UNIVERSITY OF TURKU
Turku School of Economics

# Does timing matter?

Empirical Study on the Effects of First Equity Funding Timing in Deep-Tech Startups (Europe and North America, 2010–2020)

Finance
Master's thesis


Author:
Henri Mäkivirta

Supervisor:
PhD Antti Miihkinen


2.6.2025
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

**Subject**: Finance

**Author**: Henri Mäkivirta

**Title**: Does timing matter? Empirical Study on the Effects of First Equity Funding Timing in Deep-Tech Startups (Europe and North America, 2010–2020)

**Supervisor**: PhD Antti Miihkinen

**Number of pages**: 52 pages + appendices 23 pages

**Date**: 2.6.2025


This study examines how well the timing between deep tech startup's launch and first round of equity funding can predict the startup's success outcome. The empirical study uses proprietary data of more than six thousand observations on European and U.S. deep tech startups founded between 2010 and 2020. The success in the study is defined by both binary and quantifiable outcomes. In binary outcomes, success is defined as an outcome where a startup has achieved a unicorn valuation (€1billion), been a subject to merger or acquisition of over €500M or raised funding over €100M in its lifetime. The valuation outcome uses the valuation of the startup's latest funding round.

The literary review of this thesis captures the theoretical foundations of venture capital. It covers how VC funding cycles work in general and how signaling theory plays a key role in each funding round. Deep tech startups' unique funding dynamics are compared to more traditional software startups. This study also investigates the role of non-dilutive funding, which is important in research heavy deep tech innovation.

The empirical study measures the impact of time between deep tech startup's launch and first round of equity funding by using that time as a predictor variable to the outcome. This study uses logistic regression, linear regression and multiple regression to estimate the impact of each variable for predicting the success outcome.

The results show that timing matter when predicting the binary outcome with logistic regression model. Logistic regression reveals that the probabilities for success are the highest when raising first round of funding three to eight months after the launch. When testing for a quantifiable outcome with linear regression and multiple regression, the timing effect disappears. Controlling other variables like patent, region and industry together with timing turn out to be more robust. The timing of the first equity funding round has a strong non-linear effect on whether a deep-tech startup achieves a successful outcome or not. However, the impact on valuation is mostly shaped by other factors than timing, such as how strong the startup's IP is, where it is located, and what industry it operates in.



**Key words**: deep tech, startups, funding timing, funding round, launch, venture capital, startup success, unicorn, valuation, patents, natural splines, logistic regression, linear regression, multiple regression

Pro gradu -tutkielma

**Oppiaine**: Rahoitus

**Tekijä**: Henri Mäkivirta

**Otsikko**: Onko ajoituksella merkitystä? Empiirinen tutkimus ajoituksen merkityksestä syväteknologia startupien oman pääoman ehtoisessa rahoituksessa (Eurooppa ja Yhdysvallat, 2010–2020)

**Ohjaaja**: PhD Antti Miihkinen

**Sivumäärä**: 52 sivua + liitteet

**Päivämäärä**: 2.6.2025

Tässä Pro Gradu -tutkielmassa tarkastellaan, kuinka hyvin syväteknologiayrityksen perustamisen ja ensimmäisen oman pääoman ehtoisen rahoituskierroksen välinen aika voi ennustaa startupin menestystä. Empiirisessä tutkimuksessa käytetään datana eurooppalaisia ja yhdysvaltalaisia syväteknologiayrityksiä, jotka on perustettu vuosina 2010–2020. Menestystä mitataan sekä binäärisellä että kvantitatiivisella mittarilla. Binäärisesti startup määritetään menestykekkääksi, mikäli startup on saavuttanut yksisarvisen valuaation (1 miljardi euroa), ollut yli 500 miljoonan euron fuusion tai yritysoston kohteena tai kerännyt yhteensä yli 100 miljoonan euron rahoituksen elinkaarensa aikana. Valuaation pohjana käytetään startupin viimeisimmän rahoituskierroksen arvostusta.

Kirjallisuuskatsaus kuvaa kasvuyritysten riskipääoman teoreettisia perusteita. Se käsittelee, miten riskipääomarahoitussyklit toimivat yleisesti ja miten signalointiteorialla on keskeinen rooli jokaisessa rahoituskierroksessa. Syväteknologiayritysten ainutlaatuista rahoitusdynamiikkaa verrataan perinteisempiin kasvuyrityksiin. Tutkimus tutkii myös apurahapohjaisen rahoituksen roolia, jolla on suuri merkitys syväteknologiainnovaatioiden kaupallistamisessa. Empiirinen tutkimus mittaa deep tech -startup-yrityksen perustamisen ja ensimmäisen rahoituskierroksen välisen ajan vaikutusta käyttämällä kyseistä aikaa ennustemuuttujana tulokselle. Tässä tutkimuksessa käytetään logistista regressiota ja lineaarista regressiota arvioidakseen kunkin muuttujan vaikutusta onnistumisen ennustamiseen.

Tulokset osoittavat, että ajoituksella on merkitystä ennustettaessa binääristä lopputulosta logistisella regressiomallilla. Logistinen regressio paljastaa, että onnistumisen todennäköisyydet ovat suurimmat, kun ensimmäinen rahoituskierros kerätään 3–8 kuukautta perustamisen jälkeen. Kun testataan kvantifioitavaa lopputulosta lineaarisella regressiolla ja moninkertaisella regressiolla, ajoituksen vaikutus häviää. Muiden muuttujien, kuten patentin, alueen ja toimialan, kontrollointi yhdessä ajoituksen kanssa osoittautuu selittävän menestystä paremmin. Ensimmäisen rahoituskierroksen ajoituksella on voimakas epälineaarinen vaikutus siihen, saavuttaako deep tech -startup onnistuneen lopputuloksen vai ei. Onnistumiseen vaikuttavat kuitenkin enimmäkseen muut tekijät kuin ajoitus, kuten startupin immateriaalioikeuksien vahvuus, sijainti ja toimiala.

**Avainsanat**: deep tech, syväteknologia, startup, rahoituskierros, riskirahoitus, kasvuyritys, yksisarvinen, valuaatio, patentti, logistinen regressio, lineaarinen regressio

# TABLE OF CONTENTS

## LIST OF FIGURES



## LIST OF TABLES

# 1 Introduction

### 1.1.1 Challenges in deep tech venture

Deep tech startups turn scientific and engineering breakthroughs into companies that build real products. The companies first start research in the fields like biotechnology, quantum computing, and advanced materials, and then turn this research into commercial products. Deep tech companies face unique challenges compared to regular tech companies. Deep tech innovation is more challenging due to long research and development (R&D) timelines, high capital needs, complex regulations, and financing risks. General software startups can quickly test and scale their products at low cost. Deep tech startups on the other hand can take years to turn scientific discoveries into products that can be sold on the market for customers. As Romme et al. (2023) note, the "valley of death" in deep-tech ventures is particularly deep due to the compounded scientific, technical, and institutional barriers. Many startups face a 5–10 years before generating any revenue.

For example, innovations like fusion energy need a lot of technical testing. They require special facilities and regulatory approvals. All of which delays commercialization (Nanda & Rhodes-Kropf, 2017). The long development process adds more uncertainty, as startups must overcome technical challenges while securing funding through several stages of development. Pinkwart et al. (2015) note that this gap, or “valley of death”, is one of the most common reasons for failure in high-tech ventures.

The recent report from Dealroom highlights the fine balance between high capital needs and innovation. The 2025 European Deep Tech Report shows that deep tech startups raise more money than regular tech firms at every stage. Almost all top rounds are raised by teams who are building deep tech companies. However, over 50% of the funding for growth-stage startups in Europe comes from non-European investors. This poses as a severe risk from geopolitical and market point of view if wind changes (Dealroom et al., 2025). This reliance on outside funding reveals gaps in Europe’s ability to support deep tech growth. There is hope for a change in Europe: Europe has strong research infrastructure, including six of the top 20 global universities and nine of the top 25 research institutes (Dealroom et al., 2025).

The mix of intellectual property (IP) risks and strategic choices makes the commercialization process even more difficult for deep tech startups (Gans and Stern, 2003). Startups often rely on patents to protect their ideas when working with established companies. Obtaining and defending patents takes a lot of resources and time. Challenges with patents puts pressure on the limited funds that startups have. In industries where IP protection is weaker, startups have to negotiate carefully with larger companies. For example, Robert Kearns couldn't protect his windshield wiper invention from Ford (Gans & Stern, 2003). Slow or weak IP enforcement can weaken a startup's position and make them vulnerable to copying. Revealing ideas too early without strong IP protection can threaten a startup's chances of commercial success.

High capital need is a key challenge of deep tech startups. Developing prototypes, running clinical trials, or scaling manufacturing is capital intensive and startups usually don't have revenue yet at that point. The Dealroom report shows that deep tech companies spend 20–40% of their capital on infrastructure like gigafactories or semiconductor plants. Heavy investments in infrastructure creates strong barriers to entry but also requires patience from investors (Dealroom et al., 2025). However, raising large amounts of money early on can conflict with maintaining flexibility. The ability to change direction or stop projects is usually tied to early results. Nanda and Rhodes-Kropf (2017) show that deep tech startups with high real option value have to face a key decision: they can either try secure as much capital as possible and so reduce financing risks or keep the ability to cancel failing projects. For example, a biotech startup might wait for preclinical trial results before raising a large funding round. This way it can avoid committing too much to unproven ideas. The reliance on investor groups makes the challenge even harder. These groups can offer valuable expertise. However, they may also create coordination problems, especially during economic downturns when follow-on funding is harder to secure.

Regulatory challenges is another key of difficulty for deep tech startups, especially in fields like biotechnology, medical devices, and clean energy. Colombo et al. (2016) show that strict product market regulations and compliance requirements hit high-tech ventures hardest. Strict regulations usually mean raise in both development costs and time-to-market. Startups in regulated industries will have to navigate and go through multiple approval and assessment stages. To go through the process often means that startups need to raise additional funds. The European Deep Tech Report show that these type of regulatory delays and high energy costs slow down innovation (Dealroom et al., 2025). The situation can be devastating if the demand is growing at the same time. Colombo et al. (2016) also note that patents alone aren't enough to overcome the negative impact of strict regulations on venture capital. Highly regulated environments make investors more cautious, and they may delay funding until startups meet key regulatory milestones.

Non-dilutive funding options are crucial for reducing risk in early-stage research and development (R&D). Grant programs help startups reach technical and regulatory goals without giving up too much equity. Grants also act as a signal of quality and traction, and therefore makes startups more credible to future investors (Colombo et al., 2016). However, grants alone are not enough for scaling and pushing startups to launch their new innovations to commercial production. The timing of equity funding rounds becomes key for a startup's survival. Startups that secure initial funding after a clear milestone lower investor risk and improve their chances of survival (Nanda & Rhodes-Kropf, 2017; Colombo et al., 2016). Hitting a milestone in the context of deep tech startups can mean proving technical feasibility or validating a prototype, getting a patent, or passing Phase I trials.

Funding timing is crucial for deep tech startups. When money is invested, along with achieving technical and legal goals and considering the economy, it impacts not just individual startups but also the success of breakthrough technologies that can benefit the society the society as a whole.

### 1.1.2 Funding timing

The timing of funding in deep tech startups is influenced by a mix of market cycles, venture capital (VC) behavior, and the specific challenges of high-risk innovation. According to Kaplan and Schoar (2005), private equity fundraising and performance follow market trends. The more money flowing in during market highs, the more funding for deep tech startups and vice versa. For deep tech startups that are facing long research and development (R&D) periods, complex regulations, and high capital needs; the potential downcycle increases financing risks. Kaplan and Schoar (2005) show that VC fund returns are similar to the S&P 500. However, the performance of top funds highlights the value of working with experienced investors. The top funds are often better at timing exits and managing market changes. Working with the top funds can help reduce risks for ventures that need ongoing capital to meet technical and regulatory goals (Gompers & Lerner, 2001).

The 2025 European Deep Tech Report echoes that investments in deep tech startups generally perform better than investments in more traditional startups. Despite the European deep tech doing well, Europe still struggles and lags U.S. when it comes to exits. Most mergers and acquisitions (M&A) are led by U.S. companies (Dealroom et al., 2025). The imbalance in exits highlights the need to time funding properly to match U.S.' ability to scale and exit. Darktrace and Exscientia are great examples of recent successful exits from Europe. But due to lack of late-stage capital available in Europe, they were bought by U.S. firms (Dealroom et al., 2025).

Kaplan and Schoar (2005) show that as a fund grows larger, the funds' performance tends to decrease. The bigger the fund, the lower the returns. The large funds can provide important initial capital, but it is crucial to work also with specialized investors. Syndicating and involving more than one investor to the cap table allows the team to keep control and avoid overvaluation. Lerner's (1994) findings match to Kaplan's and Schoar's: experienced VCs team up to verify deals. Teaming up for due diligence in early-stage funding is beneficial for everyone. Technical challenges usually require careful research and having more than one expert involved can speed up the process. During economic booms, too much capital can flow into the market. Startups have more options and valuations push higher and higher (Gompers & Lerner, 2000). As a result, deep tech startups may feel pressured to accept inflated terms. If the market later turns, these high valuations can make it harder to raise more funding. Down round is a negative signal for the markets.

The fact that new venture capital firms often enter the market during economic booms makes these challenges even worse. Kaplan and Schoar (2005) show that venture capital funds raised during market peaks often perform poorly. Many inexperienced investors enter the market during boom periods. Their lack of experience lowers overall returns. It is risky for deep tech startups to depend on a new inexperienced investor who has just entered the world of venture during the upmarket. Startups might need more funding when they reach key regulatory milestones. If the market shifts and the new inexperienced investors pulls back, the startup may not be able to raise the money it needs to continue. Grants can help in the short term (Colombo et al., 2016), but raising equity is still essential for growth.

The long-term performance of venture capital (VC) firms, shown by Kaplan and Schoar (2005), highlights the value of partnering with experienced VCs. Top venture capital firms are also better at timing IPOs and acquisitions (Gompers & Lerner, 2001). This skill is especially important for deep tech startups. It is unique to deep tech startups that they usually take longer time to reach an exit. Newer VC firms may feel pressure to deliver quick returns for Limited Partners (LPs) (Gompers, 1996).

Regulatory and intellectual property (IP) risks make timing the funding again more difficult. In industries like cleantech, weak protection of ideas forces startups into risky negotiations with larger, established companies (Gans & Stern, 2003). Weaker position leads deep tech startups to raise money in stages. Gompers (1995) points out that staged funding helps to keep entrepreneurs accountable while giving investors the ability to cancel projects that aren't working. The staging can come from either from the needs and preferences of the startup or the demand and preference of the investor. The same trade-off between flexibility and risk that was discussed earlier in chapter 1.1.1, is in play here as well (Nanda and Rhodes-Kropf, 2017).

The best time to raise funding is when the market conditions are good. It helps if the round can be raised after some clear milestone. It is beneficial to partner with strong, reliable funds or partners. These partners can support the startup through bad market conditions. Tools like syndication, non-dilutive funding options, and targeted regulations are crucial for thriving deep tech startup ecosystem.

## 1.2 Research Objectives and Questions

This study examines whether the timing of first equity funding round affect the deep tech outcomes. The empirical analysis uses unique proprietary dataset of 6,226 deep-tech startups from Europe and U.S. founded between 2010 and 2020. The study explores successful outcomes both binary (success/failure) point of view and from quantitative (startup's valuation) point of view.

The empirical analysis uses logistic regression, linear regression and multiple regression to answer the following questions:

1. How does the timing of the first funding round influence startup success?
2. What role do economic cycles play in that timing?
3. Does the funding timing affect company valuation after controlling other factors?

The sample data consist of different deep tech startups across all sectors like health tech, energy, robotics and semiconductors. The goal is to identify timing strategies across all deep tech sectors and add to the broader understanding of how VCs make investment decisions.

## 1.3 Thesis Structure

Section 2 reviews literature on VC timing, deep tech funding dynamics, and prior studies linking timing factors to startup outcomes. The literature review focuses only on the top journals. For the journals to be included in this thesis, they had to either score double digits in The Journal Impact Factor (JIF) or Q1 in Quartile score.

Section 3 reviews the data collection and variable editing, methodology, and econometric models to test hypotheses. The study uses proprietary data on deep tech startups founded between 2010-2020. The empirical study tests whether the timing of the first equity funding effects the success of the startup. The empirical study uses logistic regression, linear regression and multiple regression to test the hypothesis.

Section 4 reviews empirical findings, including descriptive statistics and hypothesis testing. Logistic regression shows significant nonlinear relationship between funding timing and success. Success peaks 3–8 months post-launch, confirming an optimal timing window. Linear regression suggests modest nonlinear timing influence on valuation. Multiple regression shows that timing is insignificant when all variables are considered.

Section 5 reviews theoretical and practical implication and concludes with limitations and future research directions. The results show that the timing of the first funding round has a strong non-linear effect on whether a deep-tech startup achieves a successful outcome. The timing's impact on valuation is not as clear and is mostly shaped by other factors like strong IP and geographic location.

# 2 Literature Review

## 2.1 Theoretical Foundations of Venture Capital Timing

### 2.1.1 VC funding cycles and market cycles

Black and Gilson (1998) argue that to have a thriving venture capital and startup ecosystem, there needs to be well oiled stock market that allows exits to happen through IPOs. This way the money circles back to the investors after the exit. The investors can reinvest it to a new cycle again. Funding timing in these cycles is important for startups and even more important for deep tech startups. Deep tech startups need a lot of capital. It takes a long time to spin out the research into a business. Gompers and Lerner (2001) explain that VC funding follows the ups and downs of the economy. When the market is strong, there's lots of money available and startup valuations go up. During downturns funding becomes harder to get. Navigating downturns is tough for deep tech companies. They need steady investment to hit key technical and legal goals (Gompers & Lerner, 2000; Kaplan & Schoar, 2005).

Hochberg et al. (2007) show VC firms performance is positively correlated with the quality of its networks. Well-connected VC firms tend to perform better. Hochberg et al. (2007) goes on to show that the portfolio companies in successful funds not only survive, but they reach successful exits faster. The best VCs have more access to more people, and they find more opportunities to invest. The most experienced VCs might have gone through the same process of suffering though regulation process. A startup can be halfway through a promising clinical trial, but if the funding dries out, the company will fail. Networks can become a lifeline during a downturn

Gompers and Lerner (2001) point out that syndication among VCs helps manage risk. Experienced investors can use their networks to share capital and pool expertise. Multiple experts help to reduce the information gaps and make due-diligence less difficult. Hochberg et al. (2007) expand on earlier research by showing that being well-connected helps VCs choose and support better-performing startups. Investors are more cautious during downturns. Well-connected VCs often act as signals of quality. Their involvement can encourage other investors to join. For example, a biotech startup is more likely to raise a follow-on round during a recession if it already has a well-connected lead VC. This is especially true when that VC has a strong network of others with industry expertise and a long-term investment outlook.

Well-connected VC firms can operate well in both up and down cycles. Kaplan and Schoar (2005) show that top-performing VC funds are able to be successful regardless of the market conditions. Hochberg et al. (2007) suggest this is partly due to their network advantages. Top-performing VCs use their syndication networks to time exits well. They can speed up IPOs and acquisitions when markets are strong and provide bridge funding when conditions are tougher. Bridge funding is like bridging the gap between two different funding rounds. It can be a leap of faith from the investor side, even more than an official funding round. Well-networked VCs can also help match funding to key milestones. Timing the funding to milestones allows teams to raise more funding on better terms.

However, network benefits are not equally available to all startups. Hochberg et al. (2007) show that the value of being well-connected depends on location and industry. In regions like Europe, where there is often a shortage of late-stage funding (Dealroom et al., 2025), deep-tech startups that rely on local networks may be more at risk during market downturns than U.S. startups that can tap into larger, more global investor groups. Non-dilutive funding, like government grants, can help reduce early-stage risk (Colombo et al., 2016), but becomes less effective in later stages when equity funding is still needed to scale. At that point, timing becomes critical: delaying institutional funding until key technical milestones are reached can reduce dilution, but it also risks missing the best market conditions for attracting investors.

Market cycles and investor networks shape funding outcomes. Gompers and Lerner (2001) show the risks that come from capital flowing in and out with the market, while Hochberg et al. (2007) show that strong investor networks help reduce that risk. Investor networks enable deal flow between investors to support startups during market swings. Having a well-connected investor can be a key safety net for deep tech startups. Well-connected investors can help them stay on track through regulatory hurdles and technical milestones. Connection between timing and investor networks shows why founders should look beyond just funding terms, and consider how well-connected their investors are, especially in sectors where long development cycles can be disrupted by sudden market shifts.

### 2.1.2 Signaling theory

Spence's (1973) signaling theory explains how people use signals to communicate in situations where one side knows more than the other. In markets where the quality of a product or service is hard to judge, credible signals help reduce uncertainty. Startups use signals to show that they are technically capable, well-managed, and have strong growth potential. Hallen and Eisenhardt (2012) expand on this idea by showing how startups can take strategic steps to make their signals stronger. Their concept of "proofpoints" serves as clear signs of a startup's readiness and potential. Proofpoints are externally validated milestones. Milestones can be for example working prototypes, regulatory approvals, or strategic partnerships. Signals need to be strong and hard-to-fake.

Hallen and Eisenhardt (2012) also show strategies that can improve the effectiveness of signals. One such strategy is "timing around proofpoints," which means raising funding soon after reaching a key milestone. The idea is to create urgency and makes the startup look stronger. It is similar to Spence's view that signals work best when they match how people process recent information. In deep tech, startups that wait to raise equity until after hitting important technical goals like passing Phase I clinical trials or completing a pilot plant, can lower investor doubts and raise funding more easily. Nanda and Rhodes-Kropf (2017) show similar results in their study of high-tech startups.

The study also identifies "casual dating," a pre-funding strategy where entrepreneurs build informal, non-transactional relationships with investors through repeated interactions focused on strategic advice. “Casual dating” helps building familiarity and trust, reducing the information asymmetry between founders and investors. Teams that lack established investor networks can try to level the playing field by showing credibility through relationship tactics like this. Colombo et al. (2016) tangents the notion of signals, noting that non-dilutive funding, such as grants, also serves as a signal, demonstrating technical merit to future equity investors.

Hallen and Eisenhardt's describe equifinality as a framework where startups can reach funding success in two ways. The first way is through strong existing relationships. The second way is through smart strategies. Already successful and validated startups often rely on past investor connections (social capital) to raise money faster. This is an example of Spence's theory about the power of clear, visible signals. Emergent deep tech startups with fewer resources must often use other tactics to create those signals. They might encourage competition among investors ("crafting alternatives") to signal urgency and value. These strategies match Gompers and Lerner's (2001) findings that investor syndicates not only share risk but also signal the quality of a startup, making it easier to complete due diligence.

Building a product that takes a lot of time and money puts an emphasis on strong and frequent signals and nurtured investor relationships. Startups need to time equity rounds carefully around technical milestones. Hallen and Eisenhardt's "scrutinizing interest" strategy helps founders deal with the "cold start" problem. The strategy analyses investor behaviour and network signals in order to find investors who are genuinely interested about the startup. The European Deep Tech Report

(Dealroom et al., 2025) similarly point out the need for patient investors and longer exit timelines in deep tech.

Hsu, 2007 shows that prior entrepreneurial success acts as a credible signal. Experienced founders reduce investor uncertainty in risky innovation contexts like deep tech startups operate in. Similarly, Shane and Stuart (2002) show that early-stage equity financing in science-based startups is more likely when founders are embedded in reputable networks and affiliations.

Deep tech startups face challenging information gaps when trying to attract investors. The reply to the challenge is two folded. Fixed signals like patents to show that what they are working on is unique. Actions in building relationship shape how investors perceive the teams building new innovations. Deep tech startup's success depends on how well it aligns its milestones, build its relationships, and respond to changing market conditions.

## 2.2 Deep Tech Startups: Unique Funding Dynamics

### 2.2.1 Capital intensity vs. software startups.

The way deep tech startups raise funding is shaped by their high capital needs and the balance between investor risk and the demands of R&D-heavy innovation. Nanda and Rhodes-Kropf (2017) explain an important idea about high real option value startups. Some startups have what they call high real option value. These are startups that need repeated testing to solve technical or market problems. High real value option startups are faced with a trade-off: they must raise enough money to stay on track but also keep the flexibility to stop projects that are not working. Raising a lot of money decreases financing risks in the short term, but it might decrease the flexibility at the same time. The challenge grows larger in industries where there are added friction like long development timeliness, strict regulations or expensive infrastructure. Due to uncertain outcomes, investors often prefer staged financing. Staged financing allows investors to back out if startup's key milestones aren't met. For example, if a deep tech startup needs several rounds of funding to keep improving its prototype. In staged funding each round depends on hitting a specific goal to unlock the next stage. Sahlman (2022) emphasizes that staged financing serves as a risk management tool. It allows investors to evaluate progress and withdraw support if necessary. The consequence of the staging approach is that an important role of venture capital firms is to shut down ventures for which they are receiving bad signals (Kerr et al., 2014).

Hellmann and Puri (2000) add to this view by showing how venture capital helps startups become more professional. Venture backed startups typically tend to adopt formal management practices earlier. Formality means better operations which helps alleviate R&D and overall efficiency. Efficiency is critical in capital-intensive sectors. Poor use of resources can hinder progress. Professionalization helps reduce investor concerns by introducing structure to technical planning and financial decisions. VCs might ask startups to hire experienced C-level operators or use stage-gate processes and tie R&D spending to clear milestones. However, such oversight often leads to multiple funding rounds to verify progress, which paradoxically increases exposure to financing risks if future investors question the startup's ability to scale.

Market fluctuation makes the challenges even harder. When the financial markets are in "hot" periods (Nanda & Rhodes-Kropf, 2017), deep tech startups can raise larger funding rounds early. Investor optimism kicks in and allows startups to secure capital for long-term R&D easier. This fits with Hellmann and Puri's (2000) finding that VCs in booming markets focus on speed and pushing startups to invest in infrastructure early. However, investors become more cautious in downturns and prefer startups that can show quick commercial results. Long development timelines put startups at risk. Startups must rely on hitting technical milestones to get more funding. The problem is even worse when they've already made big, irreversible infrastructure investments. The 2025 European Deep Tech Report show that 60% of deep tech startups in regulated sectors need extra funding to handle delays. Only 35% of those startups manage to raise it during market downturns (Dealroom et al., 2025).

Non-dilutive funding helps startups lower early-stage risk without giving up equity. The most common and used forms of non-dilutive funding are government grants or corporate R&D partnerships. But as Colombo et al. (2016) note, this type of funding is usually not enough to scale capital-intensive startups. Well-timed equity rounds are still needed. Startups that raise funding right after hitting key technical milestones can use those achievements to reduce investor uncertainty. This fits with Hallen and Eisenhardt's (2012) idea of "proofpoints" as strong signals. For example, a photonics startup might wait to raise its Series A round until it proves the technology can be manufactured, showing investors that the biggest technical risks are already behind them.

Cockburn & MacGarvie (2011) show how strong incumbent-held patents can deter entry by new startups, particularly in patent-intensive industries. This further shape the strategic choices available for deep tech startup compared to more traditional software startups. Overall, the relationship between capital intensity and risk tolerance in deep tech creates a dual challenge: securing enough funding to support R&D while maintaining the flexibility to pivot or discontinue projects. Nanda and Rhodes-Kropf's (2017) framework, together with Hellmann and Puri's (2000) work on how VCs shape startups, shows that investor patience depends not just on how much capital is available but also on how well financing is managed. If startups align funding, technical milestones and broader market conditions, they are better able to balance bold innovation with long-term financial health.

### 2.2.2 Role of non-dilutive funding

Non-dilutive funding, like government grants and corporate partnerships, plays an important role in helping deep tech startups manage financial risk and shape their commercialization strategies. Grants and partnerships allow startups to achieve technical and regulatory milestones without diluting equity. Founders are able to preserve precious ownership and maintaining flexibility for future funding rounds. Colombo et al. (2016) point out that programs like the U.S. SBIR grants serve two main purposes: they lower the risk of early R&D and send strong signals to future investors. When startups use non-dilutive funding to reach proof-of-concept or preclinical results, they strengthen their position both in equity fundraising and in the "market for ideas" (Gans & Stern, 2003), where deals with larger companies depend on how well the startup can protect and use its intellectual property.

Gans and Stern's framework demonstrates how the environment for bringing innovations to market shapes how well non-dilutive funding works. Especially the ability to keep competitors out and control key resources. Using grants to build strong intellectual property (IP) improves startups position in the market for ideas. Strong IP helps making licensing deals or form partnerships with larger companies. For example, a deep tech startup that uses grant funding to patent a new technology can negotiate better terms with partnering companies by using its patents to protect its work. This fits with Gans and Stern's idea that strong IP protection reduces the risk of being copied and helps startups keep leverage after they share their ideas. In sectors where IP protection is weaker, non-dilutive funding has a different kind of role. For example, in cleantech where IP protection is weaker, startups may use grants to build unique manufacturing processes or pilot plants. Building unique infrastructure allows startups to compete directly in the market without needing to rely on partnerships with bigger firms.

Corporate partnerships are built on top of the connection between non-dilutive funding and commercialization strategy. Startups can access complementary assets without giving up equity by collaborating with established firms. The assets can be for example manufacturing capabilities, distribution networks, or regulatory expertise. Corporate partnerships fit into Gans and Stern's "market for ideas" model, where startups use the strengths of larger firms to bring their innovations to market. But these partnerships also come with risks, especially if the startup's IP is not fully protected. To manage the risk, startups may take a phased approach: using early grant-funded milestones to show technical progress before sharing sensitive information. A good example is Genentech's partnership with Eli Lilly on synthetic insulin (Gans & Stern, 2003). Genentech first used grants to secure key patents, then used those patents as the basis for a licensing deal that let them keep a share of the profits, even though Lilly controlled the path to market.

Non-dilutive funding options come with special challenges of their own. Startups have face tough trade-offs in sectors where it's harder to protect idea, sectors like advanced materials and robotics. Corporate partners may use their stronger position to take advantage of the startup, as seen in Robert Kearns' long legal battle with Ford over windshield wiper technology (Gans & Stern, 2003). Startups need to carefully time partnerships so that they have reached technical milestones that strengthen their position. Also, non-dilutive funding is usually not enough to scale capital-heavy businesses, so equity funding is still needed. Colombo et al. (2016) note out that delaying institutional investment until after key milestones can reduce dilution, but startups risk missing the right market window if broader economic conditions change.

Startups environment eventually determines whether it will choose to use some form of non-dilutive funding or if it will form a strategic partnership with some other company. In IP-heavy sectors like biotechnology, startups often use grants to strengthen their ability to protect their ideas, which helps them succeed in the market for ideas through licensing or partnerships. In contrast, startups in asset-heavy, low-IP sectors may use grants to build their own production or technical capacity. This reflects Gans and Stern's idea that when IP protection is weak, startups are more likely to compete directly rather than cooperate. Deep tech startups can find better paths to market and deal more effectively with the timing and funding challenges of high-risk innovation by matching their non-dilutive funding strategies to the level of IP protection and the need for outside resources.

## 2.3 Prior Studies on Funding Timing and Startup Outcomes

Prior research on the timing aspects of venture funding has provided valuable insights into how financing gaps and investor selection affect startup outcomes. Yet there are still significant gaps in understanding what decisions and characteristics affect the high-risk, capital-intensive ventures especially in the early stages. Ewens and Farre-Mensa (2020) report a decline in IPOs and a widening funding gap for growth-stage startups. They link this trend to rising regulatory costs and the growing influence of private "unicorn" investors. Ewens' and Fare-Mensa's analysis shows that many startups now delay going public to avoid the extra scrutiny. More and more private companies rely on late-stage private funding rounds instead of going public. Private rounds may be better deal for investors, but worse deal for the rest of the ecosystem as a whole. Avoiding going public works well for those startups who have a clear path to profitability. Deep tech typically has long R&D timelines and uncertain paths to commercialization. Ewens and Farre-Mensa note that the shift towards private capital raises systemic risks. If the money won't circle back to early stages, the speed of innovation slows down.

Puri and Zarutskie (2012) compare in their study VC-backed startups and non-VC-backed startups in relation to how long the startups survive. Their study shows that venture capital gives startups a clear survival advantage. The study shows that VC-backed startups are 25% more likely to survive over five years. The advantages come from factors like investor monitoring, access to networks, and milestone-based funding. However, the study focuses mostly on software and consumer startups, which are known for being capital-efficient and quickly scaling. Deep tech startups have to deal with high technical uncertainty and often face complex regulations. There are mixed views whether the deep tech startups benefit as much from traditional VC oversight. For example, staged funding helps control risk in asset-light startups (Gompers, 1995). But in deep tech, it can cut short the long R&D cycles that need steady, long-term investment. This points to a major gap in the research. VC success patterns from low-capital sectors are not easily applicable to high-risk, science-based innovation. Pahnke et al. (2023) add that startup success is more likely when there is resource interdependence between the startup and its investors. It aligns incentives and increases VC engagement toward achieving a successful exit.

Bernstein et al. (2016) examine how venture capital monitoring affects innovation in high-risk ventures. In their natural experiment study, they show how reduced travel costs for venture capitalists affect their behaviour and the venture outcomes. Results show that VCs become more involved on-site. As a result startups produce more patents and are more likely to reach a successful exit. Active investor oversight appears to help reduce technical and operational risks. The study focuses on IT and healthcare startups. It uncertain whether the findings apply to other deep tech sectors with longer development timelines. The study also uses geographic closeness as a stand-in for how involved an investor is.

Colombo et al. (2016) examine how the added friction of regulation affects the venture funding. Strict product market regulations are a major barrier to venture capital. The analysis shows that startups in regulated industries often face longer funding intervals. Investors hesitate to commit until critical regulatory milestones are met. Non-dilutive grants help reduce early-stage risks, but Colombo et al. warn that complex regulations weaken the signaling value of this funding. Investors may overlook early achievements if they are overshadowed by ongoing compliance challenges. Startups often need to spend a lot of capital upfront to meet regulatory requirements. If funding comes too late it can force startups to make tough choices between giving up more equity or risking their survival.

Earlier studies in venture funding and timing of the investments show that the timing plays a complex role in innovation. Timing can help drive progress but it can also create constraints. Ewens and Farre-Mensa (2020) and Puri and Zarutskie (2012) show strong links between steady financing and startup survival. Their models however, do not fully reflect the nonlinear, milestone-based paths that deep tech ventures follow. Bernstein et al. (2016) and Colombo et al. (2016) show the effects of investor involvement and regulatory barriers. There is a gap in research when it comes to funding intervals in deep tech funding. Future research should address the gap and study the links between funding timing strategies, of high-risk innovation milestones and high-risk innovation team characteristics. This study is the first step into that direction.

# 3 Methodology

This chapter explains how the data for the empirical study on European and U.S. deep tech startups was collected, cleaned, and analyzed. The data consists of deep-tech startups from Europe and the U.S. founded between 2010 and 2020. The first chapter describes how the proprietary data from Dealroom was collected and filtered. The second chapter describes how each variable was defined and coded for the empirical study. The third chapter outlines the statistical models to study the effects of timing in deep tech venture outcomes. The study uses range of statistical methods tarting with a basic logistic regression, followed by flexible spline-based linear models, and ending with a full multiple regression that includes all predictor variables.

## 3.1 Data Sources and Collection

All data used in this study comes from Dealroom's startup database. Dealroom is a commercial platform for investors, founders and other ecosystem entities. It's an industry standard as a data source for VCs researching s specific startup, industry or investment thesis. Dealroom gathers global startup ecosystem data through a hybrid approach. Data is being collected with both algorithm based web scraping and with community-sourced input from venture capital firms, startups, universities, and other ecosystem participants.

The data was exported from Dealroom on April 18, 2025. It includes all startups tagged as "deep tech," founded between 2010 and 2020, with headquarters in Europe or North America. The initial export had approximately 23,000 observations. After cleaning and removing any firm missing data required for the regressions; the final sample was reduced to about 6,200 startups. The most common missing variables that led to exclusions were the date of first funding and valuation.

## 3.2 Data Preparation

The following edits and transformations were applied to the dataset:

1. Binary Outcome Variables
    a. The binary outcome variables (Success, Unicorn, IPO>500, M&A>500, Funding>100) were imported from Dealroom using Dealroom's built-in filtering criteria to extract relevant data for the study.
2. Valuation Data

   a. In cases where valuations were listed as a range (i.e., Valuation Low and Valuation High), the lower end of the range was selected. Low end of the valuation approach maintained consistency and ensured that the valuation data reflects the most conservative estimate.

3. Calculation of Months Between Launch and First Funding

   a. The variable months_between was calculated as the number of months between the startup's launch date and its first funding date. The launch date was subtracted from the first funding date.

4. Primary Industry Selection

   a. The first industry listed in the industries column was selected as the primary industry. This was done to simplify classification.

5. Primary Client Focus

   a. The first entry from the Client_focuses list was selected as the primary client focus. This was done to simplify classification.

6. Primary Revenue Model

   a. The first entry from the Revenue_models list was selected as the primary revenue model. This was done to simplify classification.

## 3.3 Variable Definition and Operationalization

This study examines two types of outcomes: binary success/failure and quantifiable company valuation. Both driven by the timing of first funding and other variables. Gornall & Strebulaev (2020) show in their venture capital valuation study that many startup valuations are inflated due to investor-friendly contractual terms. They add that accurate valuation estimates should adjust for liquidation preferences and other rights embedded in VC deals. The empirical study of this thesis uses the lower end of the valuations for the deep tech startups.

A startup is defined as successful if it had at the time of the data collection achieved any of the following:

1. "Unicorn" status (valuation ≥ € 1 B).

2. An IPO, acquisition, or merger at a valuation ≥ € 500 M.

3. Total funding raised exceeding € 100 M.

All others are defined as unsuccessful.

The main predictor variable is the number of months_between company/product launch and first VC investment. This metric can be negative (pre-launch funding) or positive (post-launch). Variable number_of_patents is also used to proxy technological sophistication, HQ_region (Europe vs. North America) as a geography dummy, and primary_industry as categories.

For quantifiable outcome variable, this study uses natural logarithm of low-end valuation plus one:

$$log(Valuation\_low\_EUR + 1).$$

### 3.4 Statistical Modeling Strategy

To test how funding timing influences both success probability and valuation, the study estimates three regressions: logistic, linear (with and without splines), and a multivariate model.

#### 3.4.1 Logistic regression

Binary success indicator using a three-degree natural spline in funding delay:

$$\Pr(success_i = 1) = \frac{1}{1 + \exp[-(\beta_0 + \mathrm{f}(\mathrm{months\_between}_i))]1}$$

Where $f\left(months_{between_i}\right)$ is the spline basis with df = 3 and $i$ is index individual startups.

With this flexible specification, the effect of funding timing on success isn't assumed to be straight-line. It can rise or fall as months between launch and first funding get shorter or longer.

In logistic regression testing the effects of economic cycles, growth variable is added to the model:

$$\Pr(success_i = 1) = \frac{1}{1 + \exp[-(\beta_0 + \mathrm{f}(months\_between_i) + \beta_4 GDPgrowth_i)]1}$$

Where $f\left(months_{between_i}\right)$ is the spline basis with df = 3 and $i$ is index individual startups.

### 3.4.2 Linear vs. spline models

Baseline "simple" linear model regresses log-valuation on patents and region only. The "spline" model adds the same three-degree natural spline in funding timing:

$$log(Valuation_i + 1) = \alpha_0 + \alpha_1 number\_of\_patents_1 + \alpha_2 I[hq\_region_i$$
$$i$$

$$= "North\ America"] + g(months\_between_i) + \varepsilon_i$$

Where $g(months_{between_i})$ is the spline basis with df = 3 and $i$ is index individual startups.

Splines are chosen to accommodate potential non-linear effects.

**ANOVA Comparison**

To test whether the spline improves fit, the study performs an F-test comparing residual sums of squares (RSS) from the simple and spline models:

$$F = \frac{(RSS_{simple} - RSS_{spline})/\Delta df}{RSS_{spline}/df_{spline}} \sim F_{\Delta df, df_{spline}}.$$

A significant p-value indicates that timing splines explain valuation variation beyond patents and region.

**Diagnostics and Goodness-of-Fit**

The study uses three different diagnostics checks: $R^2$ comparisons, Residuals vs. Fitted plots and Normal Q-Q plots of standardized residuals. $R^2$ comparisons quantify how much additional variance the spline specification captures. Residuals vs. Fitted plots reveal whether residuals show non-random patterns. Normal Q-Q plots of standardized residuals assess whether errors approximate normality, a key assumption for valid confidence intervals and hypothesis tests.

### 3.4.3 Multiple regression

Multiple regression model extends from the linear model to include industry dummies and national GDP growth rates. It isolates the marginal effect of funding timing in a rich set of controls:

$$\log(Valuation_i + 1) =$$

$\gamma_0 + \gamma_1 months\_between_i + number\_of\_patents_i + \gamma_3 I[hq\_region_i = "North\ America"] + \sum_k \gamma_{4k} \mathrm{I}\{\mathrm{industry}_i = \mathrm{k}\} + \mathrm{u}_i.$

The model tests whether patents, geography, industry sector, and macro cycle, the timing of first funding continues to have an affect for eventual firm valuation.

# 4 Empirical Analysis

## 4.1 Data Analysis

The final sample includes 6,226 deep-tech startups, all with first funding dates between 2010 and 2020. Before presenting the main empirical results, the Data Analysis chapter provides an overview of the dataset's key dimensions: geography, time, industry, and client focus.

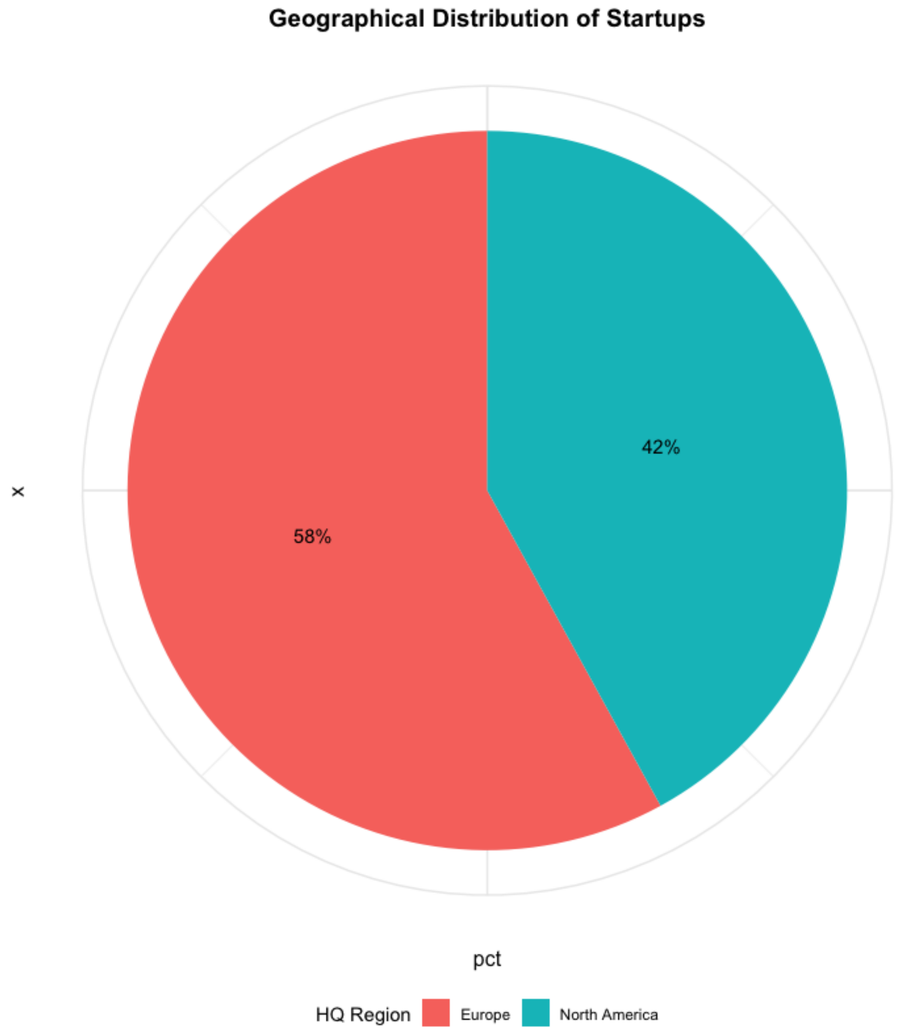


Figure 1. Geographical distribution of startups by region.

Startups in the sample are concentrated in two regions: Europe (3,611 firms; 58.0 %) and North America (2,615 firms; 42.0 %). Figure 1 shows the split to these two regions. The spllit reflects the relative maturity of venture ecosystems in these two markets, with Europe contributing slightly more deals than North America but still trailing in total VC volume.

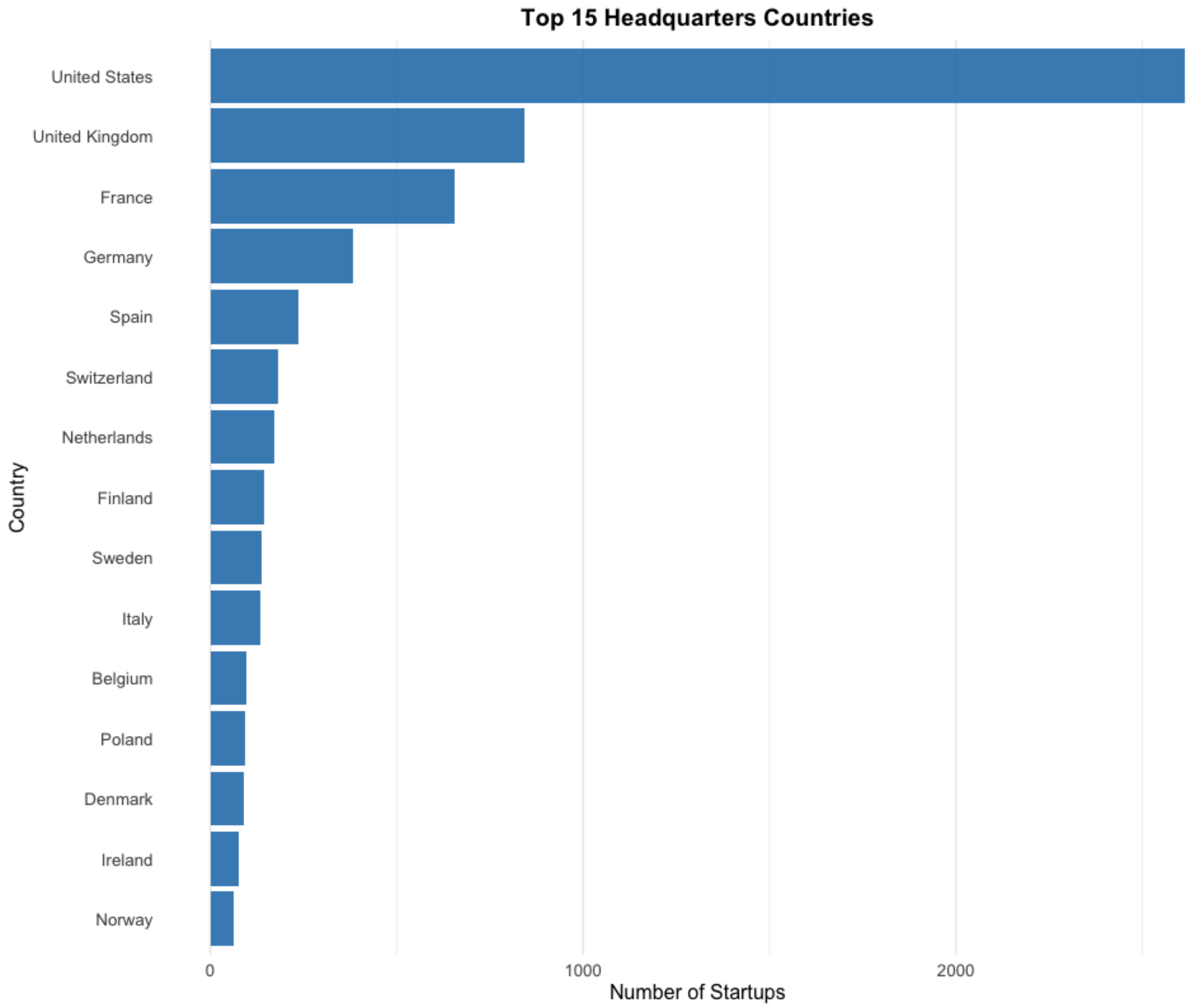


Figure 2. Top countries in represented in the data.

Within those two regions, the United States accounts for 2,615 startups (42 % of the sample). The figure 2 shows the next most common locations are the United Kingdom (841; 13 %), France (657; 10 %), and Germany (383; 6 %). Spain (236), Switzerland (183), the Netherlands (173), Finland (143), and Sweden (137) round out the top 10, with Italy (133), Belgium (96), Poland (95), Denmark (89), Ireland (75), and Norway (61) completing the top 15.

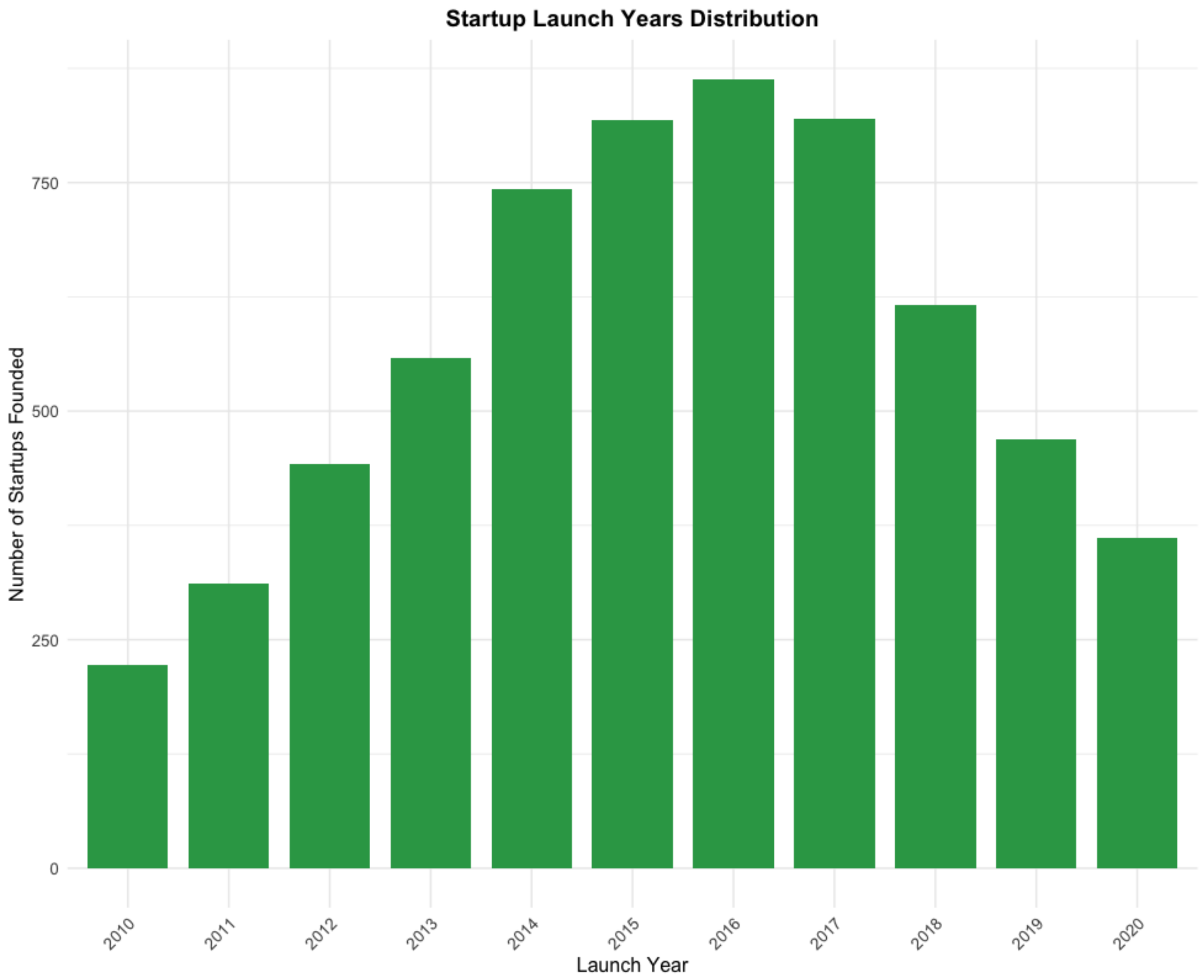


Figure 3. Distribution of launch years.

Startup founding peaked in 2016 (863 firms), following steady growth from 223 in 2010 to 819 in 2015. Figure 3 shows that after 2016, annual launch counts drift down to 820 in 2017, then 616 in 2018, 469 in 2019, and 361 in 2020. The decline in launches after 2016 may be affected by the cut off year of the data sample approaching, due the end of the investment cycle or due to early pandemic disruptions.

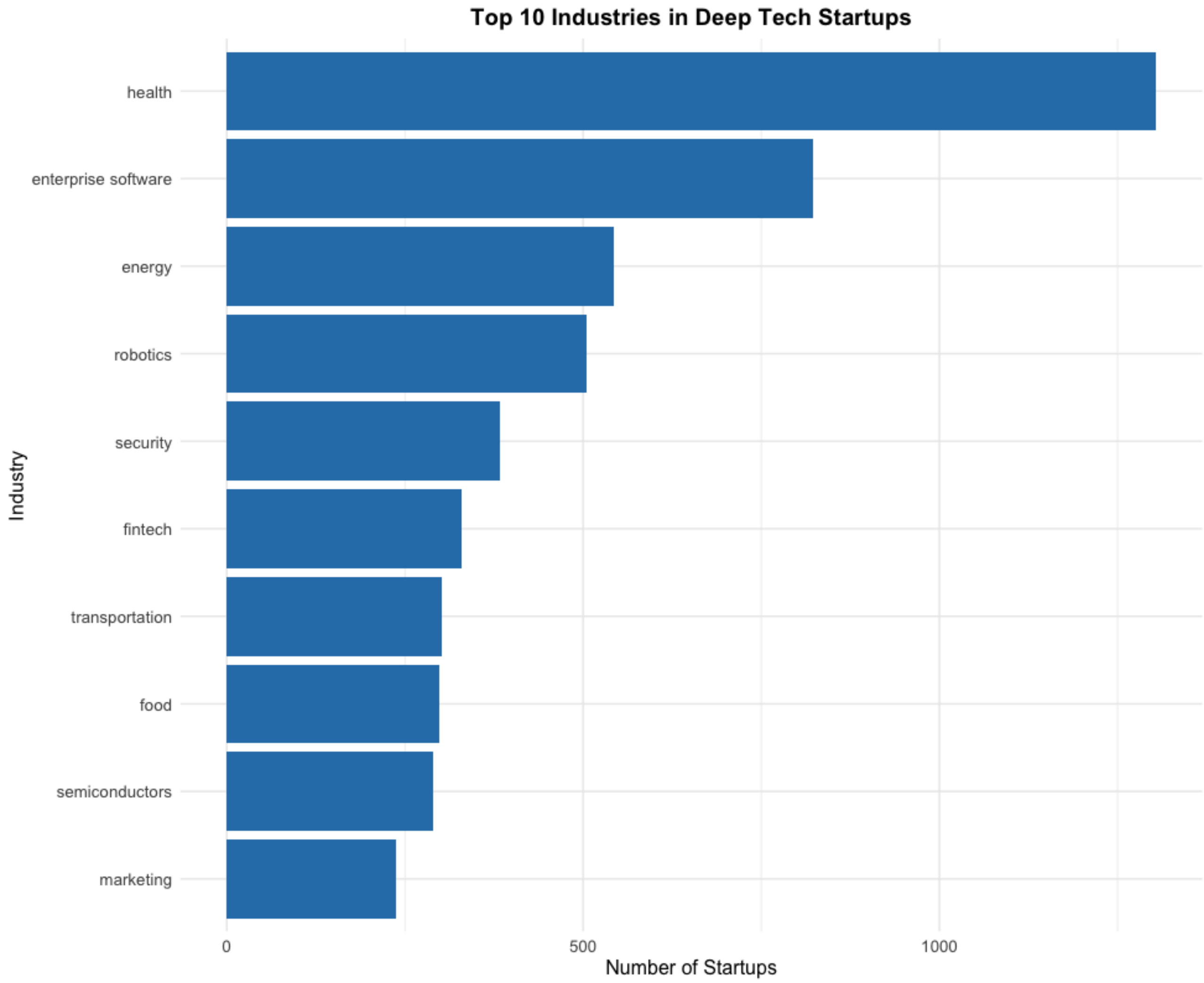


Figure 4. Top industries among deep tech startups studied.

The single largest sector is Health (1,304 firms; 21 %), followed by Enterprise Software (822; 13 %), Energy (543; 9 %), and Robotics (505; 8 %). Figure 4 shows that Security (384), Fintech (329), Transportation (301), Food (299), Semiconductors (290), and Marketing (238) complete the top 10 industries, together accounting for over 80 % of the sample.

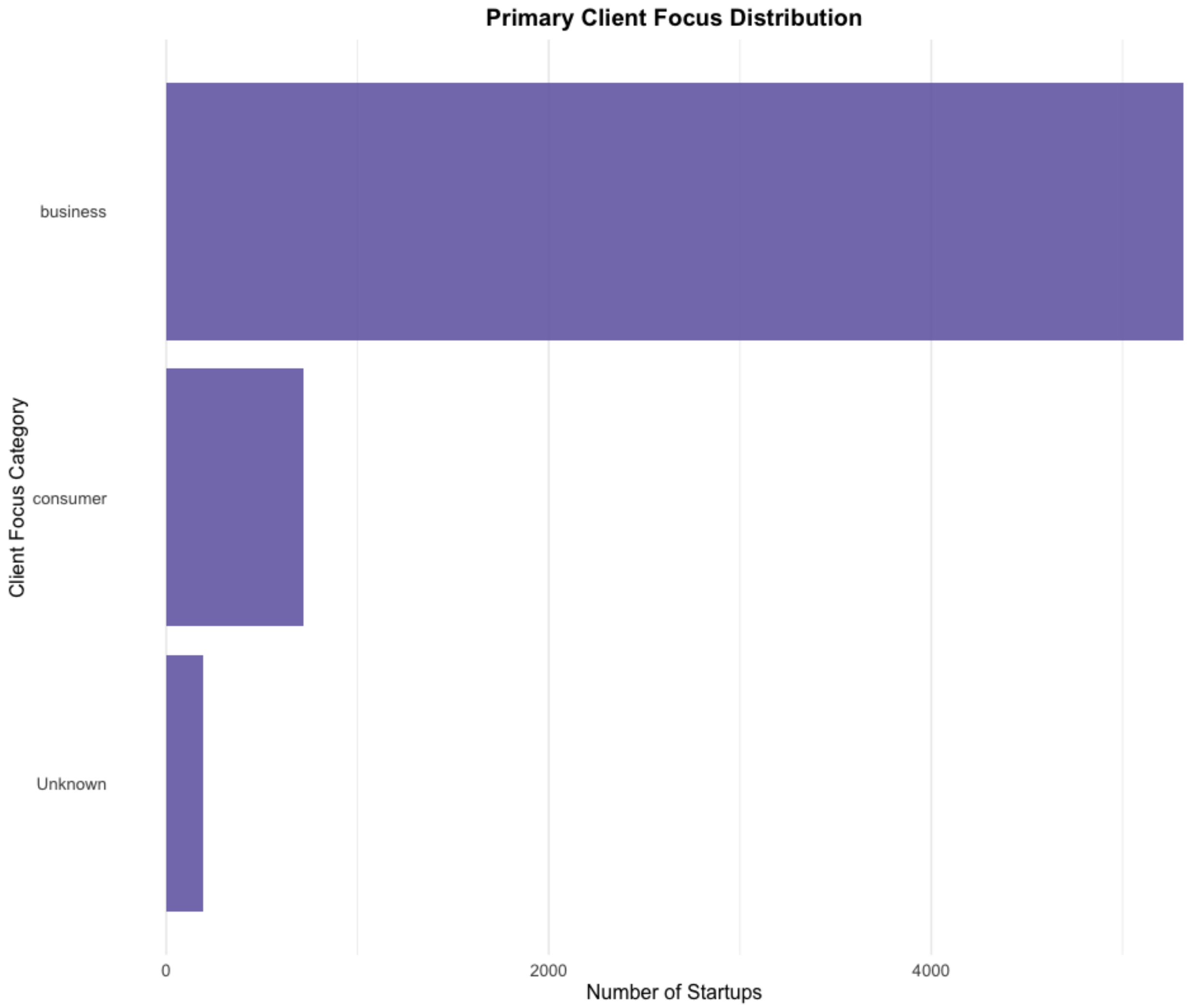


Figure 5. Distribution of primary client focus of the deep tech startups.

A vast majority of deep-tech startups target business customers: 5,316 firms (85.4 %) are B2B-oriented, while 717 (11.5 %) focus primarily on consumers. The final bar on figure 5 shows a small remainder of startups (193; 3.1 %) whose primary client focus could not be classified and are labelled “Unknown.”


Figure 6. Counts and success rates by timing of first funding relative to launch.

The first bar on figure 6 shows that firms that secure capital within six months after launch enjoy the highest success rate (about 12 percent) nearly double the sample average of 6.8 percent. This suggests that entrepreneurs who wait to demonstrate early traction or a working prototype before seeking institutional backing are able to leverage that momentum most effectively. By contrast, companies that raise their first round up to six months before launch fare poorly, with a success rate below 4 percent, indicating that pre-launch financings may leave startups under-prepared to meet market demands. Finally, ventures that obtain funding more than six months after launch, whether before or after that six-month threshold, exhibit mid-range success rates of approximately 5–6 percent, closely aligned with the overall mean. Taken together, these patterns point to a “sweet spot” immediately following product introduction: too early and firms lack the proof points investors seek; too late and they may have already burned through scarce resources without reaching key milestones. In the next chapter, we will model timing as a continuous variable, allowing us to trace this non-linear effect in greater detail and to control for industry and macroeconomic conditions.

Together, these six descriptive figures demonstrate that our deep-tech sample is heavily centered on mature venture hubs in Europe and North America, with a strong skew toward B2B-oriented, health- and enterprise-software companies founded in the mid-2010s. Next section covers the core empirical analysis of funding-timing effects.

## 4.2 Regression Results

### 4.2.1 Logistic regression

Logistic regression models the probability of a “successful” exit as a flexible function of the number of months between launch and first VC funding, using a 3-degree natural spline. Table 1 below shows the estimated spline-basis coefficients.

**Logistic Regression**

***Table 1: Logistic Regression***

| Term | Estimate | Std_Error | Z_value | P_value |
|---|---|---|---|---|
| (Intercept) | -11 | 3.298 | -3.335 | 0.001 |
| ns(months_between, df = 3)1 | 2.599 | 1.656 | 1.569 | 0.117 |
| ns(months_between, df = 3)2 | 16.625 | 6.476 | 2.567 | 0.01 |
| ns(months_between, df = 3)3 | 4.488 | 2.071 | 2.167 | 0.03 |
| Null deviance | 3096.8 | | | |
| Residual deviance | 3059.8 | | | |
| AIC | 3067.8 | | | |
| DF (null) | 6225 | | | |
| DF (residual) | 6222 | | | |

Note: Fitted logistic-regression model object estimating “success” as a non-linear spline function of months_between (3 degrees of freedom). Outcome variable success is defined as a binary indicator (0/1): (IPO or acquisition > €500M, unicorn, or raised >

| Table 1: Logistic Regression | | | | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Std_Error** | **Z_value** | **P_value** |
| €100M). Predictor variable number of months between the firm's launch date and its first institutional funding round. Computed from "launch_date" and "first_funding_date". | | | | |

Only the second and third spline coefficients differ significantly from zero, confirming that success probability is a nonlinear function of time between launch and funding. The second spline has highly elevated log-odds and an estimate of +16.625 (SE = 6.476, z = 2.567, p ≈ 0.01). It captures roughly the period following immediately after launch. The third spline has an estimate of +4.488 (SE = 2.071, z = 2.167, p ≈ 0.03). It shows a moderate positive effect on success probability for the later part of the time distribution. There is a small but significant benefit for this later time funding, but not as significant as for the time captured by the second spline.

The intercept is highly negative –11.000 (SE = 3.298), indicating very low baseline odds of success if funding timing is far from the central spline ranges. The first spline term estimate is +2.599 (SE = 1.656, z = 1.569, p ≈ 0.12) but it is not statistically significant. Early funding shows a positive but not statistically reliable effect.

To visualize where in the funding timing spectrum the odds of success rise and fall, Figure 7 plots the fitted probability curve.

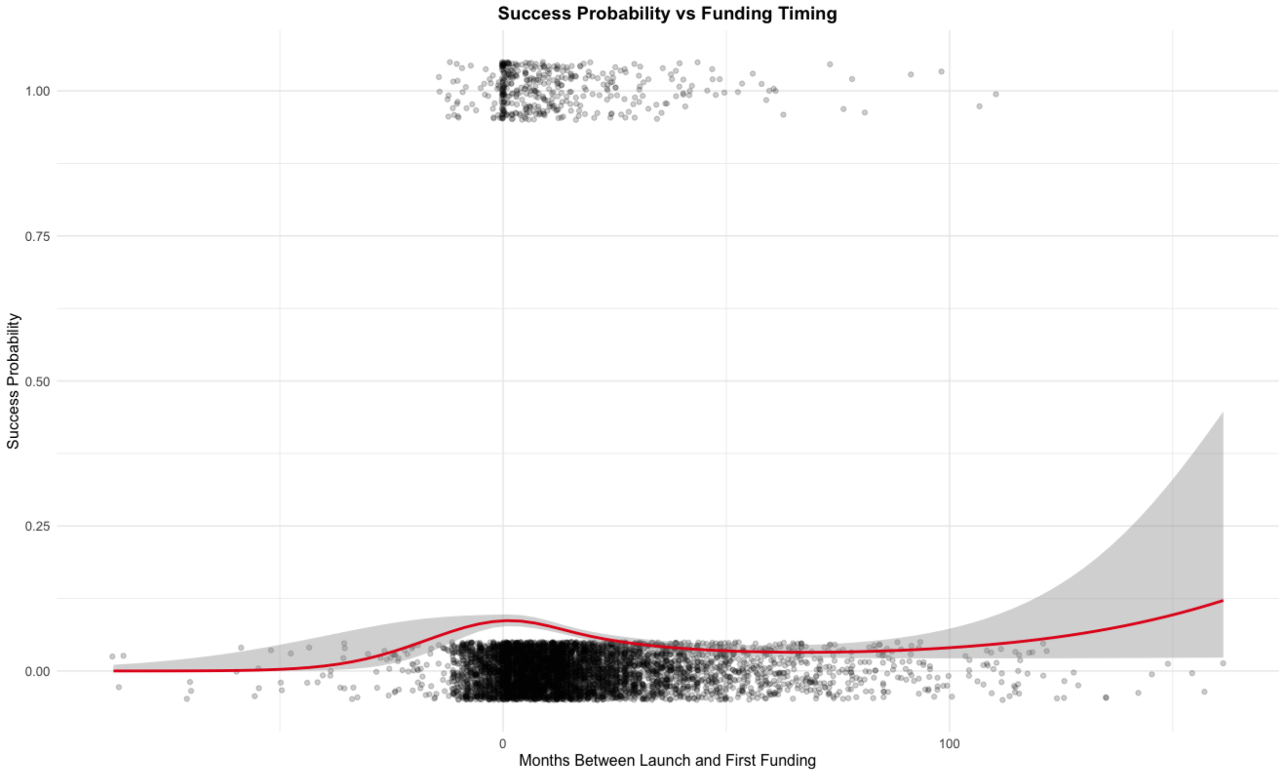


Figure 7. Success probability vs funding timing.

The curve in figure 7 peaks for the funding timing delays at roughly three to eight months post-launch and then declines, illustrating a single "sweet-spot" for first-round timing.

### 4.2.2 Linear regression models

**Simple Patent Impact Model**

| *Table 2a: Patent Impact Model (Simple)* | | | | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Std_Error** | **t_value** | **P_value** |
| (Intercept) | 15.941 | 0.034 | 467.79 | 0 |
| number_of_patents | 0.006 | 0.001 | 9.903 | 0 |
| hq_regionNorth America | 0.642 | 0.053 | 12.225 | 0 |
| Residual Std. Error | 2.046 | | | |
| Multiple $R^2$ | 0.0395 | | | |
| Adjusted $R^2$ | 0.0391 | | | |

| Table 2a: Patent Impact Model (Simple) | | | | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Std_Error** | **t_value** | **P_value** |
| F-statistic | 127.81 | | | |
| DF (model) | 2 | | | |
| DF (residual) | 6223 | | | |
| p-value (F) | <0.001 | | | |

Note: Fitted linear-regression model object predicting log_valuation from number_of_patents + hq_region. Outcome variable is the reported "low" (floor) valuation of the latest funding round, in euros. The regression uses natural log of valuation_low_eur + 1. Predictor variable number of months between the firm's launch date and its first institutional funding round. Computed from "launch_date" and "first_funding_date" Other predictor variables are total count patents held by the startup, as of the latest date and categorical variable for the startup's headquarters region.

Table 2a regresses log-valuation on the number of patents and a North America region dummy. The number of patents is strongly positive and highly significant ($p < 2e-16$). The coefficient on patents ($\approx 0.006$, $p<0.001$) implies that each additional patent is associated with a 0.6% higher valuation, all else equal.

The estimate for startup region dummy (North America) is +0.6425 (SE = 0.05255, t ≈ 12.225). This means that being based in North America corresponds to a ~90% higher median valuation ($e^{0.642} \approx 1.90$).

The intercept is high, estimate being 15.941 (SE = 0.03408, t ≈ 467.790). The baseline log-valuation level is very large (≈ 8.8 million EUR) for a hypothetical non-U.S. startup with zero patents.

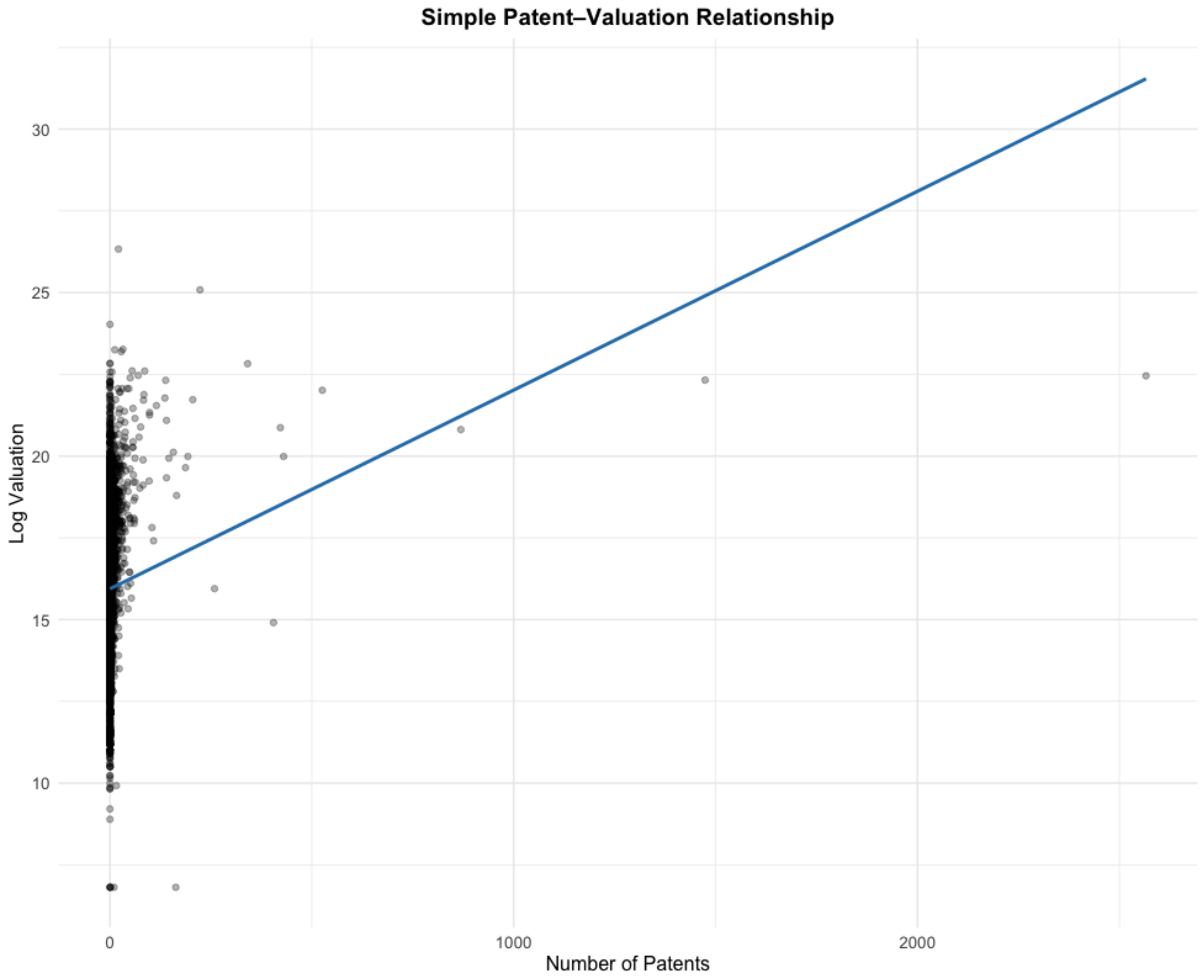


Figure 8. Simple patent–valuation relationship.

Figure 8 shows the fitted line for "Europe" startups: it shows an upward slope confirming that larger patent portfolios translate into higher valuations on a log scale.

### Spline-augmented Patent Impact

*Table 2b: Patent Impact Model (Spline)*

| Term | Estimate | Std_Error | t_value | P_value |
|---|---|---|---|---|
| (Intercept) | 16.432 | 0.672 | 24.442 | 0 |
| ns(months_between, df = 3)1 | -0.713 | 0.364 | -1.956 | 0.051 |

| *Table 2b: Patent Impact Model (Spline)* | | | | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Std_Error** | **t_value** | **P_value** |
| ns(months_between, df = 3)2 | -0.289 | 1.347 | -0.215 | 0.83 |
| ns(months_between, df = 3)3 | 0.886 | 0.568 | 1.561 | 0.119 |
| number_of_patents | 0.006 | 0.001 | 9.848 | 0 |
| hq_regionNorth America | 0.653 | 0.053 | 12.382 | 0 |
| Residual Std. Error | 2.044 | | | |
| Multiple R² | 0.0413 | | | |
| Adjusted R² | 0.0405 | | | |
| F-statistic | 53.55 | | | |
| DF (model) | 5 | | | |
| DF (residual) | 6220 | | | |
| p-value (F) | <0.001 | | | |

Note: Fitted linear-regression model object predicting log_valuation from a spline on months_between (df = 3) + number_of_patents + hq_region. Outcome variable is the reported “low” (floor) valuation of the latest funding round, in euros. The regression uses natural log of valuation_low_eur + 1. Predictor variable number of months between the firm’s launch date and its first institutional funding round. Computed from “launch_date” and “first_funding_date” Other predictor variables are total count patents held by the startup, as of the latest date and categorical variable for the startup’s headquarters region.

Table 2b adds funding-timing spline to the simple specification. None of the three spline terms reach conventional significance, although the first spline term is close to the 5% threshold ($p \approx 0.051$). The first spline estimate is +0.65305 (SE = 0.05274, $t \approx 12.382$), suggesting that raising funding well before launch is associated with marginally lower valuations.

The patent and region effects remain highly significant. Number of patents estimate: +0.00604 (SE = 0.0006135, t ≈ 9.848) and the Northe American region dummy estimate is +0.65305 (SE = 0.05274, t ≈ 12.382). The $R^2$ rises from 3.95% to 4.13%.

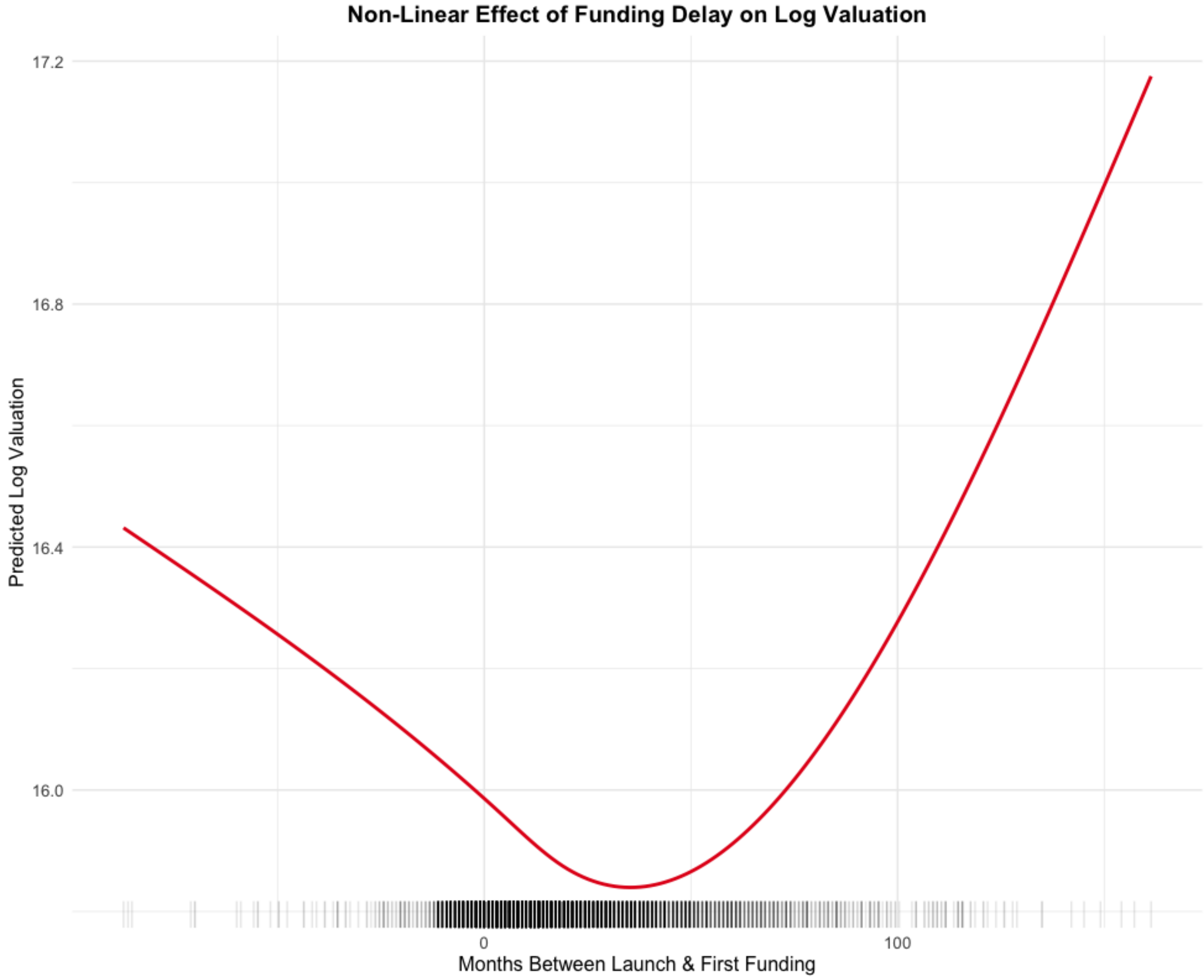


Figure 10. Non-linear effect of the time between launch and funding to log valuation.

Figure 10 shows the non-linear effect of time between funding on predicted log-valuation (holding median patent count and "Europe" region constant). The very short and very long times between launch and funding slightly depress valuation, whereas moderate delays lift it, mirroring the logistic pattern on the log-valuation scale. The "rug" of actual observations along the x-axis shows where most data lie.

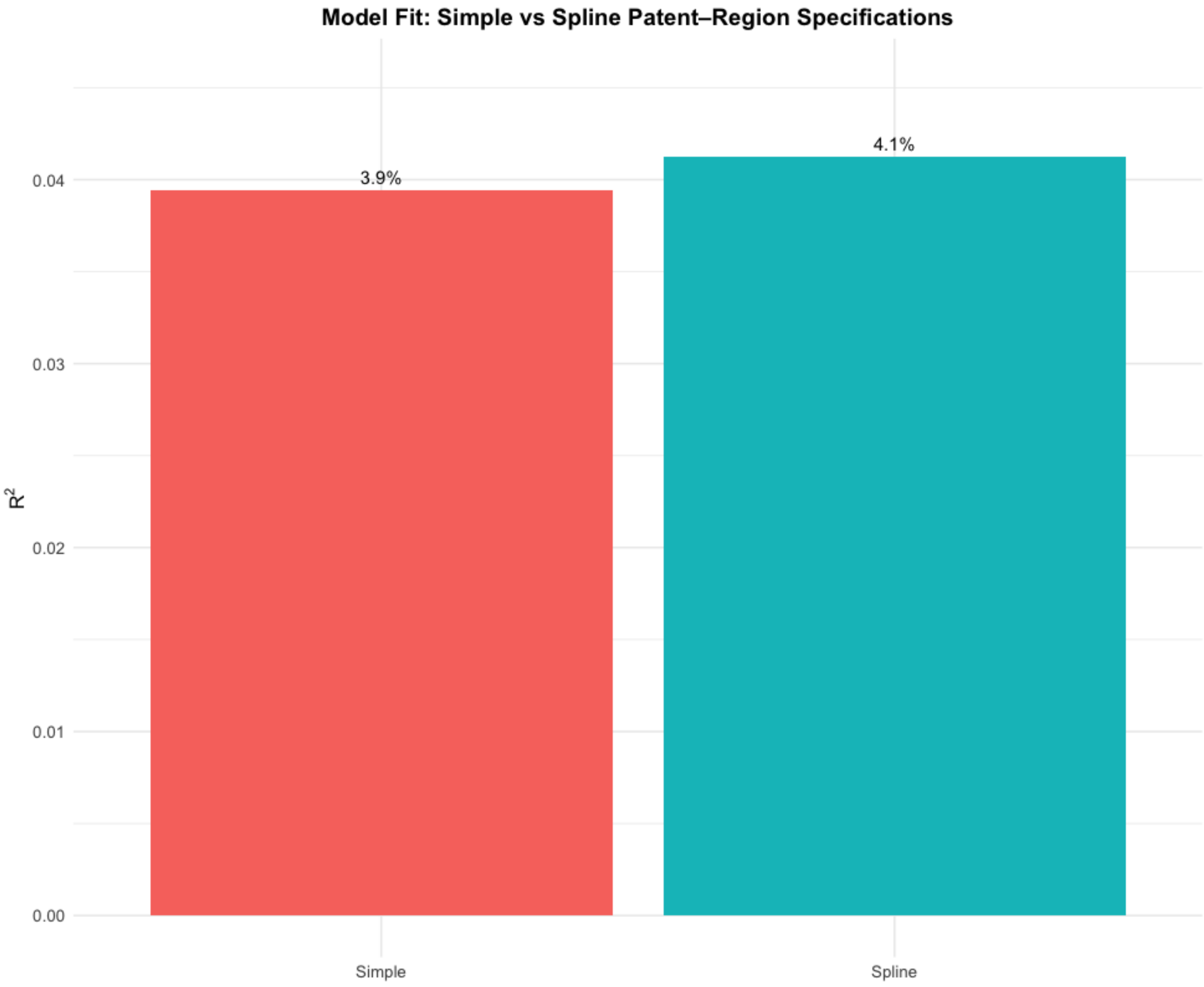


Figure 11. Model fit improvement over simple patent model.

The modest $R^2$ improvement is visually summarized in Figure 11, where the spline bar edges out the simple model. Adding funding timing splines yields a small but statistically significant $R^2$ increase ($\Delta R^2 \approx 0.0018$).

**Model Comparison via ANOVA**

An ANOVA (Table 2c) confirms that adding the three spline terms to the simple patent model yields a small but statistically significant reduction in residual sum of squares ($\Delta RSS \approx 49$, $F \approx 3.92$, $p \approx 0.008$). In practical terms, the timing features add explanatory power to the patent-valuation relationship. However, the added explanatory power is modest.

***Table 2c: ANOVA Comparison (Simple vs Spline)***

| Comparison | Residual_DF | RSS | DF_Change | SS_Change | F_value | P_value |
|---|---|---|---|---|---|---|
| Model 1 | 6,223 | 26,039.33 | | | | |
| Model 2 | 6,220 | 25,990.25 | 3 | 49.086 | 3.916 | 0.008 |

Note: where Model 1 is “log_valuation ~ number_of_patents + hq_region” and Model 2 is “log_valuation ~ ns(months_between, df = 3) + number_of_patents + hq_region”.

### 4.2.3 Multiple regression

Finally, Table 3 shows a fully adjusted log-valuation model including months-between, patents, region, and industry dummies.

***Table 3: Full Multiple Regression Valuation Model***

| **Term** | **Estimate** | **Std_Error** | **t_value** | **P_value** |
|---|---|---|---|---|
| (Intercept) | 14.981 | 0.259 | 57.888 | 0 |
| months_between | 0 | 0.001 | -0.19 | 0.849 |
| number_of_patents | 0.006 | 0.001 | 9.693 | 0 |
| hq_regionNorth America | 0.711 | 0.053 | 13.525 | 0 |
| primary_industrychemicals | 1.621 | 0.659 | 2.46 | 0.014 |
| primary_industryconsumer electronics | 1.856 | 0.718 | 2.585 | 0.01 |
| primary_industryeducation | -0.139 | 0.326 | -0.427 | 0.67 |
| primary_industryenergy | 1.342 | 0.27 | 4.962 | 0 |
| primary_industryengineering and manufacturing equipment | 1.632 | 0.422 | 3.866 | 0 |
| primary_industryenterprise software | 1.098 | 0.265 | 4.141 | 0 |
| primary_industryevent tech | 0.056 | 0.551 | 0.101 | 0.919 |
| primary_industryfashion | 0.711 | 0.422 | 1.685 | 0.092 |
| primary_industryfintech | 0.581 | 0.279 | 2.084 | 0.037 |

*Table 3: Full Multiple Regression Valuation Model*

| Term | Estimate | Std_Error | t_value | P_value |
|---|---|---|---|---|
| primary_industryfood | 0.984 | 0.281 | 3.499 | 0 |
| primary_industrygaming | 0.114 | 0.295 | 0.388 | 0.698 |
| primary_industryhealth | 1.077 | 0.262 | 4.11 | 0 |
| primary_industryhome living | -0.084 | 0.387 | -0.216 | 0.829 |
| primary_industryhosting | 0.056 | 0.422 | 0.132 | 0.895 |
| primary_industryjobs recruitment | 0.422 | 0.415 | 1.017 | 0.309 |
| primary_industrykids | -0.507 | 0.564 | -0.898 | 0.369 |
| primary_industrylegal | 0.979 | 0.352 | 2.78 | 0.005 |
| primary_industrymarketing | 0.607 | 0.287 | 2.115 | 0.034 |
| primary_industrymedia | 0.632 | 0.312 | 2.022 | 0.043 |
| primary_industrymusic | 0.406 | 0.477 | 0.851 | 0.395 |
| primary_industryreal estate | 1.008 | 0.311 | 3.245 | 0.001 |
| primary_industryrobotics | 0.664 | 0.271 | 2.451 | 0.014 |
| primary_industrysecurity | 1.419 | 0.276 | 5.149 | 0 |
| primary_industrysemiconductors | 0.983 | 0.282 | 3.483 | 0 |
| primary_industryspace | 2.283 | 0.37 | 6.173 | 0 |
| primary_industrysports | 0.222 | 0.387 | 0.573 | 0.567 |
| primary_industrytelecom | 1.284 | 0.32 | 4.008 | 0 |
| primary_industrytransportation | 0.882 | 0.281 | 3.139 | 0.002 |
| primary_industrytravel | 0.415 | 0.43 | 0.966 | 0.334 |
| primary_industrywellness beauty | 0.468 | 0.564 | 0.829 | 0.407 |
| Residual Std. Error | 2.012 | | | |

| Table 3: Full Multiple Regression Valuation Model | | | | |
|---|---|---|---|---|
| **Term** | **Estimate** | **Std_Error** | **t_value** | **P_value** |
| Multiple $R^2$ | 0.0749 | | | |
| Adjusted $R^2$ | 0.07 | | | |
| F-statistic | 15.2 | | | |
| DF (model) | 33 | | | |
| DF (residual) | 6192 | | | |
| p-value (F) | <0.001 | | | |

Note: Fitted linear-regression model object predicting log_valuation from months_between + number_of_patents + hq_region + primary_industry. Outcome variable is the reported “low” (floor) valuation of the latest funding round, in euros. The regression uses natural log of valuation_low_eur + 1. Predictor variables are all available variables, which can be seen in the table 5 in appendix 4.

Table 3 shows that the coefficient on months_between is essentially zero (–0.0002, $p \approx 0.85$), indicating that once you control for patent count and industry, the pure timing effect washes out. Number_of_patents stays highly significant (≈0.0059, $p<0.001$), and North America shows a 0.71 log-point premium.

After patents and region, the firm’s sector has the highest impact on expected valuation. Several industries like space, security and energy carry significant positive valuation shifts. Space has the highest coefficient of +2.2826 (SE = 0.3698, $t \approx 6.173$, $p \sim 7e-10$). Belonging to capital-heavy industry like space with high barriers for entry can dramatically shift valuation expectation. The overall $R^2$ rises to 7.5%, reflecting the added power of industry controls. With 33 predictors, the model explains 7.5% of valuation variation: modest but improved over simpler specifications.

### 4.2.4 Economic cycles and outcomes

**GDP Growth Model Results**

The impact of economic cycles is included in the GDP growth model shown in table 4. The logistic regression models the probability of a “successful” exit as a flexible function of the number of months between launch and first VC funding plus GDP growth rate, using a 3-degree natural spline. Table 4 below shows the estimated spline-basis coefficients.

***Table 4. Logistic Regression Including GDP Growth***

| Term | Estimate | Std_Error | Z_value | P_value |
|---|---|---|---|---|
| (Intercept) | -11 | 3.298 | -3.335 | 0.001 |
| ns(months_between, df = 3)1 | 2.598 | 1.657 | 1.568 | 0.117 |
| ns(months_between, df = 3)2 | 16.622 | 6.478 | 2.566 | 0.01 |
| ns(months_between, df = 3)3 | 4.486 | 2.072 | 2.165 | 0.03 |
| gdp_growth | 0.001 | 0.033 | 0.02 | 0.984 |
| Null deviance | 3096.8 | | | |
| Residual deviance | 3059.8 | | | |
| AIC | 3069.8 | | | |
| DF (null) | 6225 | | | |
| DF (residual) | 6221 | | | |

Note: Note: Fitted logistic-regression model object estimating “success” as a non-linear spline function of months_between (3 degrees of freedom). Outcome variable success is defined as a binary indicator (0/1): (IPO or acquisition > €500M, unicorn, or raised > €100M). Predictor variable number of months between the firm’s launch date and its first institutional funding round. Computed from “launch_date” and “first_funding_date”. The other predictor variable is annual global GDP growth rate (percent) for the funding year. (“NY.GDP.MKTP.KD.ZG”), 2000–2020.

Table 4 shows that the results are similar to the earlier logistic regression model without the growth term shown in table 1. The GDP-growth has essentially no effect (Estimate = 0.001; z = 0.02; p = 0.984). The spline terms retain their significance. This confirms that the timing effect is robust to broad cyclical conditions. The results suggest that when controlling for the nonlinear timing effect, the annual global growth rate that year does not help predict which deep-tech startups achieve successful exits in the future.

To summarize the results, timing matters for predicting binary success in logistic-only model. The effect is strong especially in the second and third spline intervals. But once the success is modeled with quantifiable valuation, patents and categorical region, timing's direct effect essentially disappears. The GDP growth model suggests that the economic cycles effect together with timing have no impact in predicting outcomes for deep tech startups. Patents and HQ region (U.S) emerge as the most robust predictors across all models. Each patent added raises valuation, and being in U.S. has a significant premium on both probability of success and valuation. After patents and region, certain industry verticals have higher valuation premiums.

# 5 Discussion and Conclusion

This chapter interprets the findings from the empirical study in Chapter 4 in relation to the research questions outlined in Chapter 1. The following chapters go over the implications for theory and practice and finally reflects on the study's limitations while proposing directions for future research.

## 5.1 Key Findings and Interpretation

The first research question asked whether the timing of initial funding rounds affects the chances of deep-tech startups reaching a successful exit. Success in this study means an IPO, a merger or an acquisition worth over €500 million, achieving unicorn status, or raising more than €100 million in funding. The Logistic spline model (Table 1, Chapter 4) shows a clear non-linear pattern: startups that receive their first institutional cheque a few months after launch, but not immediately or too late, have the highest chance of success. As shown in Figure 7, the probability of success increases sharply when funding happens around three to eight months after launch, then levels off. This "sweet spot" supports signaling theory (Hallen and Eisenhardt, 2012), which suggests that investors react most positively when funding is timed with visible and credible milestones, also known as proofpoints.

Building on signaling theory's focus on credibility over time, the hump-shaped pattern in Figure 7 suggests that raising funding too early may happen before the startup has shown enough product or market progress, leaving it underprepared. Raising funding too late, on the other hand, can signal that the startup has used up resources ineffectively without hitting key milestones. The top of the hump shows the highest success probability in the middle range in Figure 7. The top of the hump likely reflects a timing window where the startup has shown clear technical or commercial progress, but still has momentum and resources to continue moving forward.

The second research question asked whether macroeconomic cycles affect how well funding timing works. The answer to the question was explored through linear spline regression. The model was updated to include annual GDP growth (Table 4, Chapter 4). The results showed an odds ratio close to one ($p = 0.98$), which means that the overall economic growth or decline in the year of funding does not significantly change the link between timing and success. The results suggests that the timing of funding in relation to the company's own milestones matters more than the state of the global economy. This results regarding the effect of market movements differs from earlier results in more cyclical sectors like consumer and software startups (Kaplan and Schoar, 2005).

The third research question asked whether funding timing affects company valuation after controlling all other variables. The simpler model with just patents and region (Table 2a) showed that each extra patent increased valuation by about 0.6 percent. Also, if the startup was based in North America it had nearly 90 percent higher valuations. When the funding-timing spline was added, the model's $R^2$ rose slightly from 3.95 to 4.13 percent, and the ANOVA F-test was significant ($p = 0.008$, Table 2c), showing that timing still added some explanatory power beyond IP and location. The full multiple regression model (Table 3) showed that the effect of funding delay was no longer statistically significant ($p = 0.85$). The multiple regression model with all possible variables utilized suggests that timing no longer has a clear impact on valuation. Instead, variables like strong technology and being in well-funded regions weigh a lot more in affecting investor expectations for deep-tech ventures.

The timing of the first funding round has a strong non-linear effect on whether a deep-tech startup achieves a successful outcome. The impact on valuation is mostly shaped by other factors than timing, such as how strong the startup's IP is, where it is located, and what industry it operates in.

## 5.2 Contributions and Implications

### 5.2.1 Theoretical implications

This study builds on Spence's classic signaling theory (Spence, 1973; Hallen and Eisenhardt, 2012) by showing how timing can act as a dynamic signal in deep-tech startups. Previous studies in timing have focused mainly on the effects of patent counts or founder background. This study suggest that investors also read into the time between a startup's launch and its first equity funding. The timing sends a signal of its own. The findings from this study add to the existing theory by showing that there is a "sweet spot" when the first equity investment should happen. In contrast, raising money too early or too late can reduce the chances of success.

The disappearance of timing's effect in the multiple regression model helps clarify how capital-intensive innovation is valued. The results are in line with Nanda and Rhodes-Kropf (2017). Nanda and Rhodes-Kropf point in their study that for startups with high real option value, IP and thriving ecosystem are more important for valuation than the exact timing of funding rounds. In this cases patents and operating in U.S. The timing matters mainly because it signals that key milestones have been reached, not because it directly increases long-term value on its own.

### 5.2.2 Practical implications

Founding teams should aim to raise their first equity funding about three to eight months after launching the product. However, they should only raise once there is clear technical progress or early customer interest. Raising too early can lead to unnecessary equity dilution. Waiting too long can weaken the position in negotiations.

A funding round raised too early may point to unproven technology. Even if the whole thesis for the investment is built around “hype”, it should still prompt closer technical review. A round raised too late may signal cash-flow problems or missed milestones. Evaluating why the team is raising funding at the chosen timeline can help investors make better investment decisions.

Finally, policymakers and grant agencies aiming to close the “valley of death” in deep-tech innovation should design non-dilutive funding to help startups reach key early milestones without needing to raise equity too early. By linking grant funding to clear technical progress, public programs can build investor confidence and support better-timed equity rounds.

## 5.3 Limitations and Future Research Directions

### 5.3.1 Limitations

This study uses unique proprietary data of 6,226 deep-tech startups. The data consists of startups founded between 2010 and 2020. The data is from Dealroom. The initial export from Dealroom resulted to 23,000 original observations exported, more than half had to be dropped due missing data. Some key variables like first funding date or valuation were missing. The incomplete observations had to be dropped from the analysis. The results may overrepresent startups that report more transparently or startups that made it to later funding stages or exits.

Measurement validity of key variables is limited. The binary “success” variable used in the study may not fully capture the quality of the startup in the long run. Same applies to other outcome “valuation”. The cross-sectional nature of this data doesn’t tell the full story of what may happen in the future. Internal validity point of view there are unobserved factors that may impact the results and drive valuation and success. These are factors like quality of the founding team, prior experience of the founding team, or quality of the investor who invested. External validity point of view there is of course the survivorship bias.

The proprietary data sets limits to reliability of the data. The data is provided by Dealroom, and inconsistencies from their side may affect precision. Also, the manual data preparation and the data being behind a paywall sets limits to the reliability from the consistency and reproducibility point of view. Further limits to reliability are the small subgroups in some spline models. Due low number of observations in extreme ends of the time variable, some parameter estimates rely fewer observation. This may limit the statistical reliability.

### 5.3.2 Future research directions

This study analyzed the effect of timing of the first funding round to deep-tech startup's success. The results revealed more unanswered questions for future research. The background and quality of the founding team is one of the most important proxies for success. Another interesting future research questions is how milestones line up with funding rounds.

Past startup exits, degrees from top universities, or experience at major tech companies are clear signals of quality and known to affect VC decisions. It would be interesting to know whether highly experienced founders raise money earlier on better terms, or do less experienced teams benefit from waiting longer?

Deep-tech startups often move through specific milestones. It would be interesting to know whether specific timing of funding rounds lined up with milestones would have an effect on the outcome.

Other promising directions for future research include sector-specific timing strategies and the optimal sequencing between funding rounds.

Sector-specific timing could reveal valuable insights into vastly different subsections of deep tech venture. While this study all deep tech like AI, biotech, quantum, and cleantech startups into one model, each of these have their unique challenges and needs for capital.

The timing between later rounds is just as important as the timing between launch and first funding. Using methods like survival analysis or hazard-rate models to study when follow-on rounds happen could help identify the key milestones and burn-rate limits that lead to successful next-stage funding.

## References

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. *The Journal of Finance*, 71(4), 1591-1622.

Black, B. S., & Gilson, R. J. (1998). Venture capital and the structure of capital markets: banks versus stock markets. *Journal of Financial Economics*, 47(3), 243-277.

Cockburn, I. M., & MacGarvie, M. J. (2011). Entry and patenting in the software industry. *Management Science,* 57(5), 915-933.

Colombo, M. G., & Shafi, K. (2016). The impact of patenting on the size of high-tech firms: the role of venture capital and product market regulation. *Economia e Politica Industriale*, 43(1), 85-103.

Dealroom, Lakestar, Walden Catalyst, & Hello Tomorrow. (2025). *The 2025 European Deep Tech Report. Dealroom*. https://dealroom.co/reports/the-european-deep-tech-report-2025

Ewens, M., & Farre-Mensa, J. (2020). The deregulation of the private equity markets and the decline in IPOs. *The Review of Financial Studies*, 33(12), 5463-5509.

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

Gompers, P. A. (2022). Optimal investment, monitoring, and the staging of venture capital. *In Venture capital* (pp. 285-313). Routledge.

Gompers, P., & Lerner, J. (2001). The venture capital revolution. *Journal of Economic Perspectives*, 15(2), 145-168.

Gornall, W., & Strebulaev, I. A. (2020). Squaring venture capital valuations with reality. *Journal of Financial Economics*, 135(1), 120-143.

Hallen, B. L., & Eisenhardt, K. M. (2012). Catalyzing strategies and efficient tie formation: How entrepreneurial firms obtain investment ties. *Academy of Management Journal*, 55(1), 35-70.

Hellmann, T., & Puri, M. (2000). The interaction between product market and financing strategy: The role of venture capital. *The Review of Financial Studies*, 13(4), 959-984.

Hochberg, Y. V., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, 62(1), 251-301.

Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy*, 36(5), 722-741.

Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance*, 60(4), 1791-1823.

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

Lerner, J. (1994). The syndication of venture capital investments. In *Venture capital* (pp. 207-218). Routledge.

Lerner, J. (1996). The government as venture capitalist: The long-run effects of the SBIR program.

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

Pahnke, E. C., Sirmon, D. G., Rhymer, J., & Campbell, J. T. (2023). Resource interdependence and successful exit: A configurational perspective on young technology firms. *Strategic Entrepreneurship Journal*, 17(3), 507-534.

Pinkwart, A., Proksch, D., Schefczyk, M., Fiegler, T., & Ernst, C. (2015). Reasons for the failure of new technology-based firms: A longitudinal empirical study for Germany. *Credit and Capital Markets–Kredit und Kapital*, (4), 597-627.

Puri, M., & Zarutskie, R. (2012). On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms. *The Journal of Finance*, 67(6), 2247-2293.

Romme, A. G. L., Bell, J., & Frericks, G. (2023). Designing a deep-tech venture builder to address grand challenges and overcome the valley of death. *Journal of Organization Design*, 12(4), 217-237.

Sahlman, W. A. (2022). The structure and governance of venture-capital organizations. *In Venture capital* (pp. 3-51). Routledge.

Shane, S., & Stuart, T. (2002). Organizational endowments and the performance of university start-ups. *Management Science*, 48(1), 154-170.

Spence, M. (1978). Job market signaling. In Uncertainty in economics (pp. 281-306). *Academic Press*.

# Appendix 1 Data Management Plan

**Data Collection**

- Data Sources:
    - The proprietary data used in this study was exported from Dealroom on April 18, 2025. Dealroom is a closed platform requiring a paid subscription to access. This subscription was provided through my previous employer. Dealroom imposes a daily data export limit of 13,000 rows per month, but for this study, the limit was extended to accommodate an initial 23,000 rows.
    - GDP growth data was collected from the World Bank API via the *WDI* package in R, covering the period from 2000 to 2020.
- Data Collection Methods:
    - Startup data: Exported directly from Dealroom using an authorized account.
    - GDP growth data: Downloaded through the World Bank API (WDI package in R), which provides annual GDP growth figures for all countries globally.

**Data Storage**

- Storage Location:
    - The data can be re-exported from Dealroom, though continued access requires maintaining the paid subscription.
- Data Security:
    - The data is not considered sensitive and is available to any user with a paid Dealroom subscription. No special encryption or security measures are required for handling this dataset.

**Data Sharing**

- Access and Sharing:
    - The data will not be shared publicly. However, it can be re-exported from Dealroom to replicate the empirical analysis.

**Data Documentation and Metadata**

- Data Description:
  - The dataset includes various columns that provide key information on the startups, such as:
    - Success, Unicorn, IPO>500, M&A>500, Funding>100: Outcome indicators and financial milestones.
    - Launch Year, First Funding Date, Seed Year: Time-related variables.
    - Primary Industry, Sub industries: Startup industry classifications.
    - Total Funding (EUR M), Valuation Low (EUR), Valuation High (EUR): Funding and valuation data.
    - Founders, Investors Names, Lead Investors: Founders and investor data.
- Metadata Standards:
  - Each column heading is listed, along with their respective descriptions, such as how 'Primary Industry' refers to the main sector each startup operates within. These variables will be cross-referenced with their column names as defined in the code for clarity.

**Data Analysis**

- Tools Used:
  - Data analysis was done using RStudio. The primary analysis will use statistical models, including regression analysis to explore the relationship between startup funding timing and success outcomes. Relevant R packages include tidyverse, lubridate, janitor, splines, broom, and flextable.
- Analysis Methodology:
  - The regression models and visualizations used will include:
    - Logistic Regression: To model the binary success of startups based on funding timing (months_between).

    - Multivariate Linear Regression: To analyze the log-transformed valuation with predictors such as funding timing, number of patents, and industry.
    - ANOVA: To compare different regression models, including the impact of patents and funding timing on startup valuation.
- Charts and visualizations used in the analysis include geographical distributions, industry distributions, and success rates based on funding timing.

## Appendix 2 R Code

This appendix covers the R code used for the statistical analysis. The study can be replicated by 1) downloading the data set, 2) changing the file path to a new working directory and 3) running the script in RStudio. The script does not include code for building the tables and exports.

```
# --------------------------

# 0. Setup & Data Preparation

# --------------------------

# Install and load packages with verification

required_packages <- c("tidyverse", "lubridate", "janitor", "broom", "splines", "flextable", "officer")

install_missing <- function(packages) {

  new_packages <- packages[!(packages %in% installed.packages()[,"Package"])]

  if(length(new_packages)) install.packages(new_packages)

}

install_missing(required_packages)

# Load packages

library(tidyverse)

library(lubridate)

library(janitor)

library(broom)

library(splines)

library(flextable)

library(officer)

# Explicitly import pipe operator
```

```
`%>%` <- dplyr::`%>%`

# Set working directory and load data

setwd("/file/path/to/data")

data_raw <- read_csv("data.csv") %>%

  clean_names() %>%

  mutate(

    number_of_patents = replace_na(number_of_patents, 0),

    across(c(primary_client_focus, primary_revenue_model, primary_industry),

           ~replace_na(., "Unknown") %>% as.factor()),

    log_valuation = log(valuation_low_eur + 1),

    months_between = ifelse(is.na(months_between), median(months_between, na.rm = TRUE),
months_between)

  ) %>%

  filter(!is.na(valuation_low_eur))

# --------------------------

# 1. Exploratory Data Analysis

# --------------------------

# Custom theme for consistency

theme_deep_tech <- function() {

  theme_minimal() +

    theme(

      plot.title   = element_text(size = 14, face = "bold", hjust = 0.5),

      axis.title   = element_text(size = 12),
```

```
    axis.text    = element_text(size = 10),
    legend.position = "bottom"
  )
}
# 1.1 Geographical Distribution
# Regions (Top 15)
data_raw %>%
  count(hq_region) %>%
  mutate(pct = n/sum(n)) %>%
  ggplot(aes(x = "", y = pct, fill = hq_region)) +
  geom_col(width = 1) +
  geom_text(aes(label = scales::percent(pct)),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  labs(title = "Geographical Distribution of Startups",
       fill = "HQ Region") +
  theme_deep_tech() +
  theme(axis.text = element_blank())
# Countries (Top 15)
data_raw %>%
  count(hq_country, sort = TRUE) %>%
  head(15) %>%
  ggplot(aes(x = fct_reorder(hq_country, n), y = n)) +
```

```
  geom_col(fill = "#2c7fb8", alpha = 0.9) +
  coord_flip() +
  labs(title = "Top 15 Headquarters Countries",
       x = "Country",
       y = "Number of Startups") +
  theme_deep_tech() +
  theme(panel.grid.major.y = element_blank())
# 1.2 Launch Year Distribution
data_raw %>%
  filter(launch_year >= 2010) %>%  # Focus on modern era
  count(launch_year) %>%
  ggplot(aes(x = factor(launch_year), y = n)) +
  geom_col(fill = "#31a354", width = 0.8) +
  labs(title = "Startup Launch Years Distribution",
       x = "Launch Year",
       y = "Number of Startups Founded") +
  theme_deep_tech() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# 1.3 Industry Distribution
data_raw %>%
  count(primary_industry) %>%
  top_n(10, n) %>%
  ggplot(aes(x = fct_reorder(primary_industry, n), y = n)) +
```

```
  geom_col(fill = "#2c7fb8") +
  coord_flip() +
  labs(title = "Top 10 Industries in Deep Tech Startups",
       x = "Industry",
       y = "Number of Startups") +
  theme_deep_tech()
# 1.4 Client Focus Analysis
# Primary Client Focus
data_raw %>%
  mutate(primary_client_focus = fct_lump(primary_client_focus, n = 6)) %>%
  count(primary_client_focus) %>%
  ggplot(aes(x = fct_reorder(primary_client_focus, n), y = n)) +
  geom_col(fill = "#756bb1", alpha = 0.9) +
  coord_flip() +
  labs(title = "Primary Client Focus Distribution",
       x = "Client Focus Category",
       y = "Number of Startups") +
  theme_deep_tech() +
  theme(panel.grid.major.y = element_blank())
# --------------------------
# 2. Temporal Analysis
# --------------------------
# 2.1 Funding Timing Analysis
```

```
data_raw %>%
  mutate(funding_timing = case_when(
    months_between < -6 ~ ">6mo Pre-Launch",
    months_between < 0 ~ "<6mo Pre-Launch",
    months_between <= 6 ~ "<6mo Post-Launch",
    TRUE ~ ">6mo Post-Launch"
  )) %>%
  ggplot(aes(x = funding_timing, fill = factor(success))) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Success Rate by Funding Timing",
       x = "Funding Timing Relative to Launch",
       y = "Success Rate",
       fill = "Success") +
  theme_deep_tech()
# --------------------------
# 3. Regression Analysis
# --------------------------
# 3.1 Logistic Regression - Funding Timing
model_logistic <- glm(
  success ~ ns(months_between, df = 3),
  data   = data_raw,
  family = binomial
```

```r
)
# 3.2 Patent Impact Analysis
#   a) Simple patents + region
model_patent_simple <- lm(
  log_valuation ~ number_of_patents + hq_region,
  data = data_raw
)
#   b) Spline-augmented timing + patents + region
model_patent_spline <- lm(
  log_valuation ~ ns(months_between, df = 3) + number_of_patents + hq_region,
  data = data_raw
)
#   c) ANOVA comparison
anova_patent <- anova(model_patent_simple, model_patent_spline)
# 3.3 Multivariate Model
model_full <- lm(
  log_valuation ~ months_between + number_of_patents + hq_region + primary_industry,
  data = data_raw
)
# --------------------------
# 4. Model Visualization
# --------------------------
```

```
# 4.1 Logistic Regression Curve

ggplot(data_raw, aes(x = months_between, y = success)) +

  geom_jitter(height = 0.05, alpha = 0.2, width = 0.3) +

  geom_smooth(method = "glm", formula = y ~ ns(x, 3),

              method.args = list(family = "binomial"), color = "#e41a1c") +

  labs(title = "Success Probability vs Funding Timing",

       x = "Months Between Launch and First Funding",

       y = "Success Probability") +

  theme_deep_tech()

# 4.2 Patent Impact Visualizations

# 4.2a Simple Patent–Region Fit

grid_simple <- tibble(

  number_of_patents = seq(0, max(data_raw$number_of_patents, na.rm=TRUE), length.out = 200),

  hq_region         = "Europe"

) %>%

  mutate(pred_log_val = predict(model_patent_simple, newdata = .))

ggplot(data_raw, aes(x = number_of_patents, y = log_valuation)) +

  geom_point(alpha = 0.3) +

  geom_line(

    data      = grid_simple,

    aes(x     = number_of_patents, y = pred_log_val),

    color     = "#2c7fb8",

    linewidth = 1          # ← use linewidth
```

```
  ) +
  labs(
    title = "Simple Patent–Valuation Relationship",
    x     = "Number of Patents",
    y     = "Log Valuation"
  ) +
  theme_deep_tech()
# 4.2b Spline-Augmented Timing Effect
grid_spline <- tibble(
  months_between   = seq(min(data_raw$months_between), max(data_raw$months_between),
length.out = 200),
  number_of_patents = median(data_raw$number_of_patents, na.rm=TRUE),
  hq_region         = "Europe"
) %>%
  mutate(pred_log_val = predict(model_patent_spline, newdata = .))
ggplot(grid_spline, aes(x = months_between, y = pred_log_val)) +
  geom_line(color = "#e41a1c", linewidth = 1) +        # ← linewidth instead of size
  geom_rug(
    data        = data_raw,
    aes(x       = months_between),
    sides       = "b",
    alpha       = 0.1,
    inherit.aes = FALSE   # ← drop the y-mapping for the rug layer
```

```
  ) +
  labs(
    title = "Non-Linear Effect of Funding Delay on Log Valuation",
    x     = "Months Between Launch & First Funding",
    y     = "Predicted Log Valuation"
  ) +
  theme_deep_tech()
# 4.2c R² Comparison Bar Chart
r2_simple <- summary(model_patent_simple)$r.squared
r2_spline <- summary(model_patent_spline)$r.squared
r2_tbl <- tibble(
  model = c("Simple", "Spline"),
  R2    = c(r2_simple, r2_spline)
)
ggplot(r2_tbl, aes(x = model, y = R2, fill = model)) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = scales::percent(R2, accuracy = 0.1)),
            vjust = -0.5) +
  labs(title = "Model Fit: Simple vs Spline Patent–Region Specifications",
       x     = NULL,
       y     = expression(R^2)) +
  theme_deep_tech() +
  ylim(0, max(r2_tbl$R2) * 1.1)
```

```r
# --------------------------
# 5. Model Diagnostics
# --------------------------
library(broom)
model_diagnostics <- function(model, title) {
  # turn the model into a data frame of fitted / residuals
  df <- augment(model)
  # Residuals vs Fitted
  p1 <- ggplot(df, aes(x = .fitted, y = .resid)) +
    geom_point(alpha = 0.5) +
    geom_hline(yintercept = 0, linetype = "dashed") +
    labs(title = paste("Residuals vs Fitted —", title),
         x = "Fitted Values", y = "Residuals")
  # QQ-plot of standardized residuals
  p2 <- ggplot(df, aes(sample = .std.resid)) +
    geom_qq() +
    geom_qq_line() +
    labs(title = paste("Normal Q–Q —", title),
         x = "Theoretical Quantiles", y = "Standardized Residuals")
  # side-by-side
  gridExtra::grid.arrange(p1, p2, ncol = 2)
}
```

```r
# Run diagnostics

model_diagnostics(model_patent_simple, "Patent Model (Simple)")

model_diagnostics(model_patent_spline, "Patent Model (Spline)")

model_diagnostics(model_full,          "Full Valuation Model")

# --------------------------

# 6. Model Summaries

# --------------------------

# Logistic regression

summary(model_logistic)

# Patent models

summary(model_patent_simple)

summary(model_patent_spline)

# ANOVA

anova_patent

# Full model

summary(model_full)

# --------------------------

#ECONOMIC CYCLES

# --------------------------

#1. Install & load the WDI package

install.packages("WDI")     # World Bank API client

install.packages("lubridate")# date parsing

install.packages("splines")  # if not already installed
```

```r
library(WDI)

library(lubridate)

library(splines)

library(dplyr)

#2. Download annual GDP growth for 2000–2020

gdp_raw <- WDI(

  country   = "WLD",

  indicator = "NY.GDP.MKTP.KD.ZG",

  start     = 2000,

  end       = 2020

)

# Rename for convenience:

gdp_raw <- rename(gdp_raw,

                  funding_year = year,

                  gdp_growth   = NY.GDP.MKTP.KD.ZG

)

gdp_raw

#3. Prepare startup data

data_raw <- data_raw %>%

  mutate(

    first_funding_date = ymd(first_funding_date),        # parse strings as Dates

    funding_year       = year(first_funding_date)        # extract the year

  ) %>%
```

```
  left_join(
    gdp_raw %>% select(funding_year, gdp_growth),       # bring in GDP growth
    by = "funding_year"
  )
# Check:
data_raw %>% select(first_funding_date, funding_year, gdp_growth) %>% head()
#4. Re-run logistic regression adding GDP growth
model_with_gdp <- glm(
  success ~ ns(months_between, df = 3) + gdp_growth,
  data   = data_raw,
  family = binomial
)
summary(model_with_gdp)
```

## Appendix 3 Diagnostics Tables

Below are the diagnostics for linear regression and multiple regression used in the analysis. Residual plots and Q-Q plots show no significant issue with homoscedasticity or normality for the models used in the study.

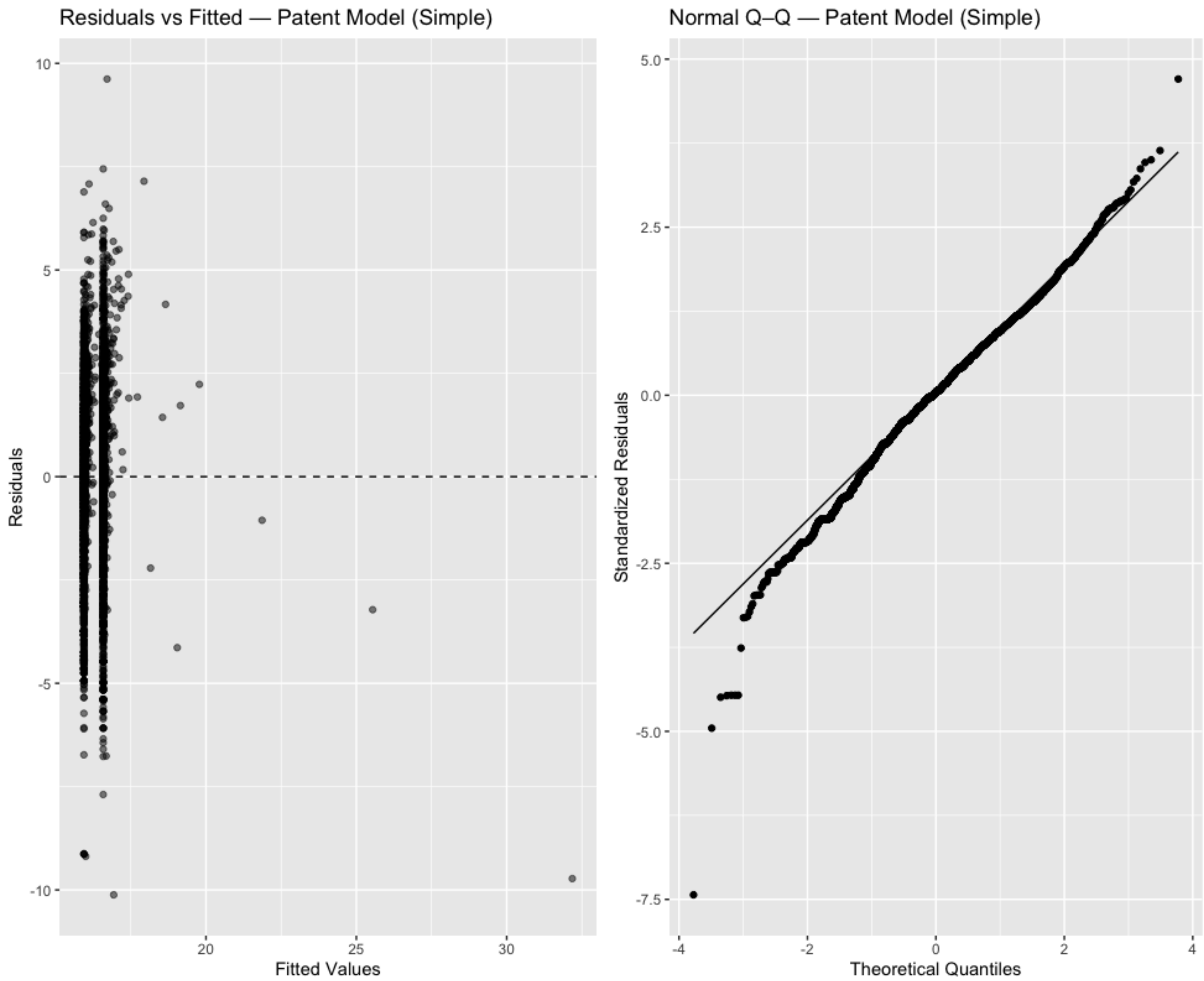


Figure 9a. Diagnostics. Residuals vs Fitted and Normal Q–Q for the Simple Patent Model.

The residual plots and Q-Q plots (Figure 9a) display a roughly even scatter around zero and adherence to the 45° line, suggesting no major departures from homoskedasticity or normality.

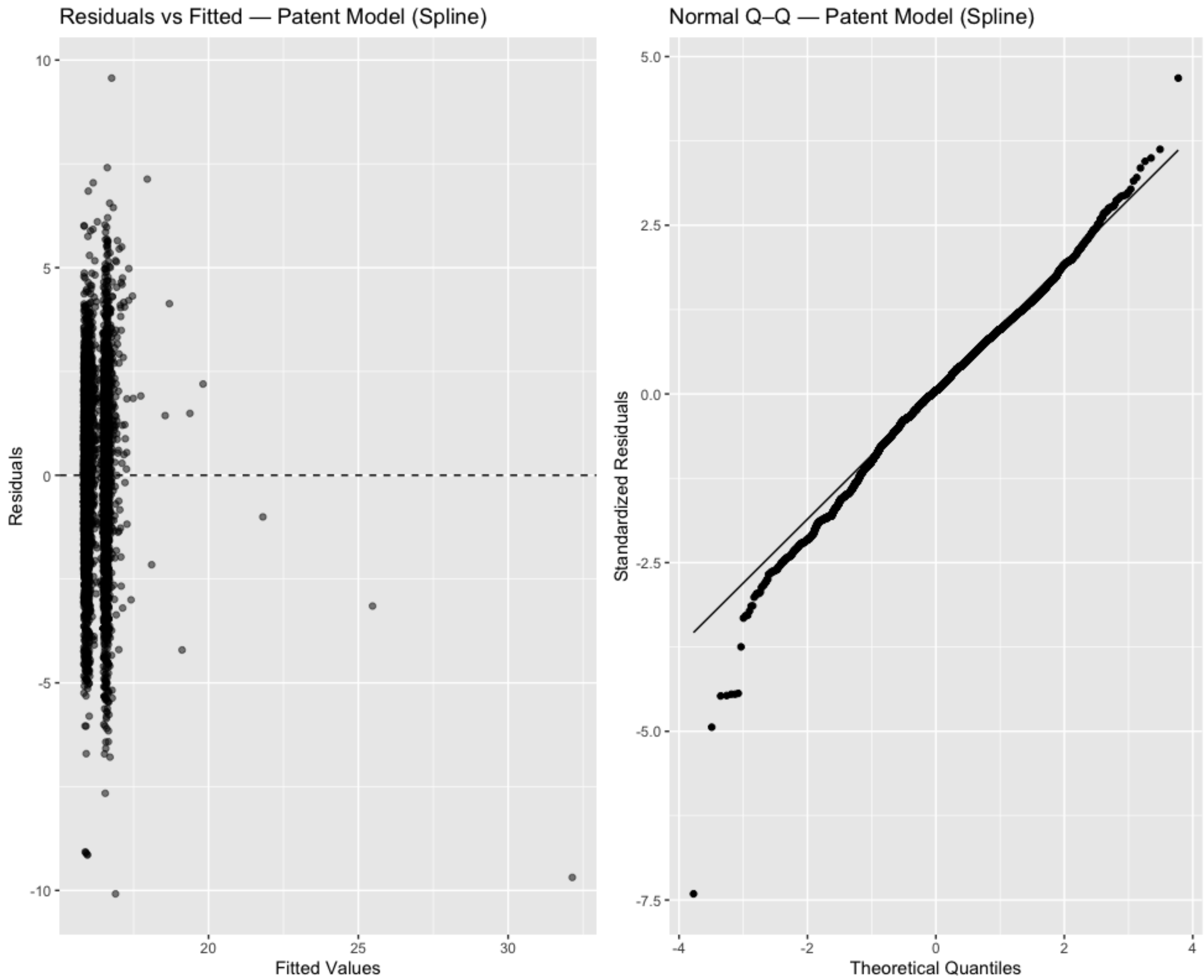


Figure 9b. Diagnostics. Residuals vs Fitted and Normal Q–Q for the Spline Patent Model.

Diagnostics for this spline model (Figure 9b) remain well-behaved.

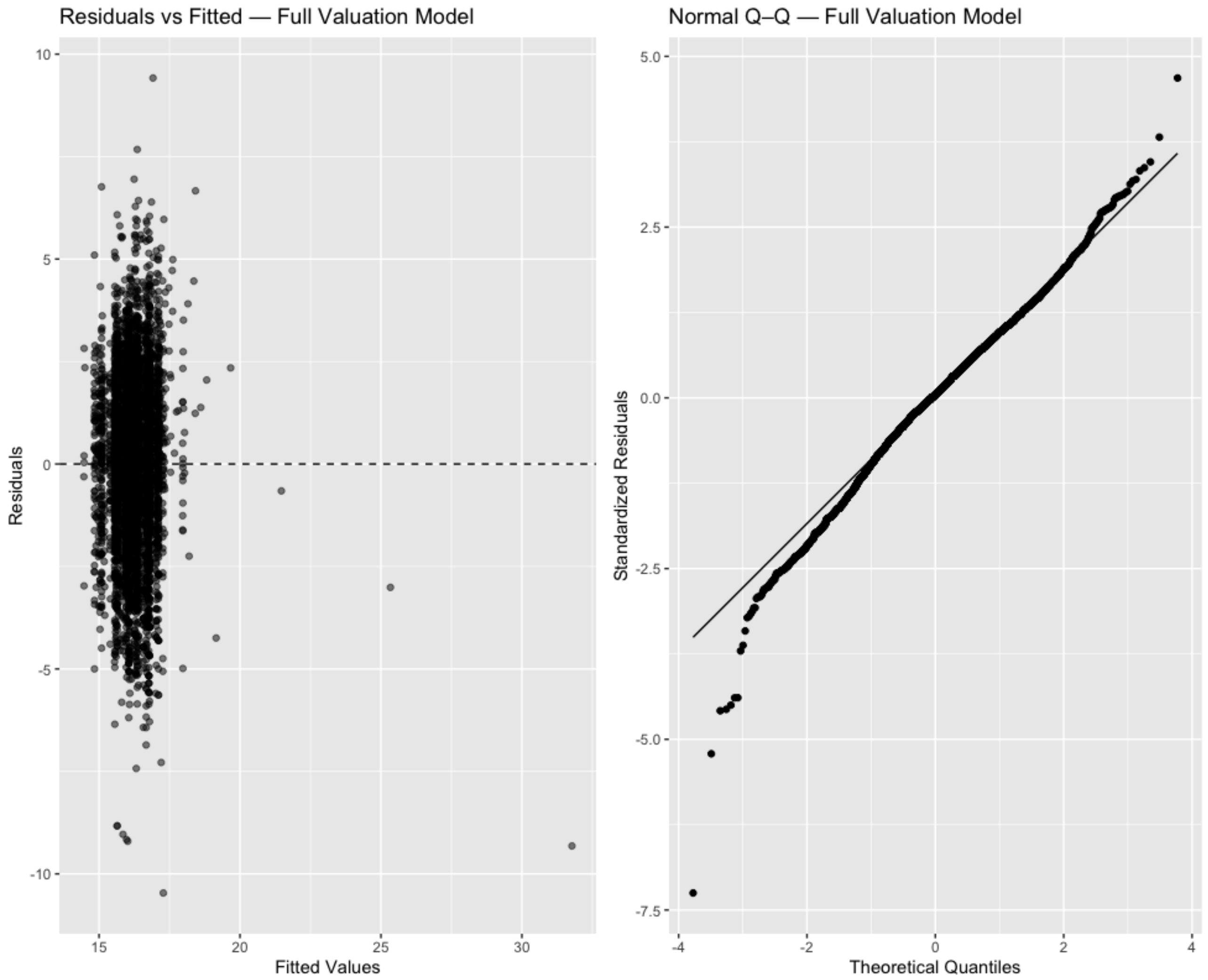


Figure 9c. Diagnostics. Residuals vs Fitted and Normal Q–Q for the Multiple Regression Model.

Figure 9c presents the residual plots and Q-Q plots for the full multiple regression model with all the available predictor variables. Though variance remains substantial, no significant violations of linear assumptions appear. This can be expected in a cross-section of deep-tech unicorn potential.

## Appendix 4. Variable Definitions

***Table 5. Variable Definitions***

| Variable | Definition | Data source |
|---|---|---|
| success | Binary indicator (0/1) of “successful exit” (IPO, acquisition > €500 M, unicorn, or raised > €100 M). | Dealroom |
| months_between | Number of months between the firm’s launch date and its first institutional funding round. Computed from “launch_date” and “first_funding_date”. | Computed in Excel |
| first_funding_date | Date of the startup’s first equity funding round. | Dealroom |
| launch_year | Year in which the startup was founded/launched. | Dealroom |
| funding_year | Year extracted from first_funding_date (used to join with GDP growth). | Dealroom |
| gdp_growth | Annual global GDP growth rate (percent) for the funding year. (“NY.GDP.MKTP.KD.ZG”), 2000–2020 | World Bank WDI API |
| valuation_low_eur | Reported “low” valuation of the latest funding round in euros. | Dealroom |
| log_valuation | Natural log of valuation_low_eur + 1. | Computed in R |
| number_of_patents | Number of patents. | Dealroom |
| hq_region | Categorical variable for the startup’s headquarters region. | Dealroom |
| hq_country | Categorical variable for the startup’s headquarters country. | Dealroom |
| primary_industry | Categorical variable for the startup’s main industry. | Dealroom |
| primary_client_focus | Categorical variable for the startup’s main client foucs. | Dealroom |
| primary_revenue_model | Categorical label for the startup’s main revenue model. | Dealroom |
| model_logistic | Fitted logistic-regression model object estimating “success” as a non-linear spline function of months_between (3 degrees of freedom). | Computed in R |
| model_patent_simple | Fitted linear-regression model object predicting log_valuation from number_of_patents + hq_region. | Computed in R |
| model_patent_spline | Fitted linear-regression model object predicting log_valuation from a spline on months_between (df = 3) + number_of_patents + hq_region. | Computed in R |

| Variable | Definition | Data source |
| --- | --- | --- |
| model_full | Fitted linear-regression model object predicting log_valuation from months_between + number_of_patents + hq_region + primary_industry. | Computed in R |
| anova_patent | ANOVA comparison object testing whether the spline-augmented patent model (model_patent_spline) significantly improves fit over the simple patent model (model_patent_simple). | Computed in R |
| Residual Std. Error | Estimate of the residual (error) standard deviation from a linear model. | Computed in R |
| Multiple $R^2$ | Proportion of variance in log_valuation explained by the regression model. | Computed in R |
| Adjusted $R^2$ | $R^2$ adjusted for number of predictors (penalizes adding irrelevant variables). | Computed in R |
| F-statistic | F-statistic testing overall significance of the regression’s explanatory variables. | Computed in R |
| DF (model) | Model degrees of freedom (number of estimated parameters – 1). | Computed in R |
| DF (residual) | Residual degrees of freedom (n – p, where p = number of parameters estimated, including intercept). | Computed in R |
| p-value (F) | p-value for the overall F-test. | Computed in R |

## Appendix 4 Use of AI tools

In the process of creating and writing this study, AI tools have been used to assist with searching and filtering the most relevant previous studies, choosing and getting familiar with the references and debugging the R code used in the empirical analysis.

1. Searching and filtering potential references
    i. ChatGPT & DeepSeek
        1. Deep research and DeepThink features from both LLM models were used to search academic literature to support the thesis. The threshold for literature to be considered as a reference was double digits in The Journal Impact Factor (JIF) or Q1 in Quartile score. The search on both LLM models performed poorly compared to manual search in Google Scholar, but some references were found and ended up being used as a result of this approach.
2. Choosing and getting familiar with the references
    i. Notebook LM
        1. Each reference was uploaded to Notebook LM and turned into AI podcast. References were turned into podcast per each reference and also per group of references + other sources. For example, three latest deep tech reports from Dealroom were turned into one podcast. Only the latest one was used as an official reference on this thesis.
    ii. DeepSeek
        1. DeepSeek was used to summarise each reference to create a quick overview before reading the full paper and manually selecting what to citate in the thesis.
3. Debugging R code
    i. ChatGPT
        1. When the R code for the empirical study ran into errors, ChatGPT was used to debug the code.