joint influence of new entry threats and firm governance on firm innovation, rather than in a reductionist approach.

Our work provides several contributions to the IS literature on innovation. The rapid rate of technological change observed in the IT industries and the associated volatility in product markets (Brynjolfsson and McAfee 2011) has brought into sharp focus the role of the firm's response to such threats. In addition, a societal focus on innovation-related IT industries entrepreneurship and the development of entrepreneurial ecosystems in the forms of incubators have radically improved the abilities of these new ventures to threaten incumbent firms. Under such contexts, it is important to understand how new entry threats may be accurately measured, and the responses that incumbent firms make to such threats in the form of their own innovative activities. In addition, since managers are usually the first line of defense for incumbents facing such threats, we study how their responses to new entry threats vary according to the extent to which they are empowered within the firm. Our work thus adds to the literature that has focused on the role of corporate governance on innovation (Becker-Blease 2011; Chemmanur and Tian 2012), by considering the role of governance in the face of threats from new entry.

Beyond the theoretical implications, we also contribute by devising and validating a new way of measuring new entry threat from entrepreneurial ventures using text analysis. We build on existing work that has used text to construct measurement schemes (Hoberg and Phillips 2010; Tetlock et al. 2008) by creating text similarity scores between incumbent firms and broad movements in the VC-backed entrepreneurial space. These similarity scores are more suitable for capturing emerging threats from the entrepreneurial space as a whole, rather than those from single entrepreneurs that are easy to ignore or discount.

**Related Literature**

**New Entry Threats and Innovation**

All firms face competition and new entry threats in their native markets. However, these are particularly pronounced and influential in the high-tech sector, specifically in the IT industry, (Fontana and Nesta 2009), where firm survival is often in question as a result of these competitive dynamics. Collectively, these are referred to as *product market threats*, defined as incipient instability and uncertainty in a firm's product market, which can threaten the sustainability of the firm's future earnings as well as viability of its product portfolio (Hoberg et al. 2014).

Within the domain of product market threats, we distinguish new entry threats specifically from competition from other existing competitors for several reasons. First, competition is contemporaneous and is usually measured by market outcomes, such as market share/concentration (Blundell et al. 1999) or Lerner Index (Aghion 2003), while new entry threats are forward-looking since they imperil the "stability and sustainability of future earnings" (Brav et al. 2005). Second, while competition describes equilibrium market outcomes, new entry threat is inherently about disequilibrium, often associated with changes in competitive dynamics (Cockburn and MacGarvie 2011). The ability to threaten future earnings, especially in the technology sector, is closely linked to technological changes and product innovation (Coad and Rao 2008), thereby making prior work in industry lifecycles of particular relevance (Abernathy and Utterback 1978). Industry lifecycles describe the processes that unfold in technologically progressive industries as they evolve from birth to maturity (Klepper 1996). In general terms, when new industries form, there is considerable new entry into the field, firms offer many varieties of similar products, prices are typically high and there is considerable product innovation that occurs within the nascent industry (Jovanovic and MacDonald 1994; Klepper 1996). However, over time, despite steady market growth, new entry into the field ceases, prices drop and there is a shakeout in the number of firms. At this stage, scholars have argued that a dominant design emerges in the industry (Jovanovic and MacDonald 1994) and the incumbent firms shift to process innovations that improve upon the dominant design, leading again to increased competition in the industry. Thus, industry lifecycles are characterized by successive waves of product and process innovations that introduce high volatility within the technology industry.
These dynamics are particularly salient in the IT sector for three reasons. First, prior work in technologically progressive industries has questioned the notion of a dominant design and the clear distinction between periods of process and product innovation (Klepper and Simons 1993). They argue that in many instances, product and process innovations emerge concurrently from different players in the industry, thereby creating product market threats from new venture firms (Abernathy and Utterback 1978). Second, the IT industry incorporates significant levels of heterogeneity in terms of firm capabilities, learning and the ability to capitalize on innovation activities (Agarwal and Gort 2002; Reinganum 1983). Therefore, given this heterogeneity in the composition of the industry, it is unlikely that industry lifecycles can be mapped as cleanly as it was possible in industries that were studied before, such as the tire (Jovanovic and MacDonald 1994) and auto industries (Argyres and Bigelow 2007). Rather, it is more likely that incumbents are bombarded with concurrent and successive waves of uncertainty from ongoing process and product innovation across the industry at any point in time, as has been documented clearly in the IT industry (Rai and Tang 2013). Finally, there is significant evidence for the increasing pace at which industry lifecycles emerge and stabilize, even briefly, in the last two decades particularly within the IT industries (Carrillo 2005; Fine 1998; Mendelson and Pillai 1998). As industry “clock speed” quickens, there is greater threat of new entry for the incumbent, based on both process and product innovations within the field.

Entry threats coming from new entrepreneurial ventures possess the ability to significantly disrupt the incumbent’s market in the future. Indeed, prior work in industry lifecycles establishes that a significant majority of product innovations (i.e. new designs and technologies, rather than process innovations) emerge from the entrepreneurial space (Agarwal and Gort 2002; Prusa and Schmitz Jr 1991; Prusa and Schmitz Jr 1994). Here however, the challenge for the incumbent is to spot the entrepreneurial firms that constitute such a threat early on and respond adequately. The large managerial literature on disruptive innovation is built on these risks faced by incumbents (Rigby et al. 2002). While this ability to spot the specific disruptor is imprecise and uncertain (Markides 2006), broad movements within the entrepreneurial space into markets that intersect with the incumbent’s product markets can still be recognized as significant threats. In other words, while an incumbent may not observe a specific startup with a specific product innovation, the incumbent will surely observe and respond to shifts across the entrepreneurial landscape into certain specific markets or technologies. Prior work studying technology fads and cascades has discussed these broader trends as being significant predictors of firm and individual behavior (Abrahamson 1991; Bikhchandani et al. 1998). The effect of such large-scale new venture formation within a certain product market is representative of new entry threat and impending competition in the future. Here however, the direct effect of this threat on the incumbent firm’s innovation investment is uncertain and remains an empirical question we address in this paper.

**Moderating Role of Governance**

New entry threats are likely to be viewed and responded to differently by firms, depending on their managers’ specific capabilities and attributes (Agarwal and Gort 2002). Corporate governance is one such firm-level characteristic that may condition the incumbent firm’s response to increased new entry threats. Corporate governance refers broadly to the mechanisms, processes and relations by which corporations are controlled and directed. A narrower definition describes governance as the core set of rules and mechanisms that regulate the power-sharing between shareholders and managers (Gompers et al. 2003). Governance is the firm’s first-level response to agency issues that emerge when firm ownership deviates from firm management (Jensen and Meckling 1976) and is a recurring theme in public firms as it influences several firm-level outcomes such as stock returns (Bebchuk et al. 2006; Gompers et al. 2003; Johnson et al. 2009), firm value (Bebchuk and Cohen 2005; Brown and Caylor 2006) and Tobin’s Q (Bhagat and Bolton 2008; Januszewski et al. 2002).

Within the specific context of innovation, interestingly, there are competing arguments for how governance may influence the firm’s decision to invest in innovation. On the one hand, the traditional

---

2 An instance of such an entrepreneurial movement may be the significant increase in startups aimed at voice-over-IP technology in the mid-2000s. While individual startups may not represent adequate threat, a collective groundswell of entrepreneurial activity is noted by incumbents (blogs.wsj.com/venturecapital/2009/03/04/voip-start-ups-scramble-to-find-niche/).
agency perspective suggests that without the corrective role of the market or shareholders, protected managers are likely to invest in assets that increase the utility of the management but not of the firm (O’Connor and Rafferty 2012). A variation of the agency perspective, called the “quiet life” hypothesis offered by Bertrand and Mullainathan (2003) argues that protected managers are likely to avoid risky investments like innovation and R&D (thereby preferring the “quiet life”). Such protected managers may also prefer to invest resources and human capital in more routine projects with quicker returns, but low overall value. These arguments imply a negative association between the protection experienced by managers and the firm’s innovation investments. On the other hand, an alternative viewpoint provided by Stein (1988) as the “managerial myopia” hypothesis, argues that when managers are not protected within the firm, the threat of takeovers or action by shareholders induces myopic decision-making. This leads to significant under-investment in activities that provide long-term payoffs. Myopia is induced by the expectation that the market (and therefore shareholders) undervalues risky innovation activities, which leads to short-term wealth loss, and in some cases, hostile takeover bids. Therefore, protecting managers allows incentives that reward long-term optimal decision making, whereby managers are more likely to initiate innovative projects (Aghion and Tirole 1994; Shleifer and Summers 1988). This reasoning implies a positive relationship between the level of protection the managers enjoy and investments in innovation. Emerging empirical evidence provides some initial support for the managerial myopia hypothesis, i.e. protected managers in a firm tend to be associated with greater innovation output, all else being equal (Becker-Blease 2011, Chemmanur and Tian 2012).

Inasmuch as prior research has indicated an association between governance and innovation, we seek to understand how governance may influence the firm’s response to increasing new entry threats, i.e. its role as a moderator. Here too, we encounter a tension that requires empirical analysis. On the one hand, increasing new entry threats could trigger a stronger response among protected managers than their unprotected counterparts for two reasons. First, protected managers are more likely to remain at their current positions into the future, when the impending threats of new entry materialize. Therefore, they have stronger incentives to react pre-emptively in innovation investments that increase the odds of fighting off any future competition (Gilbert and Newbery 1982, Goolsbee and Syverson 2008). Second, by virtue of their protection, they are likely to have the cognitive abilities and firm resources to actively investigate emerging or nascent market threats (Stein 1988). Protected managers thus have the incentives, the decision rights as well as the resources at their disposal to actively respond to such threats in the long-term (Hoberg et al. 2014). On the other hand however, if the “quiet life” hypothesis is correct, and the protection clauses do not provide enough incentives for long-term thinking, entrenched managers may not behave systematically different from unprotected ones. In addition, these threats are based on future competition; the protected manager is likely to act on shorter-term incentives to extract immediate value from the firm, consistent with the agency-theoretic perspective (O’Connor and Rafferty 2012). Given this ambiguity in the direct of moderation, we allow the empirical analysis to provide guidance.

### Research Methods and Data

#### A Text-Based Measure of New Entry Threats

Studying new entry threats empirically poses several challenges. By definition, threats from entrepreneurial 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 materialize at some point in the future. While many theoretical analyses of new entry threat exist, empirical measurement has typically relied on some variation of measures of present market competition and extrapolating from these measures. In addition, these measures typically define competition or industry concentration based on static SIC or NAICS codes (Aghion et al. 1992; Becker-Blease 2011), and suffer from multiple shortcomings: a firm is rarely reclassified even when they enter into a different industry; and they lack temporal variations both within and across industries. Furthermore, for start-ups, there are no ready-to-use and accepted industry classifications, causing difficulties in identifying true potential threats. Given the lack of a viable existing

---

3 We note that as part of the analysis, we do test for the direct effects of firm governance on innovation as well. These results are discussed together with the results of the moderation tests.
measure of the threat of new entry from startups, we develop our own measure here. New machine learning techniques with text analysis provide methods that may be used to measure such threats in a novel manner. We explore these options to construct measures for new entry threats emerging from startups.

Intuitively, we measure the extent to which a firm’s description of its product markets provided in its 10-K filings overlaps with the descriptions of new entrepreneurial firms that receive first-stage funding. We extract business descriptions of startup firms receiving first funding from VentureXpert. The focus on very early stage entrepreneurial firms is particularly appropriate here since these firms collectively represent movement in the incumbent firm’s product market towards possible future competition. In addition, the fact that they were funded by venture capitalists makes these threats particularly relevant because they tend to be highly innovative.

The use of text analysis requires reasonably descriptive text corpora from firms that may be used to construct appropriate measures. A considerable body of work that has used public filings provided by the firms, specifically firm annual reports (10-Ks) in the US, to create measures of firm fundamentals such as competitive intensity, industry classes and firm strategy (Hoberg and Phillips 2010; Tetlock 2011; Tetlock et al. 2008). These documents are useful as sources of data for two reasons. First, public firm product descriptions must be representative and significant as required by financial market regulations. Thus, product descriptions of public firms contain timely information about their products, markets and competitors that are consistent with the firm’s perceptions. Second, as firms evolve, these descriptions are modified and updated to reflect the changing nature of their businesses, thereby providing longitudinal variation. We therefore consider the use of text-based measurement schemes for capturing new entry threats in the context of IT industry.

We construct our new measure using techniques available in the text analytics literature. We use the VentureXpert dataset and focus on startups in all high-tech space that are backed by venture capital funding. Using VentureXpert data allows us to focus on IT entrepreneurs that have received some venture capital funding, ensuring that all entrepreneurial firms have baseline quality and represent credible threats to incumbents. It is important to note that 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 threat from the entrepreneurial space in two ways. First, we argue that the threat of new entry from startups does not appear from any single entrepreneur but from broad movements in the entrepreneurial space, i.e. evidence of systematic entrepreneurial movements into a specific area or sub-industry are more representative of new entry threat for an incumbent. Following Hoberg et al. (2014), we identify new entry threats at the level of the “industry”; for the purposes of our analysis, we treat the whole set of entrepreneurial ventures on VentureXpert that receive funding as the relevant “industry”, since they represent collectively the new entry threat that is faced by incumbents. Second, it is unlikely that all entrepreneurial firms represent the same level of threat to the incumbent. Therefore, we consider those entrepreneurs that receive \textit{first-stage funding} in a given year as posing new entry threat to incumbents in that year. If the entrepreneurial ecosystem observes value in a specific industry subclass or technology space and systematically invests in new ventures at the early funding stage, there is likely to be a groundswell of new ventures associated with this industry subclass entering the VentureXpert dataset in a given year, which could then potentially lead to significant realized new entry in 2-3 years, thereby representing new entry threat for the incumbent. In summary, the new entry threat measurement here is based on (a) new ventures that receive new first stage funding, and (b) collective body of \textit{all} entrepreneurs who receive first-stage funding, rather than individual entrepreneurs.

The collective body of entrepreneurial startups will represent varying levels of new entry threats to incumbents, depending on how closely the entrepreneurial ventures are related to the primary market of the incumbent. We therefore require a measure of the similarity between the funding observed in the entrepreneurial space and the incumbent’s product market. We use the cosine text similarity approach to capture this similarity (Sebastiani 2002). The primary building block to construct our text cosine similarity is the set of unique words that firms use to describe their products in their business descriptions. For publicly traded firms (incumbents), the source of business descriptions are from Section 1 of their 10-K annual filings. For entrepreneurial firms (start-ups), we use their business description from VentureXpert database. We extract all detailed business descriptions from start-ups that received first-stage funding and aggregate these descriptions for each year $t$; these business descriptions are usually
short, with the typical description consisting of 4-5 sentences. Aggregating these for a given year provides a more representative and useful document of entrepreneurial entry in a particular year. Cosine similarity between this collective entrepreneurial document and an incumbent’s business description forms the basis for measuring new entry threat by calculating the overlap in word usage.

Specifically, once we have the respective text documents available, we parse semantics at a sentence level and retain the nouns and proper nouns, which are the most meaningful words elements in product descriptions. Next, we define all incumbents’ business descriptions and the aggregated start-up document as one document corpus (or collection) for a particular year \( t \) (that is, the corpus includes \( n+1 \) documents in total, with \( n \) being the number of incumbents in year \( t \)). Then, we build document vectors for each incumbent’s text and the aggregated start-up 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 document corpus of year \( t \). Let \( \mathbf{W}_i \) 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 start-up file) in year \( t \). We use Term Frequency times Inverse Document Frequency (TF-IDF) (Tata and Patel 2007) to weight each word in the document vector. In this case, each element \( j \) in \( \mathbf{W}_i \) represents the relative importance of the word (or term) \( j \) in the document, given the within-document frequency and cross-document frequency. Term Frequency (\( f_{ij} \)) is defined as the number of occurrences of words \( j \) in document \( i \). The normalized term frequency is defined as:

\[
TF_{ij} = \frac{f_{ij}}{\max_k f_{ik}}
\]  

That is, the term frequency of term \( j \) in document \( i \) is \( f_{ij} \) normalized by the maximum number of occurrences of any term in document \( i \). This normalization process helps correct for bias caused by the length of a document (i.e. term frequency gets inflated in longer documents). The \( IDF_j \) for a term is defined in following fashion. Suppose 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), while a term that occurs in only a few documents, such as “encryption”, gets a higher IDF. The weighting score of TF-IDF for term \( j \) in document \( i \) is then defined to be \( TF_{ij} \times IDF_j \). Intuitively, words with high within-document frequency will obtain high weighting and those with high cross-document frequency are weighted less. Therefore, terms that are used infrequently in the collection of documents gain more importance in calculating the similarity of two documents while those that are common across documents are less discriminating and have lower weights. Lastly, since our main interest lies in the similarity between the document representing the incumbent and the aggregate document representing start-ups in that year, we operationalize the text-based measure of new entry threats for incumbent firm \( i \) in year \( t \) as:

\[
NET_{TFIDF_{it}} = \text{SIMc}(\mathbf{W}_{it}, \mathbf{W}_{0,t}) = \frac{\mathbf{W}_{it} \cdot \mathbf{W}_{0,t}}{||\mathbf{W}_{it}|| \times ||\mathbf{W}_{0,t}||}
\]  

where \( \mathbf{W}_{it} \) denotes incumbent firm \( i \)’s document vector in year \( t \) (\( i = 1, 2, 3,...n \)) and \( \mathbf{W}_{0,t} \) represents the aggregated start-ups’ business descriptions indicating product space from all start-ups. By definition, the cosine TFIDF-based measure of new entry threat \( NET_{TFIDF_{it}} \) is bounded between \([0, 1]\). Higher values represent greater threat of new entry for the incumbent (since the two word vectors are closer in unit vector space) while lower values represent less threat.

We also calculate the cosine similarity based on a binary weighting scheme for each element in document vector \( \text{NET}_{BIN_{it}} \). In this case, we construct the words vector of document with word vector \( \mathbf{W}_{it} \) denoting and ordered Boolean vector of length \( J_t \). Element \( j \) of \( \mathbf{W}_{it} \) equals 1 if document \( i \) includes word \( j \) in its product description and is zero otherwise. Intuitively, this measure does not weight words by their relative frequency but only by their presence in the word vector for a specific firm-year. Our result using this measure is consistent across all specifications in the paper.

In order to verify that our method of text analyses indeed accounts for changes within the entrepreneurial ecosystem in terms of the sectors associated with new entrant firms, we examine the most popular terms that emerge from the TF-IDF weights across the years of observation. In Table 1, we present the list of 20 words with highest weights for three years as exemplars: 2000 (the peak of the dot-com bubble), 2003
(the year immediately after the collapse of the dot-com bubble), and 2006 (post-dot-com bubble, and last year of our sample). We find that in year 2000, the VC-funded entrepreneurial space was dominated by firms related to Internet or software industry: words such as “online”, “Internet”, “software”, “web”, “broadband”, “ecommerce” and “portal” were among the most frequently used ones in their product descriptions. However, in year 2003, we observe a significant change of the vocabulary used to describe the startups that were funded: the occurrence of Internet-related words reduced significantly, to be replaced by words such as “disease”, “patient”, “drug”, “treatment”, “therapy”, “protein”, “biotech”, and “antibody”. These changes show that the overall movement of venture capital funding shifted from Internet/Software industry to pharmaceutical and biotech industries in 2003. Interestingly, in 2006, we see the re-emergence of software and Internet-related terms, but with a difference. Terms such as “search”, “cloud”, and “blog” become more influential while terms like “advertising”, “video” and “game” emerge but with a different angle, given the growth in the mobile and internet domains of such firms. Overall, Table 1 provides evidence for the significant longitudinal variation in the words that are used to describe the entrepreneurial firms in different years; our text-based measure of new entry threat captures and builds on these underlying trends within the entrepreneurial space.

<table>
<thead>
<tr>
<th>Year</th>
<th>Words List</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>onlin, internet, softwar, wireless, softwareprovid, servicesprovid, solut, web, content, broadband, ecommerc, softwaredevelop, servicesdevelop, platform, media, drug, email, network, enterpris, patient, portal</td>
</tr>
<tr>
<td>2003</td>
<td>diseases, patient, drug, therapeut, cancer, softwar, therapi, treatment, protein, wireless, discord, video, molecule, inhibitor, tissue, healthcar, biotechnolog, antibody, pharmaceut, cell</td>
</tr>
<tr>
<td>2006</td>
<td>onlin, diseases, patient, cancer, drug, therapeut, search, publish, media, video, therapi, cloud, blog, antibody, tissue, treatment, game, software, advertis, platform</td>
</tr>
</tbody>
</table>

Other Variables

Detailed variable definition and statistical summary are described in Table 2. The dataset was constructed using multiple sources. We restrict our analyses to core IT industries, using the 18 4-digit NAICS industry codes including only IT software, telecom and hardware industries. In addition to NET variables described in the previous section, the following variables form the key components of our analyses: for Innovation, we extract firm-level patent data from the NBER Patent Citation database; Corporate governance is measured by the G-index (Gompers et al. 2003) from RiskMetrics (formerly Investor Responsibility Research Center) database; financial data and other firm level controls are from CompuStat. We include a vector of firm characteristics that may affect a firm’s innovation productivity, including firm size, profitability, investment in R&D, asset tangibility, leverage, capital expenditure, product market competition, growth opportunity, and financial constraints. Our final dataset consists of 547 publicly traded firms over the period 1996-2007, representing an unbalanced panel.

---

4 IT industries include software and hardware, which are defined by 4 digital NAICS code: 2211, 3332, 3333, 3336, 3339, 3341, 3342, 3343, 3344, 3345, 3346, 5171, 5172, 5173, 5174, 5179, 5112 and 5181.
Empirical Analysis

Main Results

Our baseline model estimates the main effect of new entry threat (NET) and corporate governance on the incumbent firm’s innovation outcomes, using a two-way fixed effects panel data specification below:

\[
\ln(\text{Innovation}_{i,t+2}) = \eta_i + \lambda_i + \beta_1 \times \text{G-index}_{i,t} + \beta_2 \times \text{NET}_{i,t} + \beta_3 \times \text{G-index}_{i,t} \times \text{NET}_{i,t} + X_{i,t} \gamma + \mu_{i,t} \quad (3)
\]

where \(i\) indexes firms, and \(t\) indexes time periods. The dependent variable \(\ln(\text{Innovation}_{i,t+2})\) can be one of the following two measures: the natural log of patent counts, \(\ln(1 + \text{Patent}_{i,t+2})\), or the natural log of the citation-weighted patent counts, \(\ln(1 + \text{WeightedCitation}_{i,t+2})\). Building on prior work (Chemmanur and Tian 2012), we impose a two-year lag on the observed new entry threat variables and governance as well as other firm level controls relative to innovation outcome since the actual investments in these innovations are likely to have been made earlier. In the robustness tests, we also estimate the models with a one-year lag as well. The two-year lag also allows for sufficient time on the part of the firm to react to new entry threats as well as reduces the chances of reverse causality in the regression. The variable \(\text{NET}_{i,t}\) is measured by either \(\text{NET}_{\text{TFIDF}}\) or \(\text{NET}_{\text{BIN}}\), while \(X\) is a set of firm characteristics that affect a firm’s innovation productivity. We control for time-invariant unobservable firm characteristics by including firm fixed effects, \(\lambda_i\). We include year fixed effects \(\eta_i\) to control for economy wide shocks. Standard errors are clustered at the firm-level to control for serial correlation (Wooldridge 2010).

We first estimate equation (3) with \(\text{NET}_{\text{TFIDF}}\) as the measure of new entry threat using a fixed effects panel regression; the results are reported in Table 3. Panel A of the table reports results with \(\ln(1 + \text{Patent}_{i,t+2})\) as the dependent variable while Panel B repeats the analyses with variable

---

5 Following Chemmanur and Tian (2012), we use the regression coefficients from Kaplan and Zingales (1997) to compute the KZ index as: \(-1.002^\times\text{Cashflow} - 39.368^\times\text{Dividends} - 1.315^\times\text{Cashflow} + 0.28^\times\text{Q} + 3.18^\times\text{Leverage}\)
ln(1 + WeightedCitation_{t+2}) as the dependent variable. Beyond the direct effects of new entry threats and governance on innovation outcomes, we are also interested in estimating the moderating role of governance on the firm’s response to new entry threats. Therefore, we estimate models with interaction terms G-Index * NET_TFIDF. Column (1) in Panel A of Table 3 provides results for the direct effects while Columns (2) shows the interaction term of governance and NET_TFIDF. This pattern is repeated in Panel B for citation-weighted patents.

Based on the results from Panel A of Table 3, the coefficient estimates on NET_TFIDF is negative and significant at the 1% level, indicating that greater new entry threats are associated with lower number of patent applications. By our estimate, a one s.d. increase in new entry threat is associated with an 8.3% reduction in patent applications. While prior literature is divided on the direct effect of the threat of new entry on innovative outcomes, our analysis here suggests that the direct effects are negative. In the IT industry, firms appear to be responding to such threats by reducing, on average, their innovation activities. While we cannot directly observe the concurrent investment of funds into alternative activities (such as marketing or branding activities, for instance), our results are consistent with existing work that shows how firms facing market disturbance are likely to conserve cash and respond conservatively (Brav et al. 2005, Hoberg et al. 2014). This strategy of conservativeness appears to extend to innovation-related investments as well, given the high risk of failure in such activities and the inability to appropriate the positive benefits from such activities in the event that new competition does materialize (Lukach et al. 2007, Fontana et al. 2009). In Columns (2), we add interaction terms of governance and new entry threats on patenting. We see that governance indeed moderates the relationship between NET_TFIDF and innovation in that it strengthens the negative relationship between NET_TFIDF and innovation. Effect size calculations based on column (2) suggest that, for firms with low levels of G-index in the sample (1 s.d. below mean), a standard deviation increase in new entry threat is associated with 0.68% decrease in patent applications. Surprisingly, this effect is not significant, indicating that managers experiencing relatively low levels of protection do not appear to respond to new entry threats at all. However, the magnitude of this effect at high levels of G-Index (1 s.d. above mean) is a 16.63% (p<0.01) reduction in patent applications, showing that responses of protected managers are significantly stronger.

<table>
<thead>
<tr>
<th>Table 3. Innovation, Governance and New Entry Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>G-index</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>NET_TFIDF</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>NET_TFIDF * G-index</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Firm level controls</td>
</tr>
<tr>
<td>Year &amp; Firm FE</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td># of Firms</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Main variables of interest are standardized with mean of zero and standard deviation of one.

Panel B in Table 3 reports the results from the same regression specification with citation-weighted patent counts as the dependent variable. At first glance, it is clear that the coefficient estimates are of higher
magnitude, suggesting that the marginal effects of new entry threats are more influential here. Prior research suggests that influential or path-breaking innovation is harder to predict and therefore, represent additional risk for the manager (Hall et al. 2001); our results here show that the presence of new entry threats leads to a lower number of such high-quality patent applications. We see consistent results with regard to the moderating effect of governance as well; there is significant moderation of the NET_TFIDF effect on patenting (column (2), Panel B) and these effects are of particularly high magnitude in the case of highly protected managers. Across Panels A and B, the relatively high R² values and significant F-statistics indicate that the model fit overall is good.

Overall, the results suggest that the threat of future competition through new entry reduces the ability or the inclination of managers to invest in risky investments such as innovation. The effects of the new entry threats are exacerbated by the extent to which managers are protected, which appears to elevate their conservatism (Hoberg et al. 2014). These findings are visually captured in Figure 1, which shows a plot of fitted values from a regression of governance structure and new entry threats on the innovation output when other variables are held constant at their main levels.

![Figure 1. Level Plots of Fitted Values](image)

**Endogeneity of Corporate Governance**

The above analyses are predicated on the assumption that the corporate governance is exogenous. However, such assumption may not necessarily hold. For example, high quality managers may tend to adopt a greater number of governance provisions relative to low quality mangers (Chemmanur et al. 2011). The quality of managers is also an important factor in determining a firm’s innovation productivity. However, manager quality is unobservable and likely correlated with both the number of protection provisions and innovation productivity. To rule out this alternative case, we use an instrumental variables approach (Chemmanur and Tian 2012).

We construct two instrumental variables for the G-Index measure that should be ideally correlated with the firm’s decisions pertaining to governance but not directly affect the firm’s innovation productivity. The first instrumental variable we use is the average G-Index observed in the focal firm’s rivals, which is likely to influence the focal firm’s G-Index (through contagion or benchmarking) but should not directly affect the focal firm’s propensity of patenting directly. We use the text-based industry classification (TNIC) system described in Hoberg et al. (2010a) to identify the set of rivals for the firm.\(^6\) The TNIC

\(^6\) We do not describe the TNIC in detail here in the interest of space but refer the interested reader to Hoberg et al. (2010) for detailed information on the construction and validation of the TNIC.
provides two distinct advantages over static SIC/NAICS schemes here. First, the TNIC for a firm varies over time depending on the changes in the text in the firm’s 10-K and its rivals. Second, the TNIC relaxes the transitivity property typical of SIC/NAICS codes, which allows for a more firm-specific influence of rival G-Index values, rather than a general level of influence based on a common NAICS/SIC code. Additionally, it is unlikely that the average governance structure observed within a firm’s TNIC directly influences the firm’s patenting propensity.

Following the corporate governance literature (Bhagat and Bolton 2008), we use a second instrumental variable, Currently Active CEOs on Board, for the endogenous firm governance. This variable is defined as the percentage of directors on the board of the focal firm who are currently active CEOs. Hallock (1997 and Westphal and Khanna (2003) emphasize the role of networks among CEOs that serve on boards, and the adverse impact on the governance of such firms. The CEO of the focal firm has strong incentives to “capture” the board, so as to ensure that he can keep his job and increase the other benefits he derives from his position. On the other hand, being members of the board, directors have the incentives to maintain their independence and to monitor the CEO. Thus, the actual governance mechanisms opted for by the firm is driven by the inherent tension between focal CEO desiring protection and the number of independent CEOs willing to exert pressure. Prior research shows no direct, systematic relationship between the composition of the board and firm performance (Hermalin and Weisbach 2003). More to the point, there is no evidence of a relationship between this instrumental variable and the focal firm’s innovation outcomes. Therefore this variable is also a suitable instrument for G-Index.

Using these two instrumental variables, we use 2SLS to estimate the direct effect of G-Index on innovation in order to address the endogeneity issue. In the first stage, year and firm fixed effects are included and standard errors are clustered at the firm level. The coefficient estimate of the first instrument, Industry Average G-index, is positive and significant at the 1% level, consistent with the intuition that firms are influenced by their rivals’ governance structure. The coefficient estimate of the second instrument, Current Active CEOs on Board, is negative and also significant at 1% level, suggesting that more active CEOs on board curtail the extent to which managers are protected.

<table>
<thead>
<tr>
<th>Table 4. 2SLS Regressions – 2nd Stage Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>G-index</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>NET_TFIDF</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>NET_TFIDF * G-index</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Firm Level Controls</td>
</tr>
<tr>
<td>Year &amp; Firm FE</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td># of Firms</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Main variables of interest are standardized with mean of zero and standard deviation of one.

Table 4 reports the results from the second stage of 2SLS regression estimating equation (3) while accounting for the endogeneity of G-Index. The Panel A in Table 4 reports results using raw patent counts while Panel B reports citation-weighted patent counts. We are primarily interested in the extent to which endogeneity may have biased the coefficients for G-Index; the first columns in Panel A/B show a
significant and positive effect of G-Index on innovation. Relative to the panel analysis, there is a substantial increase in the magnitudes of the main coefficient of interest as well as its associated standard error. The larger estimates indicate that, accounting for endogeneity, the effect size of governance is indeed significant on innovation. The wide range of 95% percent confidence intervals for the adoption of additional governance provisions include effects that are very large; therefore, caution is recommended in using point estimates from this analysis for predictions outside the sample. Nevertheless, the IV estimates from 2SLS follow the same pattern as the baseline regressions, suggesting that the endogeneity of G-Index may not change the inferences from our analysis.

**Endogeneity of NET – Reverse Causality and Omitted Variables**

Another concern here is the endogeneity of new entry threat, potentially from reverse causality (i.e. the presence of pre-emptive patenting to deter entry) or omitted variables that may influence both entrepreneurial entry and incumbent innovation (such as technological maturity or opportunities). We address both of these concerns below.

By definition, new entry threats emerge from actions taken by new entrepreneurial firms that operate independently of incumbent firms, whose decisions we are modeling here. Additionally, note that our measure of new entry threat is composed of multiple decision-makers – the incumbent’s use of product vocabulary in its 10-K filing, the use of similar product vocabulary by the entrepreneur, and the first-round funding by an independent venture capitalist in that year. We argue that the combination of these three events occurring concurrently is unlikely to be systematically correlated with the error term in equation (3) (Goolsbee and Syverson 2008, Hoberg et al. 2014). Thus, a generalized concern of endogeneity, wherein the NET variable is systematically correlated with the error term, may not exist. If an omitted variable, such as technological opportunities emerging in a given year, is assumed to affect both entrepreneurial entry and incumbent innovation strategy, note that for this to systematically influence our NET measure, three conditions would be necessary – first, the entrepreneur should have been formed incorporated prior to this event (so as to be in existence when the technological opportunity emerges). Prior work in venture capital suggests that newly formed entrepreneurs are rarely accorded first round funding (Gompers and Lerner 2001), suggesting that this likelihood is slim. Second, this entrepreneur should receive first round funding from a VC at or around the time that the technological opportunity emerges. Finally, the incumbent firm should use a similar product vocabulary as the funded entrepreneur in its yearly 10K filing for that year. Additionally, of course, the incumbent should choose to contemporaneously increase its investments in innovation as a response to the technological opportunity, leading to increased patents in two years. We argue that such a concatenation of events in response to an emergent technological opportunity specifically is unlikely to manifest in this manner. The presence of an opportunity may result in entrepreneurial firm formation, which may then receive VC funding in later years. Alternatively, existing entrepreneurs who happen to be in business when the opportunity emerges may have used product vocabulary that differs from that used by incumbents or differ from words relating to the opportunity. In all of these cases, there is unlikely to be a systematic correlation between NET and the error term in equation (3). Prior work using text analysis to capture constructs such as competition have argued for and shown how endogeneity of this form is unlikely, for similar reasons as the ones we provide here (Hoberg et al. 2014).

A second concern is that of reverse causality, i.e. the negative relationship between NET and firm innovation may be driven by the incumbent’s pre-emptive patenting as a strategy to deter entrepreneurial entry. For example, Cockburn and MacGarvie(2011) identify a distinct “property rights” effect for patent holders to exclude competitors and raise the entry cost. However, we argue for several reasons that may not be very likely. First, we consider the patent application year as our basis for measuring innovation, i.e. we include all patents applied in year t that were eventually granted. Since there is no guarantee that these patents are granted in a timely manner, they represent weak signals of pre-emptive action, if any, to potential entrepreneurs. Of course, this assumes that patent applications from incumbents are actually visible to potential entrepreneurs in order to deter entry, which is unlikely. Pre-emptive patenting may be more reasonable in the competition between *incumbents*. Second, we impose a two-year lag in our specifications in order to avoid the possibilities of reverse causality and to allow enough time for the incumbent to respond to perceived new entry threats. If patent applications could indeed be pre-emptive, the negative relationship between NET and incumbent innovation would be highest in the same year or one year lagged. We conducted both these analyses and found no significant effects of NET on innovation.
Given the limits on space, we do not provide these detailed results here. They are available upon request.
transaction in year $t$, and zero otherwise. We estimate a panel data firm fixed effects model on $\text{Ln(TransactionValue}_{t+1})$ and fixed effect linear probability model on the binary dependent variable $\text{AcqPropensity}_{t+1}$ to probe the relationship between NET and firms' acquisition activity. The results (available upon request) show no significant relationship between NET and acquisition activities. Second, we study the firm's cash holdings and compute the firm's cash holding as the natural log of firm cash and cash equivalents, and run a panel data with 3 digit NAICS industry fixed effect model using cash holding as the dependent variable (Hoberg et al. 2014). The results (available upon request) indeed show that firms withhold more cash as new entry threats increase, consistent with Hoberg et al's (2014) finding that firms are more financially conservative when their product markets are volatile. Interestingly, we also find a significant moderating effect of corporate governance in that more protected managers tend to reserve more cash holdings in the face of greater threats. This finding provides some preliminary evidence for how protected managers may be responding to increased expected turbulence in their product markets vis-à-vis their innovation investments and cash holdings. However, we caution that more work is needed to establish these mechanisms rigorously and also represent viable avenues for future research.\(^8\)

We briefly outline some limitations of our work here as well. First, we rely on self-reported descriptions of products and services on the part of entrepreneurs, which may lead to some bias. It is possible that entrepreneurs shade the descriptions of their products based on what is considered more favorable in the venture market. Our approach of aggregating text across all entrepreneurs allows some reduction in such biases. Second, while we study innovation outcomes in the form of patent applications, prior work indicates that not all firms use patents (Ahuja and Katila 2001). To that extent, our models of innovation outcomes may be under-estimating the true innovation-related spending attributable to new entry threat. Finally, while we use the cosine method to measure new entry threat, more computationally intense text analytic strategies are available (Sebastiani 2002). We leave these explorations to future work.

Our work makes several useful contributions to the literature on innovation within the IT industry. First, existing theories are divided in their predictions of the relationship between new entry threats and innovation. One school of thought, the “Darwinian” view of competition, asserts that competition forces firms to innovate and to be more efficient (Porter 1990). In addition, Aghion et al. (2005) argue that in sectors characterized by “neck-and-neck” competition, competition reduces pre-innovation rents by more than it reduces post-innovation rents, therefore resulting in the incentive of incumbents to make R&D investments to “escape competition”. This point of view finds some support in recent empirical studies that find a positive relationship between competition and productivity growth (Blundell et al. 1999; Nickell 1996). In contrast, endogenous growth theory has stressed the Schumpeterian view that the primary driver of innovation and economic growth lies in the existence of future monopoly rents (Aghion et al. 1999). To the extent that increasing competition reduces monopoly rents, Schumpeterian point of view leads to a negative relationship between competition and innovation. One approach to reconcile these contrasting predictions, offered by Aghion et al. (2001), is to alter the technological assumptions: if competition is induced by outsider innovators that earn no rents if they fail or become monopolies if they succeed, or if incumbents can leapfrog over their rivals through technological innovation, the Schumpeterian effect dominates. However, if technological progress is characterized by “step-by-step” innovation and technological laggards need to first catch up with leaders before competing for leadership in the future, then the Darwinian effect dominates (Aghion et al. 1999). Since our measurement of new entry threats captures mostly the incipient instability induced by entry from outsiders, and it is well-known that leapfrogging over current rivals is known to occur in the IT industry, our finding is consistent and provides complementary empirical evidence to this point of view, specifically from new entry threats from entrepreneurs (Aghion et al. 2005).

Second, progressively more companies in the IT industry, such as Google, Amazon and Facebook are adopting governance structures such as dual-class shares – owners of a small portion of a company’s stock getting the majority of the voting power – in order to protect founders/managers that allegedly protect and promote the long-term prospects of the firm. While anecdotal evidence is abundant in the IT industry for how owner-managers may indeed be beneficial for the firm – consider, for example, how Apple’s

---

\(^8\) We also tried to analyze the effect of NET on the firm’s advertising and marketing spending but sparse and incomplete data prevented us from completing this analysis in full measure.
performance greatly deteriorated after its founder was ousted – there is a surprising gap in the understanding of the role these protection clauses play in firm strategy in prior IS literature. Our analyses show that corporate governance does indeed shape IT firms' innovation strategies: consistent with the "managerial myopia" view (Stein 1988), protected IT managers tend to be long-term thinking and invest more heavily in risky innovation activities that have long-term payoffs. More importantly, we reveal that protected managers are also forward-looking in terms of their reaction to new entry threats: they actively respond to potential future threats by responding to movement in the entrepreneurial space, in stark contrast to their relatively unprotected counterparts who appear to not respond at all. Clearly, protected and unprotected managers differ in their planning horizons; we show that the costs of protection, by shielding managers from the market, may be offset by their ability to respond proactively to future threats, especially in volatile markets.

In addition to theoretical contributions, we make a methodological contribution by creating a new measurement of new entry threat that is particularly useful in gauging the impact of entrepreneurial firms. 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 firms enter and exit certain product markets. As a result, using this text-mining approach is likely to reveal interesting new patterns. For example, Hoberg et al’s (2010) analyze using text-based measures showed that the text-based similarity between acquirer-target pairs of firms was significantly higher than suggested by SIC-based measures. Our new measure of new entry threat thus fills a gap in prior literature regarding appropriate measurement of potential competitive threats from new entrants. We encourage future research to use similar text-analyses based approach in the studies that involve measuring competitive dynamics.

References


