Outsourcing IT and Technological Differentiation: Evidence from Digital Startups

Stephen Michael Impink
New York University, Stern School of Business
October 28, 2022

Abstract: Does outsourcing IT impact a startup's ability to differentiate itself and grow? With the advent of cloud services, firms are licensing IT instead of developing IT in-house. Despite this growing trend, we know little about how early-stage resource acquisition decision affects technology adoption, product differentiation, and longer-term performance. When firms outsource their IT, they develop a supplier relationship with a cloud services provider and receive valuable resources related to their cloud provider's platform. However, cloud suppliers control which resources they create and share, which technologies they suggest, and how well technologies fit with their platform, potentially impacting product differentiation. Using panel data on app-developing startups, I find that startups outsourcing to cloud platforms adopt more technologies. When outsourcing, development tools that programmers use to code an app become more similar to other startups, saving time by avoiding compatibility issues. On the other hand, data analytics capabilities become more diverse, enabling startups to cultivate more robust data, which aids in product differentiation and startup growth.

I thank my committee (Robert Seamans, Deepak Hedge, Raffaella Sadun, Melissa Schilling, and Belen Villalonga), JP Eggers, Sinziana Dorobantu, Gino Cattani, Shane Greenstein, Geoffrey Borchhardt, Melody Chang, Manav Raj, Katia Meggiorin, Jorge Guzman, Bo Cowgill, Nataliya Wright, Rem Koning, Mike Ewens, Josh Lerner, Kristina McElheran, Charles Becker, Hyunjin Kim, Daniel Rock, Nan Jia, Toby Stuart, and conference participants at ECDC, MAD, and CCC for helpful feedback; Microsoft Corporation (Sue Glueck, Jacques Lawarree), Boston University TPRI (James Bessen), NYU Center for Global Economy and Business, and NYU Fubon Center for funding data collection; AIdentified (Michael Els) for providing data on startup founders; and Harvard Extension School for providing a helpful RA (Samuel Lee).

1. Introduction

Entrepreneurs develop innovations that fuel economic progress (Gans et al., 2019; Guzman and Stern, 2020; Kerr et al., 2014). Yet digitization and the use of big data raise numerous unanswered questions about the nature of entrepreneurial innovation (Greenstein et al., 2013; Lerner and Nanda, 2020). To survive and scale, young startups must acquire the Information Technology (IT) assets necessary to develop their products (Bessen et al., 2018, 2022; DeStefano et al., 2020; Jin and McElheran, 2019). Outsourcing IT asset development to cloud providers has become increasingly common, particularly for new firms that need access to IT quickly.[1] Despite this growing trend, we know little about how this early-stage resource acquisition decision affects technology adoption, differentiation, and longer-term performance. The need for IT forces firms to determine whether they "make" or "buy" these resources early in their existence before developing their products (Lacity et al., 2009; Schneider and Sunyaev, 2016; Susarla et al., 2009), potentially creating a tradeoff between efficiencies gained from outsourcing and the ability to differentiate digital products.

Outsourcing has implications for firms' organizational structure, partnerships, and ability to control production (Coase, 1937; Williamson, 1979, 1998), which may influence the technologies used as inputs to develop resulting products. When firms make IT assets in-house, they hire specialized technical labor, purchase computing hardware (e.g., servers, mainframes) from many manufacturers and distributors, acquire physical space to house their IT infrastructure, and sign long-term IT services agreements. These activities increase their initial capital expenditure and expand their firm's structure horizontally. Alternatively, when firms outsource (i.e., the "buy" scenario), they lease subscription-based cloud IT services from a single technology firm, adapt the base IT platform through co-invention, and develop their products on that infrastructure, which enables a narrow organizational structure focused on a single product. Cloud services have become more sophisticated and secure, making them increasingly challenging to

[1] The use of cloud services has become so prolific that Gartner (2018) suggests firms not using cloud services will soon be as rare as firms not using the internet. Acemoglu et al. (2022) provide census data from 2019 (Figure 4 in their paper; https://www2.census.gov/ces/wp/2022/CES-WP-22-12.pdf) showing that younger, smaller firms are more likely to adopt cloud services.

replicate and raising the cost of developing comparable IT capabilities internally.[2] The initial capital expenditure of internal IT development may be too high for cash-strapped startups and their investors to bear, especially compared to highly discounted introductory offers from cloud providers that share technological resources that lower development costs.[3] The decision to either internally develop or externally license IT assets was once a strategically important source of differentiation. However, now, due to the benefits of quickly and cheaply accessing high-quality cloud IT services, it is difficult for digital-first startups to rationalize internal IT development, even if that means forgoing aspects of technological differentiation.

My research question asks whether and how outsourcing IT development impacts startup differentiation and growth. On the one hand, since capital requirements are lower, VC investors can provide more startups with enough money to start product development (i.e., "spray and pray" investing: Ewens et al., 2017). Possibly more startups are "experimenting" with their business models and quickly testing their product ideas (Kerr et al., 2014; Koning et al., 2022). On the other hand, we have little insight into how these changes affect the nature of resulting innovations (Lerner and Nanda, 2020; Ewens et al., 2017). In fact, despite potential increases in aggregate production at the industry level, there are many reasons why one could expect outsourcing IT development to a few large technology firms to constrain startups from adopting more diverse technological inputs and producing more differentiated products than startups developing their IT internally. First, outsourcing constrains startups' ability to customize aspects of their IT to fit their specific product development needs.[4] Second, outsourcing startups use services from a single cloud supplier instead of many suppliers (e.g., distributors, original equipment manufacturers, software

[2]Similarly, Bloom and Pierri (2018) discuss the increased pace of cloud adoption for smaller, newer firms in an HBR article (https://hbr.org/2018/08/research-cloud-computing-is-helping-smaller-newer-firms-compete). I depict cloud platform adoption rates for startups in my sample that are three or more years old in Appendix Figure C.1.
[3]When outsourcing, startups incur "co-invention" costs when adapting the IT infrastructure and their processes. Bresnahan et al. (1996) describe co-invention as users adapting the initial invention to make it more economically valuable for their needs amid the shift from mainframes to client/server computing in the 1990s. In the case of internal development, a startup would be "inventing" on its own. Moreover, there are costs in determining the compatibility of technologies with one other and with the underlying IT platform.
[4]More specifically, using a cloud platform can constrain startups' abilities to customize aspects of runtime, middleware, virtualization, and networking capabilities.

vendors), limiting access to a broader array of resources and expertise. Third, outsourcing creates a supplier relationship that provides all startups with the same platform-related resources. Startups use these resources to co-invent on the platform to meet their production needs but potentially make similar adaptations to other startups using the same shared resources. Fourth, product development requires platform-specific investment in complementary assets (i.e., asset specificity; sunk cost), increasing the fit of the startup's technologies with one platform and limiting cross-compatibility with other platforms. Lastly, investors, stretched thin from funding more startups, have less bandwidth to tailor their expertise and guidance to each startup's specific needs.[5]

This paper examines unique panel data on technology adoption for ~3,400 high-tech, app-producing startups with a web-based or mobile (i.e., Android, iOS) application (app) to examine the impact of outsourcing IT development on startup differentiation. The main analysis relies on an OLS difference-in-differences research design with firm-level fixed effects to control for time-invariant aspects of firms and year-level fixed effects to control for variation correlated with time. My main analyses include Coarsened exact matching (CEM) based on observable firm characteristics to ensure that startups using cloud platform services are demographically comparable to startups not using these services.[6] Next, I examine two mechanisms related to startup differentiation: a stronger customer-supplier relationship and an increased need for technological fit. To further support my analysis, I use Heckman's (1979) selection approach, an instrumental variable approach based on Google's late 2015 open source release of TensorFlow[7], and a double machine learning model (Chernozhukov et al., 2018) based on a random forest algorithm. These methods adjust estimates for potential explainable variation in the control and treatment groups (e.g., CEM, Heckman), omitted variables (e.g., IV, ML) and reverse causality (e.g., IV).

[5]Ewens et al. (2018) note that VCs are less likely to join startups' boards.
[6]CEM models match on firm age, location (region), size, and industry.
[7]TensorFlow, an AI development framework enabling firms to train neural network algorithms, is a quasi-exogenous shock to AI production, reducing the costs associated with AI development and increasing the value of complementary AI-related labor (Rock, 2019). https://github.com/tensorflow

To gauge changes in startup differentiation, I examine the adoption of two types of technologies: product development (i.e., app development tools) and data analytics technologies (i.e., data analytic capabilities). These analyses show that startups adopt more technologies after outsourcing IT development (i.e., adding cloud platform services). However, outsourcing affects the breadth of the bundle of technologies that startups adopt, depending on the technology's type, interdependency with other technologies, and fit with the IT platform's underlying technologies for the digital app to work effectively.

First, product development tools are developer operations frameworks, such as *angular*, *next*, and *django*, which enable programmers to build, test, organize, and update code necessary in app development. Bundles of these development technologies become more similar to those used by potential rivals – other app-producing startups in my sample – as apps will only work effectively if these technologies are compatible with one another and fit with the IT platform. As an example based on traditional manufacturing processes, product development technologies are like individual machines used in an assembly line, and each technology provides some interrelated function in production. The development technology bundle represents all the interdependent "machines" a firm uses to develop its product. These machines must fit with the assembly line process and other machines used in production for products to work effectively.

Second, firms cultivate data analytics capabilities to collect, analyze, and reconfigure data, and these more unique and robust data resources aid in product design improvements and decision-making. Data resources are valuable to digital firms, particularly AI-producing startups needing data to train algorithms. For example, data analytics technologies like *matomo* and *parse* enable startups to analyze their website traffic to determine user location and demographics. Others like *improvely* and *optimizely* enable a/b testing to create experiment settings that provide valuable data as outputs. Unlike product development technologies, these analytics technologies are more modular, rendering the fit with the IT platform and compatibility of these technologies with each other less important to producing needed data resources. Moreover, they do not necessarily interact with the app or impede its functionality.[8]

[8]Technologies supporting data capabilities are connected to the startup's web domain, powered by the cloud, and used in ways that may or may not related to the product app. For instance, these technologies could analyze telemetry data

Investors reward startups for adopting more similar development tools and more distinct data analytics capabilities than others, supporting that startups may benefit from streamlining their app development "assembly line" to focus on accessing the data resources needed to enable differentiation. A secondary analysis of startup description changes before and after outsourcing to a cloud platform supports these findings, suggesting that end products become more differentiated after outsourcing to a cloud platform, driven by more diverse data analytics capabilities. Altogether, this paints a picture of a high-performing digital startup as standardizing its development tools to fit with its other technologies and supplier's platform and then cultivating more diverse data analytics capabilities to differentiate its products with richer data resources.

My paper makes several contributions. First, I contribute to a developing research agenda in high-tech entrepreneurship (Bessen et al., 2018, 2021; Dushnitsky and Stroube, 2021; Ewens et al., 2018; Lerner and Nanda, 2020) and, more broadly, digitization (Cowgill and Tucker, 2019; Furman and Seamans, 2019; Goldfarb and Tucker; 2019; Tucker, 2019) by showing how technology adoption changes when using cloud platform services. Furthermore, I show meaningful relationships between changes in technology adoption and measures of product differentiation and venture performance.[9]Next, I contribute to the literature on resource sharing and the nature of technological innovation (Baumol, 2001; Boudreau, 2012; Gulati, 1995; Katila et al., 2022. Mowery et al., 1998; Stuart, 2000; Stuart et al., 1999) by examining how suppliers' shared resources and the need for technological fit with the supplier's platform influence technology adoption and differentiation. Lastly, I use the context of digital entrepreneurship to contribute to the literature on transaction cost economics (Nagle et al., 2020; Tadelis and Williamson, 2012; Williamson, 1979, 1998) by showing that startups using cloud suppliers with a higher market (i.e., more market power) adopt relatively more similar bundles of platform-related technologies than startups using lower market share cloud providers.

---

from their app's usage. Alternatively, they could analyze their and competitors' website traffic or general market trends.

[9]Performance measures include any funding, VC funding, deal size, and duration of web domain visits.

This paper proceeds as follows. Section 2 provides an overview of related theories on outsourcing and differentiation, focusing on how outsourcing reduces a startup's control over production, affecting their partnerships, initial resources, and organizational structure. Section 3 introduces the context of high-tech startups developing apps and provides insight into two mechanisms, the strength of the customer-supplier relationship and the need for technological fit, influencing the breadth of technology adoption as a proxy for startup differentiation. The following two sections describe my sample and data, technology adoption measures (Section 4), and the research design (Section 5). Then, section 6 reports these findings, robustness analyses related to these findings, and econometric approaches used to support that my results are consistent with a causal interpretation. Section 7 examines the link between adopting more diverse data analytics technology bundles and increased product differentiation and the connection between technology adoption and firm performance. The last section (Section 8) discusses these findings and concludes.

2. The impact of outsourcing on differentiation

The decision to outsource, and expand the boundaries of the firm, relates to a broad literature on transaction cost economics that extends the neoclassical economic perspective by adapting contracting theory to address optimal organizational structures and governance models (Coase, 1937; Williamson, 1979, 1998). Coase (1937) provides the initial argument that firms will outsource when the market offers lower costs than internal production. However, drawing on learnings of coordination across multidivisional firms[10], Klein (1980) and Williamson (1979) highlight the limitations of contracts, explaining that inefficiencies arise from splitting surplus rents ex-post and that boundedly rational suppliers are potentially opportunistic. Moreover, Aghion and Tirole (1993) expand this literature by discussing optimal ownership arrangements amongst firms, suppliers, and third-party investors and the appropriability of resulting innovations. Supplier opportunism in outsourcing arrangements leads to potential inefficiencies, such as switching costs

[10]Williamson (1975) discusses the advent of the multidivisional firm, and issues of sharing and collaboration across internal departments.

(Monteverde and Teece, 1982), hostage due to credible threats (Williamson, 1983), corresponding asset-specific investments (i.e., sunk costs; Riordan and Williamson, 1985), and uncertainty associated with contract terms and the frequency of exchange (Grossman and Hart, 1986).

The literature on outsourcing pulls from research examining the inefficiencies associated with contracting and co-financing, interweaving an understanding of these inefficiencies into more nuanced discussions of the impact of outsourcing on supplier power, partnerships, and organizational structure.

First, outsourcing increases a supplier's power over which technologies customers adopt now (Rysman and Simcoe, 2008; Wen et al., 2022) and in the future (Greenstein, 1993). Firms often adopt technologies compatible with their supplier's platform or technologies (Simcoe, 2012). They make complementary asset-specific investments[11] and engage in co-invention activities with their supplier to increase their fit with the supplier's technologies, aligning with their supplier's 'ad hoc' standards (Bresnahan et al., 1996). These investments and the availability of complementary resources may lead firms down a particular technological path (Arthur, 1994; Pfeffer and Salancik, 2003; Schilling, 1999), locking firms to the platform (e.g., hostage; Williamson, 1983) and the first ability to compete in an industry (Rivkin, 2000; Siggelkow, 2001). Switching costs increase over time as these firms compound sunk investments and progress along this technological path, suggesting that outsourcing will constrain differentiation in the long run.

Second, the decision to outsource creates a supplier relationship that affects future partnerships (Combs and Ketchen, 1999; Madhok and Tallman, 1998; Young-Ybarra and Wiersema, 1999), which in turn influences the firm's technology adoption (Gulati, 1995; Katila et al., 2022; Mowery et al., 1998; Stuart, 2000, et al., 1999) and innovations (Ahuja, 2000; Baumol, 2001; Hagedoorn and Schakenraad, 1990). Suppliers determine which resources they develop and whether they share those resources with their broader ecosystem. If they share resources, they decide which resources will be shared and which partners

[11]Investments are unrecoverable (i.e., sunk) and asset-specific, fitting only with the current platform. Firms would likely have to make similar investments in complementary assets and increased compatibility again if they changed platforms, raising switching costs and "locking" the startup to their providers (Monteverde and Teece, 1982).

will receive them. Moreover, suppliers determine the compatibility of their shared technologies with other technologies, further constraining technology adoption.

However, despite this undue influence on future technologies, resource-strapped entrepreneurial ventures may cease to exist without access to these early resources (Stuart et al., 1999). These "access relationships" (Stuart, 2000), including customer-supplier relationships, technology exchange agreements, and one-directional technology flows[12], enable firms to access needed resources to develop their products. Moreover, these interfirm relationships create synergies that overwhelm the benefits of internal development in many cases (Madhok and Tallman, 1998; Silverman, 1999) and mitigate hold-up issues amongst network participants when information is dispersed widely across firms in an industry (Powell et al., 1996). These relationships complement the codification of the transaction terms, strengthening governance mechanisms (Poppo and Zenger, 2002). Moreover, the benefits from collaboration provide insight into why firms may depart from the resources-based perspective that internal development of rent-generating resources enables firms to collect excess returns from imperfections in strategic factor markets (Barney, 1986) and reduces the threat of imitation (Montgomery, 1994; Peteraf, 1993).

Third, firms' outsourcing decisions impact their organizational structure. Firms choosing to make a resource must vertically integrate the inputs of that resource's production, which diversifies the firm to focus on developing an additional product (Argyres and Zenger, 2012; Brynjolfsson et al., 1994; Teece, 1982). Having to focus on an additional product line can impede progress for two reasons. First, in digital industries, having multiple production lines may prevent benefits from scaling. This ability to scale digital resources rapidly (Fazli et al., 2018), coupled with the firm being more vertically narrow scope, enabling a greater focus on a single product, may create a situation where the potential gains of being able to scale a single product quickly[13] outweigh the transaction cost inefficiencies from outsourcing (Cachon and Harker, 2002; Giustiziero, 2021). Second, it may be difficult for firms to develop processes and products

[12]Hagedoorn and Schakenraad (1990, p.5) provide an exhaustive list, including direct investment, joint research corporations, joint ventures, and joint R&D agreements.
[13]Research suggests that productivity gains (Aral et al., 2012) and IT discounts (Benzina, 2019) are associated with scaling and increased firm size.

simultaneously (Henderson and Clark, 1990). Moreover, aspects of firm structure may impair the firm's ability to remain flexible and adjust levels of product and process R&D. This inflexibility could inhibit a firm's ability to benefit from sequencing process and product-related research activities in the short-run (Athey and Schmutzler, 1995). When nascent firms initially outsource process-related R&D, their structure would be more focused on product-related R&D. For digital firms, this focus on digital products would manifest in firms acquiring richer data resources or the means enabling enhanced data collection, recombination, and use.

Lastly, this literature also considers the technological environment's dynamism and the pace of technological change in examining the decision to outsource. Vertical integration is more effective when the likelihood of technological obsolescence is high (Balakrishnan and Wernerfelt, 1986). Outsourcing may be more beneficial in digital industries where technological development is fast-paced and quickly renders prior technologies obsolete (Aral et al., 2012; Cachon and Harker, 2002; Giustiziero, 2021). High-tech firms may benefit from outsourcing due to the rapid pace of technological change, forgoing upfront investment in the internal development of quickly deteriorating IT assets.

3. High-tech startups developing apps on cloud platforms

Though subscription-based cloud services are fairly new, since 2006, they have fundamentally changed how startups procure IT assets. This paper primarily focuses on cloud platform services (PaaS), a specific type of cloud service which enables startups to host technologies on their web domain and develop and test apps. These platforms are a "layered architecture of digital technology" (Yoo et al., 2010) with a governance model (Rochet and Tirole, 2003; Parker et al., 2017). Cloud providers also offer hardware infrastructure services (IaaS), leased cloud-based computers and servers, that are less consequential to differentiation than platform services but important in the context of IT spending, hardware development, and productivity.

High-tech startups outsourcing IT platform development may access difficult-to-develop resources from their suppliers, avoid investment in deteriorating technologies, and benefit from a more vertical organizational structure that enables quick scaling of their digital product. On the other hand, outsourcing may constrain differentiation when firms adopt technologies to fit with their suppliers' technologies, co-invent with the same shared resources as others, or alter their firms' structure to facilitate outsourcing similarly to others. Supplier power increases over time as startups make additional sunk, asset-specific investments to fit their cloud platform and as monopsony costs increase from cloud suppliers consolidation. Moreover, incentives to innovate dissipate as these platforms grow larger (Boudreau, 2012), suggesting that the emergence of several large cloud suppliers may further reduce differentiation. Still, others argue the opposite, that large platforms are multi-dimensional product spaces offering a nearly limitless possibilities for technology recombination (Caves, 2000; Parker et al., 2017; Zittrain, 2006)

Two mechanisms, the strength of the customer-supplier relationship and the need for technological fit, provide insight into how outsourcing impacts technology adoption and differentiation. These mechanisms are interrelated, influencing technological inputs in production directly and the allocation of time and programming labor among app development and data analytics activities indirectly.

Need for technological fit. Startups will adopt increasingly similar technologies when the fit amongst interdependent technologies or between the bundle of technologies and the underlying cloud platform's technologies is important for the outcome. In the case of product development technologies, this outcome is developing a working app. For data analytics technologies, the outcome is producing data resources that are unrelated to the app working effectively.

First, the need for technological fit amongst interdependent technologies impacts the breadth of technology adoption. The potential for technological incompatibility increases with the size of a firm's technology bundle, as it is increasingly difficult to combine many highly interdependent technologies in production (i.e., complexity catastrophe; Fleming and Sorenson, 2001). Firms using larger bundles of interdependent technologies incur increased coordination costs associated with managing fit, incentivizing them to use more standardized development tools with known compatibilities. Data analytics technologies do not face similar constraints as they are more modular (i.e., less interdependent). Second, a startup's initial fit with its development platform is strategically significant, as startups may be unable to incur the added cost of adapting less compatible technologies to fit the platform.[14] Moreover, the need for firm impacts which technologies are added to a startup's technology bundle. These adaptation and coordination costs are relatively higher for development technologies due to interdependencies, potentially requiring many technologies to be exchanged or adapted. Given these costs of maintaining fit, startups may benefit from entrenching themselves with a single provider's recommended development tools, following a particular technological trajectory.

Customer-supplier relationship. When digital startups are founded, they have few customers and lack relationships with other firms, However, they quickly establish a relationship with a technology firm supplying their cloud services. Larger technology firms have abundant IT and data resources. The largest cloud services providers – Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure – have startup-related corporate programs that share resources with their customers.[15] These corporate accelerator programs share generalized resources like traditional accelerators (Hochberg, 2016; Yu, 2020) yet also provide technical resources (i.e., software, compatibility documentation, expertise, and troubleshooting) related to their cloud platform. Unlike traditional accelerator programs, these corporate programs are often open to any viable startups. This relationship is a conduit for resources to pass from the supplier to their customers. Suppliers likely have a stronger relationship with startups that are potentially valuable customers, using more of their cloud service products (i.e., both platform and infrastructure services). This stronger relationship may unlock additional shared resources that are difficult for young startups to develop. Alternatively, receiving more resources may incentive adopting certain technologies

[14]Once startups develop their apps on their supplier's platform, moving their product to another platform or internalizing production is costly. There are fees for offloading data. Moreover, fit-enhancing investments in a specific platform do not necessarily transfer to another platform, and firms may have to replicate some of those investments on the new platform. Startups may even need to hire different programmers as coding requirements are not universal across platforms. It may be less costly for a firm to abandon its app on one cloud platform and start development from ground zero on a different platform than to move its app.

[15]To give a sense of the scale and breadth of these programs, Amazon AWS even has an accelerator specifically focused on providing resources to startups focused on space travel. https://www.geekwire.com/2021/amazon-web-services-launches-space-accelerator-final-frontier-startups/

that the supplier touts to be compatible with their services. Moreover, coordination costs may be higher increasing the technological constraints associated with coordinating fit with an additional cloud service.

When fit issues emerge, access to a cloud provider's shared technologies, processes, and technological compatibility guidance could save time for programmers. Programmers could adopt more standardized development tools and spend their saved time and resources cultivating data analytics capabilities. There is a fixed pie of available resources, and startups that more efficiently build a functional app have more resources to collect and analyze data resources that could provide them with a competitive advantage. Firms benefit from developing data analytics capabilities (Brynjolfsson et al., 2021; Provost and Fawcett, 2013, Tambe, 2014). These benefits may be even larger for firms with a digital app, enabling them to examine their app's usage data (Chatterji and Fabrizio, 2014), or for firms producing AI, enabling them to source, clean, and recombine training data (Furman and Seamans, 2019; Bessen et al., 2022).The potential of shifting resources from one important activity to another aligns with recent research findings that using standardized or low-code development tools can enable commercial success (Miric et al., 2021; Dushnitsky and Strobe, 2021), suggesting that using more standardized development tools does not necessarily limit the ability for firms to differentiate. In the case of digital app production, product differentiation stems from utilizing unique or more robust data that is challenging for rivals to replicate.[16]

4. Data

Before starting the quantitative analyses, I informally interviewed fifteen high-tech startup founders to understand the IT asset outsourcing decision better. Founders and early technical employees determine which cloud provider to adopt before developing their product. All the startups I interviewed used cloud platform services, and about half also used infrastructure services. [17] One startup initially developed IT in-

[16]That data then could be used to make business decisions (e.g., product design decisions, marketing decisions, and customer and partner acquisitions decisions) and, in the case of AI-developing startups, to train algorithms.
[17] A few mentioned adding a second cloud provider to access more free cloud credits for a tangential project or from "blob" storage, which lets developers store unstructured data on the cloud. This data can be accessed from anywhere in the world and can include audio, video, and text. Blobs are grouped into "containers" that are tied to user accounts.

house (from 2013 to 2017) to produce a marketing app, citing security issues of working with sensitive data and competitively significant algorithms in the cloud. Several founders mentioned that their end-customers' industry influences their decision on which cloud supplier to adopt. One healthcare startup felt pressure to develop on the Microsoft Azure platform because Azure offered HIPAA-compliant cloud services earlier, enabling Azure to develop an early foothold in that vertical. Some startups responded that they joined a particular platform to access free software, services, cloud credits, or corporate accelerator programs. However, most startups I contacted revealed that highly discounted offers facilitated through their accelerator or incubator programs influenced their decision to join a particular cloud platform.

The initial cloud decision impacts the startups' future technological compatibility and complementarity as it is challenging to switch platforms. For instance, several startups reported being unable to change cloud providers because they would have to "rebuild their entire product" on the new platform.[18] Founders discussed programming labor as a constraint: "it is hard to find (replace) a good programmer." Some mentioned having to hire a programmer with different coding preferences if they switched to Azure, a platform requiring more extensive knowledge of C# than GCP or AWS, arguing that it would be costly to find other programmers. Others suggested that they cannot backstep and start over if they already have a functioning app, which is often a milestone for investors. Developing a functioning app and landing a few early customers are clear initial goals.

4.A. Firm demographics

To examine how outsourcing IT affects startup technology adoption and differentiation, I establish a sample of startups with a digital app from multiple data providers. I use data from Crunchbase[19] and Pitchbook[20] to compile a list of active, high-tech startups in IT or software-related industries. Then I use data from

[18]Only one startup changed to a new primary cloud provider after product development; however, the change coincided with the startup founder hiring a prior executive from the new cloud provider.
[19]Crunchbase provides data about startups and sources its data in four ways: venture programs, machine learning, an in-house data team, and the Crunchbase community.
[20]Pitchbook is a software-as-a-service company that delivers data, research, and technology covering private capital markets, including venture capital, private equity, and M&A transactions.

Apptopia[21] and startup descriptions to confirm these startups have a mobile or website-based app and an active web domain. To capture higher-growth potential startups, I exclude firms that are older (i.e., >10 years old), larger (i.e., more than 500 employees), or located in China.[22] These criteria yield a sample of 3,434 startups that develop an app as their product and were founded between 2012 and 2021. Most startups in the sample operate in more developed economies (~90%) and are primarily located in the Americas (56% SD 0.05) or Europe (27% SD 0.04). The average startup in the sample is 4 (SD 2.4) years old and has 45 (SD 63) employees. One-fifth of these startups target customers in the financial services and healthcare industries. Many describe themselves as developing commercial AI products (38% SD 0.49) or using machine learning in production (9% SD 0.29) (Table 1.A. under the heading *Demographics*).

4.B. Founder measures

I build measures on startup founders from three data sources: Mantheos[23], AIdentified[24], and manually collected data from public LinkedIn[25] profiles. In our sample, 12% (SD 0.33) of startups have founders at least one founder with prior IT experience, including 5% (SD 0.22) with prior hardware development experience and 7% (SD 0.25) with prior Big Tech experience (Amazon: 1%, Google: 3%, Microsoft: 3%). On average, 44% (SD 0.50) of startups have a founder with a technical[26] undergraduate or graduate major; 21% (SD 0.41) have an advanced degree (i.e., master's or doctorate) in a field other than business administration; 24% (SD 0.43) have a master's degree in business administration (MBA). From text analysis of founder's names, I determine that 13% (SD 0.33; 95% CI) of startups have a female founder or

[21] Apptopia' s data intelligence platform enables brands to analyze critical competitive signals and gain insights across mobile applications and connected devices.
[22] As a potential limitation, third-party data sources may not accurately represent the underlying population. These English-language data sources underrepresent the number of startups in China or from certain emerging countries where the English language is less commonly used. Additionally, founders from these countries will be underrepresented on LinkedIn. Yet, even in English-language developed markets, a small number of very young startups may try to stay under the radar. Results from this sample will be more valid for developed markets, where data-representativeness issues are less of a concern.
[23] Mantheos is a business intelligence company providing accurate, clean, and structured data aggregation on demand. They are currently out of business (4/25/2022) after being sued by LinkedIn.
[24] AIdentified reveals the best paths for sales teams, account executives, and brands to connect to hyper-targeted, qualified prospects using predictive analytics and next-level AI-based relationship intelligence mapping.
[25] LinkedIn is an employment-oriented online service that operates via websites and mobile apps.
[26] Technical degrees include math, physics, computer science, statistics, and data science.

CEO.[27] These estimates reflect the low participation of females in the population of high-tech entrepreneurship (Table 1.A. under the heading *Founders*).

4.C. Cloud services provider measures

I collect firm-level cloud services and technology adoption data across time (2012-2021) for these startups from BuiltWith.[28] This provider offers information on cloud services (i.e., PaaS, IaaS, storage) and technologies connected to the startup's web domain by making "HTTP requests" and analyzing website code to determine which "back end" technologies startups adopt.[29] Since each of these startups has a digital app as its product, information on the adoption of domain-based technologies provides insight into web and mobile app development. Most startups in my sample develop their app on a cloud platform (85%) and license cloud hardware infrastructure services (79%) (Table 1.B.).

I identify ten technology firms that license cloud services. The largest suppliers are Amazon, Google, and Microsoft (Big Tech CSPs, 78%). The other seven suppliers offer more niche (e.g., fintech digital currency mining) or less expensive technology services: Alibaba, Digital Ocean, IBM, OVH, Oracle, and Linode (Other CSPs, 7%). These cloud providers also offer hardware infrastructure services that provide high-power virtual machines with processors (GPUs) and solid-state hard drives (SSD), which are particularly valuable in AI development.[30] The remaining startups (15%) do not have cloud-based platforms connected to their web domain (Table 1.C.)[31]

[27]I used the "gender" library in R, SSA method, focused on English language birth names common in the 1980s.
[28]BuiltWith returns all the technologies connected to a web domain, covering more than 59k technologies across analytics, advertising, hosting, frameworks, CMS, and more.
[29]Prior research in strategy uses similar data from BuiltWith: Koning et al., 2022 (A/B testing technologies); Dushnitsky and Stroube, 2021 (connection with Shopify technologies). I connect to BuiltWith's API to download this data on each startup's web domain; all startups included have an active website. More information on BW: https://techcrunch.com/2012/02/16/ builtwith-reveals-the-tech-used-by-the-130-million-web-sites-that-matter-most
[30]E.g., Amazon Elastic Compute (EC2), Google Compute Engine, Cloud AutoML, and Azure Machine Learning
[31]Despite the richness of this technology adoption data, there are several limitations. First, I assume that startups with a supplier relationship participate in programs that share resources. Next, cloud suppliers with startup programs, a conduit for sharing resources, share resources with greater than suppliers without programs. Third, I cannot pinpoint which technologies startups added as a direct result of sharing. Fourth, and related to the prior point, I cannot determine which other non-technology resources, like technical and business expertise, are shared formally through programs or informally through increased network connectedness.

4.D. Technology adoption measures

Though I have data on all domain-connected technologies for each startup, this paper does not focus on "front end" technologies (e.g., website hosting, fonts, e-commerce, payment, etc.) or organizational technologies (e.g., customer relationship management, sales tools, workforce management, email hosting, etc.). Instead, it focuses on product development technologies (e.g., content management systems, content delivery networks, frameworks, and security) and analytics technologies (e.g., data analytics, collection, and telemetry) that are essential for high-tech startups that develop apps and run their business.[32] I share summary statistics on these measures in Table 1.D. under the heading *Technology Adoption*.

Technology Bundle Size. First, I calculate the firm-year technology bundle size, a count of technologies connected to the web domain, using a similar approach to Berman and Israeli (2022). Startups use an average of 50 (SD 27) technologies, ranging from 1 to 252. Startups using a cloud platform have larger technology bundles (53 SD 27) than startups not using a cloud platform (36 SD 19). On average, firms use 8 (SD 4) product development technologies and 6 (SD 6) analytics technologies.[33]

Technology Bundle Dissimilarity. Next, I calculate firm-year technology dissimilarity for product development and data analytics technologies based on pairwise cosine similarity each year to address my research question, focused on relative changes to the breadth of technology adoption of app-producing firms in my sample as a proxy of startup differentiation. I calculate angular distance as a technology dissimilarity measure based on prior strategy and economics research (Seamans and Zhu, 2014; Sweeting, 2010; Wang and Shaver, 2014).[34] In my case, the coordinates are firm-year vectors of technologies, taking the value 1

[32] I list all the technologies included in the analyses and their descriptions in Appendix D, and I provide more information on cleaning the BuiltWith data in Appendix Note A.1. Even though I make all attempts to clean and organize technologies based on the provided categories and descriptions of technologies, there is the chance that startups onboard a technology and not use it. Some of these technologies have a licensing cost and are likely to be off-boarded quickly; however, free technologies, especially those not requiring substantial space, could linger. Alternatively, they could use product development technologies in a way that is unrelated to application development.
[33] Additionally, I calculate the number of Big Tech technologies (13 SD 6), premium/subscription technologies as defined by BuiltWith (3 SD 4) and open source technologies (1 SD 1) as defined by analysis of their descriptions. (Table 1.D.)
[34] An example of this measure comes from astronomy, where the angular distance is the angle between two "sightlines" of two far-away objects.

when both startups do not use the technology and 0 when both startups use the technology, configuring a positive semidefinite matrix. The resulting dissimilarity measure will be bound by [0,1], taking the value of 1, the maximum distance, when there is no overlap of any technologies in a given year, and the value of 0 when there is perfect overlap.

$$bundle\_dissimilarity_{ij,t} = \left(cos^{-1}\left(\frac{V_{it} * V_{jt}}{\parallel V_{it} \parallel\parallel V_{jt} \parallel}\right)\right)\Big/\frac{\pi}{2} \quad (1)$$

where, $bundle\_dissimilarity_{ij,t}$ refers to the pairwise angular distance of the focal firm and rival[35], i refers to the focal firm, j refers to the rival, t takes on the value of the year, and $(V_{it} * V_{jt})$ is the firm-level pairwise dot product, normalized by the length of each vector ($\parallel V_{it} \parallel\parallel V_{jt} \parallel$), so the technology bundle size will not impact measurement.

For the main analysis, I calculate the mean of the angular distance for each focal startup with respect to other app-developing startups in the comparison panel (e.g., all startups, startups using a CSP, startups using Big Tech CSP, startups using Amazon AWS, etc.) from the disaggregated data. For instance, when I calculate this mean measure for focal startups in the Amazon AWS panel, it only includes pairwise matches of firms that use the Amazon AWS platform. The average bundle dissimilarity is 0.78 (SD .07) for product development technologies and 0.66 (SD 0.12) for analytics technologies.[36]

4.E. Funding measures

I create and use indicator variables for whether startups received any funding (71% SD 0.45; including seed/angel funding), follow-on funding (52% SD 0.5; at least two rounds of funding), venture capital funding (61% SD 0.49), or higher reputation venture capital funding (9% SD 0.28). Next, I create a firm-year measure of deal size (5.1 SD 7.1, log) and an aggregate measure of total funding (11.9 SD 6.2, log).[37]

[35]Rival firms are defined as any other high-tech app-producing startup in my sample.
[36]I provide additional details on measure construction in Appendix Note A.2., kernel density estimates in Appendix Figure C.2, and summary statistics in Table 1.D. under the heading *Technology Adoption*. I depict the relationship between product development and analytics technology bundle dissimilarity before and after outsourcing to a cloud platform services in Figure C.3.
[37]Also, I create indicator variables for participating in an accelerator (19% SD 0.39) or having direct funding from a Big Tech firm (3% SD 0.26).

As another performance measure, I use data on website visit duration (3.1 SD 3.3, log minutes) from SimilarWeb[38] and an indicator variable for if startups have a patent (4% SD 0.18) from IPQwerty.[39] Though many young high-tech startups do not patent, this measure captures some aspects of proprietary innovation. Only 5% (SD 0.21) of startups have closed, and 6% (SD 0.24) of startups have been acquired. Startups in my sample are young, so we do not yet have a clear indication of which startups will survive, opening the door for future research in the years to come. I provide additional descriptive statistics in Table 1.D. under the heading *Performance*, and I report correlations of these measures with firm demographics measures (Appendix Table A.3.), technology measures (Table A.4.), and performance measures (Table A.5.)

5. Research Design

I construct a relatively homogenous sample of active startups less than ten years old with fewer than 500 employees and an existing app. Then I use Coarsened Exact Matching (CEM; Iacus et al., 2012) based on observables: age, employment size, region, industry vertical (healthcare, financial services) across comparison groups to weight regressions. This matching procedure is consistent with the argument that startups using a cloud platform are not observably different from those not using a cloud platform. I provide summary statistics on sample means before and after matching on the right side of Tables 1.A. and 1.C. and depict the standardized mean differences between the weighted and unweighted groups in Appendix Note A.6.

For the main specification, I use an OLS difference-in-differences approach with two-way fixed effects to model the impact of outsourcing IT development to a cloud platform on technology adoption. This approach assumes that the treatment and control groups have parallel trends even if no firms are treated (Abadie 2005). I compare the trends for the control and treatment groups in Appendix Figure C.4.

[38]SimilarWeb is a digital intelligence provider for enterprise and small to mid-sized business customers. The platform provides web analytics services and offers its users information on their clients' and competitors' web traffic and performance.
[39]IPQwerty applies a series of contextual clues to help differentiate between similar company names, then separates each into the correct IP profile.

Development technology bundle size and dissimilarity show parallel trends across all years, as does analytics bundle size. However, analytics differentiation shows parallel trends only up until 2019.

$$y_{it} = \beta_1 outsource_{it} + \beta_2 yearFE_t + \beta_3 firmFE_i + \varepsilon_{it} \quad (2)$$

where, y_{it} refers to the dependent variables: technology bundles size and dissimilarity measures; $outsource_{it}$ refers to an indicator variable that takes the value 1 if a startup uses a cloud platform and 0 otherwise; $yearFE_t$ refers to the year-level fixed effect; $firmFE_i$ refers to the firm-level fixed effect. I cluster standard errors at the firm level. In these models, I match and weight regression according to the treatment, adopting a cloud platform versus not adopting a cloud platform, which drops 21 unmatched firms. To overcome potential estimation issues from a staggered difference-in-differences model with two-way fixed effects, I estimate aggregated pre- and post-treatment estimates, average treatment effects, local average treatment effects, and heterogeneity-robust instantaneous treatment effects for robustness.

Next, I use the following OLS specification to compare different platforms, adjusting my CEM matching approach to compare across treatment groups (i.e., outsourcing to Platform X vs. Platform Y). For example, when comparing startups outsourcing to the AWS platform versus those outsourcing to a different platform, I would match to ensure that startups using AWS are similar to those not using AWS.

$$y_{it} = \beta_1(cloud\_providerX_{it} = 1) + \beta_2(cloud\_providerY_{it} = 1) + \beta_3 yearFE_{it} + \beta_4 firmFE_{it} + \varepsilon_{it} \quad (3)$$

where, y_{it} refers to the dependent variable: technology bundle size or dissimilarity measures, $cloud\_providerX_{it}$ refers is an indicator variable that takes the value 1 if the startup outsources to Platform X and 0 otherwise, $cloud\_providerY_{it}$ refers is an indicator variable that takes the value 1 if the startup uses cloud services from Platform Y and 0 otherwise.

6. Impact of outsourcing to a cloud platform on technology adoption

6.A. Main findings

I examine how technology adoption changes when startups outsource to cloud platform services, using panel data product development and data analytics technology bundles startups. These technologies are direct inputs in digital production.

Larger technology bundles. I find that product development technology bundles (Table 2, model (1): +0.43 SD 0.02) and analytics technology bundles (model (8): +0.52 SD 0.02) become larger after outsourcing. OLS estimates increase when adding firm-level fixed effects (models (2) and (9)) and decrease when adding year-level fixed effects (models (3) and (10)).[40] Adding both firm and year-level fixed effects reduces estimates of technology bundle size (model (4): +0.30 SD 0.02; model (11): +0.24 SD 0.02), but results remain positive and significant. These estimates remain similar when dropping 21 unmatched startups and weighting regressions based on the matching procedure, controlling for potential observable differences between the control and treatment groups (models (5) and (12)). Lastly, the ratio of development tools to all other technologies slightly increases after outsourcing to a cloud platform (model (7): +0.012 SD 0.003).[41]

More standardized tools. Technology bundle size, the total number of technologies that startups use, provides limited insight into startup differentiation. Regardless of if startups use more technologies, their apps will not work effectively unless these technologies fit with their production needs, other interdependent technologies, and their IT platform's underlying technologies. As anticipated, product development technology bundles, collectively the tools that programmers use to build digital apps, become less dissimilar (i.e., more similar; model (1): -0.041 SD 0.002) to other startups in the sample when they outsource to cloud platform services. These technologies have similar fit constraints and receive many of

[40]When firm-level fixed effects are added, 290 single observations/firms are dropped. These firms have no pre/post outsourcing variation.
[41]Additionally, in Table B.1., I report additional results for Big Tech (i.e., developed and licensed by Amazon, Google, or Microsoft), paid/subscription, and open source technology bundle size.

the same resources from their suppliers, pushing them toward a similar technological path. Building from this base model, I add firm-level fixed effects (model (2): -0.089 SD 0.002), year-level fixed effects (model (3): -0.025 SD 0.002), and both firm and year-level fixed effects (model (4): -0.028 SD 0.002). Lastly, results remain the same when I weight regressions and drop 21 unmatched firms in addition to including fixed effects (model (5) -0.028 SD 0.002; coefficients depicted in Figure 1).

Next, I examine two mechanisms, the need for technological fit and the strength of the customer-supplier relationship, to provide evidence consistent with a causal relationship. Product development technologies are often interdependent, requiring certain base technologies for the added technology to work appropriately. Startups using larger bundles of development technologies incur increased costs of coordinating among a higher number of interdependent technologies. Moreover, the larger bundle of technologies must still fit with the cloud platform's underlying technologies for the app to work effectively. As such, the development tools become even more similar (Table 3 model (6), *Outsource x H. Tech Count*: -0.056 SD 0.003; Figure C.5.A.).

As a second mechanism, startups with a stronger cloud provider relationship will likely have access to additional shared resources yet face greater fit constraints from using multiple cloud services products. Startups are more entrenched in a single supplier's technology when they license cloud hardware infrastructure services (e.g., virtual machines providing computing power) in addition to cloud platform services. These additional shared resources sources from the potentially stronger relationship may mitigate some compatibility issues but also may guide startups down the same development path or technological trajectory. Development tools become more similar when startups use multiple cloud services from the same provider (Table 3 model (7), *Outsource x IaaS*: -0.032 SD 0.002; Figure C.5.B.), which makes them face the same new set of additional fit-related constraints and provides them with the same resources that lead them to similar technical solutions.

More diverse data analytics capabilities. Technologies enabling firms to collect, analyze, and recombine data become more dissimilar when outsourcing to a cloud platform (Table 3 model (12): +0.087 SD 0.005;

coefficients depicted in Figure 1).[42] Since these technologies are not related to the mechanics of product development, startups can use large and more diverse analytics technology bundles without fit and compatibility-related issues reducing the app's efficacy. Unlike the tools used to develop the app, data analytics technologies are not a cog in a larger development process. Data analytics technologies are more modular (i.e., less interdependent) than development technologies. Without these constraints, having a larger analytics technology bundle enables the adoption of a more diverse data analytics technology bundle. Moreover, additional shared resources enable the adoption of a more diverse analytics technology bundle as they are modular and less affected by fit constraints from adopting additional cloud services products.

Substitution. Before outsourcing, there is no significant relationship between the dissimilarity of product development and data analytics technologies. However, after outsourcing, there is a significant tradeoff between these two technologies: using more standardized tools relates to adopting more diverse data capabilities. The sequencing of this tradeoff is even starker for startups with fewer employees than those with more employees, as smaller startups have more resource-deprived (Appendix Figure C.6.). Smaller startups adopt more standardized development tools in the period after outsourcing and adopt more diverse data capabilities in the following period. I interpret this as startups saving programmer time by outsourcing to a cloud platform and taking advantage of the standardized tools and guidance on how to use those tools to develop an app that integrates well with their cloud platform. Startups can focus their technical labor on other priorities, such as developing more diverse data analytics capabilities, which enable them to build more robust data resources.

6.B. Identification

In addition to matching and weighting regressions based on observable demographics, I use three additional methods to adjust regression estimates for potential endogeneity: Heckman's selection approach, an instrumental variable approach, and a double machine learning model.

[42]In Table 3 models (8)-(11), I provide a similar build-up to support these results as in the preceding section, adding firm fixed effects, year fixed effect, firm and year fixed effects, and firm and year fixed effects with matching.

Selection model. Empirical insight into firm-level decisions to outsource to a cloud platform is scarce in prior research (Schneider and Sunyaev, 2016; Yang and Tate, 2012), partly because adoption is endogenous. Startups are not randomly assigned to different IT development conditions or cloud platforms. Instead, founders and early IT employees chose to develop internally or adopt services from a specific cloud provider.

I examine selection based on founder characteristics, industry, headquarters location, age, and funding from a Big Tech cloud provider using a time-series probit specification in Table 4.[43] In model (1), I examine cloud platform adoption. The startup's age is a significant driver of cloud platform adoption (+0.65 SD 0.01), as startups make these decisions when they reach product development. Startups selling into healthcare and energy, more highly regulated industries, are less likely to outsource. Additionally, in Europe, where data regulation (e.g., GDPR) is more intense, startups are less likely to outsource to cloud platform services (-0.23 SD 0.05), raising a potential concern that European entrepreneurs are missing out on the benefits of cloud if, indeed, there are benefits.[44] Firms with earlier funding from one of the three largest cloud providers are more likely to outsource to a cloud platform (+0.30 SD 0.11). As anticipated, startups with a founder with hardware experience are more likely to develop IT in-house. Lastly, founders with an MBA or advanced technical degree are more likely to outsource.[45]

Using Table 4 model (1) as the base probit specification, I calculate the inverse Mill's ratio (IMR) based on Heckman's (1979) selection approach, using including founder characteristics in the IMR calculation but excluding them in second stage regression (Appendix Note A.7.). Results remain similar to the main results when including IMR as a control (Table B.2. model (4), *development dissimilarity*: -0.025 SD 0.09; model (8), *analytics dissimilarity*: +0.086 SD 0.005).

[43]As a limitation of this selection analysis, certain characteristics may not be observable or are not captured in the data. Moreover, the size of the founding team and combinations of the founding team's characteristics could affect selection in a way unaccounted for by my model.
[44]And this finding supports the notion that Americans may do (cloud) I.T. better than others, too (Bloom et al., 2012).
[45]I examine the selection of a Big Tech versus other smaller cloud platform providers for the treatment group in Table 4 model (2). Model (3) compares the selection of Amazon AWS versus other providers. Model (4) uses a multinomial time series probit model and finds similar results to model (2) based on the full sample.

Instrumental variable approach. To address endogeneity issues stemming from unobservable variables and reverse causality, I use the release of an open source[46] version of TensorFlow in late 2015 as a quasi-exogenous shock to AI startups' adoption of cloud services. TensorFlow is an AI framework that enables firms to develop and train deep learning algorithms. Google's decision to develop and release TensorFlow enabled AI-producing firms using cloud services to be more productive and their complementary labor more valuable, enabling them to customize this AI framework to their development needs (Rock, 2021).[47] After the open source release, TensorFlow could be used on any cloud platform, not just Google's platform. However, it took another year for Amazon to develop an Amazon Machine Image (AMI) that easily enabled the use of TensorFlow on their platform.[48]

This instrumental variable approach includes the interaction between two binary variables: startups that 1) benefit from TensorFlow (i.e., startups that adopted Google Cloud Platform in 2016 or Amazon AWS in 2017) and 2) develop AI products (i.e., startups that benefit from AI frameworks like TensorFlow), described in Appendix Note A.8. Including an interaction term between two binary variables increases the strength of the first-stage regression without biasing estimation (Aghion et al., 2005; Bun and Harrison, 2018). In this case, my first stage regression yields significant F-statistics (K-P Wald F: 102, C-D Wald F: 105), consistent with the argument that the instrument is adequately powered.

Using TensorFlow directly relates to AI startups adopting a cloud platform, the outcome variable in the first stage. (Table 5.A., *Tensor*: +0.24 SD 0.018; *Tensor x AI*: +0.29 SD 0.024). Yet, in support of the exclusion restriction, TensorFlow does not directly relate to the breadth of technology adoption.[49] The

[46]Apache 2.0 opensource license
[47]Other recent research also supports that complementarities exist between AI and human labor (Choudhury et al., 2020; Krakowski et al., 2022; Tong et al., 2021).
[48]Though other AI frameworks were released around this same time, TensorFlow was the most popular. Keras was released in March 2015; Microsoft's Cognitive Toolkit (CNTK) was released in January 2016; Facebook's PyTorch was released in September 2016. Amazon's Sagemaker, released in November 2017, and open sourced in 2019.
[49]There are two potential limitations I wanted to address. First, though I discussed TensorFlow's higher market share and the release timing of other competing frameworks, it is possible that I am gauging the effect of TensorFlow and other frameworks. I try to overcome this by building my TensorFlow measure to exclude startups on Google's cloud platform, who were less able to benefit from TensorFlow's open source release but could still benefit from the release of other platforms. Second, though many programmers tout the versatility of TensorFlow, there could be some technologies that are indeed dependencies and influence the breadth of innovation (on the margins). However, I cannot find any documentation that suggests this.

second-stage regression results remain directionally similar and significant. The impact of cloud services increases product development dissimilarity (i.e., development tools become even more similar) from -0.027 SD 0.002 (main analysis repeated in Table 5.B. model (1)) to -0.051 SD 0.016 (IV approach, model (2)). Analytics dissimilarity remains similar, slightly increasing from +0.087 SD 0.005 (main analysis repeated in model (4)) to +0.089 SD 0.018 (IV approach, model (5)).[50]

Double machine learning. As another method of addressing endogeneity from potentially omitted variables (Belloni et al., 2014), I use a double machine learning model to estimate the treatment and outcome using a random forest machine-learning algorithm trained with 64 firm-level control variables. This approach, described in Appendix Note A.9., divides the sample in half, using half the observations to train the model and the other half for prediction, and calculates Neyman orthogonal scores to estimate the causal parameter (Chernozhukov et al., 2018; Neyman, 1959; Wooldridge, 1991). I then take the first differences based on the machine learning models' prediction of the (1) treatment and (2) outcome and run an OLS regression with firm and year-level fixed effects.

Similar to the instrumental variable approach, the double machine learning approach dampens the effect of platform adoption, down about 40% from the base model (Table 5.B. model (3), *Development dissimilarity*: -0.015 SD 0.002; model (6), *Analytics dissimilarity*: +0.053 SD 0.004). However, this additional analysis support that the effect remains and is significant.[51]

6.C. Robustness

Since my panel is unbalanced, I run the main specification with a subsample of firms with data before and after the outsourcing event, dropping all firms that outsourced in their initial year of existence (Table B.2.

[50] I realize that an instrumental variable approach does not solve all related endogeneity issues; yet, in conjunction with my main analysis, these findings provide more confidence in the scale and direction of my findings. As robustness for this approach, I examined the effect of TensorFlow's release on all firms (not just AI startups), and the results are not significant in the second stage.

[51] I then test coefficient stability (Oster, 2019; *Development dissimilarity*: $\delta = 3.08$, *Analytics dissimilarity*: $\delta = 1.35$) to support that effect is unlikely to be negated by unobserved variables. A $\delta = 3.08$ is interpreted as the impact of unobserved variables would need to be 3x greater than the impact of observed variables for the effect to change signs. Impacts greater than 1x are highly unlikely (<5%).

model (1), *Development dissimilarity*: -0.027 SD 0.02; model (5), *Analytics dissimilarity*: +0.092 SD 0.005). Next, I run the main specification with firms greater than three years old with data from 2012 to 2018 to overcome potential concerns that: (a) startups have less choice over whether they outsource to cloud platforms in more recent years, (b) cloud services are fundamentally different in earlier years, and (c) parallel trends impact analytics technology bundle dissimilarity measures before 2018, and (d) product development and data analytics technologies are at different stages in their lifecycle in earlier periods (model (2), *Development dissimilarity*: -0.026 SD 0.03; model (6), *Analytics dissimilarity*: +0.082 SD 0.006). Third, I include a firm-level investor overlap measure as a control for resource sharing from investors (model (3), *Development dissimilarity*: -0.033 SD 0.04; model (7), *Analytics dissimilarity*: +0.079 SD 0.008).[52] For instance, firms with the same investors may receive similar guidance (e.g., "talk to Sue about how to build that feature," "hire the new programmer with X skill or through Y recruiting agency," or "join Z startup programs").

Fourth, to ensure a single cloud provider does not drive the outsourcing effect, I run the main specification for subsets of firms by cloud provider: AWS, GCP, MS, and Other; results remain similar (Table B.3.). Fifth, I show the results of an alternate dependent variable, a technology dissimilarity measure based on how often a particular technology is used by startups in the sample (Appendix Note A.10.; Table B.4.). Sixth, I show that my analysis holds in several key verticals: AI, ML, Financial Services, and Healthcare (Table B.5.). Seventh, to support that spatial autocorrelation, stemming from technological spillovers from technology hub locations, does not impact the validity of my results, I show that results for city-level subsamples for San Francisco, London, and New York are similar in direction and significance to the main results (Table B.6.).

Eight, as robustness for my staggered difference-in-differences model with two-way fixed effects, I use several approaches to estimate treatment effects. In the most straightforward approach, I collapse my

[52]Using this methodology, I create an investor dissimilarity measure based on if startups have overlapping investors for the firms that have investors. For example, on the extremes, this measure takes a higher value if two firms have the same investors and takes a value of 0 if they have no investors in common. Mean investor dissimilarity is 0.59 SD 0.08.

data to pre- and post-estimates for a balanced panel of 873 firms (i.e., switchers) that change from in-house to outsourcing. This two-period model estimates the outsourcing event to be more intense (Table B.7. model (1), *Development dissimilarity*: -0.086 SD 0.002; model (5), *Analytics dissimilarity*: +0.129 SD 0.005). In another approach, I report the average treatment effect (ATE), computed as the difference between the average treatment received by switchers after their first switch and the treatment they would have received if they had never switched (models (2) and (6); De Chaisemartin and d'Haultfoeuille, 2020). Additionally, I use a fuzzy difference-in-differences approach to estimate the local average treatment effect (LATE; models (3) and (7); De Chaisemartin and d'Haultfoeuille, 2018). In the last approach, I estimate heterogeneity-robust instantaneous treatment effects (ITE) to estimate the treatment in each period (models (4) and (8); Athey and Imbens, 2022; De Chaisemartin and d'Haultfoeuille, 2020).[53]

Ninth, I use disaggregated pairwise data (35 million observations) to address potential issues from the treatment effect spillovers (i.e., stable unit treatment violations assumption; SUTVA). This more granular data enables me to calculate the pairwise angular distance for *stable rivals*, including all the focal startups and only the rival startups that do not change their technologies. Results on this subset are similar, suggesting that spillovers from the treatment effect did not change the results (Table B.8. model (1), *Development dissimilarity*: -0.027 SD 0.003; model (9) *Analytics dissimilarity*: +0.083 SD). Then lastly, I include firm, rival, and year-level fixed effects to show that results are robust to all combinations of these effects (models (8) and (16)).

Tenth, I also run two analyses to ensure that differences in the technology lifecycle stage of product development technologies (older) and data analytics technologies (newer) do not impact my results. Technologies that are relatively newer would experience greater increases in diversity as the number of technologies in the category exponentially grow versus more mature technology categories. First, I run a randomization, which shows similar downward trends (i.e., increased similarity) over time. However, data

[53]Under the Common Trends Assumption, further analyses of switchers support that all switchers who received the treatment have positive weight. STATA package: *twowayfeweights*; De Chaisemartin and d'Haultfoeuille (2020). Also, this approach estimates prior periods before the outsourcing event, confirming that pre-trends are not an issue.

analytics technologies start from a higher point, aligning with my understanding that these technologies are, on average newer and at an earlier lifecycle stage (Figure C.7.). Given that there is no difference in trends, the year-level fixed effect in my main specification effectively adjusts for differences in levels. Second, I examine Shannon's (1948) entropy for both these technologies. Product development technology entropy increases at a slower rate than data analytics technology entropy, though absolute levels are similar and not significantly different from one another (Figure C.8.). I interpret this as data analytics technologies becoming slightly less concentrated over time, enabling more diversity.

6.D. Additional findings

Supplier power. Three-quarters of startups in my sample outsource to Amazon AWS. I examine heterogeneous effects adopting the largest platform, AWS, which has higher market power compared with other suppliers.[54] Based on analyses of the treatment group, using higher market share platforms relates to using more standardized tools (Table B.10. model (1): -0.08 SD 0.01) and more diverse data analytics capabilities (model (4): +0.09 SD 0.03). The effect remains similar when using specification (3) to compare startups using one of the largest three cloud providers (AWS, GCP, and MS Azure) versus other smaller cloud providers (Alibaba, Digital Ocean, IBM, OVH, Oracle, and Linode) in models (2) and (5) and comparing firms using the AWS platform versus those using any other platform (models (3) and (6)). [55]

Heterogeneity by location. Startups located in cities with higher concentrations of VC firms (e.g., San Francisco Bay Area, London, New York, Boston) use more diverse analytics technology bundles after they outsource, despite having a similar level of development technology bundle dissimilarity (Appendix Figure C.10.). This finding suggests that these hubs potentially enable increased analytics technology dissimilarity, yet there are several plausible mechanisms. For instance, more robust analytics technologies could be a hot topic for investors, increasing startup awareness. Though, cities with higher concentrations of investors also

[54]For instance, Google GCP is the next largest supplier with around 5% market share. 2,466 startups use a single cloud platform; 467 startups use more than one cloud provider's platform. Analysis in Appendix Table B.10 only includes firms that use a single cloud provider.
[55]I graph technology dissimilarity on the event timeline starting for AWS versus not AWS from year 0, the year the startup outsourced (treatment group only), in Appendix Figure C.9.

have more technology firms. These firms could increase awareness of the benefit of more robust and unique data resources amongst themselves, potentially by hiring each other employees.[56] Moreover, these technology hubs have a higher proportion of programmers and tech-focused labor than other cities.

7. Link with product differentiation and performance

Product differentiation. Analyses of technology adoption are interesting in their own right, especially in the context of app development, given the close gap between development technologies and end products. In this paper, they serve as a proxy for startup differentiation because objectively collected product data from these young, small startups are challenging to find at scale. To understand end-product differentiation, I use text analysis to measure changes to startup descriptions. These descriptions are typically a couple of sentences describing a startup's product.[57] Though this measure is likely a closer proxy product differentiation than other aspects of technology adoption, the smaller sample size is relatively small, as only 193 startups changed their descriptions after outsourcing to the cloud. Moreover, the reasons prompting startups to change their description are unclear.

The text measure of description differentiation is significantly correlated with data analytics technology bundle dissimilarity, providing additional support that data analytics capabilities relate to increased differentiation. Moreover, I find a link that descriptions become more differentiated after outsourcing (Table 6 model (1): +0.016 SD 0.004; Figure 2), and this correlation increases when the outsourcing startups use more dissimilar data analytics technology bundles (model (3), *Outsource x Analytics dissimilarity (cont.)*: +0.027 SD 0.009). This finding corroborates that more distinct analytics capabilities aid in product differentiation.

[56] Other aspects of location are not significant; however, heterogeneity by VC location suggests there is potential to explore spatial spillovers in additional research. For instance, distance from a VC funding/technology hub is likely to be more influential on technology adoption than other location features.

[57] Calculation: quanteda is an R package for managing and analyzing textual data developed by Kenneth Benoit, Kohei Watanabe, and other contributors. The European Research Council supported its initial development. Summary statistics are reported in Table 1.D. under the heading *Product Differentiation*.

Link with performance. Performance analyses are correlational, relying on OLS specification with firm and year-level fixed effects and panel data of startups that outsource (Figure 3). Decreased development technology dissimilarity relates to decreased funding (Table 7 model (5): -3.3 SD 1.2), and decreased analytics technology dissimilarity relates to increased performance (model (6): +2.3 SD 0.71). Performance results align with my interpretation that a) startups using more similar development tools benefit from using technologies that fit with the cloud platform's underlying technologies and b) startups benefit from receiving their cloud supplier's resources, which provide insight into navigating compatibility issues among interdependent technologies. Moreover, startups benefit when adopting more diverse data analytics capabilities, which produce data as outputs that aid in product differentiation.

A similar relationship holds for binary measures of positive funding and follow-on (i.e., second-round) funding. [58] Lastly, I examine the average duration spent on a startup's web domain as an additional performance measure, and I find a significant correlation between adopting more similar development tools and increased web traffic (model (7): -3.56 SD 0.48) and patenting (model (9): -0.08 SD 0.023).

8. Discussion and Conclusion

Given the importance of data-centric entrepreneurship and AI development to future macroeconomic growth, it is paramount to understand how the rise of several large cloud platforms that share lots of resources with the startup ecosystem affects startup differentiation and growth. Cloud platforms are here to stay. They continue to grow in usefulness and scale and provide services that are becoming increasingly difficult to replicate in-house. Cloud suppliers enable access to IT and share resources, making it easier for startups to fund entry and develop their digital products. However, less differentiated products may benefit the economy less, driving lower productivity growth than anticipated.

[58]For robustness, I use alternate dependent variables, including VC backed, higher reputation VC, closed, and acquired, to support this interpretation further (Table B.10). Additionally, I examine the interaction between an indicator variable for larger bundles of technologies and higher levels of technology dissimilarity (Table B.11).

My findings contribute to understanding how outsourcing to a cloud platform impacts the breadth of technologies used to develop digital innovations. More specifically, the impact of the cloud on product differentiation depends on whether interdependencies among technologies and fit issues with the platform hinder the technology's functionality in producing its intended outcome. However, when interdependencies create compatibility issues, shared resources present a solution, guiding startups down a particular technological path to alternate compatible technologies. Moreover, some evidence suggests that investors reward adopting more standardized, similar bundles of technologies when there are fit constraints. Perhaps this indicates that top-performing startups of a particular vintage converge on adopting the latest and greatest bundle of compatible product development technologies at a given time. Using more standardized tools reduces coordination and adaptation costs and saves time.

For resource-strapped startups, unspent technical resources can be redeployed on other firm objectives like building data analytics capabilities. It is plausible that startups repurpose programming labor to experiment with which data analytics technologies are necessary to ascertain needed data. These technologies are less constrained by fit and interdependencies and become more distinct from other startups that outsource to the cloud. Moreover, the substitution of programming labor, coupled with fewer technical constraints, likely enables the adoption of more diverse data analytics technologies, enabling them to collect and manipulate data resources to suit their needs and aiding in product differentiation. Furthermore, this narrative aligns with the finding from the analysis of changes in startup descriptions; more diverse data analytics technologies are important for digital product differentiation.

I employ numerous econometric approaches (i.e., matching, selection models, instrumental variable, firm fixed effects, double machine learning) to provide results consistent with a causal argument based on observed and unobserved variation to rule out alternate explanations and adjust estimates of potentially omitted variables. Though I cannot entirely dispel threats to causal identification: all analyses yield similar results; an Oster test suggests that the risk of an unobservable variable negating my findings is low; the instrumental variable approach provides some evidence that reverse causality is not a significant concern.

Evidence supports that European startups are slower to adopt cloud services than similar startups in the US. However, there is limited evidence of the heterogeneous effects of adopting cloud platform services on technology adoption by country or region. Startups in Europe may be missing the potential benefits of adopting cloud platform services more quickly, enabling programmers to focus less on development tools and more on analytics capabilities. National boundaries seem less important to cloud platform adoption's impact than startup proximity to a large technology hub city (San Francisco, New York, London). Startups in or near these cities use more diverse analytics technologies, suggesting a potentially nuanced spatial relationship between technological differentiation and distance from these hubs. Estimating these technology differentiation spillovers remains an interesting avenue for future research.

By comparing startups using cloud providers with and without startup programs, this paper contributes to the literature on interfirm resource sharing by showing the impact of the customer-supplier relationship on digital product differentiation. There are significant differences in technology adoption for startups using cloud platforms with startup programs (e.g., AWS, GCP, and Azure) compared to those using smaller cloud platforms that do not have these programs. Programs that share platform-related resources lower the costs of coinventing on the platform and selecting other compatible technologies.

Though these findings suggest that startup and product differentiation remain robust presently, they also raise concerns that cloud suppliers can curate which bundles of compatible technologies are used in high-tech product development, directly and indirectly altering technology adoption. Technology firms managing these cloud platforms control all the levers. They choose which features to build into the platform, which resources to develop and share with startups, and which technologies are more and less compatible with their platform. Using suppliers with more monopsony power (i.e., control) relates to adopting more similar development tools yet more diverse analytics capabilities, providing insight into how startups differentiate digital products while using more standardized development tools. Tying back to transaction cost economics, firms become more dependent on their suppliers over time, potentially enabling cloud suppliers with higher market share to exert control over technology adoption in a way that may not yet manifest in the inability to differentiate products. Despite increased control, the importance of analytics capabilities and richer data resources in enabling high-tech startups to differentiate their products will remain constant.

This gradual increase in supplier control will not bode well for startups competing against their cloud providers in downstream markets. It posits an interesting question for future research: Will startups be able to effectively differentiate their products when competing against one of these larger technology firms in downstream markets? Startups' analytics capabilities and data resources pale in comparison to those of the largest technology firms (Benzina, 2019; Iansiti, 2021; Khan, 2016; Scott-Morton et al., 2019). In addition to having less robust data-related technologies, capabilities, and resources, startups share tons of information about their products and industries with their cloud providers through these corporate startup programs. Cloud providers can use startup feedback and platform usage data[59] to improve their platform, products, and strategies, and this information is likely competitively valuable when aggregated across many startups. In the most egregious cases, technology firms could use this information to make acquisitions directly shaping the technological landscape (Cunningham et al., 2019; Zingales et al., 2021). In a likelier scenario, they could use this information and control over their platform's compatibilities to advance their and their largest customers' strategies and goals.

[59]Passively provided data is often referred to as telemetry and is a user's digital footprint on the platform or usage "exhaust" (Chatterji and Fabrizio, 2014).

Tables and Figures

Table 1.A. – Demographics and founders summary

	Unmatched								Matched					
	All startups				In-house		Outsource		All startups		In-house		Outsource	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Demographics														
Age	4.0	2.4	0.0	9.0	3.0	2.3	4.3	2.3	4.2	2.3	4.2	2.4	4.2	2.3
Employment	45	63	1	375	35	50	47	66	40	55	39	55	41	55
Employment (<10 emp., dummy)	0.36	0.48	0	1	0.41	0.49	0.35	0.48	0.36	0.48	0.38	0.49	0.36	0.48
Healthcare	0.09	0.29	0	1	0.10	0.30	0.09	0.28	0.09	0.29	0.10	0.30	0.09	0.29
Finance	0.09	0.28	0	1	0.07	0.25	0.09	0.29	0.08	0.28	0.07	0.25	0.09	0.28
AI	0.38	0.49	0	1	0.37	0.48	0.39	0.49	0.39	0.49	0.40	0.49	0.38	0.49
Machine learning	0.09	0.29	0	1	0.08	0.27	0.10	0.30	0.10	0.29	0.09	0.29	0.10	0.29
US	0.49	0.50	0	1	0.41	0.49	0.51	0.50	0.50	0.50	0.47	0.50	0.51	0.50
UK	0.06	0.25	0	1	0.06	0.24	0.07	0.25	0.06	0.24	0.04	0.21	0.07	0.25
France	0.03	0.17	0	1	0.02	0.14	0.03	0.18	0.03	0.16	0.01	0.11	0.03	0.18
Germany	0.03	0.16	0	1	0.04	0.20	0.02	0.15	0.02	0.15	0.03	0.17	0.02	0.15
Canada	0.04	0.20	0	1	0.04	0.21	0.04	0.20	0.04	0.20	0.05	0.22	0.04	0.20
Americas	0.56	0.50	0	1	0.47	0.50	0.58	0.49	0.58	0.49	0.54	0.50	0.58	0.49
Asia (ex. China)	0.13	0.34	0	1	0.15	0.35	0.13	0.33	0.14	0.34	0.17	0.37	0.13	0.33
Europe	0.27	0.44	0	1	0.34	0.47	0.25	0.43	0.25	0.43	0.25	0.43	0.25	0.43
Founders														
IT experience	0.12	0.33	0	1	0.10	0.30	0.13	0.34	0.12	0.33	0.11	0.31	0.13	0.33
Hardware experience	0.05	0.22	0	1	0.05	0.22	0.05	0.22	0.05	0.22	0.05	0.23	0.05	0.21
Big Tech experience	0.07	0.25	0	1	0.04	0.20	0.07	0.26	0.06	0.25	0.04	0.21	0.07	0.26
Technical major	0.44	0.50	0	1	0.32	0.47	0.47	0.50	0.44	0.50	0.33	0.47	0.46	0.50
Advanced degree	0.21	0.41	0	1	0.18	0.39	0.22	0.41	0.21	0.41	0.19	0.39	0.22	0.41
MBA	0.24	0.43	0	1	0.17	0.37	0.25	0.44	0.23	0.42	0.17	0.37	0.25	0.43
Female	0.13	0.33	0	1	0.12	0.33	0.13	0.33	0.13	0.34	0.12	0.33	0.13	0.34

Notes: Unmatched summary statistics are calculated at the firm-year level for all firms in the sample. Matched summary statistics use Coarsened exact matching (CEM): age #10, employment size #10, healthcare, financial services, and region #4 based on the treatment, cloud platform versus no cloud platform, dropping 21 unmatched firms in the main analyses. All firms included in the sample have a digital app, have an active web domain, are listed as active on Crunchbase or Pitchbook, are ten or fewer years old, have fewer than 500 employees, and are not located in China. All demographic information is from Crunchbase and Pitchbook. Information on gender is from an analysis of founder names in R. Founders' background measures are calculated at the firm-year level and based on data from AIdentified, Mantheos, Pitchbook, and manual collection of public profiles on LinkedIn.

Table 1.B. – Cloud services type panel summary

	Startups (3,434)				Observations (19,678)			
	In-house	PaaS	IaaS	IaaS/PaaS	In-house /Before Outsource	PaaS	IaaS	IaaS/PaaS
All	501	2,933	2,739		1,916/1,949	15,813	13,318	
AWS only		1,181	1087	92%		7,845	6,714	86%
GCP only		204	180	88%		867	696	80%
Azure only		48	47	98%		358	300	84%
Other supplier		121	73	60%		955	398	42%
AWS/GCP		509	503	99%		1,823	1,710	94%
AWS/Azure		133	133	100%		751	699	93%
GCP/Azure		10	10	100%		54	53	98%
AWS/GCP/Azure		75	75	100%		226	218	96%
Mixed (Big Tech & other)		652	631	97%		2,934	2,530	86%

Notes: Cloud services provider information is from BuiltWith. IaaS is cloud infrastructure services, licensed by firms to access computation capability from PCs and servers. PaaS is cloud platform services, licensed by firms to host and develop applications.

Table 1.D. – Platform comparison summary

	Amazon AWS		Google GCP		Microsoft Azure		Other CSP	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Technology adoption								
All	52.77	27.2	47.68	23.4	49.35	23.1	49.43	24.4
Development tech count	8.83	4.1	7.30	3.6	8.62	3.9	8.10	3.7
Analytics tech count	7.14	5.6	5.75	4.8	5.73	4.8	5.74	4.9
Big tech count	13.53	6.1	12.43	5.3	13.71	5.7	11.41	5.7
Premium/paid count	4.02	3.7	2.74	3.1	2.78	2.9	3.01	3.3
Open Source count	1.32	1.2	0.88	1.1	1.21	1.2	1.09	1.2
Development dissimilarity	0.77	0.07	0.77	0.07	0.80	0.08	0.78	0.07
Analytics dissimilarity	0.68	0.08	0.69	0.10	0.67	0.10	0.66	0.09
Product differentiation								
Firm description	0.86	0.04	0.86	0.05	0.87	0.03	0.86	0.03
IP description	0.93	0.02	0.93	0.01	0.93	0.01	0.93	0.01
Performance								
Funding	0.77	0.42	0.66	0.47	0.75	0.44	0.69	0.46
Follow-on funding	0.58	0.49	0.49	0.50	0.54	0.50	0.49	0.50
VC-backed	0.67	0.47	0.59	0.49	0.56	0.50	0.57	0.50
Higher rep. VC	0.10	0.30	0.09	0.28	0.06	0.25	0.05	0.22
Deal size (log)	5.10	7.1	4.76	7.0	3.98	6.4	4.29	6.6
Funds raised (cumulative, log)	12.8	5.6	11.3	6.5	11.8	5.7	11.3	6.2
Acquired	0.08	0.27	0.03	0.18	0.01	0.10	0.06	0.23
Closed	0.05	0.22	0.05	0.22	0.06	0.24	0.06	0.24
Accelerator	0.20	0.40	0.18	0.38	0.26	0.44	0.17	0.38
Big Tech funding	0.02	0.13	0.02	0.16	0.11	0.32	0.02	0.14
SimilarWeb visit duration	3.4	3.3	3.2	3.0	3.0	3.3	2.9	3.2
Patents	0.03	0.18	0.07	0.25	0.03	0.16	0.03	0.16

Notes: Summary statistics are calculated for firms in the sample that use a single primary cloud provider: AWS, GCP, Azure, or Other cloud providers (e.g., Linode, Digital Ocean, etc.)

Table 1.C. – Technology, product, and performance summary

	Unmatched								Matched					
	All startups				In-house		Outsource		All startups		In-house		Outsource	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Technology adoption														
All	49.9	26.5	1	252	36.0	18.9	53.3	26.9	49.6	25.8	39.1	20.2	52.3	26.4
Development tech count	8.2	4.1	1	27	5.5	3.1	8.8	4.1	8.1	4.0	5.9	3.2	8.7	4.0
Analytics tech count	6.4	5.5	0	39	3.5	3.2	7.1	5.7	6.3	5.3	3.9	3.6	6.9	5.5
Big tech count	12.5	6.3	0	46	8.0	4.2	13.6	6.2	12.4	6.1	8.7	4.5	13.3	6.0
Premium/paid count	3.5	3.7	0	27	1.3	1.9	4.0	3.8	3.4	3.5	1.6	2.2	3.8	3.6
Open Source count	1.2	1.2	0	7	0.9	1.1	1.3	1.2	1.2	1.2	0.9	1.1	1.3	1.2
Development dissimilarity	0.78	0.07	0	1.00	0.81	0.08	0.77	0.07	0.78	0.07	0.80	0.08	0.77	0.07
Analytics dissimilarity	0.66	0.12	0	1.00	0.59	0.17	0.68	0.09	0.67	0.11	0.61	0.16	0.68	0.09
Product differentiation														
Firm description	0.86	0.04	0.56	0.97	0.86	0.04	0.86	0.04	0.86	0.04	0.86	0.04	0.86	0.04
IP description	0.93	0.02	0.83	0.99	0.93	0.02	0.93	0.02	0.93	0.02	0.93	0.02	0.93	0.02
Performance														
Funding	0.71	0.45	0	1	0.53	0.50	0.76	0.43	0.72	0.45	0.59	0.49	0.75	0.43
Follow-on funding	0.52	0.50	0	1	0.31	0.46	0.57	0.49	0.53	0.50	0.39	0.49	0.56	0.50
VC-backed	0.61	0.49	0	1	0.42	0.49	0.65	0.48	0.61	0.49	0.48	0.50	0.64	0.48
Higher rep. VC	0.09	0.28	0	1	0.04	0.20	0.10	0.30	0.08	0.28	0.05	0.23	0.09	0.29
Deal size (log)	5.1	7.1	0	22	5.0	6.8	5.1	7.1	4.9	7.0	4.6	6.8	5.0	7.0
Funds raised (cumulative, log)	11.9	6.2	0	22	8.9	7.0	12.6	5.8	11.9	6.1	9.8	6.9	12.5	5.8
Acquired	0.06	0.24	0	1	0.02	0.14	0.07	0.26	0.06	0.24	0.03	0.17	0.07	0.25
Closed	0.05	0.21	0	1	0.04	0.20	0.05	0.22	0.05	0.22	0.04	0.19	0.05	0.22
Accelerator	0.19	0.39	0	1	0.13	0.33	0.20	0.40	0.19	0.39	0.14	0.35	0.20	0.40
Big Tech funding	0.02	0.15	0	1	0.01	0.11	0.03	0.16	0.02	0.15	0.02	0.12	0.03	0.16
SimilarWeb visit duration	3.1	3.3	0	12	1.9	2.8	3.4	3.3	3.1	3.2	2.3	3.0	3.3	3.3
Patents	0.04	0.18	0	1	0.03	0.17	0.04	0.19	0.04	0.18	0.04	0.18	0.04	0.18

Notes: Unmatched summary statistics are calculated at the firm-year level for all firms in the sample. Matched summary statistics use Coarsened exact matching (CEM): age #10, employment size #10, healthcare, financial services, and region #4 based on the treatment, cloud platform versus no cloud platform, dropping 21 unmatched firms in the main analyses. All firms included in the sample have a digital app, have an active web domain, are listed as active on Crunchbase or Pitchbook, are ten or fewer years old, have fewer than 500 employees, and are not located in China. Information on technologies is from BuiltWith. Funding information is from Crunchbase or Pitchbook. Patent information is from IPQwerty. Web traffic information is from SimilarWeb.

Table 2 - Technology bundle size

DV is log of:	Development technology count						Ratio: Count/All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
[0,1] Outsource	0.429***	0.583***	0.371***	0.295***	0.290***	0.150***	0.012***
	(0.015)	(0.018)	(0.015)	(0.019)	(0.020)	(0.025)	(0.003)
[0,1] IaaS						0.102***	
						(0.030)	
Outsource x IaaS						0.353***	
						(0.021)	
R2	0.100	0.491	0.168	0.588	0.586	0.594	0.502

	Analytic technology count						Ratio: Count/All
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
[0,1] Outsource	0.515***	0.612***	0.444***	0.235***	0.225***	0.052*	0.003
	(0.019)	(0.023)	(0.020)	(0.024)	(0.024)	(0.029)	(0.002)
[0,1] IaaS						0.118***	
						(0.041)	
Outsource x IaaS						0.301***	
						(0.027)	
R2	0.0733	0.560	0.125	0.641	0.646	0.652	0.565
Observations	19679	19389	19679	19389	18802	18802	18802
Firms	3434	3144	3434	3144	3123	3123	3123
Firm FE	No	Yes	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes	Yes
CEM	No	No	No	No	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS; standard errors are clustered at the firm level in parentheses below the coefficients. Cloud Platform is an indicator variable for adopting cloud platform services from a cloud services provider. Models (6) and (13) examine the interaction between using cloud platform services and IaaS (an indicator variable for using cloud infrastructure services) to estimate the impact of having a stronger relationship with a cloud provider. Models (5-7) and (12-14) drop unmatched startups and weight regressions based on Coarsened exact matching (CEM): age #10, employment size #10, healthcare, financial services, and region #4.

Table 3 - Technology bundle dissimilarity

DV is:	Development dissimilarity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
[0,1] Outsource	-0.041***	-0.089***	-0.025***	-0.028***	-0.028***	-0.018***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
[0,1] H. Tech Count						-0.022***	
						(0.003)	
Outsource x						-0.056***	
H. Tech Count						(0.003)	
[0,1] IaaS							-0.004
							(0.005)
Outsource x							-0.032***
IaaS							(0.002)
R2	0.0470	0.385	0.355	0.668	0.673	0.701	0.675
DV is:	Analytics dissimilarity						
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
[0,1] Outsource	0.096***	0.140***	0.084***	0.093***	0.087***	0.098***	0.097***
	(0.004)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
[0,1] H. Tech Count						0.089***	
						(0.007)	
Outsource x						0.108***	
H. Tech Count						(0.005)	
[0,1] IaaS							0.017
							(0.011)
Outsource x							0.086***
IaaS							(0.005)
R2	0.110	0.502	0.185	0.575	0.577	0.590	0.577
Observations	19679	19389	19679	19389	18802	18802	18802
Firms	3434	3144	3434	3144	3123	3123	3123
Firm FE	No	Yes	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes	Yes
CEM	No	No	No	No	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS; standard errors are clustered at the firm level in parentheses. CSP is an indicator variable for adding a CSP (PaaS). Weighting is based on CEM: age #10, employment size #10, healthcare, financial services, and region #4. In models (6) and (13), H. Tech count is an indicator variable for startup-level above-median development and analytics bundle size. In models (7) and (14) IaaS is an indicator variable for using CSP cloud infrastructure services.

Table 4 - Probit selection analysis

Treatment, DV is:	(1) Outsource vs. in-house	(2) Big Tech vs. Other CSP	(3) AWS Only vs. Not AWS	(4) mprobit In-house/before (0); Big (1); Other (2)	
				=1	=2
[0,1] Healthcare	-0.136**	0.082	-0.019	0.069	-0.014
	(0.066)	(0.076)	(0.074)	(0.046)	(0.075)
[0,1] Energy	-0.393*	0.124	0.114	-0.493***	-0.418*
	(0.200)	(0.257)	(0.254)	(0.132)	(0.229)
[0,1] Financial	0.121*	0.052	0.025	0.124***	0.179**
	(0.073)	(0.078)	(0.077)	(0.048)	(0.073)
[0,1] Prior Big Tech funding	0.297***	0.003	-0.119**	0.075	-0.252
	(0.108)	(0.069)	(0.059)	(0.087)	(0.165)
Age (log)	0.647***	0.438***	0.333***	0.254***	-0.096***
	(0.013)	(0.009)	(0.008)	(0.022)	(0.034)
[0,1] Europe	-0.233***	-0.261***	-0.242***	-0.458***	0.260***
	(0.045)	(0.051)	(0.051)	(0.030)	(0.045)
[0,1] Asia	-0.168***	0.016	0.086	-0.069*	0.093
	(0.059)	(0.068)	(0.067)	(0.040)	(0.065)
[0,1] Founder OEM/HW exp.	-0.192**	0.075	0.022	-0.095	-0.006
	(0.076)	(0.070)	(0.063)	(0.061)	(0.098)
[0,1] Founder MBA	0.186***	0.127***	0.083***	0.075**	-0.331***
	(0.041)	(0.034)	(0.030)	(0.032)	(0.057)
[0,1] Founder technical degree	0.199***	-0.037	-0.019	0.232***	0.078*
	(0.035)	(0.030)	(0.027)	(0.027)	(0.044)
Observations	19679	17762	17762	19679	

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using firm-year level data and a time series probit (average effect) model, comparing startups that use certain cloud platforms. Model (1) is the base for the Heckman selection model calculation of the inverse of the Mill's ratio and includes all startups. Models (2) and (3) include only the treatment group (17,762 observations), whereas model (4) used a multinomial probit specification, including the base case Outsource/before (0); Big (1); Other (2).

Table 5.A. – IV (first stage)

	IV
DV is:	Outsource
[0,1] Tensor	0.24***
	(0.018)
[0,1] AI	0.08***
	(0.015)
Tensor x AI	0.29***
	(0.024)
Observations	18802
Firms	3123
K-P Wald F	102
C-D Wald F	70
K-P LM	105

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients in the first stage are estimated using OLS with standard errors clustered at the firm level. Tensor is an indicator variable for if the open source release of TensorFlow benefits the startup (i.e., the startup adopted Google GCP in 2016 or Amazon AWS in 2017). AI is an indicator variable for if the startup develops a commercial AI product. This model includes an interaction for AI and Tensor to capture the firms that benefit the most (i.e., those that use TensorFlow in AI development.)

Table 5.B. – IV (second stage) and DML

	(1)	(2)	(3)	(4)	(5)	(6)
DV is:	Development dissimilarity			Analytics dissimilarity		
Model:	Base	IV	DML	Base	IV	DML
[0,1] Outsource	-0.028***	-0.051***	-0.015***	0.087***	0.089***	0.053***
	(0.002)	(0.016)	(.002)	(0.005)	(0.018)	(0.004)
Observations	18802	18802	18802	18802	18802	18802
Firms	3123	3123	3123	3123	3123	3123
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CEM Weighted	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients in the second stage are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models include Coarsened exact matching (CEM) based on age #10, employment size #10, healthcare, financial services, and region #4), dropping 21 unmatched firms. Model (1) is the main result repeated from Table 3 model (4), and model (4) is the main result repeated from Table 3 model (8); Models (2) and (5) use the first stage estimation in Table 5.B. above as an instrument to adjust coefficients. Models (3) and (6) use a double machine learning approach (DML) to estimate coefficients with ~65 potentially omitted variables.

Table 6 – Startup differentiation

	(1)	(2)	(3)
DV is:	Startup description similarity		
[0,1] Outsource	0.016***	-0.015	0.004
	(0.004)	(0.022)	(0.008)
Outsource x		0.005	
Dev. dissimilarity		(0.025)	
Outsource x			0.027***
Ana. dissimilarity			(0.009)
Observations	382	382	382
R2	0.0698	0.0694	0.0692
Firms	191	191	191
Firm FE	No	No	No
Year FE	Yes	Yes	Yes

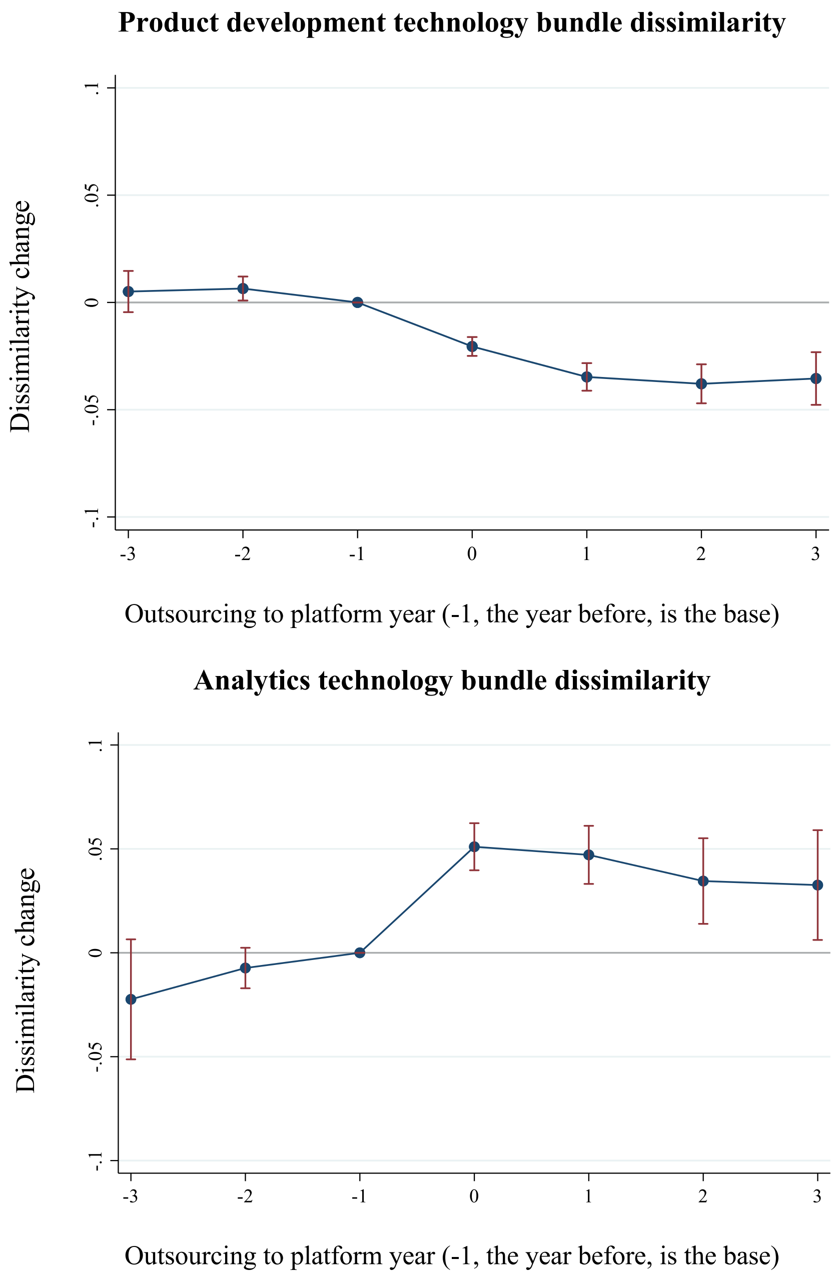
Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with year-level fixed effects in all models; standard errors are clustered at the firm level. In models (1)-(3), the panel is balanced, with one observation for each startup before and after outsourcing. Models (3) includes an interaction between outsourcing to a cloud platform and a continuous measure of analytics dissimilarity.

Table 7 - Technology dissimilarity and performance outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Funded		Follow-on Funding		Deal Size (log)		Web Visit Dur. (log)	
Development	-0.796***		-0.888***		-3.306***		-3.558***	
dissimilarity	(0.068)		(0.073)		(1.215)		(0.484)	
Analytics		0.166***		0.124***		2.309***		-0.250
dissimilarity		(0.043)		(0.045)		(0.711)		(0.342)
Observations	17628	17628	17628	17628	17628	17628	9200	9200
R2	0.691	0.686	0.690	0.685	0.165	0.165	0.428	0.422
Firms	2799	2799	2799	2799	2799	2799	2053	2053
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

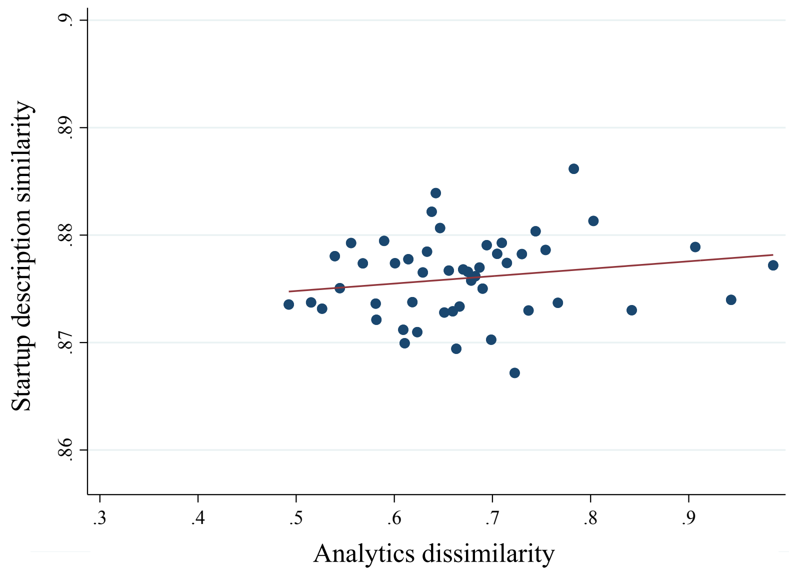
Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. Includes only the treatment group, startups using cloud platform services.

Figure 1 – Cloud platform adoption event



Coefficients are estimated using Chaisemartin and D'Haultfoeuille (2020) to account for issues arising in a two-way fixed-effect design that does not differentiate between observations that have never been treated or have not yet been treated. Each point is the coefficient of the effect based on "switchers" in a given year. These estimates are robust to dynamic effects (#5) and do not display parallel trends (#4). Standard errors are clustered at the firm level and bootstrapped (#50). Models drop and weight regressions based on CEM: age #10, employment size #10, healthcare, financial services, and region #4. The adoption event is year 0; the base period for the regressions is the year prior to the adoption, -1.

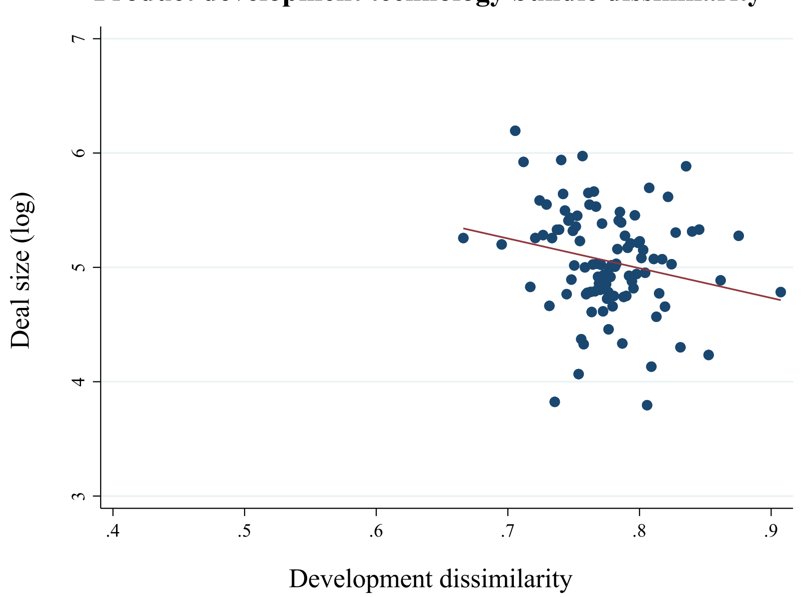
Figure 2 – Relationship between analytics dissimilarity and startup differentiation



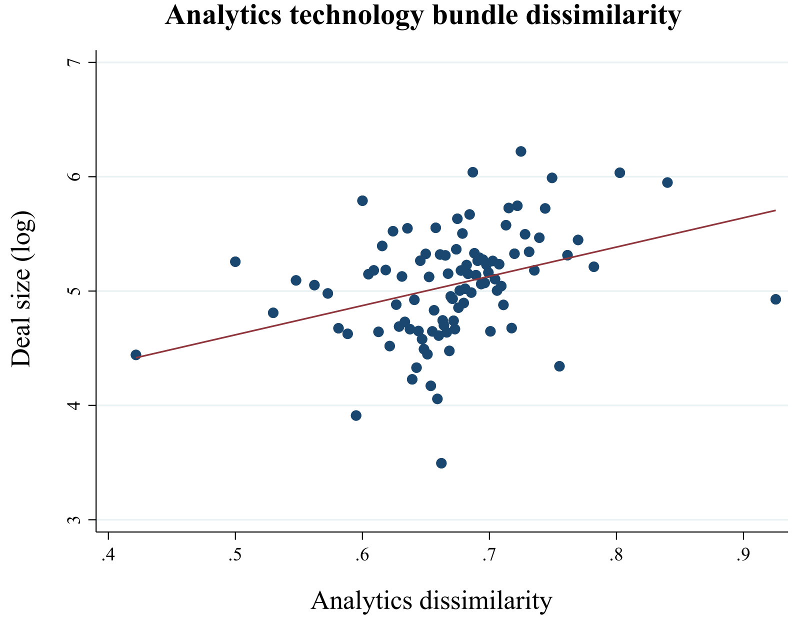
These charts use a bin scatter (50 points) with year fixed effects and coarsened exact matching (CEM), to depict the correlation before and after the treatment. In the first scenario where startups do not use a cloud platform, consider a profit function: $\Pi = (b_1 count_{dev} - c_1 dissimilarity_{dev}) + (b_2 count_{ana} - c_2 dissimilarity_{ana})$; constrained such that the $b_1 < b_2$ and $c_1 > c_2$. In the scenario where startups use a cloud platform, consider the profit function: $\Pi = (b_1 count_{dev} - c_1 dissimilarity_{dev}) + (b_2 count_{ana} - c_2 dissimilarity_{ana}) - (d_3 dissimilarity_{dev}(market\ share_{platform}))$; constrained such that the $b_1 < b_2$ and $c_1 > c_2$.

Figure 3 – Technology dissimilarity and VC Funding

Product development technology bundle dissimilarity



Analytics technology bundle dissimilarity



To visualize the regression specification, these charts use a bin scatter (100 points) residualized on firm and year effects with CEM matching.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*(1), 1-19.

Aghion, P., & Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics*, 109(4), 1185-1209.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, 120(2), 701-728.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, *45*(3), 425-455.

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, 58(5), 913-931.

Argyres, N. S., & Zenger, T. R. (2012). Capabilities, transaction costs, and firm boundaries. *Organization Science*, 23(6), 1643-1657.

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. University of Michigan Press.

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1), 62-79.

Athey, S., & Schmutzler, A. (1995). Product and process flexibility in an innovative environment. *The RAND Journal of Economics*, 557-574.

Balakrishnan, S., & Wernerfelt, B. (1986). Technical change, competition and vertical integration. *Strategic Management Journal*, 7(4), 347-359.

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, 32(10), 1231-1241.

Baumol, W. J. (2001). When is inter-firm coordination beneficial? The case of innovation. *International Journal of Industrial Organization*, *19*(5), 727-737.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29-50.

Benzina, K. (2019). Cloud Infrastructure-as-a-Service as an Essential Facility: Market Structure, Competition, and the Need for Industry and Regulatory Solutions. *Berkeley Tech. LJ*, *34*, 119.

Berman, R., & Israeli, A. (2022). The Value of Descriptive Analytics: Evidence from Online Retailers. *Marketing Science*.

Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. (2018). The Business of AI Startups. Boston University School of Law, Law and Economics Research Paper, (18-28).

Bessen, J., Impink, S. M., Reichensperger, L., & Seamans, R. (2022). The role of data for AI startup growth. *Research Policy*, 51(5), 104513.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans do IT better: US multinationals and the productivity miracle. *American Economic Review*, 102(1), 167-201.

Bloom, N., & Pierri, N. (2018). Cloud computing is helping smaller, newer firms compete. *Harvard Business Review*, 94(4).

Boudreau, K. J. (2012). Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. *Organization Science*, *23*(5), 1409-1427.

Bresnahan, T., Greenstein, S., Brownstone, D., & Flamm, K. (1996). Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers on Economic Activity*, Microeconomics, 1-83.

Brynjolfsson, E., Jin, W., & McElheran, K. (2021). The power of prediction: predictive analytics, workplace complements, and business performance. *Business Economics*, 56(4), 217-239.

Brynjolfsson, E., Malone, T. W., Gurbaxani, V., & Kambil, A. (1994). Does information technology lead to smaller firms?. *Management Science*, 40(12), 1628-1644.

Bun, M. J., & Harrison, T. D. (2019). OLS and IV estimation of regression models including endogenous interaction terms. *Econometric Reviews*, 38(7), 814-827.

Cachon, G. P., & Harker, P. T. (2002). Competition and outsourcing with scale economies. *Management Science*, 48(10), 1314-1333.

Caves, R. E. (2000). *Creative industries: Contracts between art and commerce* (No. 20). Harvard University Press.

Chatterji, A. K., & Fabrizio, K. R. (2014). Using users: When does external knowledge enhance corporate product innovation?. *Strategic Management Journal*, 35(10), 1427-1445.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal*, 21(1).

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, 41(8), 1381-1411.

Coase, R. H. (1937). The nature of the firm, 18-33.

Combs, J. G., & Ketchen, Jr, D. J. (1999). Explaining interfirm cooperation and performance: toward a reconciliation of predictions from the resource-based view and organizational economics. *Strategic Management Journal*, 20(9), 867-888.

Cowgill, B., & Tucker, C. E. (2019). Economics, fairness, and algorithmic bias. SSRN Working Paper.

Cunningham, C., Ederer, F., & Ma, S. (2021). Killer acquisitions. *Journal of Political Economy*, *129*(3), 649-702.

De Chaisemartin, C., & d'Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2), 999-1028.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964-96.

DeStefano, T., Kneller, R., & Timmis, J. (2020). Cloud computing and firm growth. SSRN Working Paper.

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

Fazli, A., Sayedi, A., & Shulman, J. D. (2018). The effects of autoscaling in cloud computing. *Management Science*, 64(11), 5149-5163.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7), 1019-1039.

Furman, J., & Seamans, R. (2019). AI and the Economy. *Innovation Policy and the Economy*, 19(1), 161-191.

Gans, J. S., Stern, S., & Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5), 736-756.

Gartner Inc. (2018). Research Report: Market Share Analysis: IaaS and IUS, Worldwide, 2017. https://www.gartner.com/en/newsroom/press-releases/2018-08-01-gartner-says-worldwide-iaas-public-cloud-services-market-grew-30-percent-in-2017

Giustiziero, G. (2021). Is the division of labor limited by the extent of the market? Opportunity cost theory with evidence from the real estate brokerage industry. *Strategic Management Journal*, 42(7), 1344-1378.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3-43.

Greenstein, S. M. (1993). Did installed base give an incumbent any (measurable) advantages in federal computer procurement?. *The RAND Journal of Economics*, 19-39.

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

Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly*, 619-652.

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), 212-43.

Hagedoorn, J., & Schakenraad, J. (1990). Inter-firm partnerships and cooperative strategies in core technologies. *New explorations in the Economics of Technical Change*, 3, 9-13.

Hagiu, A. (2009). Two-sided platforms: Product variety and pricing structures. *Journal of Economics & Management Strategy*, *18*(4), 1011-1043.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

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

Hochberg, Y. V. (2016). Accelerating entrepreneurs and ecosystems: The seed accelerator model. *Innovation Policy and the Economy*, *16*(1), 25-51.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1-24.

Iansiti, M. (2021). The value of data and its impact on competition. SSRN Working Paper.

Jin, W., & McElheran, K. (2019). Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing. Rotman School of Management Working Paper, (3112901).

Katila, R., Piezunka, H., Reineke, P., & Eisenhardt, K. M. (2022). Big fish versus big pond? Entrepreneurs, established firms, and antecedents of tie formation. *Academy of Management Journal*, 65(2), 427-452.

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

Khan, L. M. (2016). Amazon's antitrust paradox. *Yale Law Journal*, 126, 710.

Klein, B. (1980). Transaction cost determinants of "unfair" contractual arrangements. *The American Economic Review*, 70(2), 356-362.

Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and Start-up Performance: Evidence from A/B Testing. *Management Science*.

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*.

Lacity, M. C., Khan, S. A., & Willcocks, L. P. (2009). A review of the IT outsourcing literature: Insights for practice. *The Journal of Strategic Information Systems*, 18(3), 130-146.

Lerner, J., & Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives*, 34(3), 237-61.

Madhok, A., & Tallman, S. B. (1998). Resources, transactions and rents: Managing value through interfirm collaborative relationships. *Organization Science*, *9*(3), 326-339.

Miric, M., Ozalp, H., & Yilmaz, E. D. (2021). Tradeoffs to Using Standardized Tools: An Innovation Enabler or Creativity Constraint?. USC Marshall School of Business Research Paper.

Monteverde, K., & Teece, D. J. (1982). Supplier switching costs and vertical integration in the automobile industry. *The Bell Journal of Economics*, 206-213.

Montgomery, C. A. (1994). Corporate diversification. *Journal of Economic Perspectives*, 8(3), 163-178.

Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1998). Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. *Research Policy*, *27*(5), 507-523.

Nagle, F., Seamans, R., & Tadelis, S. (2020). Transaction Cost Economics in the Digital Economy: A Research Agenda. Harvard Business School Strategy Unit Working Paper, (21-009).

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. *Probability and Statistics*, 213-234.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187-204.

Parker, G., Van Alstyne, M. W., & Jiang, X. (2016). Platform ecosystems: How developers invert the firm. Boston University Questrom School of Business Research Paper, (2861574).

Peteraf, M. A. (1993). The cornerstones of competitive advantage: a resource-based view. *Strategic Management Journal*, 14(3), 179-191.

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations: A Resource Dependence Perspective*. Stanford University Press.

Poppo, L., & Zenger, T. (2002). Do formal contracts and relational governance function as substitutes or complements?. *Strategic Management Journal*, 23(8), 707-725.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 116-145.

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc.

Riordan, M. H., & Williamson, O. E. (1985). Asset specificity and economic organization. *International Journal of Industrial Organization*, 3(4), 365-378.

Rivkin, J. W. (2000). Imitation of complex strategies. *Management Science*, 46(6), 824-844.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990-1029.

Rock, D. (2019). Engineering value: The returns to technological talent and investments in artificial intelligence. SSRN Working Paper 3427412.

Rysman, M., & Simcoe, T. (2008). Patents and the performance of voluntary standard-setting organizations. *Management Science*, *54*(11), 1920-1934.

Schilling, M. (1999). Winning the standards race: Building installed base and the availability of complementary goods. *European Management Journal*, *17*(3), 265-274.

Schneider, S., & Sunyaev, A. (2016). Determinant factors of cloud-sourcing decisions: reflecting on the IT outsourcing literature in the era of cloud computing. *Journal of Information Technology*, 31(1), 1-31.

Scott Morton, F., Bouvier, P., Ezrachi, A., Jullien, B., Katz, R., Kimmelman, G., ... & Morgenstern, J. (2019). Committee for the study of digital platforms: Market structure and antitrust subcommittee report. Chicago: Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business.

Seamans, R., & Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science*, 60(2), 476-493.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423.

Siggelkow, N. (2001). Change in the presence of fit: The rise, the fall, and the renaissance of Liz Claiborne. *Academy of Management Journal*, 44(4), 838-857.

Silverman, B. S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science*, 45(8), 1109-1124.

Simcoe, T. (2012). Standard setting committees: Consensus governance for shared technology platforms. *American Economic Review*, *102*(1), 305-36.

Stuart, T. E. (2000). Interorganizational alliances and the performance of firms: a study of growth and innovation rates in a high-technology industry. *Strategic Management Journal*, *21*(8), 791-811.

Stuart, T. E., Hoang, H., & Hybels, R. C. (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44(2), 315-349.

Susarla, A., Barua, A., & Whinston, A. B. (2009). A transaction cost perspective of the" software as a service" business model. *Journal of Management Information Systems*, 26(2), 205-240.

Sweeting, A. (2010). The effects of mergers on product positioning: evidence from the music radio industry. *The RAND Journal of Economics*, *41*(2), 372-397.

Tadelis, S., & Williamson, O. E. (2012). *Transaction cost economics*. The Handbook of Organizational Economics, 159.

Tambe, P. (2014). Big data investment, skills, and firm value. Management Science, 60(6), 1452-1469.

Teece, D. J. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization*, 3(1), 39-63.

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600-1631.

Tucker, C. (2019). Digital Data, Platforms and the Usual [Antitrust] Suspects: Network effects, Switching Costs, Essential facility. *Review of Industrial Organization*, 54(4), 683-694.

Walker, G., & Weber, D. (1984). A transaction cost approach to make-or-buy decisions. *Administrative Science Quarterly*, 373-391.

Wang, R. D., & Shaver, J. M. (2014). Competition-driven repositioning. *Strategic Management Journal*, *35*(11), 1585-1604.

Wen, W., Forman, C., & Jarvenpaa, S. L. (2022). The effects of technology standards on complementor innovations: Evidence from the IETF. *Research Policy*, 51(6), 104518.

Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. *The Journal of Law and Economics*, *22*(2), 233-261.

Williamson, O. E. (1983). Credible commitments: Using hostages to support exchange. *The American Economic Review*, 73(4), 519-540.

Williamson, O. E. (1998). The institutions of governance. *The American Economic Review*, 88(2), 75-79.

Wooldridge, J. M. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics*, 47(1), 5-46.

Yang, H., & Tate, M. (2012). A descriptive literature review and classification of cloud computing research. *Communications of the Association for Information Systems*, 31(1), 2.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—the new organizing logic of digital innovation: an agenda for information systems research. *Information Systems Research*, 21(4), 724-735.

Young-Ybarra, C., & Wiersema, M. (1999). Strategic flexibility in information technology alliances: The influence of transaction cost economics and social exchange theory. *Organization Science*, 10(4), 439-459.

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

Zingales, L., Kamepalli, S. K., & Rajan, R. G. (2021). Kill Zone. CEPR Discussion Series.

Zittrain, J. (2006). *Internet governance and regulation: the future of the internet and how to stop it*. Yale University Press.

Appendices

Appendix A

Note A.1. – BuiltWith technology data

BuiltWith provides the name of the technology and the category of that technology. I create a new set of categories that include relevant development technologies from these categories. Here is a list of all technology categories, many of which are “front end” or administrative, unrelated to product development. These omitted technologies include the following BuiltWith categories: Accounting, Ads, Collaboration, Communications, Content Management, Content Marketing, CRM, Demand Generation, Design, Digital Marketing, E-commerce, Email Hosting, Finance, Hiring, Marketing Automation, Payments, Product Management, Productivity, Sales, SEO and Search Marketing, SEO Headers, Servers, Shopping, Web Hosting, Web Server, Workforce Management Additionally, I drop technologies related:

- Languages (e.g., French, Spanish, English, etc.)
- Error messages (i.e., common name invalid, domain not resolving)
- Schema

From the BuiltWith data, after the cleanups, I have the following categories of “backend” technologies: 1) Product development technologies: developer frameworks (API, developer tools, DevOps, and programming languages), security, content delivery network, and 2) Analytics technologies. I list and describe the technologies used in this study in Appendix D.

Note A.2. – Measures descriptions

Technology adoption

- *All Technologies* is a measure of any technologies connected to the startup’s domain, including front-end and back-end technologies.
- *Product Development Technologies* are backend, data infrastructure technologies (e.g., Content Management Systems, Content Delivery Networks, Frameworks, Security, and) that are core to product development
- *Analytics Technologies* are backend data collection and analysis technologies that are core to accruing and repurposing data.
- *Big Tech Technologies* are technologies providers by Amazon, Google, or Microsoft.
- *Paid/Subscription Technologies* are proprietary technologies that startups way a royalty to access, based on information provided by BuiltWith.
- *Open source Technologies* are freely available technologies that startups can adapt and customize. These technologies are described as open source in their description in BuiltWith.

Startup descriptions

I omit words used infrequently (i.e., proper nouns) or very often (e.g., the, and, but, or, he, she, it, etc.), removing "outliers" at the 5% and 95% level. I then stem words, remove numbers, punctuation, hyphens, and web addresses, and tokenize the counts of the analyzed words. This text is then vectorized by word, creating a sparse matrix: 0 if the word is not shared in a pairwise match; 1 if the word is shared. I calculate the angular distance in the same manner as above using specification (1). Mean startups description differentiation is 0.83 (SD 0.04)

Tables A.3.-A.5. – Correlations

Table A.3. – Correlation (demographics)

		(1)	(2)	(3)	(4)	(5)
(1)	Dev. dissimilarity	1				
(2)	Ana. dissimilarity	-0.21*	1			
(3)	Age	-0.40*	0.25*	1		
(4)	Employees	-0.090*	0.073*	0.067*	1	
(5)	Americas	-0.0012	0.11*	0.013+	0.018*	1
(6)	EU	-0.016*	-0.067*	-0.0066	-0.065*	-0.68*

Table A.4. – Correlation (technologies)

		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Dev. dissimilarity	1					
(2)	Ana. dissimilarity	-0.21*	1				
(3)	Dev. Tech Ct.	-0.56*	0.25*	1			
(4)	Data Tech Ct.	-0.47*	0.32*	0.65*	1		
(5)	IaaS	-0.25*	0.22*	0.32*	0.29*	1	
(6)	AI	0.24*	-0.12*	-0.11*	-0.11*	-0.071*	1
(7)	Tensor	0.070*	-0.018*	-0.013+	-0.021*	0.010	0.034*

Table A.5. – Correlation (funding)

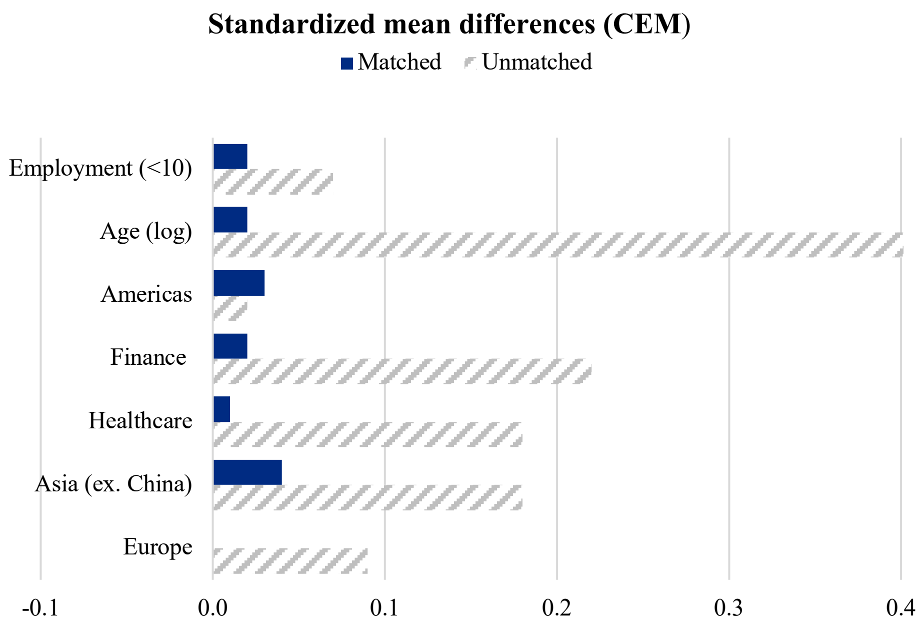
		(1)	(2)	(3)	(4)	(5)
(1)	Dev. dissimilarity	1				
(2)	Data dissimilarity	-0.21*	1			
(3)	Funds raised (log)	-0.21*	0.18*	1		
(4)	Funded	-0.19*	0.16*	0.79*	1	
(5)	VC backed	-0.21*	0.15*	0.68*	0.79*	1
(6)	Investor dissimilarity	-0.12*	0.078*	0.22*	0.078*	0.15*

Notes: + p<0.10; * p<0.05

Note A.6. – CEM matching procedure

I use the *cem* package from STATA to match and weight startups based on the following criteria:

- Age; 10 quantiles
- Employee; 10 quantiles
- Regions (4; Asia, Americas, EU, MEA)
- Healthcare (0,1)
- Finance (0,1)



Note A.7. – Heckman selection and the inverse Mill's ratio

To control for this potential selection issue, I use Heckman's selection equation to calculate the inverse of the Mill's ratio (IMR) to control regression from a probit regression comparing two outcomes (i.e., the decision to add a cloud platform versus not, or the decision to add a certain cloud platform). I use IMR as a control in the second stage of the regression to address potential selection issues.

$$outsource_i = w_i\gamma + \varepsilon_i \qquad \text{(Z.1) [Selection equation]}$$

$$\lambda = \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)} \qquad \text{(Z.2) [inverse Mill's ratio]}$$

where, $outsource_i$ refers to an indicator variable that takes the value 1 based on the treatment outcome (CSP, Other CSP, Amazon) at the observation level. For example, when comparing the impact of using a CSP, the variable takes the value 1 if a firm uses a CSP, otherwise 0. w_i is a vector of demographic (e.g., industry, age), funding (e.g., prior Big Tech funding), and founder (e.g., hardware or IT work experience, technical education) indicator variables plausibly correlated with adoption.

Note A.8. – Instrumental variable research design

Below is the first-stage regression equation of the instrumental variable:

$$outsource_{it} = \beta_1 tensor_{it} + \beta_2 AI_{it} + \beta_3(tensor_{it} \times AI_{it}) + \varepsilon_{it} \quad (Z.3)$$

where, $outsource_{it}$ refers to the binary dependent variable: adopting cloud versus not, $tensor_{it}$ refers to an indicator variable take the value 0 if there is no TensorFlow benefit and 1 if there is a Tensor Flow benefit (i.e., startup adopts Google's platform in 2016 or adopts Amazon AWS platform in 2017), AI_{it} refers to an indicator variable that takes the value 1 if the firm develops commercial AI and 0 otherwise, and $tensor_{it} \times AI_{it}$ is the interaction between the two binary variables.

Note A.9. – Double machine learning with orthogonalization

I use double machine learning following Chernozhukov et al. (2018) to examine the causal parameter θ, a scalar that adjusts the regression coefficient, by using a random forest machine-learning algorithm.[60] Specification Z.2 is the main predictive model, and specification Z.3 constructs Neyman orthogonal scores (Chernozhukov et al., 2018; Neyman, 1959; Wooldridge, 1991).

$$tech\_diff_{it} = \theta outsource_{it} + g_0(x_{it}) + year_t + \varsigma_{it} \quad \text{(Z.4)}$$

$$outsource_{it} = m_0\,(x_{it}) + year_t + v_{it} \quad \text{(Z.5)}$$

where $year_t$ is an indicator variable for all observed years (2012 through 2021); $outsource_{it}$ refers to an indicator variable that takes the value 1 if the firm adopts the cloud platform and 0 otherwise; x_{it} is a vector of covariates error terms; ς_{it} and v_{it} are normally distributed (0,1) error terms.

This approach uses orthogonalization to overcome regularization biases (i.e., issues associated with overfitting the model). The sample is initially randomly ordered, and then 50% of the sample is used from training and the remaining 50% for prediction. In the model, I include every possible covariate from the data The algorithm determines which of those variables should be added to the model. I calculate the differences between the true parameter and test estimates resulting from specifications (Z.4) and (Z.5), adjusting the main specification (2) to estimate the coefficient of interest β_1.

$$(tech\_diff_{it} - \widehat{tech\_diff}_{it}) = \beta_1(outsource_{it} - \widehat{outsource}_{it}) + \beta_2 yearFE_t + \beta_3 firmFE_i + \varepsilon_{it} \quad \text{(Z.6)}$$

Note A.10. – Average install base of technology bundles dissimilarity measure

The average install base of the technology bundle is a measure firm-level average of the technology-level average number of startups in my sample using each technology.

$$AIB_{it} = \frac{\sum_{j=1}^{x} rival\_count_{jt}}{x} \quad \text{(Z.7)}$$

Where, *i* takes on the value of the firm id, *t* takes on the value of the year, *j* takes on the value of the technology, *x* is the total count of technologies for each firm in year *t*, and $rival\_count_{jt}$ is the total number of users by technology-year in the sample.

[60] I use the *rforest* package in STATA with 100 iterations, minimum leaf sized adjusted to 10.

Appendix B – Additional analyses and robustness

Table B.1. – Other technology counts

	(1)	(2)	(3)	(4)	(5)	(6)
DV is log of :	All	Big	Ratio: Big/All	Open	Premium	Ratio: Open/Prem
[0,1] Outsource	0.249***	0.308***	0.119***	0.329***	0.021***	-0.003**
	(0.021)	(0.019)	(0.020)	(0.024)	(0.004)	(0.001)
Observations	18802	18802	18802	18802	18802	18802
R2	0.634	0.647	0.571	0.729	0.549	0.500
Firms	3123	3123	3123	3123	3123	3123
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CEM Weighted	Yes	Yes	Yes	No	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS; standard errors are clustered at the firm level. All models drop and weight regressions based on CEM: age #10, employment size #10, healthcare, financial services, and region #4.

Table B.2. – Combined robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Development dissimilarity				Analytics dissimilarity			
	Balanced	bf. 2018	Inv. diff.	IMR	Balanced	bf. 2018	Inv. diff.	IMR
[0,1] Outsource	-0.027***	-0.026***	-0.033***	-0.025***	0.092***	0.082***	0.079***	0.086***
	(0.002)	(0.003)	(0.004)	(0.002)	(0.005)	(0.006)	(0.008)	(0.005)
Investor differentiation			-0.017				0.022	
			(0.012)				(0.021)	
Inv. Mill's ratio				0.089***				-0.008
				(0.007)				(0.012)
Observations	17045	8445	9319	18802	17045	8445	9319	18802
R2	0.680	0.683	0.698	0.680	0.548	0.639	0.554	0.577
Firms	2779	2198	1799	3123	2779	2198	1799	3123
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CEM Weighted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models are weighted using CEM.

Table B.3. – Platform subsamples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Development dissimilarity				Analytics dissimilarity			
[0,1] AM	-0.029***				0.074***			
	(0.003)				(0.007)			
[0,1] GCP		-0.026***				0.086***		
		(0.006)				(0.012)		
[0,1] MS			-0.008				0.082***	
			(0.010)				(0.014)	
[0,1] Other				-0.021***				0.090***
				(0.005)				(0.012)
Observations	18802	18802	18802	18802	18802	18802	18802	18802
R2	0.671	0.668	0.667	0.667	0.562	0.553	0.551	0.554
Firms	3123	3123	3123	3123	3123	3123	3123	3123
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models are weighted using CEM. Only includes startups that use a single cloud platform provider.

Table B.4. – Alternative DV: Average installed base

	(1)	(2)	(3)	(5)	(6)	(7)
DV is:	Development tech avg. IB			Analytics tech avg. IB		
[0,1] Outsource	0.089***	0.090***	0.071***	-0.143***	-0.100***	-0.161***
	(0.020)	(0.022)	(0.024)	(0.024)	(0.025)	(0.031)
[0,1] H. size bundle		0.006			-0.282***	
		(0.016)			(0.031)	
Outsource x		0.092***			-0.351***	
H. size bundle		(0.021)			(0.027)	
[0,1] IaaS			0.011			0.005
			(0.033)			(0.044)
Outsource x			0.097***			-0.136***
IaaS			(0.021)			(0.026)
Observations	18802	18802	18802	18802	18802	18802
R2	0.818	0.818	0.818	0.702	0.715	0.702
Firms	3123	3123	3123	3123	3123	3123
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CEM Weighted	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models are weighted using CEM.

Table B.5. – Industry Subsamples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Development dissimilarity				Analytics dissimilarity			
	AI	ML	Financial	Healthcare	AI	ML	Financial	Healthcare
[0,1] Outsource	-0.030***	-0.026***	-0.020***	-0.031***	0.084***	0.075***	0.058***	0.089***
	(0.004)	(0.008)	(0.007)	(0.009)	(0.008)	(0.015)	(0.019)	(0.014)
Observations	6872	1651	1574	1719	6872	1651	1574	1719
R2	0.666	0.723	0.688	0.651	0.567	0.561	0.505	0.538
Firms	1239	319	268	313	1239	319	268	313
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models are weighted using CEM.

Table B.6. – City level subsamples

	(1)	(2)	(3)	(4)	(5)	(6)
DV is:	Development dissimilarity			Analytics dissimilarity		
	San Francisco	London	New York	San Francisco	London	New York
[0,1] Outsource	-0.028***	-0.034***	-0.050***	0.101***	0.130***	0.075***
	(0.006)	(0.010)	(0.008)	(0.013)	(0.024)	(0.023)
Observations	2765	866	1530	2765	866	1530
R2	0.668	0.655	0.689	0.547	0.560	0.536
Firms	452	142	244	452	142	244
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models are weighted using CEM.

Table B.7. – Staggered DiD with two-way fixed effects (robustness)

	Development dissimilarity				Analytics dissimilarity			
	Pre/Post	ATE	LATE	ITE	Pre/Post	ATE	LATE	ITE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
[0,1] Outsource	-0.086***	-0.011***	-0.014***	-0.021***	0.129***	0.036***	0.139***	0.051***
	(0.002)	(0.002)	(0.007)	(0.02)	(0.005)	(0.004)	(0.014)	(0.005)
Observations	1746	3284	19082	5539	1746	3284	19082	5539
R2	0.579	0.0058			0.380	0.0191		
Firms	873	3284	3284	873	873	3284	3284	873
Year FE	Yes	Yes	No	No	Yes	Yes	No	No
CEM Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE	Clustered (ID)	Clustered (ID)	Bootstrap	Clustered (ID)	Clustered (ID)	Clustered (ID)	Bootstrap	Clustered (ID)

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS. All models drop and weight regressions based on CEM: age #10, employment size #10, healthcare, financial services, and region #4, and all models cluster standard error at the firm level.

Table B.8. – Disaggregated data: SUTVA and focal/rival effects

	Stable Rivals	Base	Focal FE	Rival FE	Focal & Rival FE	Focal & Year FE	Rival & Year FE	Focal, Rival & Year FE
DV is:	Development dissimilarity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
[0,1] Outsource	-0.027***	-0.076***	-0.088***	-0.098***	-0.105***	-0.028***	-0.031***	-0.029***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
R2	0.716	0.0278	0.504	0.0571	0.526	0.710	0.262	0.710
	Analytics dissimilarity							
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
[0,1] Outsource	0.083***	0.114***	0.115***	0.132***	0.130***	0.088***	0.106***	0.096***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
R2	0.622	0.0469	0.577	0.0519	0.582	0.613	0.0811	0.614
Observations	8013783	34461926	34461926	34461926	34461926	34461926	34461926	34461926
Firms	3123	3123	3123	3123	3123	3123	3123	3123
Focal Firm FE	Yes	No	Yes	No	Yes	Yes	No	Yes
Rival Firm FE	No	No	No	Yes	Yes	No	Yes	Yes
Year FE	Yes	No	No	No	No	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level.

Table B.9. – Platform market share

	(1)	(2)	(3)	(4)	(5)	(6)
DV is:	Development dissimilarity			Analytics dissimilarity		
Market share	-0.077***			0.087***		
	(0.011)			(0.025)		
Big Tech		-0.016***			0.031***	
		(0.004)			(0.006)	
Not Big Tech		*base*			*base*	
AWS			-0.025***			0.009*
			(0.003)			(0.005)
Not AWS			*base*			*base*
Observations	11588	9588	7530	11588	9588	7530
R2	0.690	0.344	0.334	0.560	0.0799	0.0575
Firms	2225	2466	1609	2225	2466	1609
Firm FE	Yes	No	No	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CEM Matched	No	Yes	Yes	No	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. Models that use the treatment sample do not include CEM matching.

Table B.10. – Additional performance measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	VC backed		VC rep.		Closed		Acquired	
Development dissimilarity	-0.721***		-0.195***		0.016		0.011	
	(0.066)		(0.035)		(0.015)		(0.025)	
Analytics dissimilarity		0.115***		0.013		0.000		0.031***
		(0.041)		(0.021)		(0.008)		(0.011)
Observations	17628	17628	17628	17628	17628	17628	17628	17628
R2	0.751	0.747	0.817	0.816	0.948	0.948	0.960	0.960
Firms	2799	2799	2799	2799	2799	2799	2799	2799
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level.

Table B.11. – Funding outcomes (interaction)

DV is:	(1) Funded	(2) Follow-on Funding	(3) Deal size (log)	(4) Funded	(5) Follow-on Funding	(6) Deal size (log)
Development bundle						
L. size x Less diss.	0.053***	0.042***	-0.225			
	(0.009)	(0.010)	(0.202)			
L size x More diss.		*base*				
H. size x Less diss.	0.108***	0.121***	0.837***			
	(0.009)	(0.010)	(0.193)			
H. size x More diss.	0.094***	0.084***	0.825***			
	(0.010)	(0.011)	(0.199)			
Analytics bundle						
L. size x Less diss.					*base*	
L. size x More diss.				0.012	-0.000	0.330*
				(0.010)	(0.011)	(0.183)
H. size x Less diss				0.099***	0.101***	0.653***
				(0.011)	(0.012)	(0.231)
H size x More diss.				0.093***	0.115***	0.694***
				(0.010)	(0.010)	(0.190)
Observations	17628	17628	17628	17628	17628	17628
R2	0.691	0.689	0.166	0.690	0.690	0.165
Firms	2799	2799	2799	2799	2799	2799
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level.

Appendix C – Additional Figures

Figure C.1. – Percentage of startups (>3 years old) outsourcing IT to cloud platforms

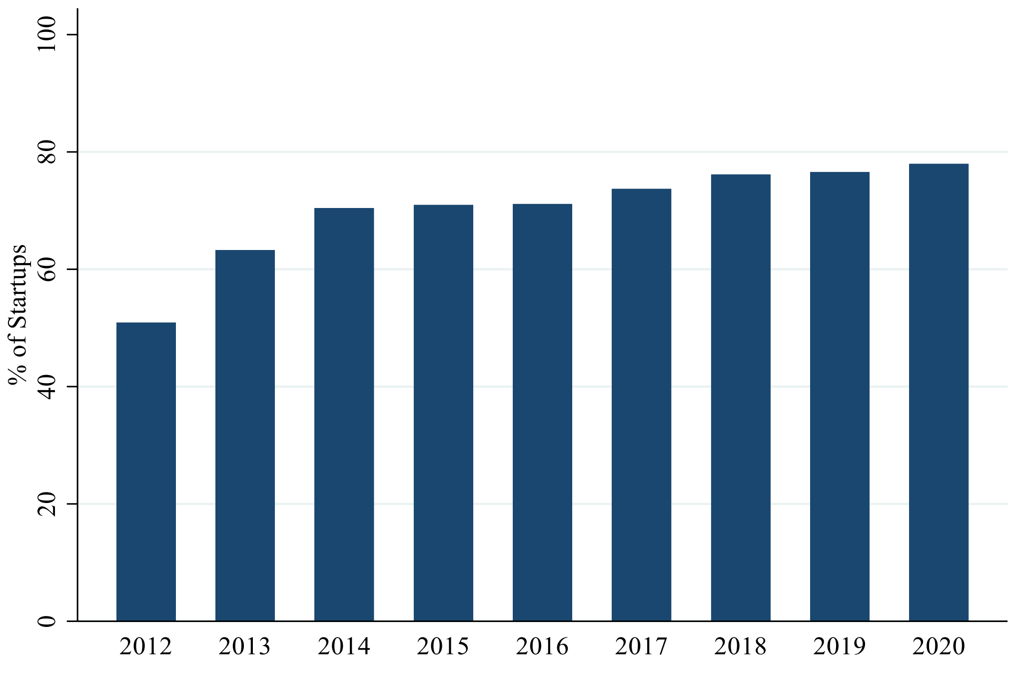


Figure C.2. – Kernel density (Outsource vs. in-house)

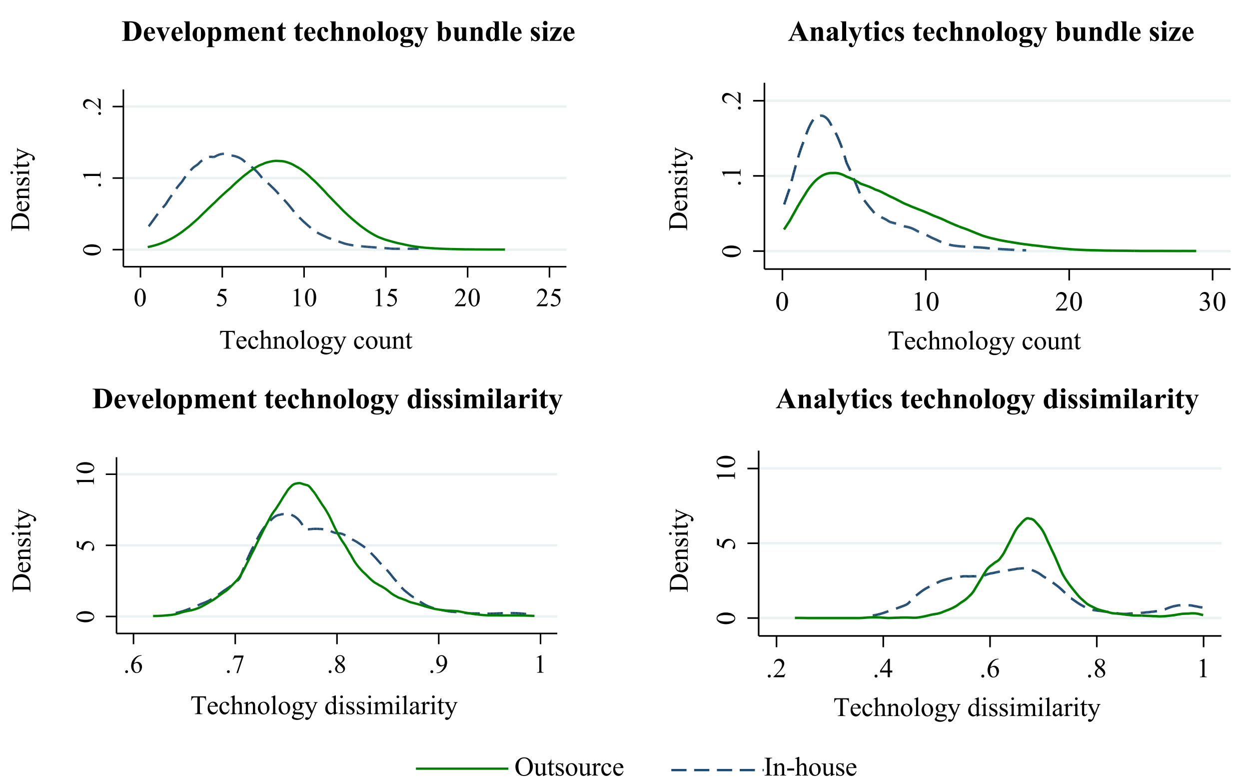


Figure C.3. – Difference in technology type correlation (Outsource vs. in-house)

In-house (Insignificant relationship) **Outsource** (Significant relationship)

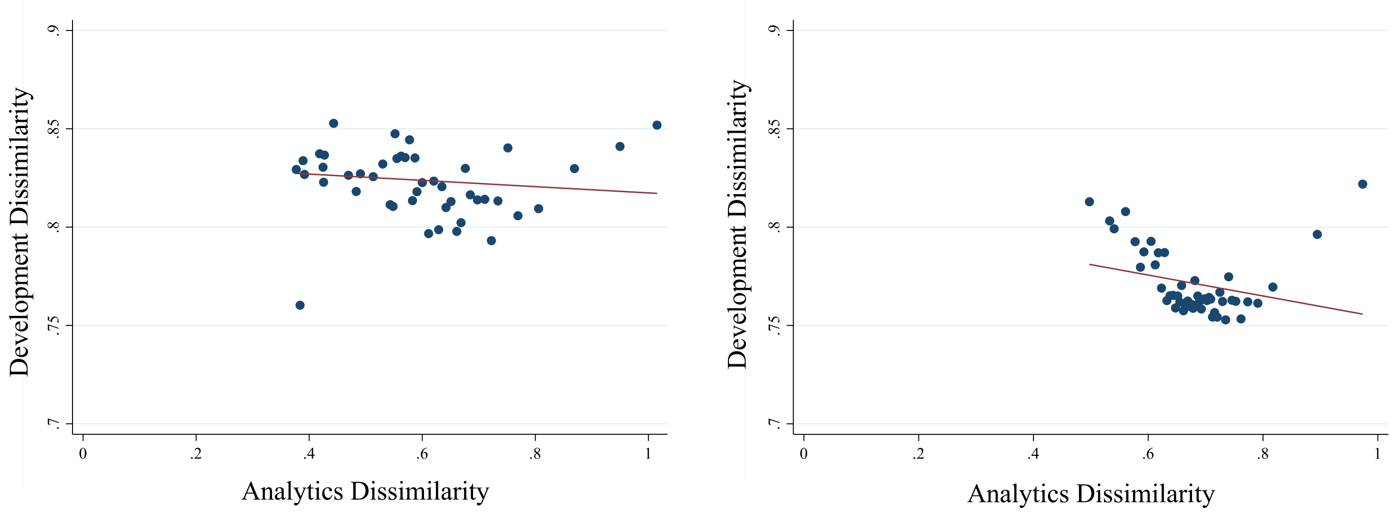


Figure C.4. – Differences in technology type trend (Outsource vs. in-house)

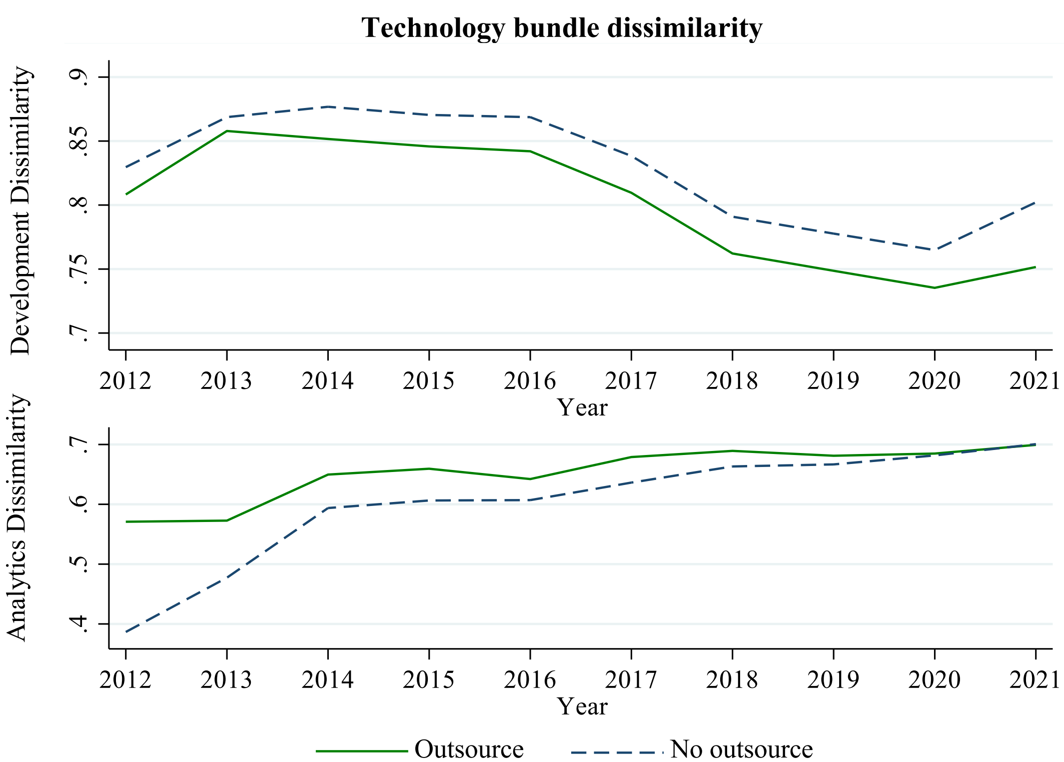
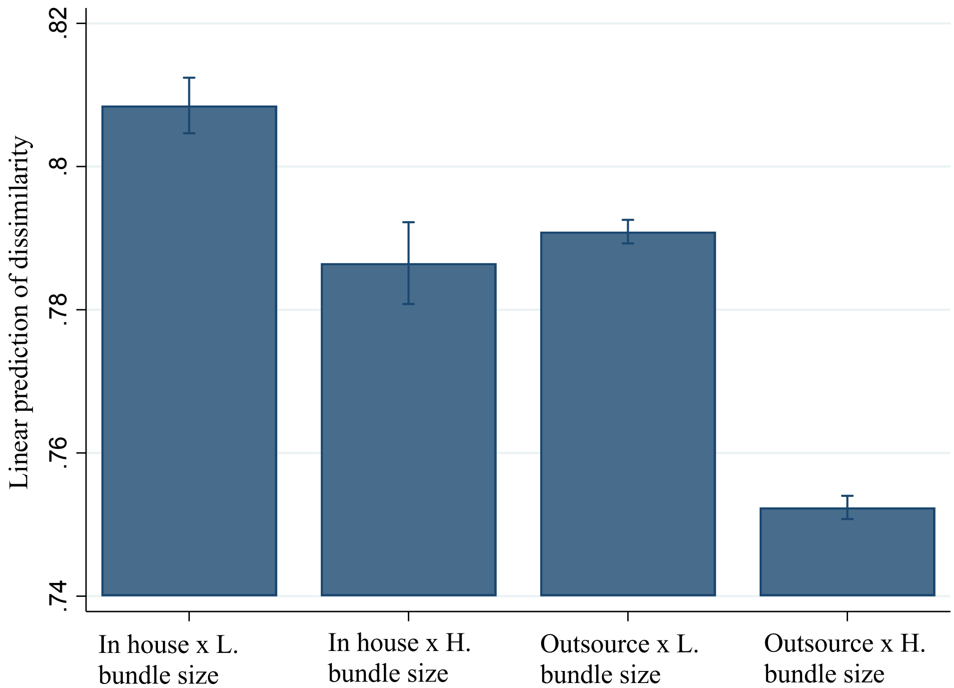


Figure C.5. – Mechanism impacting product development technology breadth

Mechanism A: Need for technological fit (Size of technology bundle)



Mechanism B: Customer-supplier relationship strength (IaaS/second cloud service)

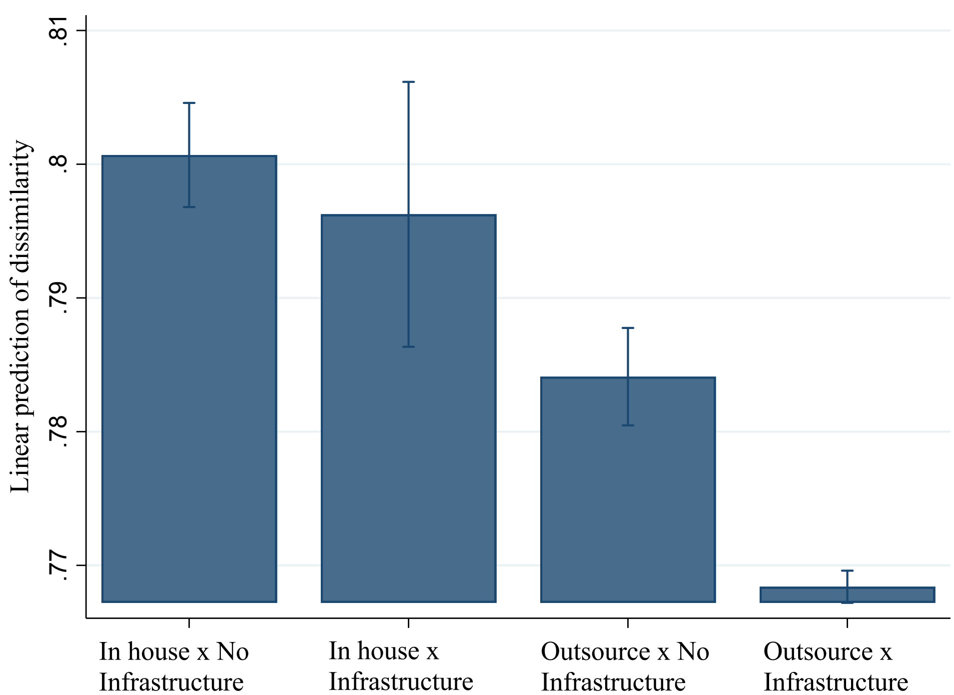
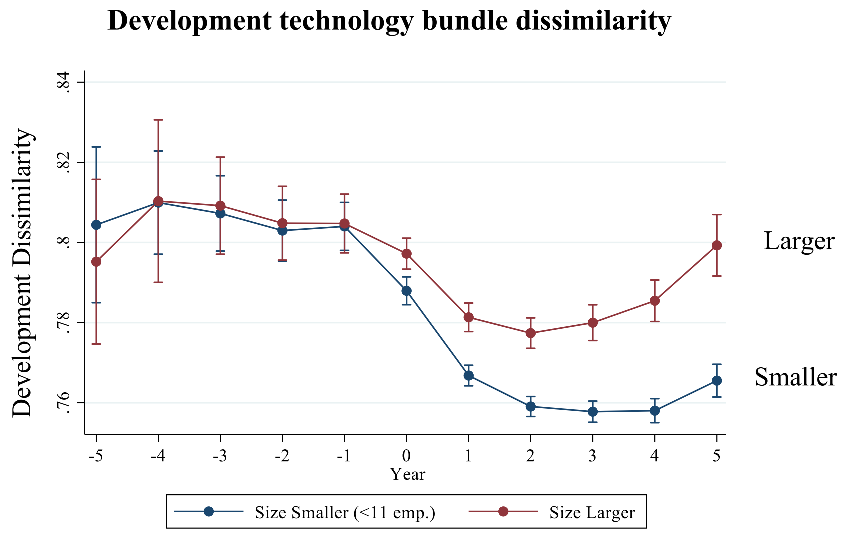
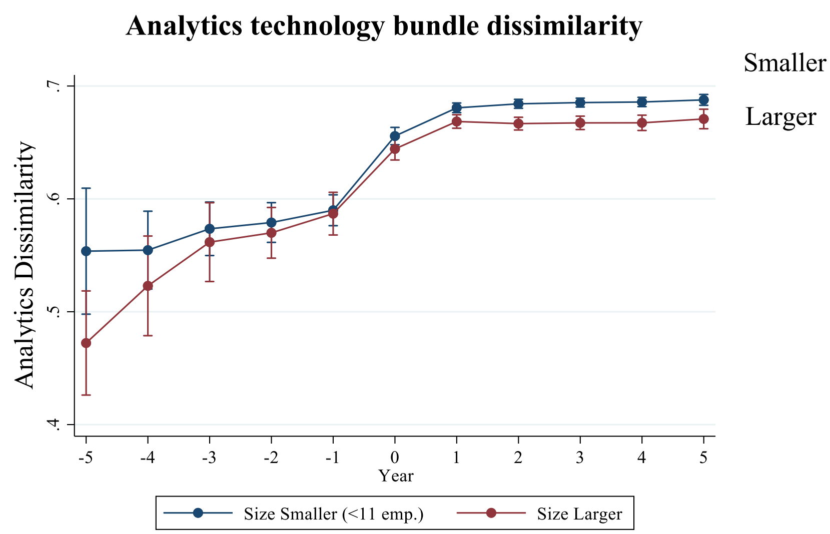


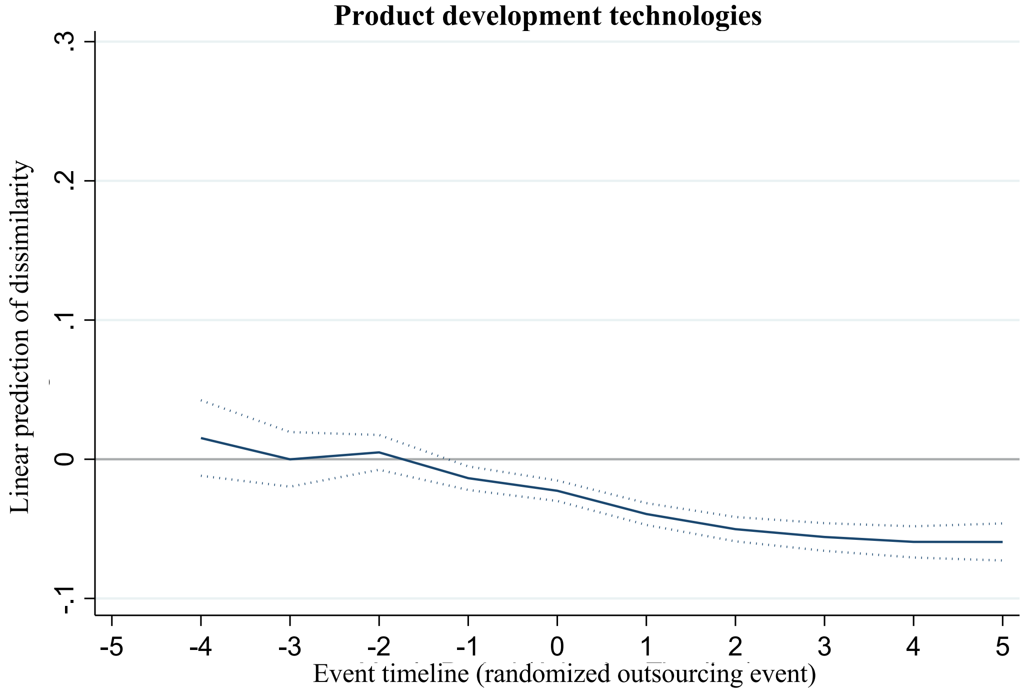
Figure C.6. – Heterogeneity in technology dissimilarity by firm size





Notes: These graphs plot the prediction of the heterogeneous treatment effects of adopting cloud platform services due to size. The specification includes year-level fixed effects and weighing. Standard errors are clustered at the firm level.

Figure C.7. – Randomization: Outsourcing event assigned to a random year



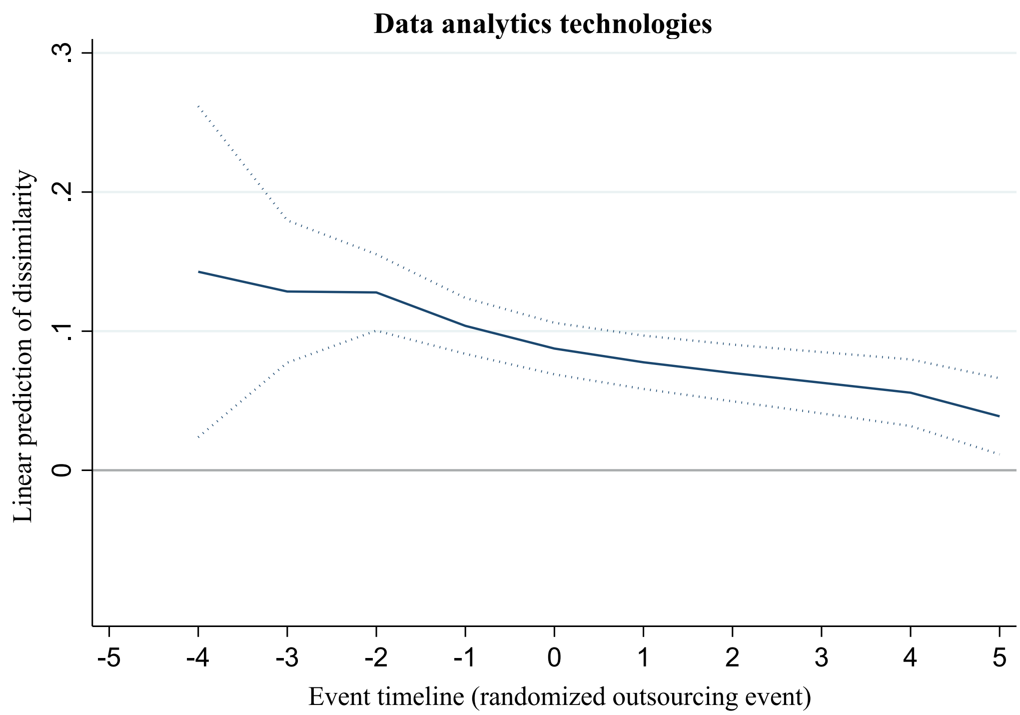


Figure C.8. – Shannon's entropy measure by year

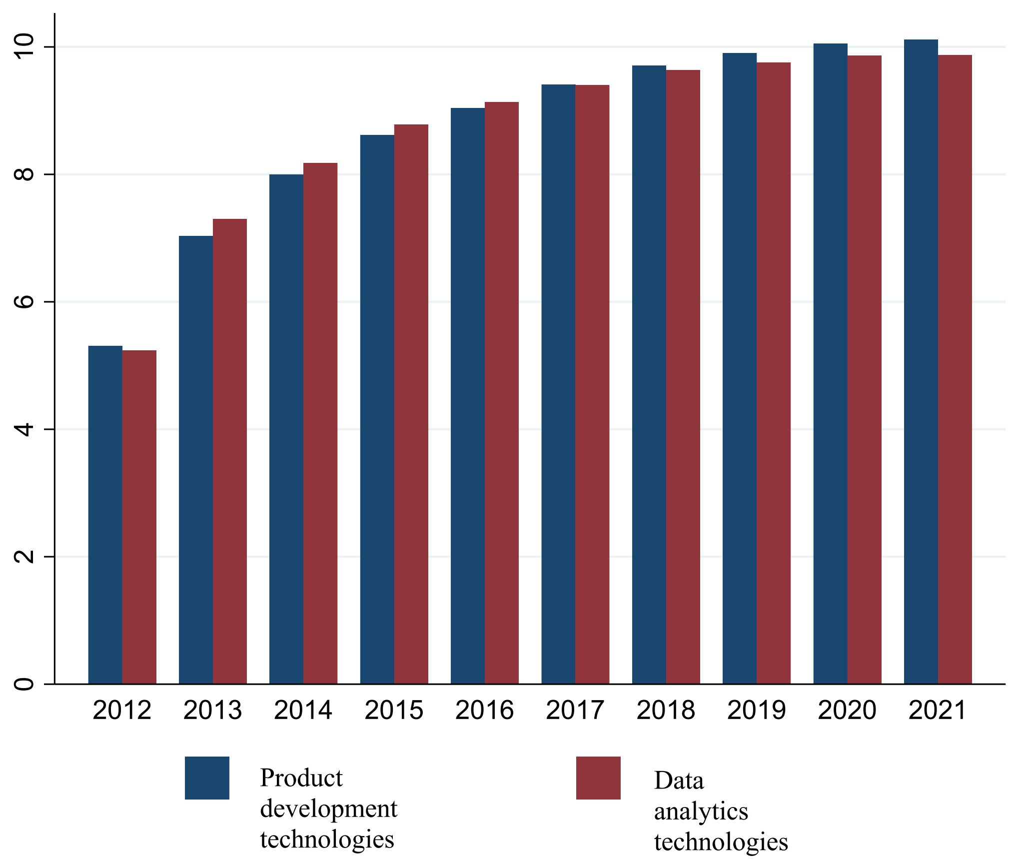
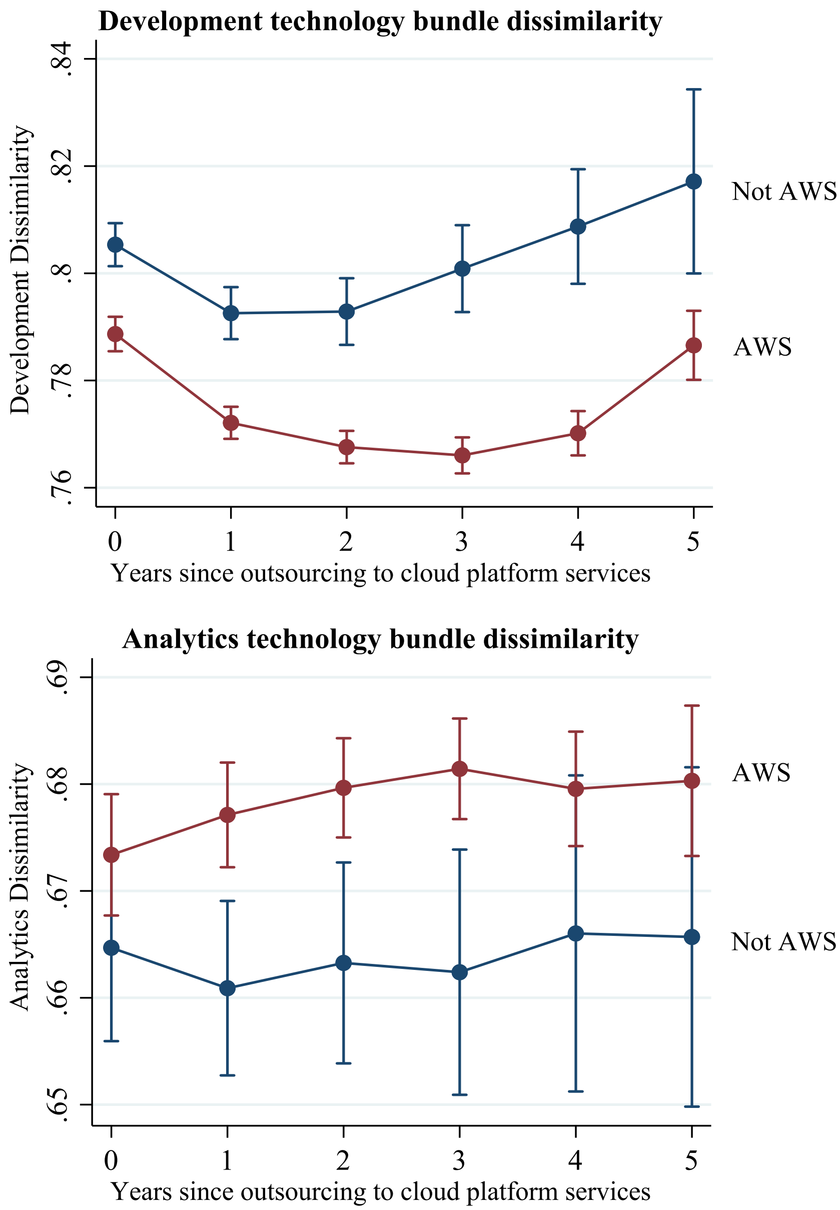
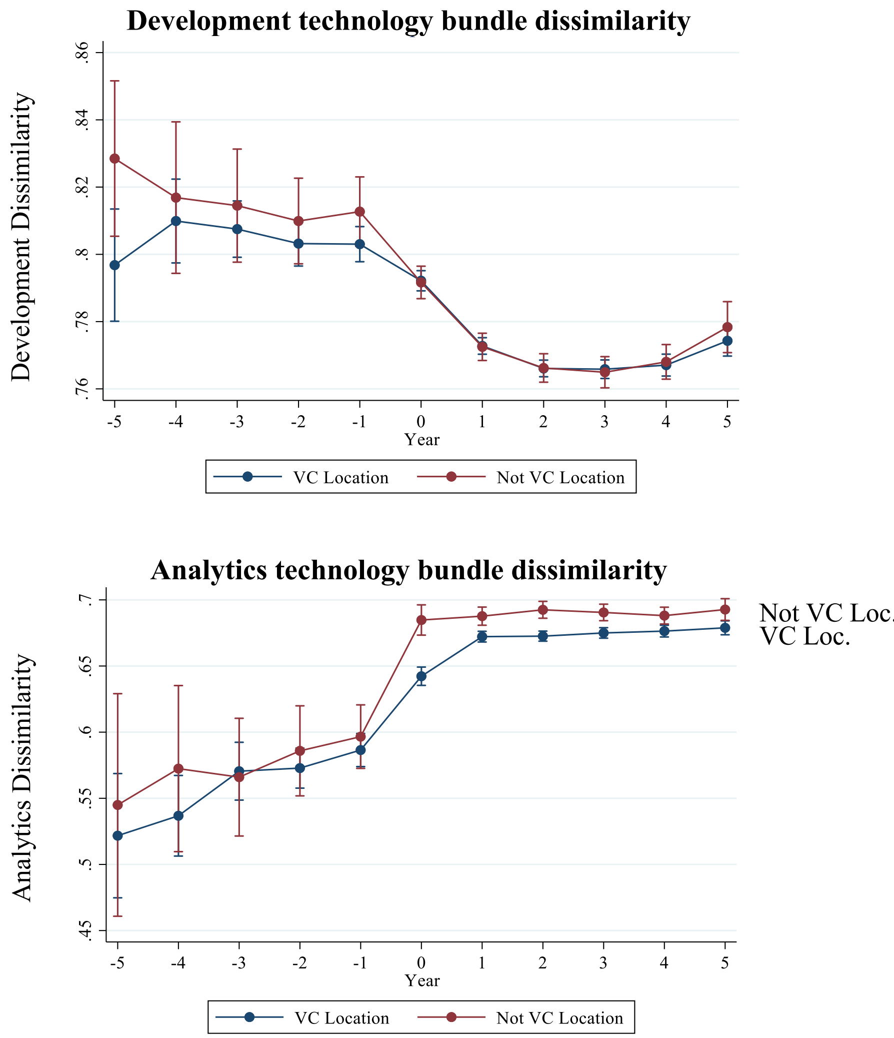


Figure C.9. – Technology dissimilarity (AWS vs. not AWS, treatment group)



Notes: These graphs plot the precited difference in technology dissimilarity for startups using AWS versus not using AWS. Includes only the treatment groups (i.e., startups using cloud platform services) based on specification (3).

Figure C.10. – Heterogeneity in technology dissimilarity (Higher vs. lower VC concentration city)



Notes: These graphs plot the prediction of the heterogeneous treatment effects of adopting cloud platform services due to VC location (San Francisco, New York City, Boston, London, Hong Kong.) The specification includes year-level fixed effects and weighing. Standard errors are clustered at the firm level.

Appendix D – Technology Descriptions

Development Technologies

Development Technology	Description
adobecoldfusion	coldfusion is an application server and software development framework used for the development of computer software in general, and dynamic web sites in particular.
adobedreamweaver	based on the use of certain javascript functions, this page contains code generated, at least initially, by dreamweaver.
ajaxlibrariesapi	the ajax libraries api is a content distribution network and loading architecture for the most popular, open source javascript libraries.
akamai	akamai provides a distributed computing platform for global internet content and application delivery.
akamaiedge	akamai's edge platform is one of the world's largest distributed computing platforms. it is a network of more than 95,000 secure servers equipped with proprietary software and deployed in 71 countries.
alphassl	certificate provided by alphassl, a globalsign company.
alternateprotocol	the server advertises alternate protocol options, most probably providing spdy support.
amazonapigateway	create, publish, maintain, monitor, and secure apis at any scale.
amazoncloudfront	amazon cloudfront is a web service for content delivery. it integrates with other amazon web services to give developers and businesses an easy way to distribute content to end users with low latency, high data transfer speeds, and no commitments.
amazons3cdn	amazon simple storage provides unlimited storage to developers and online businesses - saving costs and increase storage reliability.
amazonssl	amazon supplied ssl certificate
angular	angular version 4.2.*
antdesign	react ui kit / design framework.
apache	apache tomcat is an open source software implementation of the java servlet and javaserver pages technologies.
apollographql	app development framework.
asp	asp.net is a web application framework marketed by microsoft that programmers can use to build dynamic web sites, web applications and xml web services. it is part of microsoft's .net platform and is the successor to microsoft's active server pages (asp) technology.
authpassthrough	frontpage security module for apache.
azureedge	content delivered via azure edge network
bootstrapcdn	bootstrap cdn system - encompasses maxcdn, netdna and stackpath - donated to jsdelivr.
bugbounty	the website has some form of responsible disclosure mechanism for the reporting of security vulnerabilities.
bulma	bulma is an open source css framework based on flexbox and built with sass.
bunnycdngeneral	using content hosted at bunny cdn.
cdnjs	cloudflare's cdn with popular javascript frameworks available.
centos	centos is an enterprise-class linux distribution derived from sources freely provided to the public by a prominent north american enterprise linux vendor. centos conforms fully with the upstream vendors redistribution policy and aims to be 100% binary compatible.
classicasp	active server pages (asp) is a server-side scripting environment that you can use to create and run dynamic, interactive web server applications.
cloudflare	automatically optimizes the delivery of your web pages so your visitors get the fastest page load times and best performance.
cloudflarecdn	content owned by this site hosted on the cloudflare cdn.
cloudflaressl	ssl solutions from cloudflare
cloudinary	image management & delivery solution.
codeigniter	codeigniter is a powerful php framework with a very small footprint.
coldfusionmarkuplanguage(cfml)	cfml is the scripting language used by adobe coldfusion, bluedragon, railo, smithproject, coral web builder, ignitefusion.

comodo	comodo positive ssl certificate.
cpanelssl	cpanel certificate.
dav	a set of extensions to the http protocol which allows users to collaboratively edit and manage files on remote web servers.
ddosguard	ddos protection for your business.
debian	debian is a free operating system (os) for your computer.
digicert	certificate provided by digicert.
digitaloceanspaces	s3-compatible object storage with a built-in cdn.
djangocsrf	django is a high-level python web framework that encourages rapid development and clean, pragmatic design. this metric displays sites that are using django + csrf.
djangolanguage	this website is running the django framework and is setting a language cookie.
encryptioneverywhere	high value, low friction end-to-end security for web hosting partners from symantec.
entrustssl	certificate provided by entrust.
envoy	open source edge and service proxy.
essentialssl	certificate provided by essentialssl, a comodo company.
express	a web application framework for node node.js - expressjs.
facebookcdn	this page has content that links to the facebook content delivery network.
facebookdomainverification	domain verification provides a way for you to claim ownership of your domain in facebook business manager.
fastlycdn	links to fastly cdn based content.
firebase	a scalable real time backend system for websites.
flatsome	woocommerce responsive theme.
gandistandardssl	gandi hosting standard ssl certificate.
gatsbyjs	modern website and web apps generator for react.
geotrust	certificate provided by geotrust.
githubhosting	this site is hosted on github infrastructure.
globalsign	certificate provided by globalsign.
godaddycdn	this site has content that links to godaddy cdn.
godaddyssl	certificate provided by godaddy.
googlecloudfunctions	event-driven serverless compute platform.
googlecloudstorage	store objects of any size and manage access to their data on an individual or group basis within the google network.
googlepagespeedmodule	the pagespeed modules are open source server modules that optimize your site automatically.
googlessl	uses ssl from google
gstaticgooglestaticcontent	google has off-loaded static content (javascript/images/css) to a different domain name in an effort to reduce bandwidth usage and increase network performance for the end user.
gumby	gumby 2 is a responsive css framework.
herokussl	ssl certificate provided by heroku. the site is normally hosted on heroku for this to happen.
herokuvegurproxy	content from this page is being sent via the heroku vegur proxy.
highwindscdn	cdn built to meet the delivery needs of even the largest media and entertainment companies.
incapsulacdn	global cdn and optimizer.
ionic	ionic framework is a open source mobile sdk for developing native and progressive web apps.
javaee	java platform, enterprise edition (java ee) is the industry standard for developing portable, robust, scalable and secure server-side java applications.
jquerycdn	the jquery amazon s3 content delivery network
jsdelivr	a free cdn where javascript developers can host their files. encompasses maxcdn, and bootstrapcdn.
laravel	a php mvc framework.
letsencrypt	let’s encrypt is a free open certificate authority.
limitermodules	log byte and bandwidth limiter modules.
lottiefiles	open source animation file format providing lightweight, scalable animations .
materializecss	material design css framework

materialui	react components that implement google's material design.
maxcdn	maxcdn's dynamic site acceleration optimizes content delivery and web applications by using edge locations. previously known as netdna.
mediatemplessl	the site is using ssl certificate from media temple hosting.
meteor	meteor is an environment for building modern websites.
microsoftazureblobstorage	windows azure blob storage is a service for storing large amounts of unstructured data that can be accessed from anywhere in the world via http or https.
microsoftcdn	content delivery network services from microsoft azure.
microsoftssl	the ssl certificate is connected with microsoft.
modpagespeed	mod_pagespeed is an open source apache module that automatically optimizes web pages and resources on them.
modssl	this module provides strong cryptography for the apache 1.3 webserver via the secure sockets layer (ssl v2/v3) and transport layer security (tls v1) protocols
next	react.js framework for static site generator apps. owned by vercel.
nuxt	vue.js application framework.
oneclickssl	fully automated ssl secure site activation from gmo internet group.
openresty	application server and framework system.
openssl	the openssl project is a collaborative effort to develop a robust, commercial-grade, full-featured, and open source toolkit implementing the secure sockets layer (ssl v2/v3) and transport layer security (tls v1) protocols as well as a full-strength general purpose cryptography library.
optimole	real-time image processing and image cdn for wordpress.
osscdn	open source software cdn from maxcdn.
ovhanycast	content hosted on an anycast load balanced ip address from ovh.
ovhssl	ssl certificates from french based network provider ovh
parallelspleskpanel	host and manage websites and servers at any scale, includes virtualization software.
parallelsssl	ssl reseller program from parallels
perl	perl is a general-purpose programming language originally developed for text manipulation and now used for a wide range of tasks including system administration, web development, network programming, gui development, and more.
php	php is a widely used general-purpose scripting language that is especially suited for web development and can be embedded into html.
placeholdit	a quick and simple placeholder service.
pubnub	api that allows you to build realtime apps in minutes.
pure	a set of small, responsive css modules.
pusher	pusher is a realtime service that complements your existing server architecture.
python	python version 2.4.*
quic	quick udp internet connections, pronounced quick is a transport layer network protocol developed by google.
rackspacecdn	rackspace cdn system.
rapidssl	rapidssl certificate provider.
rawgit	serves raw files from github with the right content type headers.
reactonrails	react on rails integrates rails with (server rendering of) facebook's react front-end framework.
redhatenterpriselinux	red hat enterprise linux (often abbreviated to rhel) is a linux distribution produced by red hat and targeted toward the commercial market, including mainframes.
rubyonrails	ruby on rails is an open source web framework that is optimized for programmer happiness and sustainable productivity. note that ruby on rails has two detection techniques and this is one of them.
sectigo	ssl from sectigo formerly comodo.
semanticui	semantic empowers designers and developers by creating a language for sharing ui.
sonatype	devops automation nexus system.
ssl	certificate provided by ssl.com
stackoverflowcdn	stackoverflow and family cdn.
stackpath	accelerates websites, apps, apis, streams and downloads.
stackpathbootstrapcdn	stackpath's bootstrap cdn system - encompasses maxcdn and netdna.
starfieldtechnologies	certificate provided by starfield technologies

startssl	certificate provided by startssl.
startupframework	design framework for web developers.
stimulus	javascript framework for augmenting html from basecamp.
sucuricloudproxy	sucuri firewall (cloudproxy) is a cloud-based waf and intrusion prevention system for web sites
svelte	ui interface builder system.
symantec	verisign/symantec ssl certificates.
thawtessl	certificated provided by thawte.
total	a server side framework for node.js providing the ability to build web sites using js, html and css.
twittercdn	this page contains content sourced from the twitter cdn, either by the use of widgets or linking to image content on twimg.com currently hosted by akamai and amazon.
ubuntu	ubuntu is a free, debian derived linux-based operating system, available with both community and professional support.
unix	a *nix based operating system (undisclosed).
unpkg	unpkg is a fast, global content delivery network for everything on npm.
vimeocdn	this page uses content from the vimeo cdn.
yahooimagecdn	the website contains links to yahoo image cdn.
zencodercdn	this page has content hosted on the zencoder cdn, owned by brightcove.

Analytics Technologies

Analytics Technology	**Description**
33across	a technology that connects users content and products into the social graph.
6sense	lead generation funnel analytics tool.
accessibe	website accessibility monitoring and auditing platform.
activecampaign	marketing automation, email marketing and behavioral analysis.
acton	marketing automation software.
acxiom	technology and marketing services that enable marketers to manage audiences.
adjust	mobile app tracking system.
adobeanalytics	marketing analytics platform from adobe.
adobedynamictagmanagement	satellite puts an end to tag and technology management, letting marketers and analysts manage their tools. previously known as search discovery satellite now adobe dtm.
adobeexperienceplatformidentityservice	connects devices to people.
adobelaunch	adobe experience platform tag management system.
adobemarketingcloud	a complete set of marketing solutions from adobe.
affiliatly	affiliate tracking software for ecommerce stores.
agilecrm	agile is a fully-integrated sales & marketing suite for small businesses.
ahoy	first party analytics for rails.
airbrake	airbrake collects errors generated by other applications, and aggregates the results for review.
airpr	prtech company provides analytics and insights for what's driving engagement.
akamaimpulse	multi-channel real time analytics package - rum system by akamai previously soasta.
albacross	b2b digital marketing tool that allows you to try to identify the companies that are visiting your website.
alexacertifiedsitemetrics	alexa's certified program and pro metrics.
alexametrics	the page has embedded alexa metrics.
amazonadvertisingsizmekadsuite	campaign management analytics from amazon formerly mediamind.
ambassador	referral marketing software.
amplitude	mobile analytics platform.
appsflyer	mobile attribution & marketing analytics platform
atlasactiontags	work alongside the tracking of campaigns and track the conversion performance of your online media activity.
attentive	personalized mobile messaging platform.
augur	device and consumer recognition javascript service.
baiduanalytics	analytics tracking pixel from chinese language search engine baidu.
bingconversiontracking	help optimize search ads campaigns.

binguniversaleventtracking	universal event tracking (uet) is a simple and powerful campaign measurement solution that allows you to track key conversion goals important to your business.
bizible	multi-channel roi marketing analytics tool.
bizoinsights	bizo insight tags are installed on a partner website to enable bizo to generate and/or record anonymous analytics about the partner’s site visitors. acquired by linkedin.
boldcommerce	shopify app development and partner to help increase sales. previously shappify.
bombora	advertising analytics and tracking service.
branch	mobile deep linking system to increase engagement and retention.
braze	braze is a lifecycle marketing platform formerly known as appboy.
calltrackingmetrics	call tracking & analytics for advertising.
capterra	software tracking system and badge.
castle	deep visibility into what users are doing on your website.
chartbeat	live traffic monitoring of your website.
claritas	custom audience segments & consumer insights for over 120 million households
clearbit	sales and marketing workflow analytics.
clearbitreveal	identifies anonymous visitors to websites.
clevertap	behavioural analytics and engagement platform.
clicktale	records visitors to the website and every action as they browse the site. creates movies to allow the website to understand how it gets used.
clicky	clicky web analytics system, previously known as getclicky
cloudflareinsights	visitor analytics and threat monitoring.
cloudflarerocketloader	automatically optimizes your pages to minimize the number of network connections and ensure even third party resources won't slow down page rendering.
cloudflarewebanalytics	privacy-first web analytics from cloudflare.
comscore	market research company that studies internet trends and behavior.
convert	increase conversion and engagement of website visitors by personalizing content based on behavior. previously known as reedge.
convertflow	lead generation and on-site retargeting
crazyegg	crazy egg provides visualization of visits to your website.
crimsonhexagon	ai-powered consumer insights tracking platform.
crosspixelmedia	cross pixel is the leading provider of high performance audience data.
customer	email people automatically based on what they do (or don't do) in your app.
datadog	cloud monitoring as a service system.
datalogix	leverages the power of purchase-based audience targeting to drive measurable online and offline sales
demandbase	abm software for mid-market and enterprise b2b companies.
digitalwindow	digital window provides performance marketing solutions. providing customers with the tools and account management to get the most from their affiliate programmes
dotomi	dotomi applies personalized media practices to anonymous, user-level marketing programs.
doubleclickfloodlight	floodlight is feature of doubleclick ads that allows advertisers to capture and report on the actions of users who visit their website after viewing or clicking on one of the advertiser's ads.
dynatrace	dynatrace provides software intelligence for enterprise cloud ecosystems. dynatrace is an ai-powered, full stack and automated monitoring and analytics solution that provides insights into users, transactions, applications, and hybrid multi-cloud environments.
efficientfrontier	unified performance marketing platform that optimizes across both search and display. now owned by adobe and includes everest tech.
eloqua	marketing automation provider.
engagio	account based marketing service.
everesttechnologies	performance testing and channel strategy provider for ecommerce.
facebookconversiontracking	conversion tracking functionality from facebook, allows a user to track advertisement clicks.
facebookdomaininsights	this website contains tracking information that allows admins to see facebook insights out of facebook to this domain.
facebookpixel	facebook pixel is facebooks conversion tracking system for ads on facebook to websites.
facebookpixelforshopify	facebook pixel specifically for shopify.
facebookpixelviewcontent	calls to facebook pixel 'viewcontent'

facebooksignal	journalists use signal to surface relevant trends, photos, videos and posts from facebook and instagram for use in their storytelling and reporting.
facebooktagapi	the javascript tag api can be used to track custom audience and conversion events.
fastlycdn	real-time analytics and cdn platform. analyze your web and server traffic patterns in real-time.
firstpromoter	affiliate and referral tracking system.
freshmarketer	conversion optimization suite from freshworks.
freshworkscrm	ai-based lead scoring, phone, email, activity capture, and more.
fullstory	fullstory lets product and support teams understand everything about the customer experience.
g2crowdconversion	conversion tracking for g2 crowd pages.
gemiuspl	online research company based in poland
globalsitetag	google's primary tag for google measurement/conversion tracking, adwords and doubleclick.
googleadwordsconversion	adwords conversion tracking code.
googleanalytics	google analytics offers a host of compelling features and benefits for everyone from senior executives and advertising and marketing professionals to site owners and content developers.
googlecallconversiontracking	use phone call conversion tracking to help you see how effectively your ads lead to phone calls from your website.
googlecontentexperiments	content experiments helps you optimize for goals you have already defined in your google analytics account, and can help you decide which page designs, layouts and content are most effective.
googleconversion	this free tool in adwords can show you what happens after customers click your ad (for example, whether they purchased your product, called from a mobile phone or downloaded your app).
googledoubleclickconversion	doubleclick conversion tracking from google global site tag.
googleoptimize360	test different variations of a website and then tailor it to deliver a personalized experience that works best for each customer and for your business.
googleuniversalanalytics	the analytics.js javascript snippet is a new way to measure how users interact with your website. it is similar to the previous google tracking code, ga.js, but offers more flexibility for developers to customize their implementations.
gosquared	see who's reading, commenting, joining, or buying on your website right now.
growsumo	reward customers and people for sending referrals.
heap	heap automatically captures every user action in your web app and lets you measure it all.
heatmapit	heatmap based tools from heatmap.it.
hittaillongtailkeywordmarketing	hittail claims they are the only product that reveals in real time which keywords people use to find the website.
hotjar	a heatmap, survey, feedback and funnel application.
hubspot	hubspot provides marketing information and leads via inbounding marketing software.
hubspotads	turn hubspot lists into ads targeting audiences and track the roi of your facebook and google ads automatically.
hubspotanalytics	measure the performance of all your marketing campaigns
hubspotcalltoactions	create personalized calls-to-action that are designed to convert and measure them.
hubspotforms	marketing automation form feedback into hubspot tool.
hubspotleadflows	lead flows allow you to easily create and customize engaging lead capture forms.
igodigital	analyzes individual shopper behavior and provides personalized product recommendations. now owned by exacttarget.
improvely	conversion tracking, click fraud monitoring and a/b testing for online marketers and agencies.
innocraftcloud	all in one analytics package from matomo.
insightera	provides b2b customer acquisition with real-time inbound marketing. now marketo real-time personalization.
inspectlet	record and watch everything your visitors do.
invitemedia	automatically buy from multiple ad exchanges in real-time, all through the same interface.
ipstack	ip to geolocation apis and global ip database services.
jabmo	automated lead generation software based on website visitors. now known as jabmo.
keenio	analytics backend-as-a-service lets developers build analytics features directly into apps.

kenshoo	automates the whole process of creating and managing search-engine marketing campaigns.
kickfire	ip address-to-company api and real time visitor intent discovery.
kissmetrics	helps measure results and improve them with analytics from kissmetrics.
klaviyo	customer lifecycle management platform for web apps and ecommerce.
knowbe4	security awareness system.
kochava	unified audience attribution and analytics platform.
leadfeeder	leadfeeder shows you which companies are visiting your site.
leadforensics	visibility of which companies have visited your site, when they visited, what they searched on and the pages they viewed.
leadin	get insights into everyone who fills out a form on your site. from hubspot.
leadinfo	identify b2b website visitors.
leadlander	real time customer intelligence, a website marketing solution.
leadworx	lead discovery tool.
linkedininsights	the linkedin insight tag is a piece of lightweight javascript code that you can add to your website to enable in-depth campaign reporting and unlock valuable insights about your website visitors and for conversion optimization of ads.
loader	load testing tool for websites.
loggly	cloud-based solution that tries to makes sense of log data coming from applications, platforms, and systems. owned by solarwinds.
lotamecrowdcontrol	data driven marketing advertising program provides social media sites with advance targeting
luckyorange	lucky orange lets you see what people are doing on your website, in real time, and interact with them.
madkudu	lead scoring and signup forms.
mailitelite	email newsletters made easy signup form.
mailmunch	email marketing service and customer acquisition app.
marinsoftware	helps advertisers and agencies manage and grow their search campaigns .
marketo	marketo provides sophisticated yet easy marketing automation software that helps marketing and sales work together to drive revenue and improve marketing accountability.
marketorealtimepersonalization	allows for event tracking and dynamic customization of a webpage back to marketo.
matomo	matomo is an open source web analytics software. it gives interesting reports on your website visitors, your popular pages, the search engines keywords they used, the language they speak and so much more. previously known as piwik web analytics.
matomocloud	cloud hosted version of matomo analytics.
mautic	open source marketing automation software.
mediamath	tools that enable and empower marketing professionals.
microsoftadcenter	clicks. leads. sales. pay only when someone clicks your ad.
microsoftapplicationinsights	gain insights through application performance management and instant analytics.
microsoftclarity	free-to-use analytics product for webmasters that shows how people are using your website.
mixpanel	this is an analytic platform that is particularly optimized funnel/work-flow optimization.
moat	moat advertising metrics system. owned by oracle.
mouseflow	mouseflow records videos of your site visitors and generates heatmaps highlighting areas users are clicking, scrolling and ignoring.
mutiny	personalization platform, engage your site visitors with a tailored experience.
naveranalytics	korean based analytics service.
netfactorvisitortracker	lead generation software for your website.
newrelic	new relic is a dashboard used to keep an eye on application health and availability while monitoring real user experience.
oktopost	social media management for b2b marketing.
omnituresitecatalyst	omniture sitecatalyst provides your website with actionable, real-time intelligence regarding online strategies and marketing initiatives.
optimizely	optimizely empowers companies to deliver more relevant and effective digital experiences on websites and mobile through a/b testing and personalization.
optimonk	retargeting platform, that tries to help increase the conversion rate.
oribianalytics	web analytics and event tracking system.

outfunnel	sales marketing automation platform for pipedrive.
owneriq	enables advertisers, manufacturers and retailers to more precisely target their online message based on what consumers own.
parse	parse.ly provides web analytics tools and apis built specifically for the needs of online content sites. its flagship product, parse.ly dash, provides historical, real-time, and predictive insights for the web's best publishers.
paypalmarketingsolutions	get powerful marketing tools designed to help increase your sale. includes paypal credit, fast checkout and venmo accept options.
pendo	pendo captures user behavior, gathers feedback, and provides contextual help.
pingdomrum	real user monitoring gives insight into performance for actual users visiting the website.
pinterestconversiontracking	tag that allows you to track actions people take on your website after viewing your promoted pin.
pipedrive	sales management tool small sales teams,
plausibleanalytics	lightweight and open source web analytics tool.
poptin	create engaging web and mobile overlays to try to improve conversion rate.
posthog	self hosted analytics tool.
preact	preact is web software that takes the job of supporting customers to the next level. formally known as less neglect.
profitwell	subscription and financial metrics in one place.
proof	social proof on sales funnel to help increase conversions.
ptengine	ptengine is a heatmap and web analytics platform.
qualia	real-time insights platform to help improve conversion rates.
qualified	conversational marketing software system.
quantcastmeasurement	provides quantcast with tracking information about your site which anyone can access and view demographic information.
rapleaf	marketing automation tools with the necessary data to help brands keep their customers engaged. now towerdata.
rdstation	digital marketing lead generation tool for websites, from brazil.
redditconversiontracking	conversion tracking system from reddit.
reporturi	enable your users browsers to automatically report security threats.
sailthruhorizon	empowers marketers to turn data into insights and act on those findings quickly and automatically.
salesforce	salesforce is a leading platform for cloud based web apps.
salesforceaudiencestudio	captures, connects and monetize consumer data - previous salesforce dmp and krux digital
salesforcewebtolead	with web-to-lead, you can gather information from your company’s website and automatically generate up to 500 new leads a day.
salesloft	sales engagement platform.
salesmanago	polish based marketing automation software.
segment	segment gives you the ability to instrument your web app for analytics once, and then send your data to any number of analytics services. previously known as segment.io
sendpulse	integrated marketing messaging platform.
sessioncam	session replay, website heatmaps and web analytics.
shareaholic	browser and website analytics tools.
sharpspring	marketing automation for agencies and smbs.
siftscience	sift science monitors a site's traffic in real time and alerts you instantly to fraudulent activity.
signal	distributed data management platform helps share data the website creates with other platforms such as advertisers and audience analytics.
smartlook	records screens of real users on your website.
snowplow	open source analytics that you store yourself.
statcounter	the website uses statcounter a free yet reliable invisible web tracker, highly configurable hit counter and real-time detailed web stats.
steelhouse	behavioral commerce platform, real time onsite offers, dynamic retargeting and other technology features.
sumo	sales and marketing strategies to reduce cart abandonment and increase average order value for ecommerce.
survicate	visitors insights for lead generation & nurturing.

tapfiliate	affiliate tracking software for ecommerce and saas
tatari	tatari measures tv advertising and helps companies optimize their campaigns.
tellapart	the best customers & prospects from the rest.
terminus	account-based marketing software for b2b marketers.
thriveleads	mailing list and conversion optimization wordpress plugin.
tiktokconversiontrackingpixel	tiktok advertising conversion tracking pixel.
toutapp	toutapp live feed tells you exactly what your leads are doing with your sales emails.
trackalyzer	leadlander solution.
trendemon	conversion optimization for content.
triblio	a content marketing platform.
trustpilot	trustpilot is an open, community-based platform for sharing real reviews of shopping experiences online.
tvsquared	real-time tv attribution platform in the industry.
twitteranalytics	a tool that helps website owners understand how much traffic they receive from twitter and the effectiveness of twitter integrations on their sites. includes twitter conversion tracking.
twitterconversiontracking	twitter ads conversion tracking code.
twitterwebsiteuniversaltag	a tool from twitter that makes it possible for advertisers to track website conversions and manage tailored audience campaigns.
tynttracer	tynt tracer, monitors and watches what is being copied from your website, such as your copyrighted content.
veinteractive	ve interactive is a data-driven solutions provider for shopping cart merchants.
vero	send more targeted emails to your customers based on their personal behaviour.
visistat	visistat is a suite of tools that measures the effectiveness of website performance and activity.
visitorqueue	website visitor tracking software
visualiq	marketing attribution and optimization service.
visualstudiotracking	microsoft visual studio based tracking services.
visualvisitor	find out who is on your site and what they are looking at with this lead tool.
visualwebsiteoptimizer	vwo provides a/b, split and multivariate testing software.
whoisvisiting	lead generation from website visitors.
wizrocket	wizrocket is a user behavior analysis & targeting tool.
woopra	woopra is a real-time customer analytics service that provides solutions for sales, service, marketing and product teams.
wootric	in-app nps scoring software.
yahoodot	fives advertisers a simple way to measure and improve customer engagement across campaigns.
yahoowebanalytics	yahoo! web analytics is an enterprise site analytics tool that provides real-time insight into visitor behavior on your website.
yandexmetrika	a free russian tool for increasing the conversion of the site. watch for key performance site, analyze visitor behavior, evaluate the impact of advertising campaigns.
zarget	conversion rate optimization and ab testing software.
zohopagesense	conversion optimization and personalization platform.
zoominfo	b2b database provider and user analytics tracking.