Science, Startups, and the Problem of Value Capture: Thin Acquisition Markets, Weak Outside Options

Roger Masclans*

August 28, 2025
[Preliminary and incomplete—Latest version]

Startups commercializing science-based innovations are crucial for tackling pressing challenges, yet, in critical sectors such as energy, industrials, and materials, entrepreneurial activity remains limited. This paper investigates whether weak value capture at exit constrains these ventures. I estimate value creation and capture in startup acquisitions by combining acquisition prices with acquirer stock returns, adjusting for market noise to isolate the economic signal attributable to the acquisition. Science-based startups capture 46 cents per dollar of acquisition-induced surplus, compared to 61 cents for non-science startups—a 24% penalty. Conversely, they create 20% more joint surplus, consistent with continued entry despite the capture penalty. To explain these patterns, I examine a central mechanism: the structure of a startup's exit conditions. I argue that science-based startups face thinner, more concentrated acquisition markets and limited ability to scale independently, features that weaken the startup's bargaining power. Indeed, I find that science-based startups face up to 40% fewer potential acquirers, who are 53% larger on average, and that their value capture is more sensitive to acquirer concentration. Concentrated markets have a dual effect: large incumbents enable greater surplus creation, but also shift bargaining power away from startups, allowing acquirers to extract most of the gains from innovation. Finally, I find that the capture penalty diminishes when startups can scale commercialization independently. The results suggest that constrained exit environments limit returns to science-based entrepreneurship, highlighting the importance of competitive acquisition markets, markets for technologies, and alternative commercialization pathways in incentivizing upstream innovation.

*Duke University, The Fuqua School of Business. roger.masclans@duke.edu.
I am thankful to Ashish Arora, Wesley Cohen, Sharique Hasan, and Ramana Nanda for extensive advice and guidance. For helpful comments, I would also like to thank Sharon Belenzon, Victor Bennett, Ines Black, Marco Testoni, Jay Prakash Nagar, and Yoko Shibuya, as well as seminar participants at Wharton Innovation Doctoral Symposium, Consortium on Competitiveness and Cooperation, Strategy Science, DRUID25, and Academy of Management.

1 Introduction

Recent research has drawn attention to the limited entrepreneurial activity surrounding science-based technologies in critical sectors, despite their potential to address pressing social challenges.[1] For example, Lerner and Nanda (2020) document that between 2015 and 2019, startups in telecommunications, networking, computer hardware, semiconductors, materials, and energy industries, distinguished by their heavy reliance on scientific research, received only 5 percent of total Venture Capital (VC) investment, with the majority of funds flowing to software, consumer, and business products and services.

Most current explanations for the limited scale of science-based entrepreneurship focus on the determinants of value creation, suggesting that these ventures generate less economic value through commercialization than their counterparts. One central argument, for instance, is that weak or uncertain market demand constrains the commercial potential of science-based innovations (e.g., Van den Heuvel and Popp, 2023), thereby limiting the returns that ventures can ultimately deliver. Indeed, empirical evidence shows that innovation in these areas is often constrained by demand conditions and that shifts in demand can incentivize upstream innovation: Popp (2002) finds that rising energy prices lead to more energy-saving innovations; Aghion et al. (2016) show that higher fuel prices shift innovation by firms in the auto sector toward cleaner technologies; and Acemoglu and Linn (2004) document that increases in potential market size, driven by demographic shifts, affect pharmaceutical innovation.

A related view holds that science-based innovations often face high financing costs due to their need for significant upfront capital, high market and technical uncertainty, and long development timelines before reaching commercial proof-of-concept (Hall and Lerner, 2010; Kerr and Nanda, 2015). This has been rationalized as rigidity in experimentation (Kerr et al., 2014), which raises the cost of capital, lowers the option value of investment, and reduces the risk-adjusted expected returns from such ventures (Ewens et al., 2018; Nanda and Rhodes-Kropf, 2017).

This paper poses and empirically tests a complementary explanation for this limited activity: a problem of value capture. Science-based ventures can be both economically and socially valuable due to their novelty, technological advancements, and potential market impact. Yet, they may systematically struggle to capture the value they create at the point of exit. Under the current sequential innovation ecosystem, these startups typically conduct early-stage R&D and subsequently transfer the technology to incumbents, who hold the complementary assets required for large-scale commercialization (Arora et al., 2020; Kolev et al., 2022; Teece, 1986). The private returns to entrepreneurs and investors, which motivate initial entry and investment, are realized at this transfer stage and depend not only on the total value created—i.e., the joint surplus generated upon

[1] Entrepreneurship is an important driver for economic growth (Kortum and Lerner, 2000; Samila and Sorenson, 2011). For example, some estimates suggest that formerly VC-backed startups, now publicly listed firms, accounted for 41% of total U.S. market capitalization and 62% of corporate R&D in 2020 (Gornall and Strebulaev, 2021). In addition to driving economic growth, science-based innovations often enable transformative improvements and, as a result, are crucial for addressing unresolved, pressing social challenges (Akcigit and Kerr, 2018; Fleming and Sorenson, 2004; Dalla Fontana and Nanda, 2023).

commercialization—but, crucially, on the share of that surplus retained by the startup (Arora et al., 2024b; Green and Scotchmer, 1995; Scotchmer, 1996). When those returns are suppressed, not because of weak market demand or financing costs, but because incumbents capture most of the value, investment is distorted, shifting resources toward technologies with stronger prospects for value capture and leaving economically and socially valuable innovations undeveloped (e.g., Gans and Stern, 2000; Grossman and Hart, 1986; Holmstrom and Roberts, 1998; Lerner and Merges, 1998; Scotchmer, 2004).

In the energy sector, for example, novel photovoltaic chemistries developed by startups may generate major efficiency gains with significant economic impact. However, the large-scale production capabilities, distribution networks, and complementary technologies needed to bring these innovations to market are typically held by large industrial manufacturers or utility firms—not by the startups themselves (Kapoor and Furr, 2015). As a result, in an eventual acquisition of the technology, incumbents are well positioned to extract most of the surplus originated from these technical gains.

To date, little attention has been paid to the role that value capture plays in shaping the returns to entrepreneurship and, in turn, the development and commercialization of certain innovations. Empirical studies often overlook the distinction between value created and value captured, either using, for example, acquisition prices as a proxy for value creation or abstracting from rent sharing dynamics altogether. Aside from a few, relevant theoretical contributions (e.g., Arora et al., 2024b; Gans and Stern, 2000; Green and Scotchmer, 1995; Phillips and Zhdanov, 2013), whether value capture is systematically suppressed for certain types of innovations remains an open question—we lack empirical evidence, especially across industries, and a clear understanding of the underlying mechanisms.

In this paper, I study this question in three steps. First, I focus on a setting that both allows for clean measurement and reflects the dominant exit route for science-based ventures: the transfer of innovation through startup acquisitions (Ederer and Pellegrino, 2023).[2] Within this setting, I develop a novel measure at the startup level that estimates the joint surplus generated in an acquisition (value creation) and how that surplus is divided between the startup and the acquirer (value capture). Second, I use this measure to analyze how value creation and capture differ between science-based and non-science counterparts. As introduced, this focus is motivated by the fact that science-based ventures, some of which are referred to as Deep Tech, sit at the center of current debates around innovation and industrial policy as well as long-term economic competitiveness given their potential to tackle relevant societal challenges. Finally, I provide a theoretical explanation and supporting empirical evidence for one central mechanism: thin acquisition markets and limited ability to commercialize independently, i.e., weak outside options, constrain the bargaining power of science-based startups at the point of exit. As a result, these ventures disproportionately face

[2] Other forms of commercialization, such as licensing agreements or collaborative arrangements with incumbents, are also commonly used by startups developing advanced technologies (see, e.g., Gans et al., 2002; Hsu, 2006) and may involve different value creation and capture dynamics. In this paper, however, I abstract from these alternative pathways and focus exclusively on acquisitions.

unfavorable exit conditions, allowing incumbents to extract most of the innovation's value.

To develop the value creation and capture measures, I use data on all startup acquisitions (M&A) conducted by U.S. publicly listed firms between 1990 and 2022. The primary data source is PitchBook; after applying standard filters and cleaning procedures, the final sample includes 5,823 startup acquisitions. I use the acquisition price and accumulated private investment to estimate the surplus captured by the startup, and the acquirer's abnormal stock market return around the deal announcement date to estimate the surplus captured by the incumbent—that is, the incremental gains expected from integrating the target's technology, conditional on the incumbent's complementary assets. The sum of these two components estimates the total joint surplus generated by the transaction, while their relative magnitudes identify each party's share.[3]

One limitation of stock market prices is that they are noisy, conflating the acquirer's gains from the acquisition with unrelated firm-level, industry, or macroeconomic events. To address this issue, I adopt a parametric approach that refines surplus estimates by isolating the signal from noise. I do so by adapting to the M&A context the methodology developed by Kogan et al. (2017), modifying some of their core assumptions to account for the possibility that acquisitions can lead to negative abnormal returns if investors perceive overpayment—driven, for example, by high integration costs or organizational frictions (Benson and Ziedonis, 2010; Chondrakis et al., 2021; Higgins and Rodriguez, 2006). To estimate these parameters, I complement the main data with close to 108,000 additional acquisitions conducted by publicly listed U.S. firms, sourced from Refinitiv SDC Platinum.

Next, I classify startups by their reliance on novel scientific discoveries. I define science-based innovations as technologies that require substantial R&D and build on advances originating in fields such as life sciences, chemistry, physics, or engineering (Fleming, 2001; Fleming and Sorenson, 2004; Hall and Lerner, 2010). A science-based startup, accordingly, is one whose products or services are rooted in such scientific knowledge, whether developed internally or sourced externally from research institutions. To identify these firms, I use a large language model (LLM). The model, Llama 3.3, processes unstructured text from diverse sources to classify the extent to which each startup's technologies draw on novel scientific research. This classification approach circumvents the limitations of traditional science and innovation measures.[4] Beyond manual validation, I validate the LLM-based classification using patent-to-paper citation data, finding strong alignment where data are available and supporting the accuracy of the model-based labels. I also conduct robustness checks using alternative prompts, as suggested by Carlson and Burbano (2024), and classification thresholds.

Results show that startups commercializing scientific innovations consistently capture less value

[3]The abnormal stock return reflects the market's expectation of future cash flows from the integration of the startup's technology with the incumbent's assets. Importantly, it incorporates beliefs about both technical feasibility and market adoption at the time of the transaction, i.e., the risks.

[4]For example, patent-to-paper citations are widely used as a proxy for science-based innovation, but are are especially incomplete and prone to measurement error in the startup context. Many startups either do not patent, rely on trade secrecy, or hold patents assigned to external entities such as universities or investors, limiting the precision of firm-level attribution (Graham et al., 2009; Bryan and Williams, 2021; Lerner and Seru, 2022).

than non-science-based startups. On average, science-based ventures capture only 46 cents of every dollar of surplus generated upon acquisition, compared to 61 cents for their non-science-based counterparts—a 24% penalty. Furthermore, there is important cross-industry heterogeneity. Some industries consist primarily of startups commercializing scientific innovations (e.g., biotechnology), while others are dominated by non-science startups (e.g., software; consumer goods and services). The highest capture penalty is experienced by startups in industries with a mix of both types, such as energy (-36%), industrials, manufacturing, and materials (-36%), and hardware (-24%)—industries precisely where concerns about a dearth of activity are particularly pronounced. Interestingly, the lowest capture penalty is in the life sciences.[5]

It is worth noting that, from the startup's perspective, the greater value capture penalty faced by science-based ventures implies a higher entry threshold for commercializing scientific innovations. This is because, as pointed out, entry decisions are based on expected private returns, which depend on both the joint surplus and the share they are able to retain. That is, in equilibrium, ventures with lower extraction rates should generate proportionally greater surplus in order to yield comparable private returns across types. Indeed, I find that science-based startups in my sample generate significantly more total value upon acquisition—an estimated 20% increase relative to non-science counterparts. This highlights the economic importance of science-based ventures, but also the selection at play: only those generating higher value can overcome limited capture and attract resources.[6]

Why do science-based startups capture less value from their innovations, despite generating substantially more? I argue that a central mechanism lies in the exit environment, which determines how ventures realize returns from their technologies. Startups typically face two exit routes: (i) independent commercialization, scaling through the product market and often culminating in an IPO, or (ii) acquisition by an incumbent through M&A. Under M&A, outcomes are in turn governed by the market structure of the potential acquirers: the number, size, and capabilities of incumbents able to acquire and commercialize the focal technology. Specifically, I argue that these exit environment conditions differ systematically between science- and non-science-based ventures, and that these differences help explain the observed patterns in value creation and capture I report.

With regard to the exit route of independent commercialization, the literature suggests that

[5] The life sciences appear to be a notable exception to the broader pattern of limited value capture in science-based sectors. One possible explanation is that this domain benefits from institutional structures that mitigate core frictions associated with uncertainty and commercialization. First, demand uncertainty is lower: the availability of epidemiological and clinical data provides relatively clear signals about market need ex-ante. Second, technical risk is partially externalized through the structured and well-subsidized process of clinical trials, which offer standardized milestones (e.g., Phases 1–3) that reduce information asymmetries between developers and investors. These milestones are often funded or de-risked by public institutions, effectively subsidizing not only early-stage science but also venture capital and incumbent pharmaceutical firms. Additionally, the strength of patent protection in pharmaceuticals supports a well-functioning market for technology, enabling licensing and acquisition at earlier stages and increasing the likelihood that upstream innovators can realize returns (Arora et al., 2022). Taken together, these features suggest that life sciences do not conform to the mechanisms that generate value capture challenges in other science-based sectors—an exception that may help illustrate the rule.

[6] As such, it is important to note that these results may not necessarily generalize to the broader population, particularly ventures acquired privately or those that failed before reaching exit.

science-based ventures face greater challenges in scaling commercialization independently. The specialized complementary capabilities they require—such as advanced manufacturing, distribution networks, regulatory expertise, or integration with related technologies—are costly to develop internally and rarely available through market-based contracting (Ewens et al., 2018; Gans et al., 2002; Kapoor and Furr, 2015; Marx et al., 2014; Moeen, 2017).[7] At the same time, science-based ventures also face systematically different acquisition market structures compared to their non-science counterparts. The very same complementary assets needed to scale their technologies are typically held by a small number of large incumbents (Adner and Kapoor, 2010; Aggarwal and Hsu, 2009; Teece, 1986), which have become increasingly concentrated over time (Klepper, 1996; Sutton, 1991, 2007; Ederer and Pellegrino, 2023). As a result, the pool of viable acquirers is often thin, leaving startups dependent on a limited set of large firms that control access to commercialization.[8]

I formalize this mechanism through a simple conceptual framework of startup acquisitions, drawing on auction theory and treating these structural differences in exit conditions as primitives and thus starting points for the model.[9] Because these primitives are well documented in the literature—both theoretically and empirically—and are also supported by my data, I refer to them henceforth as stylized facts for ease of exposition.[10]

The framework yields three sets of simple predictions regarding the main variables of interest, value creation and capture. First, in concentrated markets where only a few large incumbents are viable acquirers, startup acquisitions tend to generate substantial joint surplus—since these incumbents access broader markets and are better positioned to realize the innovation's potential—but startups capture little of that value due to limited acquirer competition and weak bargaining power. Second, in more fragmented markets composed of smaller or less capable acquirers, startups may retain a greater share of the value through better bargaining, but total value is lower, as the acquirer can only reach a limited portion of the market. Third, the ability of the startup to scale independently acts as a critical outside option, conditioning this tradeoff. When credible, it improves bargaining power and enables greater value capture, regardless of downstream market

[7] Contract manufacturing and other market-based (or publicly subsidized) complements lower the effective fixed cost of commercialization when tasks are standardized across firms. If a specialized supplier can amortize large fixed costs across many clients, the per-firm cost of accessing the capability falls from F to approximately F/N plus a variable charge. This shifts the make-or-buy boundary: a startup that would otherwise face prohibitive in-house investment can credibly scale via the market at much lower upfront cost. In bargaining terms, the outside option improves (the firm can scale independently at lower expected cost), which raises the seller's threat point and increases capture in acquisition negotiations. By contrast, when tasks are highly idiosyncratic (high asset specificity), fixed costs cannot be spread, outside options remain weak, and startups rationally accept lower prices rather than fund the development of such capabilities.

[8] Additionally, the majority of science-based innovations, with the exception of biotechnology and pharmaceuticals, face weak markets for technology that would otherwise provide a competitive market for early exits (Arora et al., 2022).

[9] In this paper, I treat the startup's ability to scale independently and the structure of the acquisition market as exogenous. In practice, these factors may be endogenously determined by the nature of the innovation itself. For example, prior research highlights how R&D intensity influence the structure of downstream markets (see, e.g., Cohen, 2010; Sutton, 1991).

[10] Although similar regularities have been documented in prior work, their broader generalizability remains uncertain and warrants further validation across different empirical settings.

structure.[11] Importantly, note that the framework is agnostic to the nature of the commercialized technology in and of itself; rather, it is the systematic differences in exit conditions between science-based and non-science startups that drive the observed disparities in value creation and capture.

To characterize acquirer markets and test these predictions, I employ existing data on downstream market competition developed by Hoberg and Phillips (2016, 2025), and widely used to assess competitive overlap, product-market rivalry, and acquirer similarity in studies of M&A and vertical integration (e.g., Frésard et al., 2020). The dataset is based on a text-based analysis of the product descriptions in firms' 10-K filings, which are used to construct pairwise similarity scores across all publicly listed firms. Firms are embedded in a high-dimensional product space, where proximity captures the degree of overlap in product offerings. This allows me to identify, at the firm-year level, the set of product-market peers for each acquirer, approximating the pool of potential acquirers for a given startup at exit. I use this to measure both the thickness of the acquisition market—the number of credible acquirers—and its concentration, based on their relative size distribution.

Using these data, I start by documenting the set of stylized facts characterizing the exit environment of science-based startups. First, these ventures face systematically thinner acquisition markets than their non-science counterparts. On average, the set of potential acquirers is 9% smaller, and up to 40% smaller in sectors such as energy and manufacturing. Second, these markets are more concentrated. The typical acquirer of a science-based startup is 53% larger in terms of market capitalization at the time of acquisition, reflecting the dominance of a few large incumbents with the necessary complementary assets. Third, science-based ventures exhibit weaker independent scaling. At exit, the median non-science startup generates 71% more revenue than the median science-based one, underscoring the significantly greater constraints the latter face in scaling independently.

Most importantly, the empirical results align with the framework's predictions, providing support for the proposed mechanism. I find that the size of the potential acquirer pool strongly predicts the share of value captured by science-based startups. A one standard deviation increase in the number of potential acquirers is associated with a 26% increase in value capture. In contrast, there is no such effect for non–science-based startups, whose capture is insensitive to acquirer pool size. This asymmetry is consistent with the model. Non-science startups typically have credible outside options, making their capture less reliant on the structure of the acquisition market. For science-based ventures, by contrast, weak outside options make the size and structure of the acquirer pool

[11] One could argue that ex-ante contracting could, in principle, mitigate rent-sharing issues, some times referred to as hold-up (Grossman and Hart, 1986), by specifying the terms of transfer and division of surplus before investments are made and uncertainty is resolved. However, such contracts are rarely feasible in the context of science-based innovation. One central reason is, for example, the high degree of ex-ante uncertainty surrounding the final market application and the identity of the most efficient commercializing acquirer (Bresnahan and Gambardella, 1998; Lerner and Merges, 1998; Kapoor and Klueter, 2021; Gambardella et al., 2021). For example, a novel battery chemistry may be applied for electric vehicles, grid storage, or emergency power supply in healthcare. This makes it difficult to determine value at the time of agreement, define the relevant contingencies, and assign control rights, limiting the scope for credible and enforceable contracts (Grossman and Hart, 1986; Holmstrom, 1989; Holmstrom and Roberts, 1998). Another important reason is the weakness of intellectual property rights, which limits the startup's ability to protect its knowledge in the absence of formal control or ownership, thereby reducing too the effectiveness of ex-ante contracting (Arora and Merges, 2004; Gans and Stern, 2000; Hsu and Ziedonis, 2013).

more consequential. Consistent with this interpretation, I find that revenue at exit accounts for most of the observed variation. For example, when conditioning on startups with above-median revenue, not only do value capture levels increase, but the difference between science-based and non–science-based startups becomes statistically insignificant.

In terms of value creation, joint surplus indeed systematically correlates with the market structure of potential acquirers. For non–science-based startups, each decile increase in the number of potential acquirers is associated with a 6.9% increase in value creation, consistent with the idea that, in less-specialized markets, marginal acquirers tend to be larger and possess stronger commercialization capabilities (Bresnahan and Gambardella, 1998; David and Nagaraja, 2004). In contrast, for science-based startups, I find that value creation declines as the number of potential acquirers increases. This may seem surprising at first, but it reflects the structure of many specialized markets. In these sectors, a small number of large incumbents typically dominate and possess the capabilities needed to commercialize advanced technologies (Klepper, 1996, 2002; Cohen, 2010). When these top firms are not interested in acquiring, the remaining potential buyers are often smaller companies in fragmented submarkets (Klepper and Thompson, 2006; Sutton, 1991, 2007). These smaller firms face more limited demand and have fewer resources, reducing both their ability to scale the innovation and their willingness to pay for it. As a result, as the number of potential acquirers increases, their average size and commercialization capacity decline, lowering the value ultimately realized by the startup.[12]

The findings have significant managerial and policy implications, and are further underscored by recent evidence showing that large incumbents are increasingly withdrawing from core scientific research activities in favor of deploying their downstream capabilities to commercialize innovations (Arora et al., 2018; Fleming et al., 2019) and that market concentration in certain industries is growing (e.g., Antón et al., 2024; Cunningham et al., 2021; Nanda et al., 2015). On the policy side, distinguishing between value creation and value capture is essential for designing targeted interventions, even if both problems can exist simultaneously. If the problem lies in value creation, limited investment in science-based startups may be explained by low expected returns stemming from high financing costs or weak demand. Under this view, underinvestment is intrinsic to the innovation itself, rather than a consequence of structural dynamics of the ecosystem.[13] To a certain extent, and put simply, one could argue that even within an incumbent firm, such innovations would face similar hurdles—low expected returns stemming from inherent features like uncertainty, capital intensity, or long timelines. In this case, the issue is not who develops the innovation,

[12]For example, consider a startup developing a novel robotics component with potential applications in precision manufacturing. The most relevant acquirers may be large multinationals such as Siemens or ABB. If those firms deem the commercial opportunity too narrow or misaligned with their strategic focus, they may opt not to acquire. The remaining interest may come from smaller, more specialized firms that operate in fragmented submarkets. These firms face limited end-user demand and lack scale, thus generating lower surplus from the innovation and offering lower valuations. These dynamics align with the broader literature on submarkets and innovation incentives, including Klepper and Thompson (2006) and Sutton (1991, 2007).

[13]I abstract here from other structural frictions that may well play an important role too, such as those arising from capital markets. For example, limited collateral for intangible assets and regulatory constraints on institutional investors.

but the nature of the innovation itself, which dampens incentives across the board. Along these lines, Narain (2025) finds that venture funds and publicly listed firms that are research intensive develop technologies with similarly short gestation periods, while government funding is more likely to support the development of technologies with significantly longer timelines.

By contrast, if the problem lies in value capture, underinvestment stems not from the innovation itself, but from firm boundaries and the structure of the innovation ecosystem, one that requires coordination and transfer across multiple actors. Since entry decisions hinge on private returns, distortions in rent sharing can misallocate innovative effort, causing economically and socially valuable innovations to go undeveloped because the *first* innovator cannot earn an adequate return (Arora et al., 2024a; Scotchmer, 1996, 2004). Therefore, policy should not only focus on boosting value creation through instruments such as demand stimulation—e.g., carbon taxes, emissions standards (e.g., Gerarden, 2023) and R&D subsidies (e.g., Howell, 2017)—, but also address how this value is distributed between the parts involved in commercialization. In particular, fostering more competitive markets for technology and acquisition markets, as well as enabling alternative commercialization pathways that reduce reliance on dominant incumbents can help address these frictions. Such pathways include, for example, shared manufacturing platforms, prototyping infrastructure (e.g., pilot plants), and access to specialized infrastructure in National Labs. For example, in battery technology, a startup with modular production capabilities or access to shared facilities such as those of the National Renewable Energy Lab (NREL) may not only face a clearer route to market but also credibly threaten independent commercialization.

On the managerial side, the findings highlight the need for managers in both incumbent firms and startups to explore practices that offset the adverse effects of a lack of outside options and downstream market concentration. For example, startup managers and investors should consider manufacturing and distribution approaches that signal the startup's ability to scale independently. This may include securing funding for production capacity or geographically locating infrastructure near target customers to facilitate distribution and strengthen credibility.[14] Moreover, if incumbents systematically suppress value capture for external innovators, they risk discouraging future innovation, reducing the pool of technologies they depend on for long-term competitiveness and growth. In sectors such as biotech, semiconductors, and advanced materials, where reliance on external innovation is high (Arora et al., 2020), a constrained supply may force greater dependence on internal R&D (Ceccagnoli et al., 2010). Thus, these dynamics underscore the importance of corporate involvement in shaping upstream external innovation, whether through corporate venture capital (Ceccagnoli et al., 2018; Ma, 2020), collaboration with public industrial infrastructure, or other mechanisms, effectively balancing the short- long-run trade-off.

The paper contributes to our understanding of the structural barriers that limit the progress of science-based innovations through startups. Prior work has highlighted factors such as market demand (Dalla Fontana and Nanda, 2023; Van den Heuvel and Popp, 2023), capital intensity (Hall

[14] These considerations raise the question of whether, while necessary for startups to strengthen their bargaining position, some of these investments are nonetheless socially inefficient (Arora et al., 2024b), especially if they duplicate capabilities already held by incumbent firms.

and Lerner, 2010), risk and experimentation (Ewens et al., 2018; Kerr et al., 2014; Nanda and Rhodes-Kropf, 2017; Howell, 2017), and long time horizons (Narain, 2025) as key obstacles. This paper extends that perspective by focusing on how surplus division in a sequential innovation system suppresses startup returns (Arora et al., 2024a; Gans and Stern, 2000; Scotchmer, 1996) while emphasizing the role of complementary capabilities that are often unavailable to startups but critical for commercialization (Helfat and Lieberman, 2002; Kapoor and Furr, 2015; Teece, 1986).

The paper also contributes to the literature on startup modes of commercialization (e.g., Ceccagnoli et al., 2014; Gans et al., 2002; Marx et al., 2014), highlighting the relevance of independent commercialization pathways not only in and of itself, but as a credible threat that may lead to better terms in acquisitions and, thus, to a more competitive ecosystem. Moreover, the paper sheds light on research on M&A and corporate strategy examining how deal-, firm-, and industry-level factors shape acquisition outcomes (e.g., Barney, 1988; Capron and Shen, 2007; Feldman et al., 2019; Kaul and Wu, 2016; Testoni, 2024; Villalonga and McGahan, 2005). It advances this literature by emphasizing how structural features of the acquisition environment—specifically, buyer concentration and capability asymmetries—influence both value creation and the division of surplus between acquirers and targets.

Finally, the paper contributes to the literature on innovation, competition, and downstream market structure by providing novel empirical evidence on the structure of acquisition markets, consistent with prior theoretical and empirical work (e.g., Cohen, 2010; Klepper, 1996; Sutton, 2007). In doing so, it highlights how industry concentration—also at the center of antitrust debates—not only raises concerns about consumer welfare, but can undermine innovation performance and long-run technological progress (Federico et al., 2020; Segal and Whinston, 2007; Shapiro, 2025).

Methodologically, this paper first introduces an approach to distinguish between value creation and value capture, enabling the analysis of rent-sharing dynamics in startup acquisitions and allowing us to study questions that have received limited attention, as well as revisit established ones through a new lens. Second, the paper introduces a novel approach using Large Language Models and textual data to classify startups based on their reliance on scientific research, improving upon traditional patent-based measures. Third, this paper develops a parametric approach to isolate the market signal in acquisitions by publicly listed firms, adapting the methodology of Kogan et al. (2017). While developed to study surplus creation and sharing in startup acquisitions, the method can be broadly applied in studies of acquisitions using stock market reactions.

2 Conceptual Framework

The current innovation ecosystem is characterized by a sequential division of labor, in which different actors often specialize in distinct stages of technological development and commercialization (Arora et al., 2018; Fleming et al., 2019). Startups typically develop early-stage technologies, which are subsequently scaled and commercialized by incumbent firms possessing the complementary assets necessary to fully realize the economic potential of these innovations (Teece, 1986). In this structure, most startups do not capture the returns from innovation through direct product market competition, but rather by transferring their technologies to incumbents (Gans et al., 2002; Marx et al., 2014), often via acquisitions (Andrews et al., 2022; Ederer and Pellegrino, 2023). As a result, the returns realized at the point of acquisition play a central role in shaping entrepreneurial incentives (Scotchmer, 1996; Gans and Stern, 2000) and, thus, understanding how these are formed in the first place is crucial to explaining why some industries or technologies attract more startup activity than others (Arora et al., 2024a).

This section develops a theoretical framework that decomposes startup returns at acquisition into two components: the total, joint surplus generated through commercialization (value created) and the share of that surplus captured by the startup. I argue that these outcomes are, in part, jointly shaped by the structure of the commercialization environment. Specifically, the number, size, and capabilities of potential acquirers, as well as the viability of independent scaling by the startup. I further pose that these structural features correlate with the underlying nature of the technology, which motivates the main hypothesis of the paper: Science-based startups systematically differ in both value creation and value capture due to thinner acquisition markets and weaker outside options. The framework generates testable predictions about how value creation and capture vary across different types of startups, which I later examine empirically using acquisition data.

2.1 Decomposing Startup Returns into Value Creation and Capture

To date, most research measures startup returns as the proceedings received at exit (e.g., acquisition price) relative to accumulated investment and implicitly equates these returns with the economic value the innovation creates. Consequently, most existing literature focuses on the determinants of value creation to explain patterns of entry, investment, and growth, arguing that limitations on value creation, such as weak demand, high risks, or high financing costs, reduce the expected economic value of certain innovations and, thus, suppress returns.

Yet, theoretical accounts suggest that under a sequential innovation system, where innovations must be transferred from an upstream to a downstream firm, returns may also be shaped by a distinct factor: the ability of the initial innovator to appropriate the value their innovation will create (Scotchmer, 1996, 2004; Arora et al., 2024b). In such a system, the innovator's payoff depends not only on the total value generated by the innovation when commercialized at scale, but also on how that value is divided between the upstream and downstream parties. As such, even when an innovation has the potential to create substantial value, the startup may capture only a small share, with the majority appropriated by the incumbent that commercializes and diffuses the technology to market.

This perspective implies that low observed returns may not only reflect limited value creation, but also limited value capture—an important distinction with implications for how innovation incentives are understood. If returns are suppressed not because innovations lack value, but because startups are unable to extract it, then entrepreneurial effort may systematically be misallocated away from high-value but low-capture domains. Despite the significance of this distinction, empirical evidence remains limited on the extent to which different types of startups are able to capture value, and on the mechanisms that determine that share.

In this paper, I take a first step toward addressing that gap by examining whether certain innovations, particularly those grounded in scientific research, are systematically disadvantaged in the share of value they extract at exit, and by exploring the mechanisms that account for this variation. It is worth clarifying that the empirical focus of the paper is not on testing underinvestment itself, but on measuring in the first place the joint surplus created at exit and its division between the startup and the incumbent. Establishing whether certain types of innovations, particularly science-based ones, face systematically lower capture is a necessary precursor to evaluating whether misallocation or underinvestment follows.

To formally examine these issues, I decompose the returns a startup gets from an acquisition into two components. The first component is the total value that the innovation will create when deployed by an incumbent. Although commercialization can generate several forms of value, such as consumer surplus, knowledge spillovers, and broader social benefits, the relevant construct for the purpose of this paper is the *joint surplus* (V_t), that is, *the private economic surplus accruing to the parties directly involved in commercialization*. This is because this joint surplus, net of costs, drives their incentives to enter, invest, and transact.[15] Henceforth, I refer to this joint surplus also as value created.

The second component is the share of this joint surplus that the startup captures, which I refer to as *value capture* (λ_s). This share determines how much of the total joint surplus the startup is able to extract, ultimately determining its private surplus—or, informally, its returns. Put simply, the startup surplus (V_s) is a function of both the size of the pie (V_t), created through the combination of the startup's innovation and the incumbent's complementary assets, and the fraction of that pie the startup can extract (λ_s):

$$V_s = \lambda_s V_t \qquad \text{(Startup surplus)} \tag{1}$$

Because startups move first, selecting which technologies to develop, I pose that the expected levels of both value creation (V_t) and value capture (λ_s) influence how talent and capital are allocated across sectors and technologies. High expected creation combined with a large capture share attracts entry and investment, while low creation or weak capture discourages effort. Crucially, even when the expected value created is high, a low capture share for the startup can deter entry. In such cases, limited appropriation by the initial innovator may prevent valuable innovations from being pursued, leaving technologies that are both socially desirable and economically promising under-exploited.

[15] While research suggests that entrepreneurs and scientists may be partially motivated by non-pecuniary considerations such as social impact or recognition (Cohen et al., 2020; Sauermann and Cohen, 2010; Lazear, 2005), in this paper, I adopt the view that entry and investment decisions are governed by expected economic rents. Furthermore, this may be especially reasonable in a context where institutional venture capital plays a prominent role, since these investors are subject to capital constraints and return expectations shaped by external stakeholders—most notably, their limited partners.

Figure 1 illustrates how startup surplus can reflect very different underlying economics, depending on both the total joint surplus created and the share captured by the startup. Each bar decomposes the value created in a startup acquisition into the portion captured by the startup and that captured by the acquirer. The figure compares two cases. In the Software as a Service (SaaS) example, the startup captures half of the joint surplus, while in the carbon technology example, it captures a much smaller share. Looking only at startup surplus, one might infer that the innovation is less valuable and expect capital and talent to flow toward sectors like SaaS, where startup returns are higher. Likewise, a policymaker aiming to promote carbon-efficient technologies might also view the low startup surplus as evidence of limited economic value, concluding that a carbon tax could stimulate demand and, all else equal, raise joint surplus.

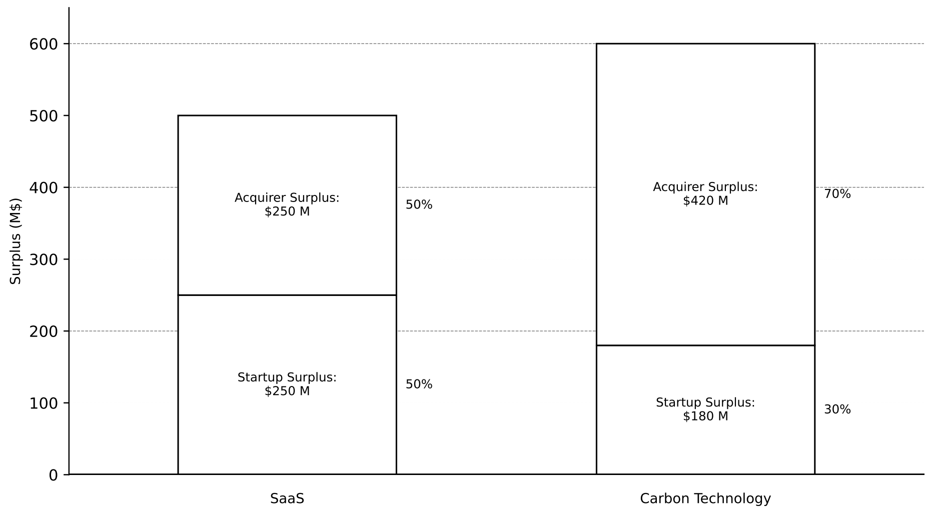


Figure 1: Illustrative examples of joint surplus creation and division at acquisition. The figure illustrates how similar observed returns can mask important differences in total value created and in the share captured by the startup, with implications for how upstream resources are allocated across technologies.

However, looking at the full distribution of joint surplus reveals a different picture. In this example, the carbon technology generates more total value than the SaaS one, but a disproportionately large share accrues to the acquirer. This distinction matters for both policy and management. From a policy perspective, when most of the rents from an innovation flow to incumbents at acquisition, stimulating downstream demand—such as through carbon taxes—may do little to encourage upstream innovation in sequential innovation systems. From a managerial perspective, entrepreneurs, investors, and managers operating in such environments must recognize that high total value creation does not necessarily translate into high private returns, and may need to adapt commercialization strategies to improve bargaining outcomes. In both cases, the challenge is not only the amount of value created, but also the fraction that the initial innovator can extract.

2.2 Determinants of Value Creation and Capture

2.2.1 Value Creation

This decomposition also offers a useful lens for understanding why startups may systematically differ in both value creation and value capture at exit, and what mechanisms may drive these differences. On the creation side, several factors can drive higher joint surplus. One key determinant is the technical performance of the innovation at the time of acquisition which shapes the per unit value delivered, either through increased willingness to pay by the end user or through reduced production costs. Importantly, this technical performance implicitly reflects any technical risks that remain unresolved, as performance is assessed relative to the degree of certainty with which it can be realized in practice. As a startup develops its innovation and gradually resolves key technical uncertainties, it reduces perceived risk, increasing the expected value of the innovation.[16]

A second important determinant of value creation is the size and structure of the addressable market the acquirer can serve. Put simply, the total surplus generated depends on the per unit value of the innovation multiplied by the scale over which that value can be realized. This scale is shaped, in turn, by the acquirer's market share, customer base, and overall competitive position.[17] Likewise, note that the market structure of the downstream firms plays a critical role: even if an innovation has broad applicability, the realized surplus will be limited if the acquirer operates in a fragmented market and controls only a small share of potential end users. Thus, the surplus induced by the transaction reflects not the full potential of the innovation, but only the portion accessible to the specific acquirer.[18]

A third determinant is the capability of the acquirer, particularly the availability and quality of complementary assets needed to realize the innovation's value. These include not only production-related resources like manufacturing infrastructure, complementary technologies and patent portfolios, and human capital, but also market-related capabilities such as regulatory expertise, distribution networks, branding, and customer relationships. If, for simplicity, one assumes that the acquiring firm is the best available match—operating at or near the technological and organizational frontier in the relevant complementary assets and capabilities—then, all else equal, the joint surplus reflects the maximum value that can be realized from the innovation in its current state. A well-matched acquirer can bring the technology to market more quickly, at lower cost, and

[16]For example, a carbon capture startup may initially face uncertainty over whether a novel chemical process can reliably absorb emissions under variable industrial conditions. Once this technical challenge is resolved and performance demonstrated under realistic scenarios, outside the lab, the innovation may experience a discrete increase in its perceived value. In this way, development progress acts as a trigger for upward revisions in expected surplus, as the probability of commercial success increases and required future investment decreases.

[17]Throughout the analysis, I abstract from the effects of acquirer market power in the traditional sense of markups or output restrictions. While buyer market power may influence acquirer surplus, willingness to pay, and bargaining dynamics, my focus is on how the ability to commercialize and scale a given innovation varies with market structure. In this context, changes in realized value reflect differences in joint surplus rather than redistributive effects arising from pricing power.

[18]Note that, as with technical risks, market risks such as uncertainty about demand volume, regulatory approval, or end-use application, are embedded in the the value of an innovation at transfer. For example, a microbial treatment in agricultural biotech may show strong technical performance, but if it is unclear whether it will be adopted in large-scale commodity farming or niche markets, the expected value will reflect that uncertainty.

at greater scale than a less capable one.[19]

2.2.2 Value Capture

On the capture side, the structure of a startup's exit environment critically shapes how the joint surplus generated by the innovation is divided between the startup and the acquirer. In acquisition settings, this division is determined through a bargaining process, whose outcome depends on two fundamental factors.

The first of these factors is the intensity of competition among potential acquirers. When the pool of interested acquirers is broad and multiple incumbents actively bid for the asset, competitive pressure drives up the acquisition price, enabling the startup to appropriate a greater share of the surplus. By contrast, in thin markets with few bidders, the startup may face a dominant acquirer with monopsony power, reducing its ability to negotiate favorable terms. The second factor is the strength of the startup's outside option. Specifically, its ability to continue developing and commercializing the technology without being acquired. A credible outside option enhances the startup's bargaining position by allowing it to reject unattractive offers, irrespective of acquirer competition. However, when such independent commercialization is infeasible, the startup becomes reliant on acquisition, weakening its leverage and diminishing its share of the surplus.[20]

2.3 Science-based Ventures: Thin Acquisition Markets, Weak Outside Options

The discussion above highlights that value creation and capture are shaped by two key features of the commercialization environment: (1) the structure of the acquisition market and (2) the startup's ability to scale independently. I now turn to the central premise of the paper: science-based startups differ systematically from non-science-based startups along both features, and these differences give rise to systematic variation in value creation and capture.

Science-based ventures are more likely to operate in concentrated acquisition markets, typically characterized by a small number of large incumbents with the capabilities required to absorb and deploy advanced technologies. At the same time, the specialized and infrastructure-intensive nature of many scientific innovations often limits their ability to commercialize independently. These systematic differences in exit conditions, I argue, generate distinct patterns of value creation and

[19] Other determinants also matter. For example, the commercialization horizon, imitation and competition, and obsolescence risk influence the duration and magnitude of expected surplus. Innovations expected to generate rents over a longer period, due to slower technological cycles or stronger intellectual property protections, create more value than those likely to be displaced in the near term.

[20] While this paper focuses on structural determinants of value capture—namely, downstream market composition and startup outside options—other factors may also influence the division of surplus in practice. These include, for example, the negotiation skill and experience of the founding team and investors, information asymmetries, and timing considerations such as urgency on either side of the deal. The bargaining process is also shaped by the incumbent's outside options, including the ability to develop a competing solution internally, acquire an alternative target, or delay action until more information becomes available (Gans and Stern, 2010). These elements introduce idiosyncratic variation in acquisition outcomes and may amplify or mitigate the effects of the underlying market structure and outside options.

capture between science-based and non-science-based ventures. In particular, they influence both the total joint surplus realized at exit and the share of that surplus retained by the startup.

In this section, I formalize these claims into a set of stylized facts—drawn from the data and consistent with prior literature—alongside theoretical propositions regarding patterns of value creation and capture. To do so, I develop a simple theoretical model of startup acquisition using a second-price sealed-bid auction, a parsimonious framework for settings with offers, negotiations, and competing bids. In the model, the incumbent with the highest valuation acquires the startup and pays the second-highest bid, subject to the startup's reservation value—its payoff from scaling independently. Market structure and outside options enter as primitives, as reflected in the stylized facts below (*1*, *2.1*, *2.2*, and *3*). Furthermore, I model startup valuations as heterogeneous and acquirer-specific, capturing heterogeneity in complementarities between the startup's technology and each incumbent's capabilities. This framework yields equilibrium prices and the division of surplus between the parties. Here I provide only the intuition and the propositions that form the basis for the paper's empirical tests; Appendix A provides full details on the model.

2.3.1 Differences in Acquisition Markets and Outside Options

Innovations differ fundamentally in their characteristics, shaping the challenges associated with their development and commercialization. Science-based startups typically pursue complex, novel, and uncertain innovations that depend on specialized capabilities and infrastructure to scale (Pisano, 1990; Gans and Stern, 2003; Henderson and Clark, 1990). Scaling these technologies, such as biopharmaceutical compounds, novel materials, or energy storage systems, often require dedicated advanced manufacturing, complex distribution networks, or regulatory expertise in order to realize their potential (Teece, 1986). In contrast, non-science-based startups, such as those in enterprise software or consumer products, are often easier to scale, less reliant on specialized assets, and customizable and applicable across a broader range of contexts (Bresnahan and Gambardella, 1998).

This difference in technological characteristics gives rise to fundamental differences in the structure of the exit environment. In terms of exit via acquisition, science-based startups frequently encounter a concentrated acquirer market characterized by a small number of large and dominant firms, resulting from the fact that only a limited number of incumbents possess the specialized complementary assets required to scale and profit from these innovations. This market structure is not a reflection of the startups' actions, of course, but stems from inherent industry dynamics tied to the technologies they develop.

High R&D intensity and large, technology-specific sunk costs create entry barriers that give rise to increasing returns to scale (Sutton, 1991; Cohen and Levin, 1989; Cohen and Klepper, 1996). Because the fixed cost of, for example, building and operating laboratories or production facilities does not fall proportionally with output, average costs decline for firms that can spread these expenses over a larger revenue base (Gans and Stern, 2000). Potential entrants must incur the same indivisible investments yet compete against incumbents that have already absorbed them, which deters entry and leads to endogenous concentration. The resulting market structure is oligopolistic:

only a few firms possess the cash flow, complementary assets, and cumulative know-how required to finance continuing R&D and to integrate new technologies (Cohen et al., 1990; Cohen and Klepper, 1996; Sutton, 2007). Consequently, acquisition markets for complex, science-based innovations are more likely to be thin, with just a handful of incumbents able and willing to absorb and commercialize at large scale such technological advances.[21]

Conversely, non-science startups face a much broader and more competitive acquisition market than science-based ventures, as the technology utility is not limited to a small, specialized group of acquirers. Because their innovations have wide applicability across many use cases and industries with little adaptation costs (Henderson and Clark, 1990; Bresnahan and Gambardella, 1998; Ewens et al., 2018), a larger and more heterogeneous set of firms view them as complements. Potential acquirers can range from small niche players to large incumbents in finance, healthcare, and retail, reducing the degree of market concentration.[22]

Stylized Fact 1. *Science-based startups tend to face more concentrated acquisition markets than non-science-based startups, i.e., a smaller number of potential acquirers.*

A feature of specialized technologies and concentrated markets is that, when a startup fails to attract interest from dominant incumbents, it tends to face a fragmented set of smaller acquirers. The specialized nature of the technology means that, if none of the large incumbents are willing to acquire it, only the remaining, smaller firms operating within similar technological domains are technically capable of doing so. These firms may be able to exploit and further develop the innovation, but they have a narrower customer base and lack the scale and market access needed to generate the same level of value as a large incumbent would. As Klepper and Thompson (2006) and Sutton (2007) pose, when industries are composed of distinct submarkets with varying fixed costs and competitive intensities, smaller firms may persist in niche segments, but their ability to extract value from frontier innovations remains limited relative to dominant incumbents operating at scale.[23] For non-science-based startups, this structure does not hold. As the pool of potential

[21] While in the framework developed here I treat these exit conditions as exogenous, they are plausibly endogenous to the nature of the innovation, consistent with models linking technical change to market structure (e.g., Sutton, 1991). The number and scale of potential acquirers, as well as the feasibility of independent scaling, are shaped by the underlying characteristics of the technology. Innovations that require tightly integrated or highly specialized complementary assets—common in science-based domains—tend to generate more concentrated downstream markets, limiting the set of viable acquirers, who are often larger incumbents. These same characteristics also influence whether critical complementary assets, such as manufacturing infrastructure or distribution networks, can be accessed through arm's-length transactions or must be developed in-house.

[22] Technologies may evolve along the specialization—generality tradeoff. Innovations that are initially specialized can become more general-purpose as access to complementary assets improves, integration costs decline, and downstream demand becomes clearer (Bresnahan and Gambardella, 1998). For example, the rise of regional ecosystems, such as the concentration of semiconductor design, manufacturing, and supporting services in Silicon Valley, can reduce commercialization frictions over time by making complementary assets more widely available (Saxenian, 1996).

[23] One may ask why large incumbents may pass on technologies that smaller firms are willing to acquire. This could be due to a combination of economic and organizational factors. Economically, the replacement effect discourages adoption when new technologies threaten to cannibalize existing products or revenue streams (Arrow, 1962). Likewise, the innovation may also be poorly aligned with the incumbent's existing asset base, limiting expected complementarities and raising coordination costs (Teece, 1986). Organizationally, innovation myopia, internal incentive structures favoring short-term performance, and a bias toward incremental improvements over disruptive change may reduce willingness to invest in certain technologies (Christensen, 2015).

acquirers expands, the average size—and especially the maximum size—of bidders increases. By order-statistics logic, a larger and more heterogeneous pool, often spanning less concentrated or adjacent industries, raises the likelihood of drawing a highly resourced incumbent. This increases the expected scale of complementarities, increasing the startup's value. In Appendix B, I provide case examples to further illustrate these dynamics.

Stylized Fact 2.1. *For science-based startups, an increase in the number of potential acquirers is associated with a decrease in the average size of acquirers.*

Stylized Fact 2.2. *For non-science-based startups, an increase in the number of potential acquirers is associated with an increase in the average size of acquirers.*

In terms of independent commercialization, science-based startups are significantly less likely to scale on their own, as the complementary assets required are costly to develop, rarely available through contracting, and often co-specialized with incumbents following years of cumulative investment and organizational learning (Kapoor and Furr, 2015; Kapoor and Klueter, 2021; Moeen, 2017). As a result, the cost and risk of independent commercialization are high, making these ventures more reliant on acquisition (Andrews et al., 2022). In contrast, non-science-based startups face lower barriers to scale independently. Their technologies typically require less capital-intensive infrastructure, allowing startups to grow in stages, with greater flexibility to experiment, iterate, and adjust to market feedback (Ewens et al., 2018; Kerr et al., 2014; Koning et al., 2022)—making independent commercialization more feasible and, in negotiations, a credible threat.

Stylized Fact 3. *Science-based startups tend to face weaker options for independent commercialization than non-science-based startups.*

2.3.2 Differences in Value Creation and Capture

Combining the mechanics of value creation and capture described in the previous section with the structure of the acquisition market and outside options, the resulting comparative statics follow. As discussed, all else equal, the size of the acquirers in the acquirer pool affects value creation. Larger acquirers can extract greater value from a given innovation, increasing total surplus. Thus, I derive the following propositions:

Proposition 1.1. *Science-based startups tend to generate greater joint surplus than non-science-based startups, as their acquirers are typically larger incumbents more capable to fully realize the innovation's potential.*

Proposition 1.2. *For science-based startups, joint surplus declines as the pool of potential acquirers increases, as additional acquirers tend to be smaller. In contrast, for non-science startups, a larger acquirer pool increases the likelihood of attracting a large acquirer, leading to higher surplus creation.*

Likewise, the number of potential acquirers, their size, and the feasibility of independent commercialization influence the share of surplus the startup is able to capture. Thus:

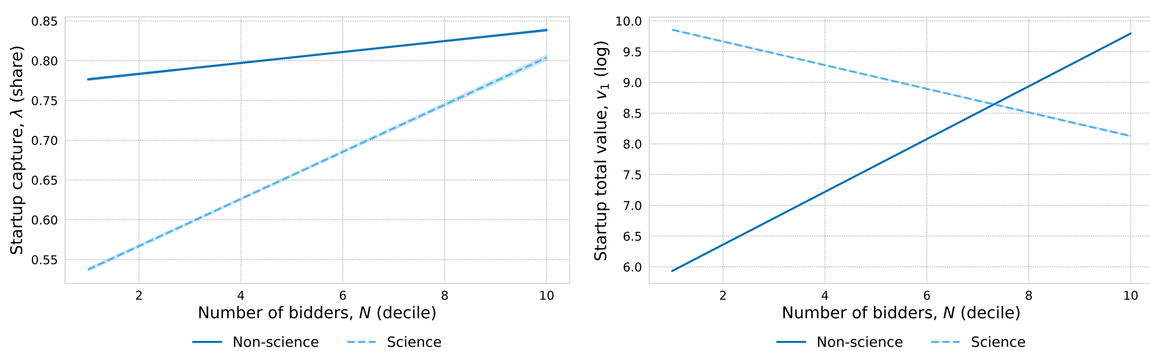
Proposition 2.1. *Science-based startups face lower value capture because they tend to operate in acquisition markets with a smaller pool of potential acquirers, which limits competition and reduces their bargaining power at exit.*

Proposition 2.2. *For both science-based and non-science-based startups, value capture increases with the number of potential acquirers, reflecting stronger competition among acquirers.*

Proposition 2.3. *The strength of the startup's outside option mediates both the level and the sensitivity of value capture to market structure. Startups with strong outside options—able to credibly pursue independent commercialization—achieve higher baseline capture and are less sensitive to the number of acquirers. Conversely, when outside options are weak, value capture is lower and more responsive to acquirer competition; in this case, science-based startups exhibit a steeper slope, indicating greater sensitivity to market structure.*

Two mechanisms further explain Proposition 2.3. First, as bidder count rises, valuation dispersion is reduced, reducing the difference between the best and second-best offers and, thus, increasing the startup surplus captured. Second, weaker external offers improve the credibility of the outside option, strengthening bargaining power. However, in both cases, this higher capture comes at the expense of lower value creation.

To illustrate the model's central predictions, I run simulations calibrated to match the moments in my data. Figure 2 summarizes the propositions described above by plotting the startup's value capture (left panel) and the joint surplus (right panel) as a function of the number of potential bidders, separately for science-based and non-science-based startups. As shown in the results section, the empirical patterns in my data closely align with the model's predictions.



(a) Startup value capture (b) Value created

Figure 2: The figure illustrates the model's predictions based on simulations with moments matched to my data. Panel (a) plots the value captured by the startup. Value capture increases with the number of bidders for both startup types, consistent with greater acquirer competition. However, the slope is markedly steeper for science-based startups, reflecting their lower reservation value, which heightens their dependence on external bids, and the smaller dispersion in acquirers' valuations. Panel (b) plots the value created (logged) as a function of the number of bidders. For science-based startups, value creation decreases as the number of bidders increases due to a decline in average acquirer size, which reduces both total surplus and willingness to pay. In contrast, value creation for non-science startups increases with the number of bidders, as a larger pool raises the likelihood of matching with a high-value acquirer.

3 Data and Measurements

This section presents the data and measurement construction for the empirical analysis. Section 3.1 describes the methodology for estimating value creation and capture for each startup acquisition in the sample; Section 3.2 describes the data sources, with the primary datasets drawn from PitchBook, CRSP, and Refinitiv; Section 3.3 details the empirical procedure to estimate value creation and capture. Finally, Section 3.4 details the classification procedure used to identify whether a startup commercializes scientific innovations, and Section 3.5 details the approach for measuring a startup's set of potential acquirers and the characteristics of those acquirers.

3.1 Estimation of Value Creation and Capture

Addressing the central question of this paper requires measuring the *joint surplus* generated by a startup at the time of acquisition, a construct that has not been systematically quantified to date. Measuring the long-run joint surplus ultimately generated by an innovation is challenging, especially if one considers the uncertainty of future developments. In many cases, the economic rents from innovation unfold over time and are shaped by unpredictable factors such as an evolving competitive landscape, emerging complementary innovations, and organizational complementarities not present at the time of acquisition.[24]

To tackle this challenge, I develop a simple methodology that relies on startup acquisitions where the acquirer is a publicly listed firm. These transactions offer a clear empirical setting to estimate both the total surplus generated by an innovation and the share captured by the startup. At the moment of acquisition, the startup exits and realizes its private return, while the acquirer assumes control of the innovation and its future benefits. Crucially, the surplus accruing to each party at this point can be estimated, providing a basis for analyzing how the total value is created and divided. To conduct this estimation, and consistent with the theoretical framework, I formally decompose the joint surplus, denoted V_t, into two measurable components: The startup's surplus, V_s, which represents the net surplus captured by the startup's shareholders (e.g., founders and investors), and the incumbent's surplus, V_i, which reflects the gains accruing to the acquiring firm. Thus,

$$V_t = V_s + V_i \qquad \text{(Joint surplus)} \tag{2}$$

On the one hand, the startup's surplus, V_s is directly observable from the acquisition price and the startup's investment history. I define it as the difference between the net acquisition

[24] For example, consider NVIDIA's 2008 acquisition of Ageia, a company that had developed the first dedicated Physics Processing Unit (PPU) for simulating real-world physical interactions in software. When acquired, this technology was designed to handle complex physics calculations in real time, such as collisions, rigid-body dynamics, and fluid motion. Following the acquisition, NVIDIA integrated Ageia's PPU engine into its GPU architecture. While initially aimed only at the gaming market, the same simulation capabilities later provided useful for high-fidelity virtual environments used in autonomous vehicle development. As NVIDIA noted, "they wound up building a solution that is so realistic, it's now used as part of the foundation of NVIDIA's self-driving car technology".

price received, P_s, and the total accumulated equity investment in the venture up to the point of acquisition, T_s:

$$V_s = P_s - T_s \qquad \text{(Startup's surplus)} \tag{3}$$

Note that the net price P_s includes all forms of consideration, such as earn-outs and stock options, and discounts the debt, if any, that the startup had contracted at the time of acquisition.[25] Likewise, the investment T_s refers specifically to the accumulated monetary capital other than debt allocated by the startup's shareholders (e.g., early investors, venture capital, and private equity), and excludes non-dilutive financing such as grants and subsidies, which do not represent equity claims and, thus, do not factor into the private surplus calculation. While non-dilutive funding may certainly affect a startup's survival or trajectory, they do not alter the residual economic rents accruing to shareholders and are therefore excluded from this measure of surplus.[26]

On the other hand, I use an event study to estimate the acquirer's surplus associated with the acquisition, V_i, based on the change in the acquiring firm's market value around the time of the deal announcement. Crucially, because the acquirer is publicly traded, I observe the market's valuation of the expected gains associated with the acquired technology. The key assumption is the stock market reaction reflects the expected net present value of future cash flows generated by combining the startup's assets and capabilities—such as its technology, intellectual property, human capital, and inventive capabilities—with the incumbent's ones—such as manufacturing capabilities, complementary technologies and patent portfolios, and distribution networks.

This assumption is grounded in the Efficient Markets Hypothesis, which holds that asset prices incorporate all publicly available information and reflect consensus expectations over future payoffs in competitive markets with rational and profit-maximizing agents (Fama, 1970; Samuelson, 1965). Accordingly, the observed market reaction provides a forward-looking, risk-adjusted estimate of the acquirer's expected surplus, capturing not only the anticipated benefits of the innovation but also the risks associated with technical feasibility, market adoption, and future development costs.[27]

[25] It is common to report the total deal value as enterprise value, which includes both equity and debt components. To compute the equity value, i.e., the actual proceeds to shareholders, the debt must be subtracted from the reported price: P_s = Net Price = Equity Value = Enterprise Value − Debt

[26] It is also worth noting that this manuscript abstracts from factors such as time, risk exposure, and opportunity costs borne by entrepreneurs and investors. Alongside monetary investments, these factors could reasonably be considered part of the economic cost, reducing the startup's surplus V_s. However, and consistent with the data, I argue that these hidden costs are likely larger for science-based ventures, which typically require more upfront capital, longer development timelines, and face greater uncertainty. As a result, omitting these costs likely renders conservative estimates of the capture penalty between the two startup types—the adjustment would be larger for science startups than for their counterparts, while it would not affect the acquirer surplus. Likewise, working capital and other deal-specific adjustments that may affect the final equity price are not incorporated, as they are generally unobservable and relatively minor. Their exclusion is not expected to bias the results.

[27] For example, in early-stage biotechnology acquisitions, the incumbent's surplus at the time of acquisition reflects not only the anticipated benefits of the acquired asset but also the risks associated with clinical trial outcomes, regulatory approval, and the substantial future investment required for commercialization. All else equal, the same drug candidate in Phase II trials will generate a larger market reaction than if it were in Phase I, as more uncertainty has been resolved. Similarly, the acquirer may be willing to pay a higher price at a later stage of development. Importantly, if the rent-sharing dynamics are affected by the degree of risk at the time of the transaction, these will be absorbed by the measure, via the differential stock market reaction and acquisition price.

Therefore, I define the acquirer surplus as

$$V_{i,d} = M_{i,d} \times \mathbb{E}[v_{i,d} \mid r_{i,d}] \qquad \text{(Acquirer's surplus)} \tag{4}$$

where $\mathbb{E}[v_{i,d} \mid r_{i,d}]$ denotes the expected change in the acquirer's firm value attributable to the acquisition ($v_{i,d}$), conditional on the observed stock market return $r_{i,d}$. Since $r_{i,d}$ is a noisy measure of the acquisition specific value, I develop a filtering approach to recover the component of the return that is attributable to the acquisition itself. The methodology used to estimate this conditional expectation $\mathbb{E}[v_{i,d} \mid r_{i,d}]$ is described in detail in the following section. $M_{i,d}$ is the acquirer's market capitalization at the time of the announcement, and the subscript d denotes the time window used to measure market returns, which is omitted throughout most of the manuscript for notational simplicity and also discussed further in subsequent sections.

The sum of these two components provides an estimate of the *joint surplus*—the total private value created by the innovation at the time of transfer, and the ratio of the startup surplus to this joint surplus yields the *startup capture share* λ_s.[28]

$$\lambda_s = \frac{V_s}{V_s + V_i} \qquad \text{(Startup capture share)} \tag{5a}$$

$$\lambda_i = 1 - \lambda_s = \frac{V_i}{V_s + V_i} \qquad \text{(Acquirer capture share)} \tag{5b}$$

Note that the methodology outlined thus far is consistent with the conceptual framework introduced in the previous section (Equation 1). As discussed, the method estimates the joint surplus at the time of acquisition using the information available to both parties—that is, it values the innovation conditional on the prevailing state of the world, including unresolved technical risks, competing technologies, projected demand, regulatory and policy constraints, capital-market and macroeconomic conditions, and expected integration costs.[29]

Likewise, because the market reaction incorporates only information available at the announcement, it excludes unforeseen shocks that materialize later. This is appropriate for the paper's purpose, as entry, investment, and acquisition decisions are made on that same information set.

[28] Note that V_i represents the net incremental value to the incumbent's shareholders and that the stock market reaction is effectively subtracting the Enterprise Price paid to acquire the startup.

[29] In some cases, the acquiring firm already holds an equity stake in the target at the time of acquisition, often through a Corporate Venture Capital (CVC) investment. In such cases, part of the surplus attributed to the startup in the baseline calculation actually reflects returns on the acquirer's pre-existing ownership. To adjust for this, I compute the incumbent's ownership share, multiply it by the acquisition price, subtract the incumbent's initial investment, and reallocate the resulting amount from the startup's surplus to the acquirer's surplus. Let α_{CVC} denote the incumbent's ownership share, P the acquisition price, and T_{CVC} the capital invested by the CVC unit. The adjusted measures are:

$$V_s' = V_s - [\alpha_{\text{CVC}} \cdot P - T_{\text{CVC}}], \quad V_i' = V_i + [\alpha_{\text{CVC}} \cdot P - T_{\text{CVC}}].$$

This treatment is consistent with the main methodology, since the market-based estimate of V_i reflects expectations of future gains and is unlikely to incorporate the sunk cost of the CVC investment. In my sample, CVC-backed acquisitions represent 3.2% of all transactions. The empirical results are robust to the inclusion or exclusion of this adjustment, but it is implemented for conceptual consistency (see Appendix).

The relevant factor that conditions incentives is therefore the division of rents *expected* at that moment—clearly for the startup and, in most cases, for the incumbent as well—rather than the long-run, ex-post realization of those rents.[30]

This perspective is consistent with other rent-sharing studies in innovation contexts (e.g., Kline et al., 2019), which likewise focus on ex-ante surplus division as the driver of investment incentives. Figure 3 illustrates the components of innovation value at the time of acquisition, distinguishing between the startup's realized surplus and the acquirer's expected surplus based on known information—which are measurable—, and additional value components, such as externalities or unresolved future uncertainties, that lie outside the scope of the measurement.

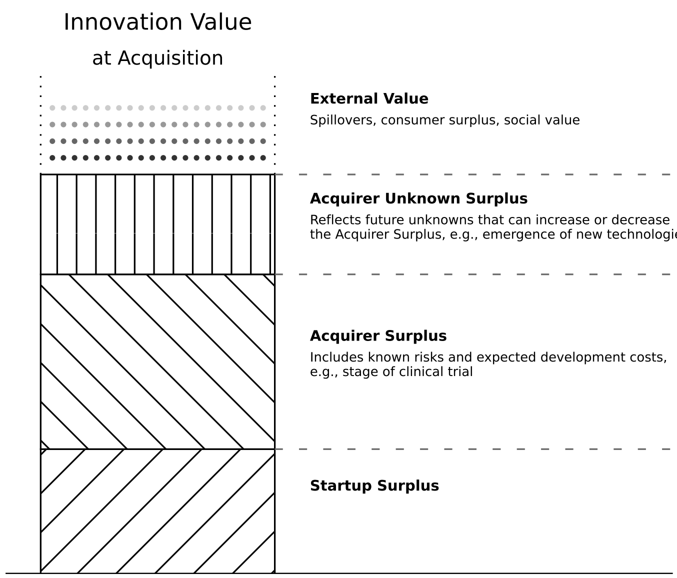


Figure 3: Schematic decomposition of innovation value at acquisition. The figure distinguishes between the startup's realized surplus, the acquirer's expected surplus based on known risks and development costs, and additional value that may accrue to the acquirer from unforeseen future events. It also highlights forms of external value not captured in private transactions, such as consumer surplus and knowledge spillovers. Only the startup and acquirer surpluses based on current information are observable and used to estimate the joint surplus at the time of acquisition.

3.1.1 Isolating the Signal from Noise

A complicating factor is that stock market abnormal returns r_i, a basis for computing the incumbent's private value V_i, can be noisy and reflect not only the acquisition's surplus, but also concurrent firm-level events and external conditions unrelated to the transaction under study.

[30]For instance, returning to the NVIDIA's 2008 acquisition of Ageia, the deal was motivated by the potential to enhance real-time physics simulation in gaming applications through the integration of PPU technology. The surplus at the time likely reflected the state of the industry, foreseeable technical developments, and anticipated commercial applications in gaming. It is reasonable to assume that subsequent uses of the same capabilities in autonomous vehicle development—applications that could not have been foreseen at the time—were not part of the original deal rationale. As such, this unknown upside was not reflected in the acquisition price or in the stock market reaction. On the other hand, it is also fair to assume that, had the parties and the market anticipated this future application, the joint surplus would likely reflect this information, and the division of rents would have adjusted accordingly. Again, this is true under the assumption of symmetric information, which I discuss further in the next section.

Large incumbents not only can release other relevant announcements but also face competitive threats and changes in macroeconomic conditions or industry-specific shocks that influence their share price around the acquisition announcement. That is, the observed abnormal return around the announcement date mixes signal with noise—spurious factors unrelated to the transaction's intrinsic value.

To isolate the acquisition-related signal from the noise and extract the expected value of the incumbent's private gains, I adopt a parametric approach inspired by Kogan et al. (2017) but adapted to the M&A context. My main adaptation is that, unlike assumed by the authors in the process of filtering out the signal from the noise,[31] I assume that acquisitions can destroy value for the incumbent and, thus, lead to negative abnormal returns. Indeed, it is well documented that acquirers often experience negative abnormal returns on acquisitions if investors believe the incumbent overpaid for the target, with the acquisition involving significant integration costs or organizational frictions. This is especially true in technology-based acquisitions (Benson and Ziedonis, 2010; Chondrakis et al., 2021; Higgins and Rodriguez, 2006), where is often argued that the value is mostly passed onto target firms (e.g., DeLong, 2001; Testoni, 2024).

The methodology is summarized as follows, with more details provided in Appendix C. I treat the observed abnormal return r_i as a noisy signal of the unobserved expected return from the acquisition, v_i. Specifically, I assume v_i follows a truncated normal distribution, truncated at an acquisition-specific lower bound $k_i < 0$. This lower bound is determined so that, while the private value for the incumbent v_i may be negative, the total value creation, V_t is non-negative. In other words, the transaction cannot destroy value in aggregate.[32]

The logic follows standard Bayesian updating: given a prior distribution for v_i and normally distributed noise, the posterior mean can be expressed in closed form, involving the normal density and distribution functions evaluated at the truncated point. The final outcome of this approach is an estimate of $E[v_i \mid r_i]$, where the parameters are estimated at the company, industry, year, and transaction level, using millions of stock prices and more than one hundred thousand M&A transactions. Formally, the expected return captured by the incumbent attributed to the acquisition, given the observed abnormal return r_i and truncation at k_i, is:

$$E[v_i \mid r_i] = \delta_i r_i + \sqrt{\delta_i}\sigma_{\varepsilon,it} \frac{\phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,it}}\right)}{1 - \Phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,it}}\right)}, \tag{6}$$

where $\phi(R_{it})$ and $\Phi(R_{it})$ denote the probability density and cumulative distribution functions of the standard normal distribution, respectively.[33]

This expression can be interpreted as follows. First, the term δ_i measures the share of the total

[31] The authors develop a methodology to compute a stock market-based measure of patent values by isolating signal from noise, under the assumption that patent grant announcements cannot destroy value.

[32] This assumption, while already an improvement over the original approach, remains somewhat restrictive as I still rule out the possibility that a transaction can destroy value in the aggregate. This simplification may overlook rare cases where the transaction could plausibly reduce total joint surplus.

[33] The results are robust to other distributional assumptions.

variance in acquisitions attributable to the incumbent's private value component—the signal-to-noise ratio. A higher δ_i implies that most of the variation in v_i is driven by private information rather than noise. Next, R_i is the standardized threshold above which v_i is truncated, reflecting the fact that we only observe v_i if it exceeds k_i. Finally, the ratio $\phi(R_i)/(1 - \Phi(R_i))$ adjusts the mean to account for this truncation, often referred to as the inverse Mills ratio. It captures how the tail of the distribution above k_i inflates the conditional expectation of v_i, given that only high realizations of v_i are observed. Finally, the expected value v_i is then used to compute the acquirer surplus, V_i, the joint surplus, $V_t = V_i + V_s$, and the corresponding share of value captured by the startup, λ_s.

It is important to note that recent research suggests this methodology may bias coefficients on the right-hand side when the private value estimate is used as the dependent variable (Arora et al., 2024a). The issue arises because the distributional assumptions in Equation 6 require all expected returns to derive from a single underlying distribution. In contrast, my econometric specification posits that the coefficient for science-based startups should differ systematically, implying that the incumbent's private value (as measured by the stock-market response) should be larger for such innovations. Arora et al. (2024a) propose a methodology to address this concern by allowing for two distinct distributions when estimating the expected returns, but I lack sufficient data to implement their correction. Instead, I conduct simulation exercises (reported in Appendix C) to assess the magnitude of this potential measurement error and the direction of the bias. Aligned with intuition, the simulations indicate that it biases the estimates downward. Therefore, the reported coefficients should be viewed as conservative, representing, in this respect, a lower bound on the true effect.

3.1.2 Measurement Assumptions and Limitations

The methodology developed in this paper interprets the acquisition price and the acquirer's stock market response as jointly revealing the expected value of the innovation at the time of transaction. Under the assumption of symmetric information and rational expectations, both parties, the startup and the incumbent, form a common belief over the distribution of future payoffs. The assumption is thus that the negotiated price reflects the seller's share of surplus, while the acquirer's abnormal return captures the acquirer's. This approach aligns with the semi-strong form of the Efficient Markets Hypothesis (EMH), which holds that publicly available information is quickly incorporated into prices (Fama, 1970), and it also mirrors the assumptions adopted in structural models of surplus division and M&A bargaining (e.g., Edmans, 2012). However, this assumption could be contested. A large body of theoretical and empirical work has documented how information frictions distort transaction outcomes in markets for complex assets. In settings marked by high uncertainty—such as science-based innovation—acquirers face serious difficulties in evaluating the quality, applicability, and long-term value of a technology (Testoni, 2022). Classic models of adverse selection predict that opaque assets will be underpriced or fail to trade altogether (Akerlof, 1970), while asymmetric information in capital markets distorts investment and financing decisions (Myers, 1984). In the M&A context, acquirers often face valuation uncertainty, leading to winner's curse dynamics and persistent mispricing (Boehmer et al., 2003; Moeller et al., 2005).

Empirical evidence confirms these theoretical concerns. Chondrakis et al. (2021) show that an institutional reform increasing patent transparency led to both a higher likelihood of acquisition and a more positive market response, suggesting that reduced information frictions raise both trade frequency and perceived value. Similarly, Higgins and Rodriguez (2006) find that acquirers with privileged access to scientific knowledge—either through prior alliances or domain specific R&D experience—earn significantly higher returns upon acquisition announcements, indicating that acquirers better informed can more accurately assess value and outbid rivals. In a related vein, research by Palermo et al. (2019) documents that externally acquired patents are more likely to be invalidated in litigation, implying that acquirers rationally discount prices to reflect unobservable quality risk.

These dynamics introduce two distinct implications, one concerning selection and the other concerning measurement and interpretation. First, asymmetric information between the transacting parties (i.e., between the startup and the acquirer) affects whether a deal occurs and how surplus is divided. High uncertainty or opacity, especially common in science-based ventures, may prevent transactions altogether if acquirers are unable to assess the technology's value or fear adverse selection. This introduces a selection bias in the observed sample: we are more likely to see deals involving lower-uncertainty innovations or situations where the acquirer has domain expertise or informational advantages. Second, even when deals do proceed, asymmetric information may distort bargaining outcomes. While in theory mispricing can go in either direction, in practice, informational disadvantages tend to suppress valuations—particularly for young, science-intensive startups that struggle to credibly signal the long-run value of their technologies without exposing themselves to appropriation. This does not introduce measurement error per se; rather, it reflects the realized outcome of bargaining under imperfect information, and is therefore part of the empirical distribution of surplus shares. I treat these distortions as features of the environment, not as flaws in the estimation, since they reflect the structural conditions under which innovation is commercialized.

By contrast, information asymmetries in capital markets, between insiders and external investors, raise concerns about measurement. Because the empirical strategy infers the acquirer's surplus share from stock market reactions, incomplete information on the part of market investors may attenuate the observed response. That is, when capital markets lack sufficient visibility into the nature or strategic relevance of a transaction, the announcement-period abnormal return may understate the true expected value to the acquirer. This type of friction leads to measurement error in the dependent variable, and, if anything, biases downwards the estimated acquirer's surplus. As such, my estimates should be viewed as conservative lower bounds.

3.2 Main Data Sources

I compile data from three primary sources. First, I utilize PitchBook, a comprehensive database of private and public equity transactions, which provides detailed information on startup funding rounds, valuations, and exit events. Using this source, I identify all startups acquired between 1990 and August 2024. During this period, from a total sample of 352,390 startups, 41,484 (11.77%) exited via acquisition.[34] This sample includes startups acquired by both public and private firms; however, for my analysis, I focus only on those acquired by publicly traded firms in the U.S.[35] This restriction is necessary because the value estimates, as defined in the previous section, rely on stock market reactions to acquisition announcements and, thus, on firms that are publicly traded.[36]

To identify which startups from PitchBook were acquired by publicly traded U.S. firms and to extract stock market data for computing estimates, I rely on CRSP, which provides detailed daily stock market information, including stock prices, shares outstanding, and turnover rates. PitchBook does not directly indicate whether an acquirer is publicly traded. However, both PitchBook and CRSP provide the Central Index Key (CIK), a unique company identifier assigned by the U.S. Securities and Exchange Commission (SEC), allowing for a direct match between PitchBook acquirers and stock market data.

Of the 46,144 acquisitions recorded in PitchBook, 20,309 (44.01%) were conducted by a company with an available CIK, indicating that the acquirer has relevant activity in the U.S. However, having a CIK does not necessarily mean that a company is publicly traded. This issue is easily resolved by matching the CIKs in PitchBook to those in CRSP, which inherently identifies publicly listed firms. By doing so, I determine which acquirers are publicly listed in the U.S. Among the 20,309 startup acquisitions with an available CIK, 13,172 are matched to a company in CRSP, representing a total of 3,605 unique incumbents.[37] These steps produce a dataset that, for each acquisition, includes information on the startup and its characteristics (e.g., founding year, funding history), details of the transaction (e.g., acquisition price, announcement date, closing date), and data on the acquirer, which includes stock market data for every trading day, particularly around the announcement

[34] For comparison, during this period, 10,303 (2.92%) exited via IPO, with the share of IPOs relative to M&A decreasing over time, a trend consistent with those documented by Ederer and Pellegrino (2023). The decline in IPO activity suggests that acquisitions have become the dominant commercialization pathway for innovation. If anything, excluding IPOs aligns the analysis more closely with the contemporary startup landscape, where most value creation and capture occur within acquisition markets rather than public listings. Empirically, the question is whether including IPOs as a route for value capture would alter the qualitative results. IPOs and acquisitions are fundamentally different exit strategies, with IPOs often representing firms that can sustain independent growth, while acquisitions may involve both startups capable of independent growth and those that require an incumbent's capabilities to scale.

[35] These firms are listed in the U.S. but are not necessarily headquartered there. Some acquisitions are made by international companies that trade on U.S. stock exchanges through Depositary Receipts, secondary listings, or subsidiaries and U.S. branches of foreign firms. In all cases, the estimation is based on the stock market reaction of the entity listed on a U.S. exchange. In the sample used for this analysis, few acquisitions are conducted by firms with non-U.S. headquarters, with the majority of these involving American Depositary Receipts (ADRs). For example, the final sample includes 20 acquisitions by Novartis AG, a Swiss multinational whose ADRs are traded on the New York Stock Exchange (NYSE).

[36] This restriction may introduce selection bias in the estimates. Regarding value creation, acquisitions by publicly traded firms are more likely to involve startups with stronger technological potential, leading to greater value creation compared to the average startup, regardless of whether the acquirer is public or private. These acquiring firms also tend to have the necessary capabilities to integrate and scale acquired technologies, reinforcing their role as key players in the acquisition market. On the other hand, the impact of public acquirers on value capture remains unclear.

[37] The CIK in the CRSP dataset is actually provided via a crosswalk table made available by Compustat. Furthermore, some minor cleaning is required on the PitchBook side, such as padding the identifiers.

period.

This dataset would be sufficient for the empirical analysis if I were not refining the estimates to isolate the signal from the noise associated with the market reaction to the acquisition. However, to achieve this, I need to estimate acquisition-specific parameters for each acquirer. Since the acquirers in the sample described have conducted far more acquisitions than those captured by PitchBook, these other transactions must be incorporated. Thus, I supplement the data on M&A transactions with Refinitiv (formerly SDC Platinum), a global M&A database. Refinitiv's data is not directly linked to CRSP or Compustat but, conveniently, Ewens et al. (2024), building on Phillips and Zhdanov (2013), developed a crosswalk. The authors provide a dataset that links Refinitiv M&A identifiers to Compustat and, in turn, CRSP.[38] This dataset includes linkages for 130,432 acquisitions available in Refinitiv. Of these, 108,181 acquisitions are matched to the CRSP sample—conducted by 13,877 acquirers. The set contains thus 108,181 acquisitions conducted by publicly traded U.S. firms, all of them matched to daily stock market data on the acquirer side.

This process results in two interrelated datasets. The first dataset contains 108,181 acquisitions conducted by publicly traded U.S. firms, supplemented with daily stock market data for each acquirer. This dataset is used to estimate the parameters needed to isolate acquisition-specific signals and compute expected abnormal returns for each M&A transaction. The second dataset contains 13,172 acquired startups, matched with estimated abnormal returns from the first dataset. It includes detailed information on both the startups and their acquirers, as well as estimates of the private values captured by acquirers and the total value created by startups.

As a last step, I further clean the final sample of 13,172 startup acquisitions by dropping observations with incomplete data, particularly those that have missing investment details or acquisition prices. After filtering out transactions with relevant missing data, the final sample consists of 5,823 startups.[39]

Figure 4 presents the temporal evolution of the number of startup acquisitions in the final sample by industry.[40] These trends align with previous studies on startup activity across industries (e.g., Lerner and Nanda, 2020). Startup acquisitions conducted by publicly listed U.S. firms over the past few decades have grown concentrated in Software & IT Services, Biotechnology & Pharmaceuticals, Consumer & Business Products and Services, and Health Care Equipment & Services. In contrast, Energy, Hardware, Industrials, Manufacturing & Materials, and Semiconductors show relatively low levels of acquisition activity, with only modest increases over time.

[38] The linkage is provided via acquirer gvkey.

[39] Missing transaction prices are common and account for the majority of observations dropped in the final sample selection. Even for acquisitions conducted by public firms, price disclosures are often absent, which may introduce selection bias. Acquisitions with disclosed prices are more likely to be those where the deal is considered strategically significant or where the acquirer wants to signal success to investors. This means that reported prices are more likely to correspond to highly successful acquisitions from the acquirer's point of view. Furthermore, price disclosure is compulsory over a certain threshold, which means that only deals under the threshold are missing. If price disclosure is systematically correlated with acquisition outcomes, the final sample may overrepresent high-value deals, leading to an upward bias in estimated value creation.

[40] The industry classification follows the Global Industry Classification Standard (GICS) developed by Morgan Stanley Capital International (MSCI). Refer to the Appendix for details on the classification methodology.

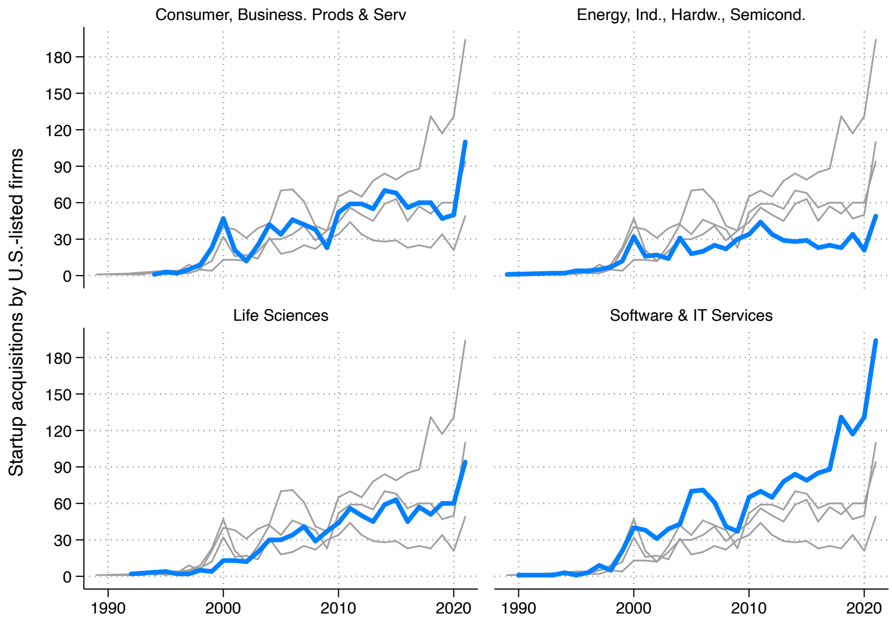


Figure 4: Number of startup acquisitions per industry and year (1990–2022). The data align with prior research on innovation activity, showing that startup acquisitions grow concentrated in life sciences (e.g., Biotechnology & Pharmaceuticals, Health Care Equipment & Services) and IT-related industries (e.g., Software & IT Services). In contrast, industries such as Energy, Industrials, Hardware, and Semiconductors display fewer acquisitions, with little growth after the 2000.

3.3 Value Creation and Value Capture Measurements

Using these data and the methodology outlined above, I compute the value created and captured by a startup. An important step is to identify first the time window of the event, i.e., for which days the stock market reacts to the acquisition announcement and, thus, I can estimate the acquirer's surplus from the acquisition. To that end, I look at the days in which the trading volume is abnormal. Figure 5 plots the share turnover around acquisition announcement days, reporting the coefficient estimates $b_l, l = [-4, 6]$ and 95% confidence intervals. While the days prior to the announcement the turnover is statistically the same, there is a significant increase in share turnover in the day of the announcement and the day after. This increased activity drops significantly two days after the announcement, although is still above pre-announcement days. By day 3, the trading volume goes back to baseline levels. These result suggest that the market significantly responds to the announcement during three days, from $[t, t+2]$. I will use this time window in my main results.[41]

[41] This time window is consistent with the literature, which notes that acquisitions are complex and the market takes a few days to assimilate all the information. In additional analysis, I test the robustness of my estimates with time windows of 1 and 2 days.

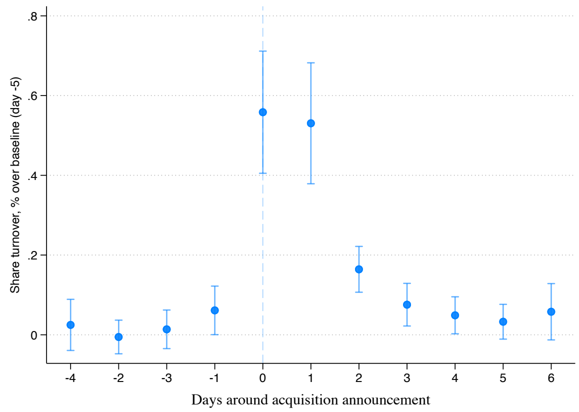
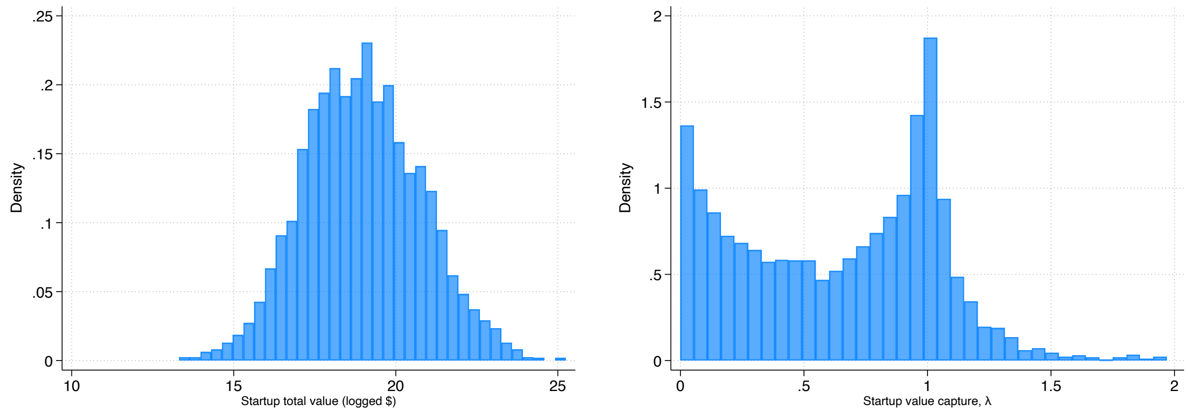


Figure 5: Share turnover around acquisition announcement days, in percentage points. Share turnover (s) is computed as the ratio of daily volume over shares outstanding. The Figure reports the coefficient estimates $b_l, l = [-4, 6]$ and 95% confidence intervals from the specification $s_{fd} = a_0 + \sum_l b_l I_{fd+l} + cZ_{fy} + \epsilon_{fd}$, where I is a binary variable indicating whether firm f announced an acquisition on day d; and Z_{fd} is a vector including firm-year and calendar day fixed effects. Standard errors are clustered at the firm-year and calendar day level. Baseline level is set at day $l = -5$. The mean and median share turnover on $l = -5$ are 1.06% and 0.61% respectively.

Within a three-day event window, I estimate the private value that acquirers derive from each acquisition. First, for every acquirer in the dataset, I calculate its observed daily abnormal return ($r_{i,d=[0,2]}$). Next, using historical acquirer-level data, I filter out noise to isolate the expected return ($v_{i,d=[0,2]}$) attributable specifically to the acquisition. This involves estimating both firm-level and deal-level parameters, with deal-level truncation defined as $k_i = -\frac{P_s}{market\_cap_{i,d=-1}}$. Next, I accumulate the return over the three-day window ($v_{i,d=[0,2]}$) and then compute the private value $V_{i,d=[0,2]} = mktcap_{i,d=-1} \times v_{i,d=[0,2]}$, where $mktcap_{i,d=-1}$ is the acquirer's market capitalization at the closing of the day previous to the acquisition announcement. Finally, I calculate the total value created (V_t) and the fraction captured by the startup, λ_s.

Figure 6 presents the distribution of total value created (panel (a)) and the fraction of value captured for the 5,823 startups in my final sample (panel (b)). The right tail of the distribution in Panel (b), where value capture exceeds one, may initially appear puzzling but is conceptually coherent. Recall that value capture reflects the share of the total surplus appropriated by the startup in the acquisition. Values above one imply that the incumbent paid more than the total realized value of the deal—i.e., the acquisition destroyed value from the acquirer's perspective. In such cases, the startup effectively walks away with more than the total surplus created, leaving the acquirer with a net loss. Beyond merely myopia, acquirers may overpay when facing competitive pressure, limited outside options themselves, or strategic motivations such as preemption or long-term positioning. Likewise, other factors common in bidding processes and prominent when competition intensifies, such as the winner curse or affiliated bidding, can lead to overpayment. Indeed, when regressing an indicator for overpayment on the number of potential acquirers, I find that the likelihood of overpayment (capture larger than 1) is significantly higher in thicker acquisition markets. This suggests that strong bargaining positions—fueled by competition among acquirers—can allow startups to extract a disproportionately large share of the surplus, even when that entails losses for the buyer.



(a) (b)

Figure 6: Panel (a) displays the dollar estimates distribution for the *total value created*, V_s. Panel (b) displays the estimates for startup *value capture* (λ_s). Values over 1 indicate that the acquisition destroyed value for the incumbent ($V_i < 0$).

Figure 7 illustrates the correlation between the incumbent's private value V_i and the VC funds raised by the acquired startup. This strong correlation suggests that market beliefs (adjusted by noise), as captured by the measure, align closely with those of venture capitalists. While a similarly strong correlation is observed when considering total joint surplus V_t, the incumbent's private value V_i provides a more precise comparison. Unlike joint surplus, which incorporates the price paid to the startup (and may indirectly reflect VC funding), the private value is influenced solely by market reactions.

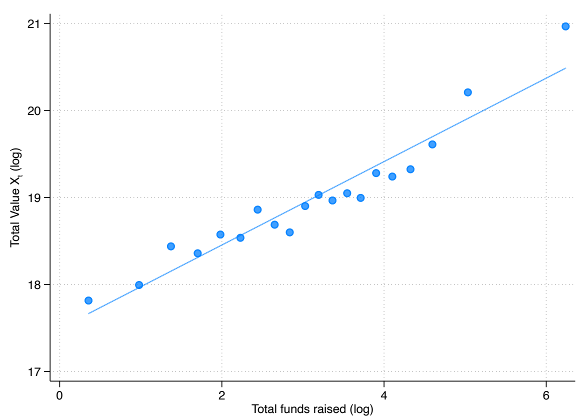


Figure 7: Cross-sectional relationship between the estimated joint surplus ($V_t = V_i + V_s$) and the VC and PE funds raised before the transaction. Variables are log-transformed.

3.3.1 Illustrative Examples

Next, I provide two examples to illustrate the dynamics of value capture in startup acquisitions based on my estimates. The first is DeepMind, acquired by Google in February 2014 for a total of $650 million, including cash and other considerations. The company had raised $61 million in venture capital prior to acquisition. Thus, the startup's surplus at acquisition was \$589 million (V_s). Founded in 2010 by U.K.-based computer scientists, DeepMind specialized in developing cutting-edge artificial intelligence and, among its major breakthroughs, later developed AlphaFold, a software that predicts protein structures. At the time of acquisition, DeepMind had approximately 75 employees, primarily scientists and personnel of technical character, and had only a promising technology, with no revenue.[42] Over the three days following the acquisition, Google's market valuation increased by \$1.59 billion, adjusted for market noise, representing its private value from the acquisition (V_i). Thus, the total estimated joint surplus, or value created (V_t), of DeepMind at the time of acquisition is given by $V_t = V_i + V_s = 1,590 + 589 = \$2,179$ million.

DeepMind's shareholders, including its founders and investors, captured 26.6% of the total value, while Google retained the remaining 73.4%. This level of value capture places the startup in the bottom 20th percentile. This division of rents suggests that the science-based startup, despite possessing a highly innovative technology, had limited bargaining power at the time of transfer, potentially driven by the lack of outside options. It is also important to note that any risks related to the uncertainty of DeepMind's AI applications, including commercialization and technological viability, are effectively accounted for in the estimation, as they are reflected in the stock market reaction.

At the other end of the spectrum, consider Postmates, a software platform for on-demand delivery and urban logistics, which connects customers with local couriers to facilitate the delivery of goods—and does not rely on scientific innovations. In 2020, Postmates was acquired by Uber for \$3,900 million. Uber's estimated private value gain from the acquisition was \$1,420 million, leading to a total economic value of \$5,320 million. In this case, Postmates captured 73.3% of the economic surplus upon transfer. Unlike DeepMind, which developed frontier technologies requiring specialized capabilities for scaling commercialization, Postmates operated in a mature and competitive market where it could credibly threaten to scale independently if necessary. Additionally, a potentially large number of acquirers further strengthened its outside options in negotiations.

3.4 Identifying Scientific Innovations

I define science-based innovations as technologies that strongly rely on the development and application of novel scientific advances, often originating from fields such as life sciences, chemistry, physics, and engineering. Unlike incremental technological improvements, scientific innovations typically expand the frontiers of knowledge and introduce novel solutions to complex problems (Fleming, 2001; Fleming and Sorenson, 2004). These innovations often require extensive R&D, specialized expertise, and long validation cycles before they can be successfully commercialized (Hall and Lerner, 2010). Therefore, I define a science-based startup as one that develops its products, technologies, and services using science-based innovations, regardless of whether the underlying knowledge originates internally or from external sources such as universities, research laboratories, or other scientific institutions. These startups translate scientific discoveries into marketable products, services, or

[42] Source: DeepMind 2014's annual return.

processes, often operating in industries such as biotechnology, advanced materials, clean energy, and artificial intelligence.

Identifying a science-based startup involves assessing the extent to which its technology is rooted in novel scientific research. While several methods exist,[43] in this paper I rely on a Large Language Model (LLM) to assess the extent to which startups rely on novel scientific research. A growing body of literature demonstrates the effectiveness of LLMs for classification tasks in economics and innovation research (e.g., Dell, 2024). These studies highlight the accuracy and efficiency of LLM-based methods, often matching or exceeding manually curated labels (Durvasula et al., 2024) and, for example, being able to classify scientific research based on its promise for commercial application with high accuracy (Masclans et al., 2025). Notably, studies also acknowledge limitations such as biases in training data and sensitivity to model selection and prompts (Ash and Hansen, 2023; Carlson and Burbano, 2024).

For each startup in the sample, I first collect extensive textual data from multiple sources, including current and historical content from the startup's website, news articles, and SEC filings related to its acquisition. This textual data provides a comprehensive view of the startup's business and innovation activities and is processed by the LLM for classification.[44]

Next, I input the collected textual data, along with the year of the startup's founding, into a Large Language Model (LLM) and instruct it to assess the extent to which each company relies on novel scientific knowledge to develop its technologies, considering the years of activity. Specifically, I use Llama 3.3 70B, a state-of-the-art, instruction-tuned model developed by Meta and released in December 2024. This model is optimized for tasks requiring nuanced language understanding and contextual reasoning. Furthermore, Meta's LLMs are open-source, significantly reducing costs while delivering performance comparable to proprietary models from other leading developers, such as OpenAI. A key advantage of Llama 3.3 is its 128,000-token context length, which allows it to process extensive, unstructured, and fragmented data from multiple sources. Because information

[43] A widely used approach involves patents, often supplemented by patent-to-paper citations to assess reliance on scientific knowledge. While this method is certainly useful, it has limitations that are particularly salient when applied to startups. First, startups have a lower propensity to patent than incumbent firms—many startups that develop novel technologies susceptible to be patented do not hold patents often due to the high costs associated not only with patenting but also with eventually enforcing them. As a result, they often rely on alternative forms of intellectual property protection, such as trade secrets (Graham et al., 2009; Bryan and Williams, 2021). Second, patents are an imperfect proxy for innovation and are prone to various measurement errors (Lerner and Seru, 2022), again, especially prevalent in startups. Among others, one key issue is that patents are frequently assigned to inventors, universities, or venture capital firms, rather than to startups themselves, complicating the task of linking patents to specific firms.

For example, in my sample of 5,823 startups, 734 are categorized as Biotechnology firms, a sector with a high propensity for patenting. However, only 417 of these firms (56.8%) are matched to patents, leaving 317 firms (43.2%) without any patent records. A closer examination of these 317 startups reveals that most are highly technological, actively developing advanced technologies, and often contributing to scientific research. This indicates that the absence of patents does not necessarily reflect a lack of scientific or technological activity. The issue is likely even more pronounced in other sectors where patenting is less common. The magnitude of this discrepancy is particularly notable given the selective nature of this dataset, which consists of successful startups acquired by publicly listed incumbents.

[44] Importantly, patent and scientific publication data are excluded from the LLM task to prevent bias. As discussed, many startups engaged in scientific innovation do not have patents or matched publications. Including these data could lead the LLM to classify startups with matched patent data as more science-based simply due to the availability of additional information, rather than an inherent difference in their reliance on science.

on a startup's reliance on science may be dispersed across website content, news articles, or SEC filings—without a predictable location—this extended context length improves the accuracy of information extraction.

The classification task involves the LLM assigning each startup a score from 1 to 5, reflecting the extent to which its technologies are grounded in novel scientific research. Each score is accompanied by a confidence metric, which enhances interpretability and robustness by quantifying the model's certainty in its classification. According to the LLM classification, 70.83% of startups receive a score of one and 5.39% of two, indicating that they do not develop scientific innovations. 5.56% of startups receive a score of three, while 15.03% receive a score of four, and 4.57% receive a score of five. The distribution is bimodal, with peaks at one and four, which is characteristic of classification tasks and suggests that the language model effectively differentiates between science-based and non-science-based startups. Scores in the middle of the distribution (two and three, representing 10.95% of the data points) correspond to cases with greater classification uncertainty, while scores at the higher extreme (five) are assigned with high confidence. This distribution indicates that the majority of startups do not engage in scientific research or develop technologies closely tied to novel scientific advancements.

For the analyses that follow, I use a binary variable to indicate whether a startup commercializes scientific innovations: I define science-based startups as those receiving a score of four or five, comprising 19.6% of the sample.[45] The Appendix provides details on the classification task, including the prompt used, the LLM configuration, and robustness analyses using alternative classification thresholds and model specifications that account for prediction uncertainty. Following Carlson and Burbano (2024), I also assess the sensitivity of the main results to variations in prompt design, which may affect the classification outcomes. Furthermore, to validate the LLM-based classification, I compare it with patent-to-paper citation data; the strong correlation between LLM-based scores and citation metrics for startups with observable patents supports the validity of the classification. Table 1 provides selected examples with the goal to illustrate the classification exercise.

[45] For comparison, using a loose measure—defining a science-based startup as one with at least one paper citation in its patents—9.6% of the sample would be classified as science-based. This upper limit in terms of patent-to-paper cites is notably lower than the 18.8% identified using the LLM classification. Additionally, it is important to acknowledge that the sample is selective, as it focuses on startups acquired by publicly traded firms, with a strong representation of biotech ventures. This composition does not necessarily reflect the composition of all startups in, for example, PitchBook, which should be considered when interpreting the results.

Table 1: Examples of startups according to the LLM-based classification. The table presents selected representative startups categorized into science (Panel A) and non-science (Panel B). While non-science startups may employ advanced technologies, they are not developing technologies based on novel scientific discoveries.

Panel A: Science Startups

Industry	**Acquirer**	**Startup**	**Startup Description**
Biotechnology, Pharmaceuticals	Astrazeneca	Alexion	Developer of biopharmaceutical drugs intended to transform the lives of people affected by rare diseases and devastating conditions. The company is involved in biotechnology research and offers therapies that treat paroxysmal nocturnal hemoglobinuria (PNH), atypical hemolytic uremic syndrome (aHUS), generalized myasthenia gravis (gMG) and neuromyelitis optica spectrum disorder (NMOSD) among others, enabling individuals to overcome the issues they face every day.
Energy	Aqua Metals	Ebonex	Manufacturer of a conductive ceramic battery designed to be used in a range of commercial cleantech applications. The company uses Ebonex conductive ceramic powder to produce a commercially viable bipolar lead acid battery, enabling users to use it in power storage, water treatment, and construction.
Hardware	Qualcomm	Qualcomm MEMS Technologies	Developer of iMoD technology for mobile products. The company offers iMoD technology, which is based on a Micro-Electro-Mechanical-Systems structure combined with thin film optics, a display technology that delivers display images with lower power consumption.
Industrials, Manufacturing, Materials	IDEX Corp	Precision Photonics	Manufacturer of optical components and coatings based in Boulder, Colorado. The company's products including high-energy laser mirrors, polarizing optics, beam-splitters, etalons, and micro-optics, as well as a custom service for difficult-to-create optics are offered to clients in telecommunications, defense, aerospace, biomedical, and semiconductor manufacturing sectors.

Panel B: Non-Science Startups

Industry	**Acquirer**	**Startup**	**Startup Description**
Software & IT Services	Oracle	Responsys	Provider of marketing software to design, execute and manage email campaigns designed to integrate comprehensive customer data analysis and management systems with targeted customer interaction systems. The company's Interact Suite comprises various integrated applications that enable the design, management and automation of tasks and processes for executing email and cross-channel marketing campaigns, providing enterprises with the ability to have personalized web interactions with customers.
Energy	SolarCity	Zep Solar	Designer of mounting hardware for photovoltaic (PV) systems. The company offers its patented Zep Groove PV module frame technology, a catalog of its own mounting and grounding hardware and third-party compatible products.
Consumer, Businesses Products & Services	Amazon	Zappos.com	Retailer of clothing, footwear and accessories based in Las Vegas, Nevada. The company offers shoes and clothing for men and women, bags and luggage, accessories, boots, slippers, electronics, eyewear, watches and jewelry.
Health Care Equipment	Hologic	Bolder Surgical	Manufacturer of surgical instruments and devices intended to elevate experiences in surgery. The company specializes in the development of right-sized surgical devices in order to improve procedural approaches and minimize incisions, enabling pediatric surgeons to safely and effectively perform minimally invasive surgery in pediatric patients from neonate to teenager.

3.5 Characterizing Potential Acquirers

In this section, I describe the measurement approach used to estimate the number of potential bidders for a given startup, a key variable for studying the mechanisms that link market structure to value capture. One approach to approximating the number of potential acquirers is to examine product-market competition in the acquirer's domain at the time of acquisition. Specifically, I define a startup's set of likely acquirers as the incumbents competing with the actual acquirer in the downstream product market, in the acquisition year.

These competitors can be measured using textual data, as product-market similarity scores derived from firm descriptions capture meaningful overlaps in market offerings and provide a systematic basis for identifying competitive relationships. The intuition is that incumbents tend to acquire startups that complement or enhance their existing products, technologies, and capabilities. For example, a startup developing advanced battery technologies for grid storage may attract interest from multiple energy incumbents, while a biotechnology venture working on novel drug delivery mechanisms could be relevant to several pharmaceutical firms seeking to expand their R&D pipelines. In software, a startup specializing in enterprise resource planning for human resources may appeal to incumbents offering complementary business applications.[46]

To operationalize this idea, I draw on the Text-Based Network Industry Classifications (TNIC) developed by Hoberg and Phillips (2010, 2016), which provide a continuous measure of firm-level product-market similarity based on cosine scores from 10-K product descriptions. This approach embeds firms in a high-dimensional space where proximity reflects overlap in product offerings and has been widely used to study competition, acquisitions, market concentration, industry relatedness, and innovation spillovers. Importantly, the measure is computed annually, offering a time-varying indicator of product-market competition well suited to this analysis. I use the most recent version, the Embeddings-Based TNIC Industry Classifications (ETNIC) released in 2024, which also extends coverage to 1989–2023 (Hoberg and Phillips, 2025).

I define the acquirer's competitive set as the number of firms with a cosine similarity score exceeding the 0.2 threshold in the year of the acquisition, consistent with Hoberg and Phillips (2016, 2025). For a given acquirer—and, by extension, the associated startup—a higher count of such firms indicates a denser competitive environment and a broader pool of potential acquirers. Conversely, a sparse ETNIC network reflects limited product-market competition, reducing the number of viable acquirers and weakening the startup's bargaining position.[47]

Let $S_{ij,t}$ denote the cosine similarity between firm i (the acquirer) and firm j in year t, where t corresponds to the year of acquisition. I define a startup's number of potential acquirers as the size of the acquirer's competitive set in the acquisiton year as follows:

[46] Conversely, some acquisitions are made specifically to eliminate competition by shutting down similar projects (Cunningham et al., 2021). In this draft, I abstract from this scenario.

[47] The data matching is straightforward. I use the gvkey identifier available in both the CRSP and Hoberg and Phillips (2025) datasets. The acquisition year is used for matching, ensuring that the pool of the acquiror's competitors reflects the market conditions at the time of the acquisition.

$$\text{Potential Acquirers}_{it} = \sum_{j \neq i} \mathbb{1}\left(S_{ij,t} > \tau\right)$$

where $\mathbb{1}(\cdot)$ is the indicator function, equal to 1 if the condition is satisfied and 0 otherwise; $\tau = 0.2$ is the similarity threshold, consistent with Hoberg and Phillips (2025); and the sum is taken over all U.S. publicly listed firms $j \neq i$ in year t.

Figure 8 presents four illustrative examples of the measure. Each panel displays the network of all US-listed firms based on product-market rivalry in 2016. The selected firms are active acquirers within my sample. In each network, the focal firm is highlighted in dark red, while its direct competitors are shown in lighter red.

Consider, first, Oracle and Cisco, two major firms in the enterprise technology sector with substantial overlap in product offerings, including software, networking, and cloud services. Oracle is associated with 33 competitors, while Cisco has 23. They share several common competitors, such as Microsoft, VMware, Arista Networks, and F5. At the same time, the measure captures differences in their competitive landscapes through firm-specific competitors. For instance, Oracle's specialization in databases and enterprise software aligns it with firms such as Red Hat, a provider of open-source enterprise solutions. In contrast, Cisco's focus on networking hardware and infrastructure is reflected in competitors like Ciena, which supplies optical networking equipment and related technologies.

Second, consider Coherent, an industrial firm specializing in novel, science-based laser and photonic solutions across industries such as automotive, electronics, and life sciences. While Coherent appears proximate to Oracle and Cisco in the product-market similarity space—likely because some of the technologies it develops have applications in computing hardware—its competitive domain remains distinct. It has only 9 direct competitors and does not share overlapping rivals with either Oracle or Cisco. This indicates that despite some technological adjacency, Coherent operates in a separate product-market space, implying a non-overlapping set of potential acquirers. Moreover, the number of competitors is substantially lower—60 to 72% fewer—than that of Oracle or Cisco, underscoring Coherent's higher degree of specialization and niche positioning. This pattern is also consistent with the idea that firms commercializing scientific innovations tend to face fewer direct rivals.

Third, consider Boeing, which is positioned much farther away in the similarity space. Its focus on large-scale aviation systems, defense technologies, and related infrastructure results in a distinct competitive environment, with 12 identified competitors, including major incumbents in defense and aerospace such as Hexcel, Lockheed Martin, General Dynamics, and Northrop Grumman.

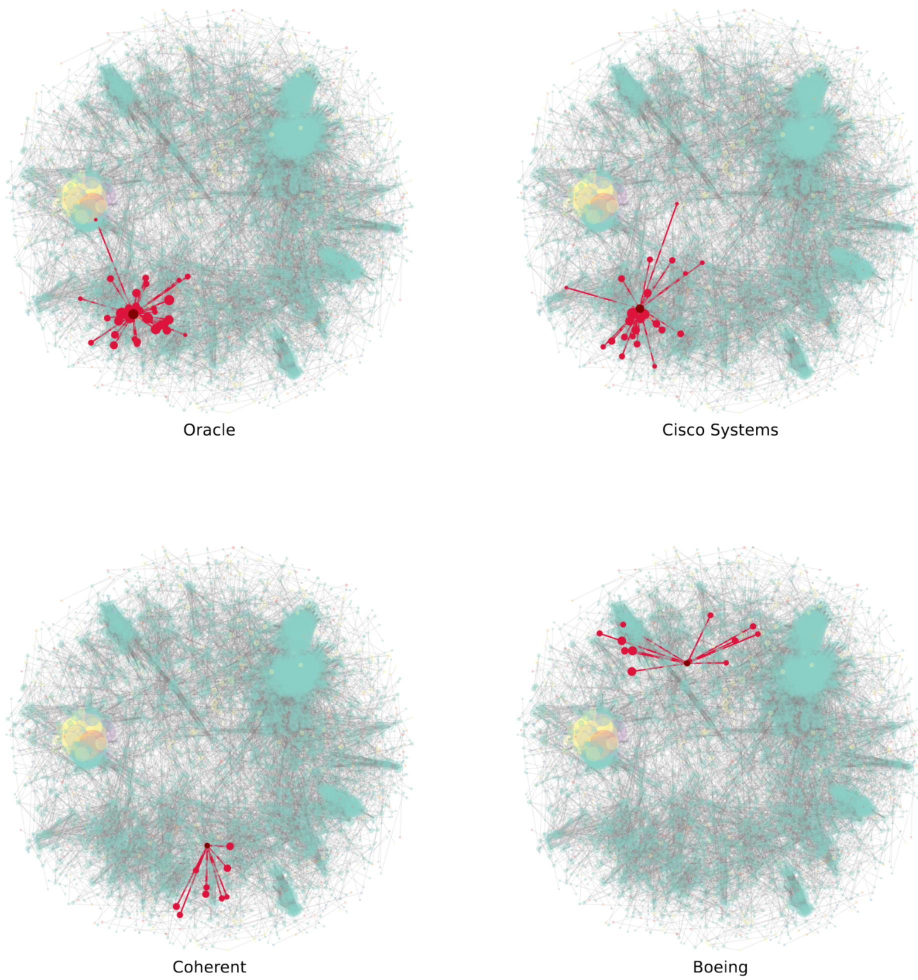


Figure 8: Product market networks for Oracle, Cisco, Coherent, and Boeing in 2016, constructed using text-based measures of product market similarity derived from 10-K filings by Hoberg and Phillips (2025). Each node represents a publicly traded U.S. firm, and edges indicate the degree of textual overlap in product descriptions. In each panel, the focal firm (shown in dark red) and its closest competitors (lighter red) form a sub-network of firms with the highest product-market similarity scores, illustrating key rivals within an industry. Oracle and Cisco occupy similar regions of the network, share some competitors, and exhibit partial overlap. By contrast, Coherent—although spatially close to Oracle and Cisco—shows no direct overlap or shared competitors, reflecting its distinct specialization. Boeing is located much farther away, with no overlapping competitors. These patterns illustrate how text-based similarity measures can identify clusters of close rivals and highlight their relative positions in the broader product market space.

3.6 Summary statistics

Table 2 reports summary statistics for the main variables of interest. Panel A presents the full sample, while Panels B and C stratify startups by non-science-based and science-based categories.

Table 2: Summary statistics for main variables of interest.

Panel A: All startups (N = 5,823)

	Mean	SD	Min	p25	p50	p75	Max
Market-adjusted return, v_i	0.010	0.027	-0.118	0.001	0.008	0.016	1.068
Acquirer surplus, V_i	365	1,432	-6,877	3	26	153	36,180
Acquisition price, P	423	2,054	0.002	20	66	231	80,269
VC, PE investment, T_s	59	474	0	2	13	42	29,505
Startup surplus, V_s	391	2,041	0	9	47	193	80,269
Joint surplus, V_t	756	2,489	0	35	127	495	78,732
Startup capture, λ_s	0.535	0.455	0.000	0.085	0.493	0.938	5.589
Acquirer market cap.	36,607	122,113	1	822	3,399	17,725	2,698,909
Science-based Startup	0.194	0.397	0	0	0	0	1
Potential acquirers	11	21	1	2	6	13	426
Startup revenue at acquisition	255	1,646	0	7	30	104	39,759

Panel B: Non-science based startups (N = 4,694)

	Mean	SD	Min	p25	p50	p75	Max
Market-adjusted return, v_i	0.010	0.022	-0.111	0.002	0.008	0.016	0.431
Acquirer surplus, V_i	368	1,495	-6,877	3	26	146	36,180
Acquisition price, P	332	1,374	0.002	19	60	201	35,000
VC, PE investment, T_s	53	458	0	1	11	38	29,505
Startup surplus, V_s	303	1,355	0	9	43	168	35,000
Joint surplus, V_t	671	2,002	0	34	119	435	36,321
Startup capture, λ_s	0.527	0.459	0.000	0.078	0.466	0.930	5.589
Acquirer market cap.	36,131	128,803	2	829	3,186	15,715	2,698,909
Potential acquirers	10	20	1	2	6	13	426
Revenue at acquisition	249	1,508	0	9	32	114	36,750

Panel C: Science-based startups (N = 1,129)

	Mean	SD	Min	p25	p50	p75	Max
Market-adjusted return, v_i	0.010	0.044	-0.118	0.001	0.007	0.014	1.068
Acquirer surplus, V_i	352	1,138	-1,719	2	30	190	12,506
Acquisition price, P	796	3,694	0.016	26	100	407	80,269
VC, PE investment, T_s	84	533	0	3	24	73	16,500
Startup surplus, V_s	757	3,688	0	10	75	355	80,269
Joint surplus, V_t	1,107	3,878	0	40	187	754	78,732
Startup capture, λ_s	0.565	0.438	0.000	0.128	0.575	0.962	2.822
Acquirer market cap.	38,567	89,514	1	811	4,968	27,087	908,886
Potential acquirers	13	22	1	2	6	13	215
Revenue at acquisition	278	2,138	0	2	18	74	39,759

Monetary values are reported in millions of U.S. dollars, except for market capitalization, which is reported in billions.

Substantial heterogeneity exists across industries. For example, in the life sciences, where science-based startups are disproportionately represented, the average value capture is significantly higher. To contextualize these patterns, Table 3 reports the distribution of science-based startups by industry. The results reveal some interesting patterns. On the one hand, some industries show minimal variation in the composition of startups regarding their reliance on science or non-science—e.g., software and consumer and business products and services.[48] On the other hand, the fields identified by current research, policymakers, and practitioners as having insufficient activity and for which concerns are often raised are those that present a more mixed composition of startups. These include startups developing novel scientific advances alongside others focusing on developing more incremental technologies.

Table 3: Frequency of science-based startups by industry.

Industry	Science (%)
Life Sciences	68.6
Semiconductors	26.2
Energy	24.2
Industrials, Manufacturing & Materials	31.4
Hardware	11.9
Software & IT Services	0.7
Consumer, Businesses Products & Services	0.0
Total	19.4

For example, in Industrial, Manufacturing & Materials, 31.4% of startups rely on science—these include startups in Aerospace & Defense, Agriculture, Farm & Water, and Construction, Electrical Equipment, and Machinery; in Energy, 24.2%—with the majority renewable energy startups; in Semiconductors, 26.2%; and in Hardware, 11.9%. For example, in Aerospace and Defense, a startup might develop cutting-edge drones that do not necessarily classify as science-driven but are highly technological; conversely, advanced aerospace technologies like propulsion systems or quantum-based navigation systems do fall within the science-based category. Similarly, in energy, startups might focus on deploying renewable energy solutions or improving grid efficiency, which may not involve new scientific advances; or develop photovoltaic materials for frontier energy storage systems.

Most revealing, Figure 9 presents the average value capture by industry, distinguishing startups that commercialize scientific innovations (light blue) from their counterparts (dark blue). Because the focus is on the distinction between science and non-science startups, the two industries with no variation are excluded (software and consumer, businesses products and services). First, the average value capture for science-based startups tends to be lower across all industries. For instance,

[48] Companies are categorized into industries based on the Global Industry Classification Standard developed by Morgan Stanley Capital International (MSCI) and aggregated into broader industry categories, following other studies of startup activity across industries (e.g., Lerner and Nanda, 2020, 2023). Notably, the classification is based on the market or industry the startup serves rather than the technology it develops. Startups under Software & IT Services are generic developers of software technologies, encompassing also Artificial Intelligence. The Consumer and Business Products and Services category encompasses Commercial & Professional Services; Communication Services; Consumer Discretionary; Consumer Staples; Financial Services; Real Estate; and Transportation. The Industrials, Manufacturing, and Materials category includes Capital Goods such as Aerospace & Defense; Agriculture, Farm, & Water Equipment; Construction & Engineering; Construction Machinery; Materials, such as Chemicals and Metals; and Light and Heavy Electrical Equipment. Hardware includes Communications Equipment; Technology Hardware, Storage & Peripherals; and Electronic Equipment, Instruments & Components. Life Sciences includes Biotechnology; Biotechnology Equipment & Software; Health Care Equipment; Health Care Providers & Services; and Pharmaceuticals.

in Energy as well as in Industrials, Manufacturing & Materials, the value capture for science-based startups is the lowest across all sectors (0.41), followed by Hardware (0.47), and Semiconductors (0.55). Furthermore, the contrast in value capture between science-based startups and their non-science counterparts in these industries can be stark, with some science-based startups capturing close to half the value of their industry counterparts (e.g., Energy).

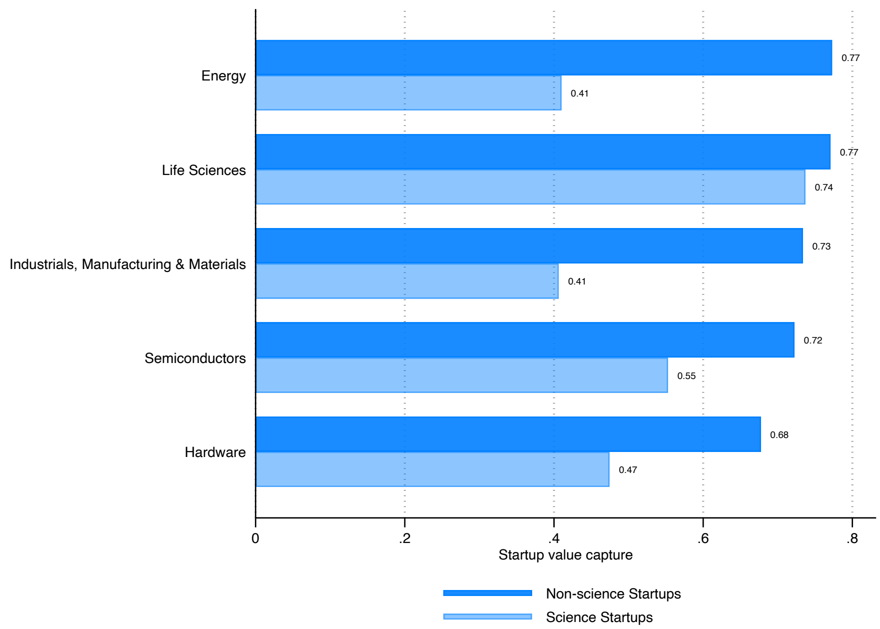


Figure 9: Average value capture by industry. The light blue bars represent startups that commercialize technologies based on novel scientific advances, while the dark bars represent non-science-based startups. The Industrials, Manufacturing, and Materials category includes Capital Goods such as Aerospace & Defense; Agriculture, Farm, & Water Equipment; Construction & Engineering; Construction Machinery; Materials, such as Chemicals and Metals; and Light and Heavy Electrical Equipment. Hardware includes Communications Equipment; Technology Hardware, Storage & Peripherals; and Electronic Equipment, Instruments & Components. Life Sciences includes Biotechnology; Biotechnology Equipment & Software; Health Care Equipment; Health Care Providers & Services; and Pharmaceuticals

Second, the life sciences show 1) substantially higher value capture for science-based innovations than that in other industries and 2) the difference between science and non-science based ones is small. Science-based innovations in the Life Science is composed mostly of Biotechnology and Pharmaceuticals startups (48%), while non-science based innovations are from Health Care Equipment (40%). As discussed in the following sections, this result could be attributed to the fact that Biotech and Pharmaceuticals have well develop markets for technologies, allowing startups to trade their technologies more efficiently (Arora et al., 2022).[49]

Further evidence on capture differences comes from simple stylized facts. Table 4 shows the

[49]Consumer and Business Products & Services and Software & IT services are omitted from the figure because these are industries with virtually non-scientific innovations, and the focus here is on differences between these. In terms of average capture, these two industries are at the bottom (0.64 and 0.59 respectively). While these results might seem puzzling at first, I examine in Section 5 differences across industries and reconcile them.

distribution of startups across value capture quartiles, by sector and by whether the startup is science-based. A clear pattern emerges: in all sectors except Life Sciences, science-based startups are disproportionately concentrated in the bottom capture quartile. For instance, in Energy, 55% of science startups fall in the bottom quartile compared to just 13% in the top; in Industrials, Manufacturing, and Materials, 51% fall in the lowest quartile compared to 11% in the top. In contrast, Life Sciences is the only sector where the distribution of science startups is relatively flat across quartiles, hovering consistently around 70% in each.

Table 4: Distribution of startups across value capture quartiles by industry and type. Each cell reports the percentage of startups in a given industry–type group (science-based or non-science) that fall into each quartile of value capture. Quartiles are defined based on the distribution of the startup's share of value capture at exit: Q1 = 0–27%, Q2 = 28–72%, Q3 = 73–98%, and Q4 = 99–559%. In all sectors except Life Sciences, science-based startups are disproportionately concentrated in the bottom capture quartile and underrepresented in the top. In Life Sciences, the distribution is notably uniform, with science-based startups accounting for roughly 70% of deals across all quartiles.

Capture Quartile	**Energy**		**Hardware**		**Industrials, Mfg. & Materials**	
	Non-science	Science	Non-science	Science	Non-science	Science
Q1	45%	55%	74%	26%	49%	51%
Q2	83%	17%	83%	17%	59%	41%
Q3	89%	11%	88%	12%	84%	16%
Q4	87%	13%	92%	8%	89%	11%
	Life Sciences		**Semiconductors**		**Total**	
	Non-science	Science	Non-science	Science	Non-science	Science
Q1	30%	70%	53%	47%	44%	56%
Q2	26%	74%	68%	32%	46%	54%
Q3	31%	69%	73%	27%	51%	49%
Q4	29%	71%	74%	26%	51%	49%

This suggests that in most sectors, science-based ventures face significant challenges in retaining value at exit, while in Life Sciences the structure of the market or institutional pathways such as markets for technology may mitigate this disadvantage. Overall, the evidence supports the interpretation that value capture constraints are tightly linked to the nature of the innovation and the surrounding commercialization environment. These statistics are further illustrated in Figure 10.

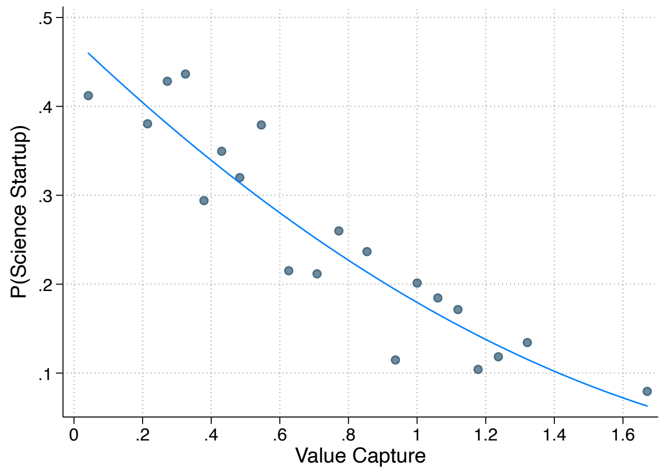


Figure 10: Share of science-based startups across the value capture distribution. This figure plots the share of startups that are science-based within bins of the value capture distribution. The share declines sharply with higher value capture: science startups are overrepresented in the lower end and underrepresented in the upper end of the distribution. A startup in the bottom quartile is 3.2 times more likely to be science-based than one in the top quartile, consistent with systematic differences in value retention at exit.

4 Empirical analysis: Science-based startups capture less, generate more value

I begin by documenting differences in value capture and creation between science-based and non-science-based startups. I employ OLS regressions with fixed effects on a sample of 5,823 startups acquired by U.S. publicly listed companies. The general specification is as follows:

$$y_i = \beta_0 + \beta_1 sci_i + \beta \sum \mathbf{X}_i + \theta_i + \xi i + \epsilon_i, \tag{7}$$

where the dependent variable y_i is either the startup's value capture, λ_s, or the startup's joint surplus, X_t, at the time of acquisition. The analysis is conducted at the startup level, i. $sci \in [0, 1]$ is an indicator for whether the startup i commercializes scientific innovations.[50] $\mathbf{X}$ is a vector of control variables, θ represents grouped industry-year fixed effects to account for dynamics across industries and years, and ξ denotes startup-level fixed effects for the country of headquarters, capturing technological, market, and regulatory differences, among other factors, based on the startup's location. The main results are reported in Table 5.

[50] The results are robust to using the discrete variable output by the LLM $sci \in [1, 5]$, which captures the likelihood that a startup is science-based, as well as to alternative classification thresholds for the indicator.

Table 5: OLS estimates of value capture and creation (joint surplus) for science-based startups. The table reports OLS regressions examining the relationship between a startup's science-based classification and its value capture—columns (1)–(3)—and (log) value creation—columns (4)–(6)—in acquisition. The specifications add controls for external financing (VC, PE), acquirer market capitalization (log), and include industry × year and country fixed effects. Standard errors are clustered at the industry-year and startup country level. The results show a consistent value capture penalty for science-based startups, even as they tend to generate higher total value.

	Value Capture			Value Creation		
	(1)	(2)	(3)	(4)	(5)	(6)
Science-based Startup	-0.147***	-0.150***	-0.143***	0.180***	0.074***	0.008
	(0.023)	(0.023)	(0.020)	(0.034)	(0.025)	(0.024)
VC, PE Investment (log)		0.007***	0.028***		0.269***	0.043***
		(0.003)	(0.003)		(0.016)	(0.008)
Acquirer market cap. (log)			-0.060***			0.729***
			(0.006)			(0.008)
Constant	0.610***	0.593***	1.861***	18.633***	17.984***	2.482***
	(0.003)	(0.007)	(0.123)	(0.005)	(0.039)	(0.156)
Industry X Year FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
N	5,823	5,823	5,823	5,823	5,823	5,823
R^2	0.057	0.058	0.101	0.095	0.145	0.774

Standard errors clustered at the Industry-Year and Country level

* p<.1, ** p<.05, *** p<.01

4.1 Value capture

Startups commercializing scientific innovations face significantly greater challenges in value capture. Table 5, column (1) shows that startups commercializing scientific innovations experience an 14.7 percentage-point reduction in value capture, from 61.0 percentage points to 46.3—a 24.1% decrease. In monetary terms, if the average science-based transaction ($796 million) had captured value at the same rate as non-science startups, average *net* surplus, or returns, for these ventures would have increased by $253 million.

Columns (2) and (3) refine the analysis by adding two key controls at both the startup and acquirer levels. Importantly, the negative coefficient on science based startup remains quite stable and highly significant. First, I control for the total investment received by the startup, a factor that plausibly influences value capture. The results demonstrate a significant relationship: the more investment a startup accumulates before the acquisition, the greater its value capture. However, this relationship is correlational and may reflect two opposing dynamics. On one hand, higher investment could de-risk the technology and market application, help find customers, or could be used to threat with independent commercialization, in all cases improving the startup's bargaining position. On the other, it might show that VCs direct greater funding toward startups with ex-ante favorable exit conditions and, thus, higher potential for value capture.[51]

[51] I examine the relationship between investment and value capture in greater detail in the Appendix. As a preview, in results not reported here, adding an interaction for *Science* × *Investment*, I find that while increased investment enhances value capture for both science and non-science startups, the rate of increase for non-science startups is

Second, I control for the acquirer's market capitalization, a proxy for firm size. The estimates show that larger acquirers capture a greater share of transaction rents, leaving less for the startup. This effect may stem from stronger complementary assets—such as distribution networks, manufacturing capacity, and prior related technological investments—that enable them to generate higher returns from the acquired technology, greater bargaining leverage in negotiations, and a correlation with the number of potential acquirers. Likewise, the coefficient on the science-based startup indicator remains stable, suggesting that other unobservables associated with commercializing scientific innovations continue to drive the effect. In the next section, I examine the underlying mechanisms and unpack these patterns in greater detail.

4.2 Value creation

Next, I turn to columns (4) to (6) (Table 5), where the dependent variable is the logged value created (joint surplus). The analysis parallels that for value capture. The results indicate that, although science-based startups in the sample capture less value, they actually generate a larger total value. Column (4) shows that startups commercializing scientific innovations generate significantly more value than their counterparts—approximately a 19.7% increase. Given an average value created of $1.11 billion for science-based startups, this corresponds to roughly $250 million more per transaction of value created than their counterparts.

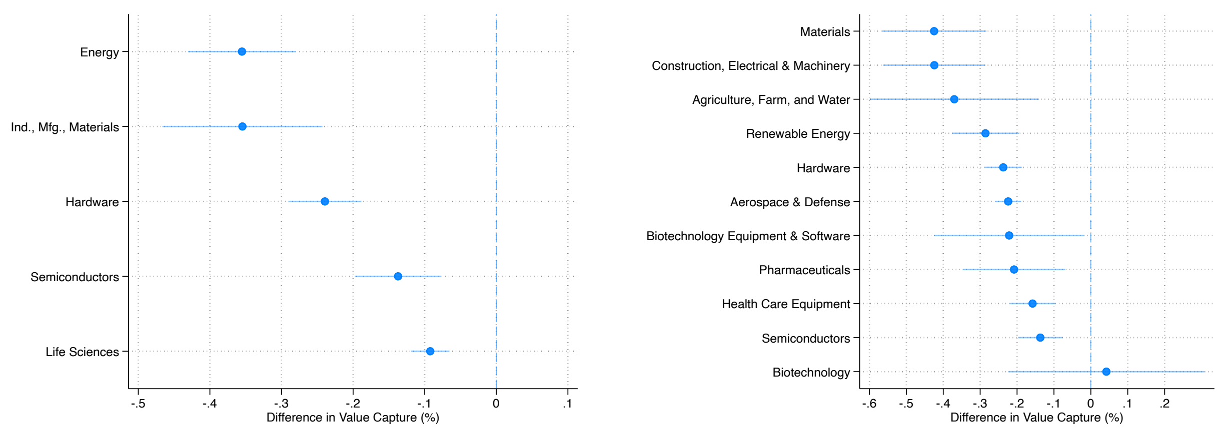
In columns (5) and (6), additional controls for total funds raised and the acquirer's size are introduced, which slightly attenuates the coefficient on Science but does not undermine the core finding. The investment coefficient remains significant, indicating that higher levels of external funding are still associated with greater total value creation. More importantly, even after accounting for the amount of investment received, science-based startups continue to generate substantially more joint surplus than their non-scientific counterparts. The influence of acquirer size further suggests that larger firms may shape the total value created, consistent with the theoretical framework and as I explore in further detail in the next section. Nevertheless, the primary conclusion, that science-based startups generate higher overall value, remains robust across specifications.

Consistent with the theory, these pattern helps explain why investment and entry into science-based ventures may remain rational in my sample of analysis, despite lower value capture for the startup itself. If the increase in total value creation is sufficiently large, it can offset weaker individual bargaining outcomes, making expected returns attractive when evaluated over a portfolio of investments. In other words, even when a smaller share of the surplus accrues to the startup, the absolute magnitude of returns may still justify entry and sustained capital flows into science-based entrepreneurship. This logic also aligns with observed patterns in venture capital, where investors often balance lower capture rates with the potential for outsized exits in high-value scientific innovations.

nearly 50% higher. This suggests that the returns to investment in terms of capture, are more efficient in non-science firms. As discussed, this is consistent with the theoretical framework and potentially due to lower commercialization barriers, faster scaling opportunities, or more direct market applications. In contrast, science-based startups face structural challenges that limit their ability to translate investment into proportional gains in value capture.

4.3 Industry Heterogeneity

The results reported so far mask important cross-industry heterogeneity. Some subsectors, such as Biotech, are dominated by science-oriented firms (where 94.5% of the startups are classified as science based), whereas others are more mixed (e.g., Energy, with 24.2% science and Industrials, Manufacturing & Materials with 31.4% science), and still others, like Consumer and Business Products & Services as well as Software and IT startups predominantly consist of non-science innovations. Consequently, the aggregate results may be attenuated or amplified by these differences. To explore how these differences play out across industries, I extend the baseline specification (equation 7) to include industry indicators and interact them with the science indicator. Figure 11, Panel A, plots the marginal effects (the change in expected value capture from being science-based, holding all other variables constant at means) at the industry level and Panel B at the subindustry level.[52]



(a) Marginal effect by industry.

(b) Marginal effect by subindustry.

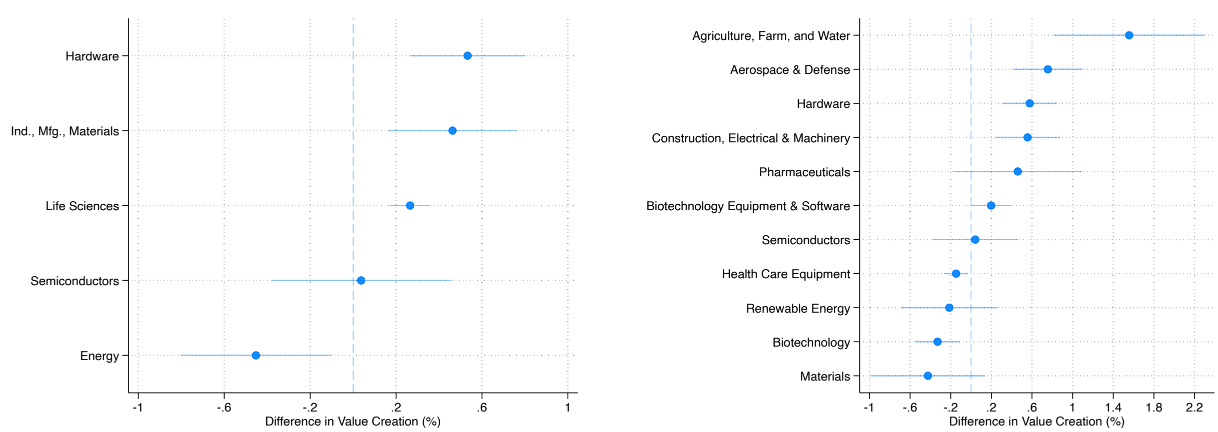
Figure 11: Comparison of how the marginal effect—the change in expected value capture associated with being science-based, holding all other variables constant at means—varies across industries (a) and subindustries (b). Confidence intervals are plotted at the 95% level. The figure shows that, across both industry and subindustry classifications, science-based startups consistently capture a smaller share of transaction rents than their non-science counterparts. The effect is generally negative and sizable in several sectors—notably energy, industrials, materials, biotechnology, and related subfields—indicating that the value capture penalty is not confined to a single domain but appears across a range of science-intensive industries.

The findings indicate that science-based startups in mixed-composition sectors tend to capture less value than their non-science counterparts. For instance, at the industry level, Energy exhibits the largest penalty, with a 35.5% difference in value capture for startups commercializing scientific innovations, followed by Industrials, Manufacturing & Materials (35.4%), Hardware (24.6%), and Semiconductors (13.7%), all of which are highly statistically significant. Notably, these industries align with those often cited as experiencing insufficient innovation activity (e.g., Lerner and Nanda, 2020). At the subindustry level, the patterns are similar, with Aerospace & Defense; Agriculture, Farm & Water; Construction, Electrical & Machinery; Hardware; Materials; and Renewable Energy all facing penalties ranging from 22.4% to 42.5% in value capture.

[52]The Appendix provides the results of the main specification.

Interestingly, the Biotechnology sector exhibits a reversed pattern. The estimated penalties are either negligible or positive, meaning that scientific innovations capture more value than their non-science counterparts. This finding aligns with the composition of the Biotechnology sector, where the majority of startups are science-based, and with the idea of more efficient markets for technology. Likewise, one could argue that, in these industries, ex-ante contracting is more feasible due to the relatively well-defined pathways from scientific discovery to commercialization. Unlike other sectors, Biotech often benefits from deep, established markets with clear demand for innovative treatments and therapies. Furthermore, intellectual property has proven to be more effective on these innovations, and the nature of Biotech innovations often allows for early-stage partnerships, licensing agreements, or milestone-based financing, enabling startups to negotiate terms that reflect the potential value of their research (Arora et al., 2022).

Figure 12 reports the marginal effects of being science-based on value creation, holding all other variables constant at their means. At the industry level, science-based startups generate significantly more total value in Hardware, Industrials, Manufacturing & Materials, and the Life Sciences. In each of these sectors, the estimated coefficient is positive and statistically significant, indicating that scientific ventures create more surplus than their non-science counterparts. By contrast, the coefficients for Energy and Semiconductors are small and statistically insignificant. At the subindustry level, the picture is more heterogeneous. Science-based startups exhibit large and significant value creation advantages in Pharmaceuticals, Biotechnology Equipment & Software, and Construction, Electrical & Machinery, while other subfields such as Renewable Energy, Biotechnology, and Materials show weaker or statistically insignificant differences.



(a) Marginal effect by industry.

(b) Marginal effect by subindustry.

Figure 12: Comparison of marginal effects of science on value creation by industry and subindustry, with 95% confidence intervals, holding all other variables constant at their means. Science-based startups show higher value creation in several sectors—most notably in hardware, industrials, manufacturing, and pharmaceuticals—while differences are statistically insignificant in others, indicating that the value creation premium is concentrated in specific science-intensive industries.

4.4 Heterogeneous Effects of Investment and Time to Exit

A natural next step is to examine how capital investment and time to exit, two central inputs at the aggregate level, shape value creation and capture. As reported in Table 5, venture capital investment is strongly and positively associated with both outcomes. I omit time to exit from the baseline specification due to its high collinearity with investment, though both variables broadly capture the intensity and duration of development prior to acquisition. Furthermore, building on the heterogeneity patterns discussed above, it is instructive to examine these effects across industries, providing initial evidence on whether investment and time operate through common channels across industries and startup types, or whether their roles differ in structurally distinct ways. This analysis also serves as a preliminary test of the mechanisms developed in the following section.[53]

The conceptual framework introduced earlier implies that in software and consumer-oriented industries, investment and time are often deployed to develop commercialization capabilities, de-risking demand, building distribution networks, acquiring customers, and scaling operations. These activities not only indeed scale the startup and drive downstream value creation, but also strengthen the startup's outside options in an eventual transfer, reducing reliance on any single acquirer and improving its bargaining position. In such settings, one would expect both value creation and capture to increase with investment and time to exit.

In contrast, science-based ventures typically require more substantial investment in R&D and greater effort to de-risk the underlying technology, and thus it is reasonable to assume that capital is primarily allocated toward technological advancement rather than market development. At the same time, if these ventures face structural frictions to scaling, such as limited access to contract manufacturing and distribution, then capital is unlikely to build or expand the external commercialization pathways they depend on. Likewise, capital and time are even less likely to affect the structure of downstream acquisition markets. As a result, in these settings, investment and time to exit may increase the total value of the innovation but do little to improve the startup's outside option and bargaining position, resulting in limited gains in value capture.

To examine these relationships, I extend the specification from Table 5. The specification I estimate is as follows:

$$y_i = \beta_0 + \beta_1 \log(Investment_i) + \beta_2 Industry_i^{'} + \beta_3 \log(Investment_i) \times Industry_i^{'} + \gamma_i + \xi_i + \epsilon_i, \quad (8)$$

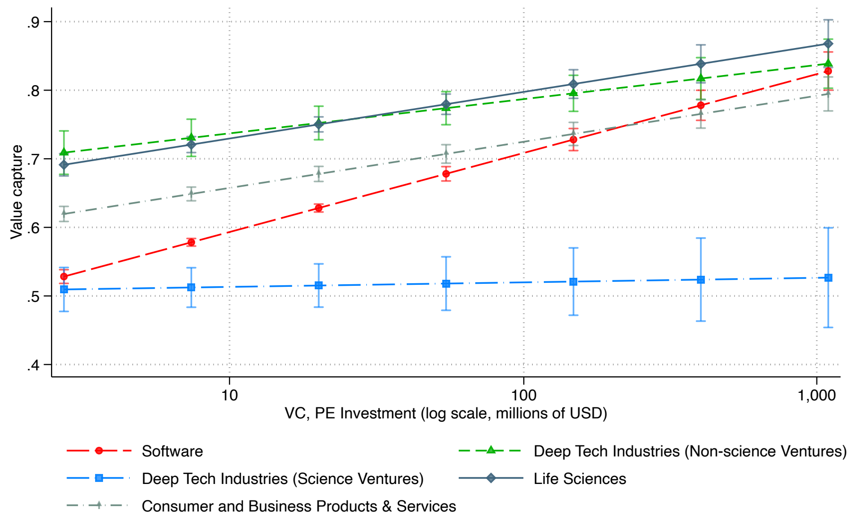
[53] The estimated effect of venture capital investment on value creation and capture should not be interpreted as causal, of course. Larger investment rounds may reflect unobserved startup quality, higher commercial potential, or stronger backing by top-tier investors—all of which could independently affect both outcomes. In this sense, the observed correlation likely captures a combination of treatment and selection effects. That said, the heterogeneity in the returns to investment across sectors, particularly the weaker relationship between investment and value capture in science-based industries, is less easily explained by selection alone. If selection were the dominant driver, one might expect stronger investors to concentrate in industries with better ex-ante commercialization prospects and, consequently, a tighter link between investment and capture. The absence of such a pattern in science-based sectors is consistent with the view that structural constraints, rather than investor or venture quality, limit the ability to translate investment into bargaining power.

where the dependent variable y_i is either the startup's value capture (λ_s) or the joint surplus at exit (X_t). *Investment* denotes the total private capital raised by the startup prior to acquisition, and *Industry'* indicates the startup's industry classification, as further defined below. Since industry enters directly into the specification, I include only year fixed effects (γ) to account for temporal dynamics and country-of-headquarters fixed effects (ξ) to control for geographic heterogeneity.

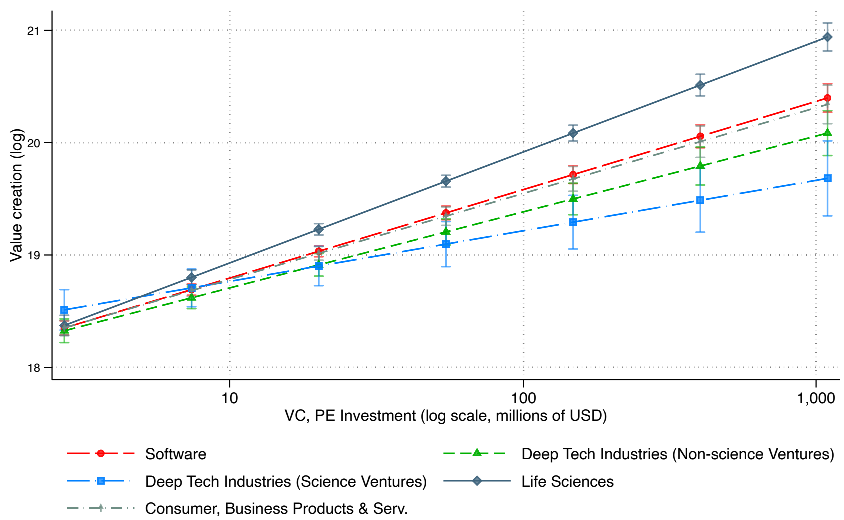
For ease of interpretation, in this analysis I aggregate startups into broader industry groups, though these aggregations simply reflect patterns observed at the level of individual industries; I denote these aggregated industries *Industry'*. Software, Consumer/Business Products and Services, and Life Sciences remain as distinct categories, retaining their original industry classifications. By contrast, I combine startups in Energy, Hardware, Industrials, Manufacturing and Materials, and Semiconductors into a single group. For brevity, I refer to this set of industries as Deep Tech.

Within this aggregated Deep Tech group, the relevant distinction for the analysis is not industry alone, but whether the startup is developing science-based innovations. As posed earlier, these industries contain a heterogeneous mix of ventures, ranging from service-oriented or engineering-driven firms to startups advancing novel scientific discoveries. For example, a digital services platform in the energy sector faces fundamentally different technical and commercialization challenges than a carbon capture venture based on advanced materials science. To account for this variation, I further divide this Deep Tech group into two subgroups: those developing science-based innovations and those that do not. This classification allows for a more precise empirical test of the theoretical framework, which predicts systematic differences in value creation and capture between science-based and non-science-based ventures.

The results are presented in Figure 13. For brevity and clarity of exposition, I plot the estimated coefficients and their 90% confidence intervals. Since the patterns are similar when using time to exit, I report in the manuscript only the results based on accumulated investment, as it more directly captures the scale and intensity of development efforts prior to acquisition. Panel (a) shows that, consistent with the framework, the returns to capital in terms of value capture are strongest for ventures in Software, where the share of value captured rises from approximately 0.55 to 0.9 as cumulative investment increases. The estimated slope is both statistically significant and economically meaningful, suggesting that investment translates directly into stronger bargaining positions and higher private returns driven by gains in capture, holding creation constant. Life Sciences and Consumer/Business Products & Services exhibit similarly positive relationships, though with somewhat more moderate slopes compared to Software. In Energy, Hardware, Industrials, Manufacturing and Materials, and Semiconductors (labeled as Deep Tech industries), ventures that do not commercialize scientific innovations still exhibit positive returns to capital in terms of value capture, with a slightly smaller slope than in other industries, although the differences are not statistically significant. Notably, the most revealing pattern emerges among ventures in these industries that do commercialize science-based innovations (Deep Tech Industries, Science Ventures). As shown, their average value capture is the lowest among all groups and it appears largely insensitive to the amount of capital deployed. Across four orders of magnitude in cumulative investment, the share of value captured remains roughly flat, around 0.50. This pattern is consistent with the theoretical prediction that, in science-based ventures, investment is primarily directed toward technical validation rather than commercialization infrastructure, and that key constraints—such as thin buyer markets and limited access to complementary assets—are not mitigated with capital.



(a) Value capture *vs.* accumulated investment (VC, PE) at exit



(b) Value creation (log) *vs.* accumulated investment (VC, PE) at exit

Figure 13: This figure plots the relationship between cumulative VC/PE investment and startup value capture and value creation, across five industry groups. Confidence intervals represent 90% bounds. The Deep Tech group comprises ventures in Energy, Hardware, Industrials, Manufacturing and Materials, and Semiconductors, and is further divided into two subgroups: ventures that commercialize scientific innovations and those that do not. The relationship between investment and value capture varies sharply across industries. In Software, capture increases with investment, consistent with capital building stronger outside options and bargaining power. In contrast, capture remains flat for science-based Deep Tech ventures, suggesting that structural commercialization constraints persist even as capital is deployed and technologies developed. All sectors exhibit a positive association between investment and value creation, though the underlying mechanisms may differ—ranging from market de-risking in Software and Consumer sectors to technological development in science-based ventures.

In terms of value creation, all industries exhibit a positive relationship with cumulative investment, indicating that greater capital deployment is indeed associated with the development of more valuable innovations. However, the underlying mechanisms may differ. In software and consumer-facing industries, returns to investment may primarily reflect the resolution of market risk, such as product–market fit or customer acquisition, while in science-based ventures, investment likely contributes more directly to reducing technological uncertainty. Importantly, while most industry slopes appear similar, only Life Sciences exhibits a statistically distinct trajectory, with a steeper and more precise relationship between investment and value creation. This suggests that capital plays an especially central role in driving innovation outcomes in this industry, potentially due to underlying technological complexity, market-related issues, or merely selection effects. The specific mechanism remains unclear and warrants further investigation.

5 Mechanism: Exit Conditions, Value Capture, and Value Creation

In this section, I examine the central mechanism described in the theoretical framework that helps explain the value creation and capture patterns outlined above: the structural conditions of the startup exit environment. These include (i) acquisition market structure—the number and size distribution of potential acquirers—and (ii) the startup's outside options, defined as its capacity to scale commercialization independently. The results are descriptive rather than causal, but they are consistent with the theoretical framework developed earlier and align with its key predictions.

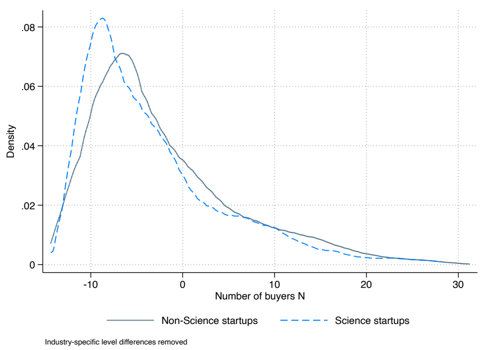
5.1 Market structure of incumbent acquirers

I begin by documenting stylized facts on how science-based and non-science-based startups differ in the characteristics of their pool of potential acquirers. As detailed in Section 3.5, I estimate the pool of potential acquirers with the acquirer's product-market competition: the count of rivals in the Hoberg and Phillips (2016, 2025)'s text-similarity network.[54] After identifying the set of potential acquirers, I characterize the size distribution and other attributes of its member firms. Furthermore, for most of the empirical analysis, I group the number of potential acquirers into deciles instead of using raw or log values. The distribution of potential acquirers is right-skewed but does not follow a log-normal, making logs problematic and raw counts unduly influenced by outliers. Decile indicators temper the leverage of extreme values and do not impose an elasticity interpretation. They also reduce sensitivity to the exact similarity cut-off used in the Hoberg and Phillips dataset that defines a potential acquirer, reducing measurement error.

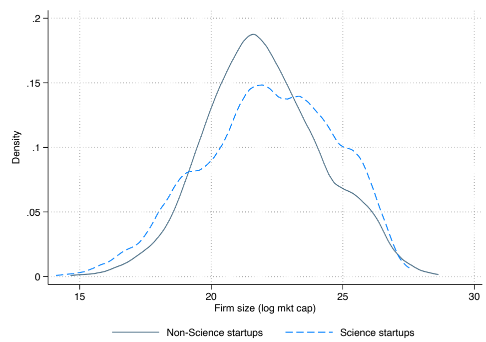
Figure 14 illustrates acquisition-market characteristics by startup type. Panel (a) depicts the distribution of the number of potential acquirers. To account for industry-level differences (e.g.,

[54]The sample size in the following analyses is slightly smaller than in previous sections due to missing data in the matching process with this dataset.

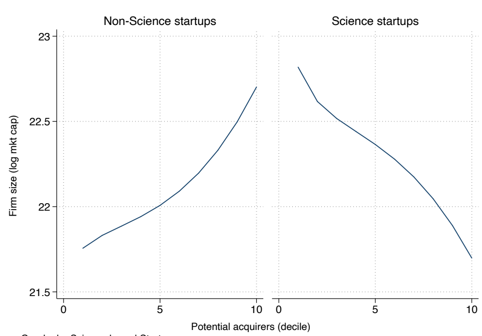
pharmaceuticals exhibit more acquirers overall than renewables), I parse out industry fixed effects.[55] The distribution for non-science-based startups lies to the right of that for science ventures, indicating systematically thicker acquisition markets for non-science ventures. On average, science-based ventures face 9.1% fewer acquirers, and this difference varies substantially across industries. For example, science-based startups in Hardware as well as in Industrials, Manufacturing, and Materials face more than 30% fewer potential acquirers than their counterparts. Panel (b) shows the log size of those potential acquirers, measured as market capitalization at the deal announcement date. Science startups acquirers tend to be larger firms—53% larger on average—, whereas non-science startups encounter relatively smaller incumbents. Together, the two panels illustrate that science-based ventures face a narrower acquirer pool, but the firms in the pool tend to be substantially larger.



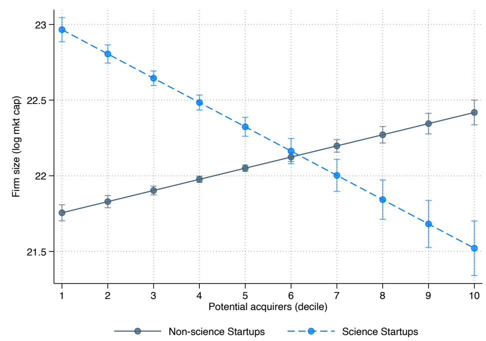
(a) Number of potential acquirers: Density



(b) Firm size: Density



(c) Potential acquirers *v.* Firm size: Non-parametric



(d) Potential acquirers *v.* Firm size: Parametric

Figure 14: Acquirer-market structure by startup type. Panel (a) plots the industry-demeaned kernel density of the number of potential acquirers; panel (b) shows the corresponding density of log market capitalization of those acquirers. Panels (c) and (d) relate acquirer size to market thickness: panel (c) plots a non-parametric estimation of firm size (log market capitalization) across deciles of the rival-count distribution separately for science and non-science startups, while panel (d) plots the marginal effect of potential acquirers decile on average firm size based on OLS estimation, with 95 percent confidence bands. Science-based startups face more numerous acquirers on average, yet additional rivals come from progressively smaller firms, reversing the positive size–number correlation observed for non-science ventures.

Panels (c) and (d) deepen in the market structure by relating acquirer size to market thick-

[55] For each observation, I regress the variable of interest on industry dummies and use the residuals. The horizontal axis therefore measures deviations from the industry mean (zero).

ness. Panel (c) plots a non-parametric estimation of average firm size as a function of number of acquirers. For non-science startups (left sub-panel), average acquirer size rises with the number of rivals, indicating that thicker markets also involve larger incumbents. Conversely, for science-based startups (right sub-panel) the slope is reversed: thin markets are populated by very large incumbents, whereas additional rivals come from progressively smaller firms. Panel (d) confirms these relationships with a parametric estimation, plotting estimates with 95% confidence intervals resulting from OLS regressions that interact the number of acquirers with the type of startup.

These empirical patterns map directly to the stylized facts outlined in the theoretical framework. Consistent with the Stylized Fact *1*, science-based startups operate in more concentrated acquisition markets, with systematically fewer potential acquirers. Moreover, the negative relationship between the number and size of potential acquirers for science-based ventures (Stylized Fact *2.1*) contrasts with the positive relationship observed for non-science-based ventures (Stylized Fact *2.2*).

To formally examine the relationship between incumbents' market structure, value capture, and value creation, I estimate OLS regressions. I start by examining the effect of market structure on capture and creation. To test for heterogeneous effects, the analysis includes interactions with an indicator for science-based startups, allowing the response of value capture to market structure to differ between science and non-science ventures. The main specification is:

$$y_i \;=\; \beta_0 + \beta_1\, sci_i + \beta_2\, acquirer\_decile_i + \beta_3\, acquirer\_decile_i \times sci_i + \theta_i + \xi_i + \varepsilon_i, \tag{1}$$

where y_i is the main variable of interest (startup's value capture share, λ, or value created, V_t, in startup i); sci_i is an indicator for science-based startups; $acquirer\_decile_i$ indexes the decile of potential acquirer counts; θi represents industry–year fixed effects; and ξ_i represents startup country fixed effects. The interaction term β_3 tests whether acquirer-pool thickness influences science and non-science startups differently, while the fixed effects absorb common shocks within technology–market–year cells and acquirer-specific heterogeneity. Standard errors are clustered at the Industry-year and country level. Table 6 presents the results for value capture.

In column (1), the dependent variable is the number of potential acquirers (decile). As shown in the facts above, the coefficient on the science-based indicator is negative and statistically significant, indicating that science-based startups face fewer potential acquirers relative to non-science startups. On average, science startups have 9.11% fewer acquirers, with the finding supporting the idea that science startups often commercialize highly specialized or nascent technologies and thus face a more limited set of acquirers that can develop the technologies further, with the necessary capabilities. Columns (2) through (5) explore how science and the number of potential acquirers relate to value capture. Across all specifications, the coefficient on science-based is negative and highly significant, aligned in magnitude with the results reported in previous sections. The coefficient on the number of potential acquirers is close to zero and not significant in column (3), suggesting that broader competition among acquirers does not enhance the startup's bargaining power. However, this coefficient masks heterogeneity. Importantly, column (5) includes an interaction term between Science and Potential acquirers, which is positive and significant. An increase in the number of potential acquirers from the mean to one standard deviation above the mean is associated with a 25.70% increase in value capture for science-based startups, compared to a null increase for non-science startups. Overall, these results support Propositions *2.1* and *2.2*.

Table 6: Value capture: Science-based startups and market structure. This table presents OLS regressions examining how acquisition market structure relates to startup type and value capture. Column (1) uses the number of potential acquirers (in deciles) as the dependent variable and shows that science-based startups face significantly thinner acquirer markets. Columns (2)–(5) use the startup's share of total value created as the dependent variable. Across all specifications, science-based startups systematically capture less value at exit, with point estimates ranging from 15 to 22 percentage points. The number of potential acquirers alone has little effect—columns (3)–(4)—, but column (5) shows that it significantly increases value capture for science-based startups through a positive interaction term. This pattern supports the theoretical prediction that acquirer competition disproportionately benefits science-based startups. All models include industry-year and country fixed effects. Standard errors are clustered at the industry-year and country level.

	Potential acquirers	Value capture			
	(1)	(2)	(3)	(4)	(5)
Science-based Startup	-0.190***	-0.153***		-0.152***	-0.215***
	(0.062)	(0.028)		(0.028)	(0.045)
Potential acquirers (decile)			0.003	0.002	0.000
			(0.002)	(0.002)	(0.002)
Science × Potential acquirers					0.012***
					(0.004)
Constant	5.281***	0.554***	0.511***	0.542***	0.554***
	(0.008)	(0.005)	(0.009)	(0.013)	(0.015)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
N	4,958	4,958	4,958	4,958	4,958
R^2	0.106	0.079	0.070	0.079	0.080

Standard errors clustered at the Industry-Year and Country level
* p<.1, ** p<.05, *** p<.01

Table 7 explores these relationships in terms of value creation. Across all specifications, science-based startups are associated with significantly higher value creation. The number of potential acquirers also has a consistently positive and significant effect on value creation. However, the interaction term in column (4) reveals a relevant asymmetry: the positive effect of acquirer pool size is significantly dampened for science-based startups, as shown by the negative interaction coefficient.

These patterns map onto the theoretical predictions from the conceptual framework. In the model, science-based startups typically face a more concentrated acquirer landscape, where the largest incumbents are most capable of realizing the innovation's full value. When the number of bidders increases, average acquirer size falls more sharply for science-based startups, decreasing the potential value created, even if competition nominally increases. In contrast, non-science startups increase acquirer's surplus as the pool expands, since their innovations are less specialized and more easily absorbed across a wide range of acquirers. Overall, these results support Propositions *1.1* and *1.2*.

Figure 15 illustrates these results. Panels (a) and (c) show that value capture rises strongly with

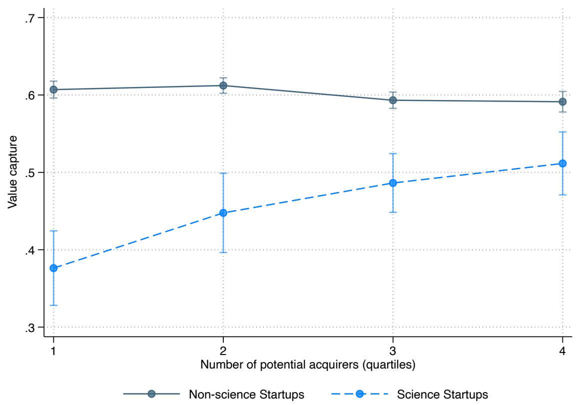
Table 7: Value creation: Science-based startups and acquisition market structure. This table presents OLS regressions where the dependent variable is the log of value creation. All models include industry-year and country fixed effects, and standard errors are clustered at the industry-year and country level. Column (1) shows that science-based startups generate, on average, significantly more value at exit than non-science startups. Column (2) adds the number of potential acquirers (decile), which has a positive and significant association with value creation. Column (4) includes the interaction term between Science and Potential acquirers, which is negative and significant, indicating that the marginal effect of expanding the acquirer pool on value creation is lower for science-based startups.

	Value creation (log)			
	(1)	(2)	(3)	(4)
Science-based Startup	0.124***		0.131***	0.979***
	(0.027)		(0.023)	(0.096)
Potential acquirers (decile)		0.036***	0.036***	0.067***
		(0.007)	(0.007)	(0.010)
Science × Potential acquirers				-0.159***
				(0.019)
Constant	18.665***	18.503***	18.476***	18.311***
	(0.005)	(0.026)	(0.026)	(0.048)
Industry × Year FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
N	4,958	4,958	4,958	4,958
R^2	0.103	0.105	0.106	0.115

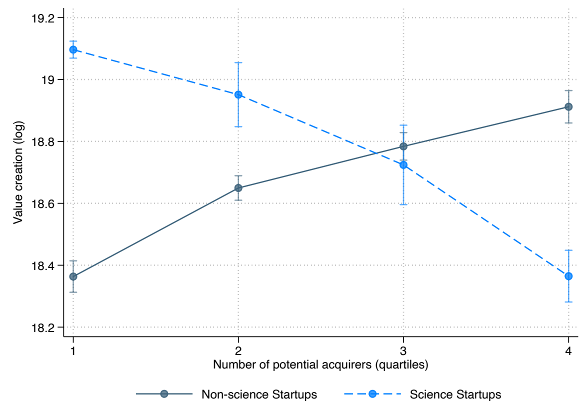
Standard errors clustered at the Industry-Year and Country level
* p<.1, ** p<.05, *** p<.01

acquirer pool size for science-based startups, while remaining flat for non-science ones. In contrast, Panels (b) and (d) reveal that value creation increases with the number of acquirers for non-science startups, but declines for science-based ventures. These opposing slopes are consistent across both non-parametric (top row) and parametric (bottom row) approaches. The non-parametric plots provide flexible visual confirmation without functional form assumptions, while the parametric specifications align with the regression estimates presented in Tables 6 and 7.

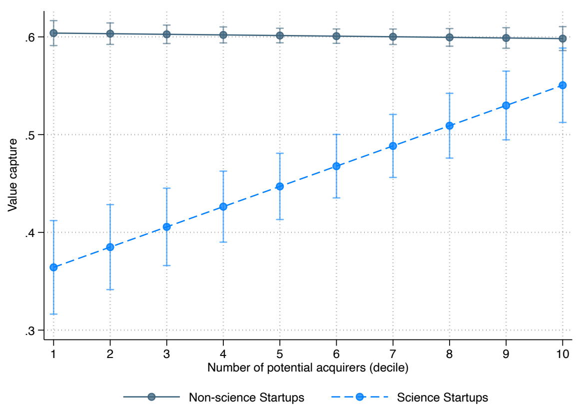
Importantly, the creation patterns reflect the structural mechanism described in the conceptual framework: in science-based markets, expanding the number of acquirers may bring in smaller, less capable incumbents, reducing the total value that can be realized from the innovation. Thus, while the number of potential acquirers boosts bargaining power for science-based startups, it simultaneously reduces average acquirer surplus and, thus, joint surplus, generating a divergence between value capture and value creation as markets thicken.



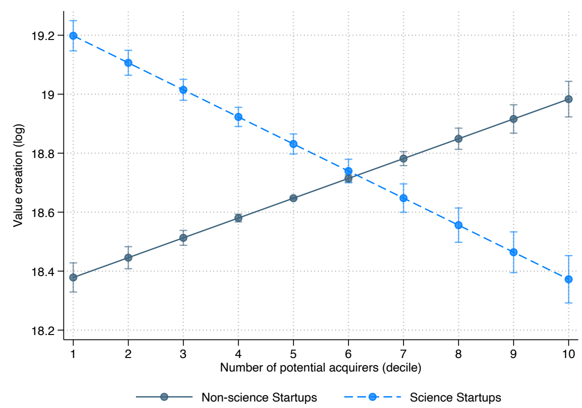
(a) Value capture *v.* Potential acquirers: Non-parametric



(b) Value creation *v.* Potential acquirers: Non-parametric



(c) Value capture *v.* Potential acquirers: Parametric, Table 6, (6)



(d) Value creation *v.* Potential acquirers: Parametric, Table 7, (5)

Figure 15: Value creation and capture by number of potential acquirers, science vs. non-science startups. The figure plots value creation and value capture (startup share of surplus) against the number of potential acquirers, separately for science-based and non-science startups. Panels (a) and (b) show non-parametric estimates by quartile bins; Panels (c) and (d) show parametric estimates using deciles. Value capture increases with the number of potential acquirers for science-based startups but remains flat for non-science ones. Value creation rises with the number of potential acquirers for non-science startups but declines for science-based ones, consistent with a drop in average acquirer size as acquirer sets expand.

5.2 Startup's Outside Options: Independent Scaling

I now turn to analyzing how outside options mediate the results, showing that they account for the remaining patterns predicted by the framework. Specifically, I examine the relationship between a startup's ability to commercialize independently, proxied by revenue generation at the time of acquisition, and its value capture and creation at exit. In line with the theoretical framework, Table 8, column (1), shows that science startups are significantly less likely to generate revenue. On average, the revenue generated at exit is reduced by more than one decile, or a 40% decrease. This is consistent with the idea that these startups face steeper barriers to independent scaling (Stylized Fact *3*).

Table 8: Value capture: Role of revenue and acquisition market structure. This table analyzes how value capture at exit varies with startups' commercialization capacity (proxied by revenue deciles) and acquisition market structure. Column (1) shows that science-based startups are significantly less likely to generate revenue. Columns (2)–(7) examine value capture. The negative effect of being science-based vanishes once revenue is included, column (3) onward, suggesting that weaker outside options fully explain the capture gap. The number of potential acquirers has a positive and significant effect, particularly for science-based ventures, columns (5)–(6), consistent with greater sensitivity to market competition when fallback options are weak.

	Revenue	Value Capture				
	(1)	(2)	(3)	(4)	(5)	(6)
Science-based Startup	-1.069***	-0.023	0.008	-0.057	-0.132	-0.047
	(0.093)	(0.024)	(0.052)	(0.036)	(0.091)	(0.081)
Revenue (decile)		0.048***	0.049***	0.049***	0.049***	0.059***
		(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Science × Revenue (decile)			-0.006		-0.001	-0.004
			(0.006)		(0.007)	(0.006)
Potential acquirers (decile)				0.009***	0.006**	0.010***
				(0.002)	(0.003)	(0.002)
Science × Potential acquirers					0.017**	0.005
					(0.008)	(0.007)
Acquirer size (log)						-0.045***
						(0.007)
Constant	5.637***	0.407***	0.399***	0.348***	0.362***	1.284***
	(0.013)	(0.004)	(0.015)	(0.016)	(0.023)	(0.168)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
N	3,539	3,539	3,539	3,539	3,539	3,539
R^2	0.215	0.180	0.180	0.193	0.195	0.236

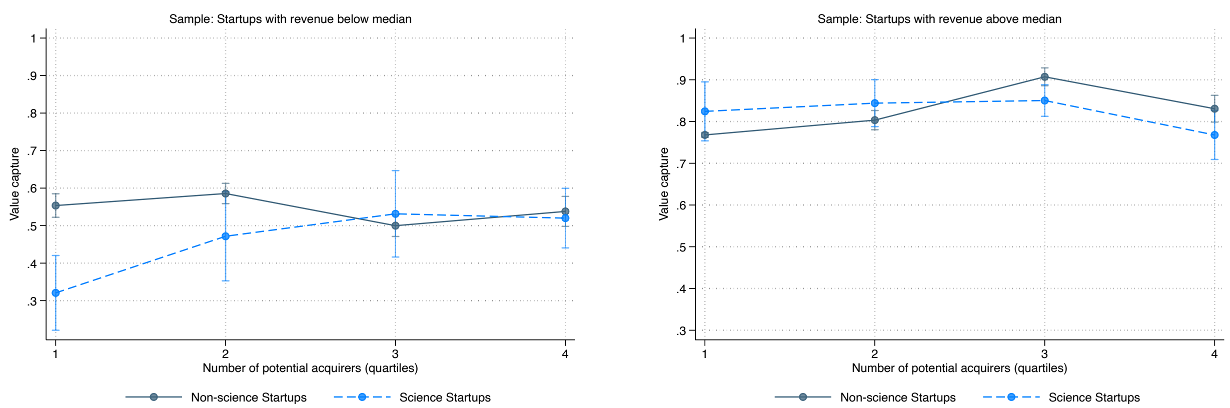
Standard errors clustered at the Industry-Year and Country level

* p<.1, ** p<.05, *** p<.01

The rest of Table 8 shows that revenue is a strong predictor of value capture. Starting in column (2), the coefficient on Science becomes statistically insignificant and remains so across all specifications, indicating that once I control for revenue generated, the raw capture gap between science and non-science startups disappears. This suggests that the differential in surplus extraction stems from systematic differences in outside options. Interestingly, the interaction between Science and Revenue decile is never significant, implying that revenue matters equally for both types of startups:

once a science-based firm reaches revenue generation, its bargaining disadvantage vanishes. Even after controlling for revenue, the number of potential acquirers continues to significantly affect value capture, particularly for science-based startups. The interaction term between Science and Potential acquirers is positive and significant in columns (4) to (6), consistent with the auction-theoretic prediction that increased bidder competition shifts surplus toward the seller.[56]

Figure 16 illustrates how the relationship between value capture and the number of potential acquirers differs based on a startup's ability to commercialize independently. I plot non-parametric conditional means by quartile bins, allowing the raw data to speak without imposing functional form assumptions.



(a) Value capture *v.* Potential acquirers: Startups with revenue below the median

(b) Value capture *v.* Potential acquirers: Startups with revenue above the median

Figure 16: Value capture by number of potential acquirers, split by revenue. The figure plots non-parametric conditional means of value capture across quartiles of potential acquirers, separately for science-based and non-science startups. Panel (a) shows startups with revenue below the median; Panel (b) shows those above. Among low-revenue startups, science-based ventures capture significantly less value, but capture rises sharply with the number of potential acquirers, consistent with stronger dependence on acquirer competition to offset weak outside options. Among high-revenue startups, capture levels are high and similar across groups, with little sensitivity to market structure. These patterns support the model's prediction that reservation values drive heterogeneity in rent-sharing outcomes.

Panel (a) shows that among startups with below-median revenue, science-based ventures capture significantly less value than their non-science counterparts, but their capture increases sharply with the number of potential acquirers, consistent with the auction model prediction that market competition helps overcome weak outside options. The flatter line for non-science ventures may

[56] It is worth noting that revenue is a noisy proxy for independent scaling. For instance, a firm may report revenue at exit that stems from a single customer—possibly even the acquirer—thus overstating its true ability to scale independently. Moreover, some startups may engage in nominal revenue—generating activities (e.g., pilot projects or non-recurring engineering contracts) that are not indicative of a sustainable business model. Others may intentionally delay commercialization in favor of technology development. These issues introduce both measurement error and potential endogeneity, which bias the estimated relationship between revenue and value capture. Likewise, revenue implicitly bundles multiple organizational capabilities. It not only reflects a startup's ability to produce a sellable good or service—often requiring significant manufacturing or regulatory readiness—but also its ability to access downstream markets through distribution, marketing, and sales infrastructure. This distinction is particularly salient in science-based ventures. A startup may be able to manufacture but still unable to generate revenue if it lacks access to distribution channels or faces regulatory or customer adoption hurdles.

reflect that, even with low actual revenue, they possess stronger credible threats, consistent with the results linking capture to accumulated investment and time to exit. Panel (b) shows that among high-revenue startups, value capture is uniformly higher for both groups and largely unresponsive to acquirer count, suggesting that once a startup has credible outside options, bargaining power equalizes and acquirer competition plays a lesser role.

Finally, Table 9 reports regressions of value creation on startup type, revenue, and acquisition-market characteristics. Consistent with earlier results, science-based startups are associated with significantly higher value creation across all specifications, even after controlling for revenue and market structure. Revenue is itself a strong predictor of value creation, as exposed in the conceptual framework's view that commercial traction reduces market risk and unlocks value. Importantly, revenue does not eliminate the science-based coefficient, which remains positive and significant through column (5). In column (6), controlling for acquirer size—capturing the potential market reach of the acquirer—renders the science coefficient insignificant, while revenue remains highly significant. This pattern is consistent with the model's prediction that commercial outcomes, rather than scientific novelty per se, explain value creation: specifically, (i) the startup's ability to scale independently (proxied by revenue) and (ii) the size of potential acquirers (proxying market reach). In the fully specified model (column 6), all interaction terms are insignificant, and the coefficients shrink markedly relative to earlier specifications, suggesting that once these commercial channels are accounted for, the residual effect of being science-based is negligible.

Table 9: Value creation: Role of revenue and acquisition market structure. This table examines how value creation varies across startups, focusing on science-based ventures, commercialization outcomes, and acquirer characteristics. Science-based startups generate higher value across most specifications, but this effect becomes insignificant once acquirer size is included (column 6). Revenue is a strong and consistent predictor of value creation, and acquirer size further explains variation, suggesting that commercial traction and acquirer market reach, rather than scientific novelty alone, drive value creation.

	Value Creation					
	(1)	(2)	(3)	(4)	(5)	(6)
Revenue (decile)	0.301***	0.308***	0.327***	0.315***	0.324***	0.180***
	(0.023)	(0.023)	(0.031)	(0.022)	(0.027)	(0.008)
Science-based Startup		0.659***	1.066***	0.558***	1.641***	0.340
		(0.084)	(0.255)	(0.062)	(0.348)	(0.219)
Science × Revenue			-0.081**		-0.051*	-0.010
			(0.035)		(0.030)	(0.015)
Potential acquirers (decile)				0.043***	0.077***	0.023***
				(0.005)	(0.006)	(0.008)
Science × Potential acquirers					-0.170***	0.005
					(0.036)	(0.022)
Acquirer size (log)						0.678***
						(0.005)
Constant	17.439***	17.273***	17.164***	17.048***	16.828***	2.780***
	(0.117)	(0.132)	(0.180)	(0.115)	(0.175)	(0.059)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
N	3,539	3,539	3,539	3,539	3,539	3,539
R^2	0.288	0.295	0.297	0.300	0.310	0.798

Standard errors clustered at the Industry-Year and Subregion level

* p<.1, ** p<.05, *** p<.01

6 Conclusion

Science-based innovations hold enormous promise for addressing critical, social challenges in sectors such as advanced materials, agriculture, energy, and manufacturing, but entrepreneurial activity in these areas remains limited. Much of the existing discussion has emphasized the challenges of value creation for science-based ventures: high financing costs, long development timelines, and uncertain or weak demand. This paper highlights a complementary and underexplored problem, value capture, that can distort incentives even when scientific startups generate substantial economic and social value. This problem is rooted in structural features of the current innovation ecosystem. In many science-intensive sectors, commercialization follows a sequential process in which early-stage innovators develop and validate the technology, but large incumbent firms—specialized in downstream manufacturing and distribution—ultimately bring it to market (Arora et al., 2018; Gans and Stern, 2000; Marx et al., 2014; Teece, 1986). While this division of labor exploits the comparative advantages of each actor, it also creates a transfer stage where the first set of innovators may hold weak bargaining positions, enabling incumbents to appropriate a disproportionate share of the returns from innovation. As a result, the structure of these relationships can disincentivize upstream innovation, shaping not only the overall rate of innovative activity but also its direction toward domains with stronger prospects for value capture (Arora et al., 2024b; Grossman and Hart, 1986; Scotchmer, 2004).

To study this problem, I develop a novel empirical methodology that measures the joint surplus an innovation creates and how that surplus is divided between the parties involved in its transfer. Using this approach, I find that science-based startups acquired by publicly listed incumbents systematically capture less of the total surplus generated at exit than comparable non-science ventures, despite creating significantly more total value. This lower capture ultimately implies that returns are suppressed not only because of a lack of potential, but because a greater share of value is extracted by acquirers. The resulting penalty—non-negligible at roughly 24%—helps explain why funding may be disproportionately allocated to non-science counterparts, alongside more commonly cited explanations such as risk, capital intensity, and time to market.

These findings are consistent with further empirical evidence and a simple theoretical framework in which two features of the exit environment—thin acquisition markets and weak outside options—erode bargaining power for science-based startups. In many science-intensive sectors, only a small number of large incumbents possess the downstream assets needed for commercialization, while independent scaling is costly and rarely feasible. This combination leaves science-based startups more dependent on a narrow acquirer set, enabling incumbents to extract a larger share of the value. At the same time, differences in market structure also shape the total joint surplus (or value) created by these startups. In non-science sectors, more fragmented acquirer markets tend to attract larger marginal acquirers, raising total surplus, whereas in science-based sectors, expanding the acquirer pool often brings in smaller, less capable firms, lowering the realized value from commercialization. Concentrated markets thus have a dual effect: large incumbents can generate greater total surplus from an innovation, but they also shift bargaining power away from startups, enabling acquirers to capture most of the gains from innovation.

Results are not without limitations. I outline four principal limitations below, while acknowledging that additional challenges may persist beyond those discussed. First, the sample is selective by design, restricted to startups acquired by publicly listed U.S. firms. This focus excludes other forms of exits—such as acquisitions by private firms or IPOs—and may bias results if the dynamics of value capture differ systematically across these pathways. Likewise, a substantial amount of deals with missing transaction prices are excluded from the analysis. If acquirers are more likely to disclose transaction values strategically—e.g., for deals perceived as more favorable—this could introduce selection bias. Second, I estimate acquirer surplus using stock price reactions. I adopt advanced techniques to isolate the causal component of these reactions and minimize noise, but this approach remains sensitive to several well-known challenges: market expectations may be shaped by incomplete or asymmetric information, investors may misprice deals due to behavioral biases, and short-term reactions may not fully reflect long-run value creation. Third, the measurement of the acquirer pool is also subject to error, as some potential acquirers may be unobserved or misclassified.

Finally, a set of limitations stems from the auction theory and rent-sharing frameworks themselves. The model assumes that acquisitions reflect optimal matches between acquirers and startups—that is, the acquirer who acquires the asset is the one that generates the highest surplus. In practice, however, matching frictions, search costs, and information asymmetries can lead to suboptimal pairings. Moreover, the framework abstracts from several important features of real-world dealmaking, including sequential negotiations, renegotiation risk, winner's curse dynamics, and affiliated bidding behavior, all of which may distort observed outcomes relative to the theoretical benchmark. As a result, the observed surplus may not reflect the true potential value under an efficient allocation, potentially affecting estimates of both total value creation and rent division. In addition, the counterfactual—what value the startup could have captured under different exit conditions—is unobserved and must be inferred indirectly through the lens of the model, which rests on assumptions about matching and equilibrium behavior. These concerns do not invalidate the core empirical patterns reported. However, they do call for cautious interpretation of the results, while pointing to promising directions for future research.

The findings carry important implications for both policy and management. On the policy side, they suggest that interventions should not focus solely on increasing value creation—through instruments such as carbon taxes, emissions standards, or R&D subsidies—but also on improving the conditions under which startups can retain a larger share of the value they generate. This requires addressing bargaining asymmetries at the point of commercialization, particularly in settings where incumbents hold disproportionate power. Of course, this assumes that incumbents remain willing to acquire these startups even under reduced rent extraction—an outcome more likely when alternative sources for the technology, such as internal development or third-party acquisition, are unavailable, and the innovation is viewed as essential for long-term growth or competitive positioning.

One potential channel for reducing these asymmetries is increasing competition in acquisition markets, which could, in principle, strengthen startup bargaining power. This connects to broader antitrust debates by highlighting how high levels of concentration in downstream markets can suppress upstream innovation (Antón et al., 2024; Federico et al., 2020; Segal and Whinston, 2007; Shapiro, 2025). When startups are unable to appropriate a fair share of the value they generate, certain technologies may never be developed or commercialized, distorting the direction of innovation. Nonetheless, as the empirical results suggest, the current market structure—thin acquisition markets for science-based ventures—is itself an equilibrium outcome of the innovation ecosystem. Most likely, acquirer concentration arises endogenously from the nature of these technologies, which are often highly specialized, capital-intensive, and dependent on complementary assets held by a few large incumbents (Teece, 1986; Klepper, 1996; Sutton, 2007). As such, shifting the structure of downstream markets is far from straightforward.

A more viable channel, thus, may lie in improving markets for technology and, especially, in fostering independent commercialization pathways. Rather than reshaping who the acquirers are, an alternative is to strengthen the startup's position—either by increasing its ability to scale independently or by expanding the institutional infrastructure that facilitates transfers outside traditional acquisition channels. Achieving this requires complementary investments and institutional changes that expand the set of capable commercializers, lower the cost for startups to build required downstream capabilities, or both. Policy tools might include targeted support for shared manufacturing platforms, public–private commercialization infrastructure, and mechanisms that reduce the fixed and transaction costs associated with scaling advanced technologies.

For startup managers and investors, the findings highlight the importance of actively managing exit conditions by strengthening the startup's bargaining position. One key strategy is to invest—either directly or through signaling—in the potential for independent scaling. For example, raising funds earmarked for pilot manufacturing or early commercialization efforts can serve as a credible threat that improves negotiating leverage. Second, startups should consider developing early on select complementary capabilities in-house, particularly those that are difficult to outsource or contract for, such as specialized manufacturing or integration with adjacent technologies. While some of these investments may be duplicative or inefficient from a broader perspective, since incumbents often already hold these downstream assets, they can be strategically necessary to signal credible outside options and improve the terms of eventual acquisition. Third, distribution and commercialization capabilities could be factored into geographic and operational decisions. Locating operations near key customer hubs—particularly within regional industrial clusters, such as those in oil and gas or automotive—, can facilitate the path to market.

For managers at incumbent firms, the findings also carry important implications. While incumbents may benefit in the short run from capturing a larger share of the surplus in acquisitions—particularly when bargaining with science-based startups lacking outside options—systematically suppressing startup returns can undermine the long-term health of the innovation ecosystem. In many science-intensive industries, incumbents increasingly rely on external innovation rather than internal R&D to access novel technologies Fleming et al. (2019). If upstream entrepreneurs and investors anticipate weak returns due to structural rent-sharing disadvantages, this may discourage entry into precisely the kinds of frontier domains that incumbents depend on for future growth. In this sense, value capture strategies that prioritize short-term extraction can erode the very pipeline of external innovation that sustains competitiveness in the long term.

To mitigate these risks, incumbents can play a more proactive role in shaping commercialization pathways by supporting startups through two complementary channels: reducing market risk and expanding access to complementary capabilities. The former includes corporate venture investments, validation partnerships, or co-development agreements that help early-stage ventures demonstrate feasibility and move closer to market readiness. The latter focuses on co-developing or enabling access to assets such as pilot manufacturing facilities, prototyping labs, or technical infrastructure that science-based startups often lack. These complementary capabilities not only improve startups' bargaining power by strengthening their outside options but also generate spillovers for the incumbent by spurring innovation within key technological domains.

A concrete example of this approach is IBM's support for quantum startups. Through the IBM Quantum Network, the company grants startups access to one of its core technological assets, quantum computing infrastructure. The goal is not necessarily to acquire these startups, but to enable them to explore applications and push the frontier of quantum technologies closer to market. Crucially, by opening access to high-cost, high-complexity infrastructure that startups could not feasibly build on their own, IBM lowers the barriers to upstream innovation in a domain where downstream capabilities are essential. While IBM may not capture direct returns through acquisition, it stands to benefit indirectly, through knowledge spillovers, improved tools and protocols, and a stronger external pipeline of advanced technologies. This model illustrates how granting access to key complementary assets can stimulate upstream innovation in ways that ultimately reinforce the incumbent's long-term position, even without formal ownership or control over the startup.

Beyond these implications, the paper makes several contributions. First, it contributes to our understanding of the structural barriers that limit the ability of science-based innovations to reach the market through startups. While prior research has emphasized challenges related to demand uncertainty (Dalla Fontana and Nanda, 2023; Van den Heuvel and Popp, 2023), capital intensity (Hall and Lerner, 2010), technological risk and experimentation (Ewens et al., 2018; Kerr et al., 2014; Nanda and Rhodes-Kropf, 2017; Howell, 2017), and long development timelines (Narain, 2025), this paper shifts attention to a complementary friction: weak value capture. Specifically, it highlights how the division of surplus in sequential innovation systems systematically suppresses startup returns (Gans and Stern, 2000; Scotchmer, 1996; Arora et al., 2024a), even when science-based startups create large total value. This distortion arises from structural features of the commercialization environment—most notably, the absence of critical complementary capabilities and the concentration of these assets in a small set of incumbent firms (Teece, 1986; Helfat and Lieberman, 2002; Kapoor and Furr, 2015). Furthermore, while I focus on startups because they offer a measurable and discrete locus of transfer, they represent only one channel for bringing science to market (Cohen et al., 2002); similar frictions may affect earlier decisions in the translation process—from scientists choosing whether to disclose and pursue further an innovation (Masclans et al., 2025), to technology transfer offices and other intermediaries deciding which ideas and innovations to prioritize (Cohen et al., 2025).

The paper also contributes to the literature on commercialization modes and entrepreneurial strategy (e.g., Ceccagnoli et al., 2014; Gans et al., 2002; Marx et al., 2014), underscoring that independent scaling pathways matter not only as alternative routes to market, but also as bargaining instruments that shape outcomes in acquisition negotiations. Likewise, it informs the literature on M&A and corporate strategy by shifting attention from deal- and firm-specific characteristics to the structural conditions under which acquisitions occur. By focusing on the role of the market structure of potential acquirers, it shows how the broader architecture of acquisition markets can systematically influence not only the total value created but also how that value is distributed between acquirer and target. This complements existing work on M&A drivers and performance across organizational and industry contexts (e.g., Barney, 1988; Capron and Shen, 2007; Feldman et al., 2019; Kaul and Wu, 2016; Testoni, 2024; Villalonga and McGahan, 2005). Finally, the paper further contributes to the literature on innovation and downstream market structure by documenting how the number and size of potential acquirers vary systematically with startup type, and how this structure influences value creation and capture. These findings provide new empirical evidence consistent with theories of market concentration and appropriability in R&D intensive sectors (Cohen, 2010; Klepper, 1996; Sutton, 2007).

Methodologically, the paper offers three innovations. First, it develops an approach to distinguish value creation from value capture in startup acquisitions, enabling the empirical study of rent-sharing under conditions of sequential innovation. This framework allows for both the analysis of underexplored questions, such as who captures value in science-based domains, and the reinterpretation of classical innovation frictions through a surplus division lens. Second, it introduces a novel classification method for identifying science-based startups using large language models applied to textual business descriptions, improving on traditional proxies such as patent and publication data. Third, the paper adapts the parametric approach of Kogan et al. (2017) to isolate the market signal in acquisitions by public firms, allowing for clean estimation of joint surplus and enabling general application to other studies of acquisition outcomes using stock market data.

References

Acemoglu, D. and Linn, J. (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, 119(3):1049–1090.

Adner, R. and Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic management journal*, 31(3):306–333.

Aggarwal, V. A. and Hsu, D. H. (2009). Modes of cooperative r&d commercialization by start-ups. *Strategic management journal*, 30(8):835–864.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., and Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, 124(1):1–51.

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Akerlof, G. A. (1970). 4. the market for 'lemons': quality uncertainty and the market mechanism. *Market Failure or Success*, 66.

Andrews, M. J., Chatterji, A., Lerner, J., and Stern, S. (2022). *The role of innovation and entrepreneurship in economic growth*. University of Chicago Press.

Antón, M., Ederer, F., Giné, M., and Schmalz, M. (2024). Innovation: the bright side of common ownership? *Management Science*.

Arora, A., Belenzon, S., Ferracuti, E., and Nagar, J. P. (2024a). Revisiting the private value of scientific inventions.

Arora, A., Belenzon, S., and Patacconi, A. (2018). The decline of science in corporate r&d. *Strategic Management Journal*, 39(1):3–32.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020). The changing structure of american innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, 20(1):39–93.

Arora, A., Belenzon, S., and Suh, J. (2022). Science and the market for technology. *Management Science*, 68(10):7176–7201.

Arora, A., Fosfuri, A., and Rønde, T. (2024b). The missing middle: Value capture in the market for startups. *Research Policy*, 53(3):104958.

Arora, A. and Merges, R. P. (2004). Specialized supply firms, property rights and firm boundaries. *Industrial and Corporate Change*, 13(3):451–475.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics: Volume two: Private enterprise and state intervention*, pages 219–236. Springer.

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15(1):659–688.

Barney, J. B. (1988). Returns to bidding firms in mergers and acquisitions: Reconsidering the relatedness hypothesis. *Strategic Management Journal*, 9(S1):71–78.

Benson, D. and Ziedonis, R. H. (2010). Corporate venture capital and the returns to acquiring portfolio companies. *Journal of Financial Economics*, 98(3):478–499.

Boehmer, E., Fishe, R. P., and Pollock, R. K. (2003). Do institutions receive favorable allocations in ipos with better long-run returns? *Journal of Financial and Quantitative Analysis*, 38(3):371–389.

Bresnahan, T. and Gambardella, A. (1998). 10 the division of inventive labor and the extent of the market. *General purpose technologies and economic growth*, page 253.

Bryan, K. A. and Williams, H. L. (2021). Innovation: market failures and public policies. In *Handbook of industrial organization*, volume 5, pages 281–388. Elsevier.

Capron, L. and Shen, J.-C. (2007). Acquisitions of private vs. public firms: Private information, target selection, and acquirer returns. *Strategic management journal*, 28(9):891–911.

Carlson, N. and Burbano, V. (2024). The use of llms to annotate data in management research: Warnings, guidelines, and an application to organizational communication. *Guidelines, and an Application to Organizational Communication (May 21, 2024)*.

Ceccagnoli, M., Graham, S. J., Higgins, M. J., and Lee, J. (2010). Productivity and the role of complementary assets in firms' demand for technology innovations. *Industrial and corporate change*, 19(3):839–869.

Ceccagnoli, M., Higgins, M. J., and Kang, H. D. (2018). Corporate venture capital as a real option in the markets for technology. *Strategic Management Journal*, 39(13):3355–3381.

Ceccagnoli, M., Higgins, M. J., and Palermo, V. (2014). Behind the scenes: Sources of complementarity in r&d. *Journal of Economics & Management Strategy*, 23(1):125–148.

Chondrakis, G., Serrano, C. J., and Ziedonis, R. H. (2021). Information disclosure and the market for acquiring technology companies. *Strategic Management Journal*, 42(5):1024–1053.

Christensen, C. M. (2015). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*, 1:129–213.

Cohen, W. M., Hasan, S., and Masclans, R. (2025). When do intermediaries distort scientific diffusion? evidence from google search. *Working Paper*.

Cohen, W. M. and Klepper, S. (1996). A reprise of size and r & d. *The Economic Journal*, 106(437):925–951.

Cohen, W. M. and Levin, R. C. (1989). Empirical studies of innovation and market structure. *Handbook of industrial organization*, 2:1059–1107.

Cohen, W. M., Levinthal, D. A., et al. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 35(1):128–152.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2002). Links and impacts: the influence of public research on industrial r&d. *Management science*, 48(1):1–23.

Cohen, W. M., Sauermann, H., and Stephan, P. (2020). Not in the job description: The commercial activities of academic scientists and engineers. *Management Science*, 66(9):4108–4117.

Cunningham, C., Ederer, F., and Ma, S. (2021). Killer acquisitions. *Journal of political economy*, 129(3):649–702.

Dalla Fontana, S. and Nanda, R. (2023). Innovating to net zero: can venture capital and start-ups play a meaningful role? *Entrepreneurship and innovation policy and the economy*, 2(1):79–105.

David, H. A. and Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons.

Dell, M. (2024). Deep learning for economists. Technical report, National Bureau of Economic Research.

DeLong, G. L. (2001). Stockholder gains from focusing versus diversifying bank mergers. *journal of Financial Economics*, 59(2):221–252.

Durvasula, M. M., Eyuboglu, S., and Ritzwoller, D. M. (2024). Distilling data from large language models: An application to research productivity measurement. *arXiv preprint arXiv:2405.08030*.

Ederer, F. and Pellegrino, B. (2023). The great start-up sellout and the rise of oligopoly. In *AEA Papers and Proceedings*, volume 113, pages 274–278. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Edmans, A. (2012). The link between job satisfaction and firm value, with implications for corporate social responsibility. *Academy of Management Perspectives*, 26(4):1–19.

Ewens, M., Nanda, R., and Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, 128(3):422–442.

Ewens, M., Peters, R., and Wang, S. (2024). Measuring intangible capital with market prices. *Management Science*.

Fama, E. F. (1970). Efficient capital markets. *Journal of finance*, 25(2):383–417.

Federico, G., Morton, F. S., and Shapiro, C. (2020). Antitrust and innovation: Welcoming and protecting disruption. *Innovation Policy and the Economy*, 20(1):125–190.

Feldman, E. R., Amit, R., and Villalonga, B. (2019). Family firms and the stock market performance of acquisitions and divestitures. *Strategic Management Journal*, 40(5):757–780.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, 47(1):117–132.

Fleming, L., Greene, H., Li, G., Marx, M., and Yao, D. (2019). Government-funded research increasingly fuels innovation. *Science*, 364(6446):1139–1141.

Fleming, L. and Sorenson, O. (2004). Science as a map in technological search. *Strategic management journal*, 25(8-9):909–928.

Frésard, L., Hoberg, G., and Phillips, G. M. (2020). Innovation activities and integration through vertical acquisitions. *The Review of Financial Studies*, 33(7):2937–2976.

Gambardella, A., Heaton, S., Novelli, E., and Teece, D. J. (2021). Profiting from enabling technologies? *Strategy Science*, 6(1):75–90.

Gans, J. S., Hsu, D. H., and Stern, S. (2002). When does start-up innovation spur the gale of creative destruction? *RAND Journal of Economics*, 33(4):571–586.

Gans, J. S. and Stern, S. (2000). Incumbency and r&d incentives: Licensing the gale of creative destruction. *Journal of Economics & Management Strategy*, 9(4):485–511.

Gans, J. S. and Stern, S. (2003). The product market and the market for "ideas": commercialization strategies for technology entrepreneurs. *Research policy*, 32(2):333–350.

Gans, J. S. and Stern, S. (2010). Is there a market for ideas? *Industrial and Corporate Change*, 19(3):805–837.

Gerarden, T. D. (2023). Demanding innovation: The impact of consumer subsidies on solar panel production costs. *Management Science*, 69(12):7799–7820.

Gornall, W. and Strebulaev, I. A. (2021). The economic impact of venture capital: Evidence from public companies. *Available at SSRN 2681841*.

Graham, S. J., Merges, R. P., Samuelson, P., and Sichelman, T. (2009). High technology entrepreneurs and the patent system: Results of the 2008 berkeley patent survey. *Berkeley Technology Law Journal*, pages 1255–1327.

Green, J. R. and Scotchmer, S. (1995). On the division of profit in sequential innovation. *The Rand journal of economics*, pages 20–33.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719.

Hall, B. H. and Lerner, J. (2010). The financing of r&d and innovation. In *Handbook of the Economics of Innovation*, volume 1, pages 609–639. Elsevier.

Helfat, C. E. and Lieberman, M. B. (2002). The birth of capabilities: market entry and the importance of pre-history. *Industrial and corporate change*, 11(4):725–760.

Henderson, R. M. and Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, pages 9–30.

Higgins, M. J. and Rodriguez, D. (2006). The outsourcing of r&d through acquisitions in the pharmaceutical industry. *Journal of financial economics*, 80(2):351–383.

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

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of political economy*, 124(5):1423–1465.

Hoberg, G. and Phillips, G. M. (2025). Scope, scale, and concentration: The 21st-century firm. *The Journal of Finance*, 80(1):415–466.

Holmstrom, B. (1989). Agency costs and innovation. *Journal of Economic Behavior & Organization*, 12(3):305–327.

Holmstrom, B. and Roberts, J. (1998). The boundaries of the firm revisited. *Journal of Economic perspectives*, 12(4):73–94.

Howell, S. T. (2017). Financing innovation: Evidence from r&d grants. *American economic review*, 107(4):1136–1164.

Hsu, D. H. (2006). Venture capitalists and cooperative start-up commercialization strategy. *Management Science*, 52(2):204–219.

Hsu, D. H. and Ziedonis, R. H. (2013). Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management Journal*, 34(7):761–781.

Kapoor, R. and Furr, N. R. (2015). Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry. *Strategic management journal*, 36(3):416–436.

Kapoor, R. and Klueter, T. (2021). Unbundling and managing uncertainty surrounding emerging technologies. *Strategy Science*, 6(1):62–74.

Kaul, A. and Wu, B. (2016). A capabilities-based perspective on target selection in acquisitions. *Strategic Management Journal*, 37(7):1220–1239.

Kerr, W. R. and Nanda, R. (2015). Financing innovation. *Annual Review of Financial Economics*, 7(1):445–462.

Kerr, W. R., Nanda, R., and Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, 28(3):25–48.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *The American economic review*, pages 562–583.

Klepper, S. (2002). Firm survival and the evolution of oligopoly. *RAND journal of Economics*, pages 37–61.

Klepper, S. and Thompson, P. (2006). Submarkets and the evolution of market structure. *The RAND Journal of Economics*, 37(4):861–886.

Kline, P., Petkova, N., Williams, H., and Zidar, O. (2019). Who profits from patents? rent-sharing at innovative firms. *The quarterly journal of economics*, 134(3):1343–1404.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The quarterly journal of economics*, 132(2):665–712.

Kolev, J., Haughey, A., Murray, F., and Stern, S. (2022). Of academics and creative destruction: Startup advantage in the process of innovation. Technical report, National Bureau of Economic Research.

Koning, R., Hasan, S., and Chatterji, A. (2022). Experimentation and start-up performance: Evidence from a/b testing. *Management Science*, 68(9):6434–6453.

Kortum, S. S. and Lerner, J. (2000). Assessing the contribution of venture capital to innovation? *RAND Journal of Economics*, 31(4):674–692.

Krishna, V. (2009). *Auction theory*. Academic press.

Lazear, E. P. (2005). Entrepreneurship. *Journal of Labor Economics*, 23(4):649–680.

Lerner, J. and Merges, R. P. (1998). The control of technology alliances: An empirical analysis of the biotechnology industry. *The Journal of Industrial Economics*, 46(2):125–156.

Lerner, J. and Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives*, 34(3):237–261.

Lerner, J. and Nanda, R. (2023). Venture capital and innovation. In *Handbook of the Economics of Corporate Finance*, volume 1, pages 77–105. Elsevier.

Lerner, J. and Seru, A. (2022). The use and misuse of patent data: Issues for finance and beyond. *The Review of Financial Studies*, 35(6):2667–2704.

Ma, S. (2020). The life cycle of corporate venture capital. *The Review of Financial Studies*, 33(1):358–394.

Marx, M. and Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, 41(9):1572–1594.

Marx, M., Gans, J. S., and Hsu, D. H. (2014). Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry. *Management Science*, 60(12):3103–3123.

Masclans, R., Hasan, S., and Cohen, W. M. (2025). Measuring the commercial potential of science. *Strategic Management Journal*, 46(9):2199–2236.

Moeen, M. (2017). Entry into nascent industries: Disentangling a firm's capability portfolio at the time of investment versus market entry. *Strategic Management Journal*, 38(10):1986–2004.

Moeller, S. B., Schlingemann, F. P., and Stulz, R. M. (2005). Wealth destruction on a massive scale? a study of acquiring-firm returns in the recent merger wave. *The Journal of Finance*, 60(2):757–782.

Myers, S. C. (1984). The capital structure puzzle. *The Journal of Finance*, 39(3):575–592.

Nagar, J. P., Breschi, S., and Fosfuri, A. (2024). Erc science and invention: Does erc break free from the eu paradox? *Research Policy*, 53(8):105038.

Nanda, R. and Rhodes-Kropf, M. (2017). Financing risk and innovation. *Management science*, 63(4):901–918.

Nanda, R., Younge, K., and Fleming, L. (2015). Innovation and entrepreneurship in renewable energy. *The changing frontier: Rethinking science and innovation policy*, 199.

Narain, N. (2025). How patient is venture capital? *Harvard University Working Paper*.

Palermo, V., Higgins, M. J., and Ceccagnoli, M. (2019). How reliable is the market for technology? *Review of Economics and Statistics*, 101(1):107–120.

Phillips, G. M. and Zhdanov, A. (2013). R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies*, 26(1):34–78.

Pisano, G. P. (1990). The r&d boundaries of the firm: an empirical analysis. *Administrative science quarterly*, pages 153–176.

Popp, D. (2002). Induced innovation and energy prices. *American economic review*, 92(1):160–180.

Samila, S. and Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *The Review of Economics and Statistics*, 93(1):338–349.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Managagement Review*, 6(2):41–49.

Sauermann, H. and Cohen, W. M. (2010). What makes them tick? employee motives and firm innovation. *Management science*, 56(12):2134–2153.

Saxenian, A. (1996). *Regional advantage: Culture and competition in silicon valley and route 128, with a new preface by the author*. Harvard University Press.

Scotchmer, S. (1996). Protecting early innovators: should second-generation products be patentable? *The Rand Journal of Economics*, pages 322–331.

Scotchmer, S. (2004). *Innovation and incentives*. The MIT Press.

Segal, I. and Whinston, M. D. (2007). Antitrust in innovative industries. *American Economic Review*, 97(5):1703–1730.

Shapiro, C. (2025). Acquisitions to enter new markets. *Journal of Economic Perspectives*, 39(1):53–76.

Sutton, J. (1991). *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. MIT press.

Sutton, J. (2007). Market structure: theory and evidence. *Handbook of industrial organization*, 3:2301–2368.

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

Testoni, M. (2022). The market value spillovers of technological acquisitions: Evidence from patent-text analysis. *Strategic Management Journal*, 43(5):964–985.

Testoni, M. (2024). Transportation networks and competition in the market for corporate control. *Strategic management journal*, 45(6):1180–1208.

Van den Heuvel, M. and Popp, D. (2023). The role of venture capital and governments in clean energy: Lessons from the first cleantech bubble. *Energy Economics*, 124:106877.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

Villalonga, B. and McGahan, A. M. (2005). The choice among acquisitions, alliances, and divestitures. *Strategic management journal*, 26(13):1183–1208.

Appendix

A A Simple Model: Startup Market Thickness, Outside Options, Value Creation, and Value Capture

The model considers a stylized setting in which a startup negotiates with a set of potential acquirers, with two main assumptions that, as I show in the results section, are supported by the data. The first assumption is that the bidder market structure differs systematically by type of innovation (science-based *vs.* non-science-based) in terms of the number of potential acquirers, their average size, and the relationship between these two dimensions.

1. Science-based startups are associated with a narrower pool of potential acquirers, typically composed of large incumbents. This is because science-based startups often require specialized assets to scale manufacturing and distribution and these are not commonly available. However, as the number of potential bidders increases, the average size and capability of these acquirers tends to decline. The intuition is that only a few large incumbents possess the specialized complementary assets required to scale commercialization. If these are not interested in acquiring the technology, the remaining potential acquirers are smaller, more fragmented, and less capable (see case examples in Appendix B).

2. Non-science-based startups, in contrast, attract a broader and more heterogeneous set of potential acquirers. As the number of bidders increases, the likelihood of encountering a large acquirer rises. The intuition is that the broader applicability of these technologies across sectors expands the pool of potential acquirers, and by order statistics, a larger pool increases the likelihood of attracting a high-valuation or large acquirer (see case examples in Appendix B).

The second assumption is that science-based startups have weak outside options, which translate into a lower reservation value—that is, a lower minimum payoff they can secure outside of acquisition. This reflects the challenges (e.g., high costs and time; access to contract manufacturing) involved in accessing or developing the necessary complementary capabilities to independently scale science-based innovations. In contrast, non-science-based startups are more likely to scale independently, due to lower capital intensity, fewer regulatory constraints, and more accessible commercialization pathways. As a result, their outside option is stronger: if acquisition terms are unattractive, they can credibly pursue growth on their own.

Note that I treat the downstream market structure and availability of complementary capabilities via contract manufacturing and distribution as exogenous, but they may well not be. In fact, these may be jointly determined by other factors, such as the nature of the innovation (e.g., potential number of market applications at discovery as well as the market demand for each of these applications (Bresnahan and Gambardella, 1998)). While very interesting, in this paper I abstract from these considerations, which I leave for future research.

Startup valuations are modeled as acquirer-specific, reflecting differences in the complementarities between the startup's technology and each incumbent's existing capabilities. The intuition is that each potential acquirer places a different value on the same startup, depending on idiosyncratic factors. This is a relevant assumption, as it entails that demand for the startup's technologies is not homogeneous and, as a result, capture outcomes will depend not only on the number of acquirers but also on the valuations each bidder places for a given startup. Finally, given that many acquisitions occur through a process of offers, negotiations, and competing bids, auction theory provides a parsimonious and tractable framework for formal analysis. In particular, I use a second-price sealed-bid auction model. In this setting, the incumbent with the highest valuation acquires the startup and pays the second-highest bid, subject to the startup's reservation value.

A.1 Model

A finite number of incumbent firms, indexed by $i \in \{1, \dots, N\}$, participate in the auction. The startup offered for sale is characterized by its type $\theta \in \{S, NS\}$, indicating whether the technology is specialized (science) or non-specialized (non-science). The type determines both the market structure and the distribution of bidder valuations. The goal is to characterize equilibrium outcomes—price, joint surplus, and capture—and how they vary with N and θ.[57]

Each bidder is risk-neutral and privately informed about its valuation for the startup. The valuation, denoted v_i, reflects the degree of complementarity between the startup's technology and the bidder's existing assets. Importantly, v_i also represents the total joint surplus (V_i) introduced in the conceptual and empirical framework if the startup is acquired and integrated by bidder i.[58] Formally, I assume

$$v_i \sim \text{Uniform}[0, \bar{v}_\theta(N)], \tag{9}$$

[57] An equilibrium here is a strategy profile $\{b_i(v_i)\}_{i=1}^N$, one for each bidder, such that no bidder has an incentive to deviate, given their beliefs about the distribution of other bidders' valuations and the strategies they follow. I study equilibrium behavior in a second-price sealed-bid auction under the standard independent private values (IPV) framework. Each bidder i selects a bid $b_i(v_i)$ to maximize expected utility given their own valuation v_i and beliefs about others' valuations. It is well known that truthful bidding $b_i(v_i) = v_i$ constitutes a weakly dominant strategy (Vickrey, 1961; Krishna, 2009). Thus, the strategy profile in which all bidders report their true valuations is a Bayesian Nash Equilibrium. Given this equilibrium, the goal is to characterize the resulting allocation, price, and surplus division. I referred to these as equilibrium outcomes, as they arise endogenously from the equilibrium bidding behavior, and I analyze how these outcomes vary with the number of bidders and the structure of valuation distributions across technology types.

[58] Put simply, bidder i values the startup at v_i and will pay a price p, where $p < v_i$. Assuming, for simplicity, zero costs, the startup's surplus is p, the acquirer's surplus is $v_i - p$, and the joint surplus is $p + (v_i - p) = v_i$.

with valuations drawn independently across bidders. The upper bound $\bar{v}_\theta(N)$ depends on both the type θ and the number of participants N. That is, bidder valuations are assumed to be drawn from a uniform distribution with an upper bound that is itself a function of θ and N.

As exposed above, supported by the data, and illustrated in the case examples in Appendix B, I also assume the following in regard to market structure, defined by the number and size of potential bidders:

$$\frac{d\bar{v}_S(N)}{dN} < 0 \quad \text{and} \quad \frac{d\bar{v}_{NS}(N)}{dN} > 0 \tag{10}$$

These assumptions capture structural asymmetries in market organization. For specialized technologies ($\theta = S$), value is concentrated among a few large acquirers. If these fail to engage in the acquisition, additional bidders are increasingly marginal, average complementarity falls and so does the startup's joint surplus (v_i). For non-specialized technologies ($\theta = NS$), applicability is broad and diffuse, and more bidders expand the probability of high-value matches (simply by reason of order statistics).

A.1.1 Bidders' Optimization Problem

Each bidder i submits a bid $b_i \in \mathbb{R}_+$. Let v_i denote bidder i's private valuation. The startup sets a reservation price r proportional to the highest valuation:

$$r = \alpha \cdot v_{(1)}, \quad \text{with } \alpha \in (0, 1),$$

where $v_{(1)} = \max\{v_1, \ldots, v_N\}$. This represents the startup's outside option (independent scaling) expressed as a fraction of what the top bidder would generate. Let $b_{(1)} = \max\{b_1, \ldots, b_N\}$ and $b_{(2)}$ denote the second-highest bid. The project is sold if $b_{(1)} \geq r$, and the winning bidder pays

$$p = \max\{r, b_{(2)}\}$$

Bidder i's utility function is:

$$u_i(b_i; v_i) = \begin{cases} v_i - \max\{r, b_{(2)}\} & \text{if } b_i > \max_{j \neq i} b_j \\ 0 & \text{otherwise.} \end{cases}$$

The optimization problem for bidder i is:

$$\max_{b_i \in \mathbb{R}_+} \mathbb{E}\left[u_i(b_i; v_i)\right] \tag{11}$$

A.1.2 Equilibrium Bidding

Under standard assumptions of independent private values (IPV) and risk neutrality, it is a classical result (Krishna, 2009; Vickrey, 1961) that truthful bidding is a weakly dominant strategy in second-price auctions. Therefore, in equilibrium,

$$b_i^*(v_i) = v_i \quad \text{for all } i$$

The winner is the bidder with the highest valuation, and the price paid is $\max\{v_{(2)}, r\}$, where $v_{(2)}$ is the second-highest valuation. Given $\alpha < 1$, the startup is always sold, and the surplus is divided between the startup and the winning bidder. While this result holds under the standard independent private values (IPV) assumption, many relevant settings—particularly those involving scientific or uncertain technologies—depart from this framework. In such cases, bidders may hold affiliated signals[59] or face common value uncertainty, giving rise to bargaining dynamics such as winner's curse behavior or more conservative bidding.[60]

A.1.3 Equilibrium Outcomes

Let $v_{(1)}$ and $v_{(2)}$ denote the first and second order statistics among N independent and identically distributed draws from the uniform distribution on $[0, \bar{v}_\theta(N)]$. It is a standard result in the theory of order statistics that if $X_1, \ldots, X_N$ are i.i.d. draws from Uniform$[0, 1]$, then the k-th order statistic $X_{(k)}$ follows a Beta distribution:

$$X_{(k)} \sim \text{Beta}(k, N + 1 - k),$$

[59] The assumptions could be relaxed to account for affiliated bidding. The current manuscript abstracts from this possibility. In auction theory, affiliation refers to a situation where bidders' signals about the value of the asset for sale are not independent but positively correlated: a high signal observed by one bidder makes it more likely that other bidders also have high signals.

[60] The winner's curse is the risk that the winning bidder overpays in auctions with uncertain or correlated values, because winning itself is a negative signal. In thin markets—especially for science-based technologies—this risk can distort outcomes in two ways. If bidders are naive, they may overpay, inflating startup surplus. But if bidders are rational and anticipate the curse, they respond by bidding more conservatively, reducing prices. Either way, the result is a distortion: inefficient allocation or weaker value capture.

with mean $\mathbb{E}[X_{(k)}] = \frac{k}{N+1}$ (see David and Nagaraja, 2004). Applying this result to the scaled distribution Uniform$[0, \bar{v}_\theta(N)]$, I obtain:

$$v_{(1)} \sim \bar{v}_\theta(N) \cdot \text{Beta}(N, 1), \quad v_{(2)} \sim \bar{v}_\theta(N) \cdot \text{Beta}(N-1, 2)$$

It follows that the expected highest and second-highest valuations are:

$$\mathbb{E}[v_{(1)}] = \frac{N}{N+1} \cdot \bar{v}_\theta(N), \quad \mathbb{E}[v_{(2)}] = \frac{N-1}{N+1} \cdot \bar{v}_\theta(N) \tag{12}$$

The price paid by the winning bidder is the maximum between the second-highest bid and the reservation price, which is set as a fraction $\alpha \in (0, 1)$ of the top valuation:

$$p = \max\{v_{(2)}, \alpha \cdot v_{(1)}\}$$

The share of total value captured by the startup is then:

$$s = \frac{p}{v_{(1)}}$$

Using the expressions in (12), I can characterize expected outcomes. The expected price is:

$$\mathbb{E}[p] = \bar{v}_\theta(N) \cdot \max\left\{\alpha \cdot \frac{N}{N+1}, \frac{N-1}{N+1}\right\},$$

and the expected share of value captured by the startup is:

$$\mathbb{E}[s] = \max\left\{\alpha, \frac{N-1}{N}\right\}$$

A.1.4 Comparative Statics and Interpretation

These expressions yield the following predictions. For specialized technologies ($\theta = S$), the function $\bar{v}_S(N)$ is decreasing in N, and thus both $\mathbb{E}[v_{(1)}]$ and $\mathbb{E}[p]$ may decline with the number of bidders. This captures the empirical regularity that value creation falls when a few large acquirers abstain and the market fragments. The startup's share $\mathbb{E}[s]$ increases in N, as competition tightens the gap between $v_{(1)}$ and $v_{(2)}$, but this share is applied to a shrinking total.

For non-specialized technologies ($\theta = NS$), the opposite logic applies. Here, $\bar{v}_{NS}(N)$ increases in N, and thus both $\mathbb{E}[v_{(1)}]$ and $\mathbb{E}[p]$ increase. The startup's share also rises in N, but the effect is flatter, as dispersion in valuations is smaller. The market structure reflects a horizontal application of the technology: adding bidders from new sectors and geographies improves the chance of finding a high match.

A.1.5 Simulations

Following the theoretical model, I simulate acquisition outcomes. For each project, the number of potential bidders N is drawn from a type-dependent distribution, reflecting the differing market structures of specialized versus non-specialized technologies. Conditional on technology type θ, bidders are symmetric. Each bidder's valuation v_i is additively separable in L_i, where L_i reflects structural complementarities between the bidder and the startup's technology—drawn from a distribution whose mean depends on θ and N. To introduce the idiosincracti component, thhat is, that demand is not homgenous and each bidder may place a unique value to a technology based on ex-ante complmentaryasset and capabilities, this shock is modeled asε_i is an idiosyncratic shock drawn i.i.d. from a normal distribution. Specifically, I define:

Following the theoretical model, I simulate acquisition outcomes. For each project, the number of potential bidders N is drawn from a type-dependent distribution, capturing differences in market structure between specialized and non-specialized technologies. Conditional on technology type θ, bidders are symmetric. Each bidder's valuation v_i is additively separable in L_i, where L_i represents structural complementarities between the bidder and the startup's technology. L_i is drawn from a distribution whose mean depends on θ and N. To capture heterogeneity in bidder-specific valuations of the technology, arising from differences in ex-ante complementary assets and capabilities, an idiosyncratic component ε_i is added. This shock ε_i is drawn i.i.d. from a normal distribution. Formally:

$$v_i = L_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2), \quad \text{i.i.d. across } i$$

Parameters are calibrated to match empirical moments of the data presented in the paper. The startup conducts the second-price auction with the endogenous reservation price r.

The valuation v_i should be interpreted as the bidder's private value from acquiring the project. While L_i reflects structural complementarities with the project, the shock ε_i captures firm-specific conditions, ranging from specific capabilities to other idiosyncratic aspects. For example, a mid-size acquirer that recently raised capital may report an elevated ε_i relative to its size, thereby increasing its bid despite a moderate L_i.

The reservation price r represents the startup's outside option to continue development or commercialization without being acquired, as discussed. Rather than being fixed across deals, I allow r to depend on the project and context. Specifically, I define $r = \alpha \cdot v_{(1)}$ where $\alpha \in (0, 1)$ is a type-dependent random variable capturing the strength of the startup's fallback. This reduced-form specification reflects the idea that the value of self-commercialization is positively correlated with the potential value to the top bidder, but at a discount. For example, a startup with a partially validated oncology platform may see 20% of the top bidder's value in a standalone scenario, while a mature SaaS product may capture 70%.

Given this structure, and following the results of the model introduced above, the transaction occurs if and only if:

$$v_{(1),N} \geq r = \alpha \cdot v_{(1),N} \iff \alpha < 1,$$

which is always true under the assumed support for α. That is, a sale always occurs. However, the binding constraint is whether the second-highest valuation $v_{(2),N}$ exceeds the reservation. If $v_{(2),N} \geq r$, the price is set by bidding competitive pressure; if $v_{(2),N} < r$, the startup sells at the fallback (outside option) price.

This payoff structure implies that the startup benefits from competition when $v_{(2),N}$ is high, and falls back on its reservation value when competition is weak. The distributional assumptions on valuations—driven by the relationship between θ, N, and bidder sizes L_i—imply sharp comparative statics for both the price and the surplus share as functions of N and project type.

For example, under a specialized technology ($\theta = S$), one expects that:

- The distribution of N is concentrated on low values.
- Bidder sizes L_i are high for small N, but decrease in expectation with N.
- Consequently, $v_{(1),N}$ may be decreasing in N, while the share $s = v_{(2),N}/v_{(1),N}$ increases.
- The share s increases in N, reflecting greater competitive pressure, but the total surplus may decrease.

In contrast, under a non-specialized technology ($\theta = NS$), larger N correlates with higher expected L_i, so both $v_{(1),N}$ and $v_{(2),N}$ increase with N, and the startup's share varies less steeply. Figure A.1 illustrates the model's predictions based on the outcomes of a simulation with 6,000 transactions.

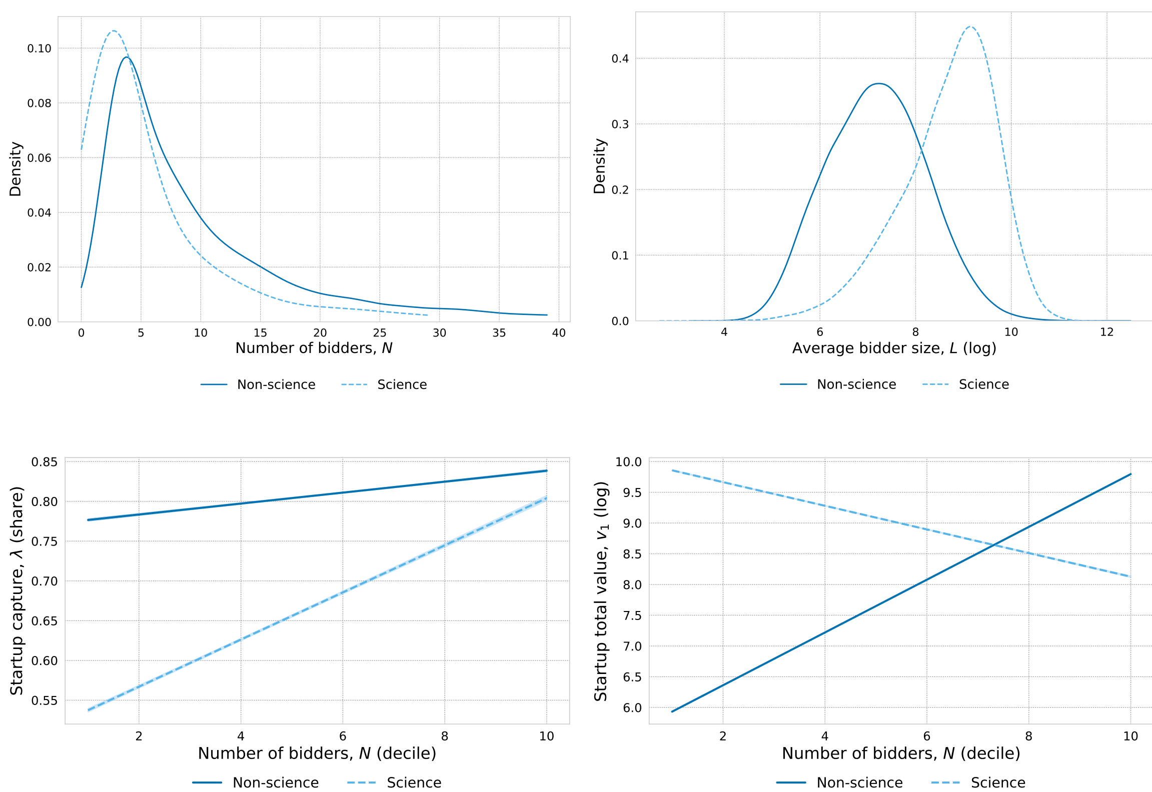


Figure A.1: The figure plots results from simulations based on model.

B Case Examples: Fragmented Adoption of Specialized Technologies

Founded in 2001 as a spin-off from the Massachusetts Institute of Technology, A123 Systems aimed to commercialize a lithium-iron-phosphate (LFP) battery chemistry that offered superior power density and thermal stability relative to conventional lithium-ion cells. Early venture financing from Sequoia Capital and GE Energy, along with a $249 million grant from the U.S. Department of Energy, the company inaugurated a 291,000 ft^2 manufacturing facility in Livonia, Michigan, in 2010, at that time the largest lithium-ion plant in North America.[61] A123's growth strategy hinged on securing at least one of the three dominant U.S. automakers as an anchor customer and potential acquirer.

General Motors tested the cells for the first-generation Chevrolet Volt, but ultimately awarded the contract to LG Chem instead, citing cost and scale considerations.[62] Chrysler announced an alliance with A123 in 2009, yet the program was shelved during Chrysler's own bankruptcy and restructuring.[63] Because the U.S. auto industry is highly concentrated, A123 faced a limited pool of potential acquirers. Having failed to secure any of the dominant incumbents as acquirers or large-scale customers, the company struggled to scale commercialization independently. As financial constraints mounted, A123 was acquired in early 2013 by Wanxiang Group, a mid-sized Chinese firm considerably smaller than the U.S. automakers the company had originally targeted.[64] Post-sale, the technology diffused mainly through niche or mid-tier applications: Fisker Automotive's low-volume Karma plug-in[65] and BAE Systems' HybriDrive propulsion kits for city buses.[66] These acquirers purchase in the thousands rather than the millions, limiting aggregate demand and depressing the economic surplus A123's IP could generate—an outcome consistent with the model's prediction that, in concentrated markets, rejection by a small set of large incumbents channels specialized technology toward fragmented, lower-value niches.

A similar example is Soraa. Founded in 2008 by University of California, Santa Barbara researchers Shuji Nakamura (Nobel Prize receipient), Steven DenBaars, and James Speck, Soraa sought to commercialize a proprietary gallium-nitride-on-gallium-nitride (GaN-on-GaN) light emitting diode architecture that delivers a violet pump and full-spectrum emission with unusually high color ren-

[61] *Wired*, "A123 Systems opens huge battery factory," 14 September 2010; Reuters, "Obama heralds opening of A123's Michigan plant," 13 September 2010.

[62] Reuters, "Battery cells for the Volt will be supplied by Korea's LG Chem," July 13 2009.

[63] Green Car Congress, "Chrysler LLC forms strategic alliance with A123Systems," April 2009; The Truth About Cars, "Fiat/Chrysler EV program loses battery supplier A123," August 2010.

[64] Reuters, "Battery maker A123 Systems files for bankruptcy," Oct 16 2012; Bloomberg, "Wanxiang wins CFIUS approval to buy bankrupt battery maker A123," Jan 29 2013.

[65] Car and Driver, "Fisker Karma production halted by A123 Systems bankruptcy," Oct 2012.

[66] Green Car Congress, "BAE Systems to offer A123Systems Li-ion unit in HybriDrive," May 2007.

dering.[67] Early venture financing from Khosla Ventures, New Enterprise Associates, and NGEN Partners underwrote process development and pilot production at the firm's facility in Fremont, California.[68] In November 2013, the company announced plans for a $400 million, state-subsidized wafer-fabrication plant in Buffalo, New York, expected to create nearly 400 jobs and to begin operations in 2015.[69] Between 2014 and 2015, Soraa withdrew from the Buffalo project, which was subsequently reassigned to SolarCity under New York's "Buffalo Billion" initiative, eliminating the anticipated scale-up path for Soraa.[70]

The LED lamp market has long exhibited oligopolistic structure, with Signify (Philips), Osram, and GE/Savant controlling a major share of global sales.[71] Soraa entered this arena with a differentiated, high-color-rendering GaN-on-GaN architecture that delivered superior spectral quality but at a substantially higher costs and with incompatibilities with existing fixtures, transformers, and dimmers—a significant hurdle to adoption. [72] To address this incompatibility, the company launched the "Works with Soraa" program, aiming to validate and ensure compatibility with various lighting components. Despite these efforts, the integration into existing systems was not seamless. Later, Soraa's started focusing on niche markets with high-end applications for museums, galleries, and premium hospitality venues.

This niche focus, combined with the higher costs associated with their advanced LED technology, may have limited their appeal to major OEMs prioritizing cost-effectiveness and broad compatibility, potentially deterring large OEMs seeking plug-and-play solutions. Persisting capacity constraints and cash-constrained, together with the absence of adoption by dominant lamp OEMs, culminated in the March 2020 sale of Soraa's assets and intellectual property to Ecosense Lighting.[73]

[67] *Optics.org*, "Soraa to build LED factory in Buffalo," 22 November 2013.
[68] NGEN Partners, "Portfolio: Soraa Corp." (accessed May 2025).
[69] *Optics.org*, ibid., lines L3–L10.
[70] *Investigative Post*, "Tesla's solar factory in Buffalo fizzles," 11 January 2023, lines 13–14.
[71] Inc. Magazine, "The lighting industry: Philips, Sylvania and GE constituted a long-standing oligopoly," Sept 2019.
[72] LEDinside, "Soraa makes new record on GaN-on-GaN LEDs," Feb 18 2013.
[73] Ecosense press release, "Ecosense acquires assets from Soraa," 24 March 2020.

C Estimating Acquisition Value from Market Returns: A Methodology to Isolate the Signal from the Noise

A central challenge in estimating the economic value of acquisitions lies in the noisy nature of market responses. Observed stock price reactions around acquisition announcements reflect both the underlying value of the transaction and idiosyncratic movements in the market. To extract more accurate estimates of acquisition value, I develop a signal extraction approach that builds on the methodology developed by Kogan et al. (2017). This section provides the intuition and details of this methodology.

Kogan et al. (2017)'s methodology, developed to estimate patent values based on stock market reactions upon granting announcement, is based on the assumption that patent grant announcements cannot destroy value. Accordingly, they model stock market responses using a normal distribution truncated at zero, a choice that also offers convenient tractable properties. In this paper, I modify their methodology by allowing the market signal to be negative through the introduction of a truncated distribution that bounds the downside of market reactions by transaction prices, which, in turn, assumes the market observes. In this case, I still rely on a relevant assumption, although more relaxed: that transactions cannot destroy value in global. Specifically, any value destruction on the incumbent's side is bounded by the price paid to the startup (put simply, if an acquirer pays \$100 million for a target, the most it can lose is \$100 million—hence, the market signal is bounded below at \$100 million).[74] All derivations that follow from this assumption, including the adjusted likelihood function, identification strategy, and moment structure, are novel. The remainder of the setup and derivation follows the framework developed by Kogan et al. (2017), which I reproduce here for clarity and completeness.

C.1 Modeling the Observed Return

Let r_i denote the observed abnormal stock return for acquiring firm i following the announcement of an acquisition. I model this return as the sum of a latent value component v_i and a noise term ε_i:

$$r_i = v_i + \varepsilon_i, \qquad \text{(Observed return)} \qquad (13)$$

where v_i represents the component of the return attributable to the fundamental value of the transaction, such as the knowledge or technology being acquired, while ε_i captures random market fluctu-

[74]Note that the transaction price is observed without noise.

ations that obscure this signal, i.e., the market noise or measurement error I want to get rid of. The econometric problem is thus to recover v_i, the underlying signal I care about, given that I observe r_i with noise ε_i that contaminates the measurement.

To do so, I start by imposing some distributional assumptions on both the signal and noise terms. First, I assume the signal follows a normal distribution with mean zero and variance σ_x^2, but with a firm-specific truncation threshold k_i, such that:

$$v_i \sim \mathcal{N}(0, \sigma_x^2) \quad \text{truncated at } v_i > k_i \qquad \text{(True signal)} \tag{14}$$

Assuming a mean of zero for the latent value distribution can be interpreted as reflecting a neutral prior: on average, the ex ante expectation is that an acquisition creates no net value. This is a conservative assumption, consistent with a setting where realized outcomes are highly uncertain. Such a prior does not rule out the possibility of substantial upside in particular cases, but it captures the idea that, absent further information, the baseline expectation is zero net surplus on the acquirer side.

The truncation captures the assumption that losses from a transaction are bounded below by the price paid. That is, even if an acquisition turns out to destroy value for the acquirer, the most that can be lost is the amount invested, i.e., negative realized surplus cannot exceed the purchase price.[75]

The noise term is assumed to be independent of the signal and normally distributed. That is, the value acquirer gets from acquisition is orthogonal to the market or other events:

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \varepsilon_i \perp v_i \qquad \text{(Noise)} \tag{15}$$

Under these assumptions, the total variance in the observed return r_i can be decomposed into the variance of the signal σ_x^2 and the variance of the noise σ_ε^2. The total variance of r_i is then $\text{Var}(r_i) = \sigma_x^2 + \sigma_\varepsilon^2$. The key parameter summarizing their relative magnitudes is the signal-to-noise ratio, defined as the proportion of the total variance that comes from the signal:

$$\delta = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \qquad \text{(Signal-to-noise ratio)} \tag{16}$$

With these, given the observed return r_i, the objective is to estimate the underlying latent value v_i.

[75] One could argue that some acquisitions may destroy value beyond this amount, for example by locking the firm into a misguided strategic direction or crowding out better alternatives. However, for modeling purposes, I impose this bound to maintain a well-defined support for the value distribution and to reflect the idea that the immediate economic loss is capped by the acquirer's upfront investment.

The optimal estimator under squared loss is the conditional expectation $\mathbb{E}[v_i \mid r_i]$, which I derive from properties of the truncated normal distribution. Since v_i and ε_i are assumed to be independent, their joint distribution implies that the posterior distribution of $v_i \mid r_i$ is a normal distribution with mean $\mu^* = \delta\, r_i$ and variance $\sigma^{*2} = \delta\, \sigma_\varepsilon^2$, truncated at $v_i > k_i$. The conditional expectation is given by:

$$\mathbb{E}[v_i \mid r_i] = \delta\, r_i + \sqrt{\delta}\, \sigma_\varepsilon \cdot \frac{\phi\left(\frac{k_i - \delta\, r_i}{\sqrt{\delta}\, \sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{k_i - \delta\, r_i}{\sqrt{\delta}\, \sigma_\varepsilon}\right)}, \qquad \text{(Expected signal)} \qquad (17)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal probability density and cumulative distribution functions, respectively. I present the full mathematical derivation of this expression at the end of this section, while now I first focus on providing intuition.

The signal-to-noise ratio δ is a key concept here because it tells us how much of what we observe in r_i is truly the signal v_i (the acquisiton value) versus random noise ε_i (market fluctuations). If δ is close to 1, it means most of the variation in r_i is due to the signal v_i and the noise is relatively low. In this case, the shrinkage towards the prior (recall I imposed a prior of 0) is small and r_i is mostly signal. Conversely, if δ is close to 0, it means most of the variation in r_i comes from noise and that the signal is weak relative to the noise. With low δ the shrinkage is strong and the observed return is heavily influenced by noise, so the estimate of v_i is pulled toward zero, i.e., toward the prior mean.

C.2 Signal-to-noise ratio intuition

To start simply, assume no truncation, which simplifies the conditional expectation considerably. I treat v_i as coming from an untruncated normal distribution. Let the rest of the assumptions introduced above hold. In that case, $v_i \sim N(0, \sigma_x^2)$ and there is no lower bound k_i to consider. The derivation of $E[v_i \mid r_i]$ then becomes a standard linear regression problem in the context of a joint normal distribution:

$$E[v_i \mid r_i] = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}\, r_i = \delta\, r_i$$

Notice that there is no additional term involving the normal PDF (ϕ) or CDF (Φ), since these components arise solely because of the truncation correction. Without truncation, the conditional expectation is linear in r_i and does not include that second component. The second component—the adjustment term—is only necessary when accounting for the fact that v_i is restricted (truncated) to be above a certain threshold. Without that constraint, the distribution is symmetric and unbounded, and the conditional expectation reduces to the simple linear form.

More formally, in the jointly normal case, the expression holds as follows. First, when two random variables v_i and r_i are jointly normally distributed, a key property is that the conditional expectation $E[v_i \mid r_i]$ is a linear function of r_i. That is, there exist constants a and b such that $E[v_i \mid r_i] = \alpha + \beta\, r_i$. If, for simplicity, I assume that both v_i and r_i have mean zero (i.e., $E[v_i] = 0$ and $E[r_i] = 0$), then the best linear predictor simplifies. Under these conditions, the intercept a is zero, and we have $E[v_i \mid r_i] = \beta\, r_i$.

In this last expression, the slope β is determined by minimizing the mean squared error (MSE) of predicting v_i using a linear function of r_i. This leads to the classical result $\beta = \frac{\mathrm{Cov}(v_i, r_i)}{\mathrm{Var}(r_i)}$, a standard outcome from linear regression theory. Thus, given the zero mean assumptions, we can write:

$$E[v_i \mid r_i] = \beta r_i = \frac{\mathrm{Cov}(v_i, r_i)}{\mathrm{Var}(r_i)}\, r_i$$

Since $r_i = v_i + \varepsilon_i$ and v_i and ε_i are independent: $\mathrm{Cov}(v_i, r_i) = \mathrm{Cov}(v_i, v_i + \varepsilon_i) = \mathrm{Var}(v_i) + \mathrm{Cov}(v_i, \varepsilon_i)$. Because $\mathrm{Cov}(v_i, \varepsilon_i) = 0$, $\mathrm{Cov}(v_i, r_i) = \sigma_x^2$. And, recall that $\mathrm{Var}(r_i) = \sigma_x^2 + \sigma_\varepsilon^2$. Then,

$$E[v_i \mid r_i] = \frac{\mathrm{Cov}(v_i, r_i)}{\mathrm{Var}(r_i)}\, r_i = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}\, r_i. \quad \text{Let } \delta = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \to E[v_i \mid r_i] = \delta\, r_i$$

Intuition:

- Covariance ($\mathrm{Cov}(v_i, r_i)$): Measures how much v_i and r_i vary together. If r_i contains a lot of signal from v_i, this covariance will be high.

- Variance ($\mathrm{Var}(r_i)$): Measures the total variability in r_i, which includes both the signal v_i and the noise ε_i.

The ratio $\frac{\mathrm{Cov}(v_i, r_i)}{\mathrm{Var}(r_i)}$ tells us what proportion of the variability in r_i is due to v_i, and is the signal-to-noise ratio, denoted by δ.

C.2.1 Negative truncation

I now turn to the second term, the truncation adjustment in the conditional expectation $E[v_i \mid r_i]$. Recall that without truncation, we have $E[v_i \mid r_i] = \delta\, r_i$, but tat the full derivation with truncation is

$$\mathbb{E}[v_i \mid r_i] = \delta\, r_i + \sqrt{\delta}\, \sigma_\varepsilon \cdot \frac{\phi\left(\frac{k_i - \delta\, r_i}{\sqrt{\delta}\, \sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{k_i - \delta\, r_i}{\sqrt{\delta}\, \sigma_\varepsilon}\right)}$$

The full derivation follows the assumption that the acquisition signal v_i has lower bound at k_i ($v_i > k_i$), i.e., I assume the true v_i is drawn from a truncated normal distribution and that, in turn, the truncation point is below 0 ($k_i < 0$). In a truncated normal, the expectation is no longer just the untruncated mean, it must account for the fact that the left tail (below k_i) is cut off.

For a random variable Z that follows a normal distribution $N(\mu, \sigma^2)$ truncated from below at a, the expected value is given by: $E[Z \mid Z > a] = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}$, where $\phi(\cdot)$ is the standard normal PDF and $\Phi(\cdot)$ is the standard normal CDF. The ratio $\frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}$ is often called the inverse Mills ratio. Thus, in my setting, after combining the signal and noise components and conditioning on r_i, the conditional distribution of v_i becomes a truncated normal with:

- Conditional Mean (if untruncated): $\mu^* = \delta\, r_i$
- Conditional Standard Deviation: $\sigma^* = \sqrt{\delta}\, \sigma_\varepsilon$
- Truncation Point: k_i

The intuition behind this second term is as follows:

- Adjustment for missing mass: The term $\sigma^* \frac{\phi\left(\frac{k_i-\mu^*}{\sigma^*}\right)}{1-\Phi\left(\frac{k_i-\mu^*}{\sigma^*}\right)}$ corrects for the fact that the distribution of v_i is cut off below k_i. Without this adjustment, one would underestimate the true average value of v_i due to ignoring all the values below k_i.
- Inverse Mills Ratio: The fraction $\frac{\phi(z)}{1-\Phi(z)}$ (with $z = \frac{k_i-\mu^*}{\sigma^*}$) increases as the truncation point k_i gets closer to μ^*. In other words, if the truncation is severe (i.e., k_i is not far below the untruncated mean μ^*), then a significant portion of the distribution is being cut off, and the adjustment is larger (this is the case for example in the original Kogan et al. (2017) methodology).
- Scaling by σ^*: The term is scaled by the standard deviation σ^* of the conditional distribution, which adjusts the magnitude of the correction to match the dispersion of the distribution.

Notice that δ and $\sqrt{\delta}$ appear in both the main term and the adjustment. This means that the quality of the signal relative to the noise also affects how much correction is needed for the truncation:

- If the signal is strong (δ high): The adjustment is smaller because r_i is already a good indicator of v_i.
- If the signal is weak (δ low): The adjustment is larger, reflecting greater uncertainty about the true v_i.

In sum, the final expression consists of two components: the first term, $\delta\, r_i$, represents a standard linear shrinkage estimator that adjusts the observed return toward zero depending on the relative precision of the signal. When the signal-to-noise ratio δ is high, the observed return is a reliable proxy for value, and little adjustment is needed. When δ is low, the return is more heavily discounted. The second term is a non-linear adjustment that arises from the truncation of the signal distribution. This term corrects for the fact that we only consider values of v_i exceeding a threshold k_i. Intuitively, when the truncation point is close to the conditional mean μ^*, the adjustment is large and positive, reflecting the asymmetric censoring of the lower tail. In contrast, when the truncation point lies far below μ^*, the adjustment becomes negligible.

C.3 Full derivation

Assumptions

1. Distribution of Acquirer Surplus:

$$V_i \sim N(0, \sigma^2_{x,ft}), \quad k_i < 0$$

2. Distribution of Measurement Errors:

$$\varepsilon_i \sim N(0, \sigma^2_{\varepsilon,ft})$$

3. Constant Proportions (constant signal-to-noise ration):

$$\frac{\sigma^2_{x,ft}}{\sigma^2_{\varepsilon,ft}} = \text{constant}$$

4. Independence:

$$V_i \perp \varepsilon_i$$

Objective

I want to calculate:

$$E[V_i \mid r_i]$$

where $r_i = V_i + \varepsilon_i$.

Derivation

1. Joint Density Function

Since V_i and ε_i are independent, their joint density is:

$$f(V_i, \varepsilon_i) = f(V_i) f(\varepsilon_i)$$

Given:

$$f(V_i) = \frac{1}{\sqrt{2\pi}\sigma_{x,ft}} \exp\left(-\frac{V_i^2}{2\sigma_{x,ft}^2}\right), \quad V_i > k_i,$$

and

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon,ft}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma_{\varepsilon,ft}^2}\right)$$

Using the transformation $\varepsilon_i = r_i - V_i$, we have:

$$f(V_i, r_i) = \frac{1}{2\pi\sigma_{x,ft}\sigma_{\varepsilon,ft}} \exp\left(-\frac{V_i^2}{2\sigma_{x,ft}^2} - \frac{(r_i - V_i)^2}{2\sigma_{\varepsilon,ft}^2}\right), \quad V_i > k_i$$

2. Simplifying the Exponent

Expand the exponent:

$$-\frac{V_i^2}{2\sigma_{x,ft}^2} - \frac{(r_i - V_i)^2}{2\sigma_{\varepsilon,ft}^2} = -\frac{V_i^2}{2\sigma_{x,ft}^2} - \frac{r_i^2 - 2r_iV_i + V_i^2}{2\sigma_{\varepsilon,ft}^2}$$

Combine terms:

$$= -\left(\frac{V_i^2}{2\sigma_{x,ft}^2} + \frac{V_i^2}{2\sigma_{\varepsilon,ft}^2} - \frac{r_iV_i}{\sigma_{\varepsilon,ft}^2} + \frac{r_i^2}{2\sigma_{\varepsilon,ft}^2}\right)$$

Define $\sigma^2 = \sigma_{x,ft}^2 + \sigma_{\varepsilon,ft}^2$. Then:

$$\frac{1}{2\sigma_{x,ft}^2} + \frac{1}{2\sigma_{\varepsilon,ft}^2} = \frac{1}{2\sigma^2}$$

So the exponent becomes:

$$-\left(\frac{V_i^2}{2\sigma^2} - \frac{r_iV_i}{\sigma_{\varepsilon,ft}^2} + \frac{r_i^2}{2\sigma_{\varepsilon,ft}^2}\right)$$

3. Completing the Square

Focus on the V_i terms:

$$\frac{V_i^2}{2\sigma^2} - \frac{r_i V_i}{\sigma_{\varepsilon,ft}^2} = \frac{1}{2\sigma^2}\left(V_i^2 - 2\frac{\sigma^2 r_i}{\sigma_{\varepsilon,ft}^2} V_i\right)$$

Complete the square:

$$V_i^2 - 2\frac{\sigma^2 r_i}{\sigma_{\varepsilon,ft}^2} V_i = \left(V_i - \frac{\sigma^2 r_i}{\sigma_{\varepsilon,ft}^2}\right)^2 - \left(\frac{\sigma^2 r_i}{\sigma_{\varepsilon,ft}^2}\right)^2$$

Thus the exponent can be written as:

$$-\frac{1}{2\sigma^2}\left(V_i - \frac{\sigma^2 r_i}{\sigma_{\varepsilon,ft}^2}\right)^2 + (\text{terms in } r_i \text{ only})$$

These r_i-only terms are constants with respect to V_i and can be absorbed into the normalization.

4. Conditional Density Function

After completing the square, the conditional density $f(V_i \mid r_i)$ is proportional to:

$$f(V_i \mid r_i) \propto \exp\left(-\frac{(V_i - \mu^*)^2}{2\sigma^{*2}}\right), \quad V_i > k_i,$$

where

$$\mu^* = \frac{\sigma_{x,ft}^2}{\sigma^2} r_i = \delta_i r_i,$$

and

$$\sigma^{*2} = \frac{\sigma_{x,ft}^2 \sigma_{\varepsilon,ft}^2}{\sigma^2} = \delta_i \sigma_{\varepsilon,ft}^2$$

Here,

$$\delta_i = \frac{\sigma_{x,ft}^2}{\sigma_{x,ft}^2 + \sigma_{\varepsilon,ft}^2}$$

Thus, $f(V_i \mid r_i)$ is a normal density $N(\mu^*, \sigma^{*2})$ truncated at $V_i > k_i$

5. Conditional Expectation

For a truncated normal $N(\mu^*, \sigma^{*2})$ truncated at $V_i > k_i$, the expectation is:

$$E[V_i \mid r_i] = \mu^* + \sigma^* \frac{\phi\left(\frac{k_i - \mu^*}{\sigma^*}\right)}{1 - \Phi\left(\frac{k_i - \mu^*}{\sigma^*}\right)},$$

where $\phi(\cdot)$ is the standard normal PDF and $\Phi(\cdot)$ is the standard normal CDF.

Define:

$$R_i = \frac{k_i - \mu^*}{\sigma^*}$$

Then:

$$E[V_i \mid r_i] = \mu^* + \sigma^* \frac{\phi(R_i)}{1 - \Phi(R_i)}$$

6. Final Formula

Substitute back $\mu^* = \delta_i r_i$ and $\sigma^* = \sqrt{\delta_i}\sigma_{\varepsilon,ft}$:

$$E[V_i \mid r_i] = \delta_i r_i + \sqrt{\delta_i}\sigma_{\varepsilon,ft} \frac{\phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,ft}}\right)}{1 - \Phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,ft}}\right)}$$

Interpretation

- $\delta_i = \frac{\sigma^2_{x,ft}}{\sigma^2_{x,ft} + \sigma^2_{\varepsilon,ft}}$ is the signal-to-noise ratio, representing the fraction of variance from the incumbents' private value.

- $R_i = \frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,ft}}$ is the standardized truncation point.

- The ratio $\frac{\phi(R_i)}{1-\Phi(R_i)}$ adjusts the expected value to account for the truncation at $V_i > k_i$.

D Correlation Table

Table D.1: Correlation Matrix of Key Variables

	Acq. price P	Exp. return v_i	Acq. surplus V_i	Joint surplus V_t	Startup capture λ_s	VC, PE invest.	Acq. mkt cap	Science startup	Potential acquirers
Acq. price P	1.000								
Exp. return v_i	-0.063	1.000							
Acq. surplus V_i	0.014	0.069	1.000						
Joint surplus V_t	0.866	-0.019	0.513	1.000					
Startup capture λ_s	0.110	-0.389	-0.311	-0.061	1.000				
VC,PE invest.	0.367	-0.039	0.003	0.316	0.056	1.000			
Acq. mkt cap	0.118	-0.002	0.891	0.547	-0.260	0.038	1.000		
Science startup	0.081	-0.006	-0.005	0.067	0.060	0.025	0.012	1.000	
Potential acquirers	-0.030	0.006	0.065	0.006	-0.072	-0.013	0.083	-0.436	1.000

E Classifying Science-Based Startups with a Large Language Model

In this section, I describe in more detail how I construct and validate the classification of science-based startups—that is, whether a startup's products, services, or technologies rely on novel scientific research at the time of founding. This classification serves as a main input throughout the empirical analysis. The objective is to distinguish ventures grounded in scientific advances in the natural sciences or engineering from those commercializing either non-technological offerings or standard technologies, such as basic software, incremental product improvements, or technologies that were already widely adopted by the time the firm was founded. To that end, I use a large language model (LLM) to read unstructured descriptions of each startup and return a structured assessment. This approach yields a classification methodology that is replicable, consistent across cases, and scalable to large samples. Furthermore, for studying startups and as I show below, using unstructured text together with a large language model produces more accurate classifications than commonly used alternatives, such as indicators based on patent filings or scientific publications.

E.1 Methodology

Each classification is based on two inputs: (1) the founding year of the startup and (2) a block of descriptive text, which may include the company's website, news coverage, and SEC filings. The model is instructed to evaluate whether the technology was based on scientific advances that were still novel at the time of founding—not merely whether it was once novel, or is considered high-tech today. The model's output is structured, requiring to return a JSON object that contains exactly three fields:

- A score from 1 (very unlikely to be science-based) to 5 (very likely),
- A confidence level between 0 and 100, and
- A short reasoning string summarizing the evidence behind the classification.

Model and Inference Setup. To generate the classifications, I use the Llama 3.3 70B Instruct model, developed by Meta and released in December 2024. This is an open-weight, instruction-tuned model that accepts long inputs (up to 128,000 tokens), allowing it to read both the full evidence block and the detailed instructions in a single pass. I run the model using Ollama, which serves the model locally loaded through a simple API.[76] All inference is performed on a single NVIDIA A100 80GB GPU, which supports the memory demands of this model and context length.

While proprietary models such as OpenAI's GPT-4 (used in ChatGPT) or Anthropic's Claude 3 are also competitive in language understanding and reasoning tasks, Llama 3.3 70B performs comparably across many benchmarks and offers full transparency, reproducibility, and local deployment options. This is especially useful in academic settings, where control over the model environment, inference process, reproducibility, computational costs, and data access is critical. Moreover, the ability to run inference without API constraints enables fine-grained experimentation, such as schema-constrained decoding and automated retries, which are important for structured classification tasks like this one.

Each input to the model includes (1) a system prompt, which defines what constitutes a science-based innovation and how the founding year should affect the judgment (see the full prompt below), and (2) a user message containing the startup's founding year and descriptive text. The prompt explicitly instructs the model to return only a JSON object in a fixed format. To ensure consistency and ease of downstream processing, I enable Ollama's built-in JSON mode by setting format="json" in the API call. This ensures that the model is constrained to return a valid JSON object. On the receiving end, I use a typed schema validator to check the output, verifying that the score is one of the five allowed integers, that the confidence is a number between 0 and 100, and that no additional text is present within these fields. If the output is malformed, I automatically send one retry prompt asking the model to reformat its response.

[76] Ollama is an open-source tool that provides a local interface to run large language models like Llama 3, exposing a lightweight HTTP API that simplifies integration into Python or other environments. It supports custom decoding strategies (e.g., grammar-constrained output) and can operate entirely on local infrastructure.

All runs are done with deterministic decoding: temperature is set to zero, and all other parameters are fixed. This ensures that the same input always yields the same classification, which is important for reproducibility.

Prompt. The design of the prompt emphasizes temporal grounding: the model is instructed to assess whether the startup's technology depended on scientific work that was still novel at the time of founding. This avoids labeling as "science-based" those ventures that commercialize technologies that were once cutting-edge but had become widely adopted or commoditized by the time the firm entered. This distinction is important for my empirical strategy, which focuses on how the structure of commercialization markets affects value capture for new science. To support robustness and validation, I log the prompt, model version, and decoding parameters for each call. I also retain the model's reasoning field to enable human audits and interpretability. The prompt follows:

```
You are an expert in assessing science-based innovations. Your task is to evaluate
    whether the technology commercialized by a startup substantially relies on novel
    scientific innovations at the time of its founding. You will receive two inputs:
1. Startup Description: Unstructured text containing detailed information about a startup
    , including a description and unstructured text from its website, news coverage, and
    SEC filings. Carefully analyze all provided text to determine if the startups
    technology is based on innovations based on recent scientific advances. Make use of
    your long context window to analyze all the text.
2. Founding Year: An integer indicating the year the startup was founded.

Definition (Science-Based Innovations):
Science-based innovations significantly depend on the development and application of
    novel scientific advances typically emerging from life sciences, chemistry, physics,
    computer science, or engineering. Unlike incremental improvements or straightforward
    technological applications, science-based innovations represent substantial
    breakthroughs that meaningfully expand scientific knowledge and introduce novel
    solutions to previously unsolved or complex problems. Such innovations usually result
     from extensive research and development (R&D), involve specialized expertise, and
    require rigorous validation before commercialization. A science-based startup
    transforms such scientific discoveries (regardless of whether they originate
    internally or externally) into commercially viable products, services, or processes.
```

```
Clarification on Technological Innovations:
1. Clearly Non-Scientific Innovations: Examples include standard software applications, consumer apps, basic hardware such as most wearables, or incremental improvements to existing devices.
2. Clearly Scientific Innovations: Innovations involving cutting-edge or novel applications such as advanced biotechnology and pharmaceuticals, LiDAR navigation systems, groundbreaking battery chemistries, novel weaponry, or completely new materials and processes.
3. Borderline Cases: Certain technologies, like drones or robotics, may or may not be science-based. If a drone technology involves novel sensor systems, advanced LiDAR navigation, innovative propulsion methods, or revolutionary battery technologies at the time of founding, it should be classified as science based. However, if it merely provides incremental features, improved usability, or integrates existing commoditized components without substantial novel scientific breakthroughs, it should not be classified as science based.

Temporal Context (IMPORTANT):
Assess strictly based on whether the startups technology represented a scientific innovation at the specific time it was founded. Technological innovations built upon previously scientific advances that had become commoditized by the founding year should not be considered science based. Consider only the state of scientific innovation as of the founding year. For instance, Google's PageRank would qualify as a scientific innovation in 1999 but likely would not by 2024 standards.

Your Task:
Evaluate the startup based strictly on the criteria above, returning the evaluation exclusively in the following JSON structure:
{
  "score": <integer from 1 to 5>,
  "confidence": <numeric confidence score (0-100)>,
  "reasoning": "<concise explanation explicitly referencing key evidence from the description and clearly linking your assessment to the scientific context at the
```

```
    founding year>"
}

Explanation of Scores:
Score (1-5):
1: Very low likelihood of being science-based.
2: Low likelihood of being science-based.
3: Moderate likelihood of being science-based.
4: High likelihood of being science-based.
5: Very high likelihood of being science-based.
Confidence (0-100): Explicitly indicate your certainty in your score. Use lower values if
     evidence is unclear or insufficient.

Provide no additional text outside the JSON structure.
```

E.2 Classification Results

Table E.1 summarizes the distribution of LLM-based science classification scores across the sample. The classification is discrete, taking integer values from 1 to 5, with 1 indicating a very low likelihood that the startup relies on scientific research and 5 indicating a very high likelihood. The distribution is notably skewed, with over 70% of startups receiving a score of 1, suggesting that the majority of firms in the sample do not commercialize products or services based on recent scientific advances. This is consistent with expectations, given that most startups operate in sectors like software, services, or low-tech consumer products where scientific content is minimal or absent.

Importantly, approximately 19% of ventures are classified as likely or very likely to be science-based (receiving scores of 4 or 5). These firms are concentrated in sectors such as biotechnology, advanced manufacturing, materials science, energy, and semiconductors. In contrast, no startups in the consumer and business products and services category receive scores in this upper range, and only 18 software startups are classified as science-based. These software firms tend to fall into two specific groups: either early-stage startups from the early 2000s developing foundational software technologies, or more recent ventures focused on artificial intelligence and machine learning research.[77]

[77] This pattern raises a broader question about the role of scientific innovation in consumer-facing markets. If no startups in consumer products are classified as science-based, does that imply that science-based consumer innova-

Table E.1: LLM-based classification of startups by their reliance on scientific research. Scores range from 1 (very unlikely to be science-based) to 5 (very likely). The distribution is grouped into three meaningful bands: Low (1), Borderline (2-3), and High (4–5). This structure is used in subsequent empirical analysis to test robustness to classification thresholds.

Score	Frequency	Percent	Cumulative
1	4,350	70.74%	70.74%
2	308	5.01%	75.75%
3	301	4.90%	80.65%
4	978	15.91%	96.55%
5	212	3.45%	100.00%

This pattern is also visible in the kernel density plot (Figure E.1), which reveals a secondary mode around score 4. The bimodal shape of the distribution suggests that the classifier is not simply assigning noise or reacting to superficial features in the text, but is instead identifying meaningful variation in the underlying scientific content of the startups' technologies. The presence of two well-separated regions in the distribution—a dominant mass near score 1 and a smaller but pronounced tail near scores 4 and 5—is consistent with the underlying binary distribution assumed in the classification task.

Finally, 11% of firms receive scores of 2 or 3, reflecting borderline or ambiguous cases. Within this 11%, 48% are startups in energy, hardware, industrials, manufacturing, materials, or semiconductors, and another 33% are in life sciences. These sectors are often science-adjacent but heterogeneous: many firms in these categories commercialize technologies that rely on engineering or scientific infrastructure, but not necessarily on novel or non-commoditized discoveries at the time of founding. However, this pattern also implies that the overall classification results may be sensitive to how borderline cases are handled. Because these intermediate scores are concentrated in technically intensive sectors, small changes in classification thresholds could meaningfully affect the composition of the science-based sample (see next section for robustness analyses to account for these borderline cases).

Based on the distribution of scores and the underlying classification logic, I define startups as science-based if they receive a score of 4 or 5 from the language model, and non-science-based

tion is rare, or simply not pursued through startups? It is possible that firms commercializing scientific advances in consumer markets—such as novel materials in wearables, biomedical sensors, or advanced food technologies—are more often large incumbents or vertically integrated manufacturers. This suggests a potential asymmetry in how science reaches different end users: new scientific breakthroughs may diffuse into consumer markets, but not through entrepreneurial entry. Instead, incumbents may be better positioned to absorb and scale these innovations, particularly when manufacturing, regulatory, or distribution capabilities are required. Identifying concrete examples of science-based consumer innovations outside the startup ecosystem (e.g., Apple Watch blood oxygen sensors, Nestle's food tech R&D) could help clarify whether this absence in the startup sample reflects a market structure issue, a classification problem on the LLM side or on the industry side, or a broader innovation pattern.

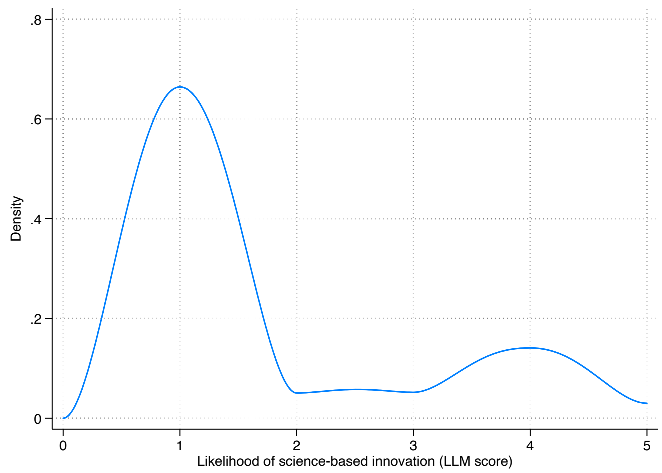


Figure E.1: Kernel density estimate of LLM-based science classification scores. The distribution exhibits two distinct regions: a dominant mode near score 1, corresponding to startups unlikely to rely on scientific research, and a secondary mode near score 4, indicating a smaller but concentrated group of science-based ventures. The bimodal shape suggests that the classifier is capturing meaningful variation in the underlying technological content of startups, rather than assigning noise.

otherwise. This threshold reflects a conservative classification strategy: only firms with strong, clear indications of reliance on novel scientific research are labeled as science-based. The intent is to reduce false positives—cases where general technical language or sector affiliation might be mistaken for genuine scientific content. This definition aligns with the conceptual goal of identifying ventures that meaningfully depend on scientific advances, rather than on engineering implementation, commoditized technologies, or abstract language. Thus, the results featured throughout the main analysis in this paper use this binary 4–5 cutoff as the primary science indicator.

E.3 Sensitivity of Main Results to Science-Based Classification Cutoffs

Nonetheless, it is important to acknowledge that the classification is not inherently binary, and that intermediate scores may contain meaningful information or reflect potential misclassifications. This section investigates whether the empirical results are robust to reasonable variations in the threshold used to classify startups as science-based. As discussed earlier, there are two conceptual challenges that motivate this analysis. First, any automated classification—particularly one using a language model to interpret unstructured descriptions—inevitably carries some degree of measurement error. Second, and more importantly, the concept of "science-based" is not inherently binary. Even if classification were perfect, it would still represent a latent continuous trait: some startups are clearly rooted in novel scientific research, others are not, and many fall somewhere in between. In this context, any binary cutoff—though helpful for interpretation and tractability—is a simplification.

These considerations motivate a set of robustness checks designed to evaluate the sensitivity of the main results to the specific threshold chosen. In particular, I explore how the results change when (i) expanding the science-based category to include borderline cases, and (ii) leveraging the full ordinal score produced by the classifier rather than imposing a binary split. These analyses serve both as a test of robustness and as a way to assess whether the classifier captures a meaningful gradient of science intensity.

These alternative specifications are reported in Table E.2. As shown in those results, the main findings of the paper are stable: the sign, magnitude, and significance of the coefficients on value capture and value creation do not meaningfully change across classification schemes. This suggests that the empirical conclusions are not driven by the specific treatment of ambiguous cases and are robust to reasonable variations in the science definition.

Table E.2: This table evaluates the sensitivity of the main results to different thresholds for defining a startup as science-based. Columns (1) and (4) reproduce the baseline specification used in the main analysis of the paper, which classifies startups with LLM-assigned scores of 4 or 5 as science-based, in a binary variable (Tight Binary [4,5]). Columns (2) and (5) apply a looser threshold, expanding the science-based group to include startups with scores of 2 to 5—adding 609 ambiguous cases (Loose Binary [2,5]). As expected, coefficients are smaller, but the effects remain statistically and directionally consistent. Columns (3) and (6) use the full 1–5 discrete, ordinal score, instead of a binary variable (Discrete [1,5]). In both dependent variables—value capture and creation—results remain in line with the main findings. All models include industry-year and startup country fixed effects.

	Value Capture			Value Creation		
	(1)	(2)	(3)	(4)	(5)	(6)
Tight Binary [4,5]	-0.147***			0.180***		
	(0.023)			(0.034)		
Loose Binary [2,5]		-0.072***			0.074*	
		(0.012)			(0.040)	
Discrete [1,5]			-0.029***			0.102***
			(0.006)			(0.016)
Constant	0.610***	0.603***	0.632***	18.633***	18.646***	18.490***
	(0.003)	(0.002)	(0.008)	(0.005)	(0.009)	(0.024)
Industry × Year FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Observations	5,823	5,823	5,823	5,823	5,823	5,823
R^2	0.057	0.054	0.054	0.095	0.095	0.096

Standard errors clustered at the industry-year and country level.

* p<0.1, ** p<0.05, *** p<0.01

When using the binary classification with a strict threshold (scores 4 or 5), the estimated coefficients are large, statistically significant, and substantively meaningful—these are the main results reported in the paper. In columns (2) and (5), I broaden the science-based category to include all startups with scores of 2 or higher. This more inclusive threshold brings an additional 609 startups into the science-based group, many of which are borderline or ambiguous cases. As expected, the estimated effects are attenuated: the coefficients fall between 50% and 60%, to 0.072 (column (2)) and 0.074 (column (5)). Nonetheless, both remain statistically significant. This pattern is consistent with a dilution effect. Under a looser definition, the science-based group likely includes ventures in science-adjacent sectors or those that use technical language but are not meaningfully driven by new scientific discoveries. Consequently, the differences in value creation and capture become more muted.

More formally, relaxing the classification threshold reduces the likelihood of false negatives—startups that are in fact science-based but receive a low score under the strict criterion. This improves recall and reduces type II error, which is desirable from the perspective of inclusive measurement. However, this comes at the cost of introducing false positives—startups that are not truly science-based but now fall above the threshold. This inflates the treatment group with firms that do not share the core scientific characteristics of interest, leading to attenuation bias in the estimated coefficients. The treatment effect is diluted, as the average within-group difference between science and non-science-based startups now narrows.

This mechanism helps explain the observed reduction in coefficient magnitudes when shifting from a strict (scores 4–5) to a loose (scores 2–5) binary split. Given that scores of 2 and 3 collectively account for 609 startups—more than a 40% increase in the number of "treated" units—this redefinition significantly alters the composition of the science-based group. The resulting estimates remain statistically significant and directionally consistent, suggesting the core effect is robust, but the average treatment effect (not causal) weakens due to the inclusion of borderline cases. Still, the fact that coefficients remain significant under both definitions reinforces the main empirical conclusion: startups with meaningful scientific content—however defined—perform differently on key commercialization dimensions, especially in their ability to generate value and capture rents.

Finally, columns (3) and (6) use the full 1–5 ordinal score as a continuous regressor. The results continue to show significant effects in the expected direction, providing further evidence that the classifier captures a structured and meaningful signal rather than noise. Furthermore, in this case, the magnitude of the coefficients is closer to that of the main binary variable. For example, in column (3), the coefficient on the ordinal science score is 0.029. This is broadly consistent with the binary effect in column (1), where the estimated coefficient is 0.147. Most startups classified as non-science have a score of 1, while science-based startups are concentrated at scores 4 and 5. Therefore, moving from a score of 1 to 4 corresponds to a four-times increase on the ordinal scale. Multiplying the ordinal coefficient by this gap (0.029×4) yields an implied effect of roughly 0.116—still slightly smaller, but approaching the binary estimate of 0.147. The gap is expected, as the ordinal regression distributes effects more continuously across the scale, while the binary split isolates the upper tail.

Taken together, these results underscore the robustness of the main findings to alternative operationalizations of science intensity. While the binary indicator based on scores 4–5 is intuitive and useful for exposition, the results hold under both a looser binary threshold (scores ≥ 2) and when using the full 1–5 ordinal score as a continuous variable. Importantly, the coefficient estimates remain statistically significant and directionally consistent across specifications. The continuous specification, in particular, lends further support to the validity of the classification: it captures a meaningful gradient in science reliance, and the estimated coefficients are of comparable magnitude to the binary model when interpreted over relevant score ranges.

E.4 Robustness of Science-Based Classification to Prompt Variations

In additional analyses not reported in the manuscript, I test the robustness of the science-based classification to changes in prompt design. Specifically, I assess whether the LLM-based scores are sensitive to variations in wording, structure, framing, and the temporal reference point used in the prompt. To do so, I follow best practices from the prompt engineering literature and adopt a strategy similar to that of Carlson and Burbano (2024), who show that prompt format can meaningfully influence downstream inferences in large-scale classification tasks. In particular, I systematically vary three dimensions of the prompt to evaluate the sensitivity of the resulting classifications.

First, I generate multiple alternative prompts using LLMs themselves, each reformulating the core task—assessing whether a startup's technology relies on novel scientific advances—using different phrasings, clarifying instructions, and types of examples. This includes both more concise and more verbose versions of the prompt, slight changes in the definition of "science-based" innovations (e.g., emphasizing novelty vs. research intensity), and alterations in the granularity of examples provided for borderline cases. Second, I vary how temporal context is introduced into the prompt. The baseline specification evaluates scientific novelty as of the startup's founding year. I test the sensitivity of results to shifting this reference point to the year of acquisition, acknowledging that what constitutes science may vary over time, especially when the time to exit is large. Third, I explore structurally distinct prompt templates.

Across all these variations, the results are remarkably stable. The classification distribution barely shifts, and the correlation of alternative score vectors with the baseline exceeds 0.90 in many cases. The set of ventures classified as science-based (scores 4–5) is virtually unchanged. As such, while the main results rely on a single prompt for simplicity and replicability, the broader classification exercise appears robust to reasonable changes in prompt design and framing.

E.5 Benchmarking the LLM-Based Classification with Patent and Publication Data

While the LLM-based classification demonstrates strong empirical performance in predicting differences in science reliance, it is useful to assess whether the score aligns with more conventional indicators of scientific content. Beyond manual validation—where I read and inspect the model's outputs and reasoning chains for selected startups—I conduct a systematic external assessment exercise using established innovation metrics: patents and scientific publications. This allows me to assess whether the LLM score correlates with observable and widely used science-linked outputs at scale.

Before presenting the results, it is worth revisiting why patent and publication data are not used as the primary classification input in this study. Patents and publications data remain foundational and very useful in innovation research and are widely used to track technological and scientific activity. However, their limitations become particularly salient when applied to startups, where such indicators often fail to reflect the underlying reliance on science. Many early-stage firms choose not to patent, even when their technologies are patentable, due to the high costs of filing and, more importantly, enforcing intellectual property rights—costs that are especially burdensome for resource-constrained ventures (Graham et al., 2009; Bryan and Williams, 2021). Instead, startups often rely on secrecy, speed, or first-mover advantages to protect their knowledge (Lerner and Seru, 2022). Furthermore, even when patents do exist, they may formally be held by inventors, universities, or venture capital firms, rather than the startup itself, complicating the task of linking IP to specific firms. Publications pose similar challenges: coverage is sparse, and affiliations are often misreported or missing entirely.

Despite these limitations, patent and publication data remain valuable for benchmarking the LLM classifier—especially for the subset of startups where reliable matches can be established. To do so, I evaluate the extent to which high LLM scores are associated with greater patenting activity and stronger linkages to science via patent-to-paper citations. I match startups to patents using the procedure developed in Nagar et al. (2024), which handles entity disambiguation and name variation in startup records. Patent-to-paper citations are derived using the dataset from Marx and Fuegi (2020), allowing me to trace the scientific roots of patented inventions.

As posed, the issues with patents identify science startups are non-trivial. For example, upon visual inspection, in my sample of 5,823 startups, 734 operate in Biotechnology—a domain with high patent propensity—yet only 417 of these (56.8%) are successfully matched to patents. The remaining 317 biotech firms (43.2%) have no patent records despite many being acquired by major pharmaceutical incumbents like Pfizer and Novartis, often for hundreds of millions of dollars. Indeed, manual inspection reveals that these firms are scientifically sophisticated and actively engaged in research. Their lack of patents reflects data incompleteness, not an absence of science. Thus, as posed, the LLM-based measure helps overcome this limitation.

Formally, I assess the external validity of the LLM-based science score using OLS regressions, with results presented in Table E.3. The aim is not to treat patents or publications as a definitive "ground truth", but rather to test for convergence: do startups that receive higher science intensity scores from the LLM also exhibit observable scientific activity, such as holding more patents or citing academic research in those patents? Across all specifications, the LLM-based score is strongly and significantly associated with both the number of patents and the volume of patent-to-paper citations. These relationships hold even after controlling for year, country, and—under stricter specifications—industry-year fixed effects. This provides robust evidence that the classifier is capturing a meaningful signal consistent with conventional bibliometric indicators, while offering broader coverage and finer granularity, particularly in the startup context where standard metrics are often missing or incomplete.

Columns (1)–(4) regress the LLM-based science score on patent-based measures across two samples: the full sample of startups (Columns 1–2) and the subset of firms matched with at least one patent (Columns 3–4). Across both samples, the results indicate a robust positive association between the LLM score and the log number of patents as well as log patent-to-paper citations, with all coefficients statistically significant. For instance, in column (1), the coefficient on the log patent count is 0.242, while in column (2), the coefficient on patent-to-paper citations is 0.246. These magnitudes are economically meaningful and suggest that startups classified as more science-intensive by the LLM tend to hold more extensive patent portfolios, with patents that more heavily draw on scientific research. The estimates remain positive and significant in the restricted sample of patenting startups.

In Panel B of Table E.3, I introduce industry-by-year fixed effects to account for cross-industry variation. That is, I estimate the relationship within industry-year cells, which provides a stricter test of the model's explanatory power by comparing startups operating in similar technological and temporal contexts. The results remain robust. The LLM score continues to be significantly associated with both patenting activity and patent-to-paper citations when using the full startup sample (columns (1) and (2)). In the restricted sample of "patenting startups" (columns (3) and (4)), the coefficient on patent count remains positive and significant, even after conditioning on industry

Table E.3: Startup reliance on science: LLM-based measure vs. patent-based measures.

Panel A: Specifications Without Industry Fixed Effects

	Sample: All Startups		Sample: Patenting Startups	
	(1)	(2)	(3)	(4)
Patent count (log)	0.242***		0.177***	
	(0.010)		(0.011)	
Patent-paper cites (log)		0.246***		0.104***
		(0.014)		(0.015)
Constant	1.481***	1.701***	1.692***	2.076***
	(0.010)	(0.002)	(0.025)	(0.005)
Year FE	Yes	Yes	Yes	Yes
Industry × Year FE	No	No	No	No
Country FE	Yes	Yes	Yes	Yes
N	6,133	6,133	2,760	2,760
R^2	0.111	0.043	0.072	0.043

Panel B: Specifications With Industry × Year Fixed Effects

	Sample: All Startups		Sample: Patenting Startups	
	(1)	(2)	(3)	(4)
Patent count (log)	0.064***		0.033***	
	(0.002)		(0.008)	
Patent-to-paper cites (log)		0.039***		0.000
		(0.007)		(0.012)
Constant	1.671***	1.734***	2.035***	2.114***
	(0.000)	(0.001)	(0.018)	(0.003)
Year FE	No	No	No	No
Industry X Year FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
N	6,114	6,114	2,738	2,738
R^2	0.625	0.620	0.634	0.632

Panel A clusters standard errors at the Year and Country level.

Panel B clusters standard errors at the Industry-Year and Country level.

* p<.1, ** p<.05, *** p<.01

and year. However, the coefficient on patent-to-paper citations becomes statistically insignificant. This attenuation is not surprising: once conditioning on industry-year, there is limited residual variation in citation intensity across startups that all already hold patents. Moreover, patent-to-paper citation data is noisier and less comprehensive at the firm level, and its distribution is highly skewed. To illustrate these results, Figure E.2 plots the correlation between the LLM-based science score and the log number of patent-to-paper citations, without the industry fixed effects (Panel A, column (4)). The figure reveals a strong and approximately linear relationship: startups with higher

LLM-assigned science scores tend to have a greater volume of scientific citations embedded in their patents. This pattern reinforces the view that the classifier captures meaningful underlying scientific content.

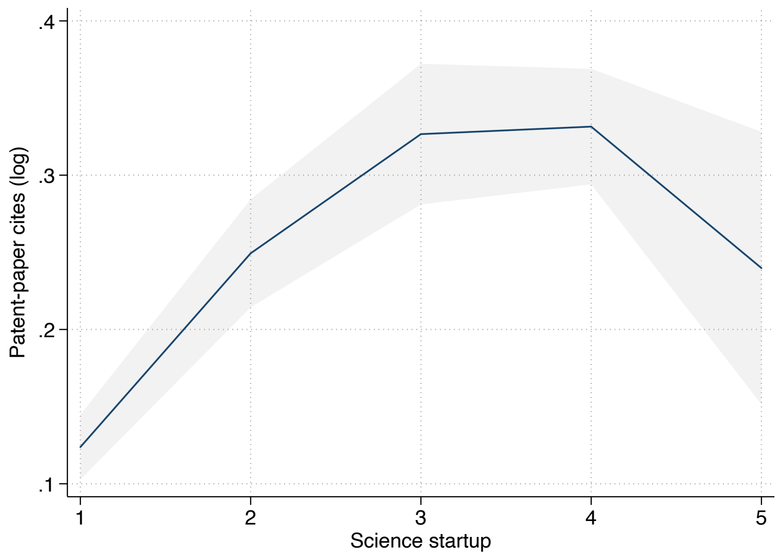


Figure E.2: Correlation between reliance on science measures: LLM *vs.* patent to paper cites.

Taken together, these results suggest that the LLM-based measure of science reliance is well-aligned with standard indicators and can serve as a valid proxy for scientific content at scale. Importantly, the measure retains explanatory power even in specifications that control tightly for industry and temporal heterogeneity. The consistent statistical significance of the coefficients across specifications reinforces the notion that the LLM classifier is not simply replicating patent-based signals but rather identifying structured, granular information embedded in unstructured textual data.

Moreover, given the known limitations of patent data for startups—such as under-patenting and measurement error—the LLM-based approach offers a scalable and potentially more inclusive alternative. This is particularly relevant when trying to capture early-stage or non-traditional forms of measuring scientific innovation that may not be reflected in patent records.