Outsourcing IT and Technological Differentiation: Evidence from Digital Startups [1]

Stephen Michael Impink

New York University, Stern School of Business

August 1, 2022

Abstract: Does outsourcing IT impact the breadth of a firm's technological innovation? With the advent of cloud services, data-centric startups are licensing IT assets as a service instead of developing them internally. Despite this growing trend, we know little about how this early-stage resource decision impacts technology adoption, product innovation, and longer-term performance. When startups outsource their IT, they develop an early "access relationship" with a cloud services provider that shares valuable resources related to their cloud platform. However, these cloud providers control which resources they share and how well their platforms fit with other technologies, potentially impacting the nature of innovation. Using panel data on app-developing startups, I find that: startups that outsource their IT adopt more technologies; their bundles of technologies core to developing products become more similar to each other to fit with the cloud platform's underlying technology; their data-related analytics technologies become more dissimilar from each other aiding in more differentiated product innovations. Lastly, having more similar production technologies to others on the same cloud platform (i.e., having a better technological fit with the cloud platform) relates to increased funding, patenting, and web traffic.

[1]I thank: my committee (Robert Seamans, Deepak Hedge, Raffaella Sadun, and Belen Villalonga), JP Eggers, Sinziana Dorobantu, Gino Cattani, Melissa Schilling, Beth Bechky, Geoffrey Borchardt, Melody Chang, Manav Raj, Katia Meggiorin, Jorge Guzman, Bo Cowgill, Nataliya Wright, Mike Ewens, Kristina McElheran 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 are valuable to the economy (Gans et al., 2019; Guzman and Stern, 2020; Kerr et al., 2014). Yet digitization and the use of big data raise numerous questions for 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). The need for computing power and web development infrastructure in digital production forces high-tech startups to determine whether they "make" or "buy" IT early in their existence before developing their products (Schneider and Sunyaev, 2016; Susarla et al., 2009) and has implications for the firms' structure and partnerships (Coase, 1937; Williamson, 1979, 1998). When startups make IT assets in-house, they hire specialized IT-focused 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, increasing their capital expenditure and expanding their structure horizontally. Alternately, when startups outsource (i.e., the "buy" scenario), they lease subscription-based cloud IT services from a single technology firm and develop their products on that firm's IT infrastructure, focusing more narrowly on product development.

Outsourcing aspects of IT asset development to cloud providers has become increasingly common; Gartner (2018) suggests that firms not using cloud services will soon be as rare as firms not using the internet. Despite this growing trend, we know little about how this early-stage resource decision impacts technology adoption, product innovation, and longer-term performance. Cloud services have become more sophisticated and challenging to replicate, raising the cost of developing comparable IT capabilities internally.[2] The initial costs of internal IT development may be too high for cash-strapped startups to bear. Furthermore, it may be difficult for startups to find investors willing to fund the higher initial capital

[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). Also, I depict cloud platform adoption rates for startups in my sample that are 3 or more years old in Appendix Figure C.2.

expenditure of internal IT development if subscription-based cloud IT services are a viable option. The decision to either internally develop or externally license IT assets was once a strategically important source of differentiation. However, now, due to myriad factors: including the data-intensive nature of digital startups' production, better cloud security options, increased cloud compliance with regulations, reduced introductory cloud pricing, and access to the cloud providers' platform-related resources, it is difficult for emerging digital-first startups to rationalize internal IT development.

My research question asks whether and how outsourcing IT development impacts the breadth of technological innovation. Subscription-based cloud services are operating expenses, lowering the initial capital requirements to enter the industry. Since capital requirements are lower, VC investors can provide more startups with enough money to start product development (i.e., "spray and pray;" Ewens et al., 2017). And possibly more startups are developing products. To a similar effect, leveraging existing platforms and technology solutions may enable startups to focus more on product development through increased experimentation (Kerr et al., 2014; Koning et al., 2019).

On the other hand, we have little insight into how these changes impact the nature of resulting innovations (Lerner and Nanda, 2020; Ewens et al., 2017). In fact, despite potential increases in aggregate production, there are several reasons why one could expect that outsourcing most IT development to a few large technology firms would lead startups to adopt more similar technologies and produce less differentiated products. At the startup level: first, outsourcing creates an access relationship that provides all startups with similar resources related to their suppliers' platform; second, startups rely more heavily on a single cloud supplier instead of the many disparate suppliers, limiting access to a broader array of resources and expertise; third, outsourcing reduces startups' ability to customize their IT assets; fourth, product development requires platform-specific investment in complementary assets (i.e., asset specificity; sunk costs), increasing the fit of technologies with one platform and limiting compatibility across platforms. At the industry level, outsourcing IT development lowers startups' initial capital needs, enabling investors to provide smaller amounts of capital to more firms. Investors, stretched thin from funding more startups,

have less bandwidth to tailor their expertise and guidance to each startup's specific needs.[3] And lastly, market share is highly concentrated with a few large suppliers, potentially limiting the breadth of IT assets developed and the resources shared.

This paper uses unique panel data on technology adoption for ~3400 app-producing startups to examine startups with a web-based or mobile (i.e., Android, iOS) application (app). These startups need IT, computer programming expertise, and data to produce their apps. I find that startups use larger bundles of technologies after adding cloud platform services (PaaS). Moreover, using cloud platform services and creating an access relationship with a cloud provider affects the breadth of technology adoption, depending on if technologies must fit with the underlying IT platform and related technologies for the resulting products to work effectively. Product development technology (i.e., development ops, frameworks, and languages) bundles become more similar to those used by potential rivals – other app-producing startups in my sample. These product development technologies are like independent machines used in a traditional manufacturing assembly line, and each technology provides some specific function in production. The bundles of technologies represent all the machines a firm uses to develop its products. These "machines" must fit with the assembly line and other machines used in production for products to work effectively. Similarly, for digital production, these technologies must fit well with the cloud platform (i.e., external fit) and the other technologies in the bundle (i.e., internal fit).

App-developing startups also use technologies tangential to production, not directly related to product development. The fit between these technologies and the underlying cloud platform is less important to product effectiveness. For instance, bundles of technologies enabling data collection, analysis, and reconfiguration are unrelated to the production "machinery" yet provide data as outputs that are important for product differentiation. Bundles of these analytics technologies become more dissimilar from those of other startups in the sample for two reasons: startups realize they need to create more differentiated

[3] Ewens et al. (2018) notes that VCs are less likely to join startups' board.

products since their product development technologies are more similar to their rivals' technologies; startups that use more standardized product development technologies (i.e., out of the box production technologies sponsored by their cloud supplier) have more time to experiment with these periphery analytics technologies. And using more distinct bundles of analytic technologies from rivals can enable startups to differentiate their product by sourcing more unique, representative, and proprietary data. Furthermore, an analysis of end products for a subsample of startups with available data suggests that product innovations become more differentiated after adding cloud platform services, driven by bundles of analytics technologies becoming more distinct from rivals.

Investors reward startups for adopting more similar bundles of product development technologies and more distinct bundles of analytics technologies than rivals and for differentiating their products. Adopting more similar product development technology bundles relate to increased funding, patenting, and web traffic measures, suggesting that startups benefit from being able to focus less on the "assembly line" and more on product differentiation. And adopting more distinct data analytics technologies, enabling product differentiation, and increased product differentiation relates to increased funding. Altogether, this paints a picture of a high-performing digital startup as a firm that standardizes its production technologies to fit better with its other production technologies and supplier's platform, then differentiating its products through adopting more distinct data collection, usage, and reconfiguration technologies.

I rely 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. I use 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. I address potential endogeneity in two ways. I use an instrumental variable approach based on the open-source release of TensorFlow, an AI development framework that enables firms to train neural network algorithms, in late 2015. Using a valid instrument enables me to adjust regression estimates for potentially omitted variables

and support the argument that reverse causality is not a concern.[4] Google's unanticipated decision to release an open-source version of TensorFlow compatible with other cloud platforms was an exogenous shock to AI production, reducing the costs associated with AI development and increasing the value of complementary AI-related labor (Rock, 2019). Then, as an alternate way to address potentially omitted variables, such as hard-to-observe measures of startup quality and founder ability that may relate to technology adoption, I use a double machine learning model based on a random forest algorithm (Chernozhukov et al., 2018). Lastly, I explore two mechanisms that increase the effect of adding cloud platform services. First, startups licensing cloud-based computing hardware (IaaS, infrastructure services) in addition to licensing platform services increases the strength of their relationship with their cloud provider, increasing the quantity and influence of the supplier's shared resources. Second, startups using larger bundles of technologies (i.e., more technologies) experience increased complexity of coordinating interdependent product development technologies while ensuring a fit with the cloud platform. As the coordination costs increase, startups likely rely on guidance and known expertise about which technologies are more compatible, leading them down a more similar development path, adopting more similar product development bundles to rivals.

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 startups' technology adoption and resulting innovations change when they use cloud platform services. Furthermore, I show a meaningful relationship among aspects of technological dissimilarity, product differentiation, and measures of startup performance, including venture funding, patenting, and website traffic. Next, I contribute to the literature on resource sharing and the nature of technological innovation (Baumol, 2001; Boudreau, 2012; Gulati, 1995; Mowery et al., 1998; Stuart, 2000, et al., 1999), by examining how shared resources and the fit of the technology

[4] https://github.com/tensorflow

with the supplier's platform influence technology adoption. 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 control) adopt more similar bundles of platform-related technologies.

This paper proceeds as follows. Section 2 provides an overview of related theories on outsourcing and innovation, focusing on how outsourcing reduces a startup's control over production and impacts a firm's partnerships and structure. Section 3 introduces the context of high-tech startups developing apps and provides insight into two mechanisms, resource sharing and the need for technology fit, influencing technology adoption and the breadth of innovation. The following two sections describe the technology adoption measures (Section 4) and the main analyses' research design (Section 5). Then, section 6 reports these findings, robustness analyses related to these findings, and econometric approaches used to support that my findings are consistent with a causal interpretation. Section 7 examines the implications of technology adoption with a focus on product differentiation and performance. The last section (Section 8) discusses these findings and concludes.

2. The impact of outsourcing on innovation

The decision to outsource IT development 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[5], Klein (1980; et al., 1978) and Williamson (1979) highlight limitations of contracting, explaining that inefficiencies arise

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

from splitting surplus rents ex-post and boundedly-rational suppliers are potentially opportunistic. Based on this literature, outsourcing leads to potential inefficiencies stemming from supplier opportunism, such as switching costs (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 contracting and the frequency of exchange (Grossman and Hart, 1986). These inefficiencies increase over time as customers invest in improving the fit between their technologies and their suppliers' technologies. Moreover, the research addressing the impact of outsourcing on innovation interweaves many of these inefficiencies in their discussion.

First, outsourcing the development of production inputs increases a supplier's control over their customer's current technologies (Rysman and Simcoe, 2008) and future technologies (Greenstein, 1993). Firms often adopt technologies compatible with their supplier's platform or technologies (Simcoe, 2012). Firms make complementary asset-specific investments[6] to increase the fit between their and their supplier's technologies, and these investments may force firms down a specific technological path (Arthur, 1994; Pfeffer and Salancik, 2003; Schilling, 1999), increasing their reliance on that supplier for future technologies (e.g., hostage; Williamson, 1983) and impacting competitive outcomes (Rivkin, 2000; Siggelkow, 2001). And switching costs increase as they progress along this path and make compounding investments.

Second, the decision to outsource creates a supplier relationship that impacts future partnerships (Combs and Ketchen, 1999; Madhok and Tallman, 1998; Young-Ybarra and Wiersema, 1999), which in turn influences the firm's technologies (Gulati, 1995; Mowery et al., 1998; Stuart, 2000, et al., 1999) and innovations (Ahuja, 2000; Baumol, 2001; Hagedoorn and Schakenraad, 1990). However, despite this undue influence, this relationship is particularly valuable for resource-strapped entrepreneurial ventures (Stuart et al., 1999) as the startup may shut down without these resources. For instance, "access relationships" (Stuart,

[6] 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).

2000, et al. 1999), including customer-supplier relationships, technology exchange agreements, and one-directional technology flows[7], enable firms to access needed resources to develop their products. 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), departing from the resources-based perspective that internal development of rent-generating resources enables firms to collect excess returns from market imperfections (i.e., strategic factor markets; Barney, 1986) and reduced imitation (Montgomery, 1994; Peteraf, 1993). In addition to suppliers determining the compatibility of their technologies with other technologies, they decide which resources they develop and share with their customers, further influencing technology adoption.

Third, firms' outsourcing decisions also impact their 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; Chandler, 1962; Coase, 1937; Teece, 1982). And vertical integration is more effective when the likelihood of technological obsolescence is high (Balakrishnan and Wernerfelt, 1986). In contrast to Balakrishnan and Wernerfelt (1986), vertical integration may be less effective in digital industries known for their fast pace of technological development, quickly rendering prior technologies obsolete (Aral et al., 2012; Cachon and Harker, 2002; Giustiziero et al., 2021). In the case of digital industries, having multiple production lines may impede scaling, reducing benefits from the productivity gains (Aral et al., 2012) and IT discounts (Benzina, 2019) associated with the firm size. This ability to scale digital resources rapidly (Fazil et al., 2018) coupled with the firm being more vertically narrow scope (i.e., enabling a greater focus on product development) may create a situation where the potential gains of being able to scale a single product quickly

[7]Hagedoorn and Schkenraad (1990, p.5) provide an exhaustive list, including direct investment, joint research corporations, joint ventures, and joint R&D agreements.

outweigh the transaction cost inefficiencies from outsourcing (Cachon and Harker, 2002; Giustiziero et al., 2021).

3. 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 platform services (PaaS), a specific type of cloud service on which startups develop their apps. Infrastructure services (IaaS) are less consequential to the nature of digital app innovation than platform services but important in the context of IT spending, productivity, and IT hardware development. Cloud platforms differ from many more traditional two-sided platforms, such as Airbnb, Shopify, or Uber. Yet they connect users with supplier services and fit a broader definition of a platform as a "layered architecture of digital technology" (Yoo et al., 2010) with a governance model (Parker et al., 2017).

Contextualizing broader theory on outsourcing to the setting of cloud app development suggests that startups will outsource IT asset development to access resources that are difficult to internally develop from suppliers, focus more time on products, and enable the potential for scaling their digital products. However, outsourcing may indirectly constrain the breadth of innovation when firms adopt technologies to be more compatible with their suppliers' technologies, take advantage of similar shared resources, and alter their firm's structure to facilitate outsourcing. Moreover, the impact of increased supplier control will grow over time or as a supplier's market share increases. Two mechanisms, supplier-to-customer resource sharing and the need for technological fit, provide insight into how outsourcing impacts the breadth of innovation.

Young digital startups have few customers and lack relationships with other firms but quickly establish a relationship with the technology firm supplying their cloud services. They need IT resources and data to develop their products, and larger technology firms have these resources in abundance. The largest cloud services providers – Amazon Web Services (AWS), Google Cloud Platform (GCP), and

Microsoft Azure – have startup-related corporate programs[8] that share resources with their customers. These corporate programs mimic accelerator programs (Hochberg, 2016; Yu, 2020) in sharing generalized resources but differ from traditional accelerators by additionally providing technical resources (i.e., software, compatibility documentation, expertise, troubleshooting) focused on their cloud platform. These shared resources are a mechanism that impacts technology adoption and the nature of innovation. Cloud suppliers share more resources with startups with whom they have a stronger relationship, further impacting the nature of innovation. Using both cloud platform and infrastructure services indicates a stronger relationship, more contact points, and additional integration of the supplier's technologies.

As more startups join cloud platforms, they may have fewer incentives to innovate (Boudreau, 2012). The technological landscape is more complex when platforms grow larger, making it more difficult to combine multiple highly interdependent technologies in innovation (i.e., complexity catastrophe; Fleming and Sorenson, 2001). Startups will adopt increasingly similar bundles of technologies for two reasons: when the fit amongst technologies is important for product efficacy and when the fit between the bundle of technologies and the underlying cloud platform is important for product efficacy. This perspective points to the emergence of several large cloud suppliers as reducing the breadth of innovation. Still, others argue the opposite, that large platforms are multi-dimensional product spaces offering limitless possibilities for technology recombination (Caves, 2000; Parker et al., 2017; Zittrain, 2006) and that standardized or low-code development tools (Miric et al., 2021; Dushnitsky and Strobe, 2021) support more innovation.

Switching costs are high for cloud platform services. Once startups develop their applications on their supplier's platform, moving their product to another platform or internalizing production is challenging. There are fees for offloading data and different coding requirements based on the platform. So, in most cases, it is easier for startups to abandon an app on one cloud provider's platform and start development from ground zero on a different platform than to move their app. As a result, this initial fit is

[8] To give a sense of the scale and breadth of these programs, Amazon AWS recently announced that they have an accelerator specifically focused on startups developing products related to space travel. https://www.geekwire.com /2021/amazon-web-services-launches-space-accelerator-final-frontier-startups/

strategically significant and has outsized implications for product development. Startups may be unable to incur the added cost of changing ill-fitting production technologies or adapting production technologies to fit better with their cloud platform through investing in complementary assets, incentivizing them to rely on standardized, "out-of-the-box" technologies.

Adopting compatible technologies is paramount since switching costs are high, and startups avoid potential incompatibilities by adopting technologies their cloud provider shares or suggests. To prevent compatibility issues, firms may avoid technologies from other suppliers as the cross-platform compatibility of their technologies may be unknown. The potential for incompatibility increases with the size of a firm's technology bundle. As firms use larger bundles of technologies, they incur additional coordination costs of managing an appropriate fit amongst interdependent technologies and with the underlying cloud platform, further incentivizing the use of technologies known to be compatible. The need for fit amongst interdependent technologies is another mechanism that impacts the nature of innovation. As the size of the technology bundles increases, bundles will likely become more similar as startups consider guidance on technology compatibility. It benefits these startups to entrench themselves with a single provider and their platform's ecosystem, following a particular technological trajectory and incurring fewer costs of maintaining fit.

Using existing or standardized product development tools or technologies shared by the cloud provider may enable startups to focus more on differentiating their products instead of worrying about compatibility and fit-related issues in production. Startups may adopt more distinct bundles of data analytics technologies than rivals to enable this increased product differentiation. Technologies enabling data collection, analysis, and recombination are indirectly related or unrelated to product development, so the fit with the underlying cloud platform is unimportant. Hence there is no pressure to adopt more similar bundles of these analytics technologies. And in the case of digital app production, product differentiation

stems from utilizing unique or more representative data that is challenging for rivals to replicate.[9] It follows that startups with larger, more standardized bundles of development technologies and larger, more distinct bundles of analytics technologies than rivals will perform better than comparable firms.[10] Cloud platforms facilitate the standardization of these product development technologies and enable startups to differentiate products by adopting a distinctive bundle of analytics technologies.

4. Data

Before starting the quantitative analyses, I informally interviewed fifteen high-tech startup founders to understand the IT asset outsourcing decision better. Most founders and early technical employees determine which CSP to use before product development. All the startups I interviewed used PaaS, and about half used hardware/IaaS services. One startup initially used internally developed IT to produce its marketing app from 2013 to 2017, 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 decisions. One healthcare startup felt pressure to develop on the Microsoft Azure platform because Azure offered HIPAA-compliant cloud services earlier, which enabled 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 their decision to join a particular cloud platform was influenced by highly discounted offers facilitated through their accelerator or incubator programs. A few mentioned adding a second cloud provider to access more free cloud credits for a tangential project or from "blob" storage[11].

[9] That data could then be used to make business decisions (e.g., product design decisions, marketing decisions, customer, and partner acquisitions decisions) and, in the case of AI-developing startups, to train algorithms.
[10] Consider a profit function: $\Pi = (b_1 count_{dev} - c_1 fit_{dev}) + (b_2 count_{ana} - c_2 fit_{ana})$; constrained such that the $b_1 < b_2$ and $c_2 > c_4$.
[11] Blob storage 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.

The initial cloud decision impacts the startups' future technological compatibility and complementarity as it is challenging to switch platforms. For instance, several startups report being unable to change cloud providers because they would have to "rebuild their entire product" on the new platform. Some mentioned they would have to hire a new programmer with different coding experience or preferences if they switched to Azure, a platform requiring more extensive knowledge of C# than GCP or AWS. Only one startup changed to a new primary cloud provider after product development; however, the change coincided with the startup founder hiring a new CEO, previously an executive at the new cloud provider.

4.A. Firm demographics.

To examine how outsourcing IT development impacts innovation in this context, I establish a sample of startups with a digital application from multiple data providers. I use data from Crunchbase[12] and Pitchbook[13] to compile a list of high-tech startups in information technology or software-related industries and then use data from Apptopia[14] and startup descriptions to confirm the startups have a mobile or website-based application. Next, I use BuiltWith[15] to determine if the startup has an active web domain. To capture higher-growth potential startups, I exclude older firms (>10 years old), firms in China[16], and firms with more than 500 employees. These criteria yield a sample of 3,434 high-tech startups that develop an app as their product. [17] Most of the startups in the sample operate in more developed economies (~90%), with the bulk of the startups located in the United States (49% SD 0.05) or the United Kingdom (6% SD 0.03).

[12] 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.
[13] Pitchbook is a SaaS company that delivers data, research and technology covering private capital markets, including venture capital, private equity, and M&A transactions
[14] Apptopia's data intelligence platform enables brands to analyze critical competitive signals and gain insights across mobile applications and connected devices.
[15] Builtwith returns all the technologies connected to a web domain, covering more than 59k technologies across analytics, advertising, hosting, frameworks, CMS, and more. Prior research in strategy relies on similar data from research: Koning et al., 2019; Dushnitsky and Stroube, 2021.
[16] Our English-language data sources underrepresent the number of startups in China; also, founders from these firms are underrepresented on LinkedIn.
[17] As a potential limitation, third-party data sources may not be representative of the underlying population. For example, they may not capture very young startups, intentionally trying to stay under the radar, or under-represent startups from certain emerging countries where the English language is less common. I do not include firms in China and stress that my results only hold for startups in more developed markets.

Startups in the sample were founded between 2011 and 2020, are 4 (SD 2.4) years old, and have about 45 (SD 63) employees. Around a fifth of these startups target customers in the financial services and healthcare industries. And many describe themselves as developing commercial AI products (38% SD 0.49) or as using machine learning in production (9% SD). I provide additional details on the startup demographics of the sample of startups in Table 1.A.

4.B. Founder measures.

I build measures on startup founders from three data sources: Mantheos[18], AIdentified[19], and manually collected data from public LinkedIn[20] profiles. In our sample, 12% (SD 0.33) of startups have founders 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 undergraduate or graduate major, including math, physics, computer science, statistics, or data science; 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 in R[21], I determine (with 95%) confidence that 13% (SD 0.33) of startups have a female founder or CEO. I provide additional details on the founder demographics of the sample of startups in Table 1.A.

4.C. Cloud services provider measures.

I collect firm-level technology data across time for these startups from BuiltWith. BuiltWith provides 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 what technologies startups

[18] 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.
[19] AIdentified reveals best paths for sales teams, account execs and brands to connect to hyper-targeted, qualified prospects using predictive analytics and next level AI-based relationship intelligence mapping.
[20] LinkedIn is an American business and employment-oriented online service that operates via websites and mobile apps.
[21] I used the "gender" library in R, SSA method, focused on English language birth names common in the 1980s.

adapt "back end."[22] 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 (85%) develop their application on cloud platforms (PaaS); 79% of startups also use cloud infrastructure services (IaaS), often added after startups adopt PaaS. From BuiltWith, I identify ten technology firms that license cloud services: Three suppliers are the largest IT services firms: Amazon, Google, and Microsoft (Big Tech CSPs, 78% share), and the other seven suppliers offer more niche technology services: Alibaba, Digital Ocean, IBM, OVH, Oracle, and Linode (Other CSPs, 7%). Many of these cloud providers offer IaaS services that provide high-power virtual machines with processors (GPUs) and solid-state drive (SSD), such as Amazon Elastic Compute Cloud (EC2), Google Compute Engine, Cloud AutoML, and Azure Machine Learning, which are particularly valuable in AI development. The remaining 15% of startups do not have cloud-based platforms connected to their web domain. I provide additional details on the cloud providers in Table 1.C.

I assume there are differences between how some cloud providers, particularly the largest ones, share resources with startups. Though suppliers often share products and technologies developed internally (i.e., Google is more likely to share Google technologies), I cannot pinpoint which technologies startups added as a direct result of sharing. Nor can I pinpoint other shared non-technology resources, like technical and business expertise, that are shared formally through programs or informally through increased network connectedness. Alternately, this may not be a shared resources story. For instance, the supplier-customer relationship could signal to investors that a startup's products are of higher quality and the startup is of higher reputation. I cannot gauge these signals of reputation and quality from my current data.

[22] 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

4.D. Technology adoption measures.

From the pairing with BuiltWith data, I can determine how startups' technology adoption changes over time.[23] Though BuiltWith provides information on all domain-connected technologies, this paper does not focus on "front end" technologies (e.g., website hosting, fonts, e-commerce, payment, etc.) or organizational technologies (e.g., CRM, sales tools, workforce management, email hosting, etc.). Instead, I focus on product development (e.g., content management systems, content delivery networks, frameworks, and security) and analytics technologies (e.g., data analytics, collection, and telemetry) adopted by startups, which I report in Appendix D. Additionally, I provide more information on cleaning the BuiltWith data in Appendix A.1

Technology Bundle Size. I calculate firm-level technology count measures at the firm x year level, similar to the measures calculated by Berman and Israeli (2022). Startups use an average of 50 (SD 27) technologies, ranging from 1 to 252. Startups using a cloud platform use more technologies (53 SD 27) than startups not using a cloud platform (36 SD 19). On average, firms use 8 (SD 4) product development technologies and 7 (SD 6) analytics technologies.[24] I provide additional details on measure construction in Appendix A.2., kernel density estimates in Appendix Figure C.4, and summary statistics in Table 1.B. under Technology Bundle Size.

Technology Bundle Dissimilarity. Next, I calculate firm-level technology dissimilarity measures for development and analytics technologies based on pairwise cosine similarity at the firm x year level to address my research question, focused on relative changes to the breadth of innovation of app-producing firms in my sample. I adapt these similarity measures to calculate the angular distance, also referred to as

[23] 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. Alternately, they could use product development technologies in a way that is unrelated to application development.

[24] Additionally, I calculate the number of Big Tech technologies (13 SD 6) and premium/subscription technologies (4 SD 4) from additional information provided by BuiltWith. For the final count measure, from analyzing technology descriptions to code open-source technologies (1 SD 1).

angular separation; identically constructed measures have been discussed and used in prior strategy and economics research (Seamans and Zhu, 2014; Sweeting, 2010; Wang and Shaver, 2014). This type of measure is often used in astronomy to calculate the distance of far-away objects by determining the angle between two "sightlines" described by the coordinates of the firm-level vectors of technologies. Those vectors take the value 1 when both startups do not use the technology and 0 when both startups use the technology, configuring a positive semidefinite matrix. The resulting angular distance measure will be bound by [0,1], taking the value of 1 when there is no overlap and 0 when there is perfect overlap.

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

Where, $bundle\_dissimilarity_{ij,t}$ refers to the pairwise angular distance of the focal firm and rival, *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. The dot product is normalized by the length of each vector ($\| V_{it} \| \| V_{jt} \|$), so the size of the technology bundle will not impact the measure.

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 angular distance or differentiation is 0.78 (SD .07) for development technologies and 0.66 (SD 0.12) for analytics technologies.

4.E. Funding measures.

I create and use indicator variables for whether startups received any funding (including seed/angel funding; 71% SD 0.45), venture capital funding (61% SD 0.49), follow-on funding (52% SD 0.5), or higher reputation venture capital funding (9% SD 0.28). Next, I create a measure of firm deal size for each firm in a given year (5.1 SD 7.1, log). As another performance measure, I use data on the duration of website visits

from SimilarWeb[25] (3.1 SD 3.3, log minutes) and an indicator variable for if startups have a patent (4% SD 0.18) to capture some aspects of innovative activity from IPQwerty[26]. Startups in my sample are young, on average, four years old, so we do not yet have a clear indication of which startups will survive. Only 5% (SD 0.21) of startups have closed, and 6% (SD 0.24) of startups were acquired. I provide additional descriptive statistics in Table 1.B. under Performance. I provide descriptive statistics on these in Table 1.B. report correlations of these measures with firm demographics measures (Table A.3.), technology measures (Table A.4.), and performance measures (Table A.5.) in the Appendix.

5. Research Design

I construct a relatively homogenous sample of startups with existing apps that are less than ten years old, have less than 500 employees, and are not located in China. 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 (i.e., cloud platform versus no cloud platform, Big Tech cloud platform versus other vendor's cloud platform, Amazon AWS platform versus other vendor's cloud platform) to weight regressions. This matching procedure supports the argument that startups using a cloud platform are not observably different from firms not using a cloud platform. I provide more details on the sample means before and after matching in Tables 1.A. and 1.B., and I depict these differences and provide more details on the matching criteria in Appendix A.6.

For the main specification, I use an OLS difference-in-differences approach with two-way fixed effects to model the impact of using 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

[25] 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.

[26] IPQwery applies a series of contextual clues to help differentiate between similar company names, then separates each into the correct IP profile. 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).[26]

the control and treatment groups' development and analytics technology bundle size (Figure C.5.A.) and technology bundle dissimilarity (Figure C.5.B.). From these depictions, 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.

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

where, y_{it} refers to the dependent variable: technology count measures or technology differentiation measures, β_1 coefficient of the change from using no cloud platform (0) to using a cloud platform (1), $yearFE_t$ refers to the year-level fixed effect, and $firmFE_i$ refers to the firm-level fixed effect. In these models, I weight regressions based on the output of the CEM procedure by treatment groups, cloud platform versus no cloud platform, which drops unmatched firms and cluster standard error at the firm level. Also, to overcome any potential issues arising from the use of a staggered difference-in-differences model with two-way fixed effects, I estimate the instantaneous treatment effect of startups that are "switchers" (i.e., startups that move from being untreated, not using a cloud platform, to being treated, using a cloud platform) in the period of treatment, following an approach from De Chaisemartin and D'Haultfoeuille (2018, 2020).

Next, I use the following specification to compare different platforms. I use difference-in-differences OLS regression models with the matching approach adjusted to the comparison level (i.e., Platform X vs. Platform Y). For example, when comparing startups on the AWS platform versus those on a different platform, CEM would be set to ensure that startups using AWS were demographically more similar to those not using AWS.

$$y_{it} = \beta_1(cloud\_provider_{it} = 1) + \beta_2(cloud\_provider_{it} = 2) + \beta_3 yearFE_{it} + \beta_4 firmFE_{it} + \varepsilon_{it} \quad (4)$$

where, y_{it} refers to the dependent variable: technology count measures or technology differentiation measures, β_1 refers to the coefficient of the change from no cloud platform (0) to Platform X (1), β_2 refers to the coefficient of the change from no cloud platform (0) to Platform Y (2), $yearFE_t$ refers to the year-

level fixed effect, $firmFE_i$ refers to the firm-level fixed effect. As with the main specification, I cluster standard errors at the firm level.

6. Impact of cloud platforms on technology adoption

6.A. Selection results.

Empirical insight into firm-level decisions to adopt cloud services is scarce in prior research (Schneider and Sunyaev, 2016; Yang and Tate, 2012). Startups are not randomly assigned to different IT development conditions or cloud platforms; founders and early IT employees choose to develop internally or use a specific cloud provider. I examine selection based on founder characteristics, industry, and headquarters location and, in Table 2, report the results of the probit specification (Appendix A.7) supporting the inverse Mill's ratio (IMR) calculation used as a robustness check. The age of the startups is a significant driver of cloud platform adoption (+0.65 SD 0.01), as startups make these decisions around the time they reach product development. Startups selling into healthcare, a more highly regulated industry, are less likely to use a cloud platform (-0.14 SD 0.07). Also, in Europe, where data regulation (e.g., GDPR) is more intense, startups were less likely to adopt cloud platform services (-0.20 SD 0.04). Firms with earlier funding from one of the three largest cloud providers are more likely to adopt a cloud platform (+0.30 SD 0.11). And as anticipated, startups with a founder with hardware experience are less likely to adopt a cloud platform as they have an increased capability to develop IT internally (-0.18 SD 0.08).

I attempt to adjust estimates of the main results through Heckman's selection procedure based on founders' observable characteristics; however, some characteristics may not be observable or are not captured in the data I collected. Moreover, the size of the founding team and combinations of the founding team's characteristics could impact selection in a way unaccounted for by my model.

6.B. Main results.

I examine how technology adoption changes when startups add cloud platform services. I find that bundles of product development technologies (model (1): +0.30 SD 0.02) and analytics technologies (model (5):

+0.23 SD 0.02) become larger. The estimate remains similar when regressions are weighted based on the output of the CEM procedure, controlling for potential differences between the control and treatment groups based on observable demographic variation (models (2) and (6)). The ratio of development technologies to all other technologies increases (models (4)). Also, in Appendix B Table B.1, I report additional results for Big Tech, paid/subscription, and open-source technologies.

Product development technology breadth. However, the technology bundle size provides limited information about the breath of innovation. Regardless of if startups use more technologies, their apps will not work effectively unless their bundle of product development technologies fits well with their production needs, other interdependent technologies, and underlying IT assets. As anticipated, bundles of product development technologies become less dissimilar to other startups in the sample when startups use cloud platform services, as these startups have similar fit constraints pushing them toward a similar path. In Table 4, I build from a model with no controls or weighting (model (1): -0.041 SD 0.002) by adding firm-level fixed effects (model (2)), year-level fixed effect (model (3)), and both firm-level and year-level fixed effects (model (4)). In Table 4, model (5), I include both firm-level and year-level fixed effects and then weight regressions based on CEM (-0.028 SD 0.002).

Next, I examine two mechanisms, resource sharing and the need for technological fit, to provide insight into how outsourcing impacts the nature of innovation and to help support that my results are consistent with a causal interpretation. First, when startups add cloud infrastructure services in addition to platform services, they are more entrenched in a single cloud provider's technologies. With the increased intensity of the relationship, suppliers share more resources with startups, and startups aware of potential interdependencies across the same provider's platform and infrastructure services are influenced by these shared resources. As such, product development technologies become more similar so startups can effectively develop their product (model (7)). Second, startups using larger bundles of development technologies incur increased costs of coordinating among a larger number of interdependent technologies. In product development, startups must also maintain compatibility among the interdependent technologies in this bundle and between the bundle of technologies with the cloud platform for their products to work

effectively. Using technologies that startups know are compatible, such as technologies developed or shared by their cloud provider or discussed in the cloud provider's compatibility guidance, reduces these associated costs, leading startups to use less dissimilar bundles of product development technologies (model (6)).

Analytics technology breadth. Alternately, bundles of analytics technologies that enable firms to collect, analyze, and recombine data become more dissimilar when using cloud platform services (Table 4 model (12): +0.087 SD 0.005), aiding in product differentiation. Since these technologies are not directly related to production, the fit between these technologies and the underlying platform is less important for product efficacy. And unlike bundles of product development technologies, bundles of analytics technologies are unconstrained by fit. In models (8)-(11), I report a similar build-up from a model with no controls or weighting as discussed for product development technologies in models (1)-(4). Also, since the fit is not a constraint, using larger bundles of technologies (model (13)) or adding cloud infrastructure services in addition to cloud platform services (model (14)) enables startups to use even more distinct bundles of analytics technologies, producing data resources that can help differentiate their products.

Cloud platform market share comparison. I examine heterogeneity by platform market share, calculated as the share of startups in my sample on each platform. Amazon AWS has the largest market share supplying about 79% of startups in my sample; Google GCP is the next largest with around 5%.[27] I find that having a higher market share relates to a reduction in product development technology bundle dissimilarity (Table B.9 model (1): -0.08 SD 0.01) and an increase in analytics technology bundle dissimilarity (model (4): +0.09 SD 0.03). The effect remains similar when comparing startups using one of the largest three cloud providers (Big Tech CSPs: AWS, GCP, and MS Azure) versus another smaller cloud provider (Other CSPs: Alibaba, Digital Ocean, IBM, OVH, Oracle, and Linode) in models (2) and (5), and comparing firms that use Amazon AWS platform to firms that do not (models (3) and (6)).[28] I graph technology dissimilarity on the event timeline starting from year 0, the year the startup outsourced, in the Appendix in Figure C.3

[27] 2,246 startups use a single CSP; 581 startups use more than one CSP

[28] I adjust the CEM matching strategy to ensure that startups using a Big Tech CSP (or AWS CSP) are demographically similar to those that use an Other CSP (or not AWS CSP) and weight regressions equations.

These matched, within-firm, difference-in-differences results are their first step in supporting a causal argument that outsourcing IT development by using a cloud platform impacts high-tech startups' technology adoption. At the firm level, I depict the relationship between development and analytics technology bundle dissimilarity before and after using cloud platform services in Figure 1.[29]

6.C. Endogeneity.

Instrumental variable approach. To address endogeneity issues stemming from unobservable variables and reverse causality, I use the release of an open-source[30] 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. Recent research supports that complementarities exist between AI and human labor (Choudhury et al., 2020; Krakowski et al., 2022; Tong et al., 2021). Google's decision to open-source TensorFlow, enabling it to be altered by startups to fit their development needs, makes these startups more productive and complementary labor more valuable (Rock, 2021). After the open-source release, TensorFlow could be used on any cloud platform. Startups already using the Google platform were less likely to benefit incrementally than startups on other platforms. Though one could argue that startups on Google Cloud Platform (GCP) are more familiar with the prior proprietary version of TensorFlow and could potentially benefit from the open-source tool more quickly, this advantage would be short-lived. A few months after its release, in early 2016, startups could use TensorFlow with any cloud provider's platform.[31]

This instrumental variable approach includes the interaction between two binary variables: startups that 1) benefit from TensorFlow (i.e., startups may benefit from TensorFlow if they are not already on the

[29] The only significant source of heterogeneity in the treatment group is due to startup size. Smaller startups adopt less dissimilar bundles of development technologies and more dissimilar bundles of analytics technologies (Appendix Figure C.1).

[30] Apache 2.0 open source license

[31] 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. However, unlike TensorFlow, these other AI frameworks, Keras, CNTK, and PyTorch, were released initially as open-source products, benefiting firms across all platforms. The only other platform that was initially proprietary and then open-sourced is Amazon's Sagemaker, which was released later, in November 2017, and not open-sourced until 2019.

Google platform in 2016) and 2) produce AI (i.e., startups may need AI frameworks, like TensorFlow).[32] Including an interaction term between two binary variables increases the strength of the first stage without biasing estimation (Aghion, 2005; Bun and Harrison, 2018). The first stage F-statistics (K-P Wald F: 389, C-D Wald F: 158) support that this instrument is adequately powered. Using TensorFlow is directly related to AI startups adopting a cloud platform. And in support of the exclusion restriction, TensorFlow is not directly related to aspects of technology adoption. 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. [33]

The first stage of the instrumental variable finds that the TensorFlow shock relates to increased cloud platform adoption (Table 5.A. model (IV), Tensor: +0.28 SD 0.01), particularly for AI-producing startups (model (1): Tensor x AI: +0.34 SD 0.02). In the second stage regression, the IV approach dampens the impact of cloud platform adoption, though the results are directionally similar and significant. In table 5.B., I report the comparison between the main result and the IV results for dissimilarity measures: the coefficient of the effect using cloud on product development dissimilarity increases by 21% (from model (1): -0.28 SD 0.002 to model (2): -0.022 SD 0.009) and analytics dissimilarity decreases by 25% (from model (4): +0.087 SD 0.005 to model (5): +0.065 SD 0.018).

Double machine learning. As another method of addressing endogeneity from potentially omitted variables (Belloni et al. 2014), I use double machine learning following Chernozhukov et al. (2018) to estimate the treatment and outcome using a random forest machine-learning algorithm trained with 64 firm-level control variables. This approach divides the sample in half, using half the observations to train the model and the other half for prediction, and uses Neyman orthogonal scores to estimate the causal parameter

[32] I provide details on the instrumental variable approach specification in Appendix A.8.

[33] For instance, the indicator variables for the ability to benefit from TensorFlow may not be not fully independent from aspects of similarity, the dependent variable in the second stage. Moreover, 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.

(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. I describe this approach comprehensively in Appendix A.9.

The double machine learning approach further dampens the impact of adding cloud platform services, down about 40% from the base model (model (3): development dissimilarity; -0.015 SD 0.002 and model (6) analytics dissimilarity; +0.053 SD 0.004) yet provides additional support that the effect remains present and significant. As robustness for this machine learning model, I use an Oster coefficient stability test (Oster, 2019; development dissimilarity: $\delta = 3.08$, analytics dissimilarity: $\delta = 1.35$) to support that effect is unlikely to be negated by unobserved variables.[34]

6.D. Robustness.

To show the robustness of my main results to potential selection issues, I report results using only treated firms, excluding firms that never used a cloud provider (Table B.2, models (1) and (6)) and include the inverse of the Mill's ratio derived from the probit analysis reported in Table 2 as a control (models (4) and (9)). The estimate of IMR is significant and positive in the model of development technology dissimilarity, but the main effect results remain similar (-0.025 SD 0.09). Next, to overcome potential concerns that startups have less choice over whether they outsource in later years and cloud services are fundamentally different in earlier periods and to focus analyses on the period where the analytics technology bundle dissimilarity measures have parallel trends, I re-run the main analysis on firms greater than three years old with data from 2012 to 2018 (models (2) and (7)). Third, I include investor dissimilarity[35] as a control for startups that received common resources from sources other than the cloud services provider (models (3) and (8)). For example, 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

[34] Based on Oster (2019), a $\delta = 3.08$ is internpreted as the impact of unobserved variables would need to be 3x greater than the impact of observed variables for the effect to change signs.

[35] 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 the value of 0 if they have no investors in common. Mean investor dissimilarity is 0.59 SD 0.08.

Z startup programs"). Fourth, to ensure a single cloud provider does not drive the effect, I run the main specification for each cloud provider: AWS, GCP, MS, and Other; results remain similar (Table B.4). Fifth, in Table B.5, I show the results of an alternate dependent variable, a technology adoption dissimilarity measure based on if a particular technology is more or less commonly used by startups in the sample.[36]

Sixth, as robustness for my staggered difference-in-differences model with two-way fixed effects, I use several approaches to estimate treatment effects (Table B.3). In the most straightforward approach, I collapse my data to pre and post estimates for a balanced panel of 873 firms (i.e., switchers) that change from not using cloud platform services to using cloud platform services. This two-period difference-in-difference analysis suggests a larger impact of outsourcing cloud platform services on technology adoption dissimilarity (model (1): development technologies, -0.086 SD 0.002 and model (5): analytics technologies, +0.129 SD 0.005). In another approach, I report the average treatment effect, 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 (models (3) and (7); De Chaisemartin and d'Haultfoeuille, 2018). And in the last approach, I estimate heterogeneity-robust instantaneous treatment effects to estimate the treatment in each period (models (4) and (8); Athey and Imbens, 2018; De Chaisemartin and d'Haultfoeuille, 2020). [37] Moreover, this approach estimates prior periods before the outsourcing event, confirming that pre-trends are not an issue. Based on models (4) and (8), I depict the coefficients for the treatment effect for first-time "switchers" in the first period they switched (i.e., untreated in the prior year and treated in the current year) and depict results for each period in Figure 2.

Seventh, I design several analyses to examine whether changes to the focal or rival firm drive this effect, using disaggregated pairwise data (35 million observations) to calculate the pairwise angular

[36] I provide more detail on how this alternate dependent variable measure is constructed in Appendix A.10.

[37] Moreover, 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)

distance between each focal startup and each other startup in the sample (i.e., rival startups) before collapsing the data to the firm-level mean. Then as a final check, I use firm, rival, and year-level fixed effects to show a limited impact of controlling for changes to rivals' development technologies (model (6): -0.028 SD 0.003 vs. model (8): -0.029 SD 0.003). Yet, controlling for changes to rival technologies increases analytics technology dissimilarity (model (14): +0.088 SD 0.005 vs. model (16) +0.096 SD 0.005), suggesting that changes to rival technologies slightly reduced the dissimilarity increase. These additional analyses at the disaggregated level support the main analyses.

Eighth, to address potential issues from the treatment effect spillovers (i.e., stable unit treatment violations assumption; SUTVA), I rerun the main analysis based on the pairwise matching between a focal startup and only rival startups that do not change their technologies. Results are very similar (Table B.6. model (1): development technologies: -0.027 SD 0.003; model (9) analytics technologies +0.083 SD), suggesting that spillovers from the treatment effect did not impact results.

7. Implications for product differentiation and performance

Analyses of technology adoption are interesting in their own right, especially given the close gap between development technologies and end-product in the app production context. Yet, in this paper, they serve as a proxy for innovation. Objectively collected product data from these young, small startups is challenging to find at scale. This is both a limitation of this current study and an opportunity for future research. As another method to proxy product differentiation, I use the quanteda[38] package in R to measure startup description and patent description differentiation. These measures of startup and patent description changes are closer proxies of product innovation; however, these analyses examine a smaller sample size as fewer startups changed their descriptions. I provide information on the construction and description of these measures in Appendix A.2.

[38] 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.

From data on startup descriptions and patent descriptions for a subset of firms, I find that innovations become more differentiated after using a cloud platform service. I have data on startup descriptions for 193 startups that changed their descriptions after treatment, and I find a correlation between using a cloud platform and having more differentiated startup descriptions (Table 6, model (1): +0.016 SD 0.004). This effect increases when startups use more dissimilar data analytics technology bundles (model (3): interaction: cloud platform x high analytics dissimilarity, +0.027 SD 0.009), corroborating that more distinct bundles of analytics technologies help startups differentiate their products from rivals. Next, I have data on patent descriptions before and after adding cloud platform services for an unfortunately small number of startups (20 startups), which yields a similar correlation (model (4): +0.007 SD 0.004). Similarly, this effect is (weakly) increased when startups use more dissimilar data analytics technology bundles (model (6): interaction: cloud platform x analytics dissimilarity, +0.048 SD 0.028). Though I only have data on a subset of startups, this analysis provides additional insight into the relationship between types of technologies, the fit with the cloud platform, and the differentiation of end-product apps.

The adoption of more or less dissimilar bundles of technologies relates to startup fundraising performance, including measures of funding [0,1], follow-on funding [0,1], and deal size (log). The increased similarity of product development technology bundles correlates with increased funding (Table 7 model (3), deal size (log): +3.3 SD 1.2), increased dissimilarity of analytics technology bundles correlates with decreased performance (model (8), deal size (log): -2.3 SD 0.71). This increase in funding suggests that startups adopting more similar development technologies benefit from increased fit and compatibility with the cloud platform. Moreover, startups raise more funding when adopting a more diverse bundle of analytics technologies, supporting that data resources help differentiate products. I depict this relationship with bin scatter plots replicating the main specification in Figure 3.

For robustness, I use alternate dependent variables, including VC funding [0,1] and higher reputation VC funding, to support this interpretation further (Table B.7). Additionally, I examine the interaction between an indicator variable for larger bundles of technologies and higher levels of technology differentiation (Table B.8). Next, I examine web traffic as an additional dependent variable and measures

of product differentiation as additional independent variables. I find a significant correlation between adopting bundles of more similar development technologies and increased web traffic (i.e., duration on the website; Table 7 model (4): -3.56 SD 0.48) and patenting (model (5): -0.08 SD 0.023). Product differentiation measures also correlate with funding outcomes, even with the smaller sample size. Increased description differentiation and increased patent differentiation relate to increased funding.

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 tons of resources with the startup ecosystem impacts the breadth of technological innovation. Cloud platforms are here to stay. They continue to grow in capabilities and scale, providing startups access to IT assets and production-related resources, making it easier for startups to begin developing their digital products. Furthermore, changes to the venture capital funding model and the advent of the pay-as-you-go cloud services make it easier for more startups to access the initial funding needed to enter this industry. However, macro-economic benefit from innovation is about more than just entry and the count of innovations. Less differentiated innovations may benefit the economy less, driving lower productivity growth than anticipated.

My findings explain that using cloud platforms impacts the breadth of technologies startups adopt, depending on the needed fit of that technology with the cloud platform and other interdependent technologies for products to work. For product development technologies, where fit amongst technologies and with the platform is important, compatibility issues constrain and shared resources influence technology adoption. Moreover, investor rewards app-developing startups for adopting more standardized bundles of development technologies that fit better with the cloud platform, suggesting that top-performing startups converge to adopt bundles of the latest and greatest compatible product development technologies.

This finding, taken on its own, would raise concerns that cloud providers curate which bundle of compatible technologies should be used to develop products, directly and indirectly altering technology adoption. However, it is only half the story. Even when using more similar development technology bundles, startups can differentiate products using a more diverse analytics technology bundle that provides better data resources as an output. These analytics technologies are unconstrained by their fit with the underlying platform. Examination of changes in startups' descriptions and patents after adding cloud services supports that using more similar product development technology bundles does not necessarily yield less differentiated products. However, using analytics technology bundles dissimilar to other app-producing startups relates to increased product differentiation. And investors reward startups with more standardized development bundles and diverse analytics technologies. I employ numerous econometric approaches (i.e., matching, instrumental variable, firm fixed effects, double machine learning) to support a causal argument based on observed and unobserved variation to rule out alternate explanations and adjust estimates. 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.

Tying back to transaction cost economics, firms become more dependent on suppliers when outsourcing the development of production inputs. Over time, firms make asset-specific investments in complementary assets, increasing their fit with the supplier but limiting their ability to switch suppliers and increasing dependence on their supplier over time. In this case, startups cede control over aspects of their production to the technology firms managing these cloud platforms. These cloud services providers 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. This increasing control over product development on their platforms will enable cloud suppliers with a high market share to influence the nature of future innovations.

A comparison of cloud providers raises two additional learnings. First, there are significant differences in technology adoption changes for startups using the largest three cloud platforms, each of which has startup-related programs that share resources, compared to startups using smaller cloud platforms, which do not have these programs. Programs that share platform-related resources lower the cost of startups accessing compatible development technologies that are guaranteed to fit well with the platform. Second, at a more granular level, using the largest market share platform, Amazon AWS, related to using more standardized development technology bundles and more distinct analytics technology bundles. These findings are particularly interesting because of the competitive concerns around a single cloud platform with a high market share wielding greater control over the startups on its platform than smaller suppliers. Yet high-tech startups can still differentiate their product by using bundles of more diverse analytics technologies that produce unique, proprietary data.

Cloud providers will consolidate market share, and this industry will evolve to meet the needs of all types of customers. It is already apparent that specific industries and verticals flock predominately to one provider over another, such as healthcare firms using Microsoft Azure or startups using Amazon AWS. What will remain constant is that unique, proprietary, and representative data are crucial to differentiation for high-tech startups. Yet startups' data resources pale in comparison to the data resources of the technology firms hosting these cloud platforms (Benzina, 2019; Iansiti, 2021; Khan, 2016; Scott-Morton et al., 2019). In addition to having less robust data technologies, processes, and resources, startups share tons of information about their product and industry with cloud providers through their corporate accelerator programs. Moreover, startups provide feedback on improving cloud IT offerings and passively provide their usage data on their providers' platform (i.e., telemetry; a user's digital footprint on the platform, or usage "exhaust;" Chatterji and Fabrizio, 2014), which the cloud provider can use to improve their own platform and products. This information is likely to be competitively valuable when aggregated across many startups.

In the most egregious cases, cloud providers could use this information to make acquisitions directly shaping the technological landscape (Cunningham et al., 2019; Zingales et al., 2021). In a much

more likely situation, they may alter the compatibility of their platforms and technologies to match their development goals or incentivize adopting technologies important to them or their largest partners. And as transaction cost economics teaches, market consolidation and the gradual increase of supplier control will not bode well for these startups in the long run, particularly if they compete against their cloud providers in unregulated downstream markets.

Tables

Table 1.A. – Firm-level Summary

	Unmatched								Matched					
	All Startups				No CSP		CSP		All Startups		No CSP		CSP	
	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: Summary statistics are calculated at the firm x year level for all firms in the sample. Matching uses CEM: age #10, employment size #10, healthcare, financial services, and region #4 based on the treatment, cloud platform versus no cloud platform, dropping 19 unmatched firms in the main analyses. Firms included in the sample have a digital application, active web domain, are ten or fewer years old, have fewer than 500 employees, and are not located in China. 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 startup level and are from AIdentified, Mantheos, Pitchbook, and manual collection of public profiles on LinkedIn.

Table 1.B. – Firm-level Summary

	Unmatched								Matched					
	All Startups				No CSP		CSP		All Startups		No CSP		CSP	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Technology Adoption														
All	49.87	26.5	0	252	36.0	18.9	53.3	26.9	49.6	25.8	39.1	20.2	52.3	26.4
Development Tech Count	8.18	4.1	0	27	5.5	3.1	8.8	4.1	8.1	4.0	5.9	3.2	8.7	4.0
Analytics Tech Count	6.40	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	12.46	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	3.45	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	1.18	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														
Startup 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 Dur.	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: Summary statistics are calculated at the firm x year level for all firms in the sample. Matching uses CEM: age #10, employment size #10, healthcare, financial services, and region #4 based on the treatment, cloud platform versus no cloud platform, dropping 19 unmatched firms in the main analyses. Firms included in the sample have a digital application, active web domain, are ten or fewer years old, have fewer than 500 employees, and are not located in China. Information on technologies is from BuiltWith.; performance information is from Crunchbase, Pitchbook, IPQwerty, and SimilarWeb.

Table 1.C. – CSP Adoption Panel Summary

	Startups (3,434)				Observations (19,678)			
	No CSP	PaaS	IaaS	IaaS/ PaaS	No/Before CSP	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 CSP		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.

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 Tech 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								
Startup 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 (unique visits)	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 CSP: AWS, GCP, Azure, or Other CSPs (e.g., Linode, Digital Ocean, etc.)

Table 2 – Probit Selection Analysis

Treatment, DV is:	(1) CSP vs. No CSP	(2) Big Tech vs. Other CSP	(3) AWS Only vs. Not AWS	(4) mprobit No/Before (0); Big (1); Other (2)	
				=1	=2
[0,1] Healthcare	-0.139**	0.078	-0.022	0.065	-0.028
	(0.065)	(0.075)	(0.074)	(0.046)	(0.075)
[0,1] Prior Big Tech Funding	0.289***	0.003	-0.120**	0.133	-0.265
	(0.107)	(0.069)	(0.059)	(0.087)	(0.163)
Age (log)	0.644***	0.438***	0.333***	-0.022	-0.136
	(0.013)	(0.009)	(0.008)	(0.055)	(0.086)
[0,1] Europe	-0.199***	-0.263***	-0.257***	-0.445***	0.233***
	(0.044)	(0.050)	(0.050)	(0.029)	(0.043)
[0,1] Founder OEM/HW Exp.	-0.183**	0.077	0.024	-0.091	-0.005
	(0.076)	(0.070)	(0.063)	(0.061)	(0.098)
[0,1] Founder MBA	0.192***	0.128***	0.083***	0.101***	-0.316***
	(0.041)	(0.034)	(0.030)	(0.032)	(0.056)
[0,1] Founder Technical Degree	0.201***	-0.037	-0.020	0.245***	0.076*
	(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 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, whereas model (4) includes the base case, No CSP = 0.

Table 3 - Technology Bundle Size

DV is log of :	Development Technology Count						Ratio: Dev. Count/All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
[0,1] CSP PaaS	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)	
PaaS x IaaS						0.353***	
						(0.021)	
R2	0.100	0.491	0.168	0.588	0.586	0.594	0.502

	Analytics Technology Count						Ratio: Ana. Count/All
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
[0,1] CSP PaaS	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)	
CSP 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 Weighted	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). Models (6) and (13) examine the interaction between CSP and IaaS (cloud infrastructure); in these models, IaaS, an indicator variable for using cloud infrastructure services. Models (5-7) and (12-14) models drop unmatched startups and weight regressions based on CEM: age #10, employment size #10, healthcare, financial services, and region #4.

Table 4 – Technology Bundle Dissimilarity

DV is:	Development Technology Dissimilarity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
[0,1] CSP PaaS	-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)	
CSP PaaS x						-0.056***	
H. Tech Count						(0.003)	
[0,1] IaaS							-0.004
							(0.005)
CSP PaaS x							-0.032***
IaaS							(0.002)
R2	0.0470	0.385	0.355	0.668	0.673	0.701	0.675

DV is:	Analytics Technology Dissimilarity						
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
[0,1] CSP PaaS	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)	
CSP PaaS x						0.108***	
H. Tech Count						(0.005)	
[0,1] IaaS							0.017
							(0.011)
CSP PaaS x							0.086***
IaaS							(0.005)
R2	0.110	0.502	0.185	0.575	0.577	0.590	0.577
Observations	18802	18802	18802	9930	18802	18802	18802
Firms	3123	3123	3123	1938	3123	3123	3123
Firm FE	No	Yes	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes	Yes
CEM Weighted	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 5.A. – IV (First Stage)

DV is:	IV CSP
[0,1] Tensor	0.28***
	(0.01)
[0,1] AI	0.08***
	(0.01)
Tensor x AI	0.34***
	(0.02)
Observations	18802
Firms	3123
K-P Wald F	389
C-D Wald F	158
K-P LM	212

Table 5.B. – IV (Second Stage) and DML

DV is:	Development Dissimilarity			Analytics Dissimilarity		
Model:	Base	IV	DML	Base	IV	DML
	(1)	(2)	(3)	(4)	(5)	(6)
[0,1] CSP PaaS	-0.027***	-0.022**	-0.015***	0.086***	0.065***	0.053***
	(0.002)	(0.009)	(.002)	(0.01)	(0.018)	(0.004)
Observations	18841	18841	18841	18841	18841	18841
Firms	3128	3128	3128	3128	3128	3128
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 matching (CEM: age #10, employment size #10, healthcare, financial services, and region #4). Model (1) is repeated from Table 4 model (4); model (4) is repeated from Table 4 model (8); Models (7) and (9) are estimated by adapting the main specification for the addition of IaaS instead of CSP/PaaS.

Table 6 – Technology Adoption Bundle Dissimilarity and Funding Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DV is:	Funded		Follow-on Funding		Deal Size (log)		Web Visit Dur. (log)		Patent	
Development	-0.796***		-0.888***		-3.306***		-3.558***		-0.084***	
Dissimilarity	(0.068)		(0.073)		(1.215)		(0.484)		(0.023)	
Analytics		0.166***		0.124***		2.309***		-0.250		0.012
Dissimilarity		(0.043)		(0.045)		(0.711)		(0.342)		(0.013)
Observations	17628	17628	17628	17628	17628	17628	9200	9200	17628	17628
R2	0.691	0.686	0.690	0.685	0.165	0.165	0.428	0.422	0.820	0.820
Firms	2799	2799	2799	2799	2799	2799	2053	2053	2799	2799
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

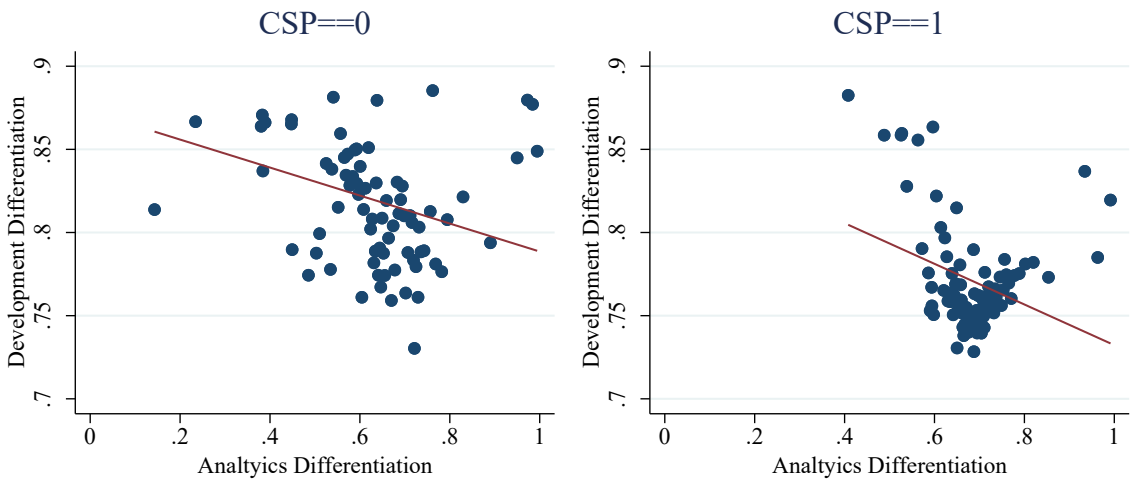
Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with firm-level and year-level fixed effects in all models; standard errors are clustered at the firm level. All models drop unmatched startups and weight regressions based on CEM: age #10, employment size #10, healthcare, financial services, and region #4. Includes only the treatment group, startups using CSP PaaS.

Table 7 – Technological Dissimilarity and Product Differentiation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DV is:	Startup Description Differentiation			IP Description Differentiation			IP Count (log)		
[0,1] CSP PaaS	0.016***			0.007*			0.279**		
	(0.004)			(0.004)			(0.117)		
Dev. Dissimilarity (cont.)		-0.015			-0.072			-1.409**	
		(0.022)			(0.063)			(0.657)	
CSP PaaS x		0.005			-0.070			-1.328*	
Dev. Dissimilarity		(0.025)			(0.065)			(0.693)	
Analytics Dissimilarity (cont.)			0.004			0.044			-0.195
			(0.008)			(0.033)			(0.380)
CSP PaaS x			0.027***			0.048*			0.029
Ana. Dissimilarity			(0.009)			(0.028)			(0.325)
Observations	382	382	382	40	40	40	880	880	880
R2	0.0698	0.0694	0.0692	0.0282	-0.0127	0.0476	0.0145	0.235	0.227
Firms	191	191	191	20	20	20	156	156	156
Firm FE	No	No	No	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

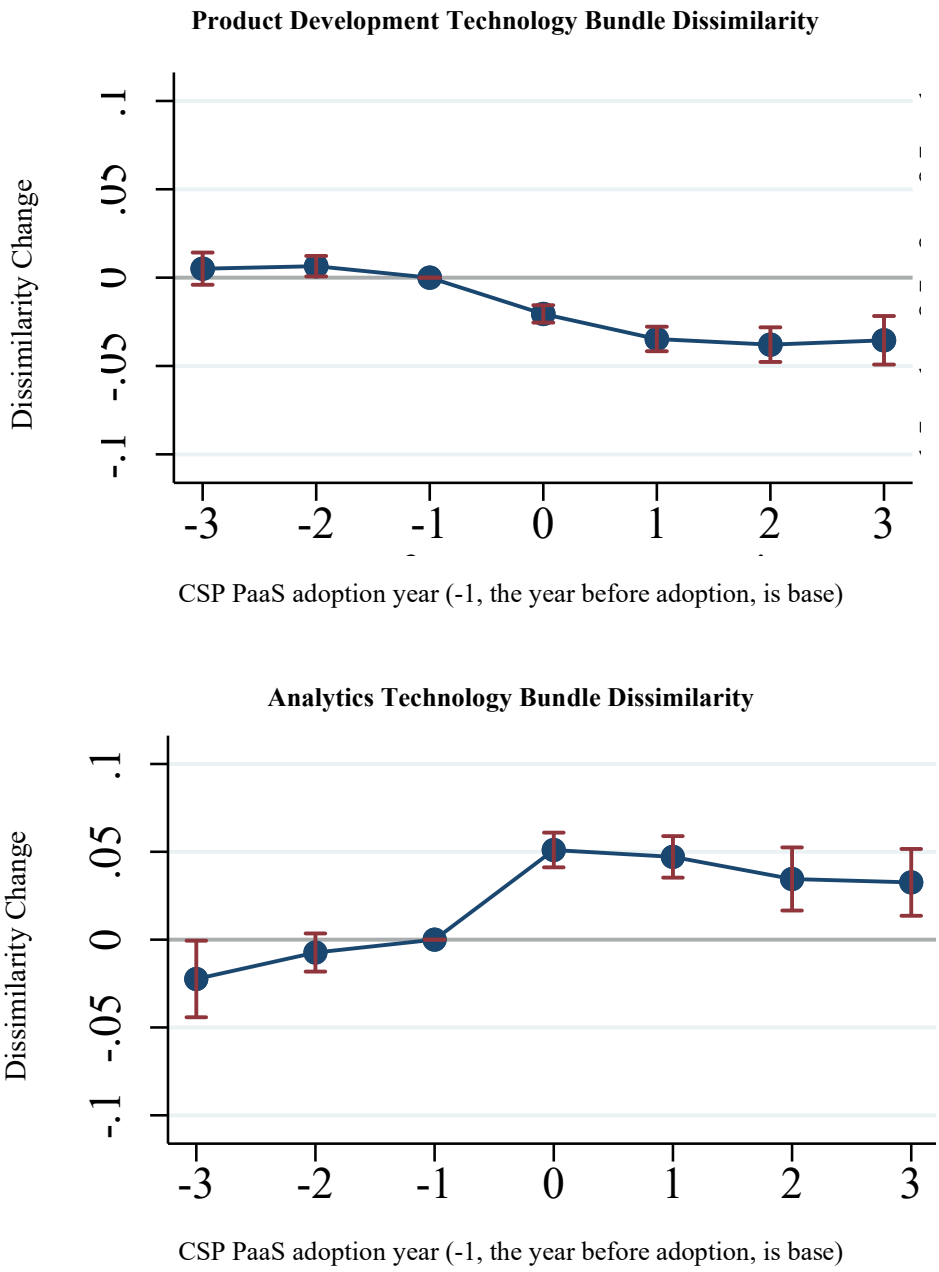
Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS with year-level fixed effects in all models; standard errors are clustered at the firm level. In models (1)-(6), the panel is balanced, with one observation for each startup before and after adding CSP PaaS. Models (7)-(9) also include firm-level fixed effects since many firms patent in multiple years.

Figure 1 – Technology Bundle Dissimilarity Comparison



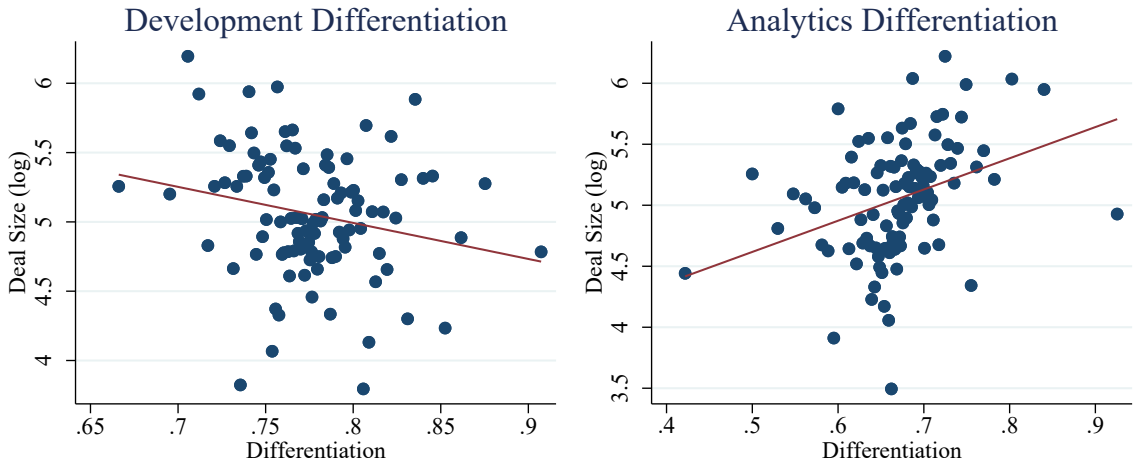
These charts use a bin scatter (100 points) with no control, only CEM matching, to depict the correlation before and after the treatment.

Figure 2 – CSP Adoption Event



Coefficients are estimated using Chaisemartin and D'Haultfoeuille 2020 to account for issues arising in a two-way fixed-effect design, which do not differentiate between observations that have never been treated or have not yet been treated. Each point is the coefficient of the effect in a given year based on the number of "switchers" in that 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 3 – Technology Differentiation and Performance



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., 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., & Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society*: Series B (Statistical Methodology), 80(4), 597-623.

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.

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.

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.

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.

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.

Chandler, A. D. (1962). Strategy and structure: Chapters in the history of the industrial empire. Cambridge, Mass.

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.

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).

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.

Klein, B., Crawford, R. G., & Alchian, A. A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *The Journal of Law and Economics*, 21(2), 297-326.

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*.

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 statsitics, 213-234.

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.

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).

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.

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.

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.

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.

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.

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

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

<u>Technology Adoption.</u>

- *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 and Patent Descriptions.

Patent descriptions are from patent abstracts (WIPO, USPTO) provided by IPQwerty. Of the sample, 156 firms have 880 patents combined. Similar methods have been used to construct patent text similarity in the strategy literature by Arts et al. (2017). 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) and patent description differentiation is .93 (SD 0.02).

Tables A.3.-A.5. – Correlation Tables

Table A.3. – Correlation (Demographics)

		(1)	(2)	(3)	(4)	(5)
(1)	Dev. Sim	1				
(2)	Data Sim	-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. Sim	1					
(2)	Data Sim	-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)

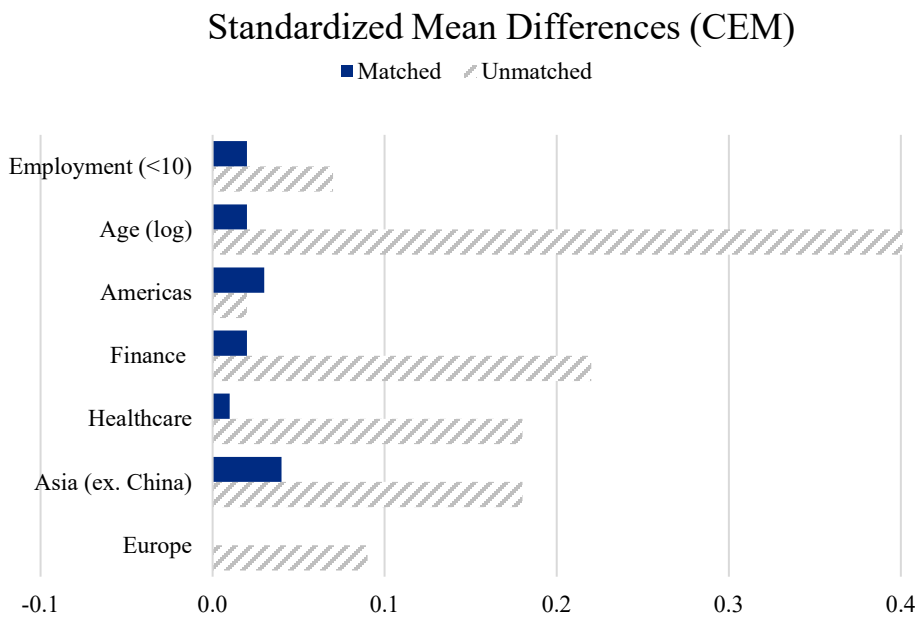
		(1)	(2)	(3)	(4)	(5)
(1)	Dev. Sim	1				
(2)	Data Sim	-0.21*	1			
(3)	Funds Raised (log)	-0.21*	0.18*	1		
(4)	Funded	-0.19*	0.16*	0.79*	1	
(5)	VC Backed	-0.21*	0.15*	0.68*	0.79*	1
(6)	Invest. Sim	-0.12*	0.078*	0.22*	0.078*	0.15*

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 CSP versus not, or the decision to add a certain CSP). I use IMR as a control in the second stage of the regression to address potential selection issues.

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

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

where, $adoption_i$ 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:

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

where, $cloud\_platform_{it}$ refers to the binary dependent variable: adopting a CSP 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., year=2016 and startup not on Google's platform), 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.[39] Specification Z.4 is the main predictive model, and specification Z.5 constructs Neyman orthogonal scores (Chernozhukov et al., 2018; Neyman, 1959; Wooldridge, 1991).

$$tech\_diff_{it} = \theta csp_{it} + g_0(x_{it}) + year_t + \varsigma_{it} \qquad (Z.4)$$

$$csp_{it} = m_0(x_{it}) + year_t + \upsilon_{it} \qquad (Z.5)$$

Where $year_t$ is an indicator variable for all observed years (2012 through 2021); x_{it} is a vector of covariates error terms; ς_{it} and υ_{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

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

training and the remaining 50% for prediction. In the model, I include every possible covariate from the data I have collected on these firms, including more than 60 measures on patents, website traffic, firm demographics, and performance. 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.

$$(tech\_diff_{it} - \widehat{tech_{diff}}_{it}) = \beta_1(csp_{it} - \widehat{csp_{it}}) + \beta_2 yearFE_t + \beta_3 firmFE_i + \varepsilon_{it} \qquad \text{(Z.6)}$$

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.

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

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.

Appendix B – Additional analyses and robustness

Table B.1. – Other Technology Counts

DV is log of :	(1) All	(2) Big	(3) Ratio: Big/All	(4) Open	(5) Premium	(6) Ratio: Open/Prem
[0,1] CSP PaaS	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. – Robustness for Staggered DiD with Two-way Fixed Effects

	Development Dissimilarity				Analytics Dissimilarity			
	Pre/Post	ATE	LATE	ITE	Pre/Post	ATE	LATE	ITE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
[0,1] Post	-0.086***				0.129***			
	(0.002)				(0.005)			
[0,1] CSP PaaS		-0.011***	-0.014***	-0.021***		0.036***	0.139***	0.051***
		(0.002)	(0.007)	(0.02)		(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			873	873			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.3. – Technology Dissimilarity Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DV is:	Development Dissimilarity					Analytics Dissimilarity				
	Balanced	bf. 2018	Inv. Diff.	Probit	IMR	Balanced	bf. 2018	Inv. Diff.	Probit	IMR
[0,1] CSP PaaS	-0.027***	-0.026***	-0.033***	-0.107***	-0.025***	0.092***	0.082***	0.079***	0.254***	0.086***
	(0.002)	(0.003)	(0.004)	(0.008)	(0.002)	(0.005)	(0.006)	(0.008)	(0.013)	(0.005)
Investor			-0.017					0.022		
Dissimilarity			(0.012)					(0.021)		
Inv. Mills Rat.					0.089***					-0.008
					(0.007)					(0.012)
Observations	17045	8445	9319	17685	18802	17045	8445	9319	17685	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	No	Yes	Yes	Yes	Yes	No	Yes
Year FE	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
CEM Weighted	Yes	Yes	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.4. – Technology Dissimilarity (by Platform)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Development Dissimilarity				Analytics Dissimilarity			
[0,1] CSP AM	-0.029***				0.074***			
	(0.003)				(0.007)			
[0,1] CSP GCP		-0.026***				0.086***		
		(0.006)				(0.012)		
[0,1] CSP MS			-0.008				0.082***	
			(0.010)				(0.014)	
[0,1] CSP Oth				-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.

Table B.5. – Alternative DV: Average Installed Base of Firm-Level Tech Bundle

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	Development Tech Bundle (Avg. IB)				Analytics Tech Bundle (Avg. IB)			
[0,1] CSP PaaS	0.089***	0.090***	0.071***	0.089***	-0.143***	-0.100***	-0.161***	-0.134***
	(0.020)	(0.022)	(0.024)	(0.020)	(0.024)	(0.025)	(0.031)	(0.024)
[0,1] H. Tech Count		0.006				-0.282***		
		(0.016)				(0.031)		
CSP PaaS x		0.092***				-0.351***		
H. Tech Count		(0.021)				(0.027)		
[0,1] IaaS			0.011				0.005	
			(0.033)				(0.044)	
CSP PaaS x			0.097***				-0.136***	
IaaS			(0.021)				(0.026)	
[0,1] H. Inv. Sim.				0.012				-0.007
				(0.028)				(0.038)
CSP PaaS x				0.102***				-0.188***
H. Inv. Sim.				(0.024)				(0.028)
Observations	18802	18802	18802	18802	18802	18802	18802	18802
R2	0.818	0.818	0.818	0.818	0.702	0.715	0.702	0.702
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
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.6. – Disaggregated Data: Technology Dissimilarity

	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] CSP PaaS	-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] CSP PaaS	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.7 – Technology Similarity and 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)
Dev.-Related						
L Tech x Less Diff.	0.053***	0.042***	-0.225			
	(0.009)	(0.010)	(0.202)			
L Tech x More Diff.		*base*				
H Tech x Less Diff.	0.108***	0.121***	0.837***			
	(0.009)	(0.010)	(0.193)			
H Tech x More Diff.	0.094***	0.084***	0.825***			
	(0.010)	(0.011)	(0.199)			
Data Related						
L Tech x Less Diff.					*base*	
L Tech x More Diff.				0.012	-0.000	0.330*
				(0.010)	(0.011)	(0.183)
H Tech x Less Diff.				0.099***	0.101***	0.653***
				(0.011)	(0.012)	(0.231)
H Tech x More Diff.				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.

Table B.8 – Additional Performance Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV is:	VC Backed		VC Rep.		Closed		Acquired	
Development	-0.721***		-0.195***		0.016		0.011	
Dissimilarity	(0.066)		(0.035)		(0.015)		(0.025)	
Analytics		0.115***		0.013		0.000		0.031***
Dissimilarity		(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.9 – Platform Market Share

	(1)	(2)	(3)	(4)	(5)	(6)
DV is:						
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.

Appendix C – Additional Figures

Figure C.1 – Heterogeneity in Technology Dissimilarity by Firm Size

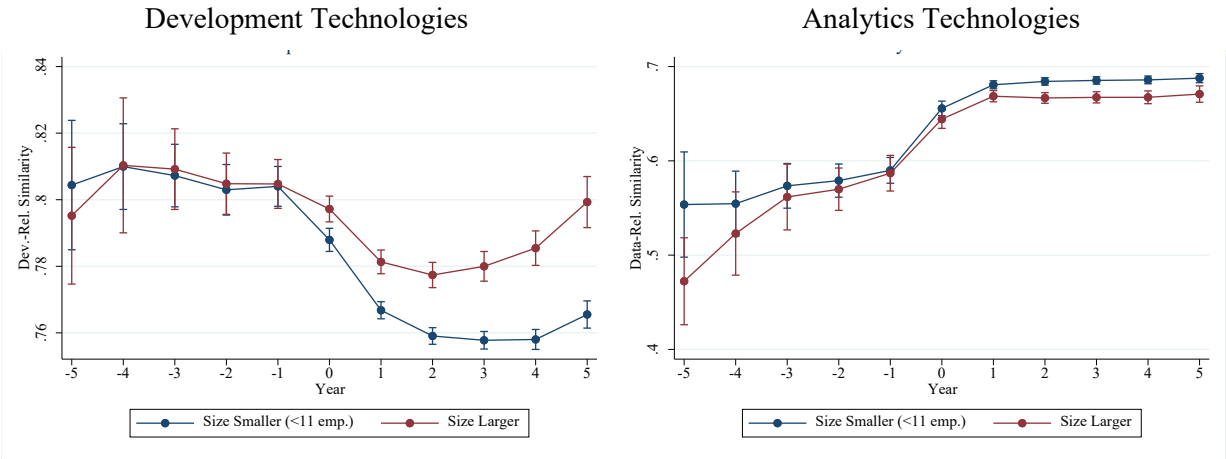


Figure C.2 – Percentage of startups (<3 years old) outsourcing IT to cloud platforms

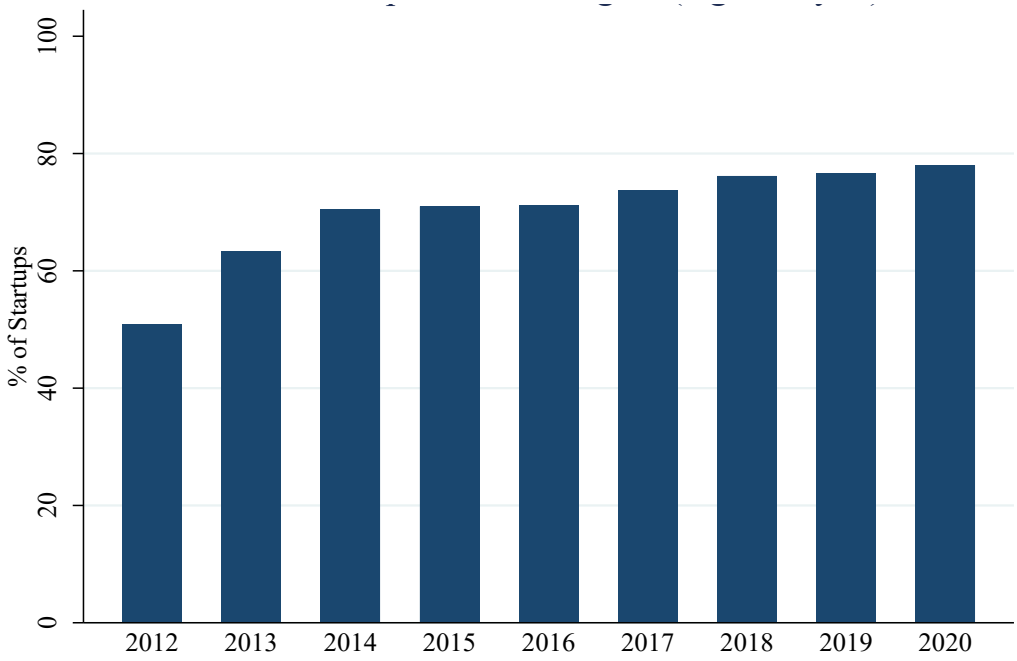


Figure C.3 – Technology Dissimilarity (AWS vs. not AWS)

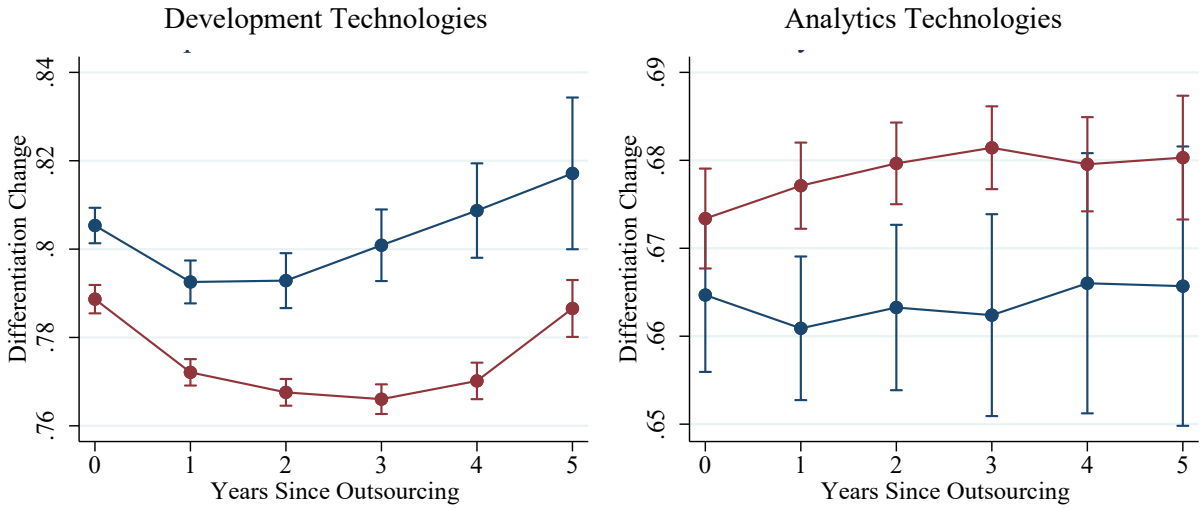


Figure C.4 – Kernel Density (CSP vs. no CSP)

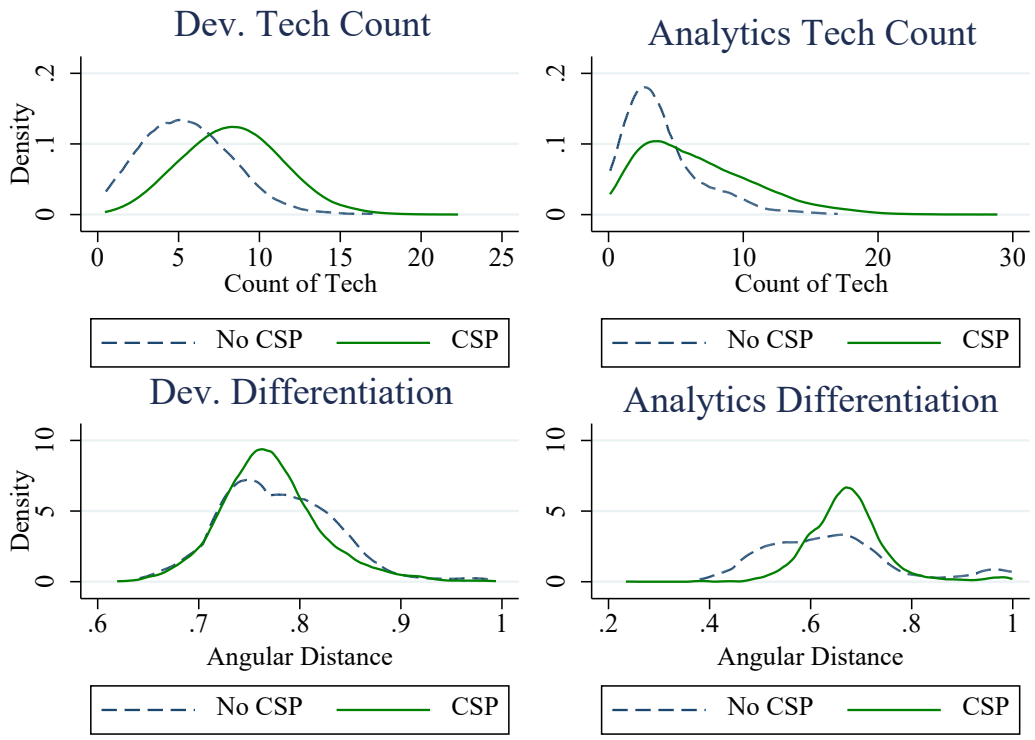


Figure C.5 – Difference in technologies by CSP PaaS versus no CSP PaaS

Figure C.5.A. – Technology Bundle Size

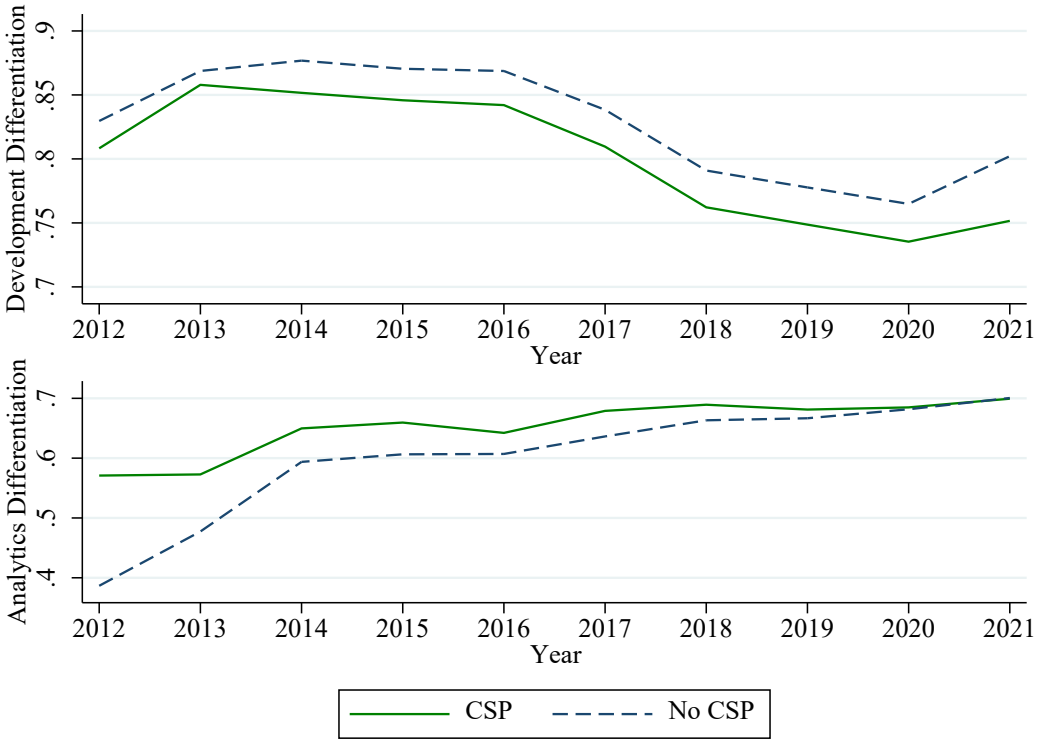
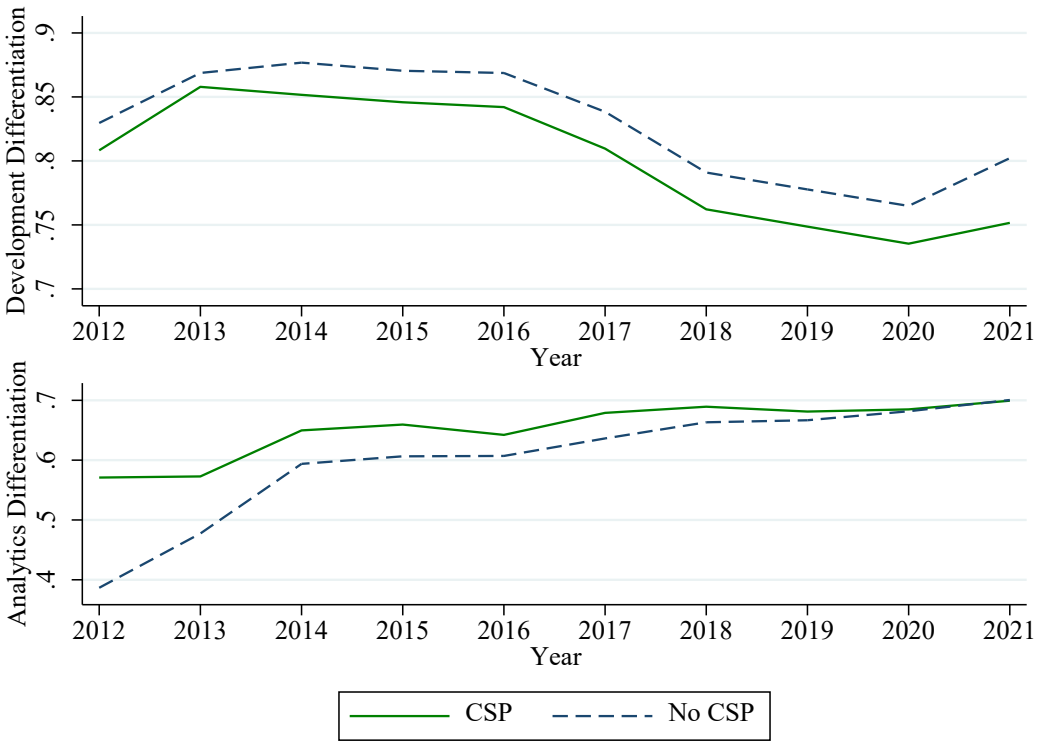


Figure C.5.B – Dissimilarity



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.