Catering Innovation:

Entrepreneurship and the Acquisition Market

Xinxin Wang*

JOB MARKET PAPER

December 2015

Abstract

Innovation in the start-up market is a key determinant of economic growth. But what determines an inventor's decision to begin a new venture and his or her subsequent innovation? This paper analyzes the role of the financial market of acquisitions. After documenting its increasing importance as the dominant exit path for entrepreneurs, I test a novel catering theory of innovation: Does the market structure of potential acquirers have a measurable impact on inventors' start-up decisions? I construct a new dataset of early stage start-ups using the uniquely broad coverage of CrunchBase data. I disambiguate and match the resulting data to employment data from LinkedIn and to the entire universe of patent data. Using the prior citation history of entrepreneurs for exogenous variation, I construct a formal proxy variable and employ the Heckman selection model to establish causality. I find that a one standard deviation increase in acquirer market concentration decreases the inventor's propensity to become an entrepreneur by 4%. This first result suggests that fragmented markets are appealing entry markets. My main finding is that a one standard deviation increase in acquirer concentration and market size increases the quality of patents, as measured by citations per patent, and the catering of entrepreneurs, as measured by technological overlap with potential acquirers. The magnitudes suggest that 5-16% of entrepreneurial innovation can be attributed to the influence of acquisition markets, particularly in the information technology and biotechnology industries.

*UC Berkeley, Haas School of Business. Email: xinxin_wang@haas.berkeley.edu

1 Introduction

In 2014, Google acquired Nest Inc. for $3.2 billion, Facebook purchased Oculus VR for $2 billion, and Johnson & Johnson obtained Alios for $1.75 billion. The common trait among these acquisitions is that the startup market provided key innovations to large corporations. Google's patent portfolio has increased from 38 patents in 2007 to over 50,000 patents within the last five years, with many of these patents purchased from the start-up market rather than produced in-house.[1] In fact, Schumpeter highlights the importance of the entrepreneur as the primary driver of innovation and economic change, labeling it "the pivot on which everything turns." Nevertheless, research on the determinants of innovation has paid little attention to the link between the acquisition market and entrepreneurial innovation. In this paper, I show that entrepreneurs and their innovation strategies are strongly affected by the market structure of acquirers. Both their initial willingness to become entrepreneurs and the positioning of their companies reflect the acquisition market and its current players.

The inventor's incentive to become an entrepreneur and to innovate depends on the rents from innovation ex-post. While an initial public offering presents an important route for entrepreneurs to diversify equity holdings and access public equity markets, an increasingly more common alternative pathway exists through the acquisition market.[2] According to the National Venture Capital Association, acquisitions constituted 89% of the value of exits of venture-backed firms in 2009. Technology giant Google alone has acquired over 180 start-ups since 2008, with Microsoft, Facebook, and Cisco following suit. This sizable proportion of acquisitions is not unique to the information technology industry but prevails in health care, financial services, and consumer goods as well. This shift in exits is recognized in practice and maintains significant implications for the decision-making of entrepreneurs. In the biotechnology industry, acquisition options are built into start-ups' strategic planning with "more than 90% of bio-entrepreneurs envision[ing] this trade-sale scenario."[3] These entrepreneurs create and grow businesses with the express vision of an acquisition exit, and innovation decisions hinge on their view of the future acquisition market.

[1] http://www.technologyreview.com/news/521946/googles-growing-patent-stockpile

[2] See Ritter and Welch (2002) for a review of the motivations to go public. A more recent paper by Bayar and Chemmanur (2011) addresses the tradeoffs between IPOs and acquisitions theoretically. They study the exit choice when the decision is made either by the entrepreneur alone or in combination with venture capitalists.

[3] Dr. Frost, CEO of Acuity Pharmaceuticals, http://www.genengnews.com/gen-articles/twenty-five-years-of-biotech-trends/1005/.

Despite this prevalent shift in exits for entrepreneurs, this area lacks academic investigation. The existing finance literature on the role of mergers and acquisitions in innovation focuses on public companies, even when the inclusion of start-up targets may alter the picture (Rhodes-Kropf and Robinson, 2008; Phillips and Zhdanov, 2013; Bena and Li, 2014; Seru, 2014). On the other hand, the innovation literature examines incentives outside of the financial market (Manso, 2011; Acemoglu et al., 2013; Balsmeier, Fleming, and Manso, 2015). The absence of academic studies regarding the positioning of entrepreneurs and their innovation in targeting specific acquisition markets, coupled with the recognition of this issue in the current financial press and among practitioners, highlights the importance of analyzing this question.

On the theoretical level, the effect of the acquisition market structure on innovation is not obvious. The entrepreneur faces tradeoffs in catering innovation to potential acquirers in order to maximize returns to scale, differentiating innovation to "escape competition," and displacing monopoly profits. These factors could affect both the incentive to start a venture and how entrepreneurs cater innovation to potential acquirers in the market. In particular, Schumpeter argues that incumbents value innovation more in concentrated industries because monopolists can more effectively appropriate the benefits of innovation and scale. On the other hand, Arrow asserts that the cannibalization of monopoly rents decreases the incentive to innovate in concentrated industries. Furthermore, the "escape competition" effect states that increased product market competition increases the incremental profits from innovating, additionally predicting a negative relationship between concentration and innovation. I describe these theories and derive direct predictions in the theoretical motivation section. The implication of these theories addresses the real economy and shows that catering innovation – innovating in the same technological areas as potential acquirers – may actually be suboptimal for overall growth.[4]

I test the competing theoretical predictions utilizing novel data on early stage start-ups collected from CrunchBase, an online aggregator of start-up data. CrunchBase, an untapped resource that captures much of the venturing of inventors and finance of start-ups, is better tuned to the innovation markets (Internet of Things, biotechnology, and electrical hardware) than later

[4]This paper has a similar flavor of catering to that of Baker and Wurgler (2004), which studies when managers pay dividends. The authors find that managers cater to investors by paying dividends when investors put a stock price premium on payers, and by not paying when investors prefer nonpayers. Here, entrepreneurs cater to potential acquirers by choosing where and how much to innovate.

staged databases such as VentureXpert. CrunchBase lists over 200,000 companies and 600,000 entrepreneurs, including extensive detail on the investments, products, and acquisitions of each company. I augment this data in three important ways. First, I scrape the employment history of each entrepreneur from LinkedIn. Second, I hand-collect SIC codes for each start-up, employing CrunchBase product market descriptions and industry categories. Last, I match the CrunchBase entrepreneurs to inventors in the universe of patent data in NBER and EPO Patstat using the employment history, location, and age of the entrepreneurs. To my knowledge, this represents the first dataset of inventors ex-ante linked to entrepreneurs and their innovation post-entrepreneurship. To this extent, my final constructed dataset comprises a panel of inventors, their entrepreneurship choices, and their patents as they move through time and across firms.

This paper includes three main contributions. My first contribution is to document the causal effect of acquirer market structure on innovation in terms of quantity, quality, and catering. The specification of interest would be one with measures of ex-ante acquirer market structure on the right-hand side and measures of innovation on the left-hand side. However, examining the causal implications of acquirer concentration on start-up innovation requires a methodology that resolves endogeneity and self-selection problems. I address these problems by exploiting plausibly exogenous differences in entrepreneurs' ex-ante acquisition markets as follows:

For each entrepreneur, I proxy for the acquisition market using the citers of the entrepreneur's prior patents. I then match entrepreneurs on observables such as prior industry and prior innovation quality. Consider the example of two inventors, Tony and Sean, who work in the same industry prior to beginning a start-up at time t. Both inventors are equally innovative and retain the same number of patents and citations. The only difference between Tony and Sean is the identity of the firms citing their patents. If Tony's citers are in heavily concentrated markets, will Tony be more or less likely to become an entrepreneur? Conditional on Tony starting a new venture, will he choose to cater to potential acquirers by innovating in technological areas that potential acquirers value?

One might argue that a potential caveat to a causal interpretation of my results is the unobserved heterogeneity in prior patents. The results may be biased if an omitted variable exists that is correlated with both prior patent citers and ex-post innovation but is uncorrelated with industry fixed effects and prior innovation fixed effects. If monopolistic companies cite Tony's prior patents because he is developing "hotter" patents, controlling for patent count and citations per patent, then the relationship between acquirer market structure and future innovation is correlational at best. However, I directly test Woolridge proxy conditions and show that, conditional on where the entrepreneur previously worked and how innovative he or she is, the assignment of citers on prior patents is orthogonal to unexplained variation in post-entrepreneurship innovation.[5]

Additionally, in order for the proxy to be informative, the ex-ante citers need to accurately forecast the ex-post acquisition market. I find that citers predict ex-post acquirers for the subset of start-ups that experience an acquisition exit event. This implies that the most likely buyers of start-ups are the prior citers of said start-ups' entrepreneurs. Interestingly, start-ups and prior citers (and thus, potential acquirers) do not necessarily compete in the same product market, indicating a difference between technological acquisitions and acquisitions to deter competition.

I find that when facing concentrated acquiring markets, entrepreneurs increase innovation quality and catering. The effects are economically and statistically significant. A one standard deviation increase in acquirer market concentration predicts 16% higher patent quality, defined as citations per patent. Additionally, a one standard deviation increase in the size of the acquisition market as measured by sales increases citations by 7%. Given that the average number of citations per patent in the sample is approximately 12, this implies that when comparing a concentrated industry, such as pharmaceuticals, to a more competitive industry, such as software, the quality of patents produced by entrepreneurs increases by two citations per patent. I also find strong evidence of catering to potential acquirers, defined as technological proximity in patent portfolios. The patent portfolios of entrepreneurs overlap those of the potential acquirers 9% more with a one standard deviation increase in acquirer market concentration and 5% more with a one standard deviation increase in market size.

The evidence supports the Schumpeter view that incumbents value innovation more in concentrated industries because monopolists can more effectively appropriate the benefits of innovation and scale. This implies that acquirers benefit more from acquiring start-ups that innovate. Furthermore, the effect of scaling intensifies with catering innovation, due to the ease with which the acquirer can apply the new technology to their existing product or technology clusters. This resembles the recent empirical work by Zhao (2009) and Bena and Li (2014). Both sets of authors find that technological

[5] This equates to checking that the proxy is redundant in the original model and that the proxy variable and the omitted variable are not jointly determined by further factors.

overlap drives mergers and acquisitions in public companies. I show that this incentive bears implications on the innovation strategy of entrepreneurs, specifically in concentrated acquisition markets due to scaling. In particular, entrepreneurs target acquisitions by catering innovation in the potential acquirers' technological area.

The recent acquisition of Gloucester Pharmaceuticals by pharmaceutical giant Celgene (CELG) illustrates the effect of scale and market structure on entrepreneurs' incentive to cater innovation in terms of technological overlap. First, Gloucester Pharmaceuticals alluded to benefits of monopoly power present in this deal, stating that, "we are thrilled with this transaction because Celgene's global leadership in the development and commercialization of innovative treatments for hematologic diseases makes them ideally suited to bring the clinical benefits of Istodax to patients."[6] Second, Gloucester acknowledged the ease with which their main compound, Romidepsin, a last-state oncology drug candidate approved for the treatment of lymphoma, "provide[s] a strategic fit and expand[s] the company's [Celgene] presence in critical blood cancers."

My second set of contributions concerns the propensity of inventors to become entrepreneurs. I test which market structures are more or less conducive in incentivizing new entrepreneurs. To address this question, I construct the same proxy variable for every inventor in the patent database and run a probit model with entry into entrepreneurship as the outcome variable. The results on entrepreneurship are interesting on their own, as they contribute to the growing research that attempts to identify the determinants of entrepreneurship.

I find that concentrated acquiring markets deter inventors from entering into entrepreneurship. A one standard deviation increase in acquirer market concentration decreases the probability that an inventor becomes an entrepreneur by 4%. I find a similar directional and significant effect of acquirer market size. This is consistent with at least two economic mechanisms. First, fragmented markets attract more entry. In a concentrated market, the risk of potential acquirers extending into the product market with or without the inventor increases. Inventors anticipate increased hesitation to "face off" against large monopolists, reducing their inclination to begin a company in the first place. Second, entrepreneurs are unlikely to extract a high acquisition price from the monopolist due to the lack of outside options and low bargaining power.

[6] See Celgene's press release, "Celgene Completes Acquisition of Gloucester Pharmaceuticals" on January 15, 2010 at http://ir.celgene.com/releasedetail.cfm?releaseid=799365.

My third contribution is to analyze the changes to innovation due to acquirer market structure and size, conditional on entry. The challenge in this analysis is, of course, that entry into entrepreneurship is not random. For example, if inventors who are low quality ex-post chose not to become entrepreneurs, then the analysis would overestimate the quality of innovation in the data. To account for non-random selection into entrepreneurship, I employ the Heckman two-stage estimation method to address potential bias. The results of the two-stage Heckman correction resemble the prior innovation results in direction and size. Both acquirer market concentration and size increase innovation quality (citations per patent) and catering (technological proximity in patents). The economic magnitudes suggest that the positioning of entrepreneurs to prepare for the acquisition market has first order effect on decision making and real output.

This paper contributes to a variety of literatures. First, this paper contributes to the long-standing industrial organization literature on market concentration and innovation (Schumpeter, 1942; Arrow, 1962; Dasgupta and Stigliz, 1980; Gilbert and Newbery, 1982; Aghion et al., 2005). Empirically, conflicting evidence exists regarding whether concentration increases innovation through economies of scale or decreases innovation through the displacement of monopoly rents (Cohen and Levin, 1989; Gayle 2003; Weiss, 2005; Aghion et al., 2014). While prior studies have focused solely on horizontal competition, this paper examines market competition in an acquirer market and its effects on start-up innovation.

Furthermore, this paper provides a link between the industrial organization literature on concentration and the corporate finance literature. Prior M&A research has demonstrated the importance of mergers and acquisitions on innovation but has devoted less attention to the role of entrepreneurship and new start-ups (Bena and Li, 2014; Seru, 2014). Theoretically, Phillips and Zhdanov (2013) demonstrate that large firms may choose to outsource innovation to avoid R&D races with smaller firms. Large firms can minimize R&D risk by only acquiring small firms that successfully innovate. This paper documents this acquisition market in a start-up setting while further investigating its implications on entrepreneurial decision-making.

Conversely, prior entrepreneurship research has primarily documented the role of funding on innovation with little focus on the role of acquisitions (Kortum and Lerner, 2000; Hirukawa and Ueda, 2011; Nanda and Rhodes-Kropf, 2013; Kerr, Lerner, and Schoar, 2014; Gonzalez-Uribe, 2014). This paper, on the other hand, documents a different mechanism for accessing equity markets and thus, a different set of innovation incentives. A recent paper by Hombert, Schoar, Sraer, and Thesmar (2014) also focuses on the effect of changing rents from entrepreneurship and innovation instead of the funding inputs by evaluating unemployment reform. However, the authors of that paper study small business entrepreneurs compared to the high-technology entrepreneurs examined in this paper.

Finally, this paper contributes to the broad literature on the various incentives to innovate. Manso (2011) shows that the optimal incentive scheme to motivate innovation exhibits tolerance for early failure and reward for later success. A separate and extensive set of papers studies how corporate governance affects innovation, focusing on determinants such as the firm's decision to go public (Bernstein, 2012), ownership structure (Ferreira, Manso, and Silva, 2012), and anti-takeover provisions (Atanassov, 2013; Chemmanur and Tian, 2013). Acemoglu, Akcigitz, and Celik (2015) focus on yet another aspect - openness to disruption - as a key determinant of creative innovation. This paper addresses a different motivating factor in an entrepreneur's decision to innovate - the market structure of acquirers.

The remainder of the paper is organized as follows. I discuss the related theoretical literature and develop the competing hypotheses for my empirical analysis in Section II. Section III describes data sources and sample construction. Section IV describes the methodology employed to causally identify the effect of acquirer market structure on innovation. Section V presents the main empirical results on both entrepreneurship and innovation. Section VI concludes the paper.

2 Theoretical Foundations

I consider the entrepreneurs' choice of effort to produce high quality innovation as well as how much to cater. High quality innovations have more widespread impact but quality may come at the expense of time consumed for additional inventions or on marketing and business development in commercialization. Entrepreneurs can cater innovation in terms of choosing to innovate proximally – innovating in the same technology area – as potential acquirers. Innovating in established technological areas contributes incrementally to the entire pursuit of science while innovating in novel areas creates new technology clusters of growth, relative to the acquirer.

In particular, I clarify that two leading types of theoretical models in the industrial organization and the M&A literature, namely the Schumpeter view and the Arrow view, lead to predictions pointing in two antithetical directions, making this an empirical question that requires resolution.

Direction 1: Concentration Increases Innovation

Multiple lines of theoretical and empirical work can be extended to predict that increases in acquirer market concentration can increase the incentive for entrepreneurs to innovate. The classical Schumpeterian argument for innovation is that reductions in competition and increases in scale both increase the incentive to invent by making it easier for firms to appropriate the benefits from innovation. The R&D scale effects have received significant attention in the organizational economics literature. For example, Henderson and Cockburn (1996) and Cohen and Klepper (1996a) identify project spillovers and cost-spreading benefits. Cohen, Levin, and Mowery (1987) argue that complementarities between innovation and non-manufacturing activities, such as distribution, marketing, and operational expertise, may be better developed within large firms. The possibility of an acquisition amplifies this potential gain from innovation since the merged entity can apply the innovation to the entire product line (Phillips and Zhdanov, 2013). Additionally, Salop (1977) and Dixit and Stiglitz (1977) argue that competition decreases the monopoly rents that reward new innovation and thus generates a positive relationship between concentration and innovation. To the extent that the most substantial gains from innovation accrue in imperfectly competitive markets, potential acquirers in concentrated industries have more incentive to purchase innovation. While whether this increases the probability or the price of acquisitions is vague, both will incentivize entrepreneurs to pursue innovation more aggressively.[7]

Concentration Increases Catering Innovation

Less competitive markets, under Schumpeter's view, also increase incentives for start-ups to engage in proximal innovation by increasing the synergies from technological overlap. Bena and Li (2014) show that technological complementarity results in increased merger incidence. They conclude that higher overlap in the same technology space leads to synergy gains above and beyond the returns to innovation conducted by each firm individually. The acquisition of the geo platform, Mixer Labs, by Twitter exemplifies the role that technological synergies play. The acquisition was

[7] If we assume that the target holds some bargaining power that yields a set proportion of the overall acquisition surplus, then increasing the surplus increases the premium.

triggered by 1) complementarity of Mixer Labs' geotag technology with that of Twitter' and 2) the ease and applicability of the technology to Twitter's own core product, tweets. Akcigit and Kerr (2010) reinforce this in their study of the tradeoffs between exploitation (similar in concept to proximal innovation) and exploration innovation. The authors find that exploitation R&D scales more strongly with firm size. Thus, the monopolist's ability to scale more efficiently implies larger returns and a higher premium for acquisitions with technological overlap.

Direction 2: Concentration Decreases Innovation

The competing hypothesis stems from Arrow (1962). Arrow shows that a monopolist that is not exposed to competition or potential competition is less likely to engage in innovation. A firm with monopoly power maintains a flow of profits that it enjoys if no innovation occurs, implying a low net profit from acquiring innovation. A monopolist can increase its profits by acquiring a start-up; however, it cannibalizes the profits from its own legacy technology in doing so. If the competitive acquirer can capture the same benefit from innovation, its differential return is higher because it has no profits to cannibalize. Furthermore, increased competition among acquirers increases the bargaining power of targets. With more competition among potential acquirers, entrepreneurs will capture a greater fraction of the acquisition surplus. Last, even minor product differentiation in contestable markets enables companies, in this case, acquirers, to capture market share (Baumol, 1982). It often is not only recommended but also necessary for firms to innovate because competition decreases pre-innovation rents, thereby increasing the incremental profits from innovating. Innovation, however marginal, can help firms "escape competition." All three arguments imply that the value of innovation increases under competition, and incumbents are more likely to acquire entrepreneurs who innovate.

Concentration Decreases Catering Innovation

The negative effect of concentration on innovation is stronger for catering innovation. Arrow's displacement effect is larger when new products intrude on the existing market for older products, than when products appeal to a new segment of the market and expand the market base. Proximal innovation builds directly on an acquirer's existing technology and increases the risk of cannibalization. In this scenario, highly concentrated markets might encourage entrepreneurs to innovate in new technologies. This implies that the acquisition premium a monopolist will be willing to pay for proximal innovation will be lower than the acquisition premium for differentiated innovation. An anecdotal example is Google's acquisition of home automation company, Nest Labs. Nest's technology created an entirely new product line in which Google had not previously invested. Google justified the high acquisition price as a way of accessing an undeveloped market. Google envisioned beyond Nest's current line, imagining a world of heating, lighting, and appliances all connected and responsive to users. Thus, the negative effect of acquirer market concentration on entrepreneurial innovation is stronger for proximal innovation.

3 Data and Sample Creation

I test the effect of acquirer market structure on the propensity of inventors to become entrepreneurs and on their subsequent innovation. One difficulty in answering this question arises from the lack of data regarding early-stage start-ups and entrepreneurs. In this section, I describe the various data sources and the methodology used to construct a novel dataset of inventors ex-ante matched to entrepreneurs and their ex-post innovation.

3.1 Institutional Setting and CrunchBase

Start-up companies, newly created companies designed to search for a scalable business model, are traditionally financed by venture capital funds. The start-up market has experienced two shifts in recent times. First, the acquisition market plays a comparatively larger role in start-ups exiting from initial financiers. Second, lower fixed costs have allowed entrepreneurs to shift away from venture capital (VC) and toward smaller angel funds. In the technology industry alone, angels fund 10,000 companies every year, while venture capitalists fund only 1,500 companies. Figure 1 shows the number of angel and venture seed rounds from the 2000s onward. The number of angel seed rounds outnumbers the VC seed rounds by a factor of 100.

The angel model of investing consists of smaller funds and thus, smaller deal sizes. This allows for quicker, smaller-dollar trade sale exits. Existing datasets such as VentureXpert only capture later stage start-ups that have already received VC investments. However, to answer my questions on early-stage innovation and the incentive to become an entrepreneur, I need to observe entrepreneur-inventors at the beginning of new venturing.

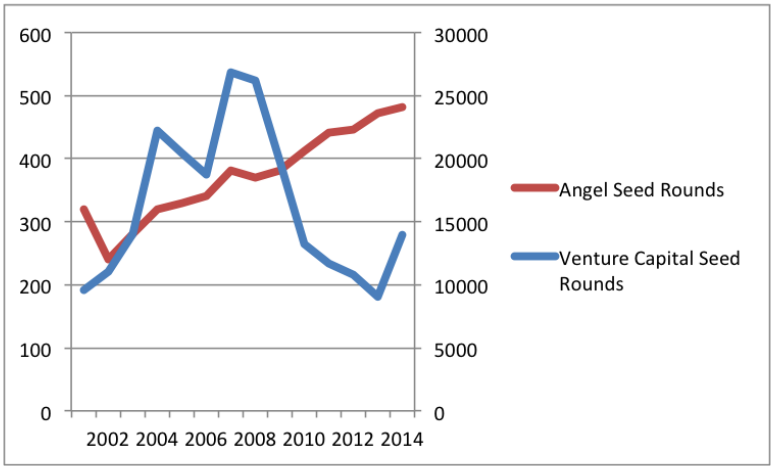

Figure 1: Angel and VC Seed Rounds 2000-2014

To address this data limitation, I collect data from CrunchBase, a database of the start-up ecosystem that tracks companies (start-ups, venture capital firms, angel groups, and accelerators) and individuals (entrepreneurs, venture capitalists, angel investors).[8] CrunchBase, which investors and analysts alike consider the most comprehensive dataset of early-stage start-up activity, describes itself as "the leading platform to discover innovative companies and the people behind them."

There are three main sources of data in CrunchBase. First, CrunchBase monitors Web-based resources such as TechCrunch, an online publisher of technology industry news, and SEC registration data. If a start-up is featured on the World Wide Web, the data is automatically collected and fed into CrunchBase. This includes real-time news on investment rounds, acquisition and IPO exits, new product offerings, and the hiring of top management.

Second, CrunchBase collects information through partnerships with venture funds, angel groups, accelerators, and university programs through the CrunchBase Venture Program.[9] Over 2,000 venture program members supply data about both legacy and new deals in exchange for better access to the CrunchBase API and resources.

The third and perhaps most innovative feature of CrunchBase is that it sources data from the crowd. CrunchBase reports more than 50 thousand individual contributors and more than 2 million active

[8] https://info.crunchbase.com/about/faqs/

[9] In addition to the Venture Program, CrunchBase has teamed up with AngelList, a platform for connecting start-ups and angel investors. AngelList start-ups, job-seekers, and angel investors may opt-in to share data with CrunchBase.

users. Data is constantly reviewed and monitored by both editors and machines to prevent against inaccurate or duplicate information.

These unique features of CrunchBase data provide several distinct advantages. First, it does not require a start-up to receive venture capital financing. This means the CrunchBase sample includes start-ups financed entirely by bootstrapping, angel investors, or crowd-funding, sources otherwise excluded in VentureOne and VentureXpert. Second, the aggregation of data from the greater web mitigates some concerns regarding data selection with self-reporting that affect existing datasets.

CrunchBase was founded in 2007 but include legacy data from the mid-1900s. I limit my sample to 1980-2010 in order to allow sufficient time for analyzing post-founding characteristics.[10] The start-up firm characteristics of interest from CrunchBase include: the entrepreneur(s), founding year, financing amount, investors, and exit event. Additionally, I hand-collect SIC industry classifications for each firm, utilizing CrunchBase product market descriptions and industry categories as additional verifications.

For each entrepreneur, I further collect employment data from LinkedIn. While CrunchBase contains some individual-level employment and demographic data, it remains largely incomplete for the less successful entrepreneurs. LinkedIn provides not only the company at which entrepreneurs were previously employed but also their tenure. Employment and tenure data are necessary for the disambiguation and matching of entrepreneurs to inventors. The ability to track entrepreneurs across time is crucial to the identification strategy explained in the next section.

Comparing CrunchBase data to other datasets, it is worth noting that, in addition to missing a significant amount of early-stage entrepreneurial activity, VentureXpert contains less data on many of the companies in the most innovative industries. Figure 1 presents the distribution of start-ups across different industry groups. While VentureXpert weighs heavily in terms of enterprise software and manufacturing companies, CrunchBase picks up the innovation economy - the biotechnology and the Internet of Things. Additionally, I compare my new data set to accelerator data (even more early-stage start-up companies), which I collect from seed-db.com, and find that accelerators

[10]To address concerns of backfill bias, I limit the sample to 1995-2010, after the dot-com bubble, and obtain economically and statistically similar results.

lack key innovation by primarily focusing on consumer software and apps.

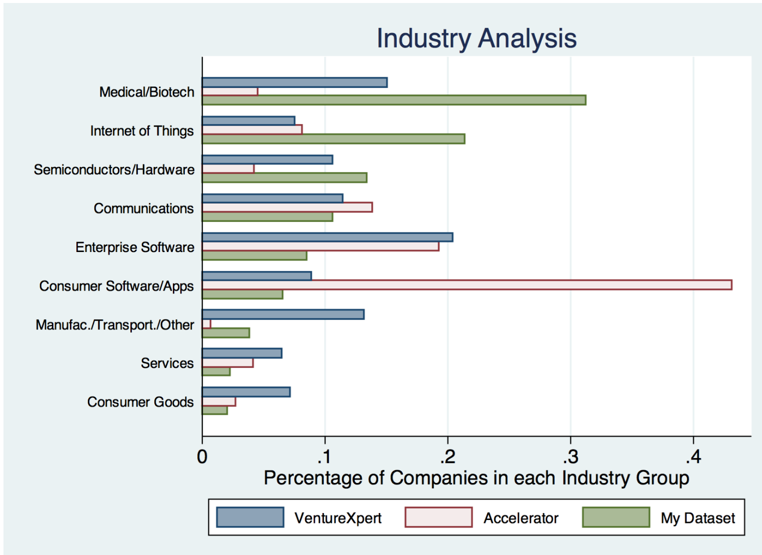

Figure 2: Dispersion of Start-up Companies Across Industries

3.2 Matching Entrepreneurs to Inventors

In order to study innovative output, I need to match the CrunchBase entrepreneurs to inventors in different patent databases. Innovation data from the NBER Patent Database, EPO Patstat, and the IQSS Patent Network database (Lai et al., 2011) comprises all patents applied for between 1975 and 2010. I extend this existing database to 2014.

The matching process proceeds in several steps. I exploit (1) the unique inventor identifiers in Lai et al. (2011), (2) the employment histories of entrepreneurs, and (3) the age and location of the entrepreneur. First, a fuzzy match of entrepreneur name to inventor name retrieves a list of potential unique inventor identifiers from the Lai inventor dataset.[11] For example, entrepreneur Jane Doe from the CrunchBase data will match multiple inventor Jane Does from the patent data. Each inventor Jane Doe will be associated with a set of patents and assignees (the corporation that owns the patent). While each inventor Jane Doe is disambiguated, defined as having a unique identifier in the patent database, the difficulty lies in assigning the correct inventor-to-entrepreneur

[11]The matching algorithm weights last names more heavily than first names since last names are much less likely to be susceptible to abbreviations or mistakes.

match.

In order to solve this problem and remove the false positive name matches, I compare the entrepreneur's past employers with the various inventors' patent assignees. If any of the entrepreneur's past employers match any of the inventors' patent assignees, then the unique inventor identifier associated with that assignee is retrieved and matched to the entrepreneur.[12] In the rare cases of multiple employer-assignee matches within the fuzzy name subsample, I verify again using the location or age of the entrepreneur. With the unique inventor-identifiers in hand, the final merge with the NBER and Patstat patent databases yields a panel that tracks entrepreneurs and their patent portfolios through time and space.

The resulting sample consists of 6,626 entrepreneur-firm pairs with 5,568 unique entrepreneurs. Conditional on patenting, each entrepreneur produces an average 8.17 patents over his or her lifetime. Within the patent database, inventors produce an average of 1.39 patents. Thus, entrepreneurial inventors are much more proficient than the average inventor. In my empirical strategy, I use the prior patent citers (forward citation assignee) as a proxy for potential acquirers. When I focus on entrepreneurs who have produced at least one patent in the four years prior to start-up founding, I am left with 2,484 entrepreneurs with an average of 9.7 patents each.

Finally, I link the patent citers to financial data from Compustat by using unique PDPASS identifiers from the NBER patent database. I match citing patent IDs in order to retrieve a PDPASS and a matched GVKEY for each citing patent assignee. While this initially limits the universe of citers to public companies, firm-level financial data is necessary to construct measures of industry concentration. In order to remove this restriction, I hand-collect SIC industry codes for each citer or potential acquirer. This expands the universe of acquirers to both private companies and companies with missing data in Compustat.

[12]The match procedure, first fuzzy string matching past employers with patent assignees in order to retrieve a firm identifier from the patent data; then, fuzzy string matching names from the firm's inventor pool with entrepreneur names in CrunchBase. This performs less effectively since assignees in the NBER patent database are only disambiguated until 2000. An initial match using last names bypasses the more common abbreviation problems that accompany company names.

3.3 Innovation Outcomes

While patents have long been recognized as a rich data source for the study of innovation and technological change, a considerable limitation is that not all inventions are patented.[13] Barring this limitation, patent citations maintain the distinct advantage of establishing invention, inventor, and assignee networks that are crucial to studying technical change and overlap. Additionally, the incentives to patent are clear. Inventors are granted monopoly rights to their innovation in exchange for disclosure.

Following Halle, Jaffe, and Trajtenberg (2005), I employ *patent application stock* and *forward citations per patent* to measure innovative output and quality. *Patent application stock* refers to the number of patent applications attributed to the inventor. Although patent count is an indicator of "knowledge stock," innovations may vary widely in their technological and economic significance. To this extent, Halle et al. argue for the usefulness of citation count as an important indicator of patent importance, which also allows for gauging the heterogeneity in the "value" of patents.[14] I use the same length of time interval to count patent and citation information, irrespective of application date, in order to allow for comparable measures.

To measure catering, I employ two measures of technological overlap. First, I utilize Jaffe's *Technological Proximity (TP)* measure to gauge the closeness of any two firms' innovation activities in the technology space using patent counts in different technology classes. Technology classes are an elaborate classification system developed by the USPTO for the technologies to which patented inventions belong. Approximately 400 three-digit patent classes and 120,000 patent subclasses exist. Each patent is assigned a class and subclass and an unlimited number of subsidiary classes and subclasses. Halle, Jaffe, and Trajtenberg (2001) further aggregate the 400 patent classes into coarser two-digit technological subcategories. I rely on both three-digit and two-digit technology classes. Since each market may comprise more than two firms, I take the average TP to obtain a product market level measure. In the appendix, I also employ the *Mutual Citation (MC)* measure, which shows the extent to which a firm's patent portfolio is directly cited by another firm. Within my paper's context, this can be interpreted in two directions. One direction, the extent

[13] See Lerner and Seru (2014) for the challenges and the potential for abuse in using patent data.

[14] In particular, the authors find that market value premia is associated with future citations. See Trajtenberg (1990), Harhoff et al. (1999) and Sampat and Ziedonis (2005) for additional support on the relationship between citations and patent quality.

to which the acquirer's patent portfolio directly cites the start-up firm's patent portfolio, captures the immediate usefulness of a start-up firm's innovative activity to a potential acquirer. The other direction, the extent to which the start-up firm's patent portfolio directly cites the acquirer's patent portfolio, captures the improvement or degree of "pushing the envelope" of the acquirer's existing technologies. Both directions represent a convergence between the entrepreneurs' and the acquirers' patent portfolios.

4 Identification Strategy

I begin by examining how the level of concentration in acquirer markets affects patent application stock, citations per patent, and technological proximity in start-up markets. I then address how concentration affects an inventor's incentive to become an entrepreneur initially and I incorporate the analysis into a Heckman specification to address potential self-selection *into* the sample.

The primitive specification of interest relates acquirer concentration to start-up innovation, as follows:

$$Innov_{i,j,t} = \beta_0 + \beta_1 Concentration_{j,t} + \epsilon_{i,j,t}$$

Concentration and measures of innovation are defined for each entrepreneur i facing acquirer market j at time t where t is the time of start-up founding. I index innovation by entrepreneur instead of by firm for two reasons. First, patent portfolios exist at the inventor level. Second, this paper addresses the incentives and innovative choices made by entrepreneurs. By matching entrepreneurs to inventor-level patent data, I construct a history of patent activity for each entrepreneur. This allows me to control for individual-specific characteristics rather than only the firm-level characteristics in prior papers - particularly when accounting for self-selection.

Without stronger exogeneity conditions, the coefficient β_1 is only evidence of correlation between concentration and innovation. One issue preventing a causal interpretation is reverse causality. For example, if innovation increased industry profits, then entry would increase as well. Another problem is self-selection of entrepreneurs into different industries. For entrepreneurs, entering differentially concentrated industries involves trade-offs in incentives, resources, and degree of entrepreneurial risk – all of which can shape innovation outcomes. Examining the causal implications of acquirer concentration on start-up innovation requires a methodology that resolves endogeneity concerns and in particular, eliminates the potential of entrepreneurs to self-select into certain markets.

4.1 Proxy Variable Method

The construction of a measure of acquirer market structure must overcome two major hurdles. First, a majority of entrepreneurs in the sample have not experienced an exit event. Hence, no acquirer exists, which renders the utilization of acquirer market structure impossible. Second, even in the case of acquisition, the final acquisition market need not be the one that the entrepreneur might have envisioned ex-ante when making decisions regarding the start-up venture and catering innovation. For example, Nest Inc.'s ultimate acquisition by Google does not imply that Nest only positioned itself for acquisition by Google. In other words, the analysis requires a variable that captures the ex-ante acquisition market at the time of start-up founding.

I construct a proxy measure of acquirer markets using the patent assignees of citations (citers) received by each entrepreneur before start-up founding. As a simple example, consider Nest Inc. founder Tony Fadell. Prior to starting Nest Inc., he worked as an engineer at Apple. During that time, he was the inventor on one patent application, cited by Samsung, IBM, and Google. By construction, the industries of Samsung, IBM, and Google comprise Tony's acquirer market.

Using the patenting history of entrepreneurs, I test and conclude that citers of prior patents are ex-ante the most likely future acquirers, and accurately predict ex-post acquisition markets.[15] I verify Woolridge proxy conditions to confirm that the coefficients on the proxy variables are estimated consistently. The first condition is that the proxy variable should be redundant in the structural equation. Using the subsample of entrepreneurs that do experience an acquisition exit, I show empirically that, given the true acquirer market, citation-based measures of market structure are not predictive of innovation ex-post. This implies that the market structure of acquirers is indeed the mechanism that incentivizes innovation and catering. The second condition is that conditional on the proxy, the acquirer market structure and the other regressors are not jointly determined by further factors.

[15] There has been extensive industry interest in employing algorithms to predict potential acquirers. See https://www.cbinsights.com/blog/acquirer-predictions/

I measure the acquirer market structure using both citers' concentration and citers' market size. The HHI of an industry k is defined as:

$$HHI_k = \sum_i s_i^2$$

where s_i represents the market share of firm i in industry k. HHI measures the size of firms in relation to the industry and indicates the amount of competition among them. Thus, HHI can range from 0 to 1, moving from perfect competition to a monopolistic industry. Increases in the HHI indicate a decrease in competition and an increase in market power.

For each start-up entrepreneur i, the acquirer market concentration is defined as:

$$HHI\_citers_{i,t} = \frac{1}{TotalCitations_{i,t-5}} \sum_{k=1}^{N} Citations_{i,k,t-5} * HHI_{k,t}$$

where $Citations_{i,k,t-5}$ represents the number of citations received by entrepreneur i from firms in industry k in the four years between $t-5$ and the year before start-up founding $t-1$. [16] Total $Citations_{i,t-5}$ represents the total number of citations received by entrepreneur i from $N = \sum k$ industries. I construct a similar measure of citers' market size:

$$Size\_citers_{i,t} = \frac{1}{TotalCitations_{i,t-5}} \sum_{k=1}^{N} Citations_{i,k,t-5} * sales_{k,t}$$

where $sales_{k,t}$ represents the sales of industry k at time t. Sales and HHI are both calculated at start-up founding time t.[17] Both measures are calculated at the entrepreneur level and represent the specific acquisition market structure that he or she faces.

I demonstrate the proxy calculation continuing with the example of Tony Fadell facing an ex-ante acquisition market that consists of the industries of Samsung, IBM, and Google. Samsung operates in SIC industry 3631, IBM operates in SIC industry 3570, and Google operates in SIC industry 7370. The concentration of Tony's acquirer market is then:

$$HHI\_citers_{Tony,t} = \frac{1}{3}[HHI_{3631} + HHI_{3570} + HHI_{7370}]$$

It is worth emphasizing three features of these measures. First, a prior citer of entrepreneur i can be the prior employer of entrepreneur i. This appears in the patent database as a self-citation. This

[16] In robustness checks, I change the time interval from four years to both three years and five years and find similar results.

[17] In previous versions, I use the max sales inside each industry instead of total sales. The results are robust to either specification.

captures the common phenomenon that many start-ups end up acquired by companies or industries at which the entrepreneurs previously worked. Second, the companies that cite entrepreneur i more frequently receive more weight in the calculation of acquirer market concentration. This captures the intuition that companies who cite a certain patent more often have more "use" for said patent and would likely experience higher returns to a possible acquisition. Last, the construction of acquirer markets does not rely solely on the traditional product market classifications (SIC) but instead accounts for the technology space of firms. Indeed, to the extent that most acquisitions cross industry lines, studying purely horizontal mergers is not informative.[18]

4.2 CEM Matching

The main empirical strategy employs the coarsened exact matching procedure (Iacus et al. 2011) to construct treatment and control groups balanced on pretreatment covariates. [19] The primary reason I chose to use CEM instead of a propensity score method was that CEM offers the ability to select the balance of the treatment and control group ex-ante. The purpose of this strategy is to identify control groups that follow a parallel trend to treatment groups, had the treatment not occurred. I exploit the employment and patent histories of entrepreneurs by focusing on two sets of pretreatment variables: entrepreneur *innovativeness* and prior *industry* before start-up founding at time t.

Specifically, I implement this by dividing the sample into two groups (high and low), based on the mean of the proxy variable, $HHI\_citers_{i,t}$. For each entrepreneur i with $HHI\_citer_{i,t}$ in the high group, I employ CEM to identify a similar entrepreneur j with $HHI\_citer_{j,t}$ in the low group. The entrepreneurs are similar in the sense that they work in the same SIC three-digit industry from $t-5$ to t, and they possess the same number of patents and citations per patent during that time.

The ideal experiment in my setting would be to flip a coin for each entrepreneur. If the coin lands on heads, the entrepreneur is assigned a concentrated acquirer market. If the coin lands on tails, the entrepreneur is assigned a competitive acquirer market. However, I am concerned that

[18] A canonical example of an acquisition for innovation that spans industry lines is retail giant Walmart's 2010 acquisition of Vudu, a content delivery and media technology company

[19] "CEM... generates matching solutions that are better balanced and estimates of the causal quantity of interest that have lower root mean square error than methods under older existing class, such as based on propensity scores, Mahalanobis distance, nearest neighbors, and optimal matching" (Iacus et al. 2011)

an entrepreneur's choice of past and future industries is correlated with his or her innovation. If this choice of prior or current industry is non-random, this will generate a bias in the β_1 coefficient of interest. For example, higher ability entrepreneurs may choose to enter into more competitive industries and maintain a higher level of ex-post innovation. Using citers of prior patents assigned to the entrepreneurs only partially alleviates this concern. Higher ability entrepreneurs may also choose to enter into prior industries differentially, in which case, the proxy remains susceptible to the same bias.

Matching on prior innovativeness at least partially addresses the potential that highly innovative people tend to *systematically* self-select into more (or less) competitive industries. Matching on prior industry addresses the potential that selection into prior industries is correlated with selection into expected industries and innovation. Matching on industry along with industry FE breaks this link between market choices elected by the entrepreneur and the acquisition market. The only variation that remains in the specification is variation that is orthogonal to innovation residuals – restricted to the proxies of innovativeness used in the analysis.

Instead of flipping a coin for each entrepreneur, I now flip a coin for each pair of matched entrepreneurs. With just one flip, I can "randomly" assign one entrepreneur to a concentrated acquirer market and another to a competitive acquirer market. The key identifying assumption is that $HHI\_citer_{i,j,t}$ is randomly assigned to entrepreneurs conditional on matching.

$$\epsilon_{i,t} \perp HHI\_citers_{i,j,t} | SIC_{i,t-1}, Innov_{i,pre-t}$$

This implies that, for a given innovativeness of entrepreneur i and a given industry that entrepreneur i works in at time $t-1$, the assigned concentration of potential acquirers resembles an assignment by coin toss. Put differently, the underlying assumption for this methodology is that there are no additional correlates of unobserved entrepreneur characteristics and the market structure of prior citers. By removing entrepreneur observations that fail to have a match, I am removing observations that are "different" and thus, most susceptible to selection bias.

To solidify our understanding of the identification strategy, imagine two entrepreneurs, Tony (again) and Sean. Tony and Sean had both worked in the same industry before pursuing entrepreneurship. They had also produced the same number of patents with the same number of forward citations before becoming start-up founders. However, their patents received citations from companies in differentially concentrated industries. Thus, they faced two different potential acquirer markets.

Figure 3 illustrates this example. The matching procedure ensures that Tony and Sean have approximately equal distributional properties in terms of prior innovativeness and industry choice, and the regression specification exploits the variation in the "treatment" (acquirer market concentration) to identify the causal impact of concentration on start-up innovation.

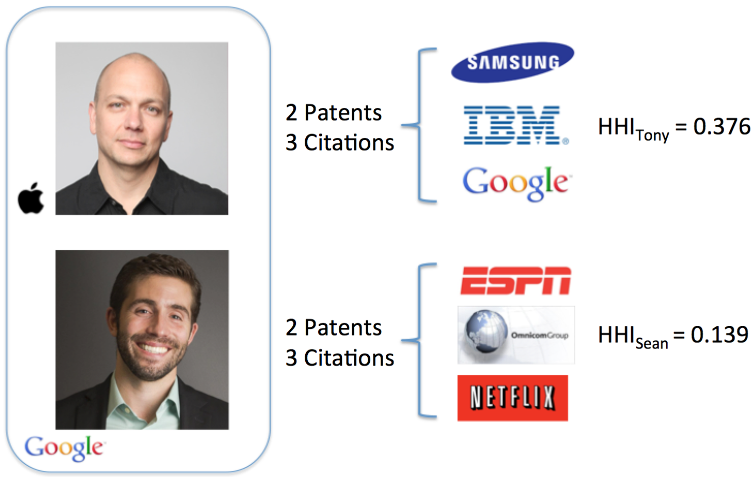

Figure 3: Matching Methodology Example

I then utilize the matched sample to isolate the causal effects of concentration on start-up innovation using the following specification:

$$Innov_{i,t} = \beta_0 + \beta_1 HHI\_citers_{i,t} + \beta_2 \overline{\mathbf{Innov}}_{\mathbf{i,t-5}} + \zeta_{j,t-5} + \zeta_{j,t} + \sigma_t + \epsilon_{i,t}$$

$\overline{\mathbf{Innov}}_{\mathbf{i,t-5}}$ is a vector of patent variables that controls for innovation before start-up founding from $t-5$ to $t-1$. In all specifications, I use both prior patent count and prior citations per patent to measure pre-innovation. I also control for SIC industry (pre- and post-) and time fixed effects - $\zeta_{j,t-5}$, $\zeta_{j,t}$, and σ_t, respectively. Note that since sample observations are already matched on prior innovation and SIC industry, controlling for $\overline{Innov}_{i,t-5}$ and pre-entrepreneurship industry fixed effects will not affect the consistency of our estimator but may improve efficiency.

I employ the same empirical methodology using $Size\_citers_{i,t}$ as a measure of acquirer market structure. Furthermore, I show results incorporating both measures.

$$Innov_{i,t} = \beta_0 + \beta_1 HHI\_citers_{i,t} + \beta_2 Size\_citers_{i,t} + \beta_3 \overline{\mathbf{Innov}}_{\mathbf{i,t-5}} + \zeta_{j,t-5} + \zeta_{j,t} + \sigma_t + \epsilon_{i,t}$$

4.3 Heckman Selection Model

The proxy variable and CEM address potential selection within the sample. Another important question concerns the degree of selection *into* the sample, i.e., the determinants of the propensity of inventors to become entrepreneurs. If entrepreneurs position their innovation to be attractive acquisition targets, they will also position their entrepreneurship choices. For example, if a concentrated acquiring market deters low-quality inventors from becoming entrepreneurs, the quality of entrepreneurs in those industries would be higher because the low end of the distribution would be missing. To address this concern, I employ the two-stage Heckman correction model for selection.

Heckman's sample selection model focuses on correcting selection bias when the dependent variable is non-randomly truncated. In my context, the incidental truncation occurs because the outcome variable, post-entrepreneurship innovation, is only observed for inventors who *choose* to become entrepreneurs. The proposed two-step model to correct for this type of selection involves 1) the selection equation considering a portion of the sample whose outcome is observed and mechanisms determining the selection process, and 2) the regression equation considering mechanisms determining the outcome variable. The goal of this model is to utilize the observed variables to estimate regression coefficients β for all inventors.

In the first stage, I construct my proxy variable for every inventor across time using the full patent database. Each observation represents an inventor-year pair associated with a specific $HHI\_citer_{i,t}$ and $Size\_citer_{i,t}$. The outcome variable is a dummy variable E_{it} for whether inventor i enters entrepreneurship at time t. The specification is as follows:

$$Prob(E_{it} = 1|Z) = \Phi(HHI\_citers_{it} + Size\_citers_{i,t} + Z\gamma_2) \qquad \text{(Selection Equation)}$$

where Z is a vector of explanatory variables including $\overline{Innov}_{i,t-5}$, industry, and time fixed effects. In the second stage, I use the transformation of the predicted individual probabilities as an additional explanatory variable:

$$\begin{aligned} Innov_{i,t} = {} & \beta_1 HHI\_citer_{i,t} + \beta_2 Size\_citers_{i,t} + \beta_3 \overline{\mathbf{Innov}}_{\mathbf{i,t-5}} \\ & + \beta_3 E_{it} + \zeta_{j,t-5} + \zeta_{j,t} + \sigma_t + u_{it} \end{aligned} \qquad \text{(Regression Equation)}$$

While this methodology directly addresses concerns about entrepreneurial selection, it also provides an answer to an important question in both the industrial organization and entrepreneurship literature. The first stage is a direct test of the impact of market structure on entry.

5 Empirical Results

5.1 Summary Statistics

The final matched sample consists of 1,910 entrepreneur-firm pairs between 1980 and 2010, including the entrepreneur's entire patenting and employment history. The majority of start-ups are located in metropolitan areas such as Silicon Valley, Boston/Cambridge, Los Angeles/San Diego, and New York. Table I shows the dispersion of start-ups across geographic space. Table II provides summary statistics on the proxy and outcome variables. On average, an inventor produces 5.368 patents before and 6.460 patents after entrepreneurship. Each patent produced before entrepreneurship elicits approximately 8.391 forward citations, whereas patents produced after entrepreneurship generate an average of 4.3 forward citations. As expected, the citations per patent distribution is heavily right-skewed. For the empirical analysis, I take logs in order to transform the distribution to a normal distribution.

In my sample, 1,471 entrepreneurs experience an external funding round. This could occur in the form of an angel seed round or a crowd-funding event. Out of that number, 918 firms receive investments from a venture capital firm. I include these variables in my analysis because the empirical literature has found a relationship between VC investment and innovation output. Among others, Gonzalez-Uribe (2013) found that VC investment increases patent innovation by increasing the number of citations to a given patent.

Furthermore, 408 out of 1,910 matched entrepreneurs exit through the acquisition market, while only 162 exit through an initial public offering. These exit frequencies are higher than start-up market average exit rates, indicating that patenting entrepreneurs are more successful than are non-patenting entrepreneurs. While this calls into question the generalizability of the results, it does not affect the interpretation of the results.

The average HHI among citers is 0.233, and the average market size among citers is $27 billion per year. To put this in context, the entertainment and games software industry, with a Herfindahl index of 0.235, generated more than $20 billion in sales in 2014. This industry is considered moderately to highly concentrated.[20] While some large companies in the market have economics of scale in manufacturing and distribution, small companies can compete successfully by developing differentiated products. Pharmaceuticals, on the other hand, maintains an average yearly HHI of 0.425. One distinction to consider is that a high concentration does not necessarily imply a large market size. The industrial organization literature has often confounded these two different dimensions of market structure. Size and HHI have a low 0.0295 correlation, which is not statistically significant.

Additionally, substantial variation exists in both measures within broader industry sectors. Table III illustrates the distribution of entrepreneurs across industry sectors and industry groups. The classifications are broad agglomerations of industry categories, as found on CrunchBase. In my empirical analysis, I use more granular measures such as three-digit and four-digit SIC codes for industry. The dispersion of industries represented in Table III indicates that CrunchBase entrepreneurs are mainly venturing in the information technology and biotech industries. I demonstrate that my empirical results are robust across industry classes.

5.2 Matched Proxy Regressions on Innovation

I investigate whether and how the acquisition market impacts innovation output and catering of the entrepreneur after start-up founding. I run the initial regressions using the patent count, citations per patent, and technological proximity measure as the dependent variables.

Table IV displays the matched regression results using post-entrepreneurship patent count as the dependent variable. Column 1 represent the baseline specification with HHI_citer as the main explanatory variable. Columns 2 and 3 add in additional industry level fixed effects. Industry (Pre) refers to the three-digit SIC industry in which the entrepreneur had been employed prior to current start-up. Industry (Post) refers to the three-digit SIC industry in which the entrepreneur and start-up currently are. Column 4 mimics the specification in Column 3 but with $Size\_citers$ as the main explanatory variable. Column 5 includes both dimensions of market structure. In

[20]The U.S. Department of Justice uses HHI for evaluating anti-competitive mergers. Industries between 0.1 and 0.2 are considered moderately concentrated.

all specifications, acquirer concentration produces no statistically significant effect on patent count after entrepreneurial founding.

The coefficient on HHI_citer is always negative but statistically indistinguishable from zero. This implies that facing a more concentrated acquisition industry does not lead to more innovation in terms of patent count. The coefficient on $Size\_citer$ is positive and slightly significant in Column 4. However, this coefficient loses significance in a specification with HHI_citer in Column 5.

In other words, post-entrepreneurship patent output appears unaffected. One potential explanation is that both escape competition and scaling forces are at play: Concentrated and large industries may generate more acquisition surplus due to scaling, while competitive industries may experience a greater need for innovation in order to capture market share. Thus, the competing forces may simply generate a net 0 effect on the incentives of the entrepreneur to pursue innovation.

In terms of the control variables, each additional prior patent increases future patent innovation, whereas prior citations produce no effect on future patent production. This is unsurprising since inventors who patent before entrepreneurship are likely to continue patenting after entrepreneurship.

Turning to my measure of innovation quality, I use log citations per patent as the right-hand side variable in Table V. Here, I estimate a significant impact of market structure. In Column 1, increasing HHI_citer from perfectly competitive to monopolistic leads to a 122% increase in average forward citation per patent. This is both large in magnitude and statistically significant at the 1% level. The average concentration for an acquiring industry is 0.233 with standard deviation 0.128 implying that a one standard deviation increase in concentration will increase citations by approximately 15-16%. Given that the average number of citations per patent in the sample is 13, this implies that, when comparing a concentrated industry such as pharmaceuticals to a more competitive industry such as software, the quality of patents produced by entrepreneurs increases by two citations per patent. Even with the addition of fixed effects in Columns 2 and 3, the coefficient on HHI_citer remains constant and significant, alleviating concerns of selection on unobservables.

In Column 4, I re-run the specification with $Size\_citer$ as a proxy for market structure. Increasing size by one standard deviation increases average citations per patent by 9%. When I account for both dimensions of acquirer market structure in Column 5, I find HHI_citer maintains a 15% effect on citations, controlling for market size. Size maintains a 7% effect on citations.

The results in Table V are consistent with the Schumpeterian hypothesis that more concentrated industry encourages innovation when innovation is measured with citations per patent as opposed to patent count. The innovation literature argues that citation-based measures more accurately reflect significant innovations (innovations that have a more widespread impact) and technological progress. To this degree, the simple patent count measure picks up a considerable amount of minor patenting, driven by the need to product differentiate in competitive industries.

The coefficient on the VC dummy is also positive and significant, implying that VC investment incentivizes innovation output by entrepreneurs. Interestingly, this effect is similar in magnitude to that which the existing literature on the role of venture capital on innovation. However, it is unclear to what extent the estimate reflects the causal impact of venture capital funding on innovation versus venture capitalists selecting highly innovative firms.

Finally, I test whether increasing concentration leads to catering. Do entrepreneurs either engage in proximal innovation relative to potential acquirers in order to increase technological synergies in an acquisition, or in differentiated innovation in order to avoid cannibalization of previous products?

Table VI shows that if acquirer concentration increases, technological proximity between the entrepreneur and potential acquirers increases as well. Entrepreneurs facing acquirers in concentrated industries tend to innovate in technology areas in which potential acquirers are also innovating. A one standard deviation increase in HHI_citer increases technological proximity by 9%, while a one standard deviation in size increases technological proximity by 5%.

These economic magnitudes suggest that entrepreneurs position for acquisitions in concentrated markets by shifting their innovation in the direction of potential acquirers. By innovating in the same technological areas as potential acquirers, entrepreneurs position their inventions for the acquirer to easily utilize and scale.

5.3 Heckman Selection Model

I now move back one step further in the inventor's decision making process and account for acquisition markets affecting the propensity of inventors to become entrepreneurs. In numerous prior studies concerning the determinants of entrepreneurship, a key challenge is to establish a starting sample of potential entrepreneurs. What is the relevant sample of people to study? Here, I have a natural starting sample – inventors. That said, I cannot speak to the entire universe of entrepreneurs, but only to patenting entrepreneurs.[21]

5.3.1 First Stage: Entrepreneurship

I employ a standard two-stage Heckman selection model to address the selection of entrepreneurs. The first stage is a probit with the outcome variable being a dummy variable for entrepreneurship. The specification is:

$$Prob(E_{it} = 1|Z) = \Phi(HHI\_citers_{it} + Size\_citers_{it} + Z\gamma_2)$$

where Z is a vector of explanatory variables including $\overline{Innov}_{i,t-5}$, industry and time fixed effects. This specification uses variation across inventors and across time to identify whether market structure affects the decisions of inventors to become entrepreneurs.

Table VII shows that entry into entrepreneurship is higher when industries are less concentrated. A one standard deviation increase in HHI_citer decreases entrepreneurship by 4%. Size produces a similar effect. This is consistent with two anecdotal facts. First, fragmented markets attract more entry. Facing concentrated acquisition markets presents a higher risk of potential acquirers extending into the product market with or without the inventor. As a result, inventors are more hesitant to "face off" against large monopolists in the case of no acquisitions. Second, even if the possibility of acquisition is high, entrepreneurs facing monopolists are unlikely to extract a high acquisition price due to the lack of outside options and low bargaining power. A low acquisition price deters inventors from entering into entrepreneurship, as compared to staying in the waged labor market.

[21] Unfortunately, this means I cannot identify what caused Mark Zuckerberg to become an entrepreneur and create Facebook.

5.3.2 Second Stage: Innovation (Conditional on Entry)

In the second stage, I rerun the previous specification but incorporate transformation of the predicted individual probabilities as an additional explanatory variable:

$$Innov_{i,t} = \beta_1 HHI\_citer_{i,t-5} + \beta_2 Size\_citer_{i,t-5} + \beta_3 \overline{Innov}_{i,t-5} + \beta_4 E_{it} + FE + u_{it}$$

Table VIII displays the second-stage Heckman results. The results are economically and statistically similar to the matched proxy model. I find that conditional on entry, citations and technological proximity increase with market concentration and size, but the effect on patent count is statistically indistinguishable from zero. The economic magnitude ranges from increases of 12% in citations and 5% in technological proximity from HHI_citer to increases of 5% in citations and 4% in technological proximity from $Size\_citer$. The stability of the coefficients lends reassurance to the strength of the identification strategy. Concentrated acquirers are best suited for scaling technologically similar innovations because of the applicability of the innovation to their entire product line.

5.4 Subsample Analysis

While my results indicate that entrepreneurs increase the quality of patents and cater technological proximity when facing concentrated acquirer markets, I test whether the results are sensitive to the different product markets in which start-ups reside. One might argue that while patents represent an important indication of innovation in the pharmaceutical industry, they have no bearing on the software industry. Furthermore, the effect of the acquisition market and its corresponding incentives may differ across industries.

To analyze intra-industry effects, I separate the sample of entrepreneurs into three broad industry sectors based on the product market of their start-up. While more granular measures of industry exist, a balance must be attained between maintaining enough observations for statistical power and identifying finer product market spaces. In my sample, 1,156 entrepreneurs operate within the information technology sector, 594 entrepreneurs in the medical/biotech sector, and 160 entrepreneurs in the non-high technology sector.

The information technology sector is the driver of the new economy and is particularly relevant to the changing landscape of entrepreneurial finance. Table IX, Panel A presents the results for this subsample. Columns (1) and (2) regress post-entrepreneurship patent count on HHI_citer and $Size\_citer$. Similar to the prior results, the coefficient is not statistically significant. Interestingly, the positive correlation between prior patents and future patents decreases to approximately 0.08 and is only significant at the 5% level. This implies that an inventor with numerous patents does not necessarily patent at the same intensity after becoming an entrepreneur. This could indicate a shift in the type of companies founded by inventors. However, despite a smaller focus on patenting, strong incentives still exist for entrepreneurs to increase patent quality and, in particular, to innovate in technologically similar areas as potential acquirers. Columns (3) and (4) show the results on log citations per patent while Columns (5) and (6) show the results on technological proximity. The magnitudes are similar to the full specification.

The results also extend to the medical/biotech sector (Table IX, Panel B). The magnitudes on citations per patents and technological proximity are larger and statistically significant at the 1% level. This can be attributed to one of two potential reasons. First, stronger acquisition incentives may exist in this sector. Anecdotally, funding for research is difficult and small biotechnology firms depend on either strategic alliances or full acquisitions by large pharmaceutical firms for survival. Second, scaling may produce non-linear benefits. Since the medical/biotech sector is dominated by heavily concentrated potential acquirers, the benefits of scale and monopoly power are even larger.

While I find that the results are generalizable across both the information technology and the medical/biotech sectors, the results do not seem to hold in the non-high technology sector. This can either be because incentives are driven by a different exit model in the non-high technology sector or because I lack sufficient observations and thus, the statistical power to obtain precision on the point estimates.

6 Conclusion

Despite recent academic and industry focus, relatively little academic work explores the determinants of innovation in finance, and in particular, within the start-up setting. In this paper, building on Schumpeter's ideas, I propose market structure as a key determinant of entrepreneurship and innovation. The distinction in this paper is to suggest a different channel for the role of market structure – specifically, by affecting acquisition surplus and premiums.

The bulk of the current paper focuses on developing and testing an empirical strategy free of endogeneity and selection problems. I construct an entirely new dataset comprising entrepreneurs from CrunchBase and their employment history from LinkedIn, which I match to their patents from EPO Patstat and the NBER patent database. I proxy for acquirer markets utilizing citers of an entrepreneur's prior patents with the intuition that ex-ante, the most likely acquirers are the people most interested in prior patents. I test and confirm these conditions.

I find consistent and causal effects of market structure on entrepreneurship and start-up innovation. First, I show that inventors are ex-ante less likely to become entrepreneurs when facing large potential acquirers in concentrated industries. Next, I find that an entrepreneur's incentive to produce high quality innovations increases with acquirer market concentration and size. However, these high quality innovations tend to occur within the same technological classes as the innovations of potential acquirers.

Overall, my results highlight how entrepreneurs position their human capital and innovation for acquiring markets. Entrepreneurs cater to and engage in proximal innovations in order to present themselves as attractive acquisition targets – evidence of the role that technological synergies play in acquisitions.

Tables

Table I: Geographic Dispersion of CrunchBase Start-Ups

Region	Frequency	Percent
Silicon Valley	642	34.756
Boston/Cambridge, MA	185	10.005
Southern California	171	9.246
New York, NY	101	5.453
Austin, TX	68	3.698
Seattle, WA	60	3.272
Boulder, CO	42	2.276
Philadelphia, PA	28	1.517
Newark, NJ	27	1.470
Other (U.S)	305	16.548
International	217	11.759
Total	1846	100

Notes Table I displays the geographic locations of CrunchBase Start-Ups for the final sample. The total number is less than 1910 because 1) geographic information is not available for every Start-Up and 2) each observation is a firm, not an entrepreneur-firm.

Table II: Descriptive Statistics

Panel A: Continuous Variables

Variable	N	Mean	Std. Dev.	Min	Max
HHI_citer	1,910	0.233	0.128	0.024	1
Size_citer	1,910	2.725	1.469	0.108	16.368
Patent Count$_{\text{t-5,t-1}}$	1,910	5.368	11.602	1	181
Forward Citations per Patent$_{\text{t,t+4}}$	1,910	8.391	16.364	.429	170
Patent Count$_{\text{t,t+4}}$	1,910	6.460	14.450	0	155
Forward Citations per Patent$_{\text{t,t+4}}$	1,910	4.298	13.071	0	137.8

Panel B: Categorical Variables

Variable	Frequency	Percent
Acquisition	408	21.361
IPO	162	8.482
Investment	1,471	77.015
VC Funding	918	48.062

Notes Table II displays summary statistics for the variables in the sample. HHI_citer and Size_citer are entrepreneur-specific proxy variables capturing the level of competition and market size of acquirers. $HHI\_citers_{i,t} = \frac{1}{TotalCitations_{i,t-5}} \sum_{k=1}^{N} Citations_{i,k,t-5} * HHI_{k,t}$ and $Size\_citers_{i,t} = \frac{1}{TotalCitations_{i,t-5}} \sum_{k=1}^{N} Citations_{i,k,t-5} * Size_{k,t}$, where k indexes the industry of the citer of entrepreneur i. Size is measured in ten billions. Investment is an indicator variable equal to 1 if the entrepreneur discloses any source of external funding. VC dummy is an indicator variable for whether an entrepreneur received venture capital investment.

Table III: Industry Dispersion of CrunchBase Start-Ups

Panel A: Industry by Sector

	Frequency	**Percent**	**Cumulative**
Information Technology	1,156	60.523	60.524
Medical/Biotech	594	31.099	91.623
Non-High Technology	160	8.377	100
Total	1910	100	

Panel B: Industry by Group

	Frequency	**Percent**	**Cumulative**
Medical/Biotech	594	31.099	31.099
Internet of Things	534	27.958	59.057
Semiconductors/Hardware	256	13.403	72.461
Communications	204	10.680	83.141
Computer Software	162	8.481	91.623
Manufac./Transport./Other	75	3.927	95.549
Services	43	2.251	97.801
Consumer Goods	42	2.198	99.999
Total	1910	100	

Notes Table III displays the product market industries of CrunchBase entrepreneurs for the final sample. Industry sector is the broadest classification. Industry group is sub-classifications under sector. In the empirical analysis, I use granular measures of industry such as three and four digit SIC codes.

Table IV: Patent Count

VARIABLES	(1) Patents$_{t,t+4}$	(2) Patents$_{t,t+4}$	(3) Patents$_{t,t+4}$	(4) Patents$_{t,t+4}$	(5) Patents$_{t,t+4}$
HHI_citer	-2.955	-1.738	-1.885		-1.721
	(2.378)	(2.581)	(1.892)		(2.550)
Size_citer				0.743*	0.485
				(0.478)	(0.539)
Patents$_{t-5,t-1}$	0.295***	0.630***	0.583***	0.625***	0.574***
	(0.028)	(0.025)	(0.016)	(0.028)	(0.021)
Citations$_{t-5,t-1}$	0.004	0.002	0.003	0.003	0.004
	(0.005)	(0.003)	(0.005)	(0.004)	(0.004)
VC Dummy	0.132	0.134	0.120	0.122	0.120
	(0.579)	(0.560)	(0.541)	(0.522)	(0.521)
Observations	1,910	1,910	1,910	1,910	1,910
R-squared	0.077	0.139	0.243	0.243	0.245
Year Time FE	YES	YES	YES	YES	YES
Industry (Pre) FE	NO	YES	YES	YES	YES
Industry (Post) FE	NO	NO	YES	YES	YES

Notes Table IV reports estimates from OLS regressions using the matched sample. HHI_citer and Size_citer are entrepreneur-specific proxy variables capturing the level of competition and market size of acquirers. The variable Patents$_{t-5,t-1}$ is the entrepreneur's patent count before start-up founding. The variable Citations$_{t-5,t-1}$ is the average citation per patent attributed to the entrepreneur before start-up founding. VC dummy is an indicator variable for whether an entrepreneur received venture capital investment. Industry (Pre) FE controls for the entrepreneur's prior three-digit SIC industry while Industry (Post) FE controls for the three-digit SIC industry after start-up founding. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table V: Patent Citations

	(1)	(2)	(3)	(4)	(5)
VARIABLES	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$
HHI_citer	1.223***	1.299***	1.125**		1.196**
	(0.356)	(0.355)	(0.474)		(0.485)
Size_citer				0.062**	0.047*
				(0.026)	(0.028)
$\text{Patents}_{t-5,t-1}$	-0.004	-0.014	-0.013	-0.017	-0.020
	(0.030)	(0.049)	(0.076)	(0.092)	(0.080)
$\text{Citations}_{t-5,t-1}$	0.363***	0.455***	0.620***	0.489***	0.486***
	(0.030)	(0.030)	(0.028)	(0.030)	(0.028)
VC Dummy	0.011**	0.011**	0.010**	0.011*	0.009*
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Observations	1,910	1,910	1,910	1,910	1,910
R-squared	0.127	0.135	0.183	0.150	0.191
Year Time FE	YES	YES	YES	YES	YES
Industry (Pre) FE	NO	YES	YES	YES	YES
Industry (Post) FE	NO	NO	YES	YES	YES

Notes Table V reports estimates from OLS regressions using the matched sample. HHI_citer and Size_citer are entrepreneur-specific proxy variables capturing the level of competition and market size of acquirers. The variable $\text{Patents}_{t-5,t-1}$ is the entrepreneur's patent count before start-up founding. The variable $\text{Citations}_{t-5,t-1}$ is the average citation per patent attributed to the entrepreneur before start-up founding. VC dummy is an indicator variable for whether an entrepreneur received venture capital investment. Industry (Pre) FE controls for the entrepreneur's prior three-digit SIC industry while Industry (Post) FE controls for the three-digit SIC industry after start-up founding. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table VI: Technological Proximity

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Tech. $Prox_{t,t+4}$	Tech. $Prox_{t,t+4}$	Tech. $Prox_{t,t+4}$	Tech. $Prox_{t,t+4}$	Tech. $Prox_{t,t+4}$
HHI_citer	0.513***	0.548***	0.681***		0.568***
	(0.171)	(0.127)	(0.203)		(0.145)
Size_citer				0.034**	0.031**
				(0.013)	(0.013)
$Patents_{t-5,t-1}$	0.021	0.038	0.022	0.041	0.035
	(0.019)	(0.022)	(0.037)	(0.050)	(0.035)
$Citations_{t-5,t-1}$	0.029	0.028	0.033	0.038	0.046
	(0.036)	(0.049)	(0.056)	(0.056)	(0.058)
VC Dummy	0.008*	0.008	0.011	0.008	0.009
	(0.005)	(0.018)	(0.019)	(0.019)	(0.018)
Observations	1,910	1,910	1,910	1,910	1,910
R-squared	0.138	0.210	0.263	0.246	0.273
Year Time FE	YES	YES	YES	YES	YES
Industry (Pre) FE	NO	YES	YES	YES	YES
Industry (Post) FE	NO	NO	YES	YES	YES

Notes Table VI reports estimates from OLS regressions using the matched sample. The technological proximity measure is an average of the patent overlap between the entrepreneur and potential acquirers. HHI_citer and Size_citer are entrepreneur-specific proxy variables capturing the level of competition and market size of acquirers. The variable $Patents_{t-5,t-1}$ is the entrepreneur's patent count before start-up founding. The variable $Citations_{t-5,t-1}$ is the average citation per patent attributed to the entrepreneur before start-up founding. VC dummy is an indicator variable for whether an entrepreneur received venture capital investment. Industry (Pre) FE controls for the entrepreneur's prior three-digit SIC industry while Industry (Post) FE controls for the three-digit SIC industry after start-up founding. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table VII -

Likelihood of Entrepreneurship

	(1)	(2)
VARIABLES	Entrepreneurship	Entrepreneurship
HHI_citer	-0.316 **	-0.313**
	(0.126)	(0.125)
Size_citer		-0.036*
		(0.020))
$\text{Patents}_{\text{t-5,t-1}}$	0.002**	0.002**
	(0.001)	(0.001)
$\text{Citations}_{\text{t-5,t-1}}$	0.000*	0.000*
	(0.000)	(0.000)
Observations	3,396,076	3,396,076
Pseudo R-squared	0.094	0.096
Year Time FE	YES	YES
Industry (Pre) FE	YES	YES
Industry (Post) FE	YES	YES

Notes Table VII reports the results from the entrepreneurship probit regression, $Prob(E_{it} = 1|Z) = \Phi(\gamma_1 HHI\_citers_{it} + \gamma_2 Size\_citers_{it} + Z\gamma)$. The sample consists of all inventors in the patent database. For each inventor in each year, I construct HHI_citer and Size_citer in the same way as in the main sample. The outcome variable, Entrepreneurship, is equal to 1 if an inventor i enters into a new venture at time t. Industry and time FE are included. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table VIII: Heckman Second Stage

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\text{Patents}_{\text{t,t+4}}$	$\text{Patents}_{\text{t,t+4}}$	$\text{Citations}_{\text{t,t+4}}$	$\text{Citations}_{\text{t,t+4}}$	Tech. $\text{Prox}_{\text{t,t+4}}$	Tech. $\text{Prox}_{\text{t,t+4}}$
HHI_citer	-1.416	-1.506	1.109***	0.939***	0.440**	0.412**
	(1.850)	(1.884)	(0.217)	(0.210)	(0.200)	(0.194)
Size_citer		0.539		0.038*		0.025*
		(0.834)		(0.026)		(0.014)
$\text{Patents}_{\text{t-5,t-1}}$	0.377***	0.367***	0.017	0.017	-0.025	-0.028
	(0.022)	(0.025)	(0.021)	(0.022)	(0.072)	(0.074)
$\text{Citations}_{\text{t-5,t-1}}$	0.004	0.004	0.438***	0.427***	0.060	0.059
	(0.015)	(0.016)	(0.074)	(0.061)	(0.096)	(0.103)
Observations	1,910	1,910	1,910	1,910	1,910	1,910
Year Time FE	YES	YES	YES	YES	YES	YES
Industry (Pre) FE	YES	YES	YES	YES	YES	YES
Industry (Post) FE	YES	YES	YES	YES	YES	YES

Notes Table VIII reports estimates from second-stage Heckman regressions with the probit of entrepreneurship as the first stage. I incorporate transformation of the predicted individual probabilities as an additional explanatory variable. Columns (1-2), (3-4), and (5-6) display results on patent count, citation per patent, and technological proximity, respectively. Pre and Post Industry and time FE are included for all specifications. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table IX - Panel A: Subsample Analysis - Information Technology Sector

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\text{Patents}_{\text{t,t+4}}$	$\text{Patents}_{\text{t,t+4}}$	$\text{Citations}_{\text{t,t+4}}$	$\text{Citations}_{\text{t,t+4}}$	Tech. $\text{Prox}_{\text{t,t+4}}$	Tech. $\text{Prox}_{\text{t,t+4}}$
HHI_citer	-1.470	-1.510	1.166**	1.148**	0.504**	0.495**
	(2.877)	(2.821)	(0.494)	(0.510)	(0.184)	(0.179)
Size_citer		0.492		0.041*		0.029*
		(0.910)		(0.029)		(0.015)
$\text{Patents}_{\text{t-5,t-1}}$	0.081**	0.083**	0.039	0.042	0.021	0.017
	(0.033)	(0.031)	(0.042)	(0.045)	(0.056)	(0.052)
$\text{Citations}_{\text{t-5,t-1}}$	0.003	0.003	0.279***	0.255***	0.051	0.048
	(0.011)	(0.011)	(0.085)	(0.091)	(0.116)	(0.114)
Observations	1,156	1,156	1,156	1,156	1,156	1,156
Year Time FE	YES	YES	YES	YES	YES	YES
Industry (Pre) FE	YES	YES	YES	YES	YES	YES
Industry (Post) FE	YES	YES	YES	YES	YES	YES

Notes Table IX-A shows matched regression results for a subsample of data. This sample is of entrepreneurs with start-ups in the Information Technology industry. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table IX - Panel B: Subsample Analysis - Medical/Biotech Sector

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\text{Patents}_{t,t+4}$	$\text{Patents}_{t,t+4}$	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$	Tech. $\text{Prox}_{t,t+4}$	Tech. $\text{Prox}_{t,t+4}$
HHI_citer	-2.128	-2.014	1.205***	1.189**	0.739***	0.727***
	(2.223)	(2.117)	(0.389)	(0.415)	(0.200)	(0.196)
Size_citer		0.678*		0.085*		0.051**
		(0.413)		(0.047)		(0.018)
$\text{Patents}_{t-5,t-1}$	0.617***	0.622***	-0.053*	-0.054*	-0.017	-0.019
	(0.031)	(0.033)	(0.029)	(0.028)	(0.067)	(0.071)
$\text{Citations}_{t-5,t-1}$	0.029**	0.024**	0.345***	0.330***	0.078	0.083
	(0.010)	(0.011)	(0.058)	(0.058)	(0.106)	(0.110)
Observations	594	594	594	594	594	594
Year Time FE	YES	YES	YES	YES	YES	YES
Industry (Pre) FE	YES	YES	YES	YES	YES	YES
Industry (Post) FE	YES	YES	YES	YES	YES	YES

Notes Table IX-B shows matched regression results for a subsample of data. This sample is of entrepreneurs with start-ups in the Medical/Biotech industry. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table IX - Panel C: Subsample Analysis - Non-High-Technology Sector

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\text{Patents}_{t,t+4}$	$\text{Patents}_{t,t+4}$	$\text{Citations}_{t,t+4}$	$\text{Citations}_{t,t+4}$	Tech. $\text{Prox}_{t,t+4}$	Tech. $\text{Prox}_{t,t+4}$
HHI_citer	0.610	0.554	0.799	0.734	0.101	0.125
	(2.789)	(2.582)	(0.720)	(0.753)	(0.209)	(0.202)
Size_citer		0.225		0.030		0.017
		(0.316)		(0.041)		(0.020)
$\text{Patents}_{t\text{-}5,t\text{-}1}$	0.117**	0.113**	0.019	0.022	0.014	0.014
	(0.052)	(0.055)	(0.079)	(0.082)	(0.059)	(0.060)
$\text{Citations}_{t\text{-}5,t\text{-}1}$	0.004	0.003	0.210**	0.207**	0.033	0.035
	(0.015)	(0.015)	(0.096)	(0.101)	(0.094)	(0.096)
Observations	160	160	160	160	160	160
Year Time FE	YES	YES	YES	YES	YES	YES
Industry (Pre) FE	YES	YES	YES	YES	YES	YES
Industry (Post) FE	YES	YES	YES	YES	YES	YES

Notes Table IX-C shows matched regression results for a subsample of data. This sample is of entrepreneurs with start-ups in the non-high-technology industry. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

References

[1] Acemoglu Daron, Ufuk Akcigit, and Murat Alp Celik, 2015,Young, Restless and Creative: Openness to Disruption and Creative Innovations, *Working Paper*

[2] Aghion, Philippe, Stefan Bechtold, Lea Cassar, Holger Herz, 2014, The Causal Effects of Competition on Innovation: Experimental Evidence, Working Paper.

[3] Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, 2005, Competition and Innovation: an Inverted-U Relationship, *Quarterly Journal of Economics* 120, 701-728.

[4] Aghion, Philippe, and Peter Howitt, 1992, A model of growth through creative destruction, *Econometrica* 60, 323-351.

[5] Ahuja, Gautam, and Riitta Katila, 2001, Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study, *Strategic Management Journal* 22, 197-220.

[6] Akcigit, Ufuk, and William R. Kerr, 2010, Growth through Heterogeneous Innovations, *Working Paper*

[7] Arrow, Kenneth, 1962, Economic welfare and the allocation of resources for invention. In R. Nelson (Ed.), *The rate and direction of inventive activity:* 609-626. Princeton, NJ: Princeton University Press.

[8] Atanassov, Julian, 2013, Do Hostile Takeovers Stifle Innovation? Evidence from Antitakeover Legislation and Corporate Patenting, *Journal of Finance* 68 (3), 1097?1131.

[9] Balsmeier, Benjamin, Lee Fleming, and Gustavo Manso, 2015, Independent Boards and Innovation, *Working Paper*

[10] Bayar, Onur, and Thomas J. Chemmanur, 2011 IPOs versus Acquisitions and the Valuation Premium Puzzle: A Theory of Exit Choice by Entrepreneurs and Venture Capitalists, *Journal of Financial and Quantitative Analysis* 46, 1755-1793.

[11] Baumol, William J., 1982, Contestable Markets: An Uprising in the Theory of Industry Structure, *American Economic Review* 71, 1-15.

[12] Bena, Jan, and Kai Li 2014, Corporate Innovations and Mergers and Acquisitions, *Journal of Finance* 69, 1923-1960.

[13] Betton, Sandra, B. Espen Eckbo, and Karin Thorburn, 2008, Corporate takeovers, in B.E. Eckbo, ed.: Handbook of Corporate Finance: Empirical Corporate Finance (Elsevier, North-Holland, Amsterdam the Netherlands).

[14] Bloom, Nicholas, Mark Schankerman, and John Van Reenen, 2013, Identifying technology spillovers and product market rivalry, Econometrica 81, 1347-1393.

[15] Cassiman, Bruno, and Reinhilde Veugelers, 2006, In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition, Management Science 52, 68-82.

[16] Chemmanur, Thomas, and Xuan Tian, 2013, Anti-takeover provisions, innovation, and firm value: A regression discontinuity analysis, Working paper, Indiana University.

[17] Cohen, Wesley M., and Steven Klepper, 1996a. "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D," *The Review of Economics and Statistics*, MIT Press, vol. 78(2), pages 232-43, May.

[18] Cohen, Wesley, Richard C. Levin, 1989, Empirical Studies of Innovation and Market Structure. In Handbook of Industrial Organization (II). Schmalensee R, Willig RD (eds). Elsevier: Amsterdam; 1059-1107.

[19] Cohen, Wesley M., Richard C. Levin, and David Mowery, 1987, ?Firm Size and R&D Intensity: A Re-Examination,? *Journal of Industrial Economics* 35, 543-63.

[20] Dasgupta, Partha, and Joseph E. Stigliz, 1980, Industrial Structure and the Nature of Innovative Activity, *The Economic Journal* 90, 266-293.

[21] Dixit, Avinash K., and Joseph E. Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity, *American Economic Review* 67, 297-309

[22] Ferreira, Daniel, Gustavo Manso, and Andre C. Silva, 2012, Incentives to Innovate and the Decision to Go Public or Private, *Review of Financial Studies*

[23] Gilbert, Richard J., and David M.G. Newbery, 1982, Preemptive Patenting and the Persistence

of Monopoly, *American Economic Review* 72, 514-26.

[24] Gonzalez-Uribe, Juanita, 2014, Venture Capital and the Diffusion of Knowledge, *Working Paper*

[25] Hall, Bronwyn H., Adam Jaffe and Manuel Trajtenberg. Market Value And Patent Citations, 2005, *Rand Journal of Economics* v36, 16-38.

[26] Harhoff, Dietmar, Francis Narin, M. Scherer F., and Katrin Vopel, 1999, Citation frequency and the value of patented innovation. *Review of Economics and Statistics* 81, 511-515.

[27] Henderson, Rebecca, and Iain Cockburn, 1996, Scale, scope, and spillovers: The determinants of research productivity in drug discovery, Rand Journal of Economics 27, 32-59.

[28] Higgins, Matthew J., and Daniel Rodriguez, 2006, The outsourcing of R&D through acquisitions in the pharmaceutical industry, Journal of Financial Economics 80, 351-383.

[29] Hirukawa, Masayuki, and Masako Ueda, 2011, Venture Capital and Innovation: Which is First?, *Pacific Economic Review* 16, 421-465.

[30] Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar, 2014, Can Unemployment Insurance Spur Entrepreneurial Activity? Evidence From France, Working Paper.

[31] Kerr, William R., Josh Lerner, and Antoinette Schoar, 2014, The Consequences of Entrepreneurial Finance: Evidence from Angel Financings, *Review of Financial Studies* 27, 20-55

[32] Kortum, Samuel, and Josh Lerner, 2000, Assessing the Contribution of Venture Capital to Innovation, *RAND Journal of Economics* 31, 674-692.

[33] Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming, 2011, Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010), http://hdl.handle.net/1902.1/15705, Harvard Dataverse, V5

[34] Lerner, Josh, and Amit Seru, 2014, The Use and Abuse of Patent Data, Working Paper, Harvard Business School.

[35] Manso, Gustavo, 2011, Motivating Innovation, *Journal of Finance* 66, 1823-1869

[36] Nanda, Ramana, and Matthew Rhodes-Kropf. Investment Cycles and start-up Innovation. Journal of Financial Economics 110, no. 2 (November 2013): 403-418.

[37] Phillips, Gordon M., and Alexei Zhdanov, 2013 R&D and the incentives from merger and acquisition activity, *Review of Financial Studies* 26, 34-78.

[38] Rhodes-Kropf, Matthew, and David T. Robinson, 2008, The market for mergers and the boundaries of the firm, *Journal of Finance* 63, 1170-1211.

[39] Ritter, Jay R., and Ivo Welch, 2012, A Review of IPO Activity, Pricing, and Allocations, *Journal of Finance*, 57, 1795-1828.

[40] Sampat, Bhaven N., and Arvids A. Ziedonis, 2005, Patent Citations and the Economic Value of Patents, Handbook of Quantitative Science and Technology Research, 277-298.

[41] Scotchmer, Suzanne, 1991. Standing on the shoulders of giants: cumulative research and the patent law. *Journal of Economic Perspectives* 5, 29-41.

[42] Schumpeter, Joseph, 1942, Capitalism, Socialism, and Democracy, *New York: Harper*

[43] Seru, Amit, 2014, Firm boundaries matter: Evidence from conglomerates and R&D activity, *Journal of Financial Economics* 111, 381-405.

[44] Trajtenberg, Manuel, 1990, A Penny for Your Quotes: Patent Citations and the Value of Innovations, *RAND Journal of Economics* 21, 172-187.

[45] Wooldridge, Jeffrey M., 2002, Econometric Analysis of Cross Section and Panel Data (MIT Press, Cambridge, MA).

[46] Zhao, Xinlei, 2009, Technological Innovation and Acquisitions, *Management Science* 55, 1170-1183.