Holding Up Innovation: Science Startups and the Problem of Value Capture

Roger Masclans*

[Preliminary and incomplete—do not circulate]

March 27, 2025

Science-based innovations increasingly rely on startups to diffuse to markets. Yet, in critical areas such as energy, industrials, and materials—crucial for addressing social challenges—entrepreneurial activity remains limited. This issue is often associated with high risks, low value, or misaligned financing models. I propose a complementary explanation: reduced value capture driven by a hold-up problem. Startups commercializing scientific innovations often cannot scale independently and depend on a few incumbents, limiting outside options and weakening bargaining power. Empirically, I estimate value creation and capture by examining stock market reactions to startup acquisitions, filtering out unrelated noise. The study reveals three main findings. First, estimates show that in industries where hold-up is more severe, science-based startups capture 46 cents per dollar of surplus generated, compared to 70 cents for non-science ones—a 34% penalty. Second, these effects are strongly associated with outside options. Third, startups with lower capture create more value, offsetting the capture penalty and explaining entry. The findings suggest that hold-up is pervasive for science-based innovations, reducing upstream incentives for investment and potentially leaving socially valuable innovations underfunded. Policy and management should focus on fostering competitive markets for technology and acquisitions, as well as enhancing independent commercialization pathways.

*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. I would also like to thank Sharon Belenzon, Ines Black, and Jay Prakash Nagar, as well as helpful feedback from seminar participants at Wharton Innovation Doctoral Symposium.

1 Introduction

Innovations grounded in scientific discoveries increasingly rely on startups to bring technologies to market, as a growing share of commercially relevant research now occurs outside incumbent firms (Arora et al., 2018; Fleming et al., 2019; Kolev et al., 2022). In addition to driving economic growth,[1] startups commercializing scientific innovations often enable transformative rather than incremental 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).[2] Yet, despite their significance, concerns are raised regarding limited entrepreneurial activity in critical areas such as energy, industrials, and materials (IEA, 2020; Moss et al., 2020; Veugelers, 2024). For example, Lerner and Nanda (2020) document that between 2015 and 2019, startups in telecommunications, networking, computer hardware, materials, and energy industries received only about 5 percent of total Venture Capital (VC) investment, with the majority of funds flowing to software, consumer, and business products and services—and, to a lesser extent, the life sciences.

Research examining why a broad range of science-based technologies attract fewer resources than their counterparts despite their social need has primarily focused on two related explanations. The first explanation argues that financing models are often inhospitable to science-based innovations, as these require significant upfront capital, entail high technical risks, and necessitate long time horizons before reaching commercial proof-of-concept (Hall and Lerner, 2010; Kerr and Nanda, 2015). This has been rationalized as a rigidity in experimentation (Kerr et al., 2014), effectively reducing the option value of investing in science-based startups and thereby reducing their risk-adjusted expected returns (Ewens et al., 2018; Nanda and Rhodes-Kropf, 2017).[3] The second explanation argues that lower market demand for certain technologies constrains activity further (Van den Heuvel and Popp, 2023), since market size is crucial for incentivizing upstream innovation and entry (Acemoglu and Linn, 2004). At face value, these arguments suggest that the problem is one of value creation. Certain ventures attract fewer resources because other opportunities offer higher expected returns driven by a greater overall economic value—whether in absolute terms or after adjusting for risks and frictions.

In this paper, I pose an empirically test an alternative explanation: A problem of value capture. Science-based startups can indeed be economically and socially valuable due to their novelty and potential market impact. However, I argue that they systematically struggle in capturing the value

[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).

[2] For instance, startups commercializing advances in chemical engineering have led to more efficient battery technologies, spurring the proliferation of modern devices ranging from smartphones to electric vehicles to advanced medical equipment. Similarly, materials science startups have allowed for the creation of next-generation aerospace and construction technologies, while molecular biology ventures have driven the development of more effective, targeted therapies.

[3] Although these constraints are common in many industries dependent on scientific advances—such as aerospace and defense, agriculture and water technologies, energy, manufacturing and industrial equipment, and materials—they are not applicable to all innovations. For example, innovations in the life sciences benefit from significantly lower commercial risk due to the availability of epidemiological data, which provides clearer demand signals ex-ante. Additionally, the technical risk and long development time horizons are mitigated by the structured process of clinical trials, which offer well-defined milestones (e.g., Phase 1, 2, and 3 trials). These well-defined validation stages, along with effective patent protections, enable robust markets for technology, in which innovations can be commercialized more efficiently (Arora et al., 2022).

of their innovations due to weak bargaining power at the time of exit. This weak capture is not only a function of risk, capital, or timing, but especially of exit conditions and structural disadvantages in how science-based startups interact with downstream markets, given the division of innovative labor in the current ecosystem.

Science-based startups often lack the capabilities to independently commercialize at scale, such as manufacturing capacity, regulatory expertise, and distribution networks. These complementary assets are both expensive to build in-house and rarely available through subcontracting markets (Aggarwal and Hsu, 2009; Gans et al., 2002; Marx et al., 2014; Moeen, 2017).[4] As a result, startups rely on downstream incumbents to diffuse their innovations and realize their full economic value (Teece, 1986)—typically through mergers and acquisitions (Ederer and Pellegrino, 2023; Lerner and Nanda, 2020). However, this reliance on incumbents introduces two additional frictions rooted in the transfer process that disproportionately affect science-based ventures.

First, I pose that the nature of science-based technologies alone makes these startups particularly susceptible to a hold-up problem (Grossman and Hart, 1986; Holmstrom, 1989). Such ventures typically have incurred in substantial sunk costs by the time the technical risk is sufficiently resolved so that incumbents are willing to acquire them, while they also operate under significant uncertainty regarding the eventual market application of their innovations.[5] When transferring their technologies via acquisition, these startups face a weak bargaining position due to the lack of credible outside options, i.e., independent commercialization, which significantly limits their ability to capture the value of the innovation.

Second, while this hold-up problem could, in principle, be mitigated by a broad pool of potential acquirers, I argue that it is in fact exacerbated in the case of science-based ventures. These startups face a narrower set of prospective buyers than their counterparts, as only a limited number of incumbents possess the specialized capabilities required to scale such innovations. Moreover, in many critical areas, these incumbents tend to operate in highly concentrated markets, further shrinking the pool of viable acquirers. In sum, when dependence on a few, large incumbents in concentrated industries is combined with the substantial upfront capital required for technology development, science-based startups are left with weak bargaining power. As a result, they capture only a small share of the value they create—a dynamic that ultimately reduces upstream incentives for entry and investment, potentially leaving economically valuable and socially desirable innovations underfunded (Holmstrom and Roberts, 1998; Scotchmer, 1996, 2004).

Needless to say, value creation and value capture problems are not mutually exclusive, and both can hold at the same time. However, distinguishing between the two is important. If the problem is one of value creation, limited investment in science-based startups may be justified by low expected returns or high inherent risk. This perspective often treats underinvestment as a feature of the innovation itself, abstracting from broader structural dynamics. To some extent, one could argue that

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

[5] In principle, this risk could be reduced through detailed ex-ante contracts, but these are often unfeasible due to uncertainty about how the innovation will be used or who the most suitable acquirer will be (Bresnahan and Gambardella, 1998; Gans and Stern, 2000; Kapoor and Klueter, 2021; Gambardella et al., 2021).

even if the same innovation were developed within an incumbent firm, it would still face similar investment hurdles—since the expected returns would remain low due to the intrinsic characteristics of the innovation. In this view, the problem is not who undertakes the innovation, but rather the nature of the innovation itself, which limits incentives across the board. In such cases, if the innovations are socially desirable, policy interventions might focus on mitigating risk or directly increasing the value created—for example, by subsidizing upstream R&D or stimulating downstream demand.

By contrast, if the problem lies in value capture, there is systematic underinvestment risk not only in socially valuable innovations, but also in economically viable ones—not due to a lack of potential, but as a consequence of the structure of the innovation ecosystem. Since entry decisions are based on private rather than social returns, distortions in value capture can misallocate innovative effort. When the first innovator expects to retain only a small share of the value upon transfer, the project may be abandoned—even if the innovation is economically valuable upon transfer. The result is a market failure: high-potential innovations go unfunded not due to a lack of overall value, but because the *first* innovator only cannot secure a sufficient return.

Consider, for example, a venture commercializing a novel method of producing non-fossil energy, such as nuclear or renewables. Bringing such technology from laboratory success to commercial scale requires enormous upfront capital to develop the necessary capabilities—e.g., constructing gigawatt-scale solar panel manufacturing plants, developing extensive wind-turbine distribution and maintenance networks, or navigating complex regulatory approval processes associated with nuclear energy. Moreover, unlike incremental innovations, these science-based ventures attempt to validate new scientific principles in real-world settings and at scale, adding layers of technical and market risk (Arora et al., 2024).[6] When these risks and financing frictions are factored into the venture's expected returns, the result may be relatively modest risk-adjusted payoffs. Unless the total economic value generated is sufficiently large, the technology may remain unfunded—reflecting the value creation problem.

Yet, even if such ventures do create substantial value, they face the additional challenge of constrained exit conditions. Lacking these specialized capabilities to commercialize energy technologies independently, the startups rely on downstream incumbents for deployment at scale. However, by the same token, only a small set of incumbents possess these requisite assets, many operating in concentrated markets. For instance, Nanda et al. (2015) find that the top 20 renewable-energy firms hold 40% of all patents, demonstrating significant industry concentration. Moreover, renewable-energy technologies tend to be highly specialized; a startup focused on solar photovoltaics is unlikely to pivot toward wind or nuclear power. Consequently, its pool of viable acquirers is small, limiting its bargaining position and thereby reducing the share of value it can capture.

The inability to commercialize independently, combined with industry concentration among potential acquirers, significantly restricts startups' bargaining power and reduces their prospects for value capture. Consequently, entrepreneurs and investors tend to shy away from these ventures, concentrat-

[6] On the one hand, the technology faces high technical risks when scaling unproven processes: it must demonstrate both competitive unit economics and that the performance and efficiency gains validated at small scale persist at full scale. On the other hand, significant market risks—from macroeconomic trends to the challenges of identifying and effectively reaching potential customers (e.g., data centers, steel manufacturers, retail, or all of the these)—obscure the commercialization pathway (Bresnahan and Gambardella, 1998; Kapoor and Klueter, 2021).

ing their activity in a narrower set of opportunities where the potential for hold-up is less severe—such as smart-grid technologies and related software solutions, which are largely incremental and, although they yield efficiency and productivity gains, fall short of being truly transformative (Nanda et al., 2015).

Empirically, I focus on demonstrating that, conditional on entry, science-based startups disproportionately face these conditions. First, I show that science-based startups in all industries but biotechnology capture less value than their non-science counterparts. Second, I show that this effect is strongly associated with a lack of outside options and, especially, with the pool of potential acquirers. Compared to their counterparts, science-based startups typically face a narrower pool of potential acquirers, which tend to be larger firms operating in more concentrated industries. I test these hypotheses relying on M&A transactions data. Consistent with the literature, I assume that M&A is the most common exit route for ventures and, therefore, the main mechanism by which the first innovator realizes the value (e.g., Ederer and Pellegrino, 2023).

I use data from 4,854 startup acquisitions by U.S. publicly listed firms. In the analysis, the acquisition price and stock market reaction of the acquiring firm serve as key measures to estimate the total private value created by the startup and the incumbent.[7] With these, I next compute the share of the total value that each captures. This methodology allows me to estimate the total surplus generated upon acquisition—i.e., the total economic, private value of a startup—as well as how these rents are shared between the two parts.

One problem with stock market prices is that they often mix the intrinsic acquisition surplus with unrelated firm-level events, competitive threats, or macroeconomic shocks. To address this issue, I adopt a parametric approach that refines the value estimates by isolating the signal from the noise. I do so by following the methodology developed by Kogan et al. (2017), but tailored for the M&A context. Namely, I acknowledge that acquisitions can lead to negative abnormal returns if investors perceive overpayment, 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.

In regard to outside options, these can take two forms: 1.) The ability of a startup to independently develop and commercialize its own innovations; and 2.) the number of incumbents possessing the requisite commercialization capabilities. To construct these measures, I employ existing data on downstream market competition developed by Hoberg and Phillips (2010, 2016) and commonly used to assess competitive overlap, product market rivalry, and potential acquirer relationships in studies of market competition, M&A, and vertical integration (e.g., Cunningham et al., 2021; Frésard et al., 2020). This dataset uses text-based analysis of the product descriptions in firms' 10-K filings to construct pairwise measures of product-market similarity across all publicly listed firms. Firms are embedded in a high-dimensional product space, where proximity reflects the degree of overlap in their product offerings. I use a recent version of this dataset released in 2024 which applies word embedding

[7] The abnormal stock market reaction of the acquiring firm around the acquisition announcement reflects the market's expectations of future cash flows generated by combining the startup's technologies with the incumbent's existing capabilities, providing an estimate of the total private value appropriated by the incumbent and, crucially, accounting for the risks related to both technical feasibility and market adoption at the time.

techniques to improve the precision of product similarity measures (Hoberg and Phillips, 2025). These data allow me to construct a firm-year level measure of the number of incumbents that are product-market similar to a given acquirer, thereby approximating the set of potential buyers for a startup's innovation at the time of acquisition.

The results indicate, first, that startups commercializing scientific innovations systematically capture less value than their counterparts, with conservative specifications estimating a 20.2% value capture penalty for science based startups. Furthermore, these results mask 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 most striking results emerge in industries with a mix of both startup types, such as energy, industrials, materials, hardware, and semiconductors—industries precisely where concerns about a dearth of activity are particularly pronounced. These mixed industries exhibit the largest value capture penalties: on average, science-based startups capture only 46 cents of every dollar of surplus generated upon acquisition, compared to 70 cents for their non-science-based counterparts—a 34% penalty.

Importantly, the higher value capture penalty associated with science-based startups would suggest that the entry threshold for startups commercializing scientific innovations should be higher. Indeed, I show that, in my sample, science-based startups generate significantly more value overall, with an estimated 27.3% increase compared to non-science startups. This finding underscores the economic and technological relevance of science-based startups while also highlighting the entry penalty they face. The higher threshold for entry suggests that only those science-based ventures capable of generating substantial value are able to attract resources.

Second, I identify limited outside options as a potential mechanism of the value capture disadvantage for science based startups. The results indicate that science-based startups tend to face a more limited set of potential acquirers compared to their non-science counterparts, with these startups having approximately 9.11% fewer buyers on average as per this measure. The pattern likely reflects the highly specialized and nascent nature of the technologies they commercialize, which require specific capabilities that not all incumbents possess. Likewise, I find that an expanded pool of potential acquirers substantially improves a startup's value capture, especially for science-based ones. An increase of one standard deviation above the mean in the number of potential acquirers is associated with a 25.70% increase in value capture for science-based startups, compared to only 4.18% for non-science startups.[8]

These 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 (Arora et al., 2018; Fleming et al., 2019) in favor of deploying their downstream capabilities to commercialize innovations. As these large firms focus on growth, scale, and efficiency, while simultaneously becoming more vertically integrated (Frésard et al., 2020), they contribute to higher market concentration and power (Antón et al., 2024; Cunningham et al., 2021), potentially

[8]The current version of the manuscript reports only correlational results and does not include any causal analysis. This will be included in future versions.

diminishing the competitive pressure necessary to incentivize upstream innovation (Scotchmer, 2004; Sutton, 1991).

Policymakers should consider interventions that not only bolster R&D funding, mitigate technological risks, and stimulate demand, but also promote more competitive markets for technology and acquisition markets, or initiatives to support alternative commercialization pathways that can facilitate startups to scale independently, from the use of National Labs or pilot plants for prototyping and scaling manufacturing to the use of shared distribution platforms. For instance, in battery technology, a startup with modular production techniques or access to shared manufacturing facilities could credibly threaten independent commercialization, strengthening its bargaining position and securing better acquisition terms. Enhancing such alternative pathways can incentivize upstream efforts towards socially relevant, missing innovations.

Similarly, managers in both incumbent firms and startups should explore collaborative strategies that offset the adverse effects of a lack of outside options and downstream market concentration. Moreover, if incumbents systematically suppress value capture for external innovators, they not only discourage potential innovation from emerging firms, but also leave themselves with fewer options to access critical technologies in the long run necessary for long-term competitiveness and growth.

The paper contributes to research on the economics of innovation and entrepreneurship. It joins a few recent papers arguing that value capture problems, rather than high capital requirements or long time horizons alone, limit the viability of science-based startups (Arora et al., 2024; Dalla Fontana and Nanda, 2023). Additionally, it extends theories of the boundaries of the firm and transaction costs by empirically demonstrating how hold-up problems impact commercialization across industries, emphasizing the role of market concentration as well as downstream capabilities in shaping innovation outcomes. Methodologically, first, it introduces a novel approach using Large Language Models and textual data to classify startups based on their reliance on scientific research. This improves upon traditional patent-based measures and offers broader applications for studying startup innovation dynamics. And, second, to the best of my knowledge, this paper is the first to apply a parametric approach, adapting the methodology first developed by Kogan et al. (2017), to isolate the market signal in acquisitions conducted by listed firms.

2 Theoretical Framework

Innovation is mostly an economic activity and, as such, resources are often allocated based on market rewards—that is, towards innovations that generate the highest economic returns (Bryan and Williams, 2021; Schmookler, 1966). In theory, if the market is efficient, one could reasonably assume that private incentives align with economic value, ensuring that innovations that create the greatest total value receive funding and progress toward commercialization without distortions.

However, the current innovation ecosystem is increasingly characterized by sequential innovation, where multiple actors contribute at different stages of the commercialization process and innovations are transferred between them (Arora et al., 2018; Fleming et al., 2019). In such a system, where one set of actors develops early-stage innovations and another set scales and diffuses them into markets,

the required transfer can create a misalignment of incentives (Gans and Stern, 2000; Scotchmer, 2004). The total economic value generated by the innovation must now be split between the parties, and the first innovator must capture a sufficient share upon transfer to justify entry and investment (Adner and Kapoor, 2010; Arora et al., 2024; Scotchmer, 1996).

Since entry decisions are based on private rather than social returns, distortions in value capture can lead to a misallocation of upstream innovative efforts. If the first innovator expects to capture too small a share of the rents upon transfer, the project may not be pursued. In such cases, insufficient expected returns drive innovators to reallocate resources to alternative projects with higher private payoffs. The result is a market failure, where high-value innovations go unfunded not because they lack economic potential, but because the first innovator cannot secure an adequate return under the expected transfer.

In this paper, I argue that scientific innovations are especially subject to incentive distortions rooted in the sequential nature of the ecosystem, explaining, in part, why there is little activity compared to non-scientific innovations. I study the issue in one key channel where the problem may be more acute: Startups. Startups drive early-stage innovation, conducting R&D, developing new technologies, and demonstrating feasibility, while incumbents hold the complementary capabilities necessary for large-scale commercialization. These capabilities—such as manufacturing infrastructure, distribution networks, and regulatory expertise—are often essential to diffuse an innovation to market efficiently (Teece, 1986). As a result, many startups are designed from the outset for eventual transfer to an incumbent through acquisition or licensing rather than scaling independently and competing in the product market (Gans et al., 2002).[9]

The problem originates because startups commercializing scientific innovations often lack outside options. Certain scientific discoveries require highly specialized capabilities for large-scale commercialization challenging to develop in-house and, thus, with little ability to scale commercialization independently. Furthermore, and relatedly, only a small number of incumbent firms can offer the downstream capabilities required to scale manufacturing and commercialization. This lack of outside options, combined with large, sunk upfront R&D investments required to develop the technology and the inability to contract ex ante, creates a hold-up problem.[10] Startups commercializing these innovations enter transfer negotiations with weak bargaining power, which translates into systematically lower capture.

To structure these ideas, consider a startup's decision to commercialize an innovation. The startup decides whether to enter based on the expected returns from transferring its innovation to an incumbent. Since the focus is exclusively on innovations that are transferred, the relevant question is how the surplus from commercialization is divided between the two parties and whether the startup anticipates capturing a sufficient share to justify entry. Upon transfer and further commercialization under an incumbent, the innovation generates total economic value (i.e., surplus or rents), V_I. The startup realizes the returns via the price received in the transaction, P and, thus, captures the fraction of the total value $\lambda_S = \frac{P}{V_I}$—while the incumbent captures the remainder, $\lambda_I = 1 - \lambda_S$. That is, the

[9] Also other reasons for transfer, such as monopoly profits being larger than the aggregate profits of a duopoly.

[10] See Gans et al. (2002) for related discussions on ex-ante contracting. A key issue is that uncertainty over the final application makes contracting infeasible, further reinforcing the hold-up problem.

startup's returns from entering and subsequently transferring the innovation are given by $P = \lambda_S V_I$. The advantage of this simple decomposition is that it explicitly characterizes the share of total value the startup is able to capture over the total economic value it creates, rather than simply treating the acquisition price as a proxy for economic value.[11] The startup enters if the expected payoff from transfer ($\lambda_S V_I$) exceeds its reservation value $\bar{R}$, which represents the opportunity cost of pursuing commercialization—or the necessary investment to realize the payoff. The entry condition is thus:

$$\lambda_S V_I \geq \bar{R} \tag{1}$$

This condition highlights the role of λ_S in shaping entry incentives. If λ_S is low, meaning the startup expects to capture only a small fraction of the total value it generates, there are two possible outcomes. First, if the expected returns from transfer fall below $\bar{R}$, the startup does not enter, even if the innovation has high potential V_I. Second, entry can still occur, but only if V_I is sufficiently large to compensate for the small share captured by the startup. In this case, the startup requires a higher total value to offset the low λ_S, effectively raising the bar for entry. As a result, valuable innovations may not be pursued due to weak value capture, while lower-value innovations with higher λ_S may still enter, leading to a misalignment between private and socially desirable innovation incentives. Table 1 summarizes the idea.

Table 1: Value entry thresholds conditional on value capture and reservation value $\bar{R}$ (investment).

	Low Capture (λ)	High Capture (λ)
Low $\bar{R}$	V	V^-
High $\bar{R}$	V^+	V

Note that this framework is agnostic to whether a startup is commercializing a scientific or non-scientific innovation and does not consider any parameter explicitly tied to it. The key point here is that value capture, λ, is primarily influenced by broader parameters related to outside options, and, while these operate for any startup in the same way, science startups tend to systematically have lower outside options.

A startup's value capture is determined by its bargaining power, which, in my setting, depends on two factors: the number of incumbents in the market and the startup's ability to credibly threaten independent scalability. Consider the baseline bargaining power as $\theta = \frac{n}{n+1}$, where n is the number of incumbents. When $n = 1$, the startup faces a monopsony, and $\theta = \frac{1}{2}$. This reflects the intuition that with only one buyer, the startup and the incumbent split the surplus equally in the absence of other

[11] It is important to distinguish what is often referred as a startup's "valuation"—the price paid to its shareholders during investment rounds or at exit—from its total value created (economic value, social value, surplus...), which encompasses the broader economic or social benefits generated by the startup and its technology. Under a simple framework, one might treat valuation as a proxy for total economic value (e.g., Total value = $\alpha \times$ valuation, where α is a constant multiplier across startup types). However, the premise of this paper is, precisely, that α is not constant and varies substantially between science based and non-science ventures and, thus, valuation is not a reliable proxy and more elaborated measures for value are required.

factors.[12] As $n \to \infty$, competition among incumbents intensifies, and $\theta \to 1$, meaning the startup can effectively capture the entire surplus. This framework captures the idea that greater competition among incumbents improves the startup's bargaining position by forcing incumbents to offer more favorable terms.

In addition to baseline bargaining power, the startup's ability to credibly threaten independent scalability strengthens its position during negotiations. While the startup does not commercialize independently in equilibrium, the threat to do so compels the incumbent to concede a higher share of the value. Thus, the startup's value capture is given by

$$\lambda = \max\left(\theta, \frac{V - \delta}{V_I}\right), \tag{2}$$

where $\frac{V-\delta}{V_I}$ represents the fraction of value capture required to the incumbent gained from the credible threat of scalability. δ represents the cost the startup would incur to scale the innovation independently and V the total value it would generate alone, which, in this setting, is assumed to be lower than that generated via transfer to an incumbent ($V < V_I$).[13] When $\delta \to \infty$, the threat is not credible, this term goes to zero, and the startup captures only θ, the baseline share—which depends only on the number of potential buyers. Conversely, when δ is small, the scalability threat is effective, and the startup captures more than its baseline share.[14] For example, a startup developing next-generation batteries might negotiate with a vehicle manufacturer. If the startup has relatively low scalability costs, such as access to shared manufacturing facilities or modular production techniques, it can credibly threaten to scale independently, forcing the incumbent to offer a more favorable deal.

Recall that the startup decides to enter if its payoff from transferring to the incumbent meets or exceeds the entry threshold $\bar{R}$. Substituting the expression 1 for λ, the threshold total value V_I^* required for entry is

$$V_I^* = \max\left(\frac{\bar{R}}{\theta}, \delta + \bar{R}\right). \tag{3}$$

If δ is high, the threat of scalability is not credible, and the entry threshold depends on the baseline bargaining power. In this case, $V_I^* = \frac{\bar{R}}{\theta}$, meaning entry becomes more likely when θ is large—and so the number of potential buyers. Conversely, if δ is small, the startup's threat to scale independently improves its value capture, and the entry threshold is $V_I^* = \delta + \bar{R}$.

Several conclusions follow from this framework. First, independent scalability costs δ play a critical role in determining the startup's value capture—and entry. In industries with low scalability costs, such as software development, startups can use the threat of scalability to negotiate better terms with incumbents, reducing the entry threshold. However, in industries like advanced manufacturing or

[12] This is, of course, a broad simplification. The current framing, in development, abstracts from additional complexities—such as incumbents' outside options and other considerations that may arise under monopsony conditions—that could further influence acquisition outcomes.

[13] The relationship between δ and the number of buyers n algins with intuition. If there are more buyers—n is large—, then θ increases. For the second term to be relevant, then δ must be smaller.

[14] The idea here is very simple. If the startup alone can generate value V and has to incur in δ for scaling commercialization, the payoff to do so is $V - \delta$. Thus, the required payoff to transfer the innovation to the incumbent is, at least, $\frac{V-\delta}{V_I} \times V_I$, where the first term represents the capture share.

energy, where independent scaling not only requires significant capital investment but also capabilities such as manufacturing or distribution that are not reasonably at hand, the threat is less credible, and startups rely primarily on baseline bargaining power. Second, the number of potential acquirers, n, affects the startup's bargaining position. Greater competition among incumbents, i.e., a lower degree of market concentration, reduces the capture problem, improving the startup's share of the value captured and lowering the threshold V_I^*.

Finally, in equilibrium, startups that enter and transfer their innovation satisfy $\lambda V_I \geq \bar{R}$. Startups with weaker threats of independent commercialization (high δ) or fewer incumbents to sell to (low n), that is, with prospects of low value capture, require higher V_I to enter, consistent with the idea that low value capture must be offset by higher total value creation. That is, the value capture penalty faced by science-based startups implies a higher entry threshold for these ventures—only the most promising science-based startups attract investment and move forward. In other words, while the hold-up problem may discourage some potentially valuable innovations, the ones that do enter are those that generate enough value to offset the expected capture penalty. Figure 1 represents these relationships.[15]

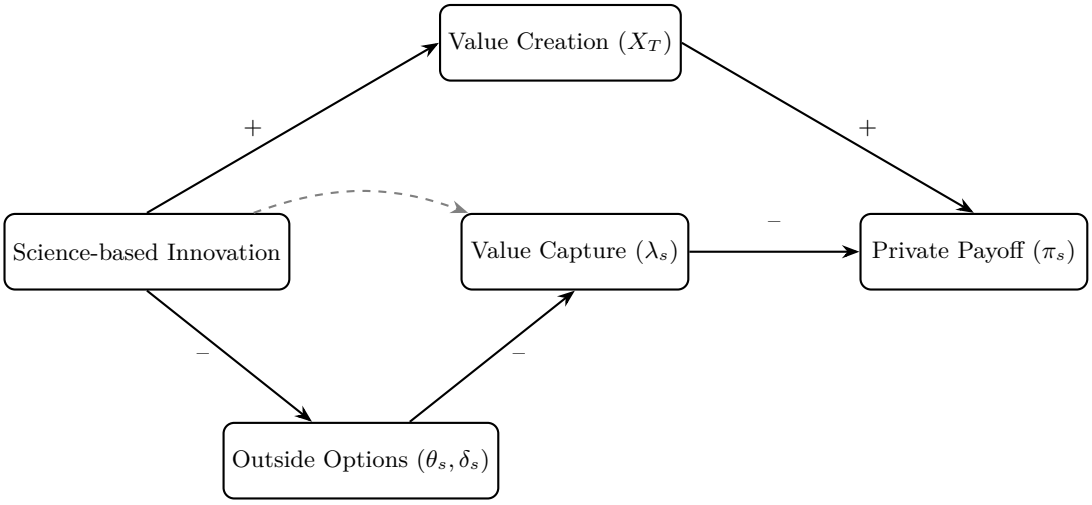


Figure 1: Directed acyclic graph representing the relationships between science-based innovation, outside options, value capture, value creation, and payoffs. Science-based innovations capture less value, but this effect is mitigated by their limited outside options. In turn, if one conditions on a sample of startups that enter, the value created must be higher so that, in equilibrium, choices remain efficient and private payoffs are equalized.

The framework is consistent with the activity patterns that prior research highlights, explaining why we observe more startups in some industries than others, especially when these startups are subject to hold-up. It also highlights the importance of practices that reduce scalability costs or foster competition among incumbents to promote entry and innovation. For example, shared production facilities in the energy sector could lower δ, enhancing the credibility of scalability threats and improving startups' value capture. Similarly, attenuating monopolistic behavior among incum-

[15] Other relationships not described here may be in place. For example, value creation may influence value capture. For now, I abstract from these.

bents around certain acquisitions could increase n, reducing the entry threshold and fostering a more dynamic innovation ecosystem.[16]

3 Data and Measurements

This section presents the data and measurement construction for the empirical analysis. Section 3.1 describes the theoretical methodology for estimating value creation and capture for each startup in the sample using stock market reactions, as well as the methodology employed to filter out noise; Section 3.2 details the data sources; Section 3.3 details the empirical procedure to estimate value creation and capture. Finally, Sections 3.4 and 3.5 detail respectively the classification procedure to determine whether a startup commercializes scientific innovations and the measurement of a startup's outside options.

3.1 Estimation of Value Creation and Capture

To empirically study the question, I develop a simple methodology for estimating the total economic value of a startup and the share of this value captured by the startup and the incumbent. I rely on acquisitions of startups by publicly traded firms. Such transactions provide a natural framework to measure the estimates because they reveal two key observables. First, acquisitions are the dominant exit mode for startups, especially after 2000 (Ederer and Pellegrino, 2023). Given that early-stage investors and founders aim at a successful exit and that M&A is the highest expected exit route, focusing on M&A transactions aligns with their ex-ante incentives and long-run objectives. Furthermore, this empirical setting is suitable for estimating value capture because acquisitions finalize the process by which a startup's shareholders realize the economic returns from their venture via the price they receive. Because the acquisition price is typically reported, this allows for a direct measure of the startup's absolute captured value at the time of exit (X_s).

Second, by focusing on publicly traded firms, I can rely on an event study: the incumbent's abnormal stock market reaction around the acquisition announcement, r_i, helps in estimating the total private value that the incumbent appropriates by integrating the startup under its umbrella (X_i). The assumption is that the stock market reaction reflects the market's expectation of future cash flows generated by combining the startup's assets—e.g., its technology, intellectual property, human capital, inventive capabilities, or customers—with the incumbent's existing assets. Furthermore, crucially for my theory, this estimate (X_i) captures the expected benefits while also discounts the risks associated with technical feasibility and market adoption.

Given the value captured by the startup (X_s) and the incumbent (X_i), the total value created by

[16] A premise of both the theoretical framework and the empirical analysis that follows is that science-based startups characteristically face a limited set of outside options for commercializing their innovations. Although the framework presented here treats outside options as exogenously given, the lack of outside options for scientific innovations might be tied to structural and economic conditions related to the nature of the technology and the state of the market. For example, δ and n may be endogenous. While in this paper I do not empirically address these aspects, future versions of this manuscript will provide a discussion to that regard and some potential explanations.

the startup is defined as $X_t = X_i + X_s$.[17] Intuitively, X_t provides a snapshot of the beliefs of the risk-adjusted value the startup's assets will generate when deployed under an incumbent's. Finally, the share of value captured by the startup is then $\lambda_s = \frac{X_s}{X_t}$, which characterizes how economic gains are split between the startup (its founders and investors) and the acquirer's shareholders.

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 X_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. 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 or only indirectly related to the transaction's intrinsic value.

To isolate the acquisition-related signal from the surrounding 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 Kogan et al. (2017),[18] in the process of filtering out the signal from the noise, 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 or if the acquisition involves 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).

The methodology is summarized as follows, with more details provided in the Appendix. I treat the observed abnormal return r_i as a noisy signal of the unobserved expected return from the acquisition, x_i. Specifically, I assume x_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 X_i may be negative, the total value creation, X_t is non-negative. In other words, the transaction cannot destroy value in aggregate.

The logic follows standard Bayesian updating: given a prior distribution for x_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 $E[x_i \mid r_i]$, 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:

[17] Note that X_i represents the net incremental value to the incumbent's shareholders and that the stock market reaction is effectively subtracting the acquisition price X_s.

[18] 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.

$$E[x_i \mid r_i] = \delta_i r_i + \sqrt{\delta_i}\sigma_{\varepsilon,it} \frac{\phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,it}}\right)}{1 - \Phi\left(\frac{k_i - \delta_i r_i}{\sqrt{\delta_i}\sigma_{\varepsilon,it}}\right)}, \tag{4}$$

where $\phi(R_{it})$ and $\Phi(R_{it})$ denote the probability density and cumulative distribution functions of the standard normal distribution, respectively.[19] This expression can be interpreted as follows. First, δ_i captures the relative importance of the incumbents' private value variance in the total variance of the valuation process (the signal-to-noise ratio). A higher δ_i indicates that private-value information, rather than noise, is driving most of the variability in x_i. Next, R_i is the standardized threshold above which x_i is truncated, reflecting the fact that we only observe x_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 x_i, given that only high realizations of x_i are observed.[20]

Finally, the expected value x_i is then used to compute the private value for the incumbent, X_i, the total value for the startup, $X_t = X_i + X_s$, and the corresponding share of value captured by the startup, λ_s.

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.[21] 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.[22] This restriction

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

[20] Recent research highlights that this methodology may introduce measurement error, particularly when the estimate is used as the dependent variable (Arora et al., 2024). To evaluate the potential impact of such measurement error, I conducted a series of simulation (details not reported here). The results of these simulations confirm that, as intuition would suggest, the presence of measurement error in my setting biases the estimates downward. Consequently, the reported coefficients should be viewed, in that regard, as conservative, potentially representing a lower bound on the true effect.

[21] 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. Furthermore, if science-based startups consistently relied on IPOs as a viable and successful exit route, the issues discussed in this paper would be significantly attenuated, if not nonexistent. 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.

[22] 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).

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.[23]

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.[24] 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 period.

This dataset would be sufficient for the empirical analysis if I were not refining the estimates to isolate the signal from the noise. 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.[25] 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.

[23] 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. Moreover, acquisitions by public firms often generate higher returns for startups, making them an attractive outcome for entrants. These firms also tend to have the strongest necessary capabilities to integrate and scale acquired technologies, reinforcing their role as key players in the acquisition market. If a selection must be made, focusing on public firms is justified, as they are where the most significant deals and innovation-driven acquisitions occur. However, the impact of public acquirers on value capture remains unclear. On one hand, their market power could enable them to exert more control and reduce the share of value captured by startups. On the other hand, they may be willing to pay a premium for high-potential acquisitions, increasing the captured value. The direction of this effect is uncertain and requires further discussion to fully understand the implications of selection bias and acquisition dynamics.

[24] 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.

[25] The linkage is provided via acquirer gvkey.

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 4,878 startups.[26]

Figure 2 presents the temporal evolution of the number of startup acquisitions in the final sample by industry.[27] These trends align with previous studies on startup activity across industries. 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.

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. To that end, I look at the days in which the trading volume is abnormal. Figure 3 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 respond to the announcement during three days, from $[t, t + 2]$. I will use this time window in my main results.[28]

[26] 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. 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. In any case, for missing data to affect the estimates, the patterns would need to differ between science and non-science startups, and there is no clear reason for this to be the case.

[27] 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.

[28] 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.

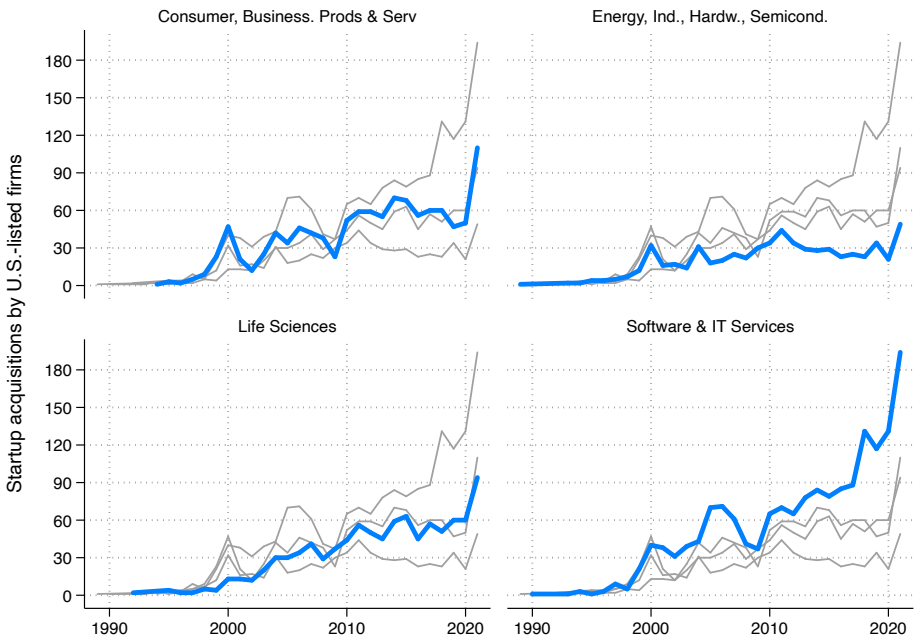


Figure 2: Number of startup acquisitions per industry and year (1990–2022). The data align with prior research on innovation activity, showing that startup acquisitions grow concentrated in life sciences (e.g., Biotechnology & Pharmaceuticals, Health Care Equipment & Services) and IT-related industries (e.g., Software & IT Services). In contrast, industries such as Energy, Industrials, Hardware, and Semiconductors display fewer acquisitions, with little growth after the 2000.

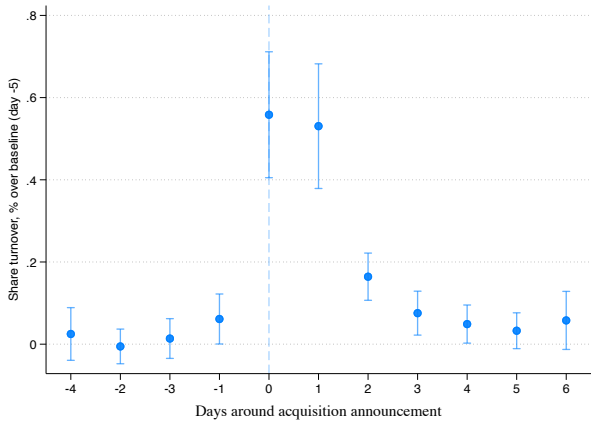
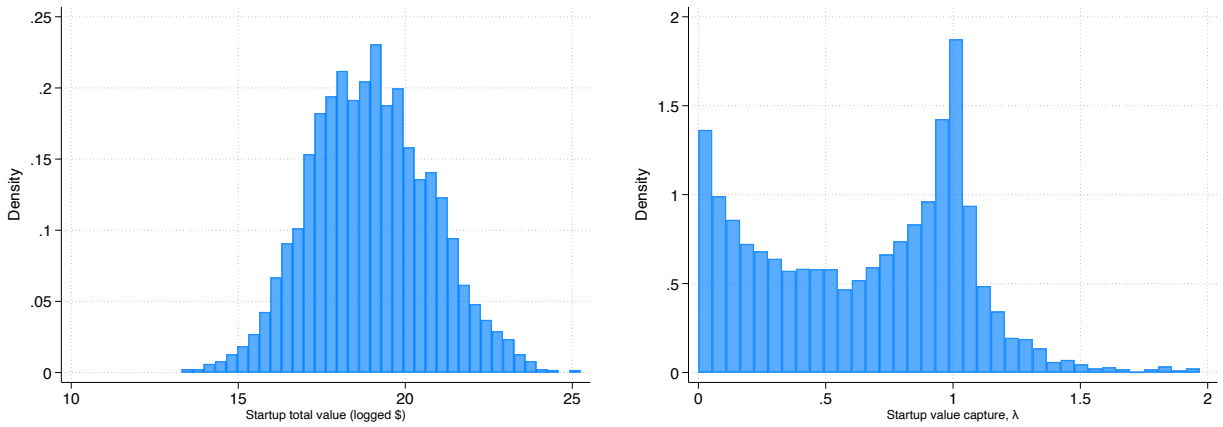


Figure 3: Share turnover around acquisition announcement days, in percentage points. Share turnover (s) is computed as the ratio of daily volume over shares outstanding. The Figure reports the coefficient estimates $b_l, l = [-4, 6]$ and 95% confidence intervals from the specification $s_{fd} = a_0 + \sum_l b_l I_{fd+l} + cZ_{fy} + \epsilon_{fd}$, 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). Next, using historical acquirer-level data, I filter out noise to isolate the expected return (x_i) attributable specifically to the acquisition. This involves estimating both firm-level and deal-level

parameters, with deal-level truncation defined as $k_i = -\frac{price_i}{market_cap_i}$.

Next, I accumulate the return over the three-day window ($x_{i,t=[0,2]}$) and then compute the private value $X_{i,t=[0,2]} = mktcap_{-1} \times x_{i,t=[0,2]}$, where $mktcap_{-1}$ is the acquirer's market cap at the closing of the previous day. Finally, I calculate the total value created (X_t) and the fraction captured by the startup, $\lambda_s = \frac{X_s}{X_t} = \frac{X_s}{X_s + X_i}$, where X_s is the price paid to the startup, including all forms of considerations, such as earn outs and stock options. The ratio λ_s indicates how much of the total surplus flows to the startup in the acquisition (and $1 - \lambda_s$ indicates how much is kept by the incumbent). Figure 4 presents the distribution of total value created and the fraction of value captured for the 5,847 startups in my final sample.



(a) (b)

Figure 4: Panel (a) displays the dollar estimates distribution for the *total value created*, X_s. Panel (b) displays the estimates for startup *value capture* (λ_s). Values over 1 indicate that the acquisition destroyed value for the incumbent ($x_i < 0$).

Figure 5 illustrates the correlation between the incumbent's private value X_i and the VC funds raised by the acquired startup. This strong correlation suggests that adjusted market beliefs, as captured by the measure, align closely with those of venture capitalists. While a similarly strong correlation is observed when considering total value X_t, the incumbent's private value X_i provides a more precise comparison. Unlike total value, which incorporates the price paid to the startup (and may indirectly reflect VC funding), the private value is influenced solely by market reactions.

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 (X_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.[29] The company had raised $61

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

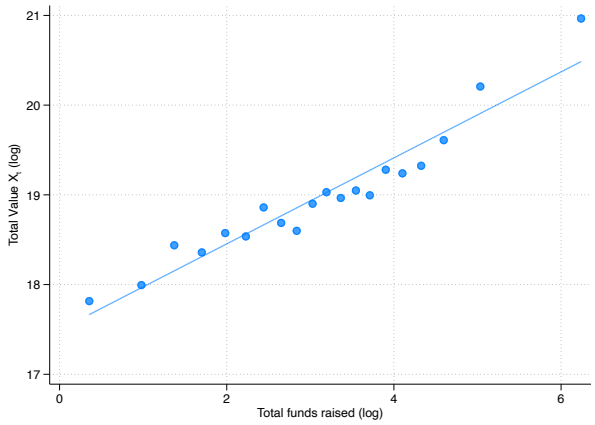


Figure 5: Cross-sectional relationship between the estimated total economic value ($X_t = X_i + X_s$) and the VC and PE funds raised before the transaction. Variables are log-transformed.

million in venture capital prior to acquisition. 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 (X_i). Thus, the total estimated economic value (X_t) of DeepMind at the time of acquisition is given by $X_t = X_i + X_s = 1,590 + 650 = \$2,240$ million.

DeepMind's shareholders, including its founders and investors, captured 29.0% of the total value, while Google retained the remaining 71.0%. This level of value capture places the startup in the bottom 20th percentile.[30] This division of rents suggests that the science-based startup, despite possessing a highly innovative, breakthrough 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. For example, a technology for which the risk of market adoption is high will see an attenuated stock market reaction, where the expected future cash flows are discounted at a higher rate.

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, needless to say, does not develop 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 73.31% 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 buyers further strengthened its outside options in negotiations. This variation in value capture aligns with the theoretical framework presented earlier, where the startup's outside options and the structure of the acquisition market shape the share of the rents captured.

[30] These value capture computations are gross and do not account for the investment required to develop the technology before acquisition. If investment costs are considered, the results adjust to $X_{t_net} = 1,590 + 650 - 61 = 2,179$ million, and the net capture by DeepMind's shareholders is given by $X_s - R = 650 - 61 = 589$, implying a capture share of $\frac{589}{2179} = 26.62\%$). The Appendix provides results using net rather than gross values.

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 involve breakthroughs that 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 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 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. Several methods exist.[31] I use 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 being able to predict, for example, the progression of scientific innovations towards commercialization in one Technology Transfer Office (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.[32] This textual data provides a comprehensive view of the startup's business and innovation activities and is processed by the LLM for classification.[33]

[31] 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 5,341 startups, 646 are categorized as Biotechnology firms, a sector with a high propensity for patenting. However, only 417 of these firms (64.6%) are matched to patents, leaving 229 firms (35.4%) without any patent records. A closer examination of these 229 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.

[32] See Appendix for details on this data collection process.

[33] 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

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 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.[34] 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 3.20% 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 18.8% of the sample. The Appendix provides robustness analyses using alternative classification thresholds and model specifications that take into account the model uncertainty. Furthermore, to validate the LLM-based classification, I compare it with patent-to-paper citation data for startups. The results, detailed in the Appendix, show a strong correlation between LLM-based scores and patent-to-paper citation metrics for startups with observable patents, supporting the validity of the LLM-based classification.[35] Table 2 provides selected examples with the purpose to illustrate the

additional information, rather than an inherent difference in their reliance on science.

[34]The Appendix provides details on the classification task, including the prompt used and the configuration in which the LLM is set. Additionally, following the methodology of Carlson and Burbano (2024), I assess the sensitivity of the main results of the paper to variations in prompt design, which may lead to differences in the classification task described in this Section.

[35]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.

classification exercises.

Table 2: 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.
Semiconductors	Cisco Systems	Luxtera	Developer of fabless silicon photonics semiconductors designed to offer low-cost, high-volume optics solutions with high performance, reach, and reliability.

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.
Industrials, Manufacturing, Materials	Shopify	6 River Systems	Developer of a mobile robot and software designed to redefine fulfillment automation for e-commerce and retail operations. The company's turnkey fulfillment software leverages system-directed workflow to make new associate onboarding instantaneous and helps to realize pick rates at half the cost of warehouse investments of the past.
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 Assessing Outside Options

Measuring outside options is challenging because a startup's full set of alternatives at the time of transfer is not directly observable. These options, which determine a firm's reservation value in bargaining situations, include 1) the number of potential acquirers in the market and 2) the viability of independently scaling the commercialization of its technology. Stronger outside options improve a firm's negotiation position, influencing both the likelihood and the terms of acquisition deals. However, the literature provides few established methods to measure a company's outside options at the time of acquisition, particularly in a way that is consistent across industries.

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 text-based similarity measures derived from product 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 instance, a startup developing advanced battery technologies for grid storage may attract interest from multiple energy incumbents aiming to strengthen their storage offerings. In biotechnology, a startup focused on novel drug delivery mechanisms could be relevant to several pharmaceutical firms seeking to expand their R&D pipelines. Similarly, in software, a startup specializing in enterprise resource planning for human resources may be of interest to incumbents providing complementary business applications.[36]

Crucially, Hoberg and Phillips (2010, 2016) develop and extensively validate a text-based measure of product-market similarity—known as the Text-Based Network Industry Classifications (TNIC)—, which has been widely employed in prior research to quantify firm-level competition, market concentration, industry relatedness, and innovation spillovers. Their approach constructs a continuous measure of similarity based on cosine scores derived from the textual analysis of 10-K product descriptions, embedding firms in a high-dimensional space where proximity reflects overlap in product offerings. Furthermore, the measure is computed annually, providing a time-varying indicator of product-market competition. This dynamic structure is particularly well-suited to the present analysis. I use a more recent version developed by the authors—the Embeddings-Based TNIC Industry Classifications (ETNIC)—released in 2024, which also extends coverage to the period 1989 to 2023.

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). 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.[37]

[36] Conversely, recent research has shown that 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.

[37] The data matching is straightforward. I use the gvkey identifier available in both the CRSP and HP datasets. The acquisition year is used for matching, ensuring that the pool of the acquiror's competitors reflects the market conditions

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 buyers as the size of the acquirer's competitive set in the acquisiton year as follows:

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

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 HP; and the sum is taken over all U.S. publicly listed firms $j \neq i$ in year t.

Figure 6 presents four illustrative examples of the product-market competition measure. Each panel displays the network of all US-listed firms based on product-market rivalry in 2016, constructed using the text-based similarity scores derived from 10-K filings by Hoberg and Phillips (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—identified using the HP measure and the specified similarity threshold of 0.2—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. Based on the HP measure and the 0.2 similarity threshold, 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 potential 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.

at the time of the acquisition.

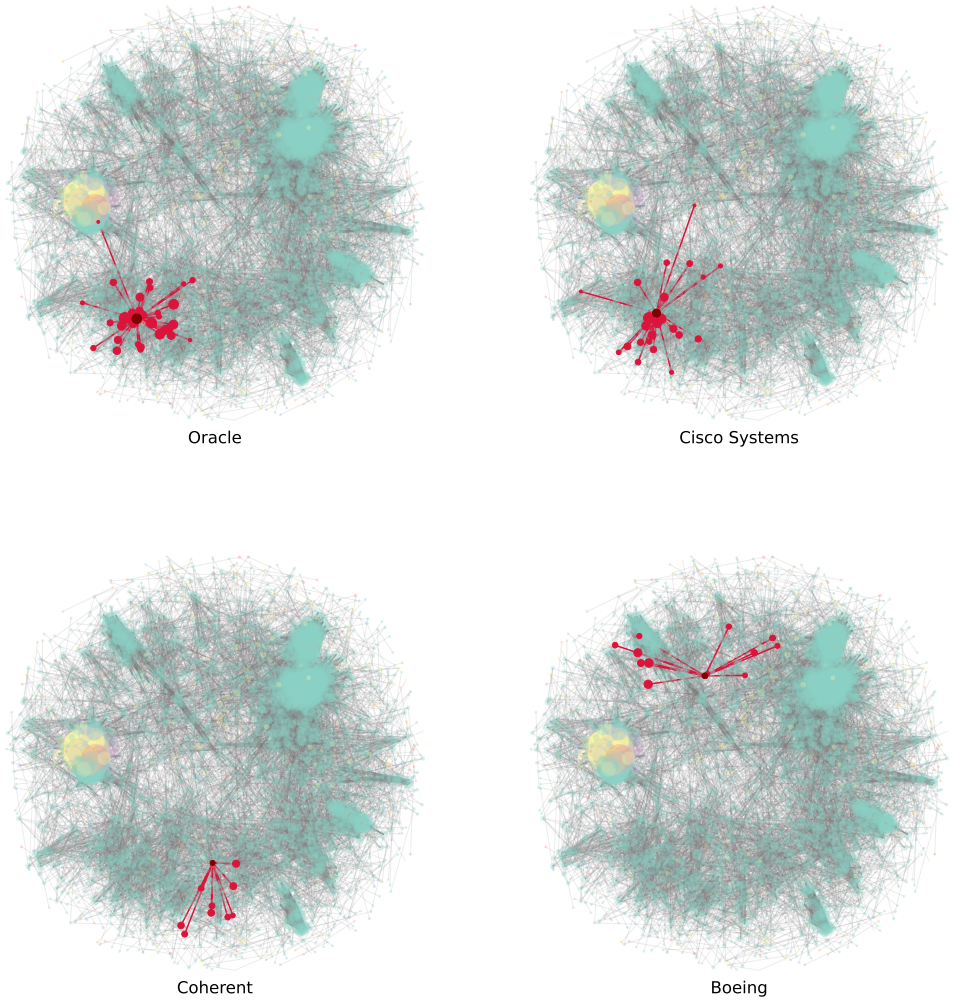


Figure 6: Product market networks for Oracle, Cisco, Coherent, and Boeing in 2016, constructed using text-based measures of product market similarity derived from 10-K filings by Hoberg and Phillips (2016). Each node represents a publicly traded U.S. firm, and edges indicate the degree of textual overlap 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 3 presents the summary statistics for the main variables of interest and Table 4 provides the distribution of science startups by industry.[38]

[38] 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;

Table 3: Summary statistics for main variables of interest.

	Mean	SD	Min	Max
Acquisition price, X_s	482	2,363	0.603	80,269
Expected market-adjusted return, x_i	0.010	0.031	-0.111	1.068
Incumbent private value, X_i (M$)	378	1,468	-6,880	36,177
Total value created, X_t (M$)	859.902	2,810	0.603	78,732
Startup value capture, λ	0.658	0.432	0.000	5.573
Total VC funds raised (M$)	78	536	0.005	29,505
Incumbent market cap (M$)	40,042	129,102	2.339	2,698,909
Science startup (1-5 scale)	1.74	1.26	1	5
Potential acquirers (HP)	9.50	19.26	0	425

Table 4: Distribution of science startups by technology.

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	18.8

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. 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. 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 (with the majority renewable energy startups), 24.2%; 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 novel scientific advances; or develop breakthroughs like next-generation photovoltaic materials or innovative energy storage systems.

Most interesting, Figure 7 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

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.

variation are excluded (software and consumer, businesses products and services). First, the value capture for science-based startups tends to be lower across all industries. For instance, 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 (0.56). 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).

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 can 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).[39] Overall, these statistics suggest that commercializing scientific innovations is associated with lower value capture. The next sections examine these dynamics in greater detail, as well as the extent to which outside options drive these results.

4 Empirical analysis

4.1 Science Startups: Low Capture, High Value

First, I analyze the extent to which science based startups face value capture issues. I employ OLS regressions with fixed effects, using the sample of 4,855 startups acquired by U.S. publicly listed companies. The dependent variable is the value captured by a startup (λ_s), i.e., the share of the rents the startup appropriates out of the total economic suprlus generated by the M&A transaction. The analysis is conducted at the startup level. The general specification is as follows:

$$\lambda_i = \beta_0 + \beta_1 sci_i + \beta \sum \mathbf{X}_i + \theta_i + \xi i + \epsilon_i, \tag{5}$$

where $sci \in [1, 5]$ is a discrete variable capturing the extent to which startup i commercializes scientific innovations. $\mathbf{X}$ is a vector of control variables that may influence value capture, θ represents grouped industry-year fixed effects to account for dynamics across industries and years, while ξ represents subregion fixed effects at the startup level, capturing regulatory and market differences, among other factors, based on the startup's location.

Results are reported in Table 5. Columns 1 to 3 analyze the entire sample of startups, while Columns 4 to 6 focus on a subset consisting only of startups that develop technologies, regardless of whether their innovations are scientific. The concern addressed here is that the results could be influenced by differences between technology-developing startups and those that do not develop tech-

[39] 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.

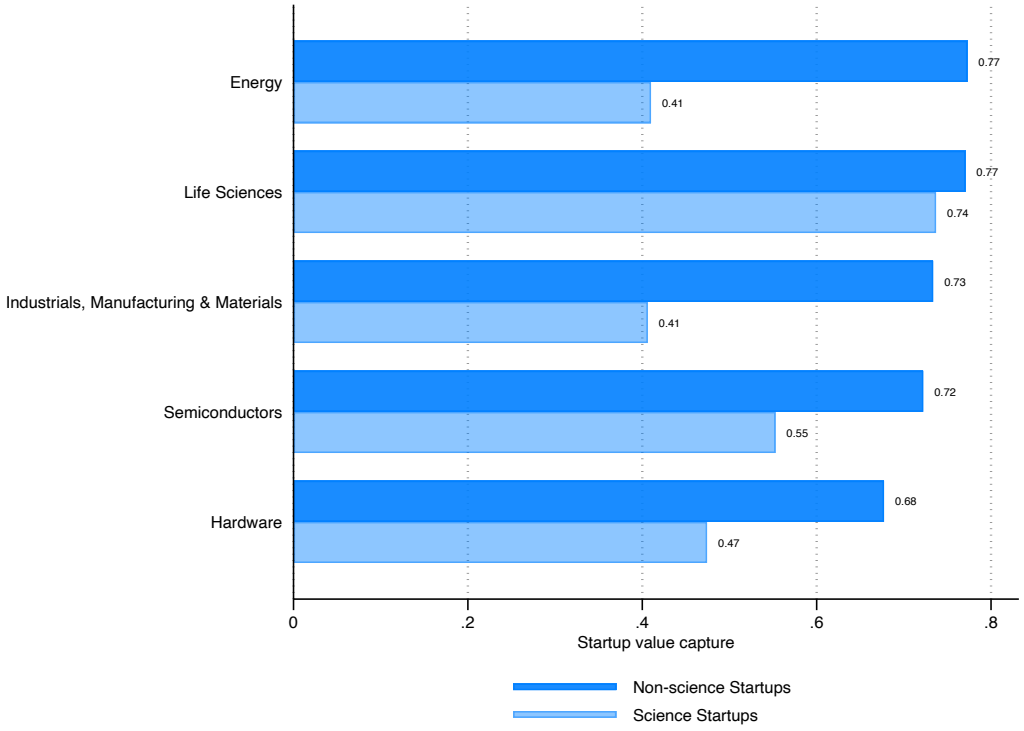


Figure 7: Average value capture by industry. The light blue bars represent startups that commercialize technologies based on novel scientific advances, while the dark bars represent non-science-based startups. The Industrials, Manufacturing, and Materials category includes Capital Goods such as Aerospace & 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

nologies at all but are still included in the sample, such as firms providing consulting services or selling simple consumer goods like food and apparel. This separation allows to examine not only the full sample of startups, but also the differences between those engaged in technological innovation and those with a technological component whose innovations are not necessarily scientific (refer to Table 2 for examples of startups commercializing scientific technologies *vs.* startups commercializing non-scientific technologies). Results indicate that startups commercializing scientific innovations face significantly greater challenges in value capture. Among technological startups, where the effects are more conservative and coefficients are slightly attenuated, Column 4 shows that startups commercializing scientific innovations experience an 11.3 percentage-point reduction in value capture, from 67.3 percentage points to 56—a 20.2% decrease. In monetary terms, if the average science-based transaction ($929 million) had captured value at the same rate as non-science startups, returns would have increased by $187.5 million.

Columns 4 and 5 refine the analysis by adding two key controls at both the startup and acquirer levels. 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 corre-

Table 5: OLS Models Regressing Value Capture on Science. This table presents OLS estimates examining how a startup's science-based classification relates to its value capture in acquisition. Columns (1)–(3) use the full sample of startups, while columns (4)–(6) focus on technological startups only. The regressions control for external financing (VC, PE), acquirer market capitalization, and include industry × year as well as subregion fixed effects. The value capture penalty is present in all specifications.

	All Startups			Technological Startups		
	Value Capture			Value Capture		
	(1)	(2)	(3)	(4)	(5)	(6)
Science	-0.117***	-0.121***	-0.098***	-0.113***	-0.117***	-0.090***
	(0.010)	(0.013)	(0.013)	(0.013)	(0.016)	(0.016)
Investment (VC, PE)		0.066***	0.100***		0.065***	0.102***
		(0.001)	(0.002)		(0.002)	(0.002)
Acquirer Market Cap			-0.092***			-0.093***
			(0.003)			(0.003)
Constant	0.679***	0.476***	2.417***	0.673***	0.473***	2.428***
	(0.001)	(0.003)	(0.075)	(0.001)	(0.006)	(0.075)
Industry × Year FE	Y	Y	Y	Y	Y	Y
Subregion FE	Y	Y	Y	Y	Y	Y
N	4,854	4,854	4,854	4,529	4,529	4,529
R^2	0.081	0.123	0.330	0.082	0.121	0.331

Standard errors clustered at the Industry-Year and Subregion level

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

lational 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.[40]

Second, I control for the acquirer's market capitalization, which serves as a reasonable proxy for firm size. The results indicate that larger acquirers appropriate a greater share of the transaction rents, thereby leaving less value for the startup. This outcome may be driven by the stronger complementarities inherent in larger firms—for example, their more developed distribution networks and manufacturing capabilities. In effect, larger firms benefit from economies of scale and have often made prior investments that enhance their ability to extract higher returns from these technologies. Consequently, while the reduced share of value for startups might seem unfavorable, it is consistent with the notion that larger acquirers are better positioned to realize a larger proportion of the overall value created by the transaction. Conversely, it could also be driven by a greater bargaining power that large incumbents may exert during negotiations. Furthermore, methodologically, controlling for the acquirer's market capitalization is also relevant, as the value capture measure is based on the market capitalization of the acquirer, which may be mechanically correlated with firm size. Higher market

[40] I examine the relationship between investment and value capture in greater detail in Section 5. 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 nearly 50% higher. This suggests that the returns to investment in terms of capture are more efficient in non-science firms. As discussed, potentially due to lower commercialization barriers, faster scaling opportunities, or more direct market applications. In contrast, science-based startups may face structural challenges that limit their ability to translate investment into proportional gains in value capture.

value can result in greater absolute private returns (X_i), thereby reducing the startup's relative share of the overall value.

Most important, when controlling for both total VC investment and market capitalization, and with the fixed effects at the industry-year and subregion level, the negative coefficient on science based startup remains quite stable and highly significant. This suggests that the value capture challenges faced by science based startups in the sample are not mostly driven by the amount of investment they receive or by the size of the acquirer, a finding consistent with the idea that these challenges lie elsewhere.

Next, I turn to Table 6, where the dependent variable is the logged value created. 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, which focuses on technological startups, shows that startups commercializing scientific innovations generate significantly more value than their counterparts—a 27.3% increase. Given an average value created of $1.27 billion for science-based startups, this corresponds to roughly $230 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 economic value than their non-scientific counterparts. The influence of acquirer size further suggests that larger firms may shape the total value created, potentially through stronger complementarities or bargaining power. Nevertheless, the primary conclusion—that science-based startups produce higher overall value—remains robust across these specifications. Consistent with the theory, this helps clarify why investment and entry into science based startups might still be rational: if higher total value creation compensates for lower individual capture, then overall returns can still be attractive enough to incentivize entry.

This conclusion, however, does not prevent from a market failure. These models reflect only science based startups that ultimately succeeded in attracting acquirers, leaving open the question of what happens to those that never reach this stage or those that fail to attract sufficient resources. If there are many science based startups that fail to raise enough funding or exit successfully, then underinvestment may remain a concern. The observed equilibrium—where startups accept a lower share but benefit from a larger overall "pie"—could mask a broader pool of unobserved ventures that never make it to acquisition.

4.1.1 Industry Heterogeneity

The results reported so far mask important cross-industry heterogeneity. Some sectors, 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 21.4% science and Industrials, Manufacturing & Materials with 30.1% science), and still others, like Consumer and Business Products & Services as

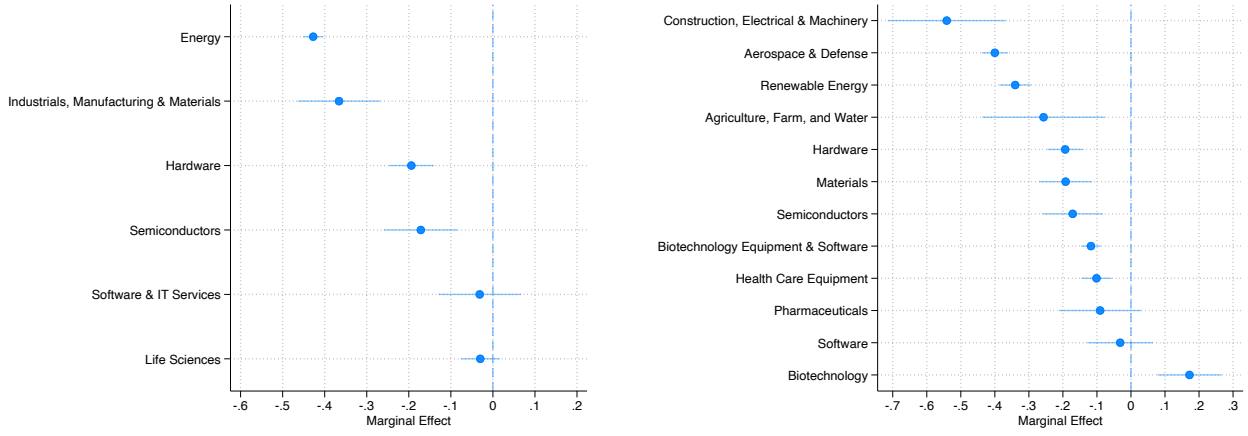
Table 6: OLS Models Regressing Value Creation on Science. This table presents OLS estimates examining how a startup's science-based classification relates to its value creation. Columns (1)–(3) use the full sample of startups, while columns (4)–(6) focus on technological startups only. The regressions control for external financing (VC, PE), acquirer market capitalization, and include industry × year as well as subregion fixed effects. Science startups tend to create more value than their counterparts.

	All Startups			Technological Startups		
	Value Creation			Value Creation		
	(1)	(2)	(3)	(4)	(5)	(6)
Science	0.146**	0.123**	-0.027	0.242***	0.208***	0.032
	(0.055)	(0.048)	(0.021)	(0.053)	(0.048)	(0.019)
Investment (VC, PE)		0.505***	0.277***		0.516***	0.276***
		(0.006)	(0.005)		(0.009)	(0.004)
Acquirer Market Cap			0.617***			0.618***
			(0.004)			(0.005)
Constant	18.943***	17.392***	4.439***	18.927***	17.344***	4.403***
	(0.008)	(0.011)	(0.089)	(0.008)	(0.016)	(0.098)
Industry × Year FE	Y	Y	Y	Y	Y	Y
Subregion FE	Y	Y	Y	Y	Y	Y
N	4,854	4,854	4,854	4,529	4,529	4,529
R^2	0.110	0.253	0.780	0.118	0.261	0.789

Standard errors clustered at the Industry-Year and Subregion level

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

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 5) to include industry indicators and interact them with the science indicator. Figure 8, Panel A, plots the marginal effects at the industry level and Panel B at the subindustry level.[41]



(a) Marginal effect by industry.

(b) Marginal effect by subindustry.

Figure 8: Comparison of marginal effects of science on value capture by industry and subindustry.

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

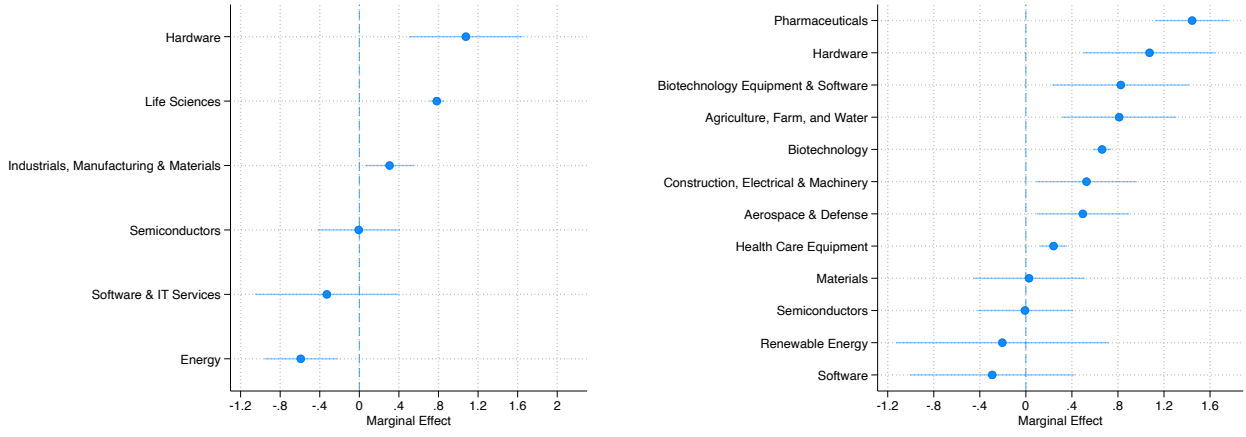
[41] The Appendix provides the results of the main specification.

the largest penalty, with a 40.0% difference in value capture for startups commercializing scientific innovations, followed by Industrials, Manufacturing & Materials (37.2%), Hardware (17.6%), and Semiconductors (16.6%), 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; Renewable Energy; and Materials all facing penalties ranging from 20% to 40% in value capture.

Interestingly, the Biotechnology and Pharmaceuticals sectors exhibit a reversed pattern. In these industries, the estimated penalties are either negligible or negative, 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.

These results can be attributed to the unique characteristics of the Biotech industry, where 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—such as drugs or medical devices—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.

Figure 9 reports the marginal effects on value creation. Science-based startups generate significantly more total value in four industries—Biotechnology & Pharmaceuticals, Hardware, Health Care Equipment & Services, and Industrials, Manufacturing & Materials. In each of these sectors, the coefficient on science-based ventures is positive and statistically significant, indicating that scientific startups create more value than their non-science counterparts. By contrast, science-based startups in Energy and Semiconductors exhibit insignificant coefficients.



(a) Marginal effect by industry.

(b) Marginal effect by subindustry.

Figure 9: Comparison of marginal effects of science on value creation by industry and subindustry.

4.2 Potential buyers, outside options, and value capture

In this section, I analyze the extent to which value capture is influenced by the number of potential buyers and how this effect varies between science and non-science startups. As introduced previously, I use as a measure of potential acquirers the number of competitors that a startup's acquirer has. Table 7 presents the results.

Table 7: OLS Models Regressing Value Capture on Science and Potential Acquirers. The table presents results from OLS regressions examining how science-based startups and the number of potential acquirers relate to value capture. Column (1) uses the (logged) number of potential acquirers as the dependent variable, while Columns (2) through (4) use value capture. All specifications include Industry × Year and Subregion fixed effects.

	Potential acquirers	Value Capture			
	(1)	(2)	(3)	(4)	(5)
Science	-0.862***	-0.133***		-0.132***	-0.166***
	(0.103)	(0.014)		(0.014)	(0.013)
Potential acquirers			0.001***	0.001***	0.001***
			(0.000)	(0.000)	(0.000)
Science × Potential acquirers					0.004***
					(0.000)
Constant	9.703***	0.670***	0.636***	0.662***	0.678***
	(0.006)	(0.002)	(0.003)	(0.002)	(0.000)
Industry × Year FE	Y	Y	Y	Y	Y
Subregion FE	Y	Y	Y	Y	Y
N	3,908	3,908	3,908	3,908	3,908
R^2	0.478	0.158	0.154	0.159	0.163

Standard errors clustered at the Industry-Year and Subregion level

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

In Column (1), the dependent variable is the number of potential acquirers. The coefficient on the Science indicator is negative and statistically significant, indicating that science-based startups are acquired by incumbents with fewer competitors and, in turn, implying that these startups have fewer potential acquirers relative to non-science startups. On average, science startups have 9.11% fewer buyers, with the finding supporting the idea that science startups often commercialize highly specialized or nascent technologies and thus face a more limited set of buyers 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 is negative and highly significant, aligned in magnitude with the results reported in previous sections. The coefficient on the number of potential acquirers is positive and significantly associated with value capture, indicating that broader competition among acquirers enhances the startup's bargaining power. An increase in the number of potential acquirers from the mean (9.5) to one standard deviation above the mean (28.76) is associated with a 6.0% increase in value capture, regardless of whether the startup is science-based or not (Column (3)). However, this coefficient masks also 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 4.18% increase for

non-science startups.

Figure 10 illustrates the relationship between the number of potential acquirers and value capture for science and non-science startups, based on Column (5)'s specification. The figure shows that value capture increases with the number of potential acquirers for both groups, but the slope is significantly steeper for science-based startups (more than 4x). At low levels of potential acquirers, science startups capture substantially less value than their non-science counterparts. However, as the number of potential acquirers increases, this gap narrows and eventually closes—illustrating that denser acquirer markets substantially mitigate the value capture disadvantage faced by science-based startups.

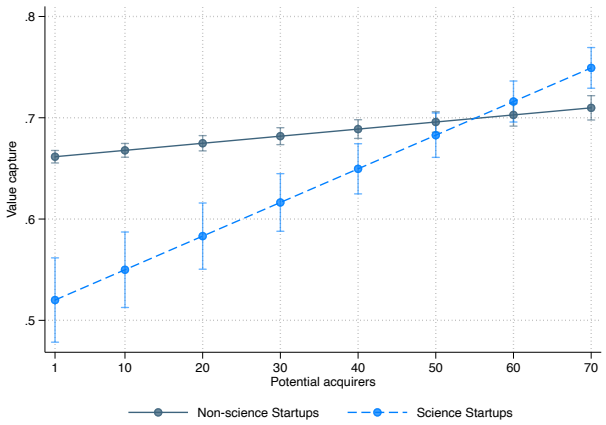


Figure 10: Predicted value capture as a function of potential acquirers.

These findings align with the theoretical framework. First, across all levels of potential buyers, science-based startups systematically capture less value than their non-science counterparts. Second, they are more sensitive to variations in the number of buyers. One explanation is that non-science ventures often have credible alternative pathways for scaling commercialization independently, enabling them to rely less on external acquirers. This stronger outside option compels buyers to offer a greater share of the rents even when the acquirer pool is small, resulting in only a modest slope as the number of buyers increases. By contrast, science-based ventures, lacking this fallback, depend more heavily on downstream incumbents and therefore experience larger gains from any expansion in the buyer pool.

5 Conclusion

Scientific innovations are essential for addressing some of the most pressing challenges, for example, in sectors like materials, energy, agriculture, and the environment. However, bringing these innovations to market is uniquely challenging, especially for startups that often lead the way in translating scientific breakthroughs into practical applications. While much of the focus has been on the problems of value creation—such as financing constraints and technological risk—this paper highlights that an equally significant challenge lies in value capture.

Specifically, I pose that when commercialization hinges on large-scale manufacturing, distribution, or other downstream capabilities held primarily by a few incumbent firms, startups founded on novel

scientific research often face weak bargaining power. Empirically, I focus on testing the extent to which value capture is a problem and its potential mechanism: a lack of outside options. To do so, I develop a simple methodology to measure both total value creation and value capture, based on an event study to stock market reactions upon acquisition announcement that, in turn, adapts a parametric approach developed by Kogan et al. (2017) to isolate the acquisition signal from market noise. Likewise, using textual information on startups and a fine-tuned Large Language Model, I assess a startup's outside options. The findings reveal that science-based startups generate significantly more total value than their counterparts, underscoring their economic importance. However, these consistently capture a smaller share of that total value, driven by a lack of outside options.

Furthermore, I find significant heterogeneity across industries, with value capture problems being more pronounced in sectors where the composition of science-based startups is mixed. These findings align with the areas where concerns about insufficient innovation activity are most pressing: advanced manufacturing, climate, agriculture, energy technologies, computer hardware, materials, and defense.

It is worth noting that my sample is highly selective: these are startups that were acquired by publicly listed incumbents, and, naturally, I do observe innovation and entry into these sectors despite the potential hold-up problem. However, my results indicate that this entry occurs because the same startups that face value capture constraints also create significantly more value overall, effectively offsetting the negative impact of the hold-up and, thus, generating sufficient returns. Nonetheless, speculative, but consistent with classical hold-up theory, I propose that the value capture dynamics not only constrain startups' ability to retain value during acquisitions but may also suppress upstream incentives to enter these sectors. Consequently, this dynamic could leave potentially valuable and socially desirable innovations underfunded, limiting their broader impact.

The results of the paper have implications for both management and policy. On the management side, these startups should aim to develop strategies that increase their bargaining power, such as diversifying their potential acquirers by cultivating relationships with multiple incumbents or alternative buyers, investing in complementary capabilities that reduce reliance on any single partner, and exploring alternative commercialization pathways like partnerships, licensing, or public-private collaborations. On the policy side, the results emphasize that solutions should extend beyond increasing R&D funding, reducing technological risks, or subsidizing demand. Addressing the hold-up problem requires policy interventions that foster more competitive acquisition markets, expand commercialization pathways, and strengthen startups' outside options so that they are not locked into partnerships with a single or a limited set of incumbents.

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.

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

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. (2024). Revisiting the private value of scientific inventions. Technical report, National Bureau of Economic Research.

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

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

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.

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.

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)*.

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.

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.

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.

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*.

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.

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.

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.

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.

IEA (2020). Clean energy innovation. Technical report, International Energy Agency.

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

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.

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.

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

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.

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. Technical report, National Bureau of Economic Research.

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.

Moss, E., Nunn, R., and Shambaugh, J. (2020). The slowdown in productivity growth and policies that can restore it. *The Hamilton Project, Brookings Institution, Washington, DC*.

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.

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.

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

Schmookler, J. (1966). Invention and economic growth.

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.

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

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

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.

Veugelers, R. (2024). Powering the clean energy innovation system. Technical report, Peterson Institute for International Economics.

Appendix

A Correlation

Table A.1: Pairwise correlations between main variables of interest

Variables	Acquisition price, X_s	Expected return, x_i	Incumbent value, X_i	Total value created, X_t	Startup capture, λ	Total VC funds raised	Incumbent market cap	Science startup	Potential acquirers
Acquisition price, X_s	1.000								
Expected return, x_i	-0.063	1.000							
Incumbent private val., X_i	0.014	0.069	1.000						
Total value created, X_t	0.866	-0.019	0.513	1.000					
Startup value capture, λ	0.110	-0.389	-0.311	-0.061	1.000				
Total VC funds raised	0.367	-0.039	0.003	0.316	0.056	1.000			
Incumbent market 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

B Modified KPSS Estimation Methodology with Truncation

In the original Kogan et al. (2017) methodology (KPSS), the unobserved private value x_i is assumed to follow a normal distribution truncated at zero ($x_i > 0$). I modify this assumption to allow for a variable truncation point k_i, where $k_i < 0$ and varies across observations. Most of the derivation is the same as that of KPSS. Specifically, I assume:

$$x_i \sim N(0, \sigma^2_{x,ft}) \quad \text{with} \quad -\infty < k_i < x_i$$

The objective is to derive the expected value $E[x_i \mid r_i]$ under this new truncation point, where $r_i = x_i + \varepsilon_i$

Modified Assumptions

1. Distribution of Incumbent Private Values:

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

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:

$$x_i \perp \varepsilon_i$$

Objective

I want to calculate:

$$E[x_i \mid r_i]$$

where $r_i = x_i + \varepsilon_i$.

Derivation

1. Joint Density Function

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

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

Given:

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

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 $\varepsilon_i = r_i - x_i$, we have:

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

2. Simplifying the Exponent

Expand the exponent:

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

Combine terms:

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

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:

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

3. Completing the Square

Focus on the x_i terms:

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

Complete the square:

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

Thus the exponent can be written as:

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

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

4. Conditional Density Function

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

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

where

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

and

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

Here,

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

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

5. Conditional Expectation

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

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

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:

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

6. Final Formula

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

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

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 $x_i > k_i$.

This generalizes the original KPSS formula to allow for a variable and negative truncation point k_i.