The Adoption of Artificial Intelligence by Venture Capitalists

Maxime Bonelli*

Job Market Paper

[Click here for the latest version]

November 7, 2022

Abstract

I study how the adoption of artificial intelligence (AI) by venture capitalists (VCs) to screen startups affects the funding of early-stage companies. Using global data on VC investments, I show that after adopting AI, VCs tilt their portfolios towards startups whose business is similar to those already tested by past startups. Within this pool of startups, AI-empowered VCs become better at picking those that survive and receive follow-on funding. At the same time, these VCs' investments become 18% less likely to result in breakthrough success. I exploit plausibly exogenous variation in VCs' incentives to automate screening from the introduction of Amazon's Web Services to establish causality between AI adoption and the above effects. Overall, my results are consistent with AI exploiting past data that are not informative about breakthrough companies. AI adoption by investors may therefore reduce the capital directed towards innovation.

Keywords: artificial intelligence, big data, venture capital, entrepreneurship, innovation, capital allocation.

*HEC Paris. **Email:** maxime.bonelli@hec.edu. I am grateful to my advisor Augustin Landier as well as Bruno Biais and Denis Gromb for continuous support and guidance. I also thank Simona Abis, Pat Akey, Pierre Azoulay, Tania Babina, Jean-Noël Barrot, Shai Bernstein, Thomas Bourveau, Taha Choukmane, Lauren Cohen, Jean-Edouard Colliard, François Derrien, Marco Di Maggio, Thierry Foucault, Robin Greenwood, Jorge Guzman, Johan Hombert, Victoria Ivashina, Jessica Jeffers, Thies Ingerslev Jensen (discussant), William Kerr, Hugues Langlois, Josh Lerner, Anton Lines, Clément Mazet-Sonilhac, Evren Örs, Jonathan Parker, Noémie Pinardon-Touati, Matthew Rhodes-Kropf, Julien Sauvagnat, Lawrence Schmidt, Antoinette Schoar, Eric So, David Sraer, Scott Stern, Per Strömberg (discussant), Johnny Tang (discussant), David Thesmar, Boris Vallée, Laura Veldkamp, Adrien Verdelhan, Emil Verner, Luigi Zingales, and participants at the 2022 EFA Doctoral Tutorial, the 2022 Oxford-Man Institute Artificial Intelligence and Financial Markets Workshop, the MIT Finance Lunch Seminar, the 2022 HEC Paris Finance PhD workshop, the 2022 Trans-Atlantic Doctoral Conference, and HEC Paris Brownbag Seminar for helpful comments and feedback. This paper was awarded the 2022 EFA Doctoral Tutorial Best Paper Prize.

1 Introduction

The past two decades have witnessed rapid advances in data availability and statistical techniques such as machine learning and artificial intelligence (AI).[1] These technologies, designed to identify statistical patterns in large datasets, outperform humans in many forecasting tasks. However, their adoption by financial intermediaries has raised concerns regarding their effects on investment decisions and, more broadly, on the allocation of capital.[2]

In this paper, I focus on a key class of financial intermediaries—venture capitalists (VCs). VCs are private equity investors that provide capital for startups with high growth potential and play a crucial role in the financing of innovation.[3] In recent years, dozens of VCs have adopted AI technologies for screening startups (i.e., sourcing, evaluation and selection).[4] Put simply, these VCs employ AI algorithms to detect patterns in historical data from previous startups and extrapolate them to predict a new startup's outcome. Despite using cutting-edge advances in machine learning, these forecasting algorithms are in essence backward-looking—because they are trained using past data. As such, they may not be successful at screening startups that differ radically from past companies. This feature may be particularly significant for venture capital, a form of financing used primarily by forward-looking, innovative startups. This paper is the first to tackle two questions about how the adoption of AI by VCs affects startup financing: Does using AI affect VCs' ability to screen startups and in particular the most innovative ones? If so, what are the consequences for the composition of startups and innovation funded by VCs?

I show that VCs that adopt AI become better at identifying good quality startups, i.e., those that survive and receive follow-on funding, but only within the pool of startups whose business is similar to that developed by past companies. At the same time, VCs that adopt AI become less likely to invest in startups that achieve major success (e.g., an IPO) or a breakthrough (e.g., highly-cited patents). This finding is associated with an increase in the share of their investments being oriented towards startups developing businesses closer to those already tried-and-tested. Overall, my paper shows that AI adoption by VCs reduces the share of VC funding directed to breakthrough companies.

To study AI adoption by VCs, I build a dataset on VC investments worldwide. From Crunchbase, I collect global data on VC firms' investments as well as on acquisitions and IPOs

[1] See for example Agrawal et al. (2018).

[2] See for example O'neil (2016); Mullainathan and Spiess (2017); Fuster et al. (2021); Bartlett et al. (2021).

[3] Among public firms founded within the last fifty years, VC-backed companies account for more than 92% of R&D spending and patent value (Gornall and Strebulaev, 2021). Schoar (2010) discusses the role of transformational entrepreneurs in the economy.

[4] See for instance wsj.com[...]/vc-firms[...]are-using-it and nytimes.com[...]/vc-[...]-choice-of-targets.

of startups. In addition, I collect data on patent applications and grants of VC-funded companies from the US, Canadian, European and World patent offices. I also obtain patent citations from Google Patents. Finally, I gather VC fund-level data from Preqin.

I develop a classification of VCs to determine whether and when they adopt AI. Using employee data from Crunchbase, I identify VC firms that hire data scientists who develop machine learning algorithms for investment screening, and I call these employees AI-related employees. Using job starting dates, I classify a VC as becoming AI-empowered from the date it hires one AI-related employee.[5]

AI-empowered VCs use machine learning algorithms to identify statistical patterns in historical data on past startups (e.g., payment data, website traffic, app downloads, and online reviews), and select new startups based on comparable quantitative metrics. To proxy for the extent to which a startup can be identified and evaluated based on past data, I construct a measure of "backward-similarity". Specifically, I measure the textual similarity of a startup's business description available in Crunchbase to those of previous VC-funded startups in the same industry.[6] High backward-similarity startups run businesses similar to those that have been already tested by past startups. Therefore, historical data on comparable companies presumably exist such that algorithms can identify quantitative patterns.

As a first step, without claiming to establish causality, I document novel evidence consistent with AI-empowered VCs becoming better at screening startups similar to those of the past but not at detecting breakthrough companies. First, I find that after adopting AI, VCs deploy more capital and tilt their investments towards backward-similar startups. Specifically, following AI adoption, AI-empowered VCs increase by 30% more their number of investments per year (1.2 more deals on average) and assets under management (AUM) than do other VCs. After adopting AI, compared to other VCs, the number of investments AI-empowered VCs make in high backward-similarity startups (top quartile) doubles (1.1 more deals on average), while that of investments in low backward-similarity startups (bottom quartile) does not change.

Second, I find that AI-empowered VCs become better at picking good quality startups but only within the pool of those that are highly backward-similar. I find that after adopting AI, compared to other VCs, the likelihood that AI-empowered VCs select a startup that survives and later receives follow-on funding increases by 4 p.p. (an 8% increase compared to the

[5]My methodology does not classify as AI-empowered the VC firms employing data scientists who advise the startups in their portfolio but do not develop screening algorithms.

[6]My backward-similarity measure is similar in spirit to the measurement of technological innovation by Kelly et al. (2021). The latter exploits text data from patent documents to measure the "backward-similarity" and "forward-similarity" of a given patent, i.e., its textual similarity to previous and subsequent patents respectively.

unconditional rate of 52 p.p.) when investing in high backward-similarity startups. However, I observe no significant change when AI-empowered VCs invest in other startups.

Third, I find that, overall, AI-empowered VCs become less likely to invest in startups that achieve major success or a breakthrough. After adopting AI, compared to other VCs, the likelihood that AI-empowered VCs select a startup that eventually goes public via an IPO or is acquired for a profit decreases by 2.7 p.p. (an 18% decrease compared to the unconditional rate of 15 p.p.). In terms of innovation, following AI adoption, relative to other VCs, AI-empowered VCs invest in startups that file 6% fewer future patent applications (0.5 fewer applications on average) and obtain 3% fewer future highly-cited patents (0.1 fewer patents on average).

I do not find that the performance of AI-empowered VC funds changes significantly after AI adoption, compared to other VC funds. This suggests that adopting AI might allow VC firms to scale-up without impacting fund performance and thus to earn more fees.

The above results do not indicate whether AI adoption is causally responsible for the said changes in VCs' investments, given AI adoption is an endogenous decision made by VCs. To address this issue, I exploit plausibly exogenous variation in AI adoption induced by a shock to one often-cited determinant of a VC's decision to adopt AI: the number of potential investment opportunities it faces.[7] Indeed, given the large fixed costs of evaluating investments (Gompers et al., 2020) and the limited scalability of VC firms (Chen et al., 2010), facing more investment opportunities makes screening more onerous, creating incentives for VCs to adopt AI technology to automate screening. Specifically, my empirical strategy uses a quasi-natural experiment: the introduction of Amazon's Web Services (AWS) in 2006, i.e., cloud computing services by Amazon. Ewens et al. (2018) document that this shock lowered the cost of starting new software- and web-related businesses, leading to more startup creations in specific industries and thus an increase in investment opportunities faced by VCs. My empirical strategy exploits variation across VCs in the growth of investment opportunities and thus in incentives to adopt AI.

I define a measure of treatment intensity at the industry-level, which proxies for how likely entrepreneurs in a given industry are to benefit from the availability of cloud computing. To do so, I compute for each industry the fraction of business descriptions of companies, which are founded before the shock, mentioning cloud computing-related keywords. I confirm that the introduction of AWS leads to more startup creations in the industries with high treatment intensity. In high-exposure industries (90th percentile), the number of overall startup creations

[7] Metinko (2022) mentions "*The rate at which startups are founded is nearly impossible to keep up with and investors are turning more to data science to keep up with the volume of companies in a period where it's nearly impossible to perform due diligence on every company in a given space*".

increases by 36% more after the shock than it does in low-exposure industries (10th percentile).

Then, I exploit variation in VCs' sectoral specialization to quantify the causal effect of AI adoption. The incentive for a VC to adopt AI after the introduction of AWS depends on the extent to which its investment prospects are affected. A VC firm investing mainly in high-exposure industries (e.g., e-commerce) would observe a larger expansion of its set of investment opportunities than a VC firm investing mainly in low-exposure industries (e.g., civil engineering). For each VC for which I can observe investments made before 2006, I compute a measure of VC exposure defined as the average industry exposure of the VC's portfolio before the shock.

I study how VCs' investments change after the shock by utilizing a difference-in-differences framework with continuous treatment intensity. More specifically, my empirical specification compares investments made by VCs with different levels of exposure before versus after the shock within a given industry-country-stage segment, while controlling for common changes in the composition of the pool of funded startups in that segment. It also controls for the overall drop in the cost of adopting AI for VCs that may arise due to the availability of cloud computing.[8]

First, I find that highly exposed VCs are significantly more likely to adopt AI after the shock. High-exposure VC firms (90th percentile) are 11 p.p. (120% relative to the unconditional mean) more likely to become AI-empowered than low-exposure VC firms (10th percentile).

Second, consistent with my previous results, I also find that VCs with higher exposure become more likely to invest in backward-similar startups than are other VCs: High-exposure VC firms become twice more likely to fund high backward-similarity startups (top quartile) than to fund low backward-similarity startups (bottom quartile).

Third, I find that VC investment outcomes are also affected in a way consistent with my previous results: after the shock, high-exposure VCs become better at picking good quality startups but only among those that are highly backward-similar. The likelihood that high-exposure VCs select startups that survive and receive follow-on funding increases by 8 p.p. (15% of the unconditional mean) when investing in high backward-similarity startups but does not change significantly when investing in other startups, relative to low-exposure VCs.

Fourth, I find that high-exposure VCs are significantly less likely to invest in startups that achieve major success. After the shock, the likelihood that high-exposure VCs fund startups that eventually go public in an IPO or are acquired for a profit decreases significantly by 8 p.p. (37% of the unconditional mean), relative to low-exposure VCs.

[8] In robustness checks, I show that my results are not driven by VCs with ex-ante expertise in cloud-related sectors being more likely to adopt AI because of the availability of cloud computing services.

Taken together, my results show that AI adoption by VCs affects startup financing and are consistent with AI exploiting past data not informative about breakthrough companies. Overall, AI adoption by investors might thus hinder the allocation of capital to breakthrough innovations. If these innovations create positive externalities that investors fail to internalize, AI-driven investment processes might negatively affect long-term growth and overall welfare.

This paper contributes to several streams of research. First, it contributes to the literature exploring how VCs make investment decisions.[9] Kaplan and Stromberg (2001) and Gompers et al. (2020) study VCs' practices.[10] Kaplan and Strömberg (2004), Kaplan et al. (2009), Gompers et al. (2010) and Bernstein et al. (2017) focus on the factors shaping VCs' investment selection. Casamatta (2003) and Kaplan and Strömberg (2003) study contracts between VCs and entrepreneurs, while Lerner (1995), Hellmann and Puri (2002) and Bernstein et al. (2016) analyze VCs' post-investment value-added.[11] My results directly complement the finding of Ewens et al. (2018) that some VCs have recently adopted a "spray and pray" approach, consisting of investing small amounts in many startups they are likely to abandon. I document a novel trend—the adoption of AI by VCs. I show this to be associated with an increase in the share of VC funding directed towards startups developing businesses similar to those already tested.[12] In that regard, my findings question VCs' economic impact (Piacentino, 2019; Kamepalli et al., 2020; Gornall and Strebulaev, 2021) and VCs' ability to spur innovation (Kortum and Lerner, 2001; Mollica and Zingales, 2007; Lerner and Nanda, 2020; Howell et al., 2020).

Second, this paper adds to the literature on the consequences of the adoption of big data and machine learning in the financial industry. Fuster et al. (2021), Blattner and Nelson (2021) and Di Maggio et al. (2022) study the impact of these technologies on credit allocation.[13] Birru et al. (2019), Coleman et al. (2021), Grennan and Michaely (2020), Dessaint et al. (2021) and Chi et al. (2021) analyze how they affect financial analysts' forecasts. D'Acunto et al. (2019), Reher and Sokolinski (2021), and Abis (2020) study their impact in asset management. Dugast and Foucault (2018), Zhu (2019) and Farboodi and Veldkamp (2020) focus on how big data affects stock price informativeness, while Martin and Nagel (2022) study the implications for market

[9] Kaplan and Schoar (2005), Robinson and Sensoy (2013), Hochberg et al. (2014) and Nanda et al. (2020) study performance in the VC and private equity industry.

[10] Kerr et al. (2014a), Hellmann and Thiele (2015), Hochberg (2016), Lerner et al. (2018), Gonzalez-Uribe and Leatherbee (2018) and González-Uribe and Reyes (2021) study other financial intermediaries such as angel investors and accelerators that shape startups. Colonnelli et al. (2022) study government participation in VC.

[11] Li et al. (2022) show how VCs can redirect innovation at the startups in their portfolio.

[12] This is particularly relevant for the financing of entrepreneurs who experiment with new ideas (Kerr et al., 2014b; Manso, 2016).

[13] Buchak et al. (2018), Fuster et al. (2019), Bartlett et al. (2021) and Di Maggio and Yao (2021) examine how FinTech lenders, i.e., online and automated mortgage application platforms, differ from other lenders.

efficiency.[14] This paper studies the adoption of AI by a key class of financial intermediaries—VCs. To the best of my knowledge, it is the first to show how this technology affects the provision of capital to startups and, more broadly, to innovative projects that differ from historical data.

Third, this paper adds to the literature investigating the impact of big data and AI on innovation and growth. Brynjolfsson et al. (2018) and Furman and Seamans (2019) study the effects of AI on productivity growth. Autor et al. (2003), Acemoglu and Restrepo (2018), Aghion et al. (2019), Agrawal et al. (2019a) and Rock (2019) analyze how new technologies, such as AI, affect the division of income between labor and capital. Cockburn et al. (2019) and Agrawal et al. (2019b) explore how AI can enhance the invention and innovation processes.[15] Abis and Veldkamp (2020) investigate how big data affect the production function of firms in the financial sector. Babina et al. (2021) study how using AI technologies affects firms' growth. This paper shows that the adoption of AI by investors, such as VCs, can affect the allocation of capital among young, innovative companies, and thus potential long-term growth.[16]

Finally, this paper contributes to the literature studying the relative strengths and weaknesses of "men" versus "machines" in financial-economic decision-making. Kleinberg et al. (2018), Li et al. (2020), Erel et al. (2021), Aubry et al. (2022), Jansen et al. (2021) and Cao et al. (2021) explore the performance of a machine learning algorithm versus a human-based approach in making bail decisions, hiring, selecting directors, pricing art works, underwriting loans, and forecasting stock prices, respectively. van Binsbergen et al. (2022) use machine learning to highlight biases in financial analysts' expectations. Zacharakis and Meyer (2000), Åstebro (2002), Åstebro and Elhedhli (2006), Hunter et al. (2017), Retterath (2020), Blohm et al. (2020), Lyonnet and Stern (2022) and Davenport (2022) study how various algorithms could be used to outperform or support human investments in startups.[17] This paper provides evidence that AI adoption by VCs improves their ability to select successful startups, but only among those startups for which historical data are informative.

The paper proceeds as follows. Section 2 discusses the institutional background. Section 3 presents the data and measures. Section 4 discusses how VCs' investments change after adopting AI. Section 5 focuses on the causal effects of AI adoption. Section 6 concludes.

[14] Gao and Huang (2020) study how modern information technologies have changed how information is disseminated in financial markets. Relatedly, Bai et al. (2016) and Farboodi et al. (2022) ask whether improvements in financial technology have changed the informativeness of stock prices. Mihet (2020) studies the link between financial innovation and financial inclusion.

[15] Kogan et al. (2017) show that technological innovation accounts for significant fluctuations in aggregate economic growth.

[16] This result connects to Decker et al. (2016) documenting a decline in high-growth young firms, despite rising research effort (Bloom et al., 2020b).

[17] Guzman and Stern (2020) use predictive analytics to estimate entrepreneurial quantity and quality.

2 Institutional background: AI-empowered VCs

Central to this paper is the distinction between AI-empowered VCs and other VCs. While one can argue that many VCs are adopting technologies to varying degrees, it is clear that some are at the forefront of using quantitative techniques to automate and enrich their investment process. For these VCs, big data, machine learning and AI algorithms are at the heart of their operations and define their investment approach.[18]

VC activities can be decomposed into three main types of tasks: (i) screening, i.e., sourcing, evaluating and selecting investments, (ii) structuring investments, i.e., contracting with entrepreneurs, and (iii) post-investment value added, i.e., monitoring and advising entrepreneurs (Kaplan and Stromberg, 2001; Gompers et al., 2020). Sørensen (2007) and Gompers et al. (2020) find screening to be the most important for value creation. This constitutes a lengthy and challenging process for VCs that evaluate each year hundreds of startups, each one being a potential investment opportunity (Sahlman, 2022). In addition, VCs might be subject to biases in their investment decisions (e.g., Ewens and Townsend, 2020). In response, a number of VC firms have developed software combining big data and machine learning to automate and enrich their screening process. For example, InReach Ventures, SignalFire and EQT Ventures have each developed their own proprietary platform that automatically tracks and scores start-ups in terms of future return prospects.[19]

Traditionally, VCs source their potential investments somehow from their network (Hochberg et al., 2007; Gompers et al., 2020; Howell and Nanda, 2019), and then select a startup according to the VC partners' assessment of its business and founding team (Kaplan and Strömberg, 2004; Kaplan et al., 2009; Gompers et al., 2010). AI-empowered VCs substantially differ in their screening process. Their first goal is to identify each company as early as possible in its life cycle. They do so through the use of algorithms tracking several sources. As startups are private, they have no obligation to report and share information. However, several sources disclose information on startups. First, many companies store and commercialize large databases on startups and VC investments (e.g., PitchBook and Crunchbase in 2007, CBInsights in 2008, AngelList in 2010). In addition, several alternative data sources have emerged from the internet (e.g., Linkedin, ProductHunt, GitHub). Once

[18] The discussion of institutional details in this section draws on extensive conversations with VC professionals and industry experts, as well as on a number of resources (e.g., articles, blog posts, podcasts, videos). For evidence on how technology is reshaping the VC industry, see for instance Bhakdi (2013), Wu and Gnanasambandam (2017), Gartner (2020), Chen (2021), Åstebro (2021) and Röhm et al. (2022).

[19] See for instance ft.com[...]ed31693349d3 and wsj.com[...]/vc-firms-have-long-backed-ai-now-they-are-using-it.

identified, AI-empowered VCs seek to collect and update continuously many "growth metrics" on each startup. These metrics can include the number of employees, rating and product reviews, payment data, website traffic, news mentions, and number of followers. Finally, VCs use machine learning techniques (e.g., classification algorithms) to identify quantitative success patterns. This last step is done by exploiting historical data of startups that have already succeeded or failed and letting algorithms "learn" the links between the growth metrics and success outcomes. The final output is usually a scoring of startups that is used to filter out the set of potential investments. This approach does not eliminate the role of humans as the final decision to invest is still typically taken by a VC partner.

3 Data and measures

This section describes the data and measures used in my empirical analysis. Subsection 3.1 describes the data sources. Subsection 3.2 details how I measure investment success and innovation. Subsection 3.3 introduces my methodology to identify AI-empowered VCs. Subsection 3.4 describes how I construct a measure of startup "backward-similarity".

3.1 Data sources

VC investments My main dataset is constructed from Crunchbase, an online database providing detailed information on startups and their investors.[20] The database records each company's business description, industry, location, founding date, employees' job and profile, as well as investors, funding rounds, acquisitions and IPOs, when applicable. Crunchbase does not only cover VC-funded startups and encompasses companies that did not raise outside capital. One key advantage of using Crunchbase for my analysis is that it provides extensive information on jobs and employees at VC firms, which is critical for my identification of AI-empowered VCs. Appendix A provides additional details on the database and its accuracy.

My download of the datasets includes information through 2021. Although the data go as far back as 1975, I restrict my analysis to VC investments made after 2000. I choose this starting date because Crunchbase's coverage of startups has been validated to be most accurate in the more recent years (Wu, 2016; Ferrati and Muffatto, 2020). When studying investment outcomes, I restrict my analysis to investments completed by 2019 so that I can observe the outcome of funded startups at least two years after the VC invested. Specifically, I include all investments

[20] I obtained an export of Crunchbase's entire database through Crunchbase academic research access, which allows researchers with projects accredited by the company to access the data.

made over the period 2000-2019 and I observe outcomes through 2021.[21] When I study VCs' number of investments per year, I also include investments made in the year 2020.

I use global data on VC investments worldwide.[22] I restrict my focus to investments in private companies made by investors classified as VCs by Crunchbase.[23] Furthermore, I only include investments corresponding to equity funding rounds.[24] Finally, I restrict my analysis to investments in startups for which information on headquarters country, industry and founding date is available.

For each VC-funded company in my dataset, I collect information on its subsequent acquisition or IPO, when available. For each VC firm in my sample, I also gather additional information including founding year, headquarters location, and full employee and job histories provided by Crunchbase. This allows me to compute a VC's age and number of employees at the time it invests in any startup. I exclude investments made by VC firms that do not report having any employees over the period 2000-2020. To determine all the locations of VC firms, I use both the headquarters location and the employees' locations. Thanks to the unique person identifier provided by Crunchbase, I can also determine whether one of the VC employees is a board member or observer at any startups in which the VC invests.

My final sample at the investment level amounts to 221,765 investments made by 7,165 distinct VC firms over the period 2000-2019. It includes funding rounds of 69,863 distinct startups that are at some point VC-funded, i.e., startups can raise several funding rounds and be funded by several VCs.

Panel A in Table 1 presents descriptive statistics at the VC investment level. It shows that as of the investment date on average the VC firm is 14 years old and has 10 employees. A total of 48% of the observations correspond to early-stage investments, i.e., investments in pre-seed, seed, or series A funding rounds. Table A.1 in Appendix A shows the most represented industry groups in terms of number of funding rounds. The top three includes Software, Health Care and Financial Services.[25] Table A.2 in Appendix A shows the most represented countries. The

[21] After any given funding round in my sample, it takes on average one year for startups to receive follow-on funding when they manage to do so.

[22] Crunchbase includes information on VC investments around the world. Da Rin et al. (2013) document that by 2011 non-US investments accounted for approximately half of all VC investments, and the quality of these data has improved considerably over the last decade.

[23] Specifically, I do not consider investments made by angel groups, family offices, funds of funds, investment banks, hedge funds, accelerators and incubators, government offices, university and entrepreneurship programs, coworking spaces, startup competitions, pension funds and loyalty programs.

[24] Specifically, I drop investment observations corresponding to grant, non-equity assistance, post IPO, corporate round, secondary market, crowdfunding, initial coin offering, convertible notes, and debt financing.

[25] Crunchbase classifies companies into 744 industries and 47 industry groups. Industries are specific market segments while industry groups are broader and encompass multiple industries. Table A.3 in Appendix A shows

top three are the U.S., China and the UK. Panel B of Table 1 presents statistics at the VC-year level. The average VC makes about four investments per year, including one as lead investor.

Startups' founding team information When available, I also collect information on the founding team of the VC-funded startups in my sample. Crunchbase provides detailed information on the history of the employees of the companies in the database. Appendix A lists the information sources used by Crunchbase and describes how I identify startup founders.[26] For each founder, I check whether they served as a founding team member of another company in the past and if so I classify them as a serial founder. To identify female founders I use the person's gender provided by Crunchbase. I also collect education records provided by Crunchbase and identify founding team members holding degrees from top schools (as defined in Appendix A).

Patent data To measure innovation by VC-funded companies, I augment my dataset with information on patent applications and patent grants from IPqwery. Patent data are sourced by IPqwery from the USPTO (United States Patent and Trademark Office), CIPO (Canadian Intellectual Property Office), EUIPO (European Union Intellectual Property Office) and WIPO (World Intellectual Property Office). The dataset includes both granted patents that are active and granted patents that have expired. Importantly, it also includes patent applications that are still pending, as well as those that have been abandoned, rejected or canceled. It provides each patent's unique identifier as well as information on its assignee, its technology class, its application year and, when applicable, its grant year. One key advantage of IPqwery is that it exploits a proprietary algorithm to identify patent holders and it includes the unique Crunchbase's company identifier in the database. This allows me to match patents with the VC-funded companies in Crunchbase.

I also obtain patent citations from Google Patent, which contains information on patents from the main patent authorities around the world. Using each patent's unique identifier, I retrieve from Google Patent the number of citations it receives from subsequent patents up to September 2022. Citation measures are adjusted for right-censoring (see Jaffe and De Rassenfosse, 2019; Lerner and Seru, 2022). Specifically, I scale each patent's citation count by the average number of citations of patents applied for in the same year and in the same technology class. The group of patents considered for this adjustment comprises all those granted to companies receiving VC

the most represented industries.

[26]Startups can have several founders, i.e., founding team.

funding in the Crunchbase database. If a patent belongs to multiple classifications, each one receives equal weights adding up to one.

VC fund-level data Finally, my dataset is supplemented by information on VC funds coming from Preqin. Preqin was founded in 2003 and provides fund-level performance and benchmarking data within the private capital industry for more than 42,000 investment vehicles.[27] I focus on funds belonging to the "Venture Capital" asset class, with vintage inception year between 2000 and 2020.[28] For a fund to be included in my sample, I require the following information to be available in Preqin: the stage and core industry of the investment, the actual size (in USD millions), the number of deals carried out by the fund, and the fund's sequence number (i.e., if the fund is the first, second, and so forth, of the VC). For funds with such information, I add fund performance measures when available.[29] I then match VC firms in Crunchbase and Preqin using their names and contact information (website url and email). Panel C of Table 1 reports summary statistics for the fund sample.

3.2 Definition of investment success and innovation measures

I use VC investment outcomes to determine whether a given startup achieves major success. In addition, I use patents to measure startup innovation.[30]

Investment success measures I construct several measures of investment success. First, I define a dummy variable *Follow-on* equal to one if the VC-funded startup is still active and has received a follow-on investment by 2021 (i.e., reinvestment by VCs) and equal to zero otherwise. Whether a VC-funded startup goes on to raise a follow-on round of funding is a key early indicator of a company's future success, as the majority of startups die out and do not raise follow-on funding (Kerr et al., 2014b).

[27] Harris et al. (2014) show that Preqin provides the most comprehensive coverage for the 2000s.

[28] Among VC funds, I drop funds whose stage of investment is one of the following "Add-on", "Grant", "Venture Debt", "PIPE", "Secondary Stock Purchase" and "Merger".

[29] The performance measures I consider are the realized net IRR and the net multiple. IRR is the annualized return to Limited Partners (LPs) net of performance fees and management fees. Net multiple is the ratio between the total value that the LPs have derived from their interest in the fund and the total cash investment in the fund, expressed as a multiple. While they have well-known limitations, IRR and multiple are the most widely available fund performance measures and are commonly used to analyze PE performance.

[30] Moser (2012) shows that many significant breakthroughs are not patented. Regarding the link between startup patenting and success, Farre-Mensa et al. (2020) show that obtaining a patent boosts a startup's subsequent growth by facilitating access to funding from VCs. Howell et al. (2020) find that patents filed by VC-backed firms are of significantly higher quality and economic importance than the average patent. Bernstein (2015) shows how going public in an IPO affects a firm's innovation.

I also define more conservative measures capturing whether a VC investment can be considered as a major success. A popular indicator is whether the VC-funded startup later goes public in an IPO. In addition, many successful companies are acquired by larger companies for a profit. Following Calder-Wang and Gompers (2021), I define a *Breakthrough Success* dummy variable equal to one if by 2021 the startup in which the VC invested goes public in an IPO or is acquired for a higher value than the total VC investments in the company. I obtain acquisition values from Crunchbase. If I am unable to identify an acquisition value, I do not consider the investment a breakthrough success. I run robustness tests using (i) a more conservative measure indicating only whether the startup goes public in an IPO and (ii) a broader measure indicating either an IPO or any acquisition.

Panel A of Table 1 shows that 52% of investments are made in startups that are still active and have received a follow-on investment by 2021. A total of 15% of the investments match with my definition of breakthrough success (8% of the startups in my sample).

Innovation measures I construct several innovation measures. First, I measure innovation quantity using the number of patent applications filed by a startup after the VC invested. I use the logarithm of one plus this variable, that is $Log(1 + Nb.\ New\ Patent\ Applications)$. I use patent applications because there is usually a gap between the dates at which a patent is applied for and granted. Because my download of the patent datasets includes patent records through January 2022, patents filed in the later years may still be in the approval process. To mitigate this right-truncation issue, I study patent applications up to January 2022 for companies receiving VC funding by 2019. Thus, I can observe patents filed in the two years after the most recent VC investments in my sample. I verify that my results are robust to using the number of future patent applications that are ultimately successful, that is $Log(1 + Nb.\ New\ Patent\ Grants)$.[31]

Second, I measure innovation quality using the number of highly-cited patents filed by a startup after the VC invested. I define a patent as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and the same technology class. I use the logarithm of one plus this variable, that is $Log(1 + Nb.\ New\ Highly\text{-}cited\ Patents)$. I also use the number of future patent citations (adjusted for right-censoring as described in subsection 3.1), that is $Log(1 + Nb.\ New\ \ Patents'\ Citations)$.

Panel A of Table 1 shows that on average a VC-funded startup files 8 patent applications

[31] I use a patent's year of application instead of the year it is granted because the former better captures the actual timing of innovation.

and is granted 5 patents after the VC investment. On average a startup obtains 2 patents that are highly-cited and the average number of received citations is 6.5. I observe that 14,039 out of the 69,863 startups (20%) in my sample apply for at least one patent after the VC investment.

3.3 Identifying AI-empowered VCs

Methodology I develop a methodology to identify the VCs using AI in their investment screening, which I call AI-empowered VCs. I do so by exploiting detailed data on VC jobs and employee descriptions available in Crunchbase. The rationale is that VCs using AI technologies rely on human capital and expertise. Hence, the employment of professionals with AI-related skills can be viewed as a proxy for using AI-driven investment screening.[32]

Crunchbase data on employees at VC firms not only include top executives and partners but also encompass other staff members such as analysts, engineers and back office professionals (cf., Appendix A). For each VC firm in my sample, I collect the whole history of jobs at the firm in Crunchbase (including past jobs that are no longer active). There are a total of 72,269 distinct jobs associated with the VC firms in my sample. For each job, Crunchbase reports the starting date, end date (or whether the job is current), the job type (employee, executive, advisor, board member or observer), the job title, and the unique identifier of the person. Using the latter, I also gather the corresponding person's profile description available on Crunchbase.

My approach to identify AI-empowered VCs consists of several steps. First, by analyzing several online resources (press articles, blog posts, podcasts, and videos), I compile an initial list of VCs that explicitly mention their use of big data and machine learning. This first list includes approximately a hundred VC firms. Then, I identify these VCs in Crunchbase using name matching. For each one, I obtain the full history of jobs and the corresponding employees' profile descriptions. I manually identify the jobs that mention core AI skills (e.g., "machine learning"). VC firms might also hire data scientists to advise their portfolio companies, i.e., the startups they have funded. I rely on people descriptions to make sure that the final list only contains jobs that unambiguously relate to the use of AI technologies in the VC firm's screening operations.[33] This procedure allows me to identify an initial list of 128 different AI-related jobs.

The second step of my identification procedure is to identify the VC firms' AI-related jobs

[32]Several papers have used job posting data to infer technology adoption and in particular AI at the firm level in other sectors. See for instance Alekseeva et al. (2021), Goldfarb et al. (2021), Abis and Veldkamp (2020), Babina et al. (2021) and Bloom et al. (2020a). The advantage of using actual jobs available in Crunchbase is that it allows to observe the stock of and not only the demand for AI-related employees.

[33]When additional information is needed to classify a job, I also directly check the Linkedin profile of the person.

that are not part of my initial list. For that purpose, I first "clean" all job titles following standard text cleaning procedures.[34] Then, I extract every single word from my initial list of AI-related job titles and I search for those in the full set of job titles at all VC firms available in Crunchbase. This gives me a list of VC firms jobs with at least one AI-related keyword match.

In three steps, I ensure that I do not consider jobs that relate to advising portfolio companies specialized in AI. First, among jobs featuring at least one keyword match, I only keep those whose job type is either "employee" or "executive" and drop those that are classified as "advisor", "board member", or "board observer". Second, I also remove jobs associated with people whose profile description mentions that the person advises startups.[35] Third, I perform a manual verification to keep only AI-related jobs that are unambiguously related to the use of the technology in the VC firm's process. In total, I am able to identify 247 different job titles that are AI-related. Figure G.1 in Appendix G displays the 30 most frequent job titles that are AI-related.

Finally, using job starting dates, I classify a VC as AI-empowered as soon as it starts employing one person in an AI-related job.[36] Otherwise, VCs are considered "traditional". This gives me a classification of each VC at each point in time, including some VCs switching from traditional to AI-empowered. In robustness checks, I also consider a continuous measure of AI adoption defined as the logarithm of one plus the current number of AI-related jobs at the VC.

The main advantage of this method, combining both a keywords-based approach and additional checks, is that it allows me to avoid false positives. Indeed, the presence of a single keyword such as "AI" in a job title would not be sufficient to determine whether a VC firm actually uses AI technologies internally.[37] Figure G.2 in Appendix G provides examples of profiles and jobs I classify as AI-related on Crunchbase. In Table H.1 in Appendix H, I list the top-10 AI-empowered VCs, based on their number of AI-related jobs.

Stylized facts Figure 1 illustrates how the presence of AI-empowered VCs has evolved over time. Panel A plots their number and fraction each year based on my classification. It shows that the fraction of AI-empowered VCs rises from virtually zero in 2000 to more than 4% in 2020. This upward trend has to be put in perspective with the growing number of investments

[34] I only keep the English words by removing numbers, symbols and special characters. I also remove all the stop words and I stem each word to its root using the Porter stemmer algorithm.

[35] Specifically, I search for the terms "advis startup", "advis compani", "advis portfolio" and "advis ventur" in people profile descriptions that have been previously "cleaned" following standard text cleaning procedures.

[36] In case a job starting date is missing, I replace it with the date at which the job was added to the Crunchbase database. This information is never missing and is always posterior to the job starting date.

[37] One example is the job title "Partner AI Fund", held by a VC partner that advises venture-backed startups in the AI sector but does not use AI technologies for investment screening. Using only a keyword search without additional verification would lead in that case to a categorization error.

they make compared to other VCs. Indeed, Panel B of Figure 1 shows that AI-empowered VCs invest in 20% of all funding rounds of startups raising capital in 2020 versus about 5% in years before 2010. Figure G.3 in Appendix G also confirms that the presence of AI-empowered VCs has been growing in terms of number of funds (9% as of 2020) and AUM (12% as of 2020).

3.4 Backward-similarity measure

AI-empowered VCs use algorithms to forecast the outcome of a given startup by extrapolating from historical data about prior startups. AI algorithms process historical data on startups that have already succeeded or failed and attempt to find any links between observable growth metrics and startups' outcomes. This process requires a large enough quantity of data to uncover empirical patterns. Intuitively, if a startup develops an idea that is similar to those of many prior ventures, it is more likely that a significant amount of data on this particular business model can be processed. AI algorithms are thus more likely to generate statistically significant predictions of the startup's future success. In contrast, if a startup is developing a radically new business idea, presumably no relevant past data can be processed by AI to predict whether that specific startup will be successful.[38]

To proxy for the amount of past data that can be processed by AI to screen a startup, I compute a measure of "backward-similarity". To create this indicator, I use a textual analysis of high-dimensional data from startups' business descriptions available in Crunchbase. I leverage natural language processing techniques to measure how backward-similar a startup is when it receives funding. Put simply, I compute the textual similarity of a startup's business description with respect to descriptions of the set of previously VC-funded startups in the same industry. Startups ranked high on backward-similarity presumably develop businesses that are similar to those which, in the past, have already been tested. AI algorithms can thus "learn" from past data to screen this type of companies.

My measure of backward-similarity builds on the methodology found in Kelly et al. (2021) that is used to measure the textual similarity of a given patent to previous ones. I summarize the different steps of my procedure below and provide more details in Appendix B. First, I convert each startup's business description into a vector of word count. Each word in a given vector is weighted by its importance using a "term frequency inverse document frequency" (TFIDF) transformation, which controls for the extent to which a term becomes more or less widely used

[38] Furthermore, because most VC-funded companies fail while some become runaway successes, returns on VC investments are highly skewed (Cochrane, 2005; Korteweg and Sorensen, 2010). Whether AI algorithms are effective at forecasting "tail" events is unclear.

over time in a given industry. Then, for a given startup i raising capital in year t, I compute the total pairwise (cosine) similarity between its description vector and those of companies that raised capital in the 5 calendar years preceding year t, in the same industry. This gives me a measure of the backward-similarity of startup i at time t. The backward-similarity of a given startup can change over the sequence of funding rounds if it raises capital several times.

I define the *Similarity Rank* of a startup as its overall percentile rank in terms of backward-similarity when it receives VC funding. I use the *Similarity Rank* in my empirical analysis because it lies in the interval [0,1] and is therefore easier to interpret. I also compute the backward-similarity quartile of each startup when it raises funding. This metric indicates whether the similarity rank of the startup is in the range [0; 0.25), [0.25; 0.5), [0.5; 0.75) or [0.75; 1]. All specifications in sections 4 and 5 include fixed effects to control for factors that could mechanically make startups appear more or less similar over time, industries or countries (e.g., the fact that terms become less novel or that the number of startups is expanding).

In Appendix B, I show that well-known "disruptive" startups and startups that are copies of existing businesses score respectively low and high on backward-similarity. I also discuss the evolution of the distribution of my backward-similarity measure over time.

4 VC investments and outcomes after AI adoption

In this section, without claiming to establish causal links, I provide novel evidence on how VC investments change after adopting AI. First I study the number and type of investments made by AI-empowered VCs. Second, I focus on VC investment outcomes. In Appendix C, I test whether specific VC firms' characteristics are associated with AI adoption. I show that VC experience and previous success do not correlate with AI adoption.

4.1 VCs scale-up towards backward-similar startups after adopting AI

Adopting AI might allow VCs to "scale-up", i.e., to evaluate more startups because of their use of algorithms. This can be useful if the VC firm seeks to manage more assets and fund a larger number of companies.[39] However, using algorithms might also be associated with increased focus on startups for which historical data are informative. I test whether AI adoption is associated with a scale-up of VC investments, and if so in what kind of startups. I show that

[39] Roberto Bonanzinga, the co-founder of InReach Ventures—a VC firm disclosing publicly its use of AI—mentioned in the press: "What used to be a handcrafted job has become significantly scalable. You become 10 times more productive". Cf., ft.com[...]ed31693349d3.

after adopting AI, VCs increases their number of investments per year and tilt their portfolios towards backward-similar startups.

Empirical specifications I estimate the following specification using VC-year observations:

$$Y_{j,t} = \beta AI_{j,t} + X_{j,t} + \alpha_j + \gamma_{c\times s\times t} + \epsilon_{j,t}, \tag{1}$$

where $Y_{j,t}$ measures the number of investments made or the AUM by VC j in year t. The main independent variable is $AI_{j,t}$, which is a dummy equal to one if VC j is classified as AI-empowered in year t, and zero otherwise. $X_{j,t}$ are time-varying control variables including the logarithm of the VC firm's age and the logarithm of its number of employees, both measured as of year t. α_j are VC firm fixed effects, and $\gamma_{c\times s\times t}$ are country-stage-year fixed effects. Country denotes the country of the VC's headquarters, and stage denotes the stage of funding in which the VC firm makes the largest number of investments over the sample period (among six categories, i.e., pre-seed, seed, series A, series B, series C, or series D and onward). Including country-stage-year fixed effects forces the coefficient β in (1) to be estimated by comparing AI-empowered VCs before versus after adopting AI relative to other traditional VCs headquartered in the same country and focused on the same investment stage. All specifications discussed in this subsection double-cluster standard errors at the VC and year levels.

The estimated β coefficient in (1) may not capture the causal effect of AI adoption. Indeed, there might be unobserved variables driving the decision to adopt the technology. For instance, a VC firm might decide to adopt AI because it *expects* to do more investments in the future and thus seeks to automate investment screening. Nevertheless, observing whether VC firms "scale-up" their investments after adopting AI is informative about how they use this technology.

To test whether AI-empowered and traditional VCs have parallel pre-trends, I also estimate the link with AI adoption using the following event study specification:

$$Y_{j,t} = \sum_{l=-4,\ l\neq -1}^{5+} \beta_l \left\{AI\ Year(l)\right\}_{j,t} + X_{j,t} + \alpha_j + \gamma_{c\times s\times t} + \epsilon_{j,t}, \tag{2}$$

where $\left\{AI\ Year(l)\right\}_{j,t}$ is a dummy variable equal to one if VC j adopts AI over the sample period and if year t corresponds to l years before/after the adoption. The omitted category is the year before the adoption.

Increase in number of investments and in AUM Table 2 reports the estimated coefficients of regression (1) using three dependent variables that are the logarithm of (i) one plus the number of investments made by the VC in the year (for years with no investment activity the count is zero), (ii) the cumulative number of funds managed by the VC firm as of the year, and (iii) the total AUM of the funds managed by the VC firm. Note that (ii) and (iii) are only defined for VC firms in Crunchbase that can be matched with funds in Preqin.

I find that the number of investments of AI-empowered VCs compared to those of traditional VCs increases significantly after adopting AI. The coefficient on AI adoption in column (1) is 0.39, translating to a 48% increase in number of yearly investments. When I control for the VC's age and number of employees, the coefficient becomes smaller in magnitude (0.25, i.e., 28% increase) but remains highly significant. Given that VCs make on average 4.2 investments per year (cf., Panel B of Table 1), the coefficient in column (2) suggests that AI adoption is associated with 1.2 additional investments each year. Columns (3) to (6) show that AI adoption is associated with more funds and AUM. The magnitudes of coefficients in columns (4) and (6) correspond to an increase by approximately 30%.

Figure 2 shows the estimated coefficients β_l in equation (2), with 95% confidence intervals. Panel A plots the coefficients for the number of investments, Panel B for the number of funds, and Panel C for AUM. All panels show no pre-trend. The increase in the number of investments appears in the year after AI adoption and persists even 5 years after. The rises in the number of funds and AUM take a similar time to materialize. These results are consistent with AI adoption being associated with a scaling-up of VC investments.

Increase in number of investments in backward-similar startups I test whether AI-empowered VCs expand their investments relatively more towards a specific type of startups. My conjecture is that AI-empowered VCs are more likely to benefit from AI when screening backward-similar startups. They might thus tilt their investments towards this type of company. To test this, for each VC firm and year, I count the number of investments made in startups belonging to each backward-similarity quartile, and I estimate regression (1) with the logarithm of (one plus) the number of investments in each quartile as the dependent variable.

Table 3 presents the estimation results. Columns (1) and (2) show that AI adoption is not associated with any change in the number of investments in the lowest quartile of similarity. However, when moving from column (1) through (8), one observes a monotonous increase in the coefficient on AI adoption. This suggests that, relative to other VCs, AI-empowered VCs increase their investments in the most backward-similar startups significantly more. In terms of magnitude, the estimated coefficients correspond to a 94% increase of investments in the fourth quartile. Figure 3 presents graphically the estimated coefficients in columns (2), (4), (6), and (8) of Table 3 and illustrates the magnitude of the difference across the different similarity quartiles.

Panels A and B of Figure 4 show the estimated coefficients β_l in equation (2) for the number of investments in the first and fourth similarity quartile respectively. While no effect is detected in Panel A, Panel B shows that the number of investments in startups with backward-similarity in the top quartile rises in a way that is consistent with the increase in the total number of investments in Panel A of Figure 2.

Additional tests I discuss additional tests in Appendix D. First, I run an analysis of the number of investments made by VCs at the fund level. The results are very consistent with those estimated using VC-year observations. Second, I estimate the same specifications as in Table 3 but I add as control variables the logarithm of the VC's number of funds and the logarithm of its AUM as of the investment year. The coefficients remain similar. I also test whether AI adoption is associated with changes in the industry composition of VCs' portfolio. I do not find any significant variation in sectoral specialization of VC firms after adopting AI. Finally, I test whether AI adoption is associated with changes in VCs' stage of investments. My results do not indicate that VCs focus less on early-stage investments after adopting AI.

Robustness I also discuss robustness checks in Appendix D. First, I show that my results are robust to using a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm. Second, I show that my findings remain similar when using the estimator of Borusyak et al. (2021), which is robust to heterogeneous effects between treated VCs or over time. Finally, I show that my conclusions are unaffected when employing a Poisson model, i.e., replacing the right-hand side of equation (1) by its exponent.

4.2 Backward-similar startups are unconditionally less successful

The results presented above show that after adopting AI, VCs tilt their portfolio towards high backward-similarity startups. In this subsection, I investigate whether these startups are better investments. I show that, in the sample of VC-funded companies, high backward-similarity startups are on average less likely succeed than low backward-similarity startups.

Likelihood of success Figure 5 shows bin-scatter plots that relate the backward-similarity of a startup to its future success. For each VC-funded startup, I plot a dummy indicating whether the startup has closed down by 2021 as a function of the startup's backward-similarity percentile rank when it raises its first round of funding (Panel A). I do the same with dummies indicating whether the startup eventually achieves breakthrough success (Panel B), survives and raises follow-on funding (Panel C), and files a patent application after the funding round (Panel D). All variables are first regressed on industry-country-year fixed effects to absorb factors that could make startups appear more or less similar.

Panel A shows that the higher the backward-similarity of the startup is, the higher its likelihood of being closed by 2021. Panels B, C and D show a negative relationship between the backward-similarity of the startup and its future success or innovation.[40] In particular, startups with backward-similarity above 0.6 are significantly less likely to achieve breakthrough success, receive follow-on funding or file a patent application. These results highlights that, on average, "pioneers" are more likely to be successful than "followers".[41]

Regressions I estimate regressions using startups' backward-similarity percentile rank as the dependent variable and measures of innovation, success and human capital as independent variables. The results are reported in Table 4. To avoid double counting, when a startup raises capital through several funding rounds, I only consider the observation corresponding to the first round of funding. Regressions include industry-country-year fixed effects. Standard errors are double clustered at the industry and country levels.

In column (1) of Table 4, the independent variables are dummies indicating whether the company has filed a patent application as of the date of the funding round or whether it will file an application afterward. Column (1) reveals a strong negative correlation between backward-similarity and innovation. The magnitude of the coefficient on future patent application is twice as large as that of the coefficient on past patent application. This result suggests that my measure of backward-similarity incorporates information regarding the ability of companies to innovate in the future. This is confirmed by the results in column (2), where the independent variables are the logarithm of (one plus) the number of past and future patent applications.

In column (3) and (4), the independent variables are a dummy indicating whether the company will achieve breakthrough success and a dummy indicating whether the startup will

[40] Guzman and Li (2022) estimate the text-based distance between a startup and its five closest incumbents (public firms), and find this measure to predict an increase in early-stage financing.

[41] This is in line with research in entrepreneurial strategy that posits an important role for early positioning (e.g., Gans et al., 2018).

survive and receive follow-on funding respectively. Column (5) includes both. The estimated coefficients confirm the graphical evidence discussed above: they are all negative and statistically significant, confirming that a higher backward-similarity is associated with a lower likelihood of future success. Column (6) shows that there is no significant relationship between a startup's backward-similarity and its number of employees reported by Crunchbase as of 2021. No significant correlation emerges with the amount raised by the company in column (7).

In column (8), I consider the subset of startups for which I can gather information on the founding team to investigate the relation between backward-similarity and human capital. Startups ranked high on backward-similarity are significantly less likely to have founders holding top school degrees. They are also less likely to be run by female founders. The relationship with dummy variables indicating whether one founder holds a PhD or a MBA are negative.

In columns (9) and (10), I test whether backward-similar startups are more likely to be headquartered out of the main VC hubs. Column (9) shows that high backward-similarity startups are less likely to be located in the U.S. and Europe. Column (10) confirms that they are also less likely to be located in one of the five main VC hubs in my sample which are California, New York, Massachusetts, England and Beijing.

In Column (11), I do not observe a systematic relationship with the company age at the time it raises funding. However, when I also include observations corresponding to subsequent funding rounds of startups that raise follow-on funding and I add startup fixed effects, column (12) shows that startups become more backward-similar as they become older. Using the same sample and a similar specification, column (13) shows that companies raising capital at a late stage (i.e., more advanced funding rounds) are more likely to score high in terms of backward-similarity.[42]

Distribution of outcomes conditional on success The results discussed above show that the backward-similarity of a startup is on average negatively related to its future success and innovation. However, conditional on success, backward-similar startups might be successful to a larger extent. To test this possibility, I investigate the distribution of the logarithm of IPO valuations among startups that achieve a successful exit. Panel A of Figure 6 shows the distribution separately for startups in the top and bottom quartile of backward-similarity when raising their first round of funding. Clearly, startups that are more backward-similar tend to obtain lower valuations when they go public than less backward-similar companies: The distribution is shifted to the left for the fourth quartile of backward-similarity.

[42]Table H.2 in Appendix H shows that my results are the same when I consider only the subset of observations corresponding to funding rounds without any AI-empowered VC.

In Panel B of Figure 6, I show the distribution of the logarithm of the number of new patents' citations among startups that manage to obtain a patent after the initial VC investment. Again, the distributions are presented separately for startups in the top and bottom quartile of backward-similarity. One observes that the patents obtained by startups that are more backward-similar tend to be less cited than those obtained by less backward-similar companies.

4.3 VCs' investment outcomes change after adopting AI

My previous results show that VCs adopting AI scale-up their investments towards backward-similar startups. In this subsection, I test whether AI adoption is associated with changes in VC investment outcomes and if so, when investing in what type of startups. I show that after adopting AI, VCs become better at selecting startups that survive and later receive follow-on funding, but only within the pool of those that are highly backward-similar. This is consistent with "cream-skimming": AI-empowered VCs are able to select the good quality startups from the pool of those that are backward-similar although these companies are unconditionally less likely to succeed. However, VCs' investments become less likely to result in major success or a breakthrough.

Empirical specifications I estimate the following specification at the investment-level:

$$Y_{j,k,t} = \beta AI_{j,t} + X_{j,k,t} + \alpha_j + \gamma_{i\times c\times t\times s} + \epsilon_{j,k,t}, \tag{3}$$

where $Y_{j,k,t}$ is an outcome variable measured for a given investment made by VC j in startup k in year t. The main independent variable is $AI_{j,t}$, which is a dummy variable equal to one if VC j is classified as AI-empowered as of the investment date and equals zero otherwise. $X_{j,k,t}$ are time-varying control variables that include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. α_j are VC firm fixed effects. $\gamma_{i\times c\times t\times s}$ are startup's industry $\times$ country $\times$ funding year $\times$ funding stage fixed effects. These stringent fixed effects force the β coefficient in (3) to be estimated by comparing VC firm j's investments before versus after AI adoption relative to other VC firms' investments in the same industry-country-year-stage segment. In all specifications presented in this subsection, standard errors are double-clustered at the VC firm and startup company levels.

To test for parallel trends, I estimate the link between AI adoption and investment outcomes using the following event study specification:

$$Y_{j,k,t} = \sum_{l=-4,\ l\neq -1}^{5+} \beta_l \left\{AI\ Year(l)\right\}_{j,t} + X_{j,k,t} + \alpha_j + \gamma_{i\times c\times t\times s} + \epsilon_{j,k,t}, \tag{4}$$

where $\{AI\ Year(l)\}_{j,t}$ is a dummy variable equal to one if VC j adopts AI over the sample period and if year t corresponds to l years before/after the technology adoption. The omitted category is the year before the adoption.

Because AI algorithms may be more effective for screening startups that can be assessed based on extrapolations from past data, I evaluate whether the changes in investment outcomes are concentrated on backward-similar startups. I estimate the following specification:

$$Y_{j,k,t} = \sum_{i=1}^{4} \beta_q \left\{AI_{j,t} \times SimilarityQuartile(q)_{k,t}\right\} + X_{j,k,t} + \alpha_j + \gamma_{i\times c\times t\times s} + \epsilon_{j,k,t}, \tag{5}$$

where $SimilarityQuartile(q)_{k,t}$ is a dummy variable equal to one if the backward-similarity of startup k when receiving funding at time t belongs to quartile q. $X_{j,k,t}$ includes the same time-varying control variables as specification (3) but adds $SimilarityQuartile(q)_{k,t}$.

Finally, I exploit within VC-year variations across investments. Two investments made by a VC firm in a given year can be made in startups that are more or less backward-similar. Thus these investments might benefit differentially from the adoption of AI. These variations allow me to consider a specification including VC × year fixed effects, which control for any unobserved time-varying characteristics at the VC firm level. Specifically, I estimate the following regression:

$$Y_{j,k,t} = \sum_{q=2}^{4} \beta_q \left\{AI_{j,t} \times SimilarityQuartile(q)_{k,t}\right\} + X_{j,k,t} + \alpha_{j\times t} + \gamma_{i\times c\times t\times s} + \epsilon_{j,k,t}, \tag{6}$$

where $AI_{j,t}$ is a dummy variable equal to one if VC j is classified as AI-empowered as of the year of investment and equals zero otherwise. $SimilarityQuartile(q)_{k,t}$ is defined as in specification (5). The key difference between regressions (5) and (6) is that the latter includes $\alpha_{j\times t}$, which are VC firm × year fixed effects controlling for time-varying VC-level variables potentially correlated with the adoption of AI. In this specification the coefficient on $\{AI_{j,t} \times SimilarityQuartile(1)_{k,t}\}$ cannot be estimated as it will be absorbed by the VC-year fixed effects. However, the coefficients on the remaining interactions $\{AI_{j,t} \times SimilarityQuartile(q)_{k,t}\}$ (for $q = 2, 3, 4$) can be estimated, as long as

$SimilarityQuartile(q)_{k,t}$ varies across deals made by VC j in year t. This specification allows me to test whether AI adoption affects more or less investment outcomes depending on the backward-similarity of the funded startup, holding VC-year omitted variables constant. I also consider a variation using a continuous measure of backward-similarity: I replace the dummies for each quartile in (6) by the startup's backward-similarity percentile rank.

Heterogeneous increase in the likelihood of survival and follow-on funding I start by estimating regressions with a dummy indicating whether the funded startup survives and receives follow-on funding as the dependent variable. Table 5 presents the estimation results. Columns (1) and (2) show that compared to traditional VCs, VCs that adopt AI technology observe a significant increase in the likelihood that the startups they fund survive and receive subsequent follow-on funding. The magnitude of the coefficient in column (2) is 2.6 p.p., which corresponds to a 5% increase compared to the unconditional mean of 52%. Panel A of Figure 7 illustrates the dynamics of the effects by showing the estimated coefficients β_l in specification (4). There is no pre-trend and the increase shows up in the years after AI adoption.

Column (3) of Table 5 shows that the increase in the likelihood of follow-on funding is concentrated on investments made in startups with backward-similarity in the top quartile (4 p.p.). Figure 8 illustrates graphically the coefficients on the interactions with each backward-similarity quartile. Column (4) of Table 5 drops the interaction with the first quartile and replaces it by the dummy AI-empowered. The latter captures the common effect across all quartiles of backward-similarity, while the interacted terms capture the additional effect in quartiles 2, 3 and 4. The coefficient on the interaction with the top quartile is positive and significant suggesting that the increase in the likelihood of follow-on funding is significantly larger for investments in the most backward-similar startups. Column (5) indicates that the relative increase in the follow-on funding rate for backward-similar startups continues to hold when comparing investments within a given VC-year. Column (6) shows that using a continuous measure of backward-similarity does not change the conclusion.

These results are consistent with "cream-skimming": after adopting AI, VCs are able to select the "most successful" startups from the pool of backward-similar companies, despite the fact that these startups are on average less likely to succeed. These results rule out the possibility that the increase in the number of investments after VCs adopt AI (cf. section 4.1) is simply the product of an investment strategy called "spray and pray" (Ewens et al., 2018). This strategy consists in early-stage investors providing funding to a large number of startups, with only a small proportion receiving follow-on funding.

Decrease in the likelihood of breakthrough success I estimate the same regressions using as the dependent variable a dummy indicating whether the funded company achieves breakthrough success, i.e., goes public through an IPO or is acquired for a profit. Table 6 presents the estimation results. The key takeaway is that after adopting AI, compared to traditional VCs, AI-empowered VCs observe a significant drop in the likelihood that the startups they fund achieve breakthrough success. The magnitude of the coefficient in column (2) is -2.7 p.p., which corresponds to 18% decrease compared to the unconditional mean. Panel B of Figure 7 illustrates the dynamics of the effects. There is no pre-trend and the drop in the likelihood of breakthrough success shows up after AI adoption.

Columns (3) and (4) suggest that the drop is homogeneous across all backward-similarity quartiles. Column (5) and (6) indicate that the relative drop is not significantly different across investments made in the same year by a given VC firm. Overall these results show that after adopting AI, VCs observe a statistically and economically meaningful decrease in the likelihood that the companies they select achieve breakthrough success. Table H.3 in Appendix H shows that this result remains valid when using more conservative measures of breakthrough success.

Less innovative startups In Table 7, I consider two dependent variables related to the innovation of a startup after it obtains funding: (i) the number of new patent applications and (ii) the number of new highly-cited patents. Columns (1) and (5) show that after adopting AI, compared to traditional VCs, AI-empowered VCs observe a significant drop in both the quantity and the quality of future innovation produced by the startups in which they invest. The magnitudes correspond to -6% for new patent applications and -3% for new highly-cited patents. Figure 9 illustrates the dynamics of the effects. Again, there is no pre-trend and the drops appear after AI adoption. Columns (2) and (6) suggest that, if anything, the decrease in innovation quantity and quality is larger for investments made in the most backward-similar startups. Columns (3)-(4) and (7)-(8) show that there is no significant difference across investments made in the same year by a given VC firm. Table H.4 in Appendix H shows that the conclusions remain similar when I use as the dependent variable the number of new patent grants and the number of new patents' citations.

Infrequent outcomes One might wonder why AI-empowered VCs are not able to improve their ability to pick startups that will achieve breakthrough success or impactful innovation.

Figure G.4 in Appendix G provides a potential answer. It shows that outcomes such as breakthrough success or a patent grant are much rarer than follow-on funding. This is true in the overall sample (Panel A) and in each quartile of backward-similarity (Panel B). Because AI algorithms require a large enough quantity of observations to uncover empirical patterns, it is likely that they are not able to predict with high confidence rare outcomes such as breakthrough success.

Sourcing, screening or value-added? Distinguishing whether the variations in VC investment outcomes are driven by changes in their ability to (i) find new startups (sourcing), (ii) better select startups (screening), or (iii) provide better support to their portfolio companies (value-added) is an important question. In Table H.5 in Appendix H, I estimate specifications (3) and (6) using a range of additional dependent variables. Columns (1) and (2) suggest that VCs become less likely to invest out of the regions where they have employees after adopting AI.[43] This is true all the more so when the investment is made in more backward-similar startups. Thus, it is unlikely that after adopting AI, a VC becomes better at finding successful startups that are far from its location and thus "off the radar screen".

Columns (3) and (4) consider a dummy indicating whether the startup's founding team includes a serial entrepreneur as the dependent variable. The coefficients are not significant. In columns (5) and (6), the dependent variable is a dummy indicating whether one of the founding team members holds a degree from a top school. While not statistically significant, the coefficient on the interaction between the AI-empowered dummy and the startup's backward-similarity is relatively large (0.026 which corresponds to 7% of the unconditional mean). This suggests that among investments made in a given year by the same AI-empowered VC, high backward-similarity startups tend to display a larger increase in the likelihood that their founder is a top school graduate. This result does not support that AI-empowered VCs become more likely to bet on "unproven" entrepreneurs after adopting AI.

Columns (7) and (8) use as the dependent variable a dummy indicating whether the VC is the lead investor in the funding round. Although after adopting AI VCs become overall more likely to lead deals than do other VCs, they become significantly less likely to do so when investing in backward-similar startups. Because the lead investor is usually the one that most supports the fund-raising startup, the better investment outcomes observed by AI-empowered VCs when investing in backward-similar startups is unlikely to be driven by more value added

[43] Regions are composed by surrounding states, countries, or cities (e.g., the San Francisco bay area).

to the portfolio company. Finally in columns (9) and (10), the dependent variable is a dummy indicating whether the VC sits on the board of the startup. While not statistically significant, the coefficient on the interaction between the AI-empowered dummy and the startup's backward-similarity is quite large and negative (-0.029 which corresponds to 9% of the unconditional mean). Again, this suggests that after adopting AI VCs do not become more likely to provide guidance to their portfolio companies that are backward-similar. These results indicate that the better outcomes observed by AI-empowered VCs when investing in backward-similar startups are unlikely to be driven by changes in the VCs' value-added.

Additional tests I discuss additional tests in Appendix D. First, I show that AI-empowered VCs become less likely invest in future funding rounds of their portfolio companies. This could be either because they do not invest additional capital when the startup raises additional funding or because they sell their shares. In either case, this suggests that VCs adopting AI become less involved in the growth phase of their portfolio companies. Second, I show that after adopting AI VCs do not invest more in rounds involving more investors and in which more capital is raised by startups. Third, I show that adopting AI is not associated with more "long shot bets", i.e., investments in companies that are more likely to fail but with higher step-ups in value for those that do receive follow-on financing. Fourth, my results are not different when I focus on investments corresponding to startups' first-rounds of funding. They also remain similar when I focus on the first investment by each VC firm in any given startup. Fifth, I estimate regressions at the fund level to gauge whether adopting AI is associated with changes in fund performance. After VCs adopt AI, compared to other funds, fund performance does not change statistically significantly.[44] This suggests that adopting AI might allow VCs to scale-up their business (i.e, to manage more assets) without deteriorating fund performance and thus to collect more fees.

Robustness Appendix D also presents robustness tests. I show that my finding that AI-empowered VCs observe a decrease in the likelihood that their investments lead to breakthrough success is robust to truncating my sample at different times (i.e., dropping investments made in the last years of my sample), addressing the concern that VCs might require a long time to achieve a successful exit. Furthermore, I show that my results are robust to using a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm. My conclusions also remain the same when using the imputation estimator

[44] The size of the sample used to estimate these regressions is reduced due to the fact that performance data is based on voluntary reporting of fund information.

of Borusyak et al. (2021) robust to heterogeneous effects. Finally, they also remain the same when using a Poisson model for count variables related to innovation.

5 The causal effects of adopting AI on VC investments

My results so far do not speak to whether there is a causal link between AI adoption and changes in VC investment behavior. In particular, AI adoption might be correlated with unobserved changes at the VC firm level resulting in an omitted variable bias. Furthermore, VCs might anticipate a change in their investment strategy leading to a reverse causality issue. In this section, I develop an empirical strategy to estimate the causal impact of adopting AI on VC investments. I exploit a plausibly exogenous shock to one often-cited determinant of a VC's decision to adopt this technology: the number of potential investment opportunities it faces. Intuitively, given the large fixed costs of evaluating investments (Gompers et al., 2020) and the limited scalability of VC firms (Chen et al., 2010), more investment opportunities make screening more onerous for VCs as they have to spend more time and resources evaluating investment prospects. This creates incentives for VCs to adopt AI to automate screening with a view to saving time and costs since automated screening reduces the fixed cost per investment opportunity. This is illustrated by this quote from an article in Crunchbase News: "*The rate at which startups are founded is nearly impossible to keep up with and investors are turning more to data science to keep up with the volume of companies in a period where it's nearly impossible to perform due diligence on every company in a given space*" (Metinko, 2022).

5.1 Identification strategy

I use the introduction of cloud computing as a natural experiment. Cloud computing services were first introduced by Amazon in 2006 with the release of Amazon Web Services (AWS). Ewens et al. (2018) document that AWS were initially developed for Amazon's internal infrastructure before, without anticipation by entrepreneurs or investors, being opened to developers outside Amazon in 2006. This technological shock expanded the set of newly created startups in specific sectors. In practice, cloud computing reduced the need for startups to make large fixed IT infrastructure investments in the early stages of the firm. Indeed, cloud computing enabled young companies to rent hardware in small increments instead of making large upfront hardware investments. Ewens et al. (2018) show that the ability to rent hardware allowed entrepreneurs to start software and web-based ventures at lower costs and led to more entrepreneurial experimentation. This increased the number of startup creations and thus investment opportunities for VCs active in software- and web-related industries. My hypothesis is that this shock created incentives for VCs exposed to this startup boom to adopt AI in order to automate the screening of their investments. Indeed, selecting startups from a pool featuring more opportunities requires evaluating a larger number of startups for VCs. I test whether VCs more exposed to the shock are more likely to adopt AI and change their investments in line with my findings in Section 4.

Industry exposure I exploit the cross-sectional heterogeneity in the impact of cloud computing across industries to identify the effect on startup creations and on AI adoption by VCs. The rationale is that the advent of cloud computing more severely impacted companies operating in sectors with a strong software and web component. Crunchbase categorizes companies into 744 industries according to a more detailed sectoral classification than other VC databases. This classification does not follow any major classification system, but rather is designed to account for the heterogeneity across startups' specific market segments.[45]

I assign a treatment intensity to each industry in Crunchbase proxying for the extent to which entrepreneurs in that industry were likely to benefit from the introduction of cloud computing in 2006 given the decrease in their business set up costs. To create industry-level treatment intensities, I rely on the business descriptions of firms in the Crunchbase database, including those of firms that were not VC-funded (Crunchbase does not only cover VC-funded startups). I start by collecting the cloud computing terms defined in the Cloud Computing IT Crunchbase from Solutions Review, a technology website gathering relevant content about cloud solutions.[46] Table H.7 in Appendix H reports the terms contained in this glossary. They include keywords, such as "Cloud Computing", "Database" and "Virtual Machine". Then, I search for these terms in the business description of each company that was founded before the introduction of AWS in 2006. Finally, for each industry I compute the fraction of company descriptions featuring at least one cloud term and I rank industries according to this metric.[47] Treatment intensity (between 0 and 1) is then defined as the overall percentile rank in the industry distribution:

$$IndustryExposure_i = \text{Rank}_I \left\{ \frac{\text{Nb. Company Descriptions with Match in Industry } i}{\text{Nb. Company Descriptions in Industry } i} \right\} \quad (7)$$

[45] For instance, software-related industries include "E-Commerce", "Apps", and "Enterprise Software".

[46] Cf., solutionsreview.com/cloud-platforms/glossary

[47] I only consider industries with more than 100 company business description in order to avoid assigning industry-level treatment intensities that are too dependent on a few companies.

where I is the set of industries in the Crunchbase database. Intuitively, industries in which companies mention cloud computing terms more often were more likely to benefit from the introduction of Amazon cloud services.

Panel A of Table H.8 in Appendix H shows the ten industries with the highest treatment intensities. They include industries such as "SaaS" (Software as a Service), "IaaS" (Infrastructure as a Service) and "PaaS" (Platform as a Service). The least exposed industries are presented in Panel B and encompass industries such as "Packaging Services", "Civil Engineering" and "Plastic and Rubber Manufacturing". This is not surprising as companies in these industries tend to be less likely to offer digital services and thus to benefit from the introduction of AWS.

VC exposure The extent to which the introduction of AWS affects a VC firm's number of potential investment opportunities depends on its sectoral specialization. A VC firm investing mainly in e-commerce platforms would observe a significant expansion in its set of investment opportunities, because cloud computing impacted greatly the startup set-up costs in that sector. By contrast, a VC firm investing mainly in the civil engineering sector would not see such an increase in its number of prospects. My empirical strategy makes use of this source of variations across VC firms to identify the impact of the introduction of cloud computing on VCs' AI adoption and investments.

I construct a measure, namely "VC exposure", which denotes a VC firm's ex-ante exposure to the cloud computing technology shock. To do so, I match each VC investment in my sample to the corresponding industry exposures defined above. Then, I define a VC-level exposure to AWS as the average industry exposure of its investments made before 2006 in the period prior to the introduction of AWS. I use the overall percentile rank of this metric in the VC distribution to obtain exposures between 0 and 1:

$$VCExposure_j = \text{Rank}_J \left\{ \frac{1}{N_{j,2006}} \sum_{i \in A_{j,2006}} IndustryExposure_i \right\}, \tag{8}$$

where J is the set of VCs with investments before 2006, $A_{j,2006}$ is the set of investments made by VC firm j before 2006, $N_{j,2006}$ is the number of investments in this set, $IndustryExposure_i$ is the treatment intensity of the industry of the startup corresponding to investment i, defined in (7). VC firms with the highest exposure are those with most of their investments before 2006 in industries with a high treatment intensity.[48] However, VCs with high exposure might also make

[48]I observe that VCs' industry specializations are very persistent over time. The correlation between the VC exposure (8) and the equivalent exposures computed using VC investments made up to 2009, 2012 and 2015 are

investments in industries with low treatment intensity. This creates within-industry variations across investments made by investors with different VC-level exposures.

Empirical specifications To study the effects of the introduction of AWS on VC's AI adoption and their investments, I estimate a specification using investment-level observations:

$$Y_{j,k,t} = \beta \{VCExposure_j \times Post_t\} + X_{j,k,t} + \alpha_j + \gamma_{i \times c \times t \times s} + \epsilon_{j,k,t}, \tag{9}$$

where $Y_{j,k,t}$ is either an outcome variable for a given investment made by VC j in startup k in year t or a dummy indicating whether VC j is classified as AI-empowered as of the investment date. $VCExposure_j$ is the exposure of VC j (between 0 and 1) as defined in (8). $Post_t$ is a dummy variable equal to one after 2006 and zero otherwise. $X_{j,k,t}$ are time-varying control variables which include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. α_j are VC firm fixed effects. $\gamma_{i \times c \times t \times s}$ are startup's industry $\times$ country $\times$ funding year $\times$ funding stage fixed effects. Standard errors are double-clustered at the VC firm and startup company levels.

Changes in the investments of VCs might be driven by shifts in the composition of the pool of startups available for investments. I address this problem by including startup's industry $\times$ country $\times$ funding year $\times$ funding stage fixed effects in specification (9). These stringent fixed effects capture changes in the pool of startups in that specific segment that are common to all VCs. They force the β coefficient in (9) to be estimated by comparing investments made by VCs with high exposure before versus after the advent of AWS, with the investments of other VCs with low exposure in the same industry-country-year-stage segment. Therefore, I exploit within-segment variations across investments made by VCs with different exposures.

To test for parallel pre-trends before the shock, I also estimate the effect of the introduction of AWS using the following event study specification:

$$Y_{j,k,t} = \sum_{l=-4,\ l \neq -1}^{10+} \beta_l \{VCExposure_j \times Year(l)_t\} + X_{j,k,t} + \alpha_j + \gamma_{i \times c \times t \times s} + \epsilon_{j,k,t}, \tag{10}$$

where $Year(l)_t$ is a dummy variable equal to one if year t corresponds to l years before/after the introduction of AWS. The omitted category is the year 2005.

Finally, in order to test whether the effects I observe are concentrated on investments in

0.89, 0.83 and 0.78 respectively.

backward-similar startups, I estimate the following specification :

$$Y_{j,k,t} = \sum_{i=1}^{4} \beta_q \left\{ VCExposure_j \times Post_t \times SimilarityQuartile(q)_{k,t} \right\} \tag{11}$$
$$+ X_{j,k,t} + \alpha_j + \gamma_{i \times c \times t \times s} + \epsilon_{j,k,t},$$

where $SimilarityQuartile(q)_{k,t}$ is a dummy variable equal to one if the backward-similarity of startup k when receiving funding at time t belongs to quartile q. $X_{j,k,t}$ includes the same time-varying control variables as specification (3) but adds $VCExposure_j \times SimilarityQuartile(q)_{k,t}$, $Post_t \times SimilarityQuartile(q)_{k,t}$ and $SimilarityQuartile(q)_{k,t}$.

Sample selection The empirical specifications (9), (10) and (11) require observing the industry composition of VCs' portfolios before 2006. As a result, the VC firms in this analysis tend to be older and more established than the full set of VCs I examine in Section 4. The type of selection can be seen in Table H.6 in Appendix H: my original sample includes about 220,000 investments from almost 7,200 VC firms, while the sample used for this analysis consists of approximately 110,000 investments from 1,700 firms. This sample change is explained by the fact that many VC firms in my sample never made any investment before 2006. Therefore, my empirical strategy selects larger, more established VC firms.

Identification assumptions The empirical specifications specifications (9), (10) and (11) use a continuous measure of exposure varying across VC firms. The identification relies on two key assumptions. The first one is a parallel trends assumption that is similar to the binary treatment setup (Callaway et al., 2021). This corresponds to the existence of parallel trends across investment outcomes of VCs with different exposures before the advent of cloud computing. This would ensure that investments made by less exposed VCs provide an appropriate counterfactual for what would have happened to investments made by the most exposed VCs in the absence of the introduction of AWS.

The second key assumption is that there was no other systematic change that happened at the same time as the introduction of AWS and that impacted VC activities in a similar way across sectors. A concern is that the introduction of AWS might not affect VCs' adoption of AI only through VCs' exposure to the boom in startup creations. One possibility is that the introduction of AWS may also reduce the cost of adopting AI for VCs due to the availability of cheaper hardware infrastructures. While this does not affect the validity of the AWS shock to study how an exogenous shock to AI adoption by VCs affects their investments, it does change the interpretation of why VCs adopt AI. If AWS lowered the cost of adopting AI for all VCs without any distinction across VC firms, the time-specific fixed effects in (9) would absorb this overall drop. Because my empirical strategy exploits variations in exposure to AWS across VCs due to differences in industry specialization, the heterogeneity in VC exposure in (8) still persists because of the industry specializations of each VC: some VCs would still observe a larger increase in their number of investment prospects and thus have stronger incentives to adopt AI.

A more subtle concern might be that the introduction of AWS lowered the cost of adopting AI for some VCs but not others. In that case, the ex-ante industry composition of a VC's portfolio might be correlated with the magnitude of the drop in its cost of adopting AI. If so, the VC firms that invest in industries with the strongest cloud-related component have presumably more expertise to exploit the availability of AWS. These VCs might thus be more likely to adopt AI algorithms in their operations due to the availability of cheaper cloud infrastructures. I show that this is not what drives my findings. The results presented below remain robust to restricting my sample to the VC firms that never make any investment (before 2006) in the most cloud-related industries, i.e., industries with exposure (8) in the top 5%. This suggests that, even among the VCs without any specific cloud computing expertise before the shock, VCs facing a larger increase in their number of investment opportunities are more likely to adopt AI.

Startup creations across industries In Appendix E, I provide evidence that the introduction of AWS led to a larger increase in startup creations in the most exposed industries. I show that in the industries at the 90th percentile in terms of exposure, the number of firm creations increases by 36% more after the shock than in the industries at the 10th percentile of exposure. Figure 10 reports the results from a dynamic specification using as the dependent variable the logarithm of (one plus) the number of newly created companies in a given industry-country-year. It shows there is no pre-trend and that the timing of the increase in the number of company creations is consistent with the introduction of AWS. In Appendix E, I also discuss how AI adoption by VCs varies *across* industries depending on the extent of the boom in startup creations.

5.2 Effects of the introduction of AWS on AI adoption and VC investments

AI adoption I test whether VC firms with higher exposure are more likely to adopt AI technology after the introduction of AWS. Table 8 estimates specification (9) using as the dependent variable a dummy equal to one if the investment is made by a VC firm classified as AI-empowered as of the investment date and equal to zero otherwise. Columns (1) and (2) show that after the advent of cloud computing and within a given industry-country-stage, the likelihood of observing an investment made by an AI-empowered VC increases significantly with the VC's exposure. In other words, VCs with an ex-ante larger fraction of their investments in the industries with high treatment intensity are more likely to adopt AI after the shock. Column (2) corresponds to a rise by 13 p.p. for investments made by VCs at the 90th percentile rank in terms of exposure, corresponding to 160% increase compared to the unconditional mean. Table H.9 in Appendix H estimates a similar regression at the VC-year level. The magnitude relative to the unconditional mean are of the same order of magnitude.

Figure 11 graphs the estimated coefficients β_l in equation (10). It shows no pre-trend. The increase in the likelihood of observing an investment made by an AI-empowered VC shows up in the years after the advent of cloud computing and persists even 10 years after. Panel A of Figure F.1 in Appendix F show there is no significant variation for VCs with low exposure.

Backward-similarity Then, I investigate whether after the introduction of cloud computing VCs with higher exposure (8) are more likely to invest in backward-similar startups. Columns (3) and (4) of Table 8 present estimation results of specification (11) including triple interactions with the backward-similarity quartile of the corresponding VC-funded startup. Column (3) shows that the rise in the likelihood of observing an investment made by an AI-empowered VC is larger if the funded startup has a high backward-similarity. The magnitudes of the coefficients indicate that the effect is twice as large for investments in startups in the top versus bottom quartile of backward-similarity. In column (4), I replace the triple interaction including the first quartile by $VCExposure_i \times Post_t$. The coefficient on this term captures the increase in the likelihood that the investment is performed by an AI-empowered VC for startups with backward-similarity in the first quartile. The remaining triple interactions capture the additional effect for startups in the second, third and fourth quartile. The coefficient on the interaction between $VCExposure$, $Post$ and $SimilarityQuartile(4)$ is positive and highly significant. This implies that the increase in the likelihood of the investment being made by an AI-empowered VC is significantly larger for startups in the fourth quartile of backward-similarity than for those in the first quartile, keeping constant the VC exposure.

Survival and follow-on funding I test whether the investment outcomes of VCs with higher exposure are also affected. Table 9 presents the estimation results of the same specifications presented in Table 8 but uses as the dependent variable a dummy indicating whether the funded startup survives and receives follow-on funding. In columns (1) and (2), the coefficient on the interaction between $VCExposure$ and $Post$ is positive but not significant. Panel A of Figure 12 shows the estimated coefficients β_l in equation (10), with 95% confidence intervals. One does not detect a meaningful change in the likelihood of VCs with higher exposure investing in startups receiving subsequent funding after the shock.

Column (3) of Table 9 reveals that there is a significant rise in the likelihood that the VC-funded company later receives follow-on funding but only when the investment is made in a startup with backward-similarity in the top quartile. The magnitude of the coefficient indicates that the relative rise is approximately 8 p.p. for VCs at the 90th percentile of exposure, i.e., a 16% increase compared to the unconditional mean. Column (4) confirms that the coefficient on the triple interaction including the fourth quartile is significantly different from the one including the first quartile. These results are consistent with my findings in subsection 4.3: VCs that are the most likely to adopt AI observe improvements in their ability to screen high backward-similarity startups.

Breakthrough success In Table 10, I consider as the dependent variable a dummy indicating whether the VC-funded startup will achieve breakthrough success by 2021. The coefficient on the interaction between $VCExposure$ and $Post$ in columns (1) and (2) is negative and significant. Its magnitude is substantial: -8 p.p. (37% compared to the unconditional mean) for VCs at the 90th percentile of exposure. Panel B of Figure 12 plots the estimated coefficients β_l in equation (10). It shows no pre-trend and a drop in the likelihood of breakthrough success for startups funded by highly-exposed VCs after the advent of cloud computing, which is a trend that persists even 10 years after. Panel B of Figure F.1 in Appendix F show there is no significant variation for VCs with low exposure.

Columns (3) and (4) of Table 10 suggest that the drop in the likelihood of breakthrough success is not statistically different across the different backward-similarity quartiles. Again, these results are consistent with my findings in subsection 4.3: VCs that are the most likely to adopt AI observe a meaningful decrease in the likelihood of selecting startups that achieve breakthrough success. Note though that the magnitudes discussed above are larger than those in subsection 4.3. This could be explained by the fact that VCs in the sample considered here are more experienced and thus might undergo a larger drop after adopting AI.

Innovation In Table 11, I investigate the effect on the quantity and quality of innovation later produced by VC-funded startups (number of new patent applications and number of new highly-cited patents). None of the coefficients are statistically significant and standard errors are in the same order of magnitude as the point estimates. However, I observe large and negative coefficients on the triple interactions including the first and fourth quartiles.

Discussion of magnitudes One issue with the specification (9) is that it estimates the effect of $VCExposure$, not of AI adoption. Thus, it can make the interpretation of the magnitude of the effect of AI adoption on investment outcomes complicated. One can estimate the effect of AI adoption by dividing the effect of $VCExposure$ on investment outcomes by the effect of $VCExposure$ on AI adoption.[49] This rescaling allows me to obtain an interpretation in terms of the effect of AI adoption on investment outcomes. The corresponding effect of AI adoption on the likelihood that a VC-funded startup in the top quartile of backward-similarity survives and receives follow-on funding is $0.086/0.141 = 0.610$. The effect of AI adoption on the likelihood that a VC-funded startup achieves breakthrough success is -0.086/0.141 = -0.610.[50] In addition, my results in section 4.1 suggest that, compared to other VCs, VC firms that adopt AI increase by 28% and 94% their overall number of investments and their number of investments in high-backward similarity startups respectively. Taken together, these results allow me to compute a simple "back-of-the-envelope" calculation of how AI adoption among VCs might map to aggregate outcomes, abstracting from general equilibrium effects. In my sample, over 2000-2019, 5,721 VC-funded startups achieve breakthrough success, i.e., 286 startups per year. Over the same period, 8,014 VC-funded high backward-similarity startups (top quartile) survive and obtain follow-on funding, i.e., 401 startups per year. A simple "back-of-the-envelope" computation suggests that a 10 p.p. increase in the number of VCs adopting AI would imply an effect of $0.1 \times 286 \times [1.28 \times (1 - 0.610) - 1] = -14$ on the number of VC-funded startups achieving breakthrough success, and of $0.1 \times 401 \times [1.94 \times (1 + 0.610) - 1] = 85$ on the number of high backward-similarity VC-funded startups that survive and receive follow-on funding.[51]

[49] It can also be done with 2SLS regressions, which lead to the same estimate.

[50] These effects are larger in magnitude than those suggested by OLS estimates in subsection 4.3. This might be due to several factors. The first possibility is that the OLS estimate is biased because of omitted variables correlated with AI adoption and investment success. A second possibility is that, because they are more experienced, the VCs studied in this section would have, in the absence of AI adoption, funded more startups eventually achieving major success than the average VC in the full sample considered in 4.3.

[51] This analysis comes with several caveats. First, the causal effect of AI adoption is only estimated on the sub-sample of VCs that make investments before 2006 in my sample. Second, more generally, when aggregating

Excluding VCs with cloud computing expertise As discussed in subsection 5.1, one concern might be that the introduction of AWS has a direct effect on the cost of adopting AI for VCs, that might be heterogeneous across VCs depending on their ex-ante expertise. If my results are driven by the increase in the number of investment opportunities for VCs, they should be robust to excluding VCs with cloud computing expertise before the shock, which are the most likely to be able to use cloud computing services. I test this by estimating (9) but using only the subset of investments made by VCs that never make any investment in the most cloud-related industries before 2006 (i.e., industries with exposure (8) in the top 5%). The results are reported in Appendix F. My conclusions remain similar.

Excluding the most exposed industries One concern might be that the above effects are only driven by investments made in the most exposed industries. However, given the low marginal cost of using AI to screen an additional investment, one expects VCs adopting AI to use it to screen their investments in all industries. If this is the case, my conclusions should be robust to excluding investments made in the most treated industries. I test this by estimating (9) but only on the subset of investments in industries with exposure below 0.8. The results are reported in Appendix F. They remain similar and are if anything stronger.

Additional robustness tests Callaway et al. (2021) show that interpreting treatment effects across different treatment values (here VC exposure) might be challenging in the presence of treatment effect heterogeneity. In Appendix F, I discuss a robustness test replacing $VCExposure_i$ by a dummy $HighVCExposure_i$, which takes the value of one for VCs with exposure above the median (0.5) and equals zero otherwise. My conclusions remain the same.

6 Conclusion

This paper presents novel evidence on how the adoption of artificial intelligence (AI) by venture capitalists (VCs) to screen startups affects the funding of young, innovative companies.

Using information on employees at VC firms, I identify VCs that employ data scientists to develop machine learning tools for investment screening, and I call these VCs AI-empowered VCs. AI algorithms may be more effective to screen startups that can be assessed based on past data. Leveraging natural language processing techniques, I run a textual analysis of startups' business

the causal effect of VC exposure on investment success, I do not take into consideration the general equilibrium effects that might arise both among VCs and among entrepreneurs.

descriptions to construct a measure of backward-similarity, capturing the extent to which a startup's business is similar to those of previously VC-funded companies. This indicator proxies for the amount of historical data on comparable startups available to train AI algorithms.

I find that after adopting AI VCs increase their investments in high backward-similar startups compared to other VCs. Within this pool of startups, AI-empowered VCs become better at picking those that survive and receive follow-on funding. However, AI-empowered VCs become less likely to pick startups that eventually achieve major success or a breakthrough, e.g., IPOs or highly-cited patents. This suggests that big disruptions creating significant wealth and pushing the innovation frontier are unlikely to be predictable based on extrapolations of past data.

For causal identification, I use a plausibly exogenous shock to one determinant of a VC's decision to adopt AI: the number of potential investment opportunities it faces. The intuition is that more investment prospects make screening more costly for humans and create a stronger incentive to automate this task using AI. My empirical strategy exploits the introduction of cloud computing by Amazon (AWS) as a natural experiment. This technological shock led to a significant reduction in the cost of setting up businesses in software- and web-related industries, which significantly increased the number of startup creations in specific sectors.

Using a text-based measure of the treatment intensity of industries, I compute each VC firm's exposure to the AWS shock. This measure is based on the sectoral composition of the VC's portfolio before the shock. I show that high-exposure VCs are more likely to adopt AI after the advent of cloud computing compared to low-exposure VCs. In addition, high-exposure VCs are more likely to shift their portfolio towards backward-similar startups. Finally, high-exposure VCs become better at selecting startups that survive and receive follow-on funding but only among those that are highly backward-similar. However, high-exposure VCs become less likely to invest in startups that eventually achieve major success, e.g., IPOs.

From the perspective of entrepreneurs, my results shed light on how VC capital supply might evolve in the future. My findings suggest that the adoption of AI by VCs can change how they allocate capital across startups. AI-based screening can have a positive impact on access to capital but only for startups that can be evaluated based on machine-processable data. Thus, AI adoption among VCs might reduced the share of VC funding allocated to breakthrough innovations.[52] This raises the question of whether this might induce entrepreneurs to produce more backward-similar ventures at the expense of breakthrough innovations.

[52] According to Gartner (a technological research consulting firm), AI will be involved in 75% of VC investment decisions by 2025. Cf., gartner.com[...]data-science[...]above-gut-feel-for-investment-decisions-by-2025

References

Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. *Machine: Quantitative and Discretionary Equity Management (October 23, 2020).*

Abis, S. and Veldkamp, L. (2020). The changing economics of knowledge production. *Unpublished working paper. Columbia Business School.*

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

Aghion, P., Jones, B. F., and Jones, C. I. (2019). *Artificial Intelligence and Economic Growth.* University of Chicago Press.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). *Prediction machines: the simple economics of artificial intelligence.* Harvard Business Press.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2019a). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2):31–50.

Agrawal, A., McHale, J., and Oettl, A. (2019b). *Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth.* University of Chicago Press.

Alekseeva, L., Azar, J., Gine, M., Samila, S., and Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, page 102002.

Åstebro, T. (2002). Assessing the commercial viability of seed-and early-stage ventures. *The Journal of Private Equity*, 6(1):9–12.

Åstebro, T. (2021). An inside peek at AI use in private equity. *The Journal of Financial Data Science*, 3(3):97–107.

Åstebro, T. and Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science*, 52(3):395–409.

Aubry, M., Kraeussl, R., Manso, G., and Spaenjers, C. (2022). Biased auctioneers. *Journal of Finance, Forthcoming.*

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333.

Babina, T., Fedyk, A., He, A. X., and Hodson, J. (2021). Artificial intelligence, firm growth, and product innovation. *Firm Growth, and Product Innovation (November 9, 2021).*

Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122(3):625–654.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2021). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics.*

Bernstein, S. (2015). Does going public affect innovation? *The Journal of finance*, 70(4):1365–1403.

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

Bernstein, S., Korteweg, A., and Laws, K. (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *The Journal of Finance*, 72(2):509–538.

Bhakdi, J. (2013). Quantitative vc: A new way to growth. *The Journal of Private Equity*, 17(1):14–28.

Birru, J., Gokkaya, S., and Liu, X. (2019). Capital market anomalies and quantitative research. *Fisher College of Business Working Paper*, (2018-03):007.

Blattner, L. and Nelson, S. (2021). How costly is noise? data and disparities in consumer credit. *arXiv preprint arXiv:2105.07554*.

Blohm, I., Antretter, T., Wincent, J., Sirén, C., and Grichnik, D. (2020). It's a peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice*.

Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J., and Tahoun, A. (2020a). The geography of new technologies. *Institute for New Economic Thinking Working Paper Series*, (126).

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020b). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Breschi, S., Lassébie, J., and Menon, C. (2018). A portrait of innovative start-ups across countries.

Brynjolfsson, E., Rock, D., and Syverson, C. (2018). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *The economics of artificial intelligence: An agenda*, pages 23–57. University of Chicago Press.

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of financial economics*, 130(3):453–483.

Calder-Wang, S. and Gompers, P. A. (2021). And the children shall lead: Gender diversity and performance in venture capital. *Journal of Financial Economics*, 142(1):1–22.

Callaway, B., Goodman-Bacon, A., and Sant'Anna, P. H. (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Cao, S., Jiang, W., Wang, J. L., and Yang, B. (2021). From man vs. machine to man+ machine: The art and AI of stock analyses. Technical report, National Bureau of Economic Research.

Casamatta, C. (2003). Financing and advising: optimal financial contracts with venture capitalists. *The journal of finance*, 58(5):2059–2085.

Chen, H., Gompers, P., Kovner, A., and Lerner, J. (2010). Buy local? the geography of venture capital. *Journal of Urban Economics*, 67(1):90–102.

Chen, X. J. (2021). How AI is transforming venture capital. *BRINK*.

Chi, F., Hwang, B.-H., and Zheng, Y. (2021). The use and usefulness of big data in finance: Evidence from financial analysts.

Cochrane, J. H. (2005). The risk and return of venture capital. *Journal of financial economics*, 75(1):3–52.

Cockburn, I. M., Henderson, R., and Stern, S. (2019). *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis*. University of Chicago Press.

Cohn, J. B., Liu, Z., and Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551.

Coleman, B., Merkley, K. J., and Pacelli, J. (2021). Man versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations. *Available at SSRN 3514879*.

Colonnelli, E., Li, B., and Liu, E. (2022). Investing with the government: A field experiment in china. *Available at SSRN*.

Da Rin, M., Hellmann, T., and Puri, M. (2013). A survey of venture capital research. In *Handbook of the Economics of Finance*, volume 2, pages 573–648. Elsevier.

Dalle, J.-M., Den Besten, M., and Menon, C. (2017). Using crunchbase for economic and managerial research.

Davenport, D. (2022). Predictably bad investments: Evidence from venture capitalists. *Available at SSRN 4135861*.

De Chaisemartin, C. and D'Haultfoeuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *National Bureau of Economic Research*.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2016). Where has all the skewness gone? the decline in high-growth (young) firms in the us. *European Economic Review*, 86:4–23.

Dessaint, O., Foucault, T., and Frésard, L. (2021). Does alternative data improve financial forecasting? the horizon effect. *HEC Paris Research Paper No. FIN-2020-1402, Swiss Finance Institute Research Paper*, (20-106).

Di Maggio, M., Ratnadiwakara, D., and Carmichael, D. (2022). Invisible primes: Fintech lending with alternative data. Technical report, National Bureau of Economic Research.

Di Maggio, M. and Yao, V. (2021). Fintech borrowers: lax screening or cream-skimming? *The review of financial studies*, 34(10):4565–4618.

Dugast, J. and Foucault, T. (2018). Data abundance and asset price informativeness. *Journal of Financial Economics*, 130(2):367–391.

D'Acunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5):1983–2020.

Erel, I., Stern, L. H., Tan, C., and Weisbach, M. S. (2021). Selecting directors using machine learning. *The Review of Financial Studies*, 34(7):3226–3264.

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

Ewens, M. and Townsend, R. R. (2020). Are early stage investors biased against women? *Journal of Financial Economics*, 135(3):653–677.

Farboodi, M., Matray, A., Veldkamp, L., and Venkateswaran, V. (2022). Where has all the data gone? *The Review of Financial Studies*, 35(7):3101–3138.

Farboodi, M. and Veldkamp, L. (2020). Long-run growth of financial data technology. *American Economic Review*, 110(8):2485–2523.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2020). What is a patent worth? evidence from the us patent "lottery". *The Journal of Finance*.

Ferrati, F. and Muffatto, M. (2020). Using crunchbase for research in entrepreneurship: data content and structure.

Furman, J. and Seamans, R. (2019). AI and the economy. *Innovation policy and the economy*, 19(1):161–191.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2021). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5):1854–1899.

Gans, J., Scott, E. L., and Stern, S. (2018). Strategy for start-ups. *Harvard Business Review*, 96(3):44–51.

Gao, M. and Huang, J. (2020). Informing the market: The effect of modern information technologies on information production. *The Review of Financial Studies*, 33(4):1367–1411.

Gartner (2020). Tech providers 2025: Strategic transformation drives growth. *Gartner special report for technology providers*.

Goldfarb, A., Taska, B., and Teodoridis, F. (2021). Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings. Technical report, Working paper, University of Toronto. https://papers. ssrn. com/sol3/papers. cfm.

Gompers, P., Kovner, A., Lerner, J., and Scharfstein, D. (2010). Performance persistence in entrepreneurship. *Journal of financial economics*, 96(1):18–32.

Gompers, P. A., Gornall, W., Kaplan, S. N., and Strebulaev, I. A. (2020). How do venture capitalists make decisions? *Journal of Financial Economics*, 135(1):169–190.

Gompers, P. A., Mukharlyamov, V., and Xuan, Y. (2016). The cost of friendship. *Journal of Financial Economics*, 119(3):626–644.

Gonzalez-Uribe, J. and Leatherbee, M. (2018). The effects of business accelerators on venture performance: Evidence from start-up chile. *The Review of Financial Studies*, 31(4):1566–1603.

González-Uribe, J. and Reyes, S. (2021). Identifying and boosting "gazelles": Evidence from business accelerators. *Journal of Financial Economics*, 139(1):260–287.

Gornall, W. and Strebulaev, I. A. (2021). The economic impact of venture capital: Evidence from public companies. *Available at SSRN 2681841*.

Grennan, J. and Michaely, R. (2020). Artificial intelligence and high-skilled work: Evidence from analysts. *Available at SSRN*.

Guzman, J. and Li, A. (2022). Measuring founding strategy. *Management Science*.

Guzman, J. and 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.

Harris, R. S., Jenkinson, T., and Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, 69(5):1851–1882.

Hellmann, T. and Puri, M. (2002). Venture capital and the professionalization of start-up firms: Empirical evidence. *The journal of finance*, 57(1):169–197.

Hellmann, T. and Thiele, V. (2015). Friends or foes? the interrelationship between angel and venture capital markets. *Journal of Financial Economics*, 115(3):639–653.

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

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

Hochberg, Y. V., Ljungqvist, A., and Vissing-Jørgensen, A. (2014). Informational holdup and performance persistence in venture capital. *The Review of Financial Studies*, 27(1):102–152.

Howell, S. T., Lerner, J., Nanda, R., and Townsend, R. (2020). How resilient is venture-backed innovation? evidence from four decades of us patenting. *NBER Working Paper*, (w27150).

Howell, S. T. and Nanda, R. (2019). Networking frictions in venture capital, and the gender gap in entrepreneurship. Technical report, National Bureau of Economic Research.

Hunter, D. S., Saini, A., and Zaman, T. (2017). Picking winners: A data driven approach to evaluating the quality of startup companies. *arXiv preprint arXiv:1706.04229*.

Jaffe, A. B. and De Rassenfosse, G. (2019). Patent citation data in social science research: Overview and best practices. *Research handbook on the economics of intellectual property law*.

Jansen, M., Nguyen, H., and Shams, A. (2021). Rise of the machines: The impact of automated underwriting. *Fisher College of Business Working Paper*, (2020-03):019.

Kamepalli, S. K., Rajan, R., and Zingales, L. (2020). Kill zone. Technical report, National Bureau of Economic Research.

Kaplan, S. N. and Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. *The journal of finance*, 60(4):1791–1823.

Kaplan, S. N., Sensoy, B. A., and Strömberg, P. (2009). Should investors bet on the jockey or the horse? evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance*, 64(1):75–115.

Kaplan, S. N. and Stromberg, P. (2001). Venture capitals as principals: Contracting, screening, and monitoring. *American Economic Review*, 91(2):426–430.

Kaplan, S. N. and Strömberg, P. (2003). Financial contracting theory meets the real world: An empirical analysis of venture capital contracts. *The review of economic studies*, 70(2):281–315.

Kaplan, S. N. and Strömberg, P. E. (2004). Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *The Journal of Finance*, 59(5):2177–2210.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–20.

Kerr, W. R., Lerner, J., and Schoar, A. (2014a). The consequences of entrepreneurial finance: Evidence from angel financings. *The Review of Financial Studies*, 27(1):20–55.

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

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2018). Human decisions and machine predictions. *The quarterly journal of economics*, 133(1):237–293.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712.

Korteweg, A. and Sorensen, M. (2010). Risk and return characteristics of venture capital-backed entrepreneurial companies. *The Review of Financial Studies*, 23(10):3738–3772.

Kortum, S. and Lerner, J. (2001). Does venture capital spur innovation? In *Entrepreneurial inputs and outcomes: New studies of entrepreneurship in the United States*. Emerald Group Publishing Limited.

Lerner, J. (1995). Venture capitalists and the oversight of private firms. *the Journal of Finance*, 50(1):301–318.

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

Lerner, J., Schoar, A., Sokolinski, S., and Wilson, K. (2018). The globalization of angel investments: Evidence across countries. *Journal of Financial Economics*, 127(1):1–20.

Lerner, J. and Seru, A. (2022). The use and misuse of patent data: Issues for finance and beyond. *The Review of Financial Studies*, 35(6):2667–2704.

Li, D., Raymond, L. R., and Bergman, P. (2020). Hiring as exploration. Technical report, National Bureau of Economic Research.

Li, X., Liu, T., and Taylor, L. A. (2022). Common ownership and innovation efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Ling, Y. (2015). The impact of venture capital on the life cycles of startups. *Available at SSRN 2709243*.

Lyonnet, V. and Stern, L. H. (2022). Venture capital (mis) allocation in the age of AI. *Fisher College of Business Working Paper*, (2022-03):002.

Manso, G. (2016). Experimentation and the returns to entrepreneurship. *The Review of Financial Studies*, 29(9):2319–2340.

Martin, I. W. and Nagel, S. (2022). Market efficiency in the age of big data. *Journal of financial economics*, 145(1):154–177.

Metinko, C. (2022). As venture capital continues to change, what will 2022 bring? *Crunchbase News*.

Mihet, R. (2020). Financial innovation and the inequality gap. Technical report, Working Paper.

Mollica, M. and Zingales, L. (2007). The impact of venture capital on innovation and the creation of new businesses. *Unpublished working paper, University of Chicago*.

Moser, P. (2012). Innovation without patents: Evidence from world's fairs. *The Journal of Law and Economics*, 55(1):43–74.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Nanda, R., Samila, S., and Sorenson, O. (2020). The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics*, 137(1):231–248.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway books.

Piacentino, G. (2019). Venture capital and capital allocation. *The Journal of Finance*, 74(3):1261–1314.

Reher, M. and Sokolinski, S. (2021). Does automation democratize asset management. Technical report, Working Paper.

Retterath, A. (2020). Human versus computer: Benchmarking venture capitalists and machine learning algorithms for investment screening. *Available at SSRN*.

Retterath, A. and Braun, R. (2020). Benchmarking venture capital databases. *Available at SSRN*.

Robinson, D. T. and Sensoy, B. A. (2013). Do private equity fund managers earn their fees? compensation, ownership, and cash flow performance. *The Review of Financial Studies*, 26(11):2760–2797.

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

Röhm, S., Bick, M., and Boeckle, M. (2022). The impact of artificial intelligence on the investment decision process in venture capital firms. In *International Conference on Human-Computer Interaction*, pages 420–435. Springer.

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

Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. *Innovation policy and the economy*, 10(1):57–81.

Sørensen, M. (2007). How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance*, 62(6):2725–2762.

van Binsbergen, J. H., Han, X., and Lopez-Lira, A. (2022). Man vs. machine learning: The term structure of earnings expectations and conditional biases. *The Review of Financial Studies*.

Wu, A. (2016). Organizational decision-making and information: Angel investments by venture capital partners. In *Academy of Management Proceedings*, volume 2016, page 11043. Academy of Management Briarcliff Manor, NY 10510.

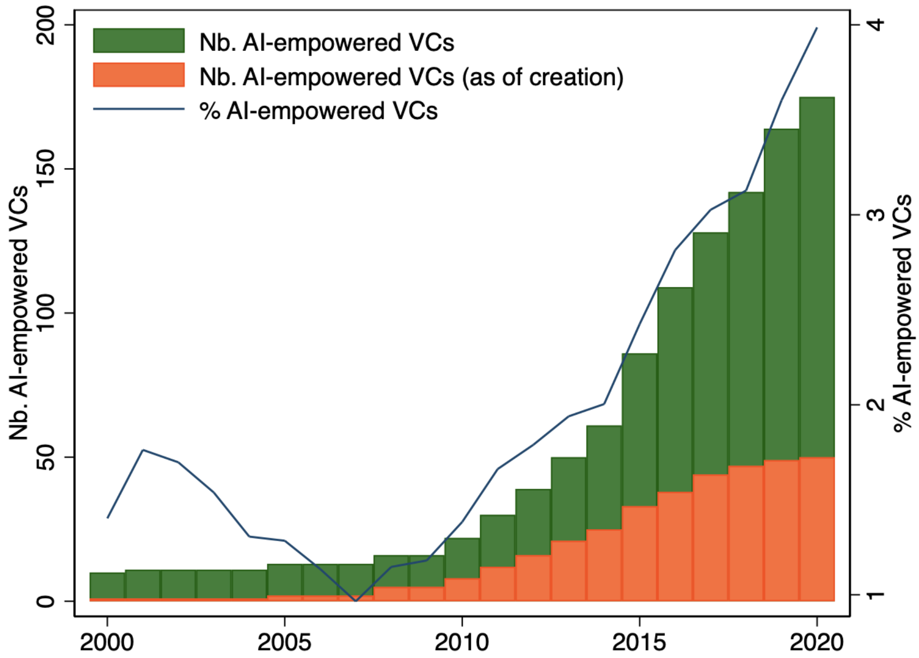
Wu, V. and Gnanasambandam, C. (2017). A machine-learning approach to venture capital. *The McKinsey Quarterly*.

Zacharakis, A. L. and Meyer, G. D. (2000). The potential of actuarial decision models: can they improve the venture capital investment decision? *Journal of Business venturing*, 15(4):323–346.

Zhu, C. (2019). Big data as a governance mechanism. *The Review of Financial Studies*, 32(5):2021–2061.

Figures

Panel A: Number and percentage of AI-empowered VCs



Panel B: Number and percentage of funding rounds involving AI-empowered VCs

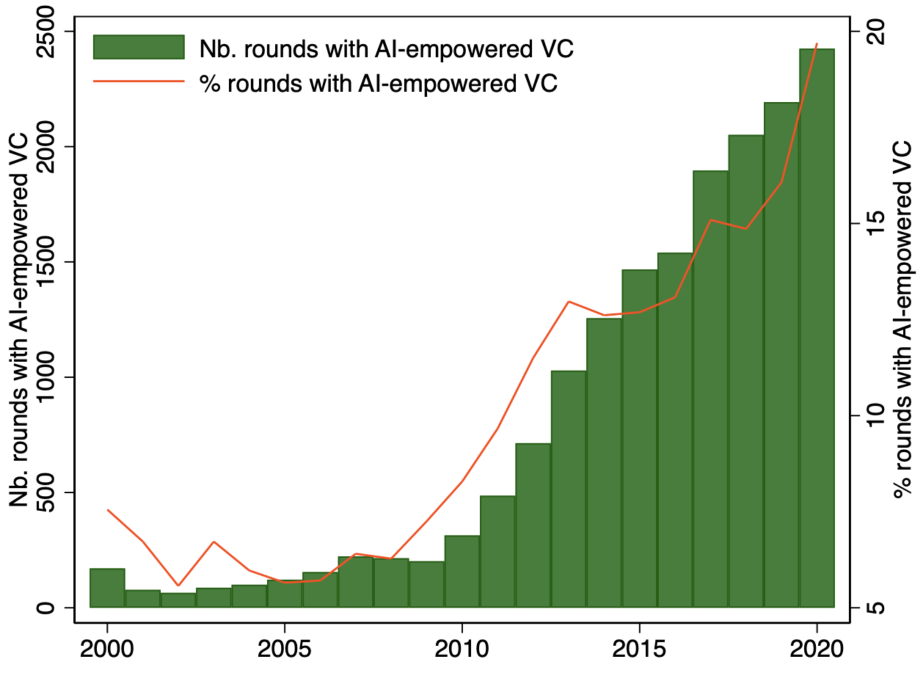


Figure 1: **Evolution of the presence of AI-empowered VCs over time**: Panel A reports for each year between 2000 and 2020 the number and percentage of VC firms classified as AI-empowered. Orange bars represent the number of newly created AI-empowered VCs, i.e., VCs classified as AI-empowered as of their creation year. Panel B reports the number and percentage of funding rounds with at least one VC investor classified as AI-empowered.

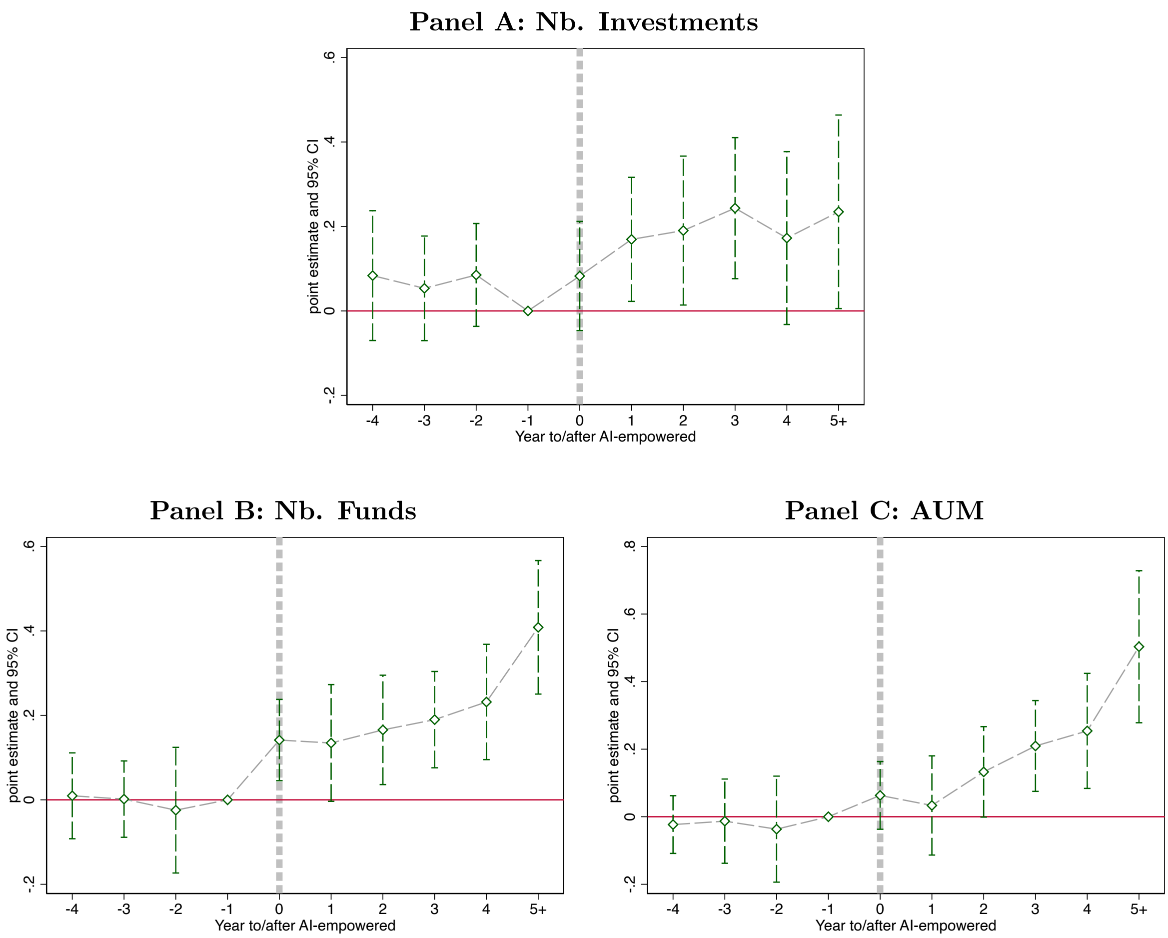


Figure 2: **Number of investments, funds and AUM before/after VCs adopt AI**: These figures plot the estimated coefficients from regressions at the VC-year level for the terms corresponding to each year relative to AI adoption. In Panel A, the dependent variable is the logarithm of (one plus) the number of investments made by the VC firm in a given year. In Panel B, the dependent variable is the logarithm of the number of funds managed by the VC firm in that year. In Panel C, the dependent variable is the logarithm of the AUM of funds managed by the VC firm in that year. The year preceding AI adoption (-1) is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and year level. Regressions include VC firm fixed effects and country-stage-year fixed effects. Country denotes the country of the VC's headquarters and stage denotes the stage of funding in which the VC firm makes the largest number of investments over the sample period (among six categories, i.e., pre-seed, seed, series A, series B, series C, or series D and onward). Regressions also control for the logarithm of the VC firm's age and the logarithm of its number of employees, both measured as of the year of investment.

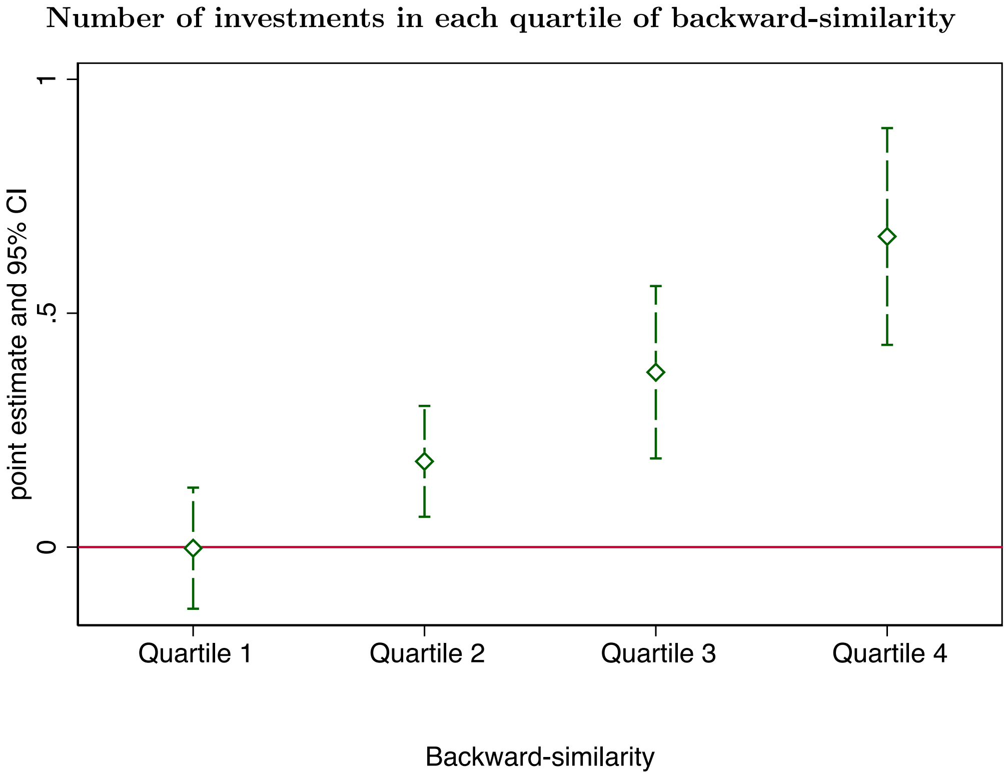
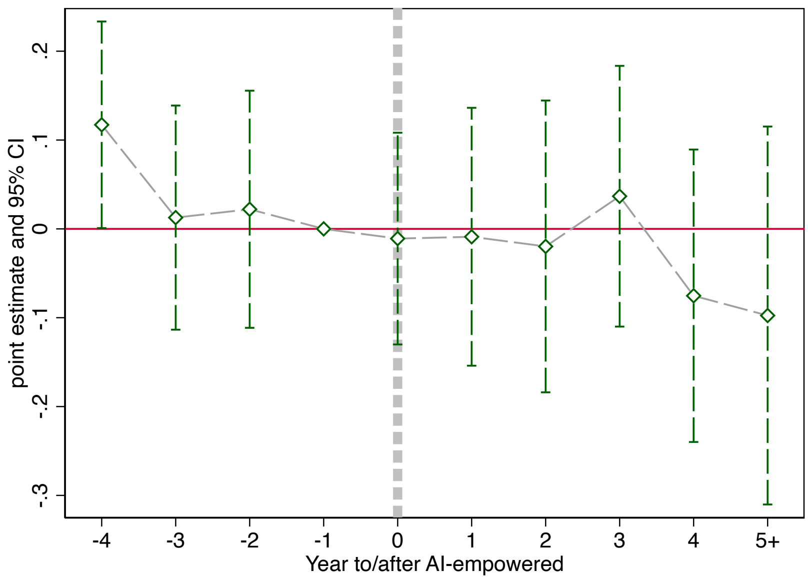


Figure 3: **Number of investments in each quartile of backward-similarity after VCs adopt AI**: This figure plots the estimated coefficients from regressions at the VC-year level. Each point represents the coefficient on a dummy equal to one if the VC is classified as AI-empowered in that year using a different dependent variable: the logarithm of (one plus) the number of investments in funding rounds of startups with backward-similarity in a certain quartile. For instance "Similarity Quartile 4" refers to investments in funding rounds of startups with backward-similarity in the top 25 percent. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and year level. Regressions include VC firm fixed effects and country-stage-year fixed effects. Country denotes the country of the VC's headquarters and stage denotes the stage of funding in which the VC firm makes the largest number of investments over the sample period (among six categories, i.e., pre-seed, seed, series A, series B, series C, or series D and onward). Regressions also control for the logarithm of the VC firm's age and the logarithm of its number of employees, both measured as of the year of investment.

Panel A: Number of investments: Backward-similarity quartile 1



Panel B: Number of investments: Backward-similarity quartile 4

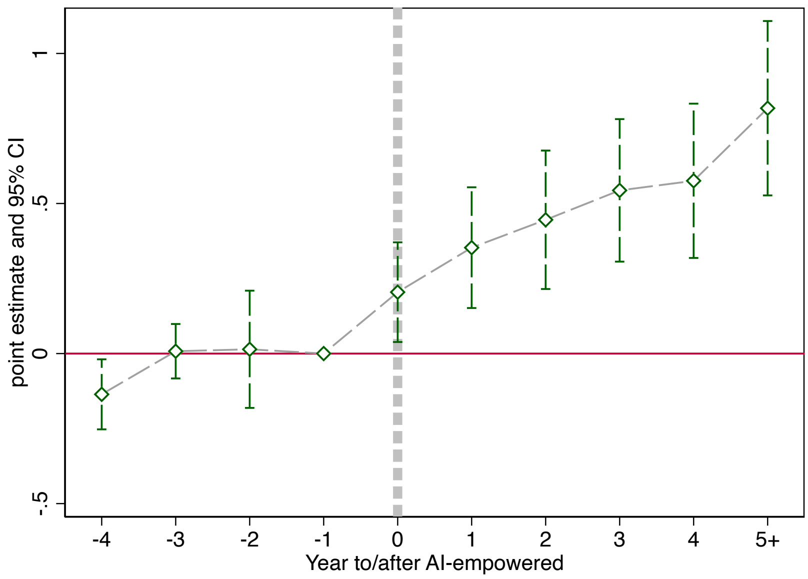


Figure 4: **Number of investments in the top and bottom quartile of backward-similarity before/after VCs adopt AI**: These figures plot the estimated coefficients from regressions at the VC-year level for the terms corresponding to each year relative to AI adoption. In Panel A, the dependent variable is the logarithm of (one plus) the number of investments in startups with backward-similarity percentile rank below 0.25, made by the VC firm in a given year. In Panel B, the dependent variable is the logarithm of (one plus) the number of investments in startups with backward-similarity percentile rank above 0.75, made by the VC firm in a given year. The year preceding AI adoption (-1) is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and year level. Regressions include VC firm fixed effects and country-stage-year fixed effects. Country denotes the country of the VC's headquarters and stage denotes the stage of funding in which the VC firm makes the largest number of investments over the sample period (among six categories, i.e., pre-seed, seed, series A, series B, series C, or series D and onward). Regressions also control for the logarithm of the VC firm's age and the logarithm of its number of employees, both measured as of the year of investment.

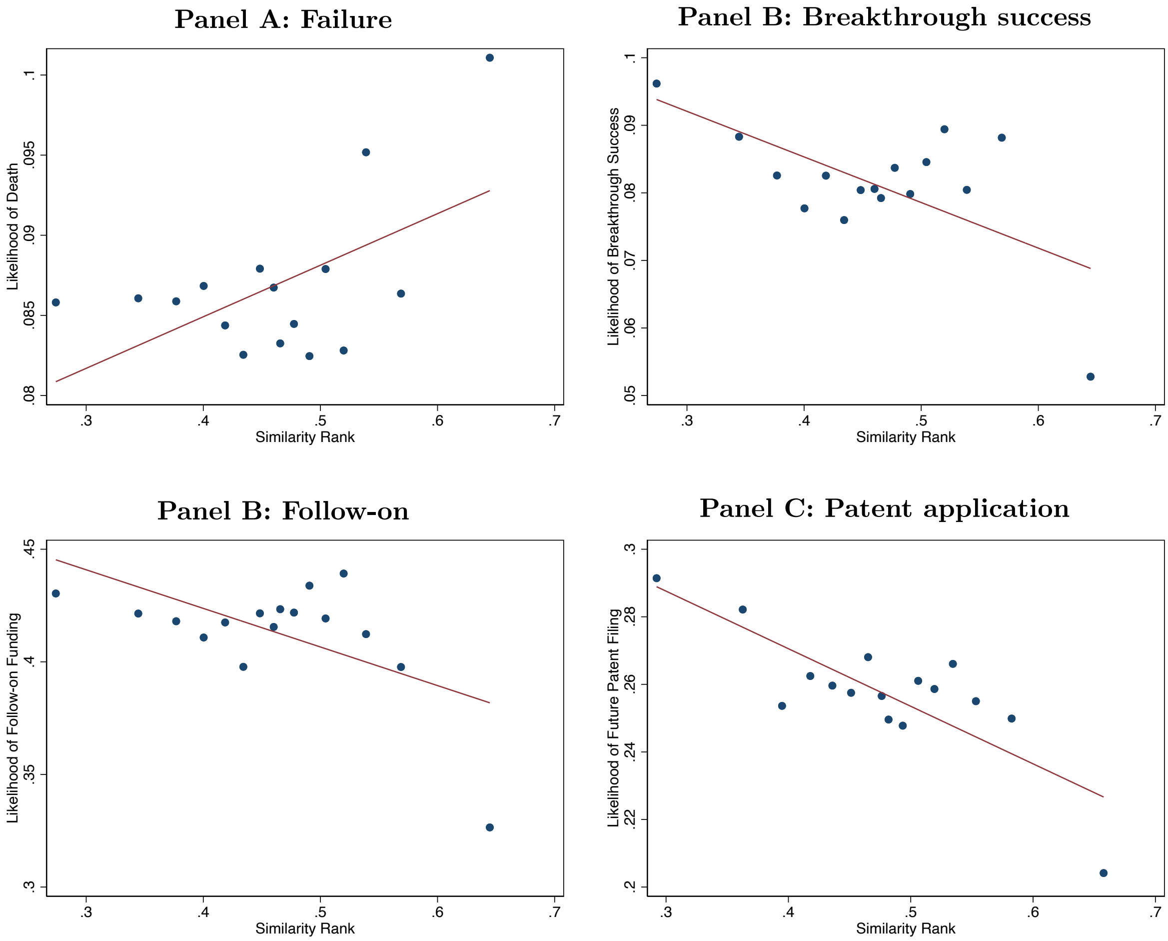


Figure 5: **The relation between a startup's backward-similarity and its future success**: These figures show bin-scatter plots that relate the backward-similarity of a startup to its future success. For each VC-funded startup, Panel A plots a dummy indicating whether the startup is dead by 2021 as a function of the startup's backward-similarity percentile rank when raising its first round of funding. Panels B, C and D do the same with dummies indicating whether the startup eventually achieves breakthrough success (IPO or profitable acquisition), survives and raises follow-on funding, and files a patent application after the funding round respectively. All variables are first regressed on industry-country-year fixed effects to absorb factors that could make startups appear more or less similar.

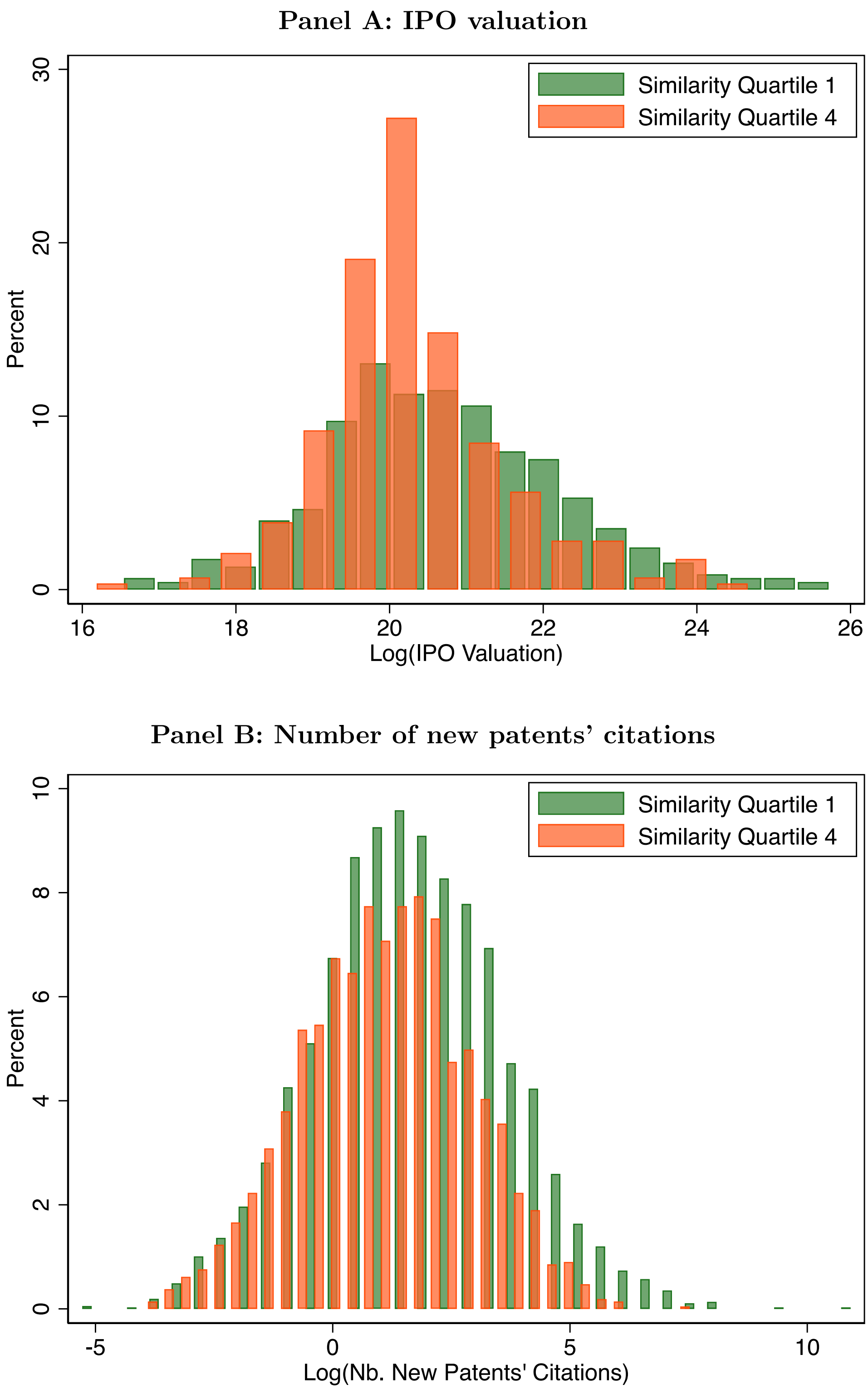
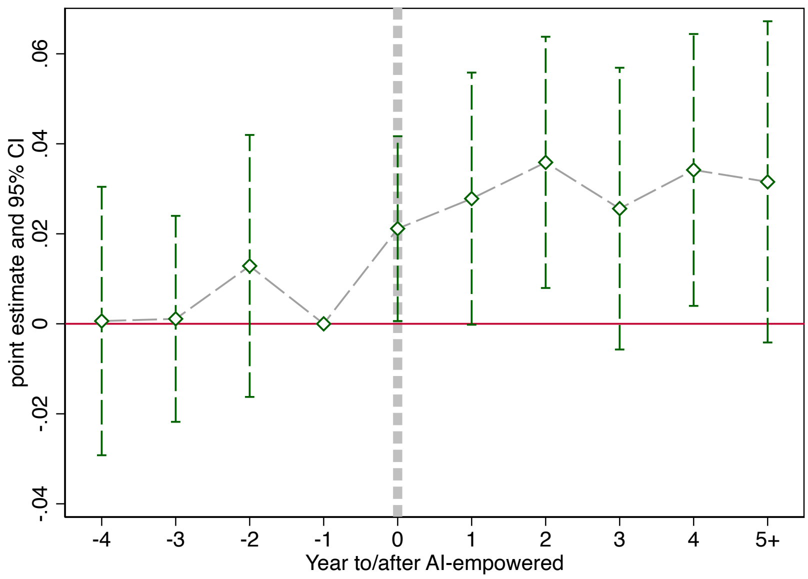


Figure 6: **Distribution of startup's outcomes conditional on success depending on backward-similarity**: These figures plot the distribution of the logarithm of IPO valuations among startups that achieve a successful exit (Panel A), and of the logarithm of the number of new patents' (adjusted) citations among startups that manage to get a patent after the initial VC investment (Panel B). Each panel shows the distribution separately for startups in the top and bottom quartile of backward-similarity when raising their first round of funding.

Panel A: Likelihood of survival and follow-on funding



Panel B: Likelihood of breakthrough success

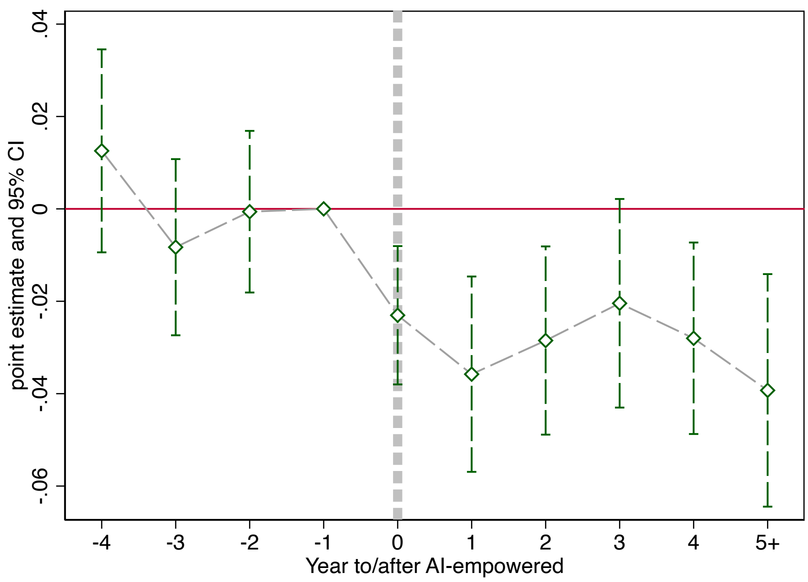


Figure 7: **Likelihood that investment leads to follow-on funding or breakthrough success before/after VCs adopt AI**: These figures plot the estimated coefficients from regressions at the VC-investment level for the terms corresponding to each year relative to AI adoption. In Panel A, the dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In Panel B, the dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. The year preceding AI adoption (-1) is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and startup level. Regressions include VC firm fixed effects and startup's industry-country-funding year-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.

Likelihood of survival and follow-on funding in each quartile of backward-similarity

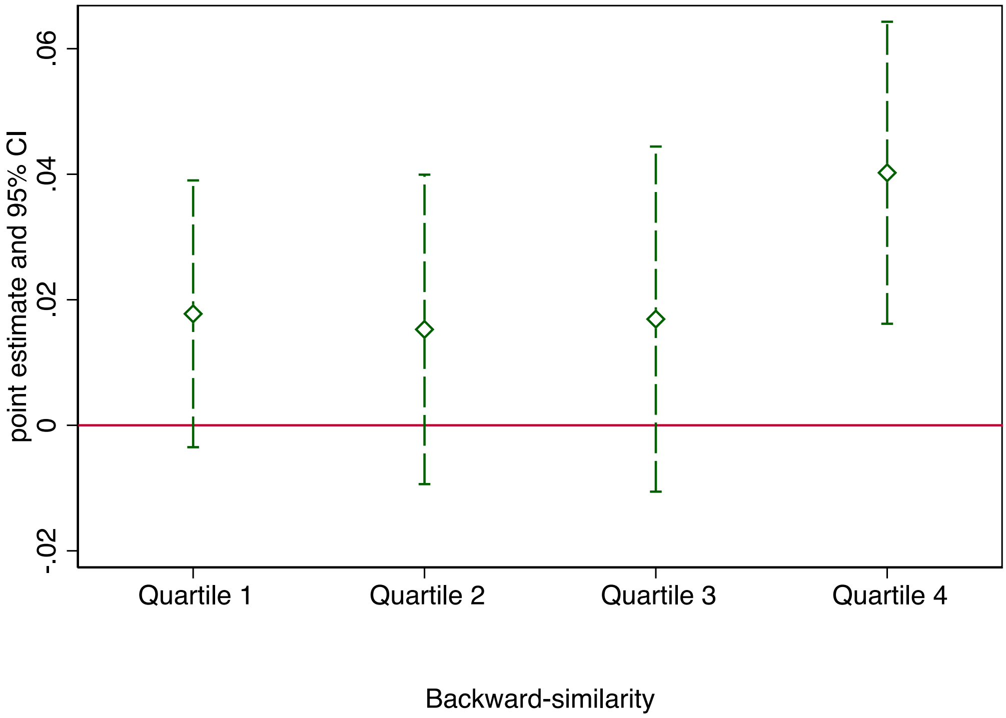
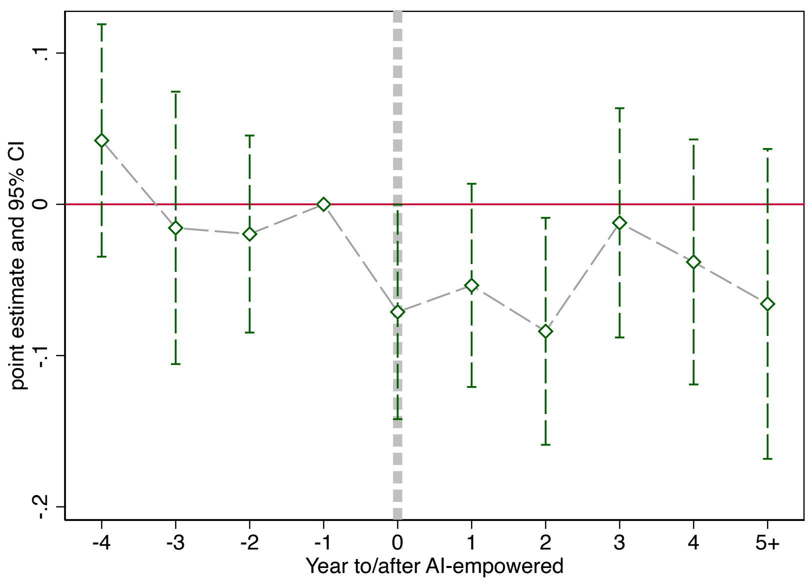


Figure 8: **Likelihood of follow-on funding after VCs adopt AI depending on the startup's backward-similarity**: This figure plots the estimated coefficients in column (4) of Table 5 from regressions at the VC-investment level. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. Each point represents a different coefficient on the interaction of a dummy indicating whether the VC is AI-empowered as of the investment date and the startup's backward-similarity quartile. For instance "Quartile 1" refers to the sub-sample of investments in funding rounds of startups with backward-similarity percentile rank below 0.25. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and startup level. Regressions include VC firm fixed effects and startup's industry-country-funding year-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.

Panel A: Number of new patent applications



Panel B: Number of new highly-cited patents

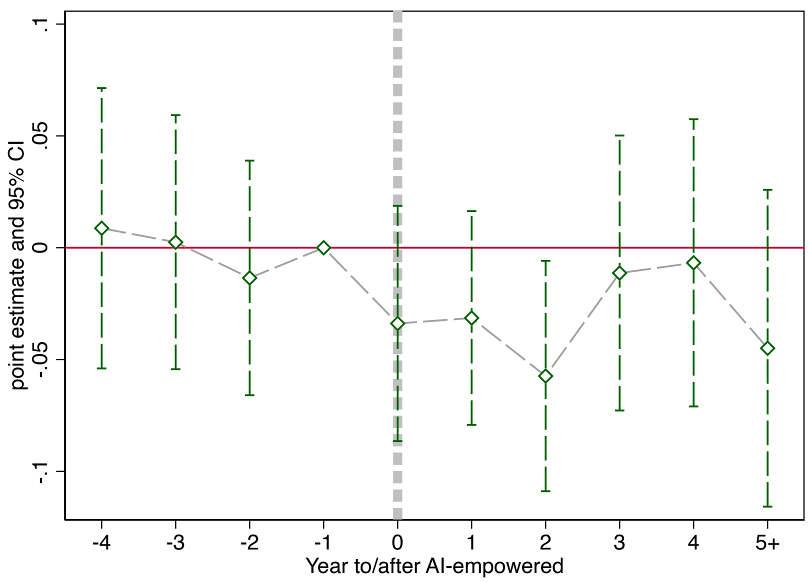


Figure 9: **Number of VC-funded startup's new patent applications and highly-cited patents after VCs adopt AI**: These figures plot the estimated coefficients from regressions at the VC-investment level for the terms corresponding to each year relative to AI adoption. In Panel A, the dependent variable is the logarithm of one plus the number of patent applications filed by the startup after the VC invested. In Panel B, the dependent variable is the logarithm of one plus the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. The year preceding AI adoption (-1) is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors double-clustered at the VC firm and startup level. Regressions include VC firm fixed effects and startup's industry-country-funding year-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.

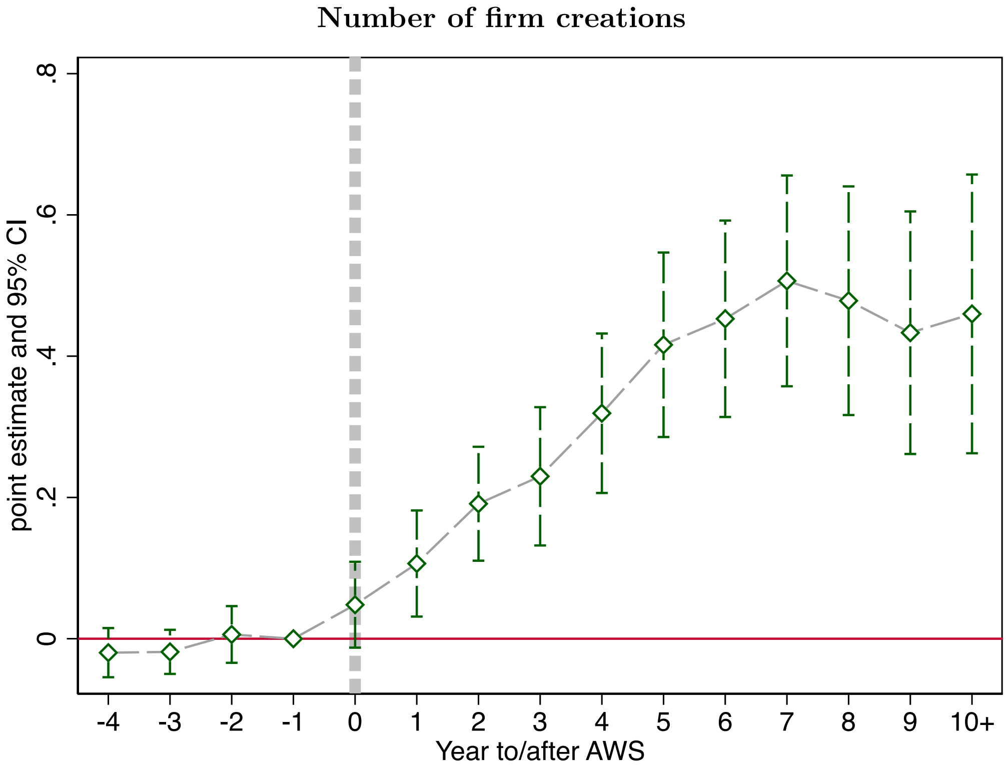


Figure 10: **The effects of the introduction of cloud computing on firm creations**: This figure plots the estimated coefficients from difference-in-differences regressions at the industry-country-year level, for the interaction terms of each year relative to 2006 and the industry exposure to the introduction of cloud computing (Amazon Web Services). The dependent variable is the logarithm of (one plus) the number of new firm creations in the country-industry-year. The 2005 interaction term is the excluded category (-1), reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors clustered at the industry and country levels. The regression includes industry fixed effects and country-year fixed effects. Additional details on the specification can be found in Appendix E.

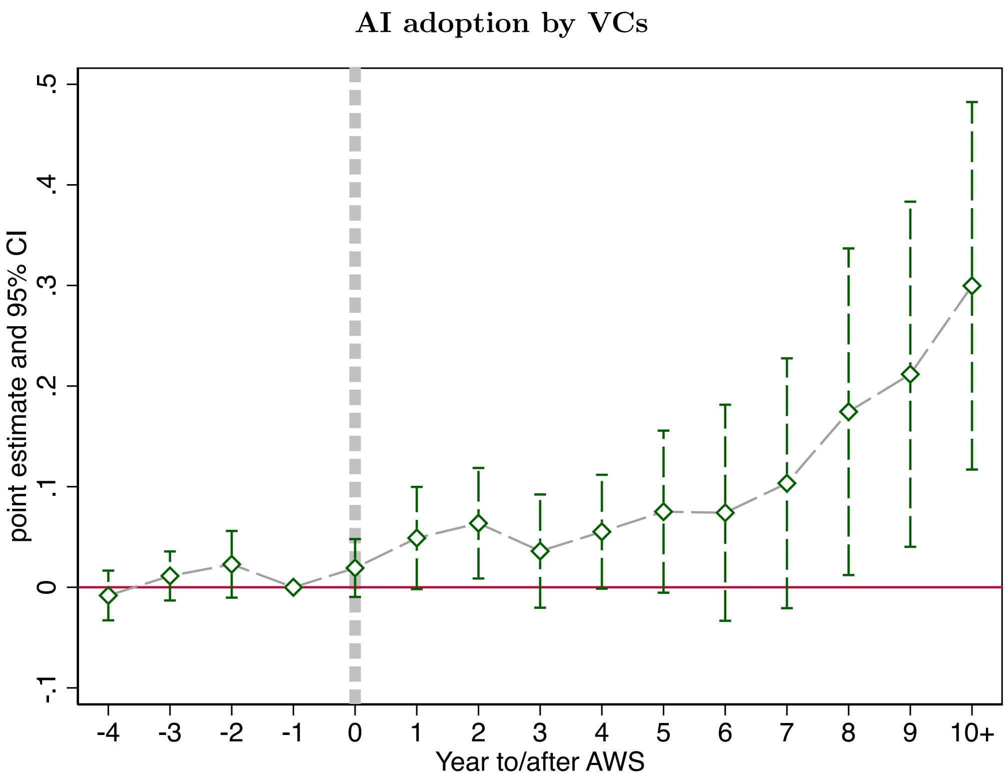
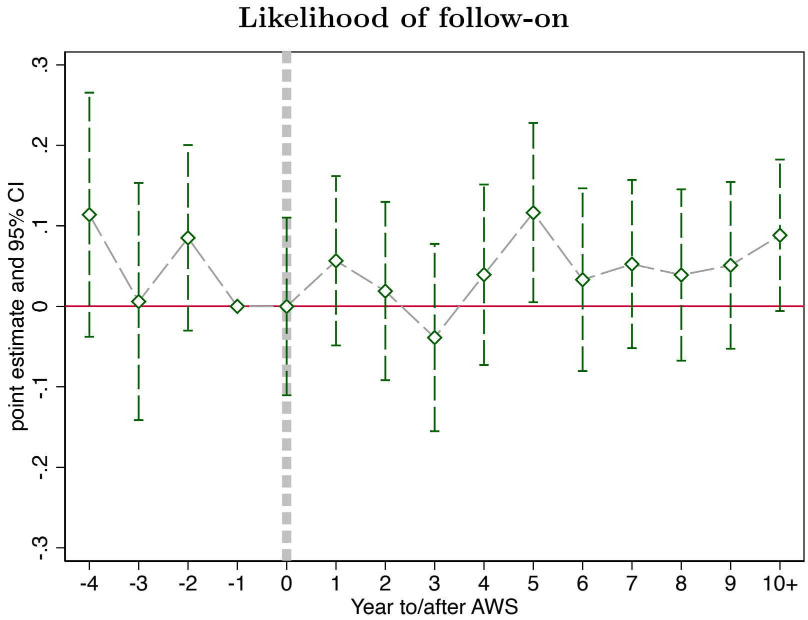


Figure 11: **The effect of VCs' exposure to cloud-related industries on their likelihood to adopt AI**: This figure plots the estimated coefficients from difference-in-differences regressions at the VC-investment level, for the interaction terms of each year relative to 2006 and the VC exposure to the introduction of cloud computing (the VC firm's exposure based on its portfolio composition before the shock). The dependent variable is a dummy indicating whether the investment is made by a VC classified as AI-empowered as of the investment data. The 2005 interaction term is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors clustered at the VC and startup levels. Regressions include VC firm fixed effects and startup's industry-country-funding year-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.



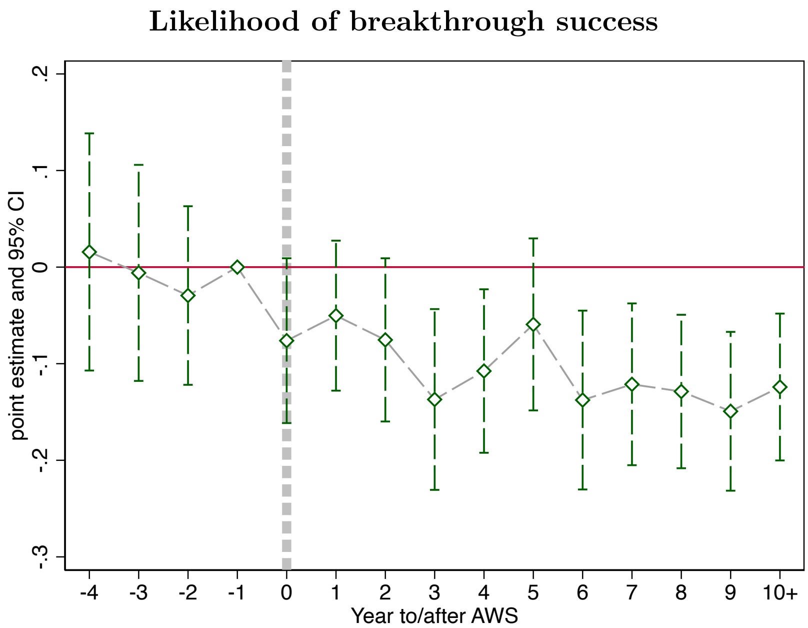


Figure 12: **The effects of VCs' exposure to cloud-related industries on their investment outcomes**: These figures plot the estimated coefficients from difference-in-differences regressions at the VC investment level, for the interaction terms of each year relative to 2006 and the VC exposure to the introduction of cloud computing (the VC firm's exposure based on its portfolio composition before the shock). In Panel A, the dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In Panel B, the dependent variable is a dummy indicating whether the VC-funded company goes public in IPO or is acquired for a higher value than the total VC investments in the company. The 2005 interaction term is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors clustered at the VC and startup levels. Regressions include VC firm fixed effects and startup's industry-country-funding year-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.

Tables

Variable	Obs	Mean	Sd	5%	25%	50%	75%	95%
Panel A: Investment level								
VC Age (Years)	221,765	13.84	15.18	2.00	5.00	9.00	18.00	39.00
VC Nb. Employees	221,765	10.31	17.87	1.00	2.00	5.00	11.00	35.00
Company Age (Years)	221,765	5.30	4.66	1.00	2.00	4.00	7.00	14.00
Early Stage	221,765	0.48	0.50	0.00	0.00	0.00	1.00	1.00
First-round	221,765	0.29	0.46	0.00	0.00	0.00	1.00	1.00
Raised Amount (USD mln)	205,005	24.41	119.16	0.48	2.51	8.00	20.00	80.00
Nb. Investors	221,765	4.32	3.69	1.00	2.00	3.00	6.00	11.00
Lead	182,706	0.40	0.49	0.00	0.00	0.00	1.00	1.00
Out of VC Region(s)	221,765	0.46	0.50	0.00	0.00	0.00	1.00	1.00
VC on Board	152,380	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Serial Entrepreneur	217,129	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Top School Founder	217,129	0.37	0.48	0.00	0.00	0.00	1.00	1.00
Female Founder	217,129	0.24	0.43	0.00	0.00	0.00	0.00	1.00
Follow-on	221,765	0.52	0.50	0.00	0.00	1.00	1.00	1.00
Breakthrough Success	221,765	0.15	0.35	0.00	0.00	0.00	0.00	1.00
IPO	221,765	0.08	0.27	0.00	0.00	0.00	0.00	1.00
IPO or Acquired	221,765	0.34	0.47	0.00	0.00	0.00	1.00	1.00
Nb. New Patent Applications	221,765	8.06	26.79	0.00	0.00	0.00	2.00	45.00
Nb. New Patent Grants	221,765	5.34	18.41	0.00	0.00	0.00	1.00	30.00
Nb. New Highly-cited Patents	221,765	2.23	8.52	0.00	0.00	0.00	0.00	12.00
Nb. New Patents' Citations	221,765	6.49	25.12	0.00	0.00	0.00	0.32	33.98
AI-empowered	221,765	0.08	0.27	0.00	0.00	0.00	0.00	1.00
Similarity Rank	221,765	0.49	0.29	0.05	0.23	0.48	0.74	0.94
VC Exposure	108,092	0.57	0.24	0.13	0.40	0.60	0.76	0.91
Panel B: VC-Year level								
US VC Firm	58,377	0.53	0.50	0.00	0.00	1.00	1.00	1.00
European VC Firm	58,377	0.26	0.44	0.00	0.00	0.00	1.00	1.00
VC Age (Years)	58,377	13.38	16.83	2.00	4.00	9.00	17.00	36.00
VC Nb. Employees	58,377	4.94	9.52	1.00	1.00	2.00	5.00	16.00
Nb. Investments	58,377	4.24	7.96	0.00	1.00	2.00	5.00	16.00
Nb. Leads	58,377	1.25	2.87	0.00	0.00	0.00	1.00	5.00
Proportion in Software	45,901	0.43	0.37	0.00	0.00	0.43	0.72	1.00
Proportion in Heath Care	45,901	0.22	0.35	0.00	0.00	0.00	0.33	1.00
Proportion in Financial Services	45,901	0.11	0.24	0.00	0.00	0.00	0.11	0.75

Continued next page

Summary statistics (continued)

Variable	Obs	Mean	Sd	5%	25%	50%	75%	95%
Proportion in Information Technology	45,901	0.20	0.29	0.00	0.00	0.00	0.33	1.00
Proportion in Internet Services	45,901	0.20	0.28	0.00	0.00	0.00	0.33	1.00
Proportion in AI and Data	45,901	0.15	0.25	0.00	0.00	0.00	0.25	0.75
Nb. Funds	31,837	3.73	5.38	1.00	1.00	2.00	4.00	10.00
AUM (USD bln)	29,798	1.28	5.30	0.01	0.07	0.21	0.74	4.19
AI-empowered VC	58,377	0.02	0.14	0.00	0.00	0.00	0.00	0.00
Nb. Investments Similarity Quartile 1	58,377	1.05	2.48	0.00	0.00	0.00	1.00	5.00
Nb. Investments Similarity Quartile 2	58,377	1.02	2.15	0.00	0.00	0.00	1.00	4.00
Nb. Investments Similarity Quartile 3	58,377	1.02	2.33	0.00	0.00	0.00	1.00	4.00
Nb. Investments Similarity Quartile 4	58,377	1.14	2.90	0.00	0.00	0.00	1.00	5.00
VC Exposure	26,096	0.50	0.28	0.05	0.26	0.50	0.75	0.94
Panel C: Fund level								
AI-empowered VC Fund	9,624	0.03	0.18	0.00	0.00	0.00	0.00	0.00
VC Age (Years)	9,624	8.72	11.05	1.00	2.00	6.00	11.00	27.00
Sequence	9,624	1.99	2.01	1.00	1.00	1.00	2.00	6.00
Fund Size (USD mln)	9,624	123.18	245.55	4.10	18.16	48.98	126.00	480.00
Nb. Investments	9,624	17.51	28.78	1.00	3.00	8.00	20.00	64.00
Net IRR (%)	1,270	13.19	23.86	-13.30	1.23	10.20	22.00	47.32
Net Multiple	1,690	1.73	2.39	0.41	0.96	1.31	1.92	4.08

Table 1: **Summary statistics**. This table presents descriptive statistics of the main variables used in the paper. Panels A, B and C present statistics at the VC investment level, VC-year level and VC fund level respectively.

	Log(1 + Nb. Investments)		Log(Nb. Funds)		Log(AUM)	
	(1)	(2)	(3)	(4)	(5)	(6)
AI-empowered	0.388***	0.249***	0.340***	0.276***	0.398***	0.285***
	(0.072)	(0.063)	(0.068)	(0.057)	(0.096)	(0.078)
Control Variables	No	Yes	No	Yes	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58,377	58,377	31,837	31,837	29,798	29,798
R^2	0.70	0.71	0.91	0.92	0.95	0.96

Table 2: **Number of investments and funds after VCs adopt AI**: This table reports results for regressions at the VC-year level, investigating whether AI-empowered VCs make more investments per year after they adopt AI than do other VCs. The dependent variables are the logarithm of one plus the number of investments in the year (columns 1 and 2), the (cumulative) number of funds as of the year (columns 3 and 4), and the (cumulative) AUM as of the year (columns 5 and 6). The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Log(1 + Nb. Investments) with Backward-similarity in							
	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	0.043	-0.002	0.244***	0.183***	0.460***	0.374***	0.787***	0.664***
	(0.065)	(0.062)	(0.061)	(0.057)	(0.092)	(0.088)	(0.116)	(0.111)
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58,377	58,377	58,377	58,377	58,377	58,377	58,377	58,377
R^2	0.60	0.61	0.56	0.56	0.57	0.57	0.64	0.65

Table 3: **Number of investments in each quartile of backward-similarity after VCs adopt AI**: This table reports results for regressions at the VC-year level, investigating whether after they adopt AI, AI-empowered VCs make more investments in companies that are more backward-similar than do other VCs. In columns 1 to 8, the dependent variable is the logarithm of (one plus) the number of investments made by the VC firm in that year in startups with backward-similarity in a given quartile. For instance, 'Quartile 1" refers to the number of investments in funding rounds of startups with backward-similarity percentile rank below 0.25. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Similarity Rank (× 100)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Has Filed Patent Application	-0.520***												
	(0.141)												
New Patent Application	-1.051***												
	(0.209)												
Log(1+Nb. Past Patent Applications)		-0.090											
		(0.090)											
Log(1+Nb. New Patent Applications)		-0.461***											
		(0.122)											
Breakthrough Success			-0.940**		-0.863**								
			(0.403)		(0.381)								
Follow-on				-0.721***	-0.692***								
				(0.209)	(0.197)								
Log(Current Nb. Employees)						-0.104							
						(0.068)							
Log(Raised Amount)							-0.063						
							(0.117)						
Serial								0.022					
								(0.099)					
Top School								-0.235***					
								(0.049)					
PhD								-0.024					
								(0.198)					
MBA								-0.179*					
								(0.090)					
Female								-0.391***					
								(0.091)					
US Startup									-2.581***				
									(0.958)				
European Startup									-2.041**				
									(0.934)				
Top-5 Region										-0.242*			
										(0.128)			
Log(Company Age)											-0.156	0.562***	
											(0.096)	(0.104)	
Log(#Funding Round)													0.736***
													(0.076)
Industry × Year FE	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Industry × Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Company FE	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Observations	69,671	64,520	64,520	64,520	64,520	62,630	45,934	52,969	69,671	69,671	69,671	130,557	130,557
R^2	0.90	0.89	0.89	0.89	0.89	0.90	0.91	0.91	0.86	0.90	0.90	1.00	1.00

Table 4: **The relation between backward-similarity and startups' characteristics, success and innovation**: This table reports results for regressions at the funding round level, investigating the link between my backward-similarity measure and a set of startup characteristics. The dependent variable is the percentile rank in terms of backward-similarity. All regressions include industry-country-year fixed effects, except column (9) that includes industry-year fixed effects. To avoid double counting, when a startup raises several funding rounds, I only consider the first round, except in columns (12) and (13). These columns also includes company fixed effects. The number of observations is lower in columns (6), (7) and (8) because these regressions are estimated only for companies with available data on raised amount, current employee count and founding team members, respectively. Top-5 Region in column (10) is a dummy equal to one if the startup is headquartered in one of the following regions: California, New York, Massachusetts, England or Beijing. Standard errors in parentheses are clustered at the industry and country levels.

	Follow-on					
	(1)	(2)	(3)	(4)	(5)	(6)
AI-empowered	0.030*** (0.010)	0.026** (0.010)		0.018 (0.011)		
AI-empowered × Similarity Quartile 1			0.018 (0.011)			
AI-empowered × Similarity Quartile 2			0.015 (0.013)	-0.002 (0.011)	0.002 (0.012)	
AI-empowered × Similarity Quartile 3			0.017 (0.014)	-0.001 (0.012)	0.003 (0.013)	
AI-empowered × Similarity Quartile 4			0.040*** (0.012)	0.022* (0.012)	0.021* (0.012)	
AI-empowered × Similarity Rank						0.027* (0.015)
Similarity Rank						-0.102*** (0.030)
Control variables	No	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	No	Yes	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	No	No
VC × Year FE	No	No	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.45	0.46	0.46	0.46	0.52	0.52

Table 5: **Likelihood that investments lead to follow-on funding after VCs adopt AI**: This table reports results for regressions at the VC-investment level, investigating whether the investments made by AI-empowered VCs after they adopt AI lead to different outcomes than those made by other VCs. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. 'Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Breakthrough Success					
	(1)	(2)	(3)	(4)	(5)	(6)
AI-empowered	-0.027*** (0.008)	-0.027*** (0.008)		-0.024** (0.010)		
AI-empowered × Similarity Quartile 1			-0.024** (0.010)			
AI-empowered × Similarity Quartile 2			-0.034*** (0.009)	-0.010 (0.010)	-0.011 (0.010)	
AI-empowered × Similarity Quartile 3			-0.023** (0.010)	0.001 (0.009)	-0.001 (0.009)	
AI-empowered × Similarity Quartile 4			-0.028*** (0.010)	-0.003 (0.009)	-0.002 (0.009)	
AI-empowered × Similarity Rank						0.006 (0.012)
Similarity Rank						-0.060** (0.028)
Control variables	No	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	No	Yes	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	No	No
VC × Year FE	No	No	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.54	0.54	0.54	0.54	0.60	0.60

Table 6: **Likelihood that investments lead to breakthrough success after VCs adopt AI**: This table reports results for regressions at the VC-investment level, investigating whether the investments made by AI-empowered VCs after they adopt AI lead to different outcomes than those made by other VCs. The dependent variable is a dummy indicating whether the VC-funded company goes public in IPO or is acquired for a higher value than the total VC investments in the company. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. 'Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Log(1+Nb. New Patent Applications)				Log(1+Nb. New Highly-cited Patents)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	-0.057**				-0.034*			
	(0.025)				(0.020)			
AI-empowered × Similarity Quartile 1		-0.037				-0.027		
		(0.034)				(0.024)		
AI-empowered × Similarity Quartile 2		-0.060**	-0.018			-0.026	0.008	
		(0.027)	(0.026)			(0.022)	(0.018)	
AI-empowered × Similarity Quartile 3		-0.067**	-0.023			-0.023	0.011	
		(0.034)	(0.026)			(0.023)	(0.016)	
AI-empowered × Similarity Quartile 4		-0.058*	-0.020			-0.049**	-0.015	
		(0.033)	(0.036)			(0.024)	(0.021)	
AI-empowered × Similarity Rank				-0.024				-0.021
				(0.042)				(0.027)
Similarity Rank				-0.352***				-0.267***
				(0.090)				(0.065)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	Yes	No	No	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	No	No	Yes	Yes	No	No
VC × Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.60	0.60	0.65	0.65	0.57	0.57	0.62	0.62

Table 7: **Innovation of VC-funded companies after VCs adopt AI**: This table reports results for regressions at the VC-investment level, investigating whether the investments made by AI-empowered VCs after they adopt AI lead to different outcomes than those made by traditional VCs. In columns (1)-(4), the dependent variable is the logarithm of one plus the number of patent applications filed by the startup after the VC invested. In columns (5)-(8), the dependent variable is the logarithm of one plus the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. 'Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	AI-empowered			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.138** (0.054)	0.141*** (0.052)		0.104** (0.044)
VC Exposure × Post × Similarity Quartile 1			0.104** (0.044)	
VC Exposure × Post × Similarity Quartile 2			0.109** (0.045)	0.005 (0.012)
VC Exposure × Post × Similarity Quartile 3			0.101*** (0.035)	-0.003 (0.023)
VC Exposure × Post × Similarity Quartile 4			0.214*** (0.074)	0.109*** (0.042)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.79	0.79	0.79	0.79

Table 8: **The effect of VCs' exposure to cloud-related industries on their likelihood to adopt AI**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy variable indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * $p<.10$; ** $p<.05$; *** $p<.01$.

	Follow-on			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.031 (0.031)	0.023 (0.031)		0.014 (0.037)
VC Exposure × Post × Similarity Quartile 1			0.014 (0.037)	
VC Exposure × Post × Similarity Quartile 2			-0.044 (0.051)	-0.058 (0.061)
VC Exposure × Post × Similarity Quartile 3			0.028 (0.109)	0.014 (0.113)
VC Exposure × Post × Similarity Quartile 4			0.086** (0.042)	0.072** (0.029)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.46	0.47	0.47	0.47

Table 9: **The effect of VCs' exposure to cloud-related industries on the likelihood that their investments lead to follow-on funding**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Breakthrough Success			
	(1)	(2)	(3)	(4)
VC Exposure × Post	-0.085***	-0.086***		-0.071**
	(0.027)	(0.027)		(0.033)
VC Exposure × Post × Similarity Quartile 1			-0.071**	
			(0.033)	
VC Exposure × Post × Similarity Quartile 2			-0.120***	-0.049
			(0.044)	(0.053)
VC Exposure × Post × Similarity Quartile 3			0.002	0.073
			(0.104)	(0.108)
VC Exposure × Post × Similarity Quartile 4			-0.105***	-0.035
			(0.038)	(0.026)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.52	0.52	0.52	0.52

Table 10: **The effect of VCs' exposure to cloud-related industries on the likelihood that their investments lead to breakthrough success**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Log(1+Nb. New Patent Applications)		Log(1+Nb. New Highly-cited Patents)	
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.024		0.047	
	(0.093)		(0.072)	
VC Exposure × Post × Similarity Quartile 1		-0.096		-0.079
		(0.121)		(0.090)
VC Exposure × Post × Similarity Quartile 2		-0.004		0.067
		(0.144)		(0.115)
VC Exposure × Post × Similarity Quartile 3		0.424		0.345
		(0.386)		(0.295)
VC Exposure × Post × Similarity Quartile 4		-0.130		-0.061
		(0.137)		(0.111)
Control Variables	Yes	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	Yes	No	Yes
Post × Similarity Quartile	No	Yes	No	Yes
Similarity Quartile	No	Yes	No	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.58	0.58	0.55	0.55

Table 11: **The effect of VCs' exposure to cloud-related industries on the innovation of the startups they select**: This table reports results for regressions at the VC-investment level. In columns (1)-(2), the dependent variable is the logarithm of one plus the number of patent applications filed by the startup after the VC invested. In columns (3)-(4), the dependent variable is the logarithm of one plus the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the similarity rank quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

ONLINE APPENDIX

ONLINE APPENDIX

The Adoption of Artificial Intelligence by Venture Capitalists

Table of Contents

A Data

Database Crunchbase was launched in July 2007 and has a global coverage of more than 1.8 million private and public companies, more than 210,000 investors and more than 730,000 investments in private companies, as of 2022. Most data are collected via Crunchbase's inhouse data team, the remaining component is crowdsourced. Crunchbase also allows users to register and to submit information that is subject to validation and reviewed by Crunchbase's staff. In particular, Crunchbase receives monthly portfolio updates from more than 3,500 global investment firms. Crunchbase is also synchronized with AngelList and, thus, has access to information regarding startups seeking capital but which have not necessarily raised funds. Crunchbase also covers closed companies and provides closing dates.

Although no data source offers a complete coverage of all venture investments, the accuracy of Crunchbase's data has been highlighted by several studies. For instance, Retterath and Braun (2020) compare frequently used VC databases and show, using verified proprietary information, that Crunchbase is among those having the best coverage and accuracy across several dimensions, such as general company information, founders and funding information. Ling (2015) also manually compares Crunchbase data for a subsample of startups with the information from major databases, and finds that the information on the funding rounds, IPO and acquisitions are similar for Crunchbase and other sources. Finally, Breschi et al. (2018) document that aggregate statistics from Crunchbase on VC funding by country and year are similar to the same figures produced with the OECD Entrepreneurship Financing Database. An increasing number of academic studies rely on Crunchbase: Dalle et al. (2017) document more than 80 academic papers in the field of economics based on this database.

Figure A.1 shows the number of existing companies on Crunchbase according to their year of foundation over the period 2000-2020.

Information on jobs and individuals Crunchbase provides extensive information on jobs and employees at companies listed in their database. Such information on jobs and individuals come from multiples sources. The latter include data partners and other companies licensing information to Crunchbase, publicly available sources that are scanned, and users who can create and update their profiles. Users can voluntarily disclose some information by creating and updating their profiles. To do so, they have to register in order to make edits. To ensure the legitimacy of user contributions, Crunchbase ask them to authenticate their user accounts via Facebook, Twitter, or LinkedIn before they can make any edits. Information is then reviewed by Crunchbase's staff. As a result, Crunchbase's data on employees do not only include information on top executives and partners but also encompass other staff members such as analysts, engineers and back office professionals.

Founding team members To gather information on founding team members, I collect the complete history of jobs associated with each VC-funded startup from Crunchbase. I classify as founding team members the people with job titles including one of the following keywords: "founder", "founding", "president", "chairman", "owner", "ceo", "chief executive", "cto", "chief technology", "chief technical", "chief software", "chief architect", "cmo", "chief marketing", "coo", "chief operating", "cso", "chief scientist", "chief scientific", "cpo", "chief product", "managing partner", "proprietor", "executive partner".

Top schools Crunchbase provides education records for people listed in the database. I follow Gompers et al. (2016) and classify as top school the following institutions: Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, Princeton University, University of Pennsylvania, Yale University, Amherst College, California Institute of Technology, Duke University, MIT, Northwestern University, Stanford University, University of California, Berkeley, University of Chicago, Williams College, Cambridge University, INSEAD, HEC Paris, London Business School, London School of Economics and Oxford University.

Panel A: Overall

Industry Group	Nb. Rounds	% Rounds	Nb. Investments	% Investments
Software	13,988	10.74	27,557	11.16
Health Care	13,512	10.37	27,056	10.96
Financial Services	8,107	6.22	16,171	6.55
Information Technology	7,588	5.83	14,282	5.78
Internet Services	6,905	5.30	12,964	5.25
Commerce and Shopping	6,669	5.12	12,315	4.99
Hardware	5,579	4.28	11,552	4.68
Biotechnology	4,630	3.55	10,343	4.19
Other	4,333	3.33	8,040	3.26
Data and Analytics	4,131	3.17	8,465	3.43

Panel B: AI-empowered VCs

Industry Group	Nb. Rounds	% Rounds	Nb. Investments	% Investments
Software	2,310	13.77	2,820	13.97
Financial Services	1,321	7.87	1,654	8.20
Health Care	1,189	7.08	1,377	6.82
Information Technology	1,125	6.70	1,369	6.78
Internet Services	963	5.74	1,154	5.72
Commerce and Shopping	941	5.61	1,118	5.54
Data and Analytics	725	4.32	896	4.44
Hardware	677	4.03	786	3.90
Artificial Intelligence	592	3.52	751	3.72
Other	569	3.39	689	3.41

Table A.1: **Main industry groups**. This table shows the top 10 industry groups in terms of number and fraction of funding rounds, in the overall sample (Panel A) and among AI-empowered VCs' investments only (Panel B). The table also reports the number and fraction of investments. Crunchbase categorizes companies into 47 distinct industry groups. Industry groups are broader subjects that encompass multiple industries.

Panel A: Overall				
Country	Nb. Rounds	% Rounds	Nb. Investments	% Investments
USA	67,064	51.49	149,349	60.48
CHN	10,323	7.93	15,463	6.26
GBR	9,888	7.59	15,212	6.16
IND	5,061	3.89	7,578	3.07
DEU	4,146	3.18	7,088	2.87
CAN	3,958	3.04	6,885	2.79
FRA	3,940	3.03	6,177	2.50
JPN	3,270	2.51	4,871	1.97
ISR	2,120	1.63	4,039	1.64
ESP	1,526	1.17	2,334	0.95
Panel B: AI-empowered VCs				
Country	Nb. Rounds	% Rounds	Nb. Investments	% Investments
USA	12,025	71.65	14,860	73.63
DEU	1,129	6.73	1,237	6.13
GBR	837	4.99	995	4.93
FRA	446	2.66	488	2.42
IND	412	2.45	439	2.18
CAN	391	2.33	461	2.28
ISR	223	1.33	245	1.21
CHN	119	0.71	123	0.61
ESP	89	0.53	96	0.48
SWE	80	0.48	100	0.50

Table A.2: **Main countries**. This table shows the top 10 countries in terms of number of funding rounds, in the overall sample (Panel A) and among AI-empowered VCs' investments only (Panel B). The table also reports the number of investments.

Panel A: Overall				
Industry	Nb. Rounds	% Rounds	Nb. Investments	% Investments
Software	8,146	6.25	15,316	6.20
Health Care	5,383	4.13	10,369	4.20
Information Technology	4,021	3.09	7,032	2.85
Biotechnology	3,428	2.63	7,564	3.06
E-Commerce	3,313	2.54	6,027	2.44
Internet	3,215	2.47	6,039	2.45
Mobile	2,369	1.82	4,623	1.87
Manufacturing	2,343	1.80	3,832	1.55
Financial Services	2,181	1.67	4,083	1.65
Medical	2,135	1.64	4,295	1.74
Panel B: AI-empowered VCs				
Industry	Nb. Rounds	% Rounds	Nb. Investments	% Investments
Software	1,198	7.14	1,441	7.14
Information Technology	485	2.89	571	2.83
Health Care	465	2.77	533	2.64
Internet	447	2.66	533	2.64
E-Commerce	444	2.64	520	2.58
SaaS	432	2.58	533	2.64
Enterprise Software	370	2.20	468	2.32
Mobile	335	1.99	394	1.95
Artificial Intelligence	320	1.90	404	2.00
Financial Services	314	1.87	389	1.93

Table A.3: **Main industries**. This table shows the top 10 industries in terms of number and fraction of funding rounds, in the overall sample (Panel A) and among AI-empowered VCs' investments only (Panel B). The table also reports the number and fraction of investments. Crunchbase categorizes companies into 744 distinct industries to account for specific market segments.

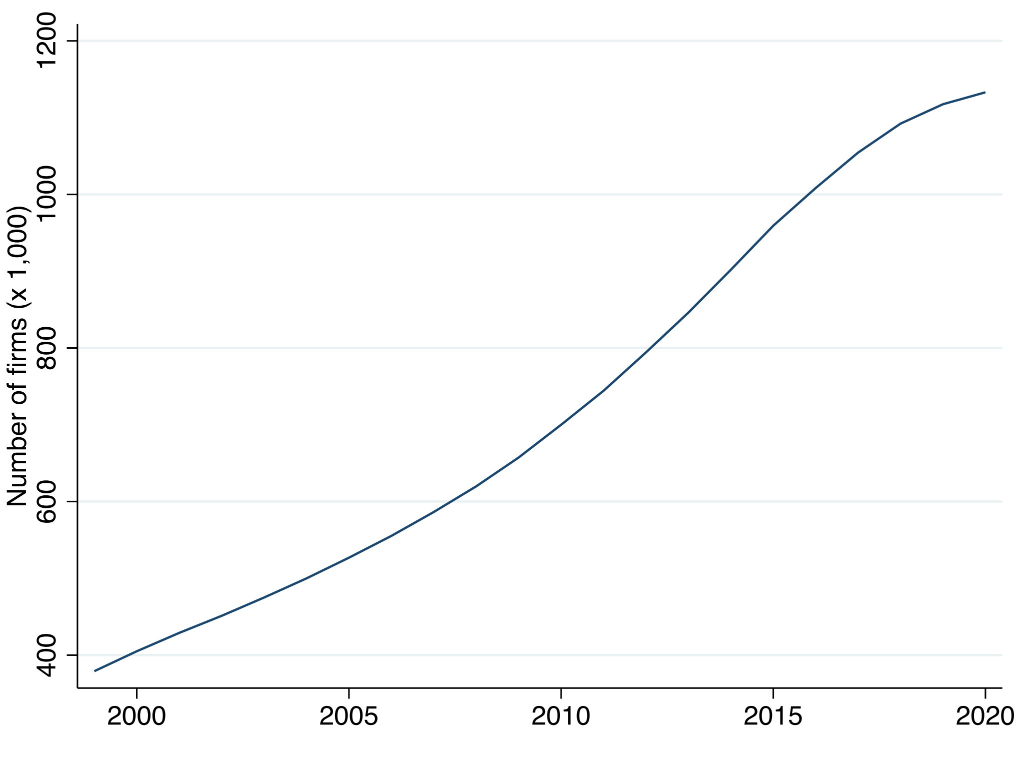


Figure A.1: **Number of companies in the Crunchbase database according to their year of foundation**: This figure shows the number of existing firms in the Crunchbase database according to companies' year of foundation. Each year, the figure corresponds to the number of firms that were founded as of the focal year, according to the Crunchbase database.

B Backward-similarity measure

In this section, I describe in more detail how I construct and validate my measure of backward-similarity.

Methodology I start by collecting from Crunchbase the business descriptions of all companies that ever raise capital. I only keep descriptions that are available in English language.[53] I "clean" each description following standard text cleaning procedures. I only keep the English words by removing numbers, symbols and special characters. I also remove the company's name and website url from its description. Finally, I remove the stop words and I stem each word to its root using the Porter stemmer algorithm (e.g. 'mathematic', 'mathematics', ... = 'mathemat'). I convert each "cleaned" description into a corresponding vector of word count. Cleaned business descriptions feature on average 45 words.

I then weight each term in a given word vector by their importance using a "term frequency inverse document frequency" (TFIDF) transformation of word counts. The purpose of this procedure is to appropriately weight words by their importance, with a view that common words used in many text descriptions are less informative. The weight given to each word w in the business description of company i raising capital in year t is defined as:

$$TFIDF_{w,i,t} = TF_{w,i} \times IDF_{w,i,t}. \tag{12}$$

The first component $TF_{w,i}$ in (12) is the count of how many times the term w appears in the business description i, divided by the total number of words in the description (term frequency). The second component $IDF_{w,i,t}$ is the inverse document frequency of the word w, defined as the logarithm of the inverse of the fraction of documents in the set of business descriptions that include the word w. The product of these two terms, $TFIDF_{w,i,t}$, describes the importance of a given word w in a given business description i. Words that appear infrequently in a description tend to have low TFIDF scores (due to low TF), as do common words that appear in many descriptions (due to low IDF). A high value of $TFIDF_{w,i,t}$ indicates that the word w appears relatively frequently in the description i but does not appear in most other descriptions.

For my purpose, I define $IDF_{w,i,t}$ as an industry-specific retrospective version of inverse document frequency. For each company i raising capital in year t, I consider the set of companies

[53] To detect whether a given description is written in English, I rely on Google's library Compact Language Detector 2, which can detect over 80 languages in plain text. In the whole Crunchbase database, 99.3% of business descriptions are detected to be written in English.

in the same industry raising capital between year t and year $t-5$. I denote this set $\mathcal{C}_{i,t}$. Then I compute $IDF_{w,i,t}$ as follows:

$$IDF_{w,i,t} = \log\left(\frac{\text{Nb. companies in } \mathcal{C}_{i,t}}{\text{Nb. company descriptions in } \mathcal{C}_{i,t} \text{ that include word } w}\right) \tag{13}$$

This frequency measure evolves as a term becomes more or less widely used over time, in a given industry. It reflects the history of company business descriptions up to, but not beyond, the startup i's fund raising year. This definition is similar to the "backward-IDF" developed by Kelly et al. (2021) for patent documents. Figure B.1 shows the ten words that are the most frequent in the business descriptions of companies in the software industry, raising capital in year 2008 (Panel A) and in year 2020 (Panel B). One can see for instance that the word "data" is the third most common word in 2020 and appears in 20% of descriptions, but is not among the top ten words in 2008.

The next step is to compute the pairwise similarity between company i raising capital in year t and the set of funded companies before t. To do so, I define the set $\mathcal{B}_{i,t}$, containing the description of companies that raised capital in the 5 calendar years preceding year t, in the same industry as company i. Unlike $\mathcal{C}_{i,t}$, $\mathcal{B}_{i,t}$ does not include year t to make sure that company i is compared to firms that raised capital previously. I choose a window of 5 years because existing computational constraints limit the window over which I can compare business descriptions. Let V_i and V_j be weighted term frequency vectors from business descriptions of company i and of another company in $\mathcal{B}_{i,t}$. I compute the (cosine) similarity of the two vectors as the normalized dot product of the term vectors:

$$\rho_{i,j} = \frac{V_i.V_j}{||V_i||\ ||V_j||}. \tag{14}$$

$\rho_{i,j}$ lies in the interval [0,1]. Companies that use the exact same set of words in their business description will have similarity of one, while firms with no overlapping terms have similarity of zero.

For each business description in the set $\mathcal{B}_{i,t}$, I compute the pairwise similarity with the business description of company i.[54] Finally, to create a measure of backward-similarity of company i at time t, I take:

$$BS_{i,t} = \sum_{j \in \mathcal{B}_{i,t}} \rho_{i,j}, \tag{15}$$

where $\rho_{i,j}$ is the pairwise similarity between companies i and j business descriptions defined in

[54] If company i has already raised capital in the past, it is not included in the set $\mathcal{B}_{i,t}$, in order to avoid comparing a company to itself.

(14) and $\mathcal{B}_{i,t}$ denotes the set of "prior" companies funded in the 5 calendar years prior to i's funding round. To remove any systematic effect due to the length of the business description, I regress the measure $BS_{i,t}$ on the description's number of words. I normalize the residuals and use them as the final backward-similarity measures for each company when raising funding.

Discussion of business descriptions in Crunchbase The set of business descriptions I obtain from Crunchbase comes from a single extract of their database and contains only one business description for each startup. One might be concerned that a startup's business description evolves over time because the company "pivots" and changes its business model. One possibility to overcome this issue could to use the "Wayback Machine" (an online platform offered by the Internet Archive) to download the description of each startup as of the time of funding. The Wayback Machine provides access to a digital library containing web-page snapshots occurring in history. Unfortunately, these snapshots are not taken at regular time for most startup webpages. Therefore, using the Wayback Machine would not allow me to compare business descriptions corresponding to the exact same time. This could create additional issues as certain terms become more or less widely used over time, in a given industry. Therefore, I rather compute backward-similarity using the set of business descriptions available in my extract of the Crunchbase's database.

Disruptive startups I conduct a validation check for my backward-similarity measure. I identify a list of disruptive startups and copies of existing businesses and examine how they score in terms of backward similarity. In subsection 4.2, I investigate in more detail how my measure relates to startups' innovation (i.e., patents) and to future success.

From online resources, I compile a list of disruptive startups and copies of existing businesses.[55] For each of these companies, I report their percentile rank in terms of backward-similarity (15) when raising capital for the first time in my sample. For instance, a value of 0.90 indicates that the company is in the top 10 percent of most backward-similar startups. Intuitively, startups that are considered as radical innovations should score low on backward-similarity, while startups that cloned existing ideas should score high on the measure. Table B.1 reports the similarity rank using both the unconditional distribution (column "No Adjustment"), and after subtracting the mean similarity measure within each Industry-Year (column "Removing Industry-Year FE"). The startups considered as disruptive

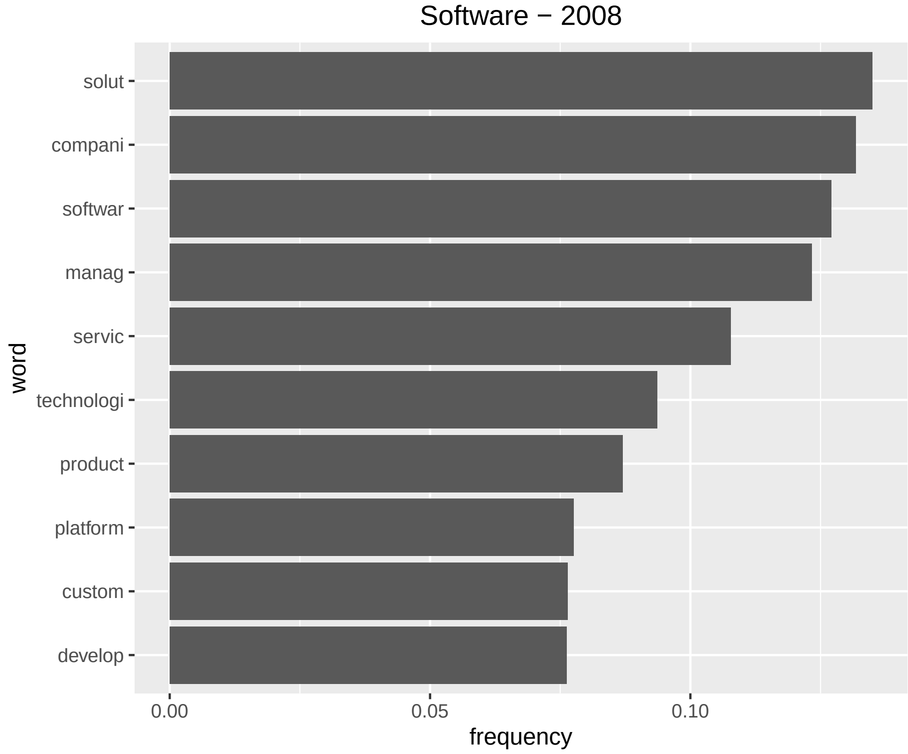
[55] These resources include grasshopper.com[...]disruptive-startups, and medium.com[...]startups-that-cloned-ideas.

are Netflix, Napster, Skype, Facebook, Digg, Youtube and Myspace. Overall, these companies score low in terms of backward-similarity rank (on average 0.09 without adjustment and 0.21 removing industry-year fixed effects). This suggests that these companies were truly different when they got funded. In contrast, Lazada Group, known to be the "copy of Amazon in Asia" has a high similarity rank: 0.68 and 0.72 with and without adjustment respectively.

In addition, I also compare the percentile ranks of companies that have developed the main online social networks we know today. Myspace and Facebook are considered as pioneers in this space, as they were among the first companies to develop this type of platforms. These two companies' ranks are virtually zero in terms of similarity without adjustment. In contrast, Instagram and Snap, which were created later and tried to improve upon Myspace and Facebook, have much higher similarity ranks: 0.30 and 0.60 respectively.

Time variation I investigate the evolution of the distribution of my backward-similarity measure over time. Panel A of Figure B.2 plots time-series of the median, 10th and 90th percentiles of similarity ranks of companies raising capital in each year from 2000 to 2020. It suggests that backward-similarity is trending up, and thus that overall VC-funded startups tend to score higher on backward-similarity over time in my sample. Panel B of Figure B.2 plots time-series of the number of disruptive startups funded each year, i.e., VC-funded startups with similarity rank below 0.25, 0.10 and 0.05. These companies are likely to be highly novel as they differ greatly from previously funded ventures. Panel B shows that the number of such startups fluctuates over time with two peaks in 2007 and 2014, and two valleys in 2002 and 2009. Since 2014, this number has been trending down.

Panel A:



Panel B:

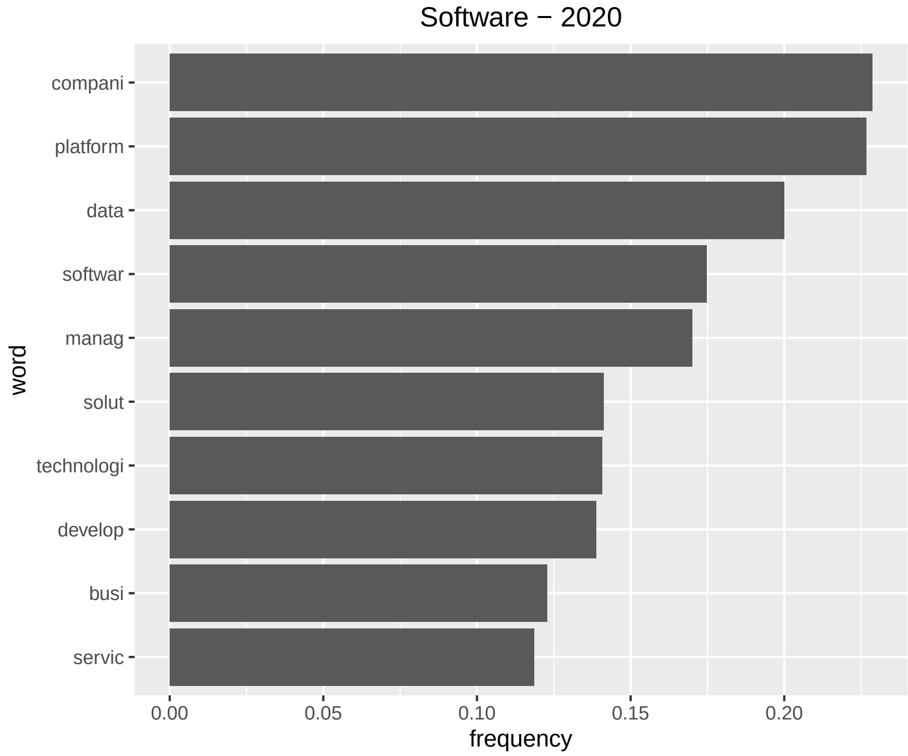
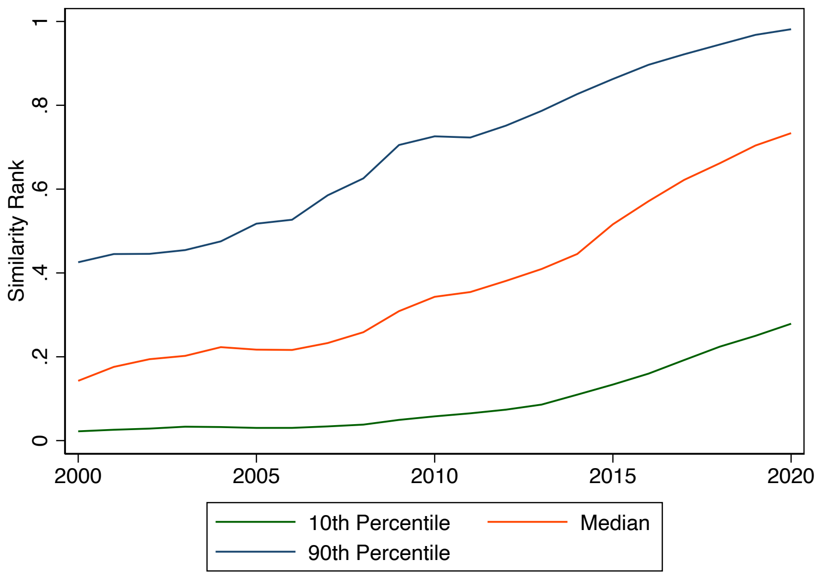


Figure B.1: **Most frequent words in the business descriptions of startups in the Software industry**: This figure displays the 10 most frequently mentioned words in business descriptions of startups in the Software industry that raised funding in 2008 (Panel A) and in 2020 (Panel B). On both panels, the frequency on the x-axis corresponds to the fraction of business descriptions mentioning the corresponding word on the y-axis.

Panel A: Distribution of similarity rank over time



Panel B: Number of disruptive startups over time

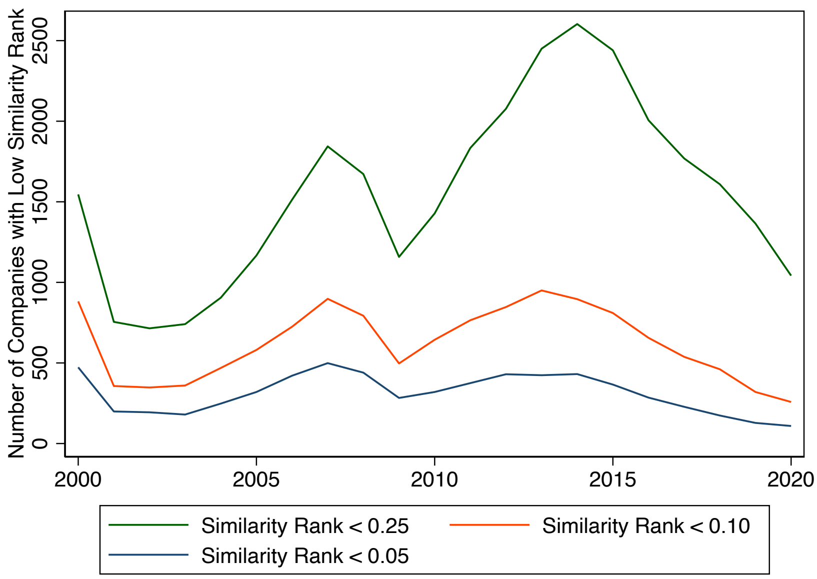


Figure B.2: **Backward-similarity over time**: This figure displays descriptive statistics of my backward-similarity measure. Panel A displays the 10th, 50th and 90th percentiles of the distribution of similarity rank of startups funded in each year between 2000 to 2020. A company's similarity rank is measured as the similarity of its business description to those of the set companies previously funded in the same industry. Panel B displays for each year the number of funded companies that have similarity rank below 0.25, 0.10 and 0.05, i.e., highly novel companies.

Company	Funding Round Date	Similarity Rank	
		No Adjustment	Remove Industry-Year FE
Netflix	2000-04-17	0.04	0.31
Napster	2000-05-23	0.06	0.21
Skype	2003-01-01	0.06	0.37
Facebook	2004-09-01	0.00	0.13
Digg	2005-10-28	0.33	0.23
YouTube	2005-11-01	0.13	0.11
Myspace	2005-03-16	0.00	0.12
Instagram	2010-10-06	0.30	0.37
Snap	2012-05-01	0.60	0.74
Lazada Group	2012-11-11	0.68	0.72

Table B.1: **Validation of the backward-similarity measure: Specific examples**. This table presents backward-similarity rank for specific companies that are either disruptive startups or copies of existing businesses, according to online sources. The startups considered as disruptive are Netflix, Napster, Skype, Facebook, Digg, Youtube and Myspace. Lazada Group is known to be the "copy of Amazon in Asia". Instagram and Snap provide online social networks trying to improve upon Myspace and Facebook, which were created before. For each of these companies, I report their percentile rank in terms of backward-similarity when they raise their first funding round in my sample. For instance, a value of 0.90 indicates that the company is in the top 10 percent of most similar companies when raising funds. I compute a company j's backward-similarity using a measure of textual similarity of its business descriptions compared to its "prior" peers, which comprise the set of companies in the same industry which raised capital in the 5 calendar years prior to company j's funding round. I report the similarity rank using both the unconditional distribution (column "No Adjustment"), and after subtracting the mean similarity measure within each Industry-Year (column "Removing Industry-Year FE"). Funding Round Date corresponds to the date of the first funding round of the company in my sample.

C Which VC firms adopt AI technology?

I test whether specific VC characteristics relate to the decision to adopt AI. I use a hazard model to study the likelihood and timing of AI adoption.[56] Specifically, I estimate a Weibull specification using VC-year observations:

$$Pr_{adoption_{j,t}} = \phi(t) \exp(\beta X_{j,t} + \gamma_y), \tag{16}$$

where j indexes VC firms and t refers to time measured as VC firm's age. $\phi(t)$ is parametrized as $\alpha t^{\alpha-1}$ ($\alpha > 0$), $X_{j,t}$ includes a wide range of variables measured at the VC-year level and γ_y are calendar year fixed effects. Standard errors are clustered at the VC firm level.

I first include in $X_{j,t}$ dummy variables indicating whether the VC firm's headquarters is located in the U.S. or Europe. Then, I add the logarithm of the number of employees at the VC firm at time t. To explore whether past investment experience and success is an important factor, I further include the logarithm of the numbers of previous investments and previously funded companies, as well as the fraction of previous investments that led to an IPO. To investigate the investment strategy of VCs adopting AI, I include the fraction of previous deals made as lead investor, the fraction corresponding to co-investments ("syndicated" deals), and the fraction corresponding to early-stage deals. Finally, I include the proportion of previous investments in the five main industry groups in my sample, which are Software, Health Care, Financial Services, Information Technology, and Internet Services. To check whether VCs adopting AI are more likely to invest in startups that are developing AI-based products, I also include the fraction of previous investments in the industry groups AI and Data.

Table C.1 presents the estimation results. The coefficients correspond to the impact on the hazard rate, i.e., the percentage change in the likelihood of AI adoption for a one-unit change in the independent variable. As exhibited by columns (3) to (5), VC experience does not correlate with AI adoption. I also do not find that the VC's previous success in terms of IPOs matters. Therefore, it is unlikely that AI-empowered VCs are the most established and experienced VC firms at the time of AI adoption. Columns (4) and (5) show that the fraction of previous deals that are syndicated and the fraction of previous lead roles do not play a significant role.

Four groups of variables appear to correlate with VC firms' decision to adopt AI. The first

[56] Hazard models of durations are models of the length of time spent in a given state before transition to another state. Here, I study the duration until adopting AI technology. The data used for this analysis include right-censored VC-year observations, i.e., observations up to the adoption of AI for AI-empowered VCs and all available observations for traditional VCs that never adopt the technology over my sample period.

one regards the VC's location. Column (5) suggests that US- and Europe-based VC firms are 79% and 69% respectively, more likely to adopt AI than similar VCs located in the rest of the world. Second, the number of employees of VC firms also correlates with AI adoption. Column (5) suggests that a 10% increase in the number of employees is associated with a 11% (1.136 $\times$ log(1.1)) increase in the hazard rate. Third, I find a strong positive relationship between AI adoption and the fraction of early-stage investments. Column (5) suggests that a 10 p.p. increase in the fraction of early-stage deals results in a 12% increase in the hazard rate. Fourth, industry specialization variables suggest that VCs adopting AI are more likely to be specialized in Software and Financial Services and less likely to be Health Care investors. The coefficients on the fraction of previous investments in Information Technology, Internet Services and AI and Data industry groups are only mildly significant (10%) and lower in magnitude.

	Hazard Model - Determinants of AI Adoption				
	(1)	(2)	(3)	(4)	(5)
US VC Firm	0.962***	0.690***	0.690***	0.734***	0.785***
	(0.237)	(0.236)	(0.237)	(0.240)	(0.244)
European VC Firm	0.707***	0.534**	0.520**	0.609**	0.690***
	(0.266)	(0.263)	(0.262)	(0.265)	(0.265)
Log(VC Nb. Employees)		0.988***	0.968***	1.077***	1.136***
		(0.075)	(0.077)	(0.084)	(0.083)
Log(Nb. Prior Investments)			-0.270	-0.112	-0.057
			(0.403)	(0.387)	(0.405)
Log(Nb. Prior Funded Companies)			0.322	0.171	0.080
			(0.435)	(0.419)	(0.439)
Proportion Prior IPO			-0.290	0.234	0.565
			(0.479)	(0.461)	(0.448)
Proportion Prior Lead				0.049	0.086
				(0.276)	(0.274)
Proportion Prior Syndicated				0.198	0.126
				(0.209)	(0.213)
Proportion Prior Early Stage				1.353***	1.260***
				(0.185)	(0.193)
Proportion Prior Software					0.916***
					(0.222)
Proportion Prior Heath Care					-0.858***
					(0.331)
Proportion Prior Financial Services					0.644***
					(0.247)
Proportion Prior Information Technology					-0.492*
					(0.285)
Proportion Prior Internet Services					0.388*
					(0.206)
Proportion Prior AI and Data					0.510*
					(0.274)
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	57,799	57,799	57,799	57,799	57,799
Log likelihood	-957.36	-899.41	-898.92	-877.33	-858.26
p value model test	0.000	0.000	0.000	0.000	0.000

Table C.1: **Determinants of AI adoption decision**: This table reports results from a hazard model to study the timing of AI adoption by VC firms. The hazard model is Weibull and is estimated with observations at the VC-year level. The coefficients correspond to the impact on the hazard rate, i.e., the percentage change in the likelihood of AI adoption for a one-unit change in the independent variable. US VC Firm and European VC Firm are dummy variables indicating whether the VC firm is located in the United States or in Europe respectively. Log(Nb. Employees) is the logarithm of the current number of employees at the VC firm. Log(Nb. Prior Investments) is the logarithm of the number of previous investments made by the VC firm. Log(Nb. Prior Funded Companies) is the logarithm of the number of distinct companies financed previously by the VC firm. Proportion Prior IPO is the percentage of previously financed companies that have managed to go public in an IPO. Proportion Prior Lead is the percentage of previous investments for which the VC firm was the lead investor. Proportion Prior Syndicated is the percentage of previous investments corresponding to syndicated deals, i.e., co-investments. Proportion Prior Early Stage is the percentage of early-stage deals (Pre-seed, Seed or Series A) among previous investments. Proportion Prior Software, Health Care, Financial Services, Information Technology, Internet Services, AI and Data are the percentages of previous investments in each industry group. Standard errors in parentheses are clustered at the VC firm level. * p<.10; ** p<.05; *** p<.01.

D Additional results on VC investments after adopting AI

Fund-level number of investments after adopting AI I run an analysis of the number of investments made by VCs in a given fund. Table D.1 presents the estimation results of a regression similar to equation (1) but estimated at the VC fund level using Preqin's dataset. I include VC firm fixed effects and fund's stage-country-industry-vintage fixed effects. In column (2), I control for the actual fund size (in $million), the fund sequence and the age of the VC firm as of the fund inception year. The coefficients in Table D.1 are very consistent with those estimated using VC-year observations. They suggest that VCs adopting AI observe an increase by 33% in the number of deals per fund relative to their own previous funds, compared to funds managed by other VCs. In columns (3) and (4), I consider as the dependent variable the logarithm of the ratio of the fund size and the number of investments made by the fund, i.e., the average deal size. The coefficient on AI adoption is negative but quite small in magnitude, and is only significant at the 10% level when I include control variables. Column (4) suggests a relative reduction by 6% in the average deal size after AI adoption.

Number of investments in backward-similar startups controlling for fund size One might be concerned that expanding investments towards backward-similar startups is simply due to the fact that VC firms adopting AI are getting bigger and cannot allocate more assets to low backward-similarity. To rule out this possibility, I estimate the same specifications as in Table 3 but I add as control variables the logarithm of the VC's number of funds and the logarithm of its AUM as of the investment year (for VC firms that can be matched with Preqin's fund data set). Table D.2 presents the estimation results. The coefficients on AI adoption are consistent with those presented in Table 3.

Specialization after adopting AI I test whether adopting AI is associated with changes in the industry specialization of VCs' portfolio. I estimate specification (1) with as the dependent variable the proportion of investments in the following industry groups: Software, Health Care, Financial Services, Information Technology, Internet Services, and AI and Data. Table D.3 presents the estimation results and shows no variation in sectoral specialization of VC firms after adopting AI. Finally, I test whether AI adoption is associated with changes in VCs' stage of investments. In Table D.4, I estimate specification (1) with as the dependent variable the proportion of early-stage investments made by the VC in a given year. The results do not indicate that VCs start focusing more or less on early-stage investments after adopting AI.

Number and type of investments using a continuous measure of AI adoption I test whether my results on VCs' number and type of investments are robust to using a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm (cf. subsection 3.3 for additional details). Tables D.5 and D.6 present the estimates of the specifications presented respectively in Table 2 and 3 but use as main independent variable the AI intensity instead of the AI dummy. They show that my conclusions remain valid when using a continuous measure of AI adoption.

Number and type of investments using an estimator robust to heterogeneous effects A fast-growing literature document that linear regressions with period and group fixed effects such as (1) and (2) may produce misleading estimates, if the effect of interest is heterogeneous between groups or over time (cf., De Chaisemartin and D'Haultfoeuille, 2022, for a survey). Several papers have proposed alternative estimators robust to this issue. Table D.7 shows that my findings on VCs' number and type of investments remain similar when using the imputation estimator of Borusyak et al. (2021).

Number and type of investments using a Poisson specification Cohn et al. (2022) discuss some issues related to estimating linear regressions of the log of one plus the outcome and recommend using instead a fixed-effects Poisson model. Table D.8 shows that my conclusions on VCs' number and type of investments are unaffected when employing a Poisson model, i.e., replacing the right-hand side of equation (1) by its exponent.

Investments in future rounds Table D.9 presents the estimation of specifications (3) and (6) for several additional dependent variables. First, I test whether VCs adopting AI become less likely to participate in future funding rounds of the startups in which they invest. Columns (1) and (2) use as the dependent variables the number of future funding-rounds of the same company in which the VC participates divided by the total number of future funding-round of the company. One sees that after adopting AI VCs become less likely be involved in future funding rounds, compared to other VCs. This could be either because they do not invest additional capital when the startup raises additional funding, or because they "cash out" and sell their shares to new investors. In either case, this suggests that VCs adopting AI become less involved in the growth phase of their portfolio companies.

Change in VC network I test whether AI-empowered VCs have access to different deals through connections they develop with other VCs. If this is the case, my results could be driven by the expansion of the VC's network. In columns (3) to (6) of Table D.9, I consider two dependent variables that are the logarithm of the number of investors participating in the funding round, and of the amount raised by the startup. The logic is that deals involving more investors and in which more capital is raised are more likely to be "mega-deals" that only a few very successful VCs can access through their network (Hochberg et al., 2007). My results show not relation between AI adoption on these two outcome variables.

Long shot bets I test whether AI adoption is associated with more "long shot bets", i.e., investments in companies that are more likely to fail, but with higher step-ups in value for startups that do in fact receive follow-on financing. To perform this test, I follow Ewens et al. (2018) and consider two dependent variables. The first one is the step-up in value across rounds, which is defined as the change in the post-money valuation between the current round of financing and the subsequent round. This variable can be computed only for startups raising several funding rounds and disclosing their valuations. The second measure is the ratio of exit value to total capital invested (or "total economic return") for startups that ultimately have a successful exit. Specifically I use as exit value the IPO or acquisition valuation for the startups that manage to go public or be acquired. Total capital invested is the sum of raised amounts across all funding rounds of the company. The estimation results are reported in columns (7)-(10) of Table D.9. I do not observe any meaningful change for any of these variables after AI adoption.

First-round I test whether my results are somehow different when I focus on investments corresponding to startups' first-round of funding. Table D.10 estimates specification (5) and (6) replacing the dummies for each backward-similarity quartile by a dummy indicating whether the investment corresponds to the startup's first-round of funding. The coefficients on the interaction between the dummy AI-empowered and First-Round are not significant.

First investment of VCs Because a VC firm can invest in the same startup across multiple rounds, I test whether my results are robust to restricting my sample to only the first investment by each VC firm in any given startup. This restriction prevents me from counting the same successful outcome more than once for any particular VC. The results are presented in Table D.11 and remains similar to those using the full sample.

Fund-level performance I attempt to gauge whether AI adoption is associated with changes in fund performance using Preqin data. In Table D.12, I estimate regressions at the VC fund level using as the dependent variables the net IRR and the multiple of the fund. Given that the performance data are mainly based on voluntary reporting from funds, the size of the sample I use is reduced and there might be a selection bias. With this caveat in mind, I do not observe a significant change in the performance of funds managed by AI-empowered VCs after they adopt AI, relative to other funds focused the same stage, country, industry and same vintage year. This suggests that adopting AI might allow VCs to scale-up their business (i.e, to manage more assets) without deteriorating fund performance and thus to collect more fees.[57]

Truncating the sample My sample includes investment outcomes observed through 2021 for investments made in the 2000 to 2019 period. I test whether my finding that AI-empowered VCs observe a decrease in the likelihood that their investments lead to breakthrough success is robust to dropping the investments made in the last years of my sample, addressing the concern that VCs might require a long time to achieve a successful exit. Table D.13 shows that the estimated coefficients remain stable when I consider only investments made up to 2018, 2017, 2016 and 2015 (for which I observe outcomes through 2021).

Investment outcomes using a continuous measure of AI adoption I test whether my results on investment outcomes are robust to using a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm. Table D.14 presents the estimation results and shows that my findings remain valid.

Investment outcomes using an estimator robust to heterogeneous effects Table D.15 presents the estimation results on investment outcomes when using the imputation estimator of Borusyak et al. (2021). The conclusions remain the same.

Investment outcomes using a Poisson specification Table D.16 estimates specifications corresponding to a Poisson model for the count variables related to innovation in Table 7. The results do not alter my conclusions.

[57]Even if I consider the negative (though insignificant) coefficient of -7.5% on IRR in column (2), the overall "effect" of AI adoption on the fees collected by VC firms would be positive. Indeed, Table 2 suggests that AUM increases by 33% following AI adoption. Given that the average fund size and management fee are \$123 million and 2% respectively, a simple back-of-the-envelope calculation suggests an average gain of $[1.33 \times (1 - 0.075) - 1] \times$ \$123 million $\times$ 0.02 = \$0.57 million per year.

	Log(Nb. Investments)		$\text{Log}(\frac{\text{Fund Size}}{\text{Nb. Investments}})$	
	(1)	(2)	(3)	(4)
AI-empowered	0.236***	0.282***	-0.111	-0.063*
	(0.083)	(0.086)	(0.067)	(0.034)
Log(Fund Size)		0.178***		
		(0.028)		
Log(Sequence)		0.253***		0.295***
		(0.046)		(0.072)
Log(VC Age)		0.060		-0.064
		(0.050)		(0.068)
Stage × Country × Industry × Vintage	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	9,624	9,624	9,624	9,624
R^2	0.73	0.75	0.67	0.68

Table D.1: **Number of investments in VC funds after VCs adopt AI**: This table reports results for regressions at the fund level, investigating whether after they adopt AI, AI-empowered VCs make more investments in their funds than do other VCs. The dependent variable in columns (1) and (2) is the logarithm of the number of deals made by the VC fund. In columns (3) and (4), the dependent variable is the logarithm of the fund's AUM divided by the fund's number of deals. The independent variable is a dummy variable equal to one if the VC managing the fund is classified as AI empowered as of the vintage inception year of the fund and equal to zero otherwise. I control for the logarithm of the fund size, the logarithm of the sequence number of the fund (whether the fund is the first, second, and so forth of the VC firms) and the logarithm of the age of the VC firm as of the vintage year. All regressions include fund's stage-country- industry-vintage year fixed effects. Standard errors in parentheses are clustered at the VC and vintage year levels. * p<.10; ** p<.05; *** p<.01.

	Log(1 + Nb. Investments) with Backward-similarity in			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
	(1)	(2)	(3)	(4)
AI-empowered	-0.055	0.156**	0.358***	0.761***
	(0.056)	(0.072)	(0.099)	(0.107)
Log(Nb. Funds)	0.180***	0.225***	0.230***	0.209***
	(0.028)	(0.027)	(0.030)	(0.035)
Log(AUM)	0.050***	0.048***	0.061***	0.084***
	(0.016)	(0.015)	(0.019)	(0.020)
Control Variables	Yes	Yes	Yes	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	29,798	29,798	29,798	29,798
R^2	0.67	0.63	0.64	0.72

Table D.2: **Number of investments in each quartile of backward-similarity after VCs adopt AI including additional control variables**: This table presents the same specifications as in Table 3 but adds as control variables the logarithm of the VC's number of funds and the logarithm of its AUM as of the investment year (for VC firms that can be matched with Preqin's fund data set). In columns 1 to 4, the dependent variable is the logarithm of (one plus) the number of investments made by the VC firm in that year with backward-similarity in a given quartile. For instance, 'Quartile 1" refers to the number of investments in funding rounds of companies with backward-similarity percentile rank below 0.25. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Proportion of Investments in Industry Group											
	Software		Health Care		Financial Services		Information Technology		Internet Services		AI and Data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AI-empowered	0.007	0.005	-0.008	-0.005	0.010	0.007	0.013	0.004	-0.022	-0.019	0.024	0.024
	(0.022)	(0.022)	(0.019)	(0.019)	(0.013)	(0.013)	(0.018)	(0.017)	(0.022)	(0.022)	(0.020)	(0.019)
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45,901	45,901	45,901	45,901	45,901	45,901	45,901	45,901	45,901	45,901	45,901	45,901
R^2	0.49	0.49	0.66	0.66	0.51	0.51	0.38	0.38	0.37	0.37	0.39	0.39

Table D.3: **Industry specialization after VCs adopt AI**: This table reports results for regressions at the VC-year level, investigating whether after they adopt AI, AI-empowered VCs make more investments in specific industry groups than do other VCs. The dependent variable is the proportion of investments in a given industry group, made by the VC in a given year. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Proportion Early Stage	
	(1)	(2)
AI-empowered	-0.033	-0.032
	(0.021)	(0.020)
Control Variables	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes
VC FE	Yes	Yes
Observations	45,901	45,901
R^2	0.61	0.61

Table D.4: **Early-stage specialization after VCs adopt AI**: This table reports results for regressions at the VC-year level, investigating whether AI-empowered VCs make more early-stage investments relative to other VCs, after they adopt AI. The dependent variable is the proportion of investments classified as early-stage (Pre-Seed, Seed or Series A), made by the VC in a given year. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Log(1 + Nb. Investments)		Log(Nb. Funds)		Log(AUM)	
	(1)	(2)	(3)	(4)	(5)	(6)
AI-intensity	0.477*** (0.082)	0.298*** (0.070)	0.449*** (0.083)	0.364*** (0.068)	0.524*** (0.116)	0.375*** (0.092)
Control Variables	No	Yes	No	Yes	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58,377	58,377	31,837	31,837	29,798	29,798
R^2	0.70	0.71	0.91	0.92	0.95	0.96

Table D.5: **AI intensity and number of investments**: This table presents the same specification as in 2 but replaces the the AI dummy by a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm (cf. subsection 3.3 for additional details). The dependent variables are the logarithm of one plus the number of investments in the year (columns 1 and 2), (one plus) the number of investments led by the VC firm in the year (columns 3 and 4), the (cumulative) number of funds as of the year (columns 5 and 6), and the (cumulative) AUM as of the year (columns 7 and 8). The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Log(1 + Nb. Investments) with Similarity Rank in							
	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-intensity	0.052	-0.007	0.282***	0.203***	0.545***	0.434***	0.946***	0.788***
	(0.073)	(0.070)	(0.074)	(0.068)	(0.111)	(0.106)	(0.133)	(0.128)
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58,377	58,377	58,377	58,377	58,377	58,377	58,377	58,377
R^2	0.60	0.61	0.56	0.56	0.57	0.57	0.64	0.65

Table D.6: **AI intensity and number of investments depending on backward-similarity of funded companies**: This table presents the same specification as in 3 but replaces the the AI dummy by a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm (cf. subsection 3.3 for additional details). In columns 1 to 8, the dependent variable is the logarithm of (one plus) the number of investments made by the VC firm in that year with backward-similarity in a given quartile. For instance, 'Quartile 1" refers to the number of investments in funding rounds of companies with backward-similarity percentile rank below 0.25. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Log(1+Nb. Investments)	Log(AUM)	Log(Nb. Funds)	Log(1 + Nb. Investments) with Similarity Rank in Quartile 1	Quartile 2	Quartile 3	Quartile 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AI-empowered	0.274***	0.306***	0.312***	0.058	0.208***	0.379***	0.633***
	(0.065)	(0.068)	(0.046)	(0.058)	(0.061)	(0.057)	(0.078)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	57,672	29,228	31,256	57,672	57,672	57,672	57,672

Table D.7: **Robust estimator and number of investments**: This table reports results for VC-year level regressions, implementing the robust estimator of Borusyak et al. (2021) (I use the STATA command `did_imputation`). The dependent variables are the logarithm of (one plus) the number of investments in the year (column 1), the (cumulative) VC's AUM as of the year (column 2) and the (cumulative) number of funds as of the year (column 3). In columns (4) to (7), the dependent variable is the logarithm of (one plus) the number of investments made by the VC firm in that year with backward-similarity in a given quartile. For instance, 'Quartile 1" refers to the number of investments in funding rounds of companies with backward-similarity percentile rank below 0.25. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	Nb. Investments	Nb. Investments with Similarity Rank in			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
	(1)	(2)	(3)	(4)	(5)
AI-empowered	0.136**	0.088	0.170**	0.219***	0.240**
	(0.053)	(0.073)	(0.086)	(0.070)	(0.104)
Control Variables	Yes	Yes	Yes	Yes	Yes
VC Country × Main Stage × Year FE	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes
Observations	58,377	58,377	58,377	58,377	58,377
Pseudo-R^2	0.60	0.47	0.41	0.45	0.53

Table D.8: **Poisson model and number of investments**: This table reports results for VC-year level regressions, using a Poisson specification for count variables. The dependent variable in column (1) is the number of investments in the year. In columns (2) to (5), the dependent variable is the number of investments made by the VC firm in that year with backward-similarity in a given quartile. For instance, 'Quartile 1" refers to the number of investments in funding rounds of companies with backward-similarity percentile rank below 0.25. The independent variable is a dummy variable equal to one if the VC is classified as AI empowered in that year and equal to zero otherwise. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors in parentheses are clustered at the VC firm and year levels. * p<.10; ** p<.05; *** p<.01.

	$\frac{\text{Nb. Future Investments}}{\text{Nb. Follow-on}}$		Log(Nb. Investors)		Log(Raised Amount)		Log(Step-up Valuation)		$\text{Log}\left(\frac{\text{Exit Value}}{\text{Capital Raised}}\right)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AI-empowered	-0.025*		-0.008		0.000		0.005		0.052	
	(0.014)		(0.024)		(0.027)		(0.013)		(0.059)	
AI-empowered × Similarity Rank		0.007		-0.039*		0.048		0.049		-0.012
		(0.016)		(0.021)		(0.035)		(0.036)		(0.098)
Similarity Rank		0.022		-0.015		0.012		0.060		-0.428
		(0.029)		(0.033)		(0.053)		(0.746)		(0.337)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
VC × Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	134,699	134,699	221,765	221,765	205,004	205,004	4,190	4,190	30,611	30,611
R^2	0.49	0.59	0.57	0.63	0.83	0.86	0.93	0.92	0.75	0.80

Table D.9: **Additional tests at the investment level**: This table reports results for regressions at the VC-investment level. Columns (1) and (2) use as the dependent variables the number of future funding-rounds of the same company in which the VC participates divided by the total number of future funding-round of the company. In columns (3) and (4), the dependent variable is the logarithm of the number of investors participating in the funding round. In columns (5) and (6), the dependent variable is the logarithm of the amount raised by the startup in the round. In columns (7) and (8), the dependent variable is the step-up in value across rounds, which is defined as the change in the post-money valuation between the current round of financing and the subsequent round. In columns (9) and (10), the dependent variable is the ratio of exit value to total capital invested (or "total economic return") for startups that ultimately have a successful exit. The independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * $p<.10$; ** $p<.05$; *** $p<.01$.

	Follow-on		Breakthrough Success	
	(1)	(2)	(3)	(4)
AI-empowered	0.029***		-0.027***	
	(0.010)		(0.008)	
AI-empowered × First-round	-0.012	-0.005	-0.002	-0.001
	(0.010)	(0.010)	(0.005)	(0.005)
First-round	-0.035***	-0.032***	-0.011***	-0.012***
	(0.007)	(0.007)	(0.004)	(0.004)
Control variables	Yes	Yes	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	No	Yes	No
VC × Year FE	No	Yes	No	Yes
Observations	221,765	221,765	221,765	221,765
R^2	0.46	0.52	0.54	0.60

Table D.10: **First-round and investment outcomes**: This table reports results for specification (5) and (6) at the VC-investment level, including the interaction with a dummy indicating whether the investment corresponds to the startup's first-round of funding. The dependent variable in columns (1) and (2) is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In columns (3) and (4), the dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. AI-empowered is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on		Breakthrough Success		Log(1+Nb. New Patent Applications)		Log(1+Nb. New Highly-cited Patents)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	0.022		-0.033***		-0.077***		-0.036*	
	(0.015)		(0.009)		(0.028)		(0.020)	
AI-empowered × Similarity Quartile 1		0.019		-0.035***		-0.048		-0.027
		(0.015)		(0.011)		(0.039)		(0.027)
AI-empowered × Similarity Quartile 2		0.009		-0.034***		-0.083**		-0.027
		(0.018)		(0.010)		(0.034)		(0.025)
AI-empowered × Similarity Quartile 3		0.010		-0.034***		-0.073**		-0.028
		(0.020)		(0.011)		(0.037)		(0.025)
AI-empowered × Similarity Quartile 4		0.039**		-0.030***		-0.089**		-0.049**
		(0.017)		(0.011)		(0.035)		(0.023)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	No	Yes	No	Yes	No	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	157,574	157,574	157,574	157,574	157,574	157,574	157,574	157,574
R^2	0.46	0.46	0.52	0.52	0.59	0.59	0.55	0.55

Table D.11: **First investment of VCs and investment outcomes**: This table reports results for specification (3) and (5) at the VC-investment level, including only observations corresponding to the first investment of each VC in a given startup. If a VC firm invests in the same startup across multiple rounds, only the first investment by that VC firm appears in this subsample. The dependent variable in columns (1) and (2) is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In columns (3) and (4), the dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. In columns (5)-(6), the dependent variable is the logarithm of one plus the number of patent applications filed by the startup after the VC invested. In columns (7)-(8), the dependent variable is the logarithm of one plus the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. AI-empowered is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. "Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Net IRR (%)		Net Multiple	
	(1)	(2)	(3)	(4)
AI-empowered	1.630	-7.522	-0.128	-0.147
	(8.572)	(7.490)	(0.295)	(0.352)
Log(Fund Size)		-1.309		-0.181**
		(1.254)		(0.076)
Log(Sequence)		3.903		0.273**
		(2.621)		(0.103)
Log(VC Age)		-6.170*		-0.432*
		(2.958)		(0.219)
Stage × Country × Industry × Vintage	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	1,270	1,270	1,690	1,690
R^2	0.81	0.81	0.69	0.70

Table D.12: **Fund performance after VCs adopt AI**: This table reports results for regressions at the fund level, investigating whether funds managed by AI-empowered VCs perform better relative to funds managed by other VCs, after they adopt AI. In columns (1) and (2), the dependent variable is the net IRR (after fees and carry) of the fund in percentage points. In columns (3) and (4), the dependent variable is the net multiple of the fund. The independent variable is a dummy variable equal to one if the VC managing the fund is classified as AI empowered as of the vintage inception year of the fund and equal to zero otherwise. I control for the logarithm of the fund size, the logarithm of the sequence number of the fund (later funds of VC firms) and the logarithm of the age of the VC firm as of the vintage year. All regressions include fund stage-VC country- fund industry-vintage year fixed effects. Standard errors in parentheses are clustered at the VC and vintage year levels. * p<.10; ** p<.05; *** p<.01.

	Breakthrough Success up to			
	2018	2017	2016	2015
	(1)	(2)	(3)	(4)
AI-empowered	-0.024**	-0.029***	-0.032**	-0.031*
	(0.010)	(0.011)	(0.014)	(0.016)
Control variables	Yes	Yes	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	195,265	168,498	145,016	123,766
R^2	0.54	0.53	0.53	0.52

Table D.13: **Robustness test truncating the sample**: This table reports results for regressions at the VC-investment level, consider only investments made up to 2018, 2017, 2016 and 2015 (for which I observe outcomes through 2021). The dependent variable in all columns is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. Each column corresponds to a different truncation of the sample. The independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on				Breakthrough Success			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-intensity	0.038***	0.027**			-0.030***	-0.031***		
	(0.010)	(0.012)			(0.010)	(0.011)		
AI-intensity × Similarity Quartile 2		-0.005	-0.001			-0.010	-0.011	
		(0.013)	(0.014)			(0.012)	(0.012)	
AI-intensity × Similarity Quartile 3		-0.003	-0.001			0.003	0.001	
		(0.015)	(0.016)			(0.011)	(0.010)	
AI-intensity × Similarity Quartile 4		0.030**	0.024*			0.004	0.007	
		(0.012)	(0.013)			(0.012)	(0.012)	
AI-intensity × Similarity Rank				0.034**				0.020
				(0.016)				(0.017)
Similarity Rank				-0.102***				-0.061**
				(0.030)				(0.028)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	Yes	No	No	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	No	No	Yes	Yes	No	No
VC × Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.46	0.46	0.52	0.52	0.54	0.54	0.60	0.60

Table D.14: **AI intensity and investment outcomes**: This table presents the same specification as in Tables 5 and 6 but replaces the the AI dummy by a continuous measure of "AI intensity", defined as the logarithm of one plus the current number of AI-related jobs at the VC firm (cf. subsection 3.3 for additional details). The dependent variable in columns (1)-(4) is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In columns (5)-(8), the dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. "Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on		Breakthrough Success		Log(1+Nb. New Patent Applications)		Log(1+Nb. New Highly-cited Patents)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	0.008*		-0.033***		-0.096***		-0.069***	
	(0.005)		(0.004)		(0.016)		(0.011)	
AI-empowered × Similarity Quartile 1		-0.016**		-0.049***		-0.165***		-0.125***
		(0.007)		(0.006)		(0.022)		(0.015)
AI-empowered × Similarity Quartile 2		-0.001		-0.031***		-0.077***		-0.060***
		(0.006)		(0.005)		(0.019)		(0.013)
AI-empowered × Similarity Quartile 3		-0.005		-0.031***		-0.121***		-0.054***
		(0.006)		(0.005)		(0.018)		(0.012)
AI-empowered × Similarity Quartile 4		0.033***		-0.030***		-0.064***		-0.061***
		(0.005)		(0.004)		(0.017)		(0.012)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	No	Yes	No	Yes	No	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	186,569	186,569	186,569	186,569	186,569	186,569	186,569	186,569

Table D.15: **Robust estimator and investment outcomes**: This table reports results for VC-investment level regressions, implementing the robust estimator of Borusyak et al. (2021) (I use the STATA command `did_imputation`). AI-empowered is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. In columns (2), (4) and (6), each coefficient AI-empowered × Similarity Quartile i corresponds to a different weighting scheme for the "treated" observations, i.e., investments made by VCs that have adopted AI. AI-empowered × Similarity Quartile i is estimated by giving a weight of 1 to each "treated" investment in backward-similarity quartile i, and 0 to other "treated" investments. Weights are not used for untreated investments. The dependent variable in columns (1)-(2) is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. In columns (3)-(4), the dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. In columns (5)-(6), the dependent variable is the logarithm of one plus the number of patent applications filed by the startup after the VC invested. In columns (7)-(8), the dependent variable is the logarithm of one plus the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. "Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. The number of observations is lower than in the main analysis because the estimator of Borusyak et al. (2021) requires fixed effects to be estimated on the control groups. If an industry-country-stage-year fixed effect cannot be estimated among traditional VCs, the (treated) observations in that segment are dropped from the sample. * p<.10; ** p<.05; *** p<.01.

	Nb. New Patent Applications				Nb. New Highly Cited Patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	-0.118***				-0.067			
	(0.042)				(0.048)			
AI-empowered × Similarity Quartile 1		-0.064				-0.027		
		(0.062)				(0.073)		
AI-empowered × Similarity Quartile 2		-0.137**	-0.120*			-0.089	-0.128	
		(0.058)	(0.070)			(0.067)	(0.089)	
AI-empowered × Similarity Quartile 3		-0.079	-0.029			0.031	0.013	
		(0.057)	(0.073)			(0.076)	(0.082)	
AI-empowered × Similarity Quartile 4		-0.161**	-0.105			-0.135	-0.171*	
		(0.068)	(0.074)			(0.083)	(0.100)	
AI-empowered × Similarity Rank				-0.099				-0.169
				(0.083)				(0.119)
Similarity Rank				-0.608***				-0.785***
				(0.232)				(0.268)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	Yes	No	No	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	No	No	Yes	Yes	No	No
VC × Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765	221,765	221,765
Pseudo-R^2	0.61	0.61	0.67	0.67	0.55	0.55	0.61	0.61

Table D.16: **Poisson model and innovation**: This table reports results for regressions at the VC-investment level, using a Poisson specification for count variables. In columns (1)-(4), the dependent variable is the number of patent applications filed by the startup after the VC invested. In columns (5)-(8), the dependent variable is the number of highly-cited patents obtained by the startup after the VC invested. A patent is considered as highly-cited if it receives a number of citations above the average number of citations of patents granted in the same year and technology class. AI-empowered is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. "Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

E Aggregate results on startup creations and AI adoption by VCs after the introduction of cloud computing

Empirical specifications I estimate the aggregate effect of the introduction of cloud computing on startup creations. I use a strongly balanced sample at the industry-country-year level. In each segment (i.e., industry-country-year), I count the number of newly created startups, including non VC-funded companies, according to their creation dates available in the Crunchbase database. Table E.1 presents summary statistics of my sample at the industry-country-year level.[58] Then, I rely on a difference-in-differences framework with continuous treatment intensity varying across industries:

$$Y_{i,c,t} = \beta \left\{ IndustryExposure_i \times Post_t \right\} + \gamma_{c \times t} + \alpha_i + \epsilon_{i,c,t}, \tag{17}$$

where $Y_{i,c,t}$ is a measure of startup creations in industry i, country c, and year t. $IndustryExposure_i$ is the treatment intensity (between 0 and 1) of industry i, defined in (7). $Post_t$ is a dummy variable equal to one after 2006 and zero otherwise, $\gamma_{c \times t}$ are country-year fixed effects, and α_i are industry fixed effects. Country-year fixed effects capture time-varying determinants of startup and VC activities in different countries, such as legal and institutional environments, aggregate shocks and trends in innovation activity. Industry fixed effects account for heterogeneity in the propensity to create and fund startups across industries. Standard errors are double-clustered at the industry and country levels.

To study the dynamics of the effects I also estimate regression (17) replacing $IndustryExposure_i \times Post_t$ by interactions of exposure and year dummies $IndustryExposure_i \times Year_t$. Finally, in order to alleviate concerns regarding noise in my industry-level exposure measurement, I also estimate (17) using a binary treatment. I replace $IndustryExposure_i$ by $HighIndustryExposure_i$ which takes the value of one for industries in the top quartile of exposure and equal to zero otherwise. I also estimate this specification using only the sub-sample of industries belonging to the top and bottom quartiles of exposure. The rationale is that industries in the top quartile are clearly benefiting more from the introduction of AWS than industries in the bottom quartile, even if exposure is measured with noise. Furthermore, the fact that those two groups of industries are very different and thus require different sets of skills makes unlikely that entrepreneurs endogenously switch across

[58] In my regression estimations, I only keep industry-country with at least two company creations and two funding rounds over my sample period. My results are robust to this choice.

industry groups.

Effects on startup creations I provide evidence that the introduction of AWS led to a larger increase in startup creations in the most exposed industries. Table E.2 reports the estimation results of specification (17) using as the dependent variable the logarithm of (one plus) the number of newly created companies in the industry-country-year. The magnitude of the coefficient in column (2) suggests that in industries at the 90th percentile in terms of exposure, the relative increase in the number of firm creations is 40% after the shock. Given that the average number of firm creations in each industry-country-year is 4.21, this corresponds to almost two additional companies. In contrast, for industries at the 10th percentile, the increase is only 4%. Column (3) replaces the continuous industry exposure by a binary treatment variable equal to one if the industry's exposure is in the top quartile. The coefficient remains positive and highly significant, though lower in magnitude. This is not surprising as this specification considers as untreated all observations corresponding to industries with exposure below 0.75. Column (4) uses only the sub-sample of observations corresponding to industries that lie in the first and fourth quartiles of exposure. The coefficient is of similar magnitude as in column (2). Although one cannot guarantee that Crunchbase provides a complete coverage of all startups worldwide, these results suggest that, after the introduction of AWS, significantly more startups were created in industries more exposed to the shock. Table E.3 estimates specifications corresponding to a Poisson model for the count of newly created startups. The results are similar and lead to the same conclusion.

Figure 10 reports the results from the dynamic version of specification (17), where each point is an estimate of the coefficient on the interaction of year dummies and industry exposure, relative to year 2006. The horizontal bars correspond to 95% confidence intervals. The patterns in the figure show that there is no pre-trend and that the timing of the increase in the number of company creations is consistent with the introduction of AWS.

Effects on AI adoption by VCs across industries I test whether AI adoption by VCs varies *across* industries depending on the extent of the boom in startup creations following the introduction of Amazon Web Services. To do so, I estimate regression (17) using several dependent variables capturing the presence of AI-empowered VCs: (i) a dummy indicating whether there is at least one active VC in that industry-country-year that is classified as AI-empowered, (ii) the logarithm of (one plus) the number of active VCs classified as AI-empowered, and (iii) the percentage of active VCs classified as AI-empowered. The results are presented in Table E.4. They show that there is a significantly larger increase in the presence of AI-empowered VCs in the most exposed industries after 2006. The coefficient in column (1) implies that there is a rise by 9.6 p.p. (96% of the unconditional mean) in the likelihood of observing a VCs using AI technologies in industries at the 90th percentile of exposure.

Figure E.1 reports graphically the results from the dynamic specification of (17) when using as the dependent variable the dummy indicating the presence of at least one active AI-empowered VC in the industry-country-year. The patterns in the figure demonstrate there is no pre-trend and that the timing of the presence of AI-empowered VCs coincides with the increase in the number of new firms triggered by the introduction of AWS.

In Table E.5, I estimate specification (17) using as the dependent variables the fraction of VC activity in the industry-country-year accounted by AI-empowered VCs in terms of (i) number of VC investments, (ii) number of funding rounds in which they participate, and (iii) funding amounts raised by startups in funding rounds in which they participate. All columns show a relative rise in the fraction of startups receiving funding from AI-empowered VCs in the industries with the highest exposure.

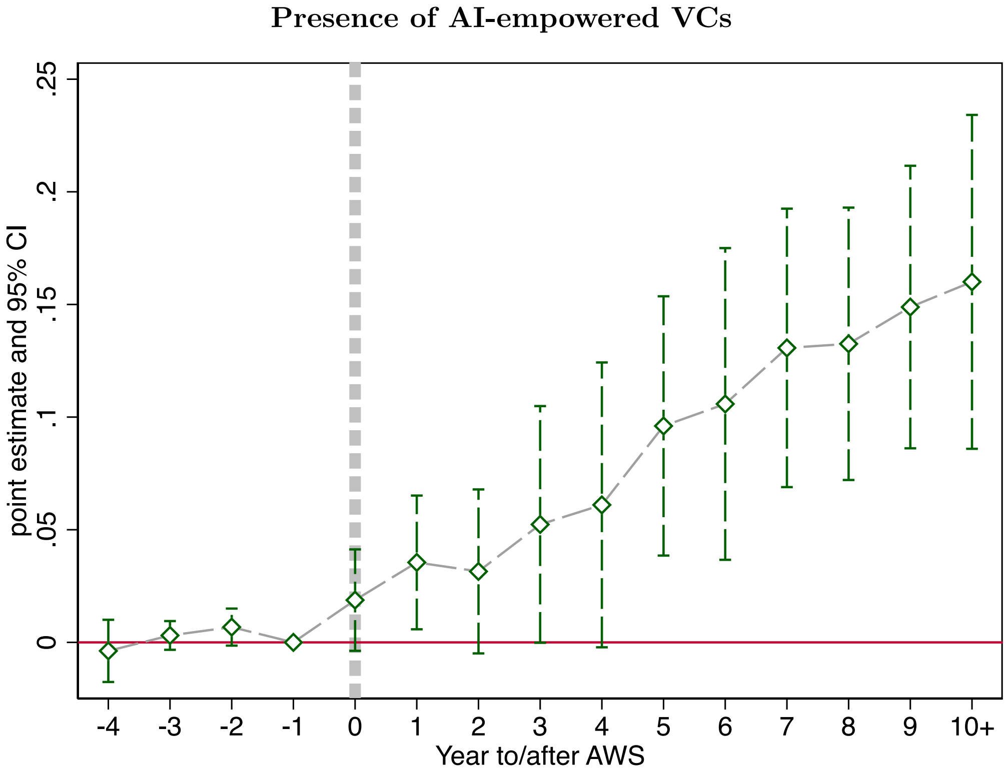


Figure E.1: **The effects of the introduction of cloud computing on the presence of AI-empowered VCs**: This figure plots the estimated coefficients from difference-in-differences regressions at the industry-country-year level, for the interaction terms of each year relative to 2006 and the industry exposure to the introduction of cloud computing (Amazon Web Services). The dependent variable is a dummy equal to one if there is at least one active VC in that industry-country-year that is classified as AI-empowered. The 2005 interaction term is the excluded category (-1), reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors clustered at the industry and country levels. The regression includes industry fixed effects and country-year fixed effects.

Variable	Obs	Mean	Sd	5%	25%	50%	75%	95%
Industry-Country-Year level								
Industry Exposure	154,560	0.55	0.26	0.13	0.33	0.53	0.76	0.95
Nb. New Companies	154,560	4.21	21.65	0.00	0.00	1.00	2.00	14.00
$\mathbb{1}$(AI-empowered VCs)	154,560	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Nb. AI-empowered VCs	154,560	0.34	2.20	0.00	0.00	0.00	0.00	1.00
%AI-empowered VCs	62,113	5.45	14.40	0.00	0.00	0.00	0.00	33.33
%Investments with AI-empowered VCs	62,113	5.39	14.52	0.00	0.00	0.00	0.00	30.77
%Rounds with AI-empowered VCs	62,113	8.98	21.68	0.00	0.00	0.00	0.00	55.56
%Funding with AI-empowered VCs	52,521	5.76	15.65	0.00	0.00	0.00	0.73	33.64

Table E.1: **Summary statistics at the industry-country-year level**. This table presents descriptive statistics at the industry-country-year level.

	Log(1 + Nb. New Companies)			
	(1)	(2)	(3)	(4)
Industry Exposure	-0.147			
	(0.113)			
Post	0.082**			
	(0.040)			
Industry Exposure × Post	0.354***	0.371***		
	(0.059)	(0.061)		
High Industry Exposure × Post			0.167***	0.309***
			(0.037)	(0.050)
Country FE	Yes	No	No	No
Country × Year FE	No	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes
Observations	154,560	154,560	154,560	65,982
R^2	0.24	0.73	0.73	0.71

Table E.2: **The effect of the introduction of cloud computing on firm creations**: This table reports results for regressions at the industry-country-year level. The specification is an OLS model including country-year and industry fixed effects. The dependent variable is the logarithm of (one plus) the number of firms created in the industry-country in a given year. The main independent variable is the interaction of Industry Exposure (the industry-level exposure to the introduction of cloud computing) and Post (a dummy equal to one after 2006). In column (4), the sample includes only industries that belong to the top and bottom quartiles in terms of exposure. High Industry Exposure is a dummy equal to one for industries in the top quartile. Standard errors in parentheses are clustered at the industry and country levels. * p<.10; ** p<.05; *** p<.01.

	Nb. New Companies			
	(1)	(2)	(3)	(4)
Industry Exposure	0.088 (0.378)			
Post	0.126 (0.095)			
Industry Exposure × Post	0.646*** (0.153)	0.807*** (0.192)		
High Industry Exposure × Post			0.391*** (0.141)	0.670*** (0.159)
Country FE	Yes	No	No	No
Country × Year FE	No	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes
Observations	154,560	154,560	154,560	65,982
R^2				

Table E.3: **Poisson model to estimate the effect of cloud computing on firm creations**: This table reports results for regressions at the industry-country-year level. The specification is a Poisson model including country-year and industry fixed effects. The dependent variable is the number of firms created in the industry-country in a given year. The main independent variable is the interaction of Industry Exposure (the industry-level exposure to the introduction of cloud computing) and Post (a dummy equal to one after 2006). In column (4), the sample includes only industries that belong to the top and bottom quartiles in terms of exposure. High Industry Exposure is a dummy equal to one for industries in the top quartile. Standard errors in parentheses are clustered at the industry and country levels. * p<.10; ** p<.05; *** p<.01.

	𝟙(AI-empowered VCs)		Log(1+Nb. AI-empowered VCs)		%AI-empowered VCs	
	(1)	(2)	(3)	(4)	(5)	(6)
Industry Exposure × Post	0.107***		0.177**		2.388***	
	(0.020)		(0.068)		(0.646)	
High Industry Exposure × Post		0.090***		0.153**		2.097***
		(0.016)		(0.059)		(0.462)
Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	154,560	65,982	154,560	65,982	62,113	26,386
R^2	0.36	0.36	0.49	0.48	0.18	0.20

Table E.4: **The effect of the introduction of cloud computing on the presence of AI-empowered VCs**: This table reports results for regressions at the industry-country-year level. The specification is an OLS model including country-year and industry fixed effects. In columns (1) and (2), the dependent variable is a dummy equal to one if at least one VC active in the country-industry-year is classified as AI-empowered. In columns (2) and (3), the dependent variable is the logarithm of (one plus) the number of active VCs classified as AI-empowered in the industry-country in a given year. In columns (4) and (6), the dependent variable is the percentage of active VCs classified as AI-empowered. The independent variable is the interaction of Industry Exposure (the industry-level exposure to the introduction of cloud computing) and Post (a dummy equal to one after 2006). In column (2), (4) and (6), the sample includes only industries that belong to the top and bottom quartiles in terms of exposure. High Industry Exposure is a dummy equal to one for industries in the top quartile and zero for those in the bottom quartile. Standard errors in parentheses are clustered at the industry and country levels. * p<.10; ** p<.05; *** p<.01.

	Proportion of AI-empowered VCs among					
	Investments		Rounds		Funding	
	(1)	(2)	(3)	(4)	(5)	(6)
Industry Exposure × Post	2.445***		4.079***		3.281***	
	(0.669)		(1.330)		(0.709)	
High Industry Exposure × Post		2.189***		3.407***		2.887***
		(0.470)		(0.914)		(0.473)
Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,113	26,386	62,113	26,386	52,521	22,439
R^2	0.19	0.21	0.20	0.22	0.17	0.20

Table E.5: **The effect of the introduction of cloud computing on the fraction of funding activity accounted for by AI-empowered VCs**: This table reports results for regressions at the industry-country-year level. The specification is an OLS model including country-year and industry fixed effects. The dependent variable is the percentage of VCs' activity in the country-industry-year that is accounted by AI-empowered, in terms of number of investments (columns (1) and (2)), number of funding rounds (columns (3) and (4)) and funding amounts (columns (5) and (6)). The independent variable is the interaction of Industry Exposure (the industry-level exposure to the introduction of cloud computing) and Post (a dummy equal to one after 2006). In column (2), (4) and (6), the sample includes only industries that belong to the top and bottom quartiles in terms of exposure. High Industry Exposure is a dummy equal to one for industries in the top quartile and zero for those in the bottom quartile. Standard errors in parentheses are clustered at the industry and country levels. * p<.10; ** p<.05; *** p<.01.

F Additional results on VC investments after the introduction of cloud computing

Excluding the VCs with cloud computing expertise As discussed in subsection 5.1, one concern might be that the introduction of AWS lowered the cost of adopting AI for some VCs but not others. If my results are driven by the increase in the number of investment opportunities for VCs, my conclusions should remain the same when excluding the VCs with cloud computing expertise before the shock. I test this by estimating (9) but only on the subset of investments made by VCs that never make any investment before 2006 in the most cloud-related industries, i.e., industries with exposure (8) in the top 5%. F.1, F.2 and F.3 report the results. My conclusions remain similar.

Excluding the most exposed industries Given the low marginal cost of using AI to screen an additional investment, when VCs adopt this technology one expects them to use it to screen their investments in all industries. If this is the case, my conclusions should be robust to excluding investments made in the most treated industries. I test this by estimating (9) but only on the subset of investments in industries with exposure below 0.8. Tables F.4, F.5 and F.6 report the results. My conclusions remain similar.

Binary treatment Callaway et al. (2021) show that interpreting differences across different values of the treatment (i.e., here VC exposure) might be challenging in the presence of treatment effect heterogeneity. Typically regressions such as (9) including time and group fixed effects put more weight on treatment near the average and less on the tails. The corresponding estimates might be misleading if the highest or lowest treatment create relatively extreme causal responses (for example if the effect is zero below a certain exposure). To make sure my conclusions are not dependent on using a continuous treatment intensity, I also estimate (9) replacing $VCExposure_i$ by a dummy variable $HighVCExposure_i$, which takes the value of one for VCs with exposure above the median (0.5) and equal to zero otherwise. Tables F.7, F.8 and F.9 report the estimation results using as the dependent variables a dummy indicating respectively whether the investment is made by an AI-empowered VC, whether the startup had survived and raised follow-on funding, and whether the startup has achieved breakthrough success. My conclusions remain the same.

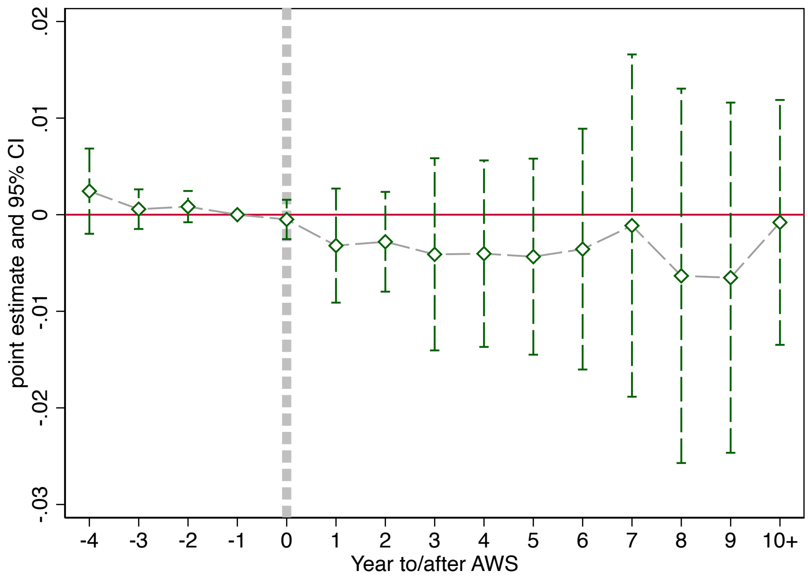
Low-exposure VCs I test whether the advent of AWS has any effect on VCs with low exposure. I estimate the following regression at the VC investment level, using the subsample of investments made by VCs with exposure to cloud-related sectors in the bottom quartile (i.e., below 0.25) and estimate:

$$Y_{j,k,t} = \sum_{l=-4,\ l\neq-1}^{10+} \beta_l \left\{Year(l)_t\right\} + X_{j,k,t} + \alpha_j + \gamma_{i\times c\times s} + \epsilon_{j,k,t}, \tag{18}$$

where $Year(l)_t$ is a dummy variable equal to one if year t corresponds to l years before/after the introduction of AWS. The omitted category is the year 2005. $Y_{j,k,t}$ is either an outcome variable for a given investment made by VC j in startup k in year t or a dummy indicating whether VC j is classified as AI-empowered as of the investment date. $X_{j,k,t}$ are time-varying control variables which include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. α_j are VC firm fixed effects. $\gamma_{i\times c\times s}$ are startup's industry $\times$ country $\times$ funding stage fixed effects. Standard errors are double-clustered at the VC firm and startup company levels.

Figure F.1 plot the estimated coefficients β_l. In Panel A, the dependent variable is a dummy indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. In Panel B, the dependent variable is a dummy indicating whether the VC-funded company goes public in IPO or is acquired for a higher value than the total VC investments in the company. Both panels show no significant variation after the advent of AWS for investments made by VCs with low exposure.

AI adoption by low-exposure VCs



Likelihood of breakthrough success for low-exposure VCs

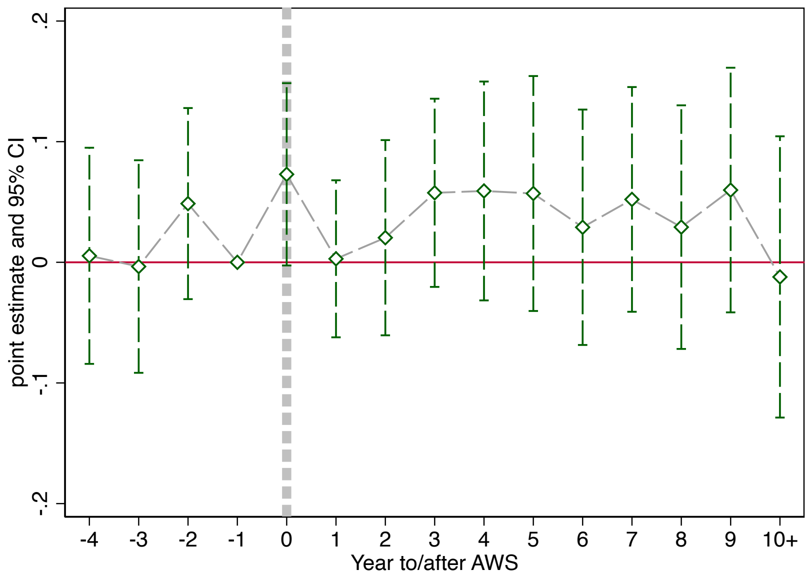


Figure F.1: **AI adoption and breakthrough success for low-exposure VCs**: These figures plot the estimated coefficients from regressions at the VC investment level, for the terms corresponding to each year relative to 2006. The sample includes only observations corresponding to investments made by VCs with exposure to cloud-related sectors in the bottom quartile (i.e., below 0.25). In Panel A, the dependent variable is a dummy indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. In Panel B, the dependent variable is a dummy indicating whether the VC-funded company goes public in IPO or is acquired for a higher value than the total VC investments in the company. The 2005 interaction term is the excluded category, reported as zero in the figures. The horizontal bars represent the 95% confidence interval for the coefficient estimates with standard errors clustered at the VC and startup levels. Regressions include VC firm fixed effects and startup's industry-country-funding stage fixed effects. Regressions also control for the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made.

	AI-empowered			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.088* (0.046)	0.091* (0.048)		0.066 (0.043)
VC Exposure × Post × Similarity Quartile 1			0.066 (0.043)	
VC Exposure × Post × Similarity Quartile 2			0.072 (0.046)	0.006 (0.011)
VC Exposure × Post × Similarity Quartile 3			0.067** (0.031)	0.001 (0.020)
VC Exposure × Post × Similarity Quartile 4			0.132** (0.062)	0.066* (0.035)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	91,340	91,340	91,340	91,340
R^2	0.83	0.83	0.83	0.83

Table F.1: **Dropping VCs with expertise in cloud computing industries to estimate the effect on VCs' likelihood to adopt AI**: This table reports results for regressions at the VC-investment level, using only the subset of investments made by VCs that never invest in industries with exposure (7) in the top 5% before 2006. The dependent variable is a dummy variable indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.046 (0.035)	0.038 (0.034)		0.029 (0.042)
VC Exposure × Post × Similarity Quartile 1			0.029 (0.042)	
VC Exposure × Post × Similarity Quartile 2			-0.039 (0.055)	-0.068 (0.066)
VC Exposure × Post × Similarity Quartile 3			0.018 (0.114)	-0.012 (0.118)
VC Exposure × Post × Similarity Quartile 4			0.112** (0.046)	0.082*** (0.031)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	91,340	91,340	91,340	91,340
R^2	0.46	0.47	0.47	0.47

Table F.2: **Dropping VCs with expertise in cloud computing to estimate the effect on the likelihood that funded-startups receive follow-on funding**: This table reports results for regressions at the VC-investment level, using only the subset of investments made by VCs that never invest in industries with exposure (7) in the top 5% before 2006. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Breakthrough Success			
	(1)	(2)	(3)	(4)
VC Exposure × Post	-0.107***	-0.108***		-0.096***
	(0.030)	(0.030)		(0.037)
VC Exposure × Post × Similarity Quartile 1			-0.096***	
			(0.037)	
VC Exposure × Post × Similarity Quartile 2			-0.150***	-0.054
			(0.049)	(0.059)
VC Exposure × Post × Similarity Quartile 3			-0.010	0.086
			(0.109)	(0.114)
VC Exposure × Post × Similarity Quartile 4			-0.112***	-0.016
			(0.041)	(0.026)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	91,340	91,340	91,340	91,340
R^2	0.52	0.52	0.52	0.52

Table F.3: **Dropping VCs with expertise in cloud computing to estimate the effect on the likelihood that investments lead to breakthrough success**: This table reports results for regressions at the VC-investment level, using only the subset of investments made by VCs that never invest in industries with exposure (7) in the top 5% before 2006. The dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	AI-empowered			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.132*** (0.049)	0.132*** (0.045)		0.098*** (0.037)
VC Exposure × Post × Similarity Quartile 1			0.098*** (0.037)	
VC Exposure × Post × Similarity Quartile 2			0.104** (0.044)	0.006 (0.014)
VC Exposure × Post × Similarity Quartile 3			0.095*** (0.031)	-0.003 (0.022)
VC Exposure × Post × Similarity Quartile 4			0.196*** (0.062)	0.098*** (0.037)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	86,661	86,661	86,661	86,661
R^2	0.80	0.81	0.81	0.81

Table F.4: **Dropping investments in high-treatment industries to estimate the effect on VCs' likelihood to adopt AI**: This table reports results for regressions at the VC-investment level, using only the subset of investments in industries with exposure (7) below 0.8. The dependent variable is a dummy variable indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on			
	(1)	(2)	(3)	(4)
VC Exposure × Post	0.054 (0.033)	0.046 (0.033)		0.034 (0.039)
VC Exposure × Post × Similarity Quartile 1			0.034 (0.039)	
VC Exposure × Post × Similarity Quartile 2			-0.009 (0.053)	-0.043 (0.064)
VC Exposure × Post × Similarity Quartile 3			0.080 (0.111)	0.046 (0.116)
VC Exposure × Post × Similarity Quartile 4			0.107** (0.045)	0.073** (0.031)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	86,661	86,661	86,661	86,661
R^2	0.49	0.50	0.50	0.50

Table F.5: **Dropping investments in high-treatment industries to estimate the effect on the likelihood that funded-startups receive follow-on funding**: This table reports results for regressions at the VC-investment level, using only the subset of investments in industries with exposure (7) below 0.8. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Breakthrough Success			
	(1)	(2)	(3)	(4)
VC Exposure × Post	-0.092***	-0.094***		-0.081**
	(0.029)	(0.029)		(0.035)
VC Exposure × Post × Similarity Quartile 1			-0.081**	
			(0.035)	
VC Exposure × Post × Similarity Quartile 2			-0.125***	-0.044
			(0.048)	(0.059)
VC Exposure × Post × Similarity Quartile 3			0.003	0.084
			(0.109)	(0.115)
VC Exposure × Post × Similarity Quartile 4			-0.108**	-0.027
			(0.042)	(0.028)
Control Variables	No	Yes	Yes	Yes
VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	86,661	86,661	86,661	86,661
R^2	0.54	0.54	0.54	0.54

Table F.6: **Dropping investments in high-treatment industries to estimate the effect on the likelihood that investments lead to breakthrough success**: This table reports results for regressions at the VC-investment level, using only the subset of investments in industries with exposure (7) below 0.8. The dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	AI-empowered			
	(1)	(2)	(3)	(4)
High VC Exposure × Post	0.030** (0.013)	0.031*** (0.011)		0.016 (0.010)
High VC Exposure × Post × Similarity Quartile 1			0.016 (0.010)	
High VC Exposure × Post × Similarity Quartile 2			0.019** (0.009)	0.004 (0.006)
High VC Exposure × Post × Similarity Quartile 3			0.021*** (0.008)	0.005 (0.010)
High VC Exposure × Post × Similarity Quartile 4			0.058*** (0.021)	0.042** (0.019)
Control Variables	No	Yes	Yes	Yes
High VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.79	0.79	0.79	0.79

Table F.7: **Binary treatment to estimate the effect on VCs' likelihood to adopt AI**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy variable indicating whether the investment is made by a VC classified as AI-empowered as of the investment date. The main independent variable is the interaction of High VC Exposure (a dummy indicating whether the VC firm's exposure is above the median) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of High VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Follow-on			
	(1)	(2)	(3)	(4)
High VC Exposure × Post	0.011 (0.013)	0.007 (0.013)		-0.002 (0.016)
High VC Exposure × Post × Similarity Quartile 1			-0.002 (0.016)	
High VC Exposure × Post × Similarity Quartile 2			-0.011 (0.023)	-0.009 (0.027)
High VC Exposure × Post × Similarity Quartile 3			0.016 (0.043)	0.018 (0.045)
High VC Exposure × Post × Similarity Quartile 4			0.033^{*} (0.018)	0.035^{**} (0.014)
Control Variables	No	Yes	Yes	Yes
High VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.46	0.47	0.47	0.47

Table F.8: **Binary treatment to estimate the effect on the likelihood that funded-startups receive follow-on funding**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy indicating whether the VC-funded startup is still active and has received follow-on funding by 2021. The main independent variable is the interaction of High VC Exposure (a dummy indicating whether the VC firm's exposure is above the median) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of High VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01

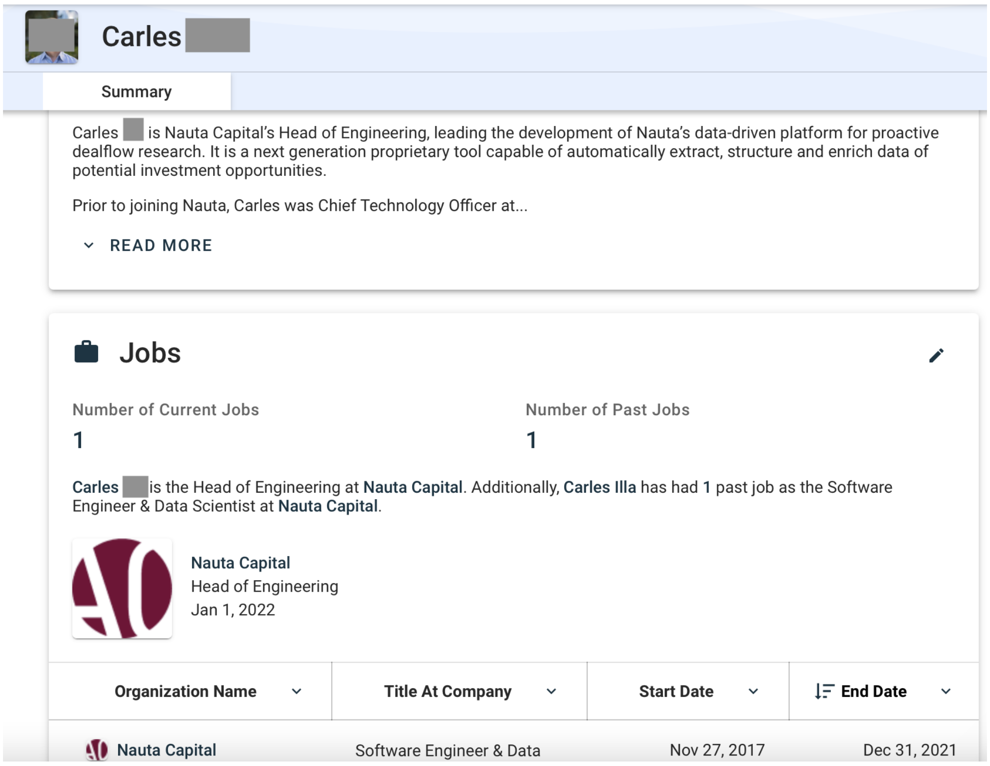
	Breakthrough Success			
	(1)	(2)	(3)	(4)
High VC Exposure × Post	-0.034***	-0.036***		-0.030**
	(0.013)	(0.013)		(0.015)
High VC Exposure × Post × Similarity Quartile 1			-0.030**	
			(0.015)	
High VC Exposure × Post × Similarity Quartile 2			-0.044**	-0.013
			(0.020)	(0.023)
High VC Exposure × Post × Similarity Quartile 3			0.002	0.033
			(0.041)	(0.044)
High VC Exposure × Post × Similarity Quartile 4			-0.049***	-0.019
			(0.019)	(0.014)
Control Variables	No	Yes	Yes	Yes
High VC Exposure × Similarity Quartile	No	No	Yes	Yes
Post × Similarity Quartile	No	No	Yes	Yes
Similarity Quartile	No	No	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	Yes	Yes
Observations	108,092	108,092	108,092	108,092
R^2	0.52	0.52	0.52	0.52

Table F.9: **Binary treatment to estimate the effect on the likelihood that investments lead to breakthrough success**: This table reports results for regressions at the VC-investment level. The dependent variable is a dummy indicating whether the VC-funded company goes public in an IPO or is acquired for a higher value than the total VC investments in the company after the VC investment. The main independent variable is the interaction of High VC Exposure (a dummy indicating whether the VC firm's exposure is above the median) and Post (a dummy equal to one after 2006). Columns (3) and (4) include interactions of High VC Exposure, Post and dummy variables indicating the backward-similarity quartile of the corresponding funded startup. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01

G Additional figures

Figure G.1: **Word cloud of AI-related job titles at VC firms**: This figures reports a word cloud of the 30 most frequent titles associated with jobs I classify as AI-related in the universe of jobs at VC firms in the Crunchbase database.

Panel A:



Panel B:

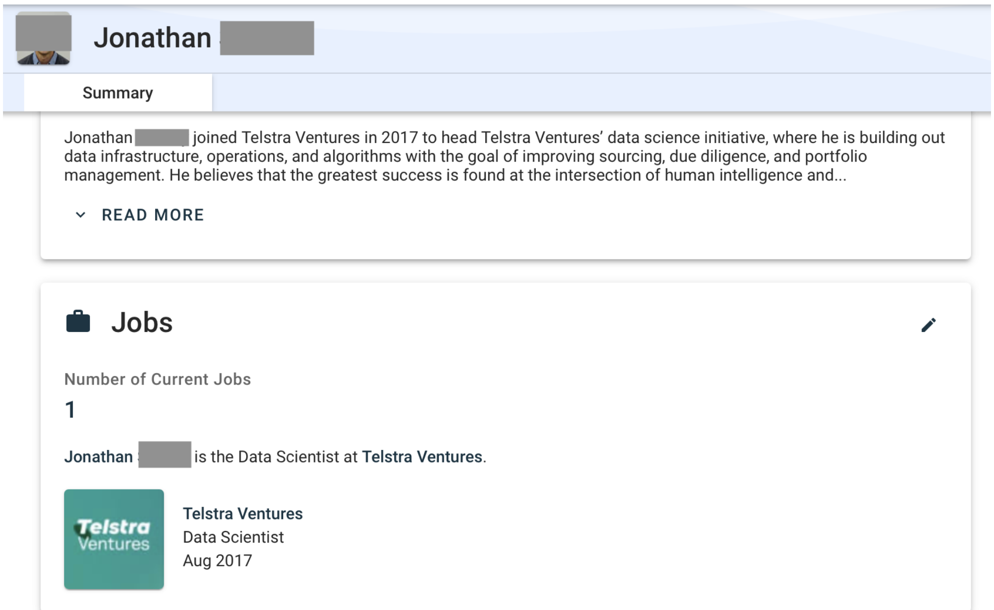
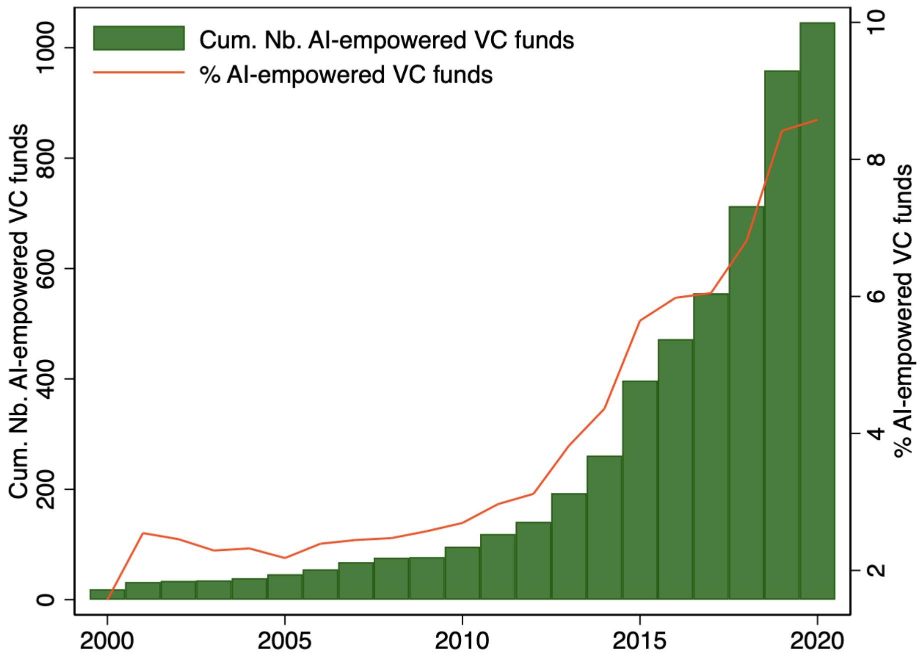


Figure G.2: **Profiles of employees with an AI-related job on Crunchbase**: This figure shows a screenshot of the Crunchbase profiles of two people that I classify as having an AI-related job in VC firms. The summary descriptions allow to establish unambiguously that each job relates to the implementation of data-driven tools that are used internally in the VC firm's screening operations. Names and pictures have been removed.

Panel A: Cumulative number and percentage of AI-empowered VC funds



Panel B: Cumulative AUM and percentage AUM of AI-empowered VCs

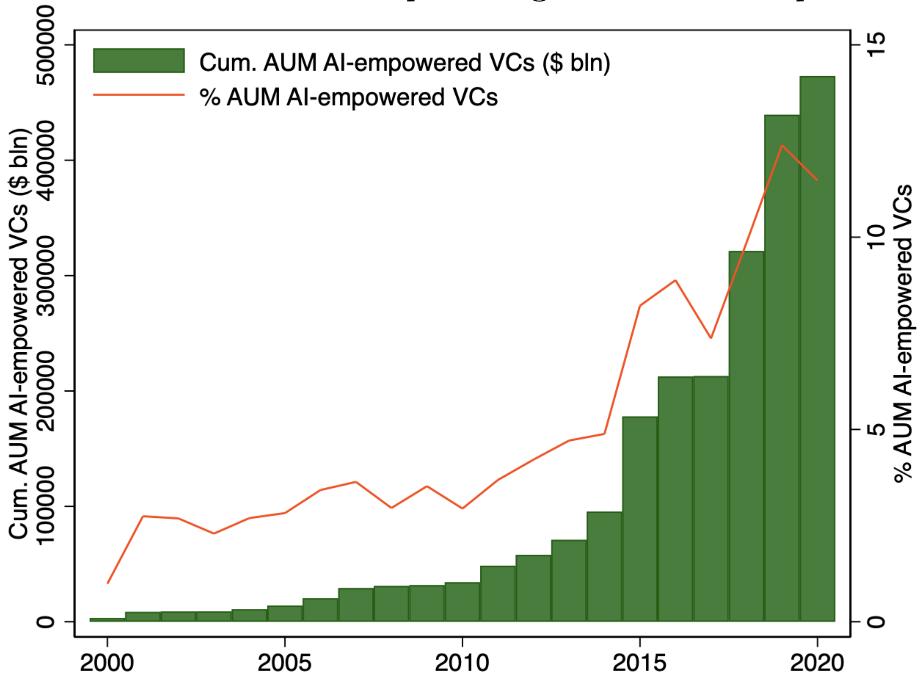
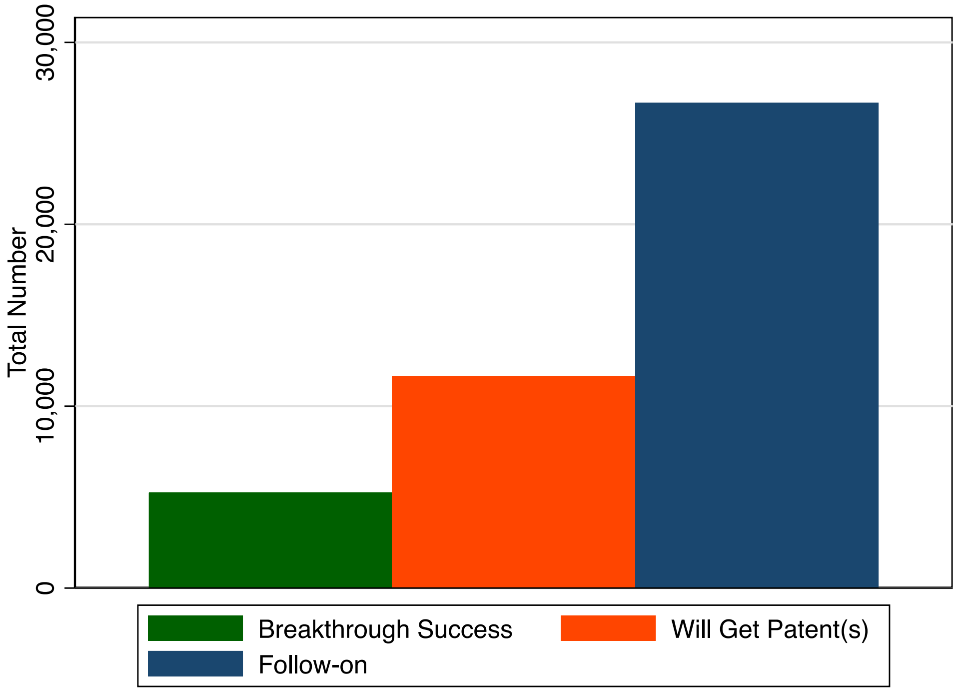


Figure G.3: **Evolution of the number of funds and AUM of VC firms classified as AI-empowered**: Panel A reports for each year between 2000 and 2020 the cumulative number and percentage of funds managed by VC firms classified as AI-empowered. Panel B reports over the same period the cumulative assets under management (AUM) and fraction of AUM of funds managed by VC investor classified as AI-empowered.

Panel A: Total number of outcomes



Panel B: Total number of outcomes in each quartile of backward-similarity

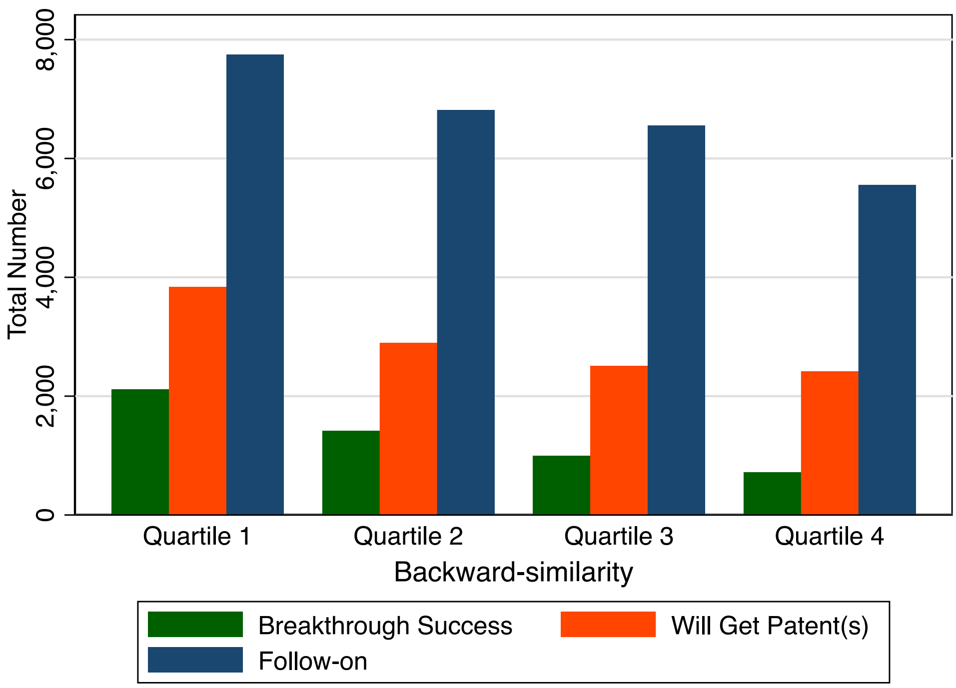


Figure G.4: **Total number of each type of outcome**: This figure shows the number of each type of outcome in my sample. I compute the total number of VC-funded startups that (i) achieve breakthrough success (i.e., go public in an IPO or get acquired for a higher value than the total VC investments in the company after the VC investment), (ii) eventual obtain a patent, and (iii) survive and receive follow-on funding by 2021. These numbers are displayed in the overall sample (Panel A) and in each quartile of backward-similarity (Panel B). "Quartile i" refers to the quartile of backward-similarity of the VC-funded startup when raising its first round of funding. for instance "Quartile 1" corresponds to the set of startups with backward-similarity percentile rank below 0.25.

H Additional tables

Rank	Name	Nb. AI jobs	Country	Nb. Investments	Nb. Funds
1	GV	8	USA	782	2
2	Andreessen Horowitz	6	USA	850	25
3	Lighter Capital	6	USA	21	1
4	EQT Ventures	5	SWE	89	1
5	InReach Ventures	5	GBR	17	1
6	BCG Digital Ventures	5	USA	10	1
7	Social Capital	4	USA	377	8
8	Correlation Ventures	4	USA	273	2
9	SignalFire	4	USA	92	6
10	645 Ventures	4	USA	57	3

Table H.1: **Top-10 AI-empowered VCs, based on their number of AI-related jobs**. This table presents the top-10 VCs classified as AI-empowered, based on the number of AI-related jobs over the sample period. "Nb. Investments" and "Nb. Funds" are respectively the total number of investments and total number of funds over the sample period 2000-2020.

	Similarity Rank (× 100)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Has Filed Patent Application	-0.534***												
	(0.128)												
New Patent Application	-1.106***												
	(0.242)												
Log(1+Nb. Past Patent Applications)		-0.100											
		(0.096)											
Log(1+Nb. New Patent Applications)		-0.467***											
		(0.138)											
Breakthrough Success			-0.981**		-0.903**								
			(0.431)		(0.410)								
Follow-on				-0.738***	-0.707***								
				(0.204)	(0.192)								
Log(Current Nb. Employees)						-0.072							
						(0.067)							
Log(Raised Amount)							-0.085						
							(0.131)						
Serial								0.001					
								(0.093)					
Top School								-0.175***					
								(0.044)					
PhD								0.011					
								(0.216)					
MBA								-0.137					
								(0.098)					
Female								-0.400***					
								(0.097)					
US Startup									-2.610***				
									(0.951)				
European Startup									-2.040**				
									(0.925)				
Top-5 Region										-0.220*			
										(0.120)			
Log(Company Age)											-0.142	0.683***	
											(0.097)	(0.098)	
Log(#Funding Round)													0.778***
													(0.094)
Industry × Year FE	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Industry × Country × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Company FE	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Observations	63,457	59,021	59,021	59,021	59,021	57,165	41,212	47,405	63,457	63,457	63,457	113,777	113,777
R^2	0.90	0.89	0.89	0.89	0.89	0.90	0.91	0.91	0.86	0.90	0.90	1.00	1.00

Table H.2: **Robustness test using only startups funded by VCs that do not adopt AI**: This table reports results for regressions at the funding round level, investigating the link between my backward-similarity measure and a set of startup characteristics. It is similar to Table 4 but only considers startups that are not funded by AI-empowered VCs. The dependent variable is the percentile rank in terms of backward-similarity. All regressions include industry-country-year fixed effects, except column (9) that includes industry-year fixed effects. Standard errors in parentheses are clustered at the industry and country levels.

	IPO				IPO or Acquisition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	-0.014** (0.007)	-0.016** (0.008)			-0.034*** (0.011)	-0.028** (0.012)		
AI-empowered × Similarity Quartile 2		-0.002 (0.006)	-0.001 (0.006)			-0.008 (0.011)	-0.005 (0.011)	
AI-empowered × Similarity Quartile 3		0.004 (0.006)	0.004 (0.006)			0.004 (0.011)	0.003 (0.011)	
AI-empowered × Similarity Quartile 4		0.001 (0.008)	0.001 (0.007)			-0.015 (0.011)	-0.008 (0.011)	
AI-empowered × Similarity Rank				0.006 (0.009)				-0.004 (0.014)
Similarity Rank				-0.022 (0.021)				-0.042 (0.032)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	Yes	No	No	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	No	No	Yes	Yes	No	No
VC × Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.56	0.56	0.62	0.62	0.56	0.56	0.62	0.62

Table H.3: **Alternative measures of breakthrough success**: This table reports results for regressions at the VC-investment level. The dependent variable in columns (1)-(4) is a dummy indicating whether the VC-funded company goes public in IPO. The dependent variable in columns (5)-(8) is a dummy indicating whether the VC-funded company goes public in IPO or is acquired. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. 'Similarity Quartile *i*" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Log(1+Nb. New Patent Grants)				Log(1+Nb. New Patents' Citations)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-empowered	-0.045*	-0.031			-0.048*	-0.033		
	(0.024)	(0.031)			(0.026)	(0.033)		
AI-empowered × Similarity Quartile 2		-0.021	-0.014			-0.013	-0.006	
		(0.022)	(0.023)			(0.024)	(0.025)	
AI-empowered × Similarity Quartile 3		-0.020	-0.013			-0.017	-0.011	
		(0.021)	(0.022)			(0.023)	(0.024)	
AI-empowered × Similarity Quartile 4		-0.013	-0.012			-0.021	-0.019	
		(0.031)	(0.031)			(0.034)	(0.032)	
AI-empowered × Similarity Rank				-0.013				-0.027
				(0.037)				(0.041)
Similarity Rank				-0.367***				-0.368***
				(0.082)				(0.086)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Similarity Quartile	No	Yes	Yes	No	No	Yes	Yes	No
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	Yes	No	No	Yes	Yes	No	No
VC × Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	221,765	221,765	221,765	221,765	221,765	221,765	221,765	221,765
R^2	0.60	0.60	0.65	0.65	0.58	0.58	0.64	0.64

Table H.4: **Additional measures of innovation**: This table reports results for regressions at the VC-investment level. In columns (1)-(4), the dependent variable is the logarithm of one plus the number of patents obtained by the startup with application date after the VC investment date. In columns (5)-(8), the dependent variable is the logarithm of one plus the number of citations received by patents obtained by the startup with application date after the VC investment date. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. 'Similarity Quartile i" refers to the quartile of backward-similarity of the VC-funded startup. for instance "Similarity Quartile 1" is a dummy indicating whether the startup's backward-similarity percentile rank is below 0.25. Similarity Rank is the backward-similarity percentile rank (between 0 and 1). Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors in parentheses are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

	Out VC Region(s)		Serial		Top School		Lead		Board	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AI-empowered	-0.039**		-0.014		-0.008		0.042***		-0.001	
	(0.020)		(0.013)		(0.011)		(0.016)		(0.019)	
AI-empowered × Similarity Rank		-0.037*		-0.007		0.026		-0.030**		-0.031
		(0.020)		(0.015)		(0.020)		(0.015)		(0.023)
Similarity Rank		-0.004		-0.025		-0.070*		-0.032*		-0.029
		(0.019)		(0.040)		(0.039)		(0.020)		(0.033)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Country × Stage × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VC FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
VC × Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	247,677	247,677	217,129	217,129	217,129	217,129	182,706	182,706	152,380	152,380
R^2	0.35	0.46	0.49	0.55	0.51	0.57	0.36	0.47	0.39	0.49

Table H.5: **Disentangling sourcing, screening and value-added**: This table reports results for regressions at the VC-investment level. In columns (1) and (2), the dependent variable is a dummy indicating whether the startup is located in a region that is not part of those where the VC has employees. In columns (3) and (4), the dependent variable is a dummy indicating whether the startup's founding team includes a serial entrepreneur. In columns (5) and (6), the dependent variable is a dummy indicating whether one of the founding team members holds a degree from a top school. In columns (7) and (8), the dependent variable is a dummy indicating whether the VC is the lead investor on the deal. In columns (9) and (10), the dependent variable is a dummy indicating whether the VC seats on the board of the startup. The main independent variable is a dummy variable equal to one if the investment is made by a VC classified as AI-empowered at that time and equal to zero otherwise. Similarity Rank is the percentile rank of the funded company's backward-similarity as of the funding date. For instance, Similarity Rank is equal to 0.9 if the funded company has a similarity with respect to previously funded companies in the top 10 percent. Control variables include the logarithm of the age of the VC firm, the logarithm of its current number of employees and the logarithm of the startup's age. All control variables are measured at the time the investment is made. Standard errors are double-clustered at the VC firm and startup levels. * p<.10; ** p<.05; *** p<.01.

Variable	Obs	Mean	Sd	5%	25%	50%	75%	95%
Panel A: Investment level								
VC Age (Years)	108,092	21.09	17.11	5.00	11.00	17.00	26.00	47.00
VC Nb. Employees	108,092	14.29	22.29	1.00	3.00	8.00	17.00	49.00
Company Age (Years)	108,092	5.96	4.97	1.00	3.00	5.00	7.00	15.00
Early Stage	108,092	0.33	0.47	0.00	0.00	0.00	1.00	1.00
First-round	108,092	0.22	0.41	0.00	0.00	0.00	0.00	1.00
Raised Amount (USD mln)	94,314	25.03	134.32	1.00	4.70	11.00	24.00	75.00
Nb. Investors	108,092	4.19	3.19	1.00	2.00	4.00	6.00	10.00
Lead	80,511	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Out of VC Region(s)	108,092	0.45	0.50	0.00	0.00	0.00	1.00	1.00
VC on Board	74,688	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Serial Entrepreneur	92,153	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Top School Founder	92,153	0.41	0.49	0.00	0.00	0.00	1.00	1.00
Female Founder	92,153	0.22	0.41	0.00	0.00	0.00	0.00	1.00
Follow-on	108,092	0.50	0.50	0.00	0.00	0.00	1.00	1.00
Breakthrough Success	108,092	0.21	0.41	0.00	0.00	0.00	0.00	1.00
IPO	108,092	0.12	0.32	0.00	0.00	0.00	0.00	1.00
IPO or Acquired	108,092	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Nb. New Patent Applications	108,092	11.76	32.83	0.00	0.00	0.00	5.00	70.00
Nb. New Patent Grants	108,092	7.92	22.74	0.00	0.00	0.00	3.00	47.00
Nb. New Highly-cited Patents	108,092	3.33	10.55	0.00	0.00	0.00	1.00	20.00
Nb. New Patents' Citations	108,092	9.71	31.02	0.00	0.00	0.00	2.16	59.69
AI-empowered	108,092	0.08	0.27	0.00	0.00	0.00	0.00	1.00
Similarity Rank	108,092	0.43	0.29	0.03	0.17	0.39	0.67	0.92
VC Exposure	108,092	0.57	0.24	0.13	0.40	0.60	0.76	0.91
Panel B: VC-Year level								
US VC Firm	26,096	0.60	0.49	0.00	0.00	1.00	1.00	1.00
European VC Firm	26,096	0.24	0.43	0.00	0.00	0.00	0.00	1.00
VC Age (Years)	26,096	19.02	19.04	3.00	9.00	15.00	22.00	44.00
VC Nb. Employees	26,096	6.27	12.70	1.00	1.00	3.00	7.00	21.00
Nb. Investments	26,096	4.39	8.97	0.00	0.00	1.00	4.00	19.00
Nb. Leads	26,096	1.38	3.49	0.00	0.00	0.00	1.00	6.00
Proportion in Software	18,579	0.39	0.37	0.00	0.00	0.33	0.67	1.00
Proportion in Heath Care	18,579	0.25	0.37	0.00	0.00	0.00	0.42	1.00
Proportion in Financial Services	18,579	0.07	0.19	0.00	0.00	0.00	0.00	0.50
Proportion in Information Technology	18,579	0.20	0.29	0.00	0.00	0.00	0.33	1.00
Proportion in Internet Services	18,579	0.20	0.28	0.00	0.00	0.00	0.33	1.00

Continued next page

Summary statistics (continued)

Variable	Obs	Mean	Sd	5%	25%	50%	75%	95%
Proportion in AI and Data	18,579	0.12	0.23	0.00	0.00	0.00	0.17	0.50
Nb. Funds	16,795	4.50	6.04	1.00	2.00	3.00	5.00	12.00
AUM (USD bln)	16,043	1.79	6.62	0.03	0.12	0.36	1.10	6.22
AI-empowered VC	26,096	0.02	0.13	0.00	0.00	0.00	0.00	0.00
Nb. Investments Similarity Quartile 1	26,096	1.43	3.17	0.00	0.00	0.00	1.00	6.00
Nb. Investments Similarity Quartile 2	26,096	1.13	2.40	0.00	0.00	0.00	1.00	5.00
Nb. Investments Similarity Quartile 3	26,096	0.91	2.41	0.00	0.00	0.00	1.00	5.00
Nb. Investments Similarity Quartile 4	26,096	0.91	3.12	0.00	0.00	0.00	0.00	5.00
VC Exposure	26,096	0.50	0.28	0.05	0.26	0.50	0.75	0.94

Table H.6: **Summary statistics: sub-sample of VCs with exposure to cloud computing**. This table presents descriptive statistics for the sample of VCs for which I am able to compute exposure to the advent of cloud computing (Amazon Web Services) before 2006. See Section 5 for details about the definition of VC exposure. Panels A and B present statistics at the VC investment level and VC-year level respectively.

Cloud computing glossary			
Amazon Web Services	AWS	App	Application Programming Interface
App Programming Interface	API	Backend-as-a-Service	BaaS
Backend database	Big Data	Cloud Application	Cloud App
Cloud Backup	Cloud Computing	Cloud Foundry	Cloud Management Platform
CMP	Cloud Marketplace	Cloud Migration	Cloud Native
Cloud Washing	Cloud Service Provider	CSP	Cloud Sourcing
Cloud Storage	Content Delivery Network	CDN	Customer Relationship Management
CRM	Data Migration	Data base	Database
DevOps	Enterprise Application	Enterprise App	Enterprise Resource Planning
ERP	Federated Database	Google Cloud Platform	GCP
Guest Machine	Host Machine	Hybrid Cloud	Hypervisor
Information technology infrastructure	IT infrastructure	Infrastructure as a Service	IaaS
Integrated Development environment	IDE	Linux	Load Balancing
Managed Service Provider	MSP	Microservices	Microsoft Azure
Middleware	Multi-Cloud	Multi-Tenancy	On-Demand Self Service
Open Source	Open Stack	Personal Cloud	Platform
Platform as a Service	PaaS	Private Cloud	Public Cloud
Service Level Agreement	SLA	Shared Resources	Software as a Service
SaaS	Software Development Kit	SDK	Software Stack
User Interface	UI	User Experience	UX
User Space	Vendor Lock-in	Vertical Cloud	Virtual Desktop Infrastructure
VDI	Virtual Machine	VM	Virtual Machine Monitor
VMM			

Table H.7: **Cloud Computing IT Glossary**: This table reports the cloud computing terms defined in the Cloud Computing IT Glossary from Solutions Review, a technology website gathering relevant content about cloud solutions (cf., solutionsreview.com/cloud-platforms/glossary).

Industry	Exposure	%Descriptions with match	Nb. Descriptions with match	Nb. Descriptions
Panel A: Most exposed industries				
SaaS	1.00	74.25	1,811	2,439
Open Source	1.00	71.90	261	363
Linux	1.00	68.53	98	143
Machine Learning	1.00	67.95	352	518
IaaS	0.99	60.91	67	110
Artificial Intelligence	0.99	60.81	647	1,064
PaaS	0.99	57.89	66	114
Database	0.99	52.51	1,036	1,973
Big Data	0.99	51.65	721	1,396
Cloud Computing	0.98	49.62	1,567	3,158
Panel B: Least exposed industries				
Fuel	0.07	0.59	3	506
Packaging Services	0.07	0.56	19	3,385
Commercial Insurance	0.07	0.54	7	1,286
Civil Engineering	0.07	0.54	23	4,274
Property Development	0.06	0.49	11	2,246
Plastics and Rubber Manufacturing	0.06	0.46	10	2,180
Resorts	0.06	0.41	2	483
Retirement	0.06	0.37	7	1,897
Laundry and Dry-cleaning	0.05	0.29	1	347
Water Purification	0.05	0.24	2	843

Table H.8: **Industry exposures to cloud computing**. This table documents the exposure of industries to the introduction of cloud computing. This proxies for the extent to which startups in a given industry are likely to benefit from the introduction of Amazon cloud services (AWS). For each industry, I use the business descriptions of companies existing before the introduction to characterize the degree to which AWS was likely to be beneficial for new startups in that industry. Specifically, in each industry I measure the fraction of firms existing before 2006 whose business description mentions at least one of the keywords from the cloud computing glossary provided by Solution Review, to provide an industry-level exposure to the introduction of cloud computing. Panels A and B present respectively the 10 industries with the largest and lowest fraction of descriptions with at least one keyword match. All industries presented in the Table received at least 20 VC investments over the sample period 2000-2020.

	AI-empowered VC	
	(1)	(2)
VC Exposure × Post	0.016**	0.015**
	(0.007)	(0.007)
Control Variables	No	Yes
VC Country × Main Stage × Year FE	Yes	Yes
VC FE	Yes	Yes
Observations	26,096	26,096
R^2	0.77	0.77

Table H.9: **Estimation at the VC-year level of the effect of VCs' exposure to cloud-related industries on their likelihood to adopt AI and their number of investments**: This table reports results for regressions using VC-years observations. The dependent variable is a dummy variable indicating whether the VC is classified as AI-empowered in a given year. The main independent variable is the interaction of VC Exposure (the VC firm's exposure to the introduction of cloud computing based on its portfolio composition before the shock) and Post (a dummy equal to one after 2006). Control variables include the logarithm of the following variables at the VC firm level as of the year: the age of the VC firm and the number of current employees. "VC Country" and "Main Stage" denote the country where the VC firm is headquartered and the stage of funding at which the VC firm invests the most over the sample period (among six categories, i.e., Pre-Seed, Seed, Series A, Series B, Series C, Series D and onward) respectively. Standard errors are double-clustered at the VC firm and years levels. * p<.10; ** p<.05; *** p<.01.