Working Paper 24-040

Contributing to Growth? The Role of Open Source Software for Global Startups

Nataliya Langburd Wright
Frank Nagle
Shane Greenstein

Harvard Business School

Contributing to Growth? The Role of Open Source Software for Global Startups

Nataliya Langburd Wright
Columbia Business School

Frank Nagle
Harvard Business School

Shane Greenstein
Harvard Business School

Working Paper 24-040

We thank Peter Cihon, Mike Linksvayer, and Kevin Xu at GitHub for help accessing GitHub's metadata to construct the key exogenous variable in this study and Noa Rosinplotz for excellent research assistance. We also thank seminar participants and discussants for comments and suggestions in the HBS Digital, Data, & Design Institute Research Day, OpenForum Academy Symposium, and SMS Open Source Extension. All errors remain our own.

Funding for this research was provided in part by Harvard Business School.

Contributing to growth? The role of open source software for global startups[1]

January 2024

Nataliya Langburd Wright [a], Frank Nagle [b], Shane Greenstein [c]

[a]Columbia Business School, New York, NY 10027, United States of America

[b]Harvard Business School, Boston, MA 02163, United States of America

[c]Harvard Business School, Boston, MA 02163, United States of America

ABSTRACT

Does participating in open source software (OSS) communities spur entrepreneurial growth? More efficiently developing shared code, learning from what the OSS community has developed, and shaping the direction of massive projects, such as those linked to frameworks for AI algorithms, attract many participants. Yet, contributing valuable resources to OSS, such as time and code, might give away too much, making it more difficult for firms to appropriate value from innovations. To gain a deeper understanding of participation, we analyze novel data matching accounts from GitHub—the largest OSS hosting platform— to the universe of global software venture-backed firms identified by PitchBook. We find a robustly positive relationship between OSS contributions and entrepreneurial growth, driven by both selection and treatment effects. The treatment effects account for roughly one-third of the overall impact: firms that increase GitHub contributions—in terms of the number of lines of code and number of users contributing—see an increase in their valuation and funding. Human capital, OSS policies, and market size moderate the statistical relationship between contributing to OSS and valuations, suggesting that OSS complements supply-side and demand-side country endowments. This research reveals that contributing to OSS can lead to entrepreneurial growth worldwide, with implications for policy and entrepreneurial strategy.

Keywords: Entrepreneurship, open source software, GitHub, human capital, global entrepreneurship, scaling, entrepreneurial strategy

[1]nlw2136@gsb.columbia.edu, fnagle@hbs.edu, sgreenstein@hbs.edu. We thank Peter Cihon, Mike Linksvayer, and Kevin Xu at GitHub for help accessing GitHub's metadata to construct the key exogenous variable in this study and Noa Rosinplotz for excellent research assistance. We also thank seminar participants and discussants for comments and suggestions in the HBS Digital, Data, & Design Institute Research Day, OpenForum Academy Symposium, and SMS Open Source Extension. All errors remain our own.

1. Introduction

Open source software (OSS) is code publicly available for use and modification. For much of its history, OSS was the purview of tinkerers and advocates of free software attempting to replicate proprietary software functionality. It has evolved into essential software for many cutting-edge technologies, including artificial intelligence (AI), web-enabled commerce, and big data analytics. OSS has drastically reduced the costs for coders to gain access to routine and frontier code across a range of applications, and it enables this access without the hassles of negotiating over intellectual property (DiBona et al., 2005; Lakhani and Wolf, 2005). After 2015, for example, any AI coder could gain access to the frontier machine learning algorithms from the TensorFlow or PyTorch libraries, develop applications built upon the frontier features of these libraries, and collaborate with sophisticated developers worldwide (Rock, 2022). Moreover, participants who use OSS develop applications in a fraction of the time and cost it would have taken to do so less than a decade earlier. Even when developing code far within the frontier, OSS reduces the time to write software modules.

Research has shown that participating in OSS can help launch entrepreneurial ventures (Nagle, 2019a; Wen et al., 2015), that startups ramp up their involvement in OSS as they prepare to seek venture funding (Conti, Peukert, and Roche, 2021), and that increased open source activity across countries predicts more startups and venture foundings in those countries (Dushnitsky and Stroube, 2021; Nagle 2019a; Wright, Nagle, and Greenstein, 2023). This leaves a gap, however. What impact does OSS participation have on entrepreneurial firms after founding? Does participation in OSS communities continue to spur entrepreneurial growth after founding? If so, how much, and how does this differ across startups in different countries? Research does not point unambiguously to a conclusion.[2] Making additional contributions to OSS might or might not propel additional firm growth, and even if it does, no research has investigated whether startups in various circumstances worldwide receive a similar benefit from participating in OSS after their founding.

The question is interesting because it sheds light on whether OSS fosters growth, and if so, how and why it does so in some circumstances and perhaps not others. On the one hand, OSS's benefits could continue as firms grow larger. Participating in OSS may help entrepreneurs access and learn how to use additional fundamental code, lowering the costs of developing additional features for existing code and scaling their products (Conti, Peukert, and Roche, 2021; Lin and Maruping, 2022; Nagle, 2019b; von Hippel and von Krogh, 2003). It also could help attract talent and investors that the entrepreneur otherwise would not find locally. On the other hand, if entrepreneurial participants in open source communities merely gain access to free and unrestricted frontier code, most of the benefits could be exhausted soon after startups first gain initial access to OSS. Moreover, there could be drawbacks to continuing participation. Participation in open source communities might also make it more difficult for firms to appropriate value from their innovations (Fosfuri, Giarratana, and Luzzi, 2008; Nagle, 2018; Pisano, 2006; Wen, Ceccagnoli, and Forman, 2015), and existing research has shown firms in the digital space employ different appropriability strategies depending on their size (Miric, Boudreau, and Jeppesen, 2019).

It is also unclear which types of firms participate most frequently in OSS. If OSS reduces the friction associated with founding a firm, it may increase the number of firms exploring low-quality ideas. No amount of help from OSS will help low-quality firms grow after their initial founding.

[2] Lin and Maruping (2022) gets closest to such a conclusion. However, the data on OSS contributions used in that study is at the repository level (complete OSS packages), compared to the lower-level individual commit data we rely upon. Further, since that study is focused on US startups, it cannot consider cross-country influences on the impact of OSS on entrepreneurial growth.

In other words, failure after initial entry could be high in the world of low-friction OSS due to the expansion of those who self-select into experimenting with entrepreneurship for a short while.

It is challenging to disentangle these competing factors in statistical data due to the mix of treatment and selection effects that shape the entrepreneurial experience while participating in OSS. Those statistical considerations make for demanding data requirements for robust inference. Estimating a statistical relationship between OSS participation and its impact on entrepreneurial firms requires assembling data on the universe of OSS contributions matched at the entrepreneurial firm level and different stages of life for startups, and it requires measuring various levels of participation in OSS. Data collection challenges also must be overcome. Though GitHub—the most widely-used repository of OSS available today—has expanded rapidly in recent years, growing from 3 million users in 2013 to 96 million repositories and 31 million users in 2018, the composition of participation has never been documented.[3] Many firms lack an official GitHub account, and even when the focus sharpens to only startups, knowing which ventures to identify is a complex endeavor. While GitHub allows capturing individual-level user data online, matching these data to entrepreneurial firms is challenging.

This paper assembles and analyzes novel and proprietary GitHub data, and matches participants' OSS activity to their parent firms. This enables the measurement and analysis of the relationship between OSS contributions and entrepreneurial growth. The included firms represent the universe of software ventures on PitchBook, one of the most prominent and comprehensive startup databases. We begin with nearly 100k worldwide firms and measure global open source activity in 2017-2022. We relate a measure of OSS activity to conventional measures of entrepreneurial success, such as valuations and funding. We combine this approach with coarsened

[3] Nagle et al (2023), page 3.

exact matching (CEM) to account for firms selecting to be OSS contributors based on observable factors such as location, sub-industry, age, and initial funding. We also assess supply- and demand-side national endowments like human capital, OSS policies, and market size that may moderate the relationship between OSS contributions and entrepreneurial performance.

The paper finds a robustly positive relationship between OSS and entrepreneurial growth. Both selection and treatment effects drive this relationship. We measure selection directly: About a fourth of software ventures globally contribute to OSS, and these firms are not random. Compared to all software firms, these participating ventures are concentrated in hubs like San Francisco and produce more back-end software leveraging AI and less consumer-facing software like social platforms. The composition is consistent with the view that many entrepreneurial firms participate in open source communities to take advantage of access to frontier software.

We initially show that OSS contributions positively predict firm outcomes cross-sectionally and longitudinally. Cross-sectionally, a 1 percent increase in GitHub participation—i.e., code contributions labeled as "commits"—is associated with a 0.3 percent increase in total funding and valuations. This initial estimate contains both selection and treatment effects. Once we incorporate fixed effects in a matched sample to account for selection, these effects fall to a 0.03-0.1 percent increase in funding and valuations. This translates to an increase of 0.6 commits being associated with a \$160K increase in valuations and a \$1.8K increase in funding per firm per year on average.[4] We conclude that the decline in coefficient magnitudes from the cross-sectional to the firm-level matched analysis suggests that at least two-thirds of the positive relationship between OSS and entrepreneurial growth is due to selection, and the remaining share is due to treatment effects.

[4] This calculation is based on roughly 60 commits, \$6M in funding, and \$160M valuation per firm per year on average.

To assess how and why OSS has a larger impact in some circumstances, we assess how the relationship between OSS and global entrepreneurship varies by the national endowments of the firm headquarters (HQ). We find that the relationship between OSS and valuations is stronger among firms headquartered in countries with higher-quality labor supplies, adopting OSS policies mandating the use of OSS in public contracting, and larger commercial and government markets. This means that startups headquartered in countries with higher levels of these supply- or demand-side endowments see up to $240K increase in valuations. These moderators suggest that OSS's treatment effect complements supply-side and demand-side national endowments.

This study is the first to measure how OSS shapes entrepreneurial growth after a startup's founding globally, and deepens our understanding of the role of OSS in the global entrepreneurial process. To do this, the study analyzes the largest dataset ever compiled on the relationship between OSS and entrepreneurial activity after founding. It builds on prior work that shows that contributing to OSS predicts global entrepreneurial founding (Wright, Nagle, and Greenstein, 2023), the probability of getting funding in the US (Conti, Peukert, and Roche, 2021), and firm valuation in the US (Lin and Maruping, 2022). This paper shows that the benefits continue to accrue as early-stage ventures mature into ongoing commercial organizations. The analysis separates causal explanations by analyzing the contributions of selection and treatment. A large part of the observed positive effect of participation in OSS comes from higher-quality firms choosing to participate in OSS in the first place. Once analysis controlled for this selection effect, we also find that "treatment" effects are large and valuable.

This research also reveals the benefits from OSS participation do not arise in equal magnitude in every location. National policy, human capital, and market complements for OSS impact the growth caused by OSS. Prior work shows that contributing to OSS can improve access to human capital and customers through a signaling mechanism (Conti, Peukert, and Roche, 2021; von Hippel and von Krogh, 2003). This study shows that such factors also complement the development of OSS in growing ventures. As such, the positive effects of OSS contributions on entrepreneurial growth are contingent on having the right enabling environment with talent that can take advantage of the opportunity on the platform and customers who demand and gain value from the code.

The research further contributes to a growing literature on the impact of platforms on lowering the costs of entrepreneurship, where GitHub serves as the focal platform in this study. Prior works focus on the treatment effects of platforms, for example, in terms of providing access to lower-cost coding tools (Dushnitsky and Stroube, 2021). This study shows that which firms decide to adopt these digital technologies is also crucial to understanding the different roles these technologies play in their founding and after. Both selection and treatment are essential for understanding why OSS spurs growth from entrepreneurship.

Lastly, this study reveals the role of OSS in entrepreneurial scaling disparities across countries. Entrepreneurial growth faces sharp inequalities across economies (Conti and Guzman, 2021; Wright, 2023). This study offers one mechanism that explains these inequalities that extends prior research showing the importance of open knowledge (such as that embodied in OSS) for enabling private firm growth (Nagaraj, 2022). Because the benefits of OSS for post-founding growth are tied to human capital, OSS policy, and market complements, companies in less-endowed economies might benefit less than others from participating in OSS. This suggests that OSS can help nurture entrepreneurial growth worldwide, but most effectively through complementary investments into the supply side (human capital) and demand side of the local economy (commercial and government market capacity).

2. Framework and Hypotheses

In this section, we dig deeper into the theoretical underpinnings of how and when OSS contribution may play a role in entrepreneurial growth. This leads to two competing hypotheses that we then test empirically. Further, we consider market-level conditions that may moderate these main effects.

2.1 What is the role of OSS in entrepreneurial growth?

Prior research and theory suggest that the relationship between contributing to OSS and entrepreneurial growth can be positive or negative. Both selection and treatment effects can account for this relationship. Selection can be positive or negative. OSS can let in systematically lower- or higher-quality entrepreneurs into the market, which determines whether the startup survives and thrives as it grows. We also expect the treatment effects to be either positive or negative. OSS presents appropriation and differentiation costs and process benefits for startups. To unpack these possibilities, we explore the nature of the selection and treatment effects.

2.1.a The case for a negative relationship between OSS and entrepreneurial growth

OSS and entrepreneurial growth can have a negative relationship, driven by both selection and treatment effects. The low frictions associated with OSS might lower entry costs and enable predominantly lower-quality startups, and these underlying quality differences might impact subsequent firm performance. Indeed, prior work finds that OSS increases entrepreneurial founding (Nagle, 2019a; Wright, Nagle, and Greenstein, 2023). If OSS reduces the quality threshold for entry, the companies that emerge because of building on OSS (and thus those that are, in turn, contributing to OSS) might be less likely to succeed on average. Because contributing to OSS better enables learning frontier technologies (Nagle, 2018), lower-quality startups lacking this knowledge might be more likely to contribute. Because of this ex-ante lower quality, we expect these companies to perform worse in subsequent years, resulting in a negative relationship between OSS contributions and entrepreneurial growth.

Even conditional on these pre-entry quality differences, contributing to OSS might make it harder for these entrepreneurs to grow because it presents several new costs for participating firms. Perhaps most crucially, appropriating value from OSS-produced innovations can be difficult for startups. This is particularly challenging for startups whose fate runs on commercializing an—often—single product. Paying the legal and operational fees to navigate this appropriation process might not be feasible, given the scarce resources that young ventures confront (Pisano, 2006; Wen, Ceccagnoli, and Forman, 2015). When members wear multiple hats simultaneously, time, money, and talent are scarce in a startup team. Taking time from developing the core product or addressing the needs of critical customers to figure out how to capture value from contributing to OSS might be nearly impossible. Additionally, open sourcing key technology can reduce the ability of the firm to profit via licensing, an important pathway to revenue for many firms (Arora, Fosfuri, and Gambardella, 2001).

Further, when startups contribute their code to open source projects that other firms can use for free, those startups may find it challenging to differentiate their innovations—which existing research finds important for startup performance (Guzman and Li, 2022; Miric, Ozlap, and Yilmaz, 2023). Little already prevents startups in the same space from adopting similar open source code; for example, machine learning algorithms from the TensorFlow or PyTorch libraries, tools for classification from Scikit-learn, and app infrastructure from the React library. Startups contributing to these libraries expend their time and money only to expand this publicly available code. The burden is then to complement these technologies with novel elements, whether data, features, user experience, etc. However, contributing to OSS might use precious developer resources that could have otherwise been spent to develop these complementary, proprietary innovations. Further, these elements might be difficult to identify and create a moat around. When "looking" too much like other startups, capturing customers and investors becomes harder, inhibiting growth.

These possibilities suggest a negative relationship between contributing to OSS and entrepreneurial growth, driven by lower-quality firms being more likely to use and contribute to OSS (selection) and OSS presenting costs for startups as they create and capture the value that exceeds any benefits (treatment). These possibilities lead to Hypothesis 1.

H1. ***OSS contributions and entrepreneurial growth have a negative relationship.***

2.1.b The case for a positive relationship between OSS and entrepreneurial growth

Contributing to OSS and entrepreneurial growth might also exhibit a positive relationship. OSS might increase the quality threshold of companies entering the market. Prior research shows that OSS contributions complement local human capital and income endowments when considering entrepreneurial founding (Wright, Nagle, and Greenstein, 2023). This suggests that the ventures that emerge because of OSS are more likely to be well-positioned in the market and exhibit positive performance, irrespective of whether they continue contributing to OSS as they mature. This trend would result in a positive relationship between OSS contributions and entrepreneurial growth.

For a given level of quality, contributing to OSS also presents several benefits for startups that might fuel their growth after the entry decision. It reduces the burden on the OSS teams to re-learn and re-invent the basics by building on existing code (Kogut and Metiu, 2001). It could help reduce the cost of experimentation, a critical barrier to success for entrepreneurs (Agrawal, Gans, and Stern, 2021). Rather than learning to engineer a basic neural network model for a data science project from scratch, contributing to an off-the-shelf model using the TensorFlow or PyTorch libraries facilitates a more straightforward approach to generating an algorithm. Leveraging such standardized tools can create valuable products (Miric, Ozlap, and Yilmaz, 2023). When the basic library is not enough and new use cases arise, startups can contribute to shaping a more desired direction for the code. Access to software resources is unparalleled in an OSS world and prevents companies from having to reinvent the costly coding wheel. And they can do this in a way that enables coordination between team members, the broader OSS developer community, and customers. For example, OSS platforms allow participants to take advantage of version control. Even if a community member makes a coding mistake or pushes code that conflicts with another's code, it is always possible to return to an earlier version while benefiting from crowdsourcing and voluntary labor. This can save both time and money in the production of software projects. Such features make working with other team members and enterprise and government customers that may be embedded within the same projects easier. Further, contributing to OSS might attract potential users, customers, investors, and acquirers (Conti, Peukert, and Roche, 2021; Lerner and Tirole, 2002) and can position a firm as a gatekeeper in an OSS community, a role through which they can influence future contributions to their projects (Tang, Wang, and Tong, 2023).

These scenarios suggest a positive relationship between OSS contributions and entrepreneurial growth, driven by higher-quality startups contributing to OSS (selection) and OSS lowering the cost of production (treatment). We would observe the latter treatment effects by assessing whether contributing to OSS increases entrepreneurial growth within firms, therefore accounting for quality differences within firms. Hypothesis 2, which competes with Hypothesis 1, summarizes this prediction.

H2. ***OSS contributions and entrepreneurial growth have a positive relationship.***

2.2 Moderators

OSS production does not happen in a vacuum. It relies on a robust ecosystem of developers, users, and customers to generate, improve, and integrate the projects into everyday applications. This means that the human capital that can interpret, edit, and expand the OSS code is a critical asset. With prior OSS trends of reducing documentation (Lerner and Tirole, 2002), having talent with the base knowledge to operate on the platform becomes increasingly important. Therefore, if contributing to OSS boosts performance, it would only be able to do so in contexts with a sufficiently high-quality labor supply. This leads to the third hypothesis.

H3. ***The OSS contribution treatment effects are larger in country contexts with a higher-quality labor supply.***

But even with top-notch human capital, contributing to OSS can have little utility for startups if the right customers who would extract value from the software are not in place. Over the years, governments and enterprises have come to promote OSS in different capacities. As a quintessential example, Microsoft went from considering OSS a “cancer” (Chicago Sun-Times, 2001) to embracing it to the point where it acquired the largest OSS platform, GitHub. They have become one of the largest contributors to OSS. Governments, too, have become critical buyers of OSS, as shown by the passage of national policies to encourage the purchase of products built on OSS (CSIS, 2023). Therefore, we would expect startups headquartered in countries with such local government support for OSS, as well as larger commercial (B2B) and government markets to purchase the OSS products, would see a higher value from contributing to OSS. After all, having a top-notch open-source project is useless if it does not have the right users who generate value from it and integrate it into actual use cases. This leads to the fourth hypothesis.

H4a. ***The OSS contribution treatment effects are larger in country contexts where government procurement policies mandate the use of OSS.***

H4b. ***The OSS contribution treatment effects are larger in countries with larger 1) commercial and 2) government markets.***

3. Data

To test these hypotheses, we leverage a panel dataset that combines startup-level OSS data from GitHub and matches it with performance outcomes from PitchBook. This allows us to measure both the prevalence of the participation of OSS among startups worldwide and how that participation is associated with performance.

The data sample begins with 110,396 software firms from PitchBook.[5] We then narrowed this down to 95,715 firms founded since 2000. This ensures that we capture entrepreneurial firms that are approximately two decades old at most. We further filter to firms founded no later than 2016, when firms stably use GitHub as an OSS platform.[6] This ensures that every firm in our sample existed throughout the range of the data, which creates a more balanced panel of firms, allowing us to capture the intensive margin of effects better. Otherwise, analyses would drop some firms in some years if they were not yet founded, reducing our ability to disentangle the extensive from the intensive margin.[7] This leaves our final analysis with 61,379 global software firms founded from 2000 to 2016.

[5] This study with global startups has a smaller number of observations than Conti, Peukert, and Roche (2021) which examines only US startups from Crunchbase. This is because we focus on firms in the software industry and venture backed startups captured in PitchBook. We are also focusing on firms founded 2000-2016, rather than Conti et al.'s period of 2005-2020. The earlier-founded sample allows us to capture variation in the relationship between OSS contributions and performance across the life cycle of the firm, from founding to successful exit/failure, in addition to early funding rounds.

[6] Firms started to use GitHub in earnest in 2014, but stably so in 2016. Our main results hold when using 2016-2022 (when lagged log commits would be from 2015, before this stable use).

[7] The results are consistent when including firm founded 2000 to 2017.

We then match these firms to their open source GitHub activity from 2017-2022. Specifically, the GitHub data show when a user employed in the focal startup added a code contribution—hereafter, labeled as a "commit"—to an OSS project. We use the levels of "commits" in a year to measure a firm's participation in OSS in that year. This matching process allows us to create a 2017-2022 panel that shows startup OSS activity and performance outcomes across time. Our ultimate dataset is on a startup-year level to match the frequency of variables characterizing startups' countries.

Unique to this dataset is the matching between individual contributors and their parent firms. GitHub data available to the public do not distinguish the parent company of individual contributors, preventing a complete firm-level analysis and requiring reliance on "organization" accounts affiliated with companies. However, not all companies whose employees contribute to OSS utilize an organization account, thus leading to a potential underestimation of firms' OSS activities. Our data, in contrast, match the domain of the email addresses of individual contributors to that of their parent companies, enabling us to evaluate open source activity at the level of the firm.[8]

Our primary independent variable is logged GitHub commits in which users from a given company contributed to open source projects in the previous year. Each commit is approximately a line of code.[9]

Our primary dependent variables are various proxies of entrepreneurial growth drawn from PitchBook. These include the following:

[8] In OSS, it is normal to use your work email address to contribute if your firm is sponsoring your activity and your personal email address if you are contributing on your own time. Although not a requirement, this is a very strong norm. We thank GitHub for assisting in this matching process.
[9] A commit can be numerous lines of code, and a commit can represent the deletion of lines of code. On average, however, commits consist of a change to one line of code (Nagle, 2018).

- ***Log Valuation:*** This variable indicates a company's logged post-money valuation in USD (valuation after receiving funding). It reveals the potential value that investors expect from the given startup.
- ***Log Deal Size:*** This variable indicates the logged funding amount in USD that companies received. This variable is often used as a performance outcome for early-stage ventures that otherwise might not have readily available revenue data (Cao, Koning, and Nanda, 2021; Howell, 2017; Wright, Koning, and Khanna, 2023; Yu, 2020).
- ***Successful Exit:*** This variable indicates the probability that startups achieved a successful exit via an acquisition or an initial public offering (IPO). For the panel analysis, we code this variable as "1" for each year on or after a given company's initial acquisition or IPO.[10] Entrepreneurship studies commonly use this variable and timeframe to show growth-oriented entrepreneurs' ultimately highly skewed but desired outcomes (e.g., Guzman and Stern, 2020).

Table 1a shows summary statistics of all firms. Table 1b shows summary statistics comparing firms contributing to OSS and those not, illustrating the first evidence that participation in OSS is not random. It indicates that OSS-contributing firms generally see higher performance outcomes, including higher logged valuations, logged total funding, and the probability of a successful exit, suggesting a positive relationship between OSS and entrepreneurial growth. However, several systematic differences exist between these OSS-contributing and OSS-non-contributing firms. OSS-contributing firms are more likely to be headquartered in hubs like San Francisco. They are also more likely to produce back-end software often sold to enterprises relating to business

[10] This approach assumes that each company experiences only one IPO or acquisition event. This is the case for the vast majority of companies in our sample (>98%). The results hold when excluding firms that experienced more than one of such events.

productivity, AI, and software development rather than, for example, consumer-facing social platforms, according to PitchBook industry classifications.

Table 1c assesses this selection into contributing to OSS in a simple regression framework and finds similar results. Being headquartered in higher-income countries[11] and hubs[12] like San Francisco and operating in technology spaces mainly serving enterprise customers like business productivity, predict whether a firm contributes to OSS. Firms producing AI products, as detected in their company descriptions, are also more likely to contribute to OSS. These systematic differences suggest that there is likely to be a selection into which firms contribute to OSS. These results are consistent with the view that many entrepreneurial firms participate in OSS to gain access to frontier software in machine learning and generative AI, such as the TensorFlow and PyTorch libraries, which became available, respectively, from Google in 2015 and Facebook in 2016. As discussed further below, within-firm analyses help disentangle the treatment from the selection effects.

[Insert Tables 1a-c]

To understand the moderators in the relationship between OSS and entrepreneurial performance, we assess how this relationship varies by the quality of the local labor supply using the following proxies:

- **Human Capital Index:** This variable is an index of 0-1 and measures the quality of the human capital in each country in a given year.[13] Specifically, it captures the adult literacy

[11] We measure country income using GDP per capita data from the World Bank.
[12] We define hubs as the top 20 ecosystems defined by the Startup Genome Project (2023). These hub cities include San Francisco, New York City, London, Los Angeles, Tel Aviv, Boston, Beijing, Singapore, Shanghai, Seattle, Washington, D.C., Seoul, Berlin, Amsterdam, Tokyo, San Diego, Toronto, Paris, Chicago, Sydney, and Bangalore. We tag only these cities as hubs in our main results, but they are robust to using any cities within the 60-mile radius of these hub cities (where data are available) as done in the Startup Genome Project (2023).
[13] The Human Capital Index is only available in 2003, 2004, 2005, 2008, 2010, 2012, 2014, 2016, 2018, 2020, and 2022. We impute the values for the unavailable years of the year before.

rate, combined enrollment ratio of primary, secondary, and tertiary schooling, expected years, and average years of schooling. The data come from the United Nations and is available for 132/143 countries in our sample.

- **ICT Supply:** This variable indicates the number of ICT graduates in each country in a year. It comes from OECD data and is available for 36/143 countries in our sample.

We also assess variance in the relationship by the demand for OSS as proxied by government policies mandating the use of OSS in government purchasing. Our data on OSS-related government policies span from 2000-2022. OSS-related policies aimed to broadly improve government processes and economic activity rather than spur entrepreneurship per-se (CSIS, 2023). About a quarter of the headquarters countries in the sample (36/143) imposed such a mandatory policy since 2000, accounting for about two-thirds of the firms.[14] The timing of such policies varies, allowing for more precise identification. We construct two variables to measure OSS-related government policies:

- **OSS Policy:** We assess whether a startup's headquarters country has any policies in place in a given year.
- **Number of OSS Policies:** We also assess the cumulative number of policies in a given country in a given year.

Lastly, we examine the size of the markets of the country the company is headquartered in using the following proxies:

- **Commercial market size:** We measure this variable by aggregating the USD revenue of all companies in the PitchBook data in each country in each year. This variable indicates

[14] We assume that startups without a listed HQ country have not been exposed to an OSS policy, so the policy variable is 0 for them in all years. The results are similar if we instead drop these firms from the policy moderator analysis.

the size of the B2B market, which is the primary market of OSS-contributing firms. We assume the value is zero if data are missing for a country in a given year, allowing us to have coverage of all countries in our sample.

- **Government expenditure:** This variable reflects each country's total government expenditure in USD each year using World Bank and International Monetary Fund (IMF) data. The data are available for 118/143 countries in our sample.
- **GDP Per Capita:** We use data on GDP per capita in USD from the World Bank as a robustness proxy of market size. The data are available for 140/143 countries in our sample.

These human capital, policy, and market size moderators shed light on the supply- and demand-side mechanisms in the relationship between OSS and startup performance.

4. Empirical Specification

We begin with a cross-sectional analysis assessing entrepreneurial growth outcomes as the dependent variable, and GitHub commits as the "independent" variable. This analysis allows us to test between Hypotheses 1 and 2, whether there is a negative or positive relationship between OSS contributions and entrepreneurial growth. This analysis aggregates treatment and selection effects, requiring us to employ finer-grained approaches to disentangle these effects.

We conduct several additional layers of analysis to unpack these treatment and selection effects. The first is to use coarse exact matching (CEM), which allows us to limit the sample pool to firms with similar characteristics that do and do not participate in OSS across the 2017-2022 span of the data. These characteristics include firm age, headquarters country, whether the headquarters city is a hub, software sub-industry, and the size of initial funding. Comparing one firm that did participate in OSS with another similar one reduces concerns that the two firms are fundamentally different. CEM helps control for selection on observables. It does not, however, account for selection on unobservable factors.

To account for selection on unobservable factors, we also conduct a within-firm analysis to control for quality differences across firms and isolate the drivers of the treatment effect. We apply the OLS specification with firm fixed effects to the CEM-matched dataset in Equation (1).

We begin with Equation (1), which assesses how OSS contributions impact entrepreneurial growth outcomes within firms in the following year.

$$PERFORM_{it} = \alpha_{it} + \beta_1 OSS_{it-1} + f_i + \gamma_t + \varepsilon_{it} \quad (1)$$

Where $PERFORM_{it}$ indicates an array of entrepreneurial performance indicators for firm i in year *t*, including logged valuation and logged deal size. OSS_{it-1} indicates OSS contributions measured in two ways: whether or not firm i contributed to GitHub in the previous year (*t-1*) and that firm's logged total number of GitHub commits in that previous year. The former sheds light on the extensive margin—whether firms that did not contribute to OSS but then do. The latter indicates the intensive margin—whether firms already contributing to OSS are increasing their activity. We use the lag of GitHub commits since any impact on the growth outcomes would take time. f_i indicates firm fixed effects to account for observable and unobservable differences between firms. γ_t indicates year fixed effects to account for differences in trends across time. We use robust standard errors clustered at the firm level.

We next assess how the relationship between OSS contributions and entrepreneurial performance in Equation (1) varies across country-level supply- and demand-side endowments to shed light on the mechanisms behind this relationship and test Hypotheses 3-4. We do so with the following specification:

$$PERFORM_{it} = \alpha_{it} + \beta_1 OSS_{it-1} + \beta_2 ENDOW_{t-1} + \beta_3 ENDOW_{t-1} x OSS_{it-1} + f_i + \gamma_t + \varepsilon_{it}$$

(2)

Where $ENDOW_{t-1}$ indicates an array of supply- and demand-side factors that may shape the relationship between OSS contributions and entrepreneurial performance.

On the supply side, we assess two proxies of the quality of the local labor supply. The first is the Human Capital Index—a measure developed by the United Nations—in a given country in year t-1. The second is the number of technically trained individuals—a measure developed by the OECD—in a given country in year t-1.

On the demand side, we also assess two sets of proxies. The first consists of government policies enforced over the last two decades mandating OSS in government procurement to proxy government demand for OSS. These policies increased the participation in OSS across countries (Wright, Nagle, and Greenstein, 2023). We use policies that CSIS (2023) classifies as mandatory—all are related to procurement—because they are the most direct reflections of demand for OSS in government transactions. Other non-mandatory policies, such as procurement advisory, R&D, and training ("how-to-use") policies, do not directly reflect demand for OSS in commercial transactions. Instead, they might reveal general support for digital development. This policy endowment takes a value of "1" if year t is after the first OSS mandatory policy was imposed in a given country since 2000.

The second set of demand proxies measure the size of the country's markets. Specifically, we use the cumulative revenue of all companies in the PitchBook dataset per country per year to proxy the size of the commercial market and the businesses that can be the potential customers of the focal startup.[15] We also use government expenditure to proxy the size of the government market using data from the World Bank and IMF.

The coefficient of interest is that on the interaction term $ENDOW_{t-1}xOSS_{it-1}$, indicating how the relationship between OSS contributions and subsequent startup performance varies by country supply- and demand-side endowments. A positive coefficient on this interaction term suggests that OSS contribution is a complement to national endowments. A negative coefficient indicates that OSS contribution is a substitute for these endowments.

5. Results

5.1 Does OSS spur entrepreneurial growth?

We begin with a cross-sectional analysis testing the overall relationship between OSS and entrepreneurial growth. This allows us to test Hypotheses 1 and 2 against each other. We do so by assessing how OSS correlates with firm outcomes cross-sectionally. To understand the extent to which selection or treatment effects account for this relationship, we leverage the panel data set with firm fixed effects by applying the OLS specification in equation (1).

Beginning with the cross-sectional, Figures 1a-b show a binscatter with the GitHub activity software firms on Pitchbook founded 2000-2016[16] with age fixed effects. Logged commits on the x-axis positively correlate with logged funding (left) and logged valuations (right) as a proxy of entrepreneurial performance on the y-axis. These figures suggest a positive relationship between OSS contributions and entrepreneurial growth, consistent with the summary statistics in Table 1.

[Insert Figures 1a-b]

[15] If revenue is not reported for a given company in a given year in PitchBook, we assume that there was no revenue for that company in that year. This measure of market size has a high correlation with a cross-sectional measure of Software-as-a-Service (SaaS) market size from Statista Market Insights.
[16] The cross-sectional results are similar if broadening the sample to firms founded 2000-2022.

Table 2 assesses this relationship with additional performance proxies and controls. It also shows the full sample and a matched sample using CEM that limits to firms with comparable characteristics in terms of headquarters country, whether located in a hub, software sub-industry, initial financing amount, and age. The results are similar. A 1 percent increase in GitHub commits is associated with a roughly 0.3 percent increase in funding and valuations. This means that approximately a 0.6 commit increase is associated with a $18K increase in funding and a $480K increase in valuations per firm per year on average.[17] These results are consistent with Hypothesis 2: OSS and entrepreneurial growth have a positive relationship.

[Insert Table 2]

But to what extent do these results reflect selection or treatment? Is it that better quality firms contribute to OSS, or does contributing to OSS present benefits for firms that exceed any costs? While CEM can help account for selection effects based on observable factors like sub-industry, HQ country, initial funding, whether the HQ city is a hub, or age, it cannot account for unobservable differences between firms.

To help account for these unobservable differences, we next turn to a longitudinal analysis assessing how GitHub commits influence subsequent performance within a firm. This allows us to isolate the treatment effects more effectively because we can account for within-firm observable and unobservable differences. Figures 2a-b show a binscatter, like the one in Figure 1, but now capturing changes in OSS commits within firms founded 2000-2016 and their lagged commits 2017-2022. Lagged log commits on the x-axis, similarly here, are positively associated with logged funding (left) and valuations (right) in the following year when including firm fixed effects and

[17] This calculation is based on roughly 60 commits, $6M in funding, and $160M valuation per firm per year on average.

calendar year controls. Again, these figures suggest a positive relationship between OSS and entrepreneurial growth, now accounting for differences between firms.

[Insert Figures 2a-b]

Table 3 breaks this relationship down in regressions. The results are similar to those in Table 2, though the coefficients are roughly no more than one-third of the magnitude of the cross-sectional ones in Table 2 and weaken in significance for the successful exit as a performance outcome.[18,19] A 1 percent increase in GitHub commits is associated with a 0.1 percent increase in valuations in the following year and a 0.03 percent increase in funding. This means that roughly a 0.6 increase in commits is associated with a $160K increase in valuations and a $1.8K increase in funding per firm per year on average. The decline in the coefficient magnitudes suggests that earlier results captured both selection and treatment effects; thus, the difference reflects the selection effects. These results indicate that selection effects account for at least two-thirds of the results.

[Insert Table 3]

Do the returns from OSS persist as firms age? Table 4 shows how the relationship shown in Table 3 varies by the age of firms. The results are mixed. The interaction term between log commits and the age of firms is negative and significant at the 5 percent level when using logged funding as the dependent variable. It is positive and significant when using logged valuations and successful exit as the dependent variables. These results suggest that OSS may have diminishing returns for firms as they age regarding funding but increasing returns regarding valuations and successful exit.

---

[18] The differences in coefficients between the cross-sectional and longitudinal models are statistically significant.
[19] The results are positive and significant at the 5 percent level when using a Cox proportional hazard model with successful exit as the outcome.

[Insert Table 4]

5.2 How do supply- and demand-side factors moderate the OSS-performance relationship?

What could be driving these treatment effects? To address this, we next turn to understanding the moderators in the relationship between OSS contributions and entrepreneurial performance shown in Table 1. Specifically, we look at the role of supply- and demand-side factors. On the one hand, it could be that human capital complements on the supply-side—which have been shown to be important for OSS contribution's impact on entrepreneurial founding (Wright, Nagle, and Greenstein, 2023)—could be driving these results. Startups headquartered in places where there is a robust enough labor supply to produce OSS are the ones that reap the entrepreneurial growth benefits of OSS after these policies. There could also be a market channel on the demand side. Startups headquartered in places where the governments demand OSS, as proxied by national policies mandating the use of OSS in procurement, and where there are bigger commercial and government markets would be the ones that would most benefit from contributing to OSS after these policies. These mechanisms are not mutually exclusive.

To measure the drivers of these treatment effects, thereby testing Hypotheses 3-4, we assess how results in Table 1 vary across country contexts. Table 5 shows that the coefficient on GitHub commits is higher in countries with thick labor markets, as proxied by the UN Human Capital Index, and logged valuations and successful exit are the metrics of startup performance (Columns 2-3). A similar relationship holds when using the logged number of ICT graduates as a proxy of the human capital index and logged valuations and successful exit as the dependent variables (Columns 5-6). Including these interaction terms make the baseline positive coefficient on lagged log GitHub commits negative, suggesting that OSS increases entrepreneurial growth at least partly through a human capital channel. The coefficients on the same interaction terms are not statistically significantly different from zero when using logged funding as the dependent variable (Columns 1 and 4), suggesting that human capital endowments are only moderators for the OSS contribution-valuation and OSS contribution-successful exit relationships and not the OSS contribution-funding relationship. This difference may partly be due to a sample selection issue because some firms do not have valuation data available.

[Insert Table 5]

Tables 6-7 repeat this exercise but with two sets of proxies of market size. The first set reflects whether governments require OSS in procurement. We proxy this demand using government policies mandating the use of OSS in public contracting, specifying this as a binary measure of whether there is such a policy in a given country in a given year or as the cumulative number of such policies since 2000. Table 6 shows the results using these proxies. The relationship between OSS and logged valuations is larger in country contexts where there is a policy mandating the use of OSS in public contracting (Column 2) and more such policies (Column 5). The magnitude of the interaction terms is economically meaningful: about 30-50 percent of the baseline coefficient on lagged log commits. This suggests that for the average firm in the sample, a 1 percent increase in commits (an increase of 0.6 commits for the average firm) is associated with up to a $240K increase in valuations for startups headquartered in countries with an OSS policy. The relationship between OSS and successful exit is similarly larger in country contexts with more of these policies. The relationship between OSS and funding does not meaningfully vary depending on the policy in place, perhaps because of sampling differences between firms with available valuations versus the entire sample.

[Insert Table 6]

The second set of demand-size proxies reflects the size of the HQ country market of the startups. We measure both the size of the commercial market (logged B2B market) using cumulative company revenue per country per year (from PitchBook) and the size of the government market using logged government expenditure (from the IMF). Table 7 shows the relationship between OSS contributions and performance based on these market proxies. The relationship between OSS contributions and valuations is larger in bigger markets as proxied by the B2B market size (Column 2) and government expenditure (Column 5).[20] As with the human capital indicators, including these market moderators offsets the positive baseline effect of lagged OSS commits, suggesting that demand for OSS helps explain the relationship between OSS and valuations.[21] A similar trend emerges for the relationship between OSS contributions and successful exit for government expenditure as a market moderator (Column 6). The relationship between OSS and funding does not meaningfully vary by market size. As mentioned above, this may be because of a sample selection issue, as some firms are missing valuation data.

[Insert Table 7]

These results suggest that supply-side human capital and demand-side market factors moderate the relationship between OSS contributions and logged valuations but not logged funding, partially consistent with Hypotheses 3-4. The coefficient plot in Figure 3 summarizes these moderator effects and suggests that startups headquartered in more endowed contexts see roughly up to 50 percent higher returns from OSS than others.

[Insert Figure 3]

5.3 Interviews with software startups reinforce the quantitative findings.

---

[20] The results are similar when using government expenditure as a share of GDP instead of government expenditure in absolute terms.

[21] The positive moderating role of market factors in the OSS-valuation relationship holds if we use GDP Per Capita from the World Bank as a proxy of market size.

Qualitative data from interviews with software startups reinforce the conclusions about the positive relationship between OSS and entrepreneurial growth and the important role of both supply- and demand-side factors. Several startups discussed how OSS served a crucial role in their growth process, particularly in expanding their user base of developers, who then attracted commercial clients—growing the human capital supply enabled the growth of market demand. For example, one US company discussed how open source enabled the startup to attract developer users and commercial customers:

> So we created these three pillars to grow the flywheel…[with] an open core business model. The open source community is what allows you not just to build innovation but create a whole awareness about the product. This awareness is what allows us to do a bottom up approach to getting to some organizations like [large US public and private sector organizations]. Like lots of different organizations where the developers get engaged, get to know our tool, and then they bring in to their organizations. And then eventually the product becomes so important, those organizations…they engage with us to get an enterprise license…and so on.

This "flywheel" that emerged due to open source attracting developers to use the technology, who then attract their organizations to license that technology reveals a business model resting on network effects. The more users within an organization that uses the technology, the more likely their organization will license that technology. Similarly, a US-based company discussed OSS as an important channel to get its technology to commercial clients.

> We're an open source company. And we are a B2B-like enterprise SaaS company. So in order to grow along any of the dimensions you mentioned—so marketing is almost out, it doesn't make a difference, the most important thing for us is to increase our open source funnel… that plays out as a distribution channel…our commercial leads come from our open source community… So our goal, therefore, is to get our product in as many hands as possible, such that those folks, if one in 100 likes it and shows up and engages with us commercially, that's a win in our business model. So the question, then, is how do we get our product in as many hands as possible?

The US-based company reveals that OSS served as an important distribution channel of their technology that enabled creating a "flywheel." This distribution channel substituted traditional marketing tools companies might use to attract customers.

Such a bottoms-up business model resonated among startups in the US and various pockets of the world. A New Zealand-based company discussed how OSS enabled them to create a similar "flywheel" in an innovative technology that otherwise would be difficult to sell commercially right away:

> As an open source project, we have a very wide applicability to a fundamental shift in technology, which we are enabling...we have this intrinsic flywheel of open source growth...So we are sitting on an opportunity that we respect deeply, but [requires] taking something which is popular as a free tool and turning it into a paid tool.

OSS enabled the company to gain users who then attracted other users to unleash exponential growth. This "flywheel" allows the company to reap the benefits of network effects to grow the organization.

One way OSS attracted users was by educating developers about software that created demand for that technology. To this end, one Netherlands-based startup discussed how they invested time and energy into expanding knowledge within their open source community:

> We're [an] open source company. So there is a huge effort of sharing knowledge with the community. And so that is actually our biggest...investment...educating and sharing knowledge on what the technology is...And providing the free versions of our software… We're sort of investing in the open source community here. For them to also share it with their friends and colleagues in new companies and other projects and things like that.

The company reveals how investing in a knowledge-sharing capability with developers enabled the organization to build demand for their software where it might not have existed. Creating this demand allowed the company to attract more developer users who could then attract commercial clients into contracts. This process created the "flywheel" that the US- and New Zealand-based companies also discussed.

The interviews suggest that open source software serves a valuable distribution role for software startups that helps them attract developer users and commercial customers. In this way, OSS depends on the quality of developer users' human capital and the commercial market capacity to enable startup growth. The resulting "flywheel" might help explain why we estimate that OSS brings diminishing returns to funding over time and increasing returns to valuations. Funding is initially essential to grow the developer user base in lieu of commercial contracts. But as the developer users attract paying company clients, those external funding needs decline while the company's value increases. These findings are consistent with the quantitative results showing a positive relationship between OSS and entrepreneurial growth, positively moderated by supply-side human capital and demand-side market size factors.

5. Robustness

We conduct several robustness checks. To better understand whether the impact of OSS contributions on startup performance is driven by the intensive versus extensive margins, we assess how OSS impacts each firm's year-on-year change in GitHub commits, in addition to whether firms contributed to OSS at all and their logged commits. Table A1 repeats the OLS analysis from Table 3 with log annual change in commits as the main proxy of contributions and reveals similar results. An increase in commits in the last year predicts more funding and higher valuations.

To further understand the role of the intensive margin, Table A2 now repeats the analysis from Table 3 with the logged number of committers as the main proxy of contributions and shows similar results. An increase in the number of committers in the previous year predicts more funding and higher valuations. These trends provide additional support for OSS's impact on entrepreneurial growth through the intensive margin.

We also confirm that the results hold without lags. Tables A3-A6 repeat the analyses from Tables 3-6 with contemporary versions of OSS contributions. The results are generally consistent with the moderators. For the latter, the supply- and demand-side factors generally moderate the relationship between OSS and logged funding and logged valuations. We use the lagged specification in the main results to better account for potential reverse causality: more funding or higher valuations might prompt companies to contribute more to OSS. While out of the scope of this study, this reverse relationship might very well also exist. Indeed, financial resources seem to complement OSS production, as seen in the selection of higher-performing firms contributing to OSS. That being said, our lagged specifications help us disaggregate this reverse causality from our estimates.

We further confirm that the baseline and supply-side moderating results generally hold when excluding US-based firms from the sample. The results with the demand-side moderators do not in this narrowed sample. This demand-side inconsistency may be because the US ranks among the top in these demand-side measures. US firms account for the largest share of firms in the sample, so removing them substantially reduces the demand-side variance in the sample.

6. Conclusion

In this study, we assess the role of participation in open source software communities after founding a startup, focusing on global entrepreneurial participation in OSS. We find that OSS contributions and entrepreneurial growth exhibit a robust positive relationship. Selection effects account for roughly two-thirds of the magnitude of this relationship. We find no evidence of a selection of lower-quality firms. Higher-quality firms are more likely to contribute to OSS, and OSS boosts their subsequent performance. These treatment effects account for the remaining roughly one-third of the magnitude of the relationship between OSS and entrepreneurial growth. A 1 percent increase in GitHub commits is associated with a 0.1 percent increase in valuations in the following year and a 0.03 percent increase in funding. This means that roughly a 0.6 increase in commits is associated with a $160K increase in valuations and a $1.8K increase in funding per firm per year on average. Supply- and demand-side endowments moderate the relationship between OSS contributions and logged valuations: this relationship increases among startups headquartered in countries with higher-quality labor supplies, countries with policies mandating the use of OSS in public contracting, and larger overall and specifically public contracting markets.

This study also deepens our understanding of the role of OSS contributions in the entrepreneurial life cycle, complementing prior work that shows that OSS contributions predict global entrepreneurial founding (Wright, Nagle, and Greenstein, 2023) and the probability of getting funding in the US (Conti, Peukert, and Roche, 2021). This paper shows that ventures continue to experience benefits from their participation in OSS as they mature, but more so in more endowed contexts.

This research also reveals the complementary nature of human capital and market readiness for OSS contribution's impact on growth. Prior work shows that OSS can improve access to human capital and customers through a signaling mechanism (Conti, Peukert, and Roche, 2021; Nagle, 2019b; von Hippel and von Krogh, 2003). This study shows that such factors enable the positive effects of OSS contributions to ventures. As such, investing in the right enabling environment with talent that can produce on the platform and customers who can gain value from the code is an important condition to realize the benefits of contributing to OSS.

Lastly, the paper sheds light on the role of OSS in entrepreneurial scaling disparities across countries. Entrepreneurial growth faces sharp inequalities across economies (Conti and Guzman, 2021; Wright, 2023). This study reveals OSS contributions as one channel through which these inequalities might emerge and how they can be overcome. Companies in less-endowed economies might benefit less than others from contributing to OSS if the benefits are tied to human capital and market complements. This suggests that investing in human capital, government policies supporting OSS, and market capacity are important to enabling OSS to nurture entrepreneurial growth worldwide.

The conclusions of this study also raise many open questions. If OSS helps firms grow, then it implies software assets are contributing to firm growth for which there are no priced licensing transactions. That means the software does not appear on any balance sheet, and its contribution will be overlooked. That suggests the potential for managerial oversight and mismeasurement (Greenstein and Nagle, 2014). The unrecorded value is potentially enormous, given the size of GitHub and the number of participants. This motivates for further measurement to estimate that value (Calderon et al., 2022) and add it to country-level estimates of the software stock.

The findings about the type of firms that select into open source communities also raise questions about the value to different entrepreneurial clusters from gaining access to frontier software. These findings are consistent with a large and non-random group of firms taking advantage of the availability of frontier machine-learning frameworks, such as TensorFlow and PyTorch, and with many benefits accruing to Silicon Valley-based entrepreneurial ventures. That motivates measurements to estimate the value of local skilled labor markets and invest in skills for talent (Rock, 2022). It also motivates further analysis of how high-skilled labor markets and entrepreneurial ventures adjust to discreet shifts in the technical frontier.

Tables

Table 1a: Summary statistics

	Obs.	Mean	Std. Dev.	Min.	Max.
Total Commits (Thousands)	61379	0.36	14.7	0	3313.4
Log Total Commits	61379	0.95	2.02	0	15.0
Valuation (Mil. USD)	22027	159.9	1766.3	0.000011	168333.3
Log Valuation	22027	2.84	1.74	0.000011	12.0
Deal Size (Mil. USD)	61379	32.9	675.4	0	97575
Log Total Deal Size	61379	0.77	1.47	0	11.5
Successful Exit	61379	0.18	0.39	0	1
Hub HQ	61091	0.31	0.46	0	1
Human Capital Index	60667	0.87	0.091	0.11	1
ICT Talent (Thousands)	48793	103.5	81.9	0.10	183.0
Log ICT Talent	48793	10.8	1.50	4.66	12.1
OSS Policy	61379	0.69	0.46	0	1
Number of OSS Policies	61379	2.09	1.89	0	6
B2B Market Size (Bil. USD)	61260	1606.4	1570.0	0	3473.8
Log B2B Market Size	61260	13.0	2.29	0	15.1
Gov Expenditure (Bil. USD)	55314	2877.3	2569.5	0.14	5686.7
Log Gov. Expenditure	55314	27.8	1.64	18.8	29.4
GDP Per Capita (Thous. USD)	61140	47.9	22.8	0.50	204.2
Log GDP Per Capita	61140	10.5	0.88	6.21	12.2
Age (Years)	61379	10.7	4.21	6	22
US HQ	61260	0.41	0.49	0	1
San Francisco HQ	60701	0.046	0.21	0	1
East Asia HQ	61162	0.16	0.37	0	1
Europe HQ	61162	0.30	0.46	0	1
Latin America HQ	61162	0.026	0.16	0	1
MENA HQ	61260	0.032	0.18	0	1
South Asia HQ	61162	0.032	0.18	0	1
Sub-Saharan Africa HQ	61162	0.0078	0.088	0	1
Bus. Productivity Software	61379	0.38	0.48	0	1
Education Software	61379	0.055	0.23	0	1
Social Platform	61379	0.090	0.29	0	1
Financial Software	61379	0.11	0.32	0	1
Software Dev	61379	0.027	0.16	0	1
AI	61379	0.076	0.27	0	1
Observations	61379				

Table 1b: Summary statistics comparing OSS- and non-OSS-contributing firms

	(1)				
	Non-OSS Contributor Obs	Non-OSS Contributor Mean	OSS Contributor Obs	OSS Contributor Mean	Difference
Valuation (Millions USD)	13738	79.10	8289	293.91	-214.81***
Log Valuation	13738	2.38	8289	3.59	-1.21***
Total Deal Size (Millions USD)	46719	13.52	14660	94.72	-81.20***
Log Total Deal Size	46719	0.46	14660	1.78	-1.32***
Successful Exit	46719	0.16	14660	0.25	-0.08***
Hub HQ	46448	0.30	14643	0.37	-0.07***
Human Capital Index	46145	0.87	14522	0.87	-0.01***
ICT Talent (Thousands)	36700	102.41	12093	106.86	-4.46***
Log ICT Talent	36700	10.82	12093	10.92	-0.09***
OSS Policy	46719	0.68	14660	0.72	-0.03***
Number of OSS Policies	46719	2.06	14660	2.20	-0.14***
B2B Market (Billions USD)	46628	1575.32	14632	1705.58	-130.27***
Log B2B Market	46628	12.98	14632	13.22	-0.24***
Gov Expenditure (Billions USD)	41813	2833.36	13501	3013.24	-179.88***
Log Gov Expenditure	41813	27.80	13501	27.95	-0.15***
GDP Per Capita (Thousands USD)	46543	47.18	14597	50.01	-2.83***
Log GDP Per Capita	46543	10.51	14597	10.61	-0.10***
Age (Years)	46719	10.71	14660	10.71	0.00
US HQ	46628	0.40	14632	0.44	-0.04***
San Francisco HQ	46099	0.04	14602	0.07	-0.03***
East Asia HQ	46558	0.17	14604	0.15	0.02***
Europe HQ	46558	0.30	14604	0.28	0.02***
Latin America HQ	46558	0.03	14604	0.03	0.00
MENA HQ	46628	0.03	14632	0.03	0.01***
South Asia HQ	46558	0.03	14604	0.03	0.00
Sub-Saharan Africa HQ	46558	0.01	14604	0.01	0.00**
Bus. Productivity Software	46719	0.35	14660	0.48	-0.13***
Education Software	46719	0.06	14660	0.05	0.01**
Social Platform	46719	0.11	14660	0.03	0.08***
Financial Software	46719	0.11	14660	0.14	-0.03***
Software Dev	46719	0.02	14660	0.05	-0.03***
AI	46719	0.06	14660	0.12	-0.06***
N	61379				

*p<.05; **p<.01; ***p<.001

Table 1c: Firms headquartered in endowed contexts with back-end AI technologies are more likely to contribute to OSS.

	(1) Whether Contributed	(2) Whether Contributed	(3) Whether Contributed	(4) Whether Contributed	(5) Whether Contributed	(6) Whether Contributed	(7) Whether Contributed
Age	-0.001						
	(0.001)						
Hub		0.072^{***}					
		(0.009)					
Log GDP Capita			0.021^{**}				
			(0.008)				
SF HQ				0.125^{***}			
				(0.005)			
Bus. Prod.					0.098^{***}		
					(0.021)		
AI						0.144^{***}	
						(0.019)	
Social Software							-0.179^{***}
							(0.018)
_cons	0.253^{***}	0.217^{***}	0.022	0.235^{***}	0.202^{***}	0.228^{***}	0.255^{***}
	(0.015)	(0.007)	(0.081)	(0.006)	(0.008)	(0.001)	(0.002)
N	61241	60972	61140	60701	61241	61241	61241
Age FEs	No	Yes	Yes	Yes	Yes	Yes	Yes
Country FEs	Yes	No	No	No	Yes	Yes	Yes
Sub Industry FEs	Yes	Yes	Yes	Yes	No	No	No
Countries	124	143	140	140	124	124	124

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: GitHub commits predict firm performance cross-sectionally.

	(1) Log Deal Size	(2) Log Deal Size	(3) Log Deal Size	(4) Log Deal Size	(5) Log Val.	(6) Log Val.	(7) Log Val.	(8) Log Val.	(9) Success. Exit	(10) Success. Exit	(11) Success. Exit	(12) Success. Exit
Log Commit	0.302***	0.277***			0.278***	0.252***			0.016***	0.013***		
	(0.030)	(0.028)			(0.017)	(0.016)			(0.001)	(0.001)		
Whether Contributed			1.207***	1.057***			1.151***	0.959***			0.079***	0.067***
			(0.142)	(0.126)			(0.104)	(0.089)			(0.004)	(0.006)
_cons	0.485***	0.607***	0.484***	0.628***	2.392***	2.535***	2.405***	2.604***	0.169***	0.190***	0.165***	0.183***
	(0.029)	(0.051)	(0.034)	(0.058)	(0.028)	(0.041)	(0.039)	(0.052)	(0.001)	(0.002)	(0.001)	(0.003)
N	61241	31444	61241	31444	21965	13803	21965	13803	61241	31444	61241	31444
Age FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub Industry FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Countries	124	97	124	97	84	74	84	74	124	97	124	97

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: GitHub commits predict firm performance outcomes in the subsequent year.

	(1) Log Deal Size	(2) Log Deal Size	(3) Log Deal Size	(4) Log Deal Size	(5) Log Valuation	(6) Log Valuation	(7) Log Valuation	(8) Log Valuation	(9) Success. Exit	(10) Success. Exit	(11) Success. Exit	(12) Success. Exit
Lagged Log Commit	0.026***	0.027***			0.091***	0.089***			0.001+	0.001		
	(0.003)	(0.003)			(0.003)	(0.003)			(0.001)	(0.001)		
Lagged Whether Contributed			0.079***	0.079***			0.235***	0.233***			0.002	0.002
			(0.008)	(0.008)			(0.008)	(0.008)			(0.002)	(0.002)
_cons	0.189***	0.276***	0.190***	0.278***	2.653***	2.931***	2.672***	2.958***	0.143***	0.160***	0.144***	0.160***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
N	368274	189192	368274	189192	117833	74325	117833	74325	368274	189192	368274	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Firms	61,379	31,532	61,379	31,532	21,529	13,569	21,529	13,569	61,379	31,532	61,379	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: The effects of GitHub have diminishing returns as firms age for funding and increasing returns as firms age for valuations and successful exit.

	(1) Log Deal Size	(2) Log Deal Size	(3) Log Valuation	(4) Log Valuation	(5) Success. Exit	(6) Success. Exit
Lagged Log Commit x Age	-0.002**	-0.002*	0.006***	0.004***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Lagged Age	-0.001	-0.002*	0.085***	0.104***	0.015***	0.020***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)
Lagged Log Commit	0.041***	0.041***	0.045***	0.058***	-0.018***	-0.015***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.001)	(0.001)
_cons	0.196***	0.293***	2.009***	2.146***	0.034***	0.014***
	(0.005)	(0.008)	(0.012)	(0.016)	(0.002)	(0.003)
N	368274	189192	117833	74325	368274	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	No	No	No	No	No	No
Matched Sample	No	Yes	No	Yes	No	Yes
Firms	61,379	31,532	21,529	13,569	61,379	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Valuation effects are larger in HQ markets with a higher-quality labor supply.

	(1) Log Deal Size	(2) Log Valuation	(3) Success. Exit	(4) Log Deal Size	(5) Log Valuation	(6) Success. Exit
Lagged Log	0.013	0.343***	0.051***			
Commit x Human Capital Index	(0.034)	(0.045)	(0.007)			
Lagged Human	-0.187	-0.251	0.054			
Capital Index	(0.170)	(0.325)	(0.054)			
Lagged Log				0.002	0.017***	0.001*
Commit x ICT Talent				(0.002)	(0.002)	(0.000)
Lagged Log				0.033	-0.014	0.026***
ICT Talent				(0.022)	(0.057)	(0.008)
Lagged Log	0.015	-0.214***	-0.043***	0.010	-0.096***	-0.013*
Commit	(0.029)	(0.039)	(0.006)	(0.020)	(0.028)	(0.005)
_cons	0.438**	3.150***	0.114*	-0.072	3.049***	-0.117
	(0.147)	(0.285)	(0.046)	(0.240)	(0.635)	(0.085)
N	187218	73937	187218	151572	63515	151572
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,203	13,496	31,203	25,262	11,534	25,262

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Valuation effects are larger in HQ markets that mandate OSS in public contracting.

	(1) Log Deal Size	(2) Log Valuation	(3) Success. Exit	(4) Log Deals Size	(5) Log Valuation	(6) Success. Exit
Lagged Log	0.006	0.030***	-0.002			
Commit x Whether Policy	(0.006)	(0.008)	(0.001)			
Lagged	-0.020	-0.099^{+}	-0.016^{+}			
Whether Policy	(0.025)	(0.058)	(0.009)			
Lagged Log				0.002	0.021***	0.003***
Commit x Num. Policies				(0.002)	(0.002)	(0.000)
Lagged Num.				0.004	-0.011*	0.000
Policies				(0.003)	(0.005)	(0.001)
Lagged Log	0.023***	0.066***	0.002	0.025***	0.058***	-0.003***
Commit	(0.005)	(0.006)	(0.001)	(0.004)	(0.004)	(0.001)
_cons	0.290***	3.007***	0.170***	0.270***	2.946***	0.159***
	(0.017)	(0.045)	(0.006)	(0.005)	(0.008)	(0.001)
N	189192	74325	189192	189192	74325	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,532	13,569	31,532	31,532	13,569	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Valuation effects are larger in larger HQ commercial and government markets.

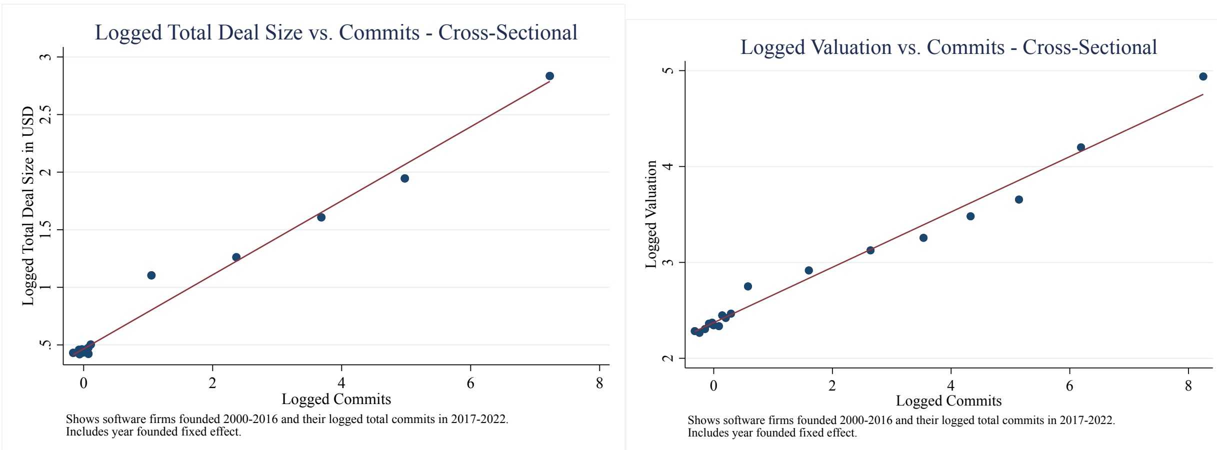
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Deal Size	Log Valuation	Success. Exit	Log Deal Size	Log Valuation	Success. Exit
Lagged Log	-0.001	0.007^{***}	-0.000			
Commit x Log B2B Market	(0.001)	(0.002)	(0.000)			
Lagged Log	-0.012^{**}	-0.013^{*}	-0.003^{***}			
B2B Market	(0.004)	(0.006)	(0.001)			
Lagged Log				0.001	0.019^{***}	0.002^{***}
Commit x Gov Expend				(0.002)	(0.002)	(0.000)
Lagged Gov				-0.095^{***}	0.076^{*}	0.004
Expend				(0.022)	(0.038)	(0.007)
Lagged Log	0.041^{**}	-0.003	0.001	-0.008	-0.460^{***}	-0.050^{***}
Commit	(0.015)	(0.023)	(0.003)	(0.050)	(0.068)	(0.012)
_cons	0.430^{***}	3.109^{***}	0.196^{***}	2.921^{***}	0.697	0.059
	(0.047)	(0.085)	(0.010)	(0.608)	(1.083)	(0.202)
N	188808	74185	188808	170130	69834	170130
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,468	13,543	31,468	28,355	12,753	28,355

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figures

Figures 1a-b: GitHub commits predict funding (left) and valuations (right) cross-sectionally.



Figures 2a-b: GitHub commits predict logged funding (left) and valuations (right) longitudinally.

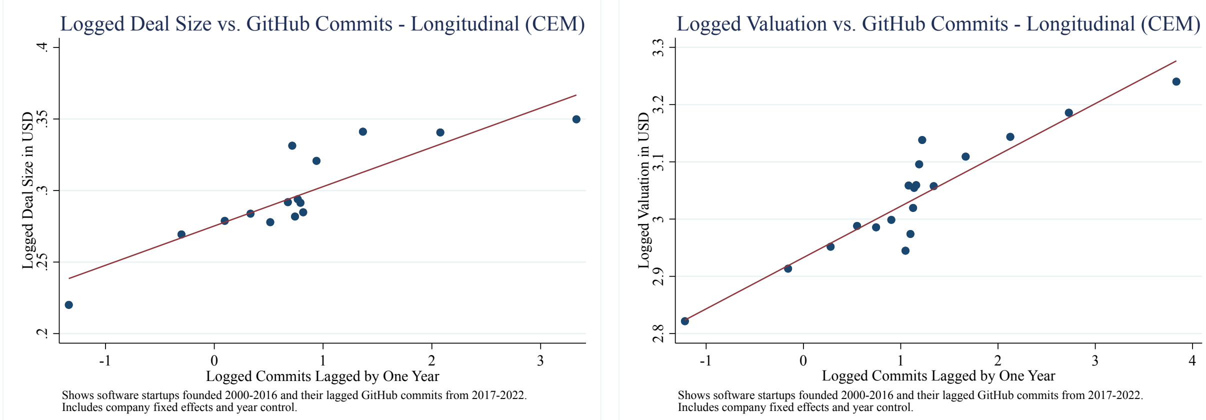
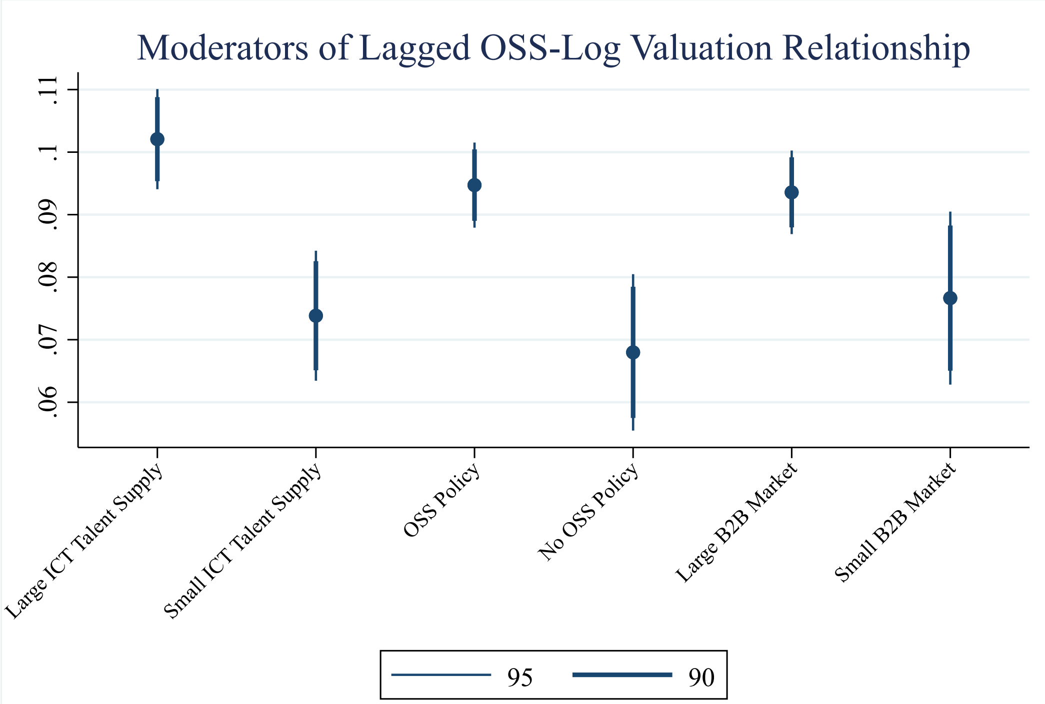


Figure 3: Effects larger in higher-quality labor markets and where there is demand for OSS.



The graph shows coefficients from regressing log valuation on lagged log commits with CEM matching.

Appendix

Table A1: The annual change in GitHub commits predicts firm performance outcomes in the subsequent year.

	(1) Log Valuation	(2) Log Valuation	(3) Log Deal Size	(4) Log Deal Size	(5) Success. Exit	(6) Success. Exit
Lagged Log	0.100***	0.096***	0.028***	0.029***	0.002**	0.002*
Annual Commit Change	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)
_cons	2.552***	2.830***	0.175***	0.261***	0.137***	0.149***
	(0.002)	(0.003)	(0.001)	(0.002)	(0.000)	(0.000)
N	104568	61372	344534	165841	344534	165841
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes	No	Yes
Firms	21,157	13,208	61,307	31,461	61,307	31,461

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Logged number of GitHub committers predict firm performance outcomes in the subsequent year.

	(1) Log Valuation	(2) Log Valuation	(3) Log Deal Size	(4) Log Deal Size	(5) Success. Exit	(6) Success. Exit
Lagged Log	0.196***	0.193***	0.057***	0.058***	0.003*	0.001
Committer	(0.006)	(0.006)	(0.006)	(0.006)	(0.001)	(0.001)
_cons	2.638***	2.908***	0.187***	0.271***	0.143***	0.160***
	(0.003)	(0.004)	(0.001)	(0.002)	(0.000)	(0.001)
N	117833	74325	368274	189192	368274	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes	No	Yes
Firms	21,529	13,569	61,379	31,532	61,379	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Baseline effects generally hold without lags.

	(1) Log Deal Size	(2) Log Deal Size	(3) Log Deal Size	(4) Log Deal Size	(5) Log Valuation	(6) Log Valuation	(7) Log Valuation	(8) Log Valuation	(9) Success. Exit	(10) Success. Exit	(11) Success. Exit	(12) Success. Exit
Log Commit	0.053***	0.052***			0.094***	0.094***			-0.007***	-0.007***		
	(0.003)	(0.003)			(0.003)	(0.003)			(0.001)	(0.001)		
Whether			0.141***	0.138***			0.229***	0.231***			-0.020***	-0.020***
Contributed			(0.008)	(0.008)			(0.009)	(0.009)			(0.002)	(0.002)
_cons	0.178***	0.256***	0.182***	0.264***	2.649***	2.923***	2.672***	2.958***	0.147***	0.166***	0.146***	0.165***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)	(0.000)	(0.001)	(0.000)	(0.000)
N	368274	189192	368274	189192	117833	74325	117833	74325	368274	189192	368274	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Firms	61,379	31,532	61,379	31,532	21,529	13,569	21,529	13,569	61,379	31,532	61,379	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Human capital moderators generally hold without lags.

	(1) Log Deal Size	(2) Log Valuation	(3) Success. Exit	(4) Log Deal Size	(5) Log Valuation	(6) Success. Exit
Log Commit x	0.048	0.384***	0.038***			
Human Capital Index	(0.035)	(0.049)	(0.007)			
Human	0.042	-0.871**	0.070			
Capital Index	(0.164)	(0.313)	(0.050)			
Log Commit x				0.010***	0.015***	-0.000
ICT Talent				(0.002)	(0.003)	(0.000)
Log ICT				0.035	-0.003	0.031***
Talent				(0.027)	(0.070)	(0.009)
Log Commit	0.010	-0.247***	-0.041***	-0.052*	-0.069*	-0.004
	(0.031)	(0.044)	(0.006)	(0.021)	(0.030)	(0.005)
_cons	0.220	3.690***	0.105*	-0.125	2.926***	-0.160
	(0.143)	(0.276)	(0.044)	(0.290)	(0.785)	(0.102)
N	187218	73937	187218	151572	63515	151572
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,203	13,496	31,203	25,262	11,534	25,262

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Policy moderators generally hold without lags.

	(1) Log Deal Size	(2) Log Valuation	(3) Success. Exit	(4) Log Deals Size	(5) Log Valuation	(6) Success. Exit
Log Commit x	0.028***	0.024**	-0.005***			
Whether Policy	(0.006)	(0.008)	(0.001)			
Whether	0.005	-0.121*	-0.013			
Policy	(0.024)	(0.059)	(0.009)			
Log Commit x				0.001	0.026***	0.002***
Num. Policies				(0.002)	(0.002)	(0.000)
Num.				-0.009***	-0.018***	-0.000
Policies				(0.003)	(0.004)	(0.001)
Log Commit	0.032***	0.076***	-0.004**	0.051***	0.052***	-0.010***
	(0.005)	(0.007)	(0.001)	(0.004)	(0.004)	(0.001)
_cons	0.252***	3.016***	0.175***	0.268***	2.949***	0.166***
	(0.017)	(0.046)	(0.006)	(0.004)	(0.007)	(0.001)
N	189192	74325	189192	189192	74325	189192
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,532	13,569	31,532	31,532	13,569	31,532

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Market moderators generally hold without lags.

	(1) Log Deal Size	(2) Log Valuation	(3) Success. Exit	(4) Log Deal Size	(5) Log Valuation	(6) Success. Exit
Log Commit x	0.005***	0.012***	-0.000			
Log B2B Market	(0.001)	(0.002)	(0.000)			
Log B2B	-0.005+	0.002	0.001			
Market	(0.003)	(0.005)	(0.001)			
Log Commit x				0.010***	0.019***	0.000
Gov Expend				(0.002)	(0.003)	(0.000)
Log Gov				-0.002	0.124***	0.016**
Expend				(0.022)	(0.032)	(0.006)
Log Commit	-0.013	-0.070**	-0.005	-0.214***	-0.454***	-0.014
	(0.015)	(0.024)	(0.003)	(0.052)	(0.071)	(0.012)
_cons	0.324***	2.892***	0.155***	0.321	-0.666	-0.284
	(0.040)	(0.063)	(0.009)	(0.601)	(0.918)	(0.174)
N	188808	74185	188808	170130	69834	170130
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Matched Sample	Yes	Yes	Yes	Yes	Yes	Yes
Firms	31,468	13,543	31,468	28,355	12,753	28,355

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Agrawal, A., Gans, J. S., & Stern, S. (2021). Enabling entrepreneurial choice. *Management Science*, *67*(9), 5510-5524.

Arora, A., Fosfuri, A., & Gambardella, A. (2001). Markets for technology and their implications for corporate strategy. *Industrial and Corporate Change*, *10*(2), 419-451.

Calderón, J., Guci, L., Robbins, C., Korkmaz, G., & Kramer, B. (2022). "Measuring the Cost of Open Source Software Innovation on GitHub," Bureau of Economic Analysis Working Paper, WP2022-10. https://www.bea.gov/system/files/papers/BEA-WP2022-10.pdf

Cao, R., Koning, R. M., & Nanda, R. (2021). Biased sampling of early users and the direction of startup innovation. National Bureau of Economic Research, No. w28882.

Chicago Sun-Times. (2001). Interview with Steve Ballmer.

Conti, A., Peukert, C., & Roche, M. (2021). Beefing IT Up for Your Investor? Open Sourcing and Startup Funding: Evidence from Github." Harvard Business School Working Paper, No. 22-001.

Conti, A, & Guzman, J. (2021). "What is the US comparative advantage in entrepreneurship? Evidence from Israeli migration to the United States." *Review of Economics and Statistics*: 1-45.

CSIS. (2023). Government Open Source Software Policies. https://www.csis.org/programs/strategic-technologies-program/resources/government-open-source-software-policies.

DiBona, C, Stone, M., & Cooper, D. (2005). Open Source 2.0: The Continuing Evolution. O'Reilly Media.

Fosfuri, A., Giarratana, M. S., & Luzzi, A. (2008). The penguin has entered the building: The commercialization of open source software products. *Organization Science*, *19*(2), 292-305.

Dushnitsky, G., & Stroube, B. K. (2021). Low-code entrepreneurship: Shopify and the alternative path to growth. Journal of Business Venturing Insights, 16, e00251.

Greenstein, S., & Nagle, F. (2014). Digital dark matter and the economic contribution of Apache. *Research Policy*, *43*(4), 623-631.

Guzman, J., & Li, A. (2022). Measuring founding strategy. Management Science, 69(1), 101 118.

Guzman, J., & Stern, S. (2020). The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US States, 1988–2014. American Economic Journal: Economic Policy, 12(4), 212-243.

Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. American Economic Review, 107(4), 1136–1164.

Kogut, B., & Metiu, A. (2001). Open-source software development and distributed innovation. *Oxford review of economic policy*, *17*(2), 248-264.

Lakhani, K, & Wolf, R. (2005). Why Hackers do What They do: Understanding Motivation and Effort in Free/Open Source Software Projects. In Feller, J., Fitzgerald, B., Hissam, S., and Lakhani, K. (Eds), Perspectives on Free and Open Source Software. MIT Press, Cambridge.

Lerner, J, & Tirole, J. (2002). Some simple economics of open source. The Journal of Industrial Economics 50, no. 2: 197-234.

Lin, Y. K., & Maruping, L. M. (2022). Open source collaboration in digital entrepreneurship. *Organization Science*, *33*(1), 212-230.

Miric, M., Boudreau, K. J., & Jeppesen, L. B. (2019). Protecting their digital assets: The use of formal & informal appropriability strategies by App developers. *Research Policy*, *48*(8), 103738.

Miric, M., Ozalp, H., & Yilmaz, E. D. (2023). Trade-offs to using standardized tools: Innovation enablers or creativity constraints?. *Strategic Management Journal*, *44*(4), 909-942.

Nagaraj, A. (2022). The private impact of public data: Landsat satellite maps increased gold discoveries and encouraged entry. *Management Science*, *68*(1), 564-582.

Nagle, F. (2018). Learning by Contributing: Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods. Organization Science 29, no. 4 (July August 2018): 569–587.

Nagle, F. (2019a). The Impact of Technology Procurement Regulations on Social Value and National Competitiveness. Harvard Business School Working Paper No. 19-103.

Nagle, F. (2019b). Open Source Software and Firm Productivity. Management Science 65, no.3: 1191–1215.

Nagle, F, Greenstein, S., Roche, M., Wright, N. L., & Mehta, S. (2023). CoPilots(s): Generative AI at Microsoft and GitHub. Harvard Business School Case N9-624-010. November 16.

Pisano, G. (2006). "Profiting from innovation and the intellectual property revolution." Research Policy 35, no. 8: 1122-1130.

Rock, D. (2022). “Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence.” Working paper. https://www.danielianrock.com/_files/ugd/3f53fa_0cbd93d4cbcd4f58b0eb1f0e0276e2ab.pdf

Startup Genome Project. (2023). “Global Startup Ecosystem Ranking 2023.” https://startupgenome.com/article/global-startup-ecosystem-ranking-2023-top-30-plus-runners-up.

Tang, S., Wang, Z., & Tong, T. (2023). Knowledge Governance in Open Source Contributions: The Role of Gatekeepers. In *Academy of Management Proceedings* (Vol. 2023, No. 1, p. 17622). Briarcliff Manor, NY 10510: Academy of Management.

Von Hippel, E., & Krogh, G. (2003). "Open source software and the “private collective” innovation model: Issues for organization science." Organization Science 14, no. 2: 209-223.

Wen, W., Ceccagnoli, M., & Forman, C. (2015). Opening Up Intellectual Property Strategy: Implications for Open Source Software Entry by Start-up Firms. Management Science. 62:2668-2691.

Wright, N. L. (2023). Helping to grow on a shoestring? The role of strategy for global startups. Harvard Business School Working Paper No. 23-041. Available at SSRN 4335210.

Wright, N. L., Koning, R., & Khanna, T. (2023). Judging foreign startups. Strategic Management Journal, 44(9), 2195-2225.

Wright, N. L., Nagle, F., & Greenstein, S. (2023). Open source software and global entrepreneurship. Research Policy, 52(9), 104846.

Yu, S. (2020). How do accelerators impact the performance of high-technology ventures? Management Science, 66(2), 530–552.