Corporate Venture Capital and Firm Scope

Yifei Zhang*

Job Market Paper

Current Version: August, 2021

Click Here for the Latest Version

ABSTRACT

This paper studies whether and how corporate venture capital (CVC) spurs changes in firm scope. Using two sets of firm scope metrics, a text-based emerging business measure and Compustat segment measures, I document that CVC investments are strongly associated with subsequent firm scope changes of the CVC corporate parent, including seeding emerging businesses, establishing new divisions, terminating obsolete divisions, and changing the primary industry. Further evidence is consistent with an experimentation view of CVC investments, with more promising ventures having a stronger impact on the scope change of parent firms. Finally, to sharpen the causality, I explore the idiosyncratic fund inflow shocks of those connected independent VCs in each CVC program, as well as the US non-stop airline routes.

Key Words: Corporate Venture Capital (CVC); Firm Scope; Emerging Businesses; Experimentation; VC Network; Textual Analysis; 10-K Business Description

JEL Classification: G24, G32, G34, O32, L22

---

*Affiliation: Toulouse School of Economics. Please address any comments to yifei.zhang@tse-fr.eu. This project is started at Yale University, School of Management, where I have an academic visit during the Fall semester of 2019. I am indebted to Ulrich Hege for his guidance and suggestions. I want to thank Milo Bianchi, Matthieu Bouvard, Catherine Casamatta, Patrick Coen, Lin William Cong, Florian Ederer, Laurent Frésard, Caroline Genc (discussant), Yassine Lefouili, Xiaojuan Liu, Song Ma, Andrew Metrick, Sophie Moinas, Filip Mrowiec, Sebastien Pouget, Konrad Raff, Armin Schwienbacher (discussant), Kelly Shue, Olav Sorenson, Xiang Zheng (discussant), as well as conference and seminar participants at the 3rd Paris Dauphine Finance PhD Workshop (2020), the 17th Corporate Finance Day (2020), the Econometric Society European Winter Meeting (2020), Wharton Innovation Doctoral Symposium (2021), Eastern Finance Association (2021), European Finance Association – Doctoral Tutorial (2021), and TSE for their helpful comments. I also want to thank the American Finance Association (2022) to include this paper into the program. All errors are mine.

I. Introduction

Understanding the firm scope and the boundary of the firm is a central topic in economics and finance.[1] However, there is little empirical work on the determinants of firm scope.[2] Perhaps even more surprisingly, despite the prominence of the "Schumpeterian" view that innovation is the key driving force behind the growth and evolution of firms and economies, there is almost no work on the relationship between corporate innovation strategies and the dynamics of firm scope. Therefore, this paper contributes to the empirical literature of firm scope by studying an open corporate innovation strategy, the corporate venture capital investment, popular among large industry leaders.

A corporate venture capital (CVC) program is a venture capital arm affiliated with an established firm, and CVC has grown in the last decades to become an important tool in the open innovation strategies of many leading companies, including tech giants, such as Apple, Google, and Microsoft. Thus, investigating the impact of CVC on the scope change of those leading firms is of great importance for both academia and practitioners. Moreover, given its special characteristics, CVC offers a unique opportunity to discover a new firm growth (and scope change) strategy, the experimentation strategy, which lies at the heart of this paper.[3]

More specifically, this paper asks whether, and more crucially, how CVC spurs the firm scope change. I start to introduce some anecdotal and survey evidence documenting that establishing a CVC program could help its parent firm (like GE) identify new business opportunities.[4] Given the newly identified business opportunities, a CVC parent firm will

[1]Important milestones in the theory of the firm include the transaction cost theory (Coase (1937) and Williamson (1985)) and the property rights approach (Grossman and Hart (1986) and Hart and Moore (1990)).

[2]As argued recently in Hoberg and Phillips (2018), the traditional conglomerate literature takes the firm scope as given and seldom explores the determinants of firm scope.

[3]Compared with other instruments that firms have at their disposal to foster innovation (e.g., in-house efforts to carry out R&D and create new intellectual property, acquisitions of research results or innovative startups, the recruitment of employees with new expertise), CVC offers the advantage that firms initial investment decisions, as well as metrics of investment outcomes, can be observed, and hence offers an exciting view on the use of experimentation in firm strategy.

[4]The evidence and surveys are documented in Section II. However, the precise mechanism by which it happens is not documented in the survey or the anecdotal evidence.

naturally integrate those new businesses into its current business domain, thus reshaping its firm scope. Further evidence is consistent with an experimentation view of CVC investments, with more promising ventures having a stronger impact on the scope change of parent firms. The organization of findings in the paper is illustrated in Figure 1.

I use two ways to gauge the firm scope change. First, I leverage on the textual analysis in defining emerging businesses and count how many emerging businesses are newly added to each public-listed firm's annual 10-K business description. The emerging businesses are proxied by "emerging phrases", top 5% most popular business word pairs taken from VC-backed startups' business descriptions. (See Figure 3 for a quick view.) Second, I use Compustat Segment data to construct variables for scope change, including establishing new divisions (segments), terminating obsolete divisions, and changing the corporate primary industry.[5]

In both cases, I find that CVC investments are strongly associated with the subsequent change of the firm scope. More specifically, a CVC parent, on average, adds 1.5 (100%) more emerging phrases into the 10-Ks than those industry-year peers without CVC investments within the two years after CVC deals. Moreover, firms are around 60% more likely to establish a new division (operating in a new industry) and around 35% more likely to remove an old division within the next two years following the CVC investment decision.[6]

Having documented these basic facts, I turn to scrutinizing the channel through which CVC helps identify new business opportunities and ultimately spurs the firm scope change. The empirical evidence is consistent with an experimentation view of CVC investments. In this view, each CVC deal could be regarded as an experiment that creates a real option for a potential new line of products or activities (Keil, Autio, and George, 2008). Through interacting with the startup managers and participating in the startup's operation in the deal,

[5]These two methods are actually complementary. While Compustat segment measures capture larger changes, they are coarse in terms of measuring the scope dynamics. The text-based emerging business measure is more granular and captures popular businesses not only new relative to the CVC parent firm but also new relative to the whole US economy.

[6]The corporate primary industry change takes effect over a longer horizon: within three to five years after investments.

a CVC parent firm could receive valuable information (called it a signal in this paper) about the future potential of this new business. Moreover, these signals help to identify promising and new business opportunities and to avoid those business "traps". The experimentation strategy is a logical response to identify good business opportunities under huge uncertainty in the VC industry (Ewens, Nanda, and Rhodes-Kropf, 2018).

There are two sets of evidence supporting the experimentation hypothesis. First, under the high uncertainty in the VC industry with numerous good and bad business opportunities, diversifying CVC deals across industries and business areas is a necessary step in experimentation and in the process of finding out promising business opportunities. Consistent with the hypothesis, diversifying investment strategy is quite popular among CVC programs, and the more diversifying the program's investment strategy is, the more likely its parent firm conducts scope changes. It implies that diversifying helps to find out great business opportunities. Second, I test the signaling and winner-picking. I estimate two discrete choice models (McFadden, 1973) for the industry choice of establishing a new division and seeding emerging business, respectively. I find that, conditional on CVC investments, receiving a good signal from the invested startup is strongly associated with the choice of establishing a new division or adding emerging phrases in the startup's industry, where the signal is only observable after CVC deals but not before.[7]

Adding to established evidence that firms experiment for future growth directions through CVC, I then sharpen the causality between CVC and firm scope and exclude an alternative story for my results: it is the business opportunity that drives CVC investments and the further firm scope changes.

I introduce a new instrument for CVC investments using the fund inflow shocks of independent venture capital firms (IVC) in each CVC program's past syndicate network.[8] The

[7]It is challenging to measure those signals since they are private information from startups to CVC firms. In the empirical exercises, I use the startups' IPO, acquisition, and patent growth information as the proxy of signals.

[8]Independent venture capital firms (IVC) are simply the traditional VC firms with a limited partnership as the organizational form. I call them IVC to distinguish them from CVC. Furthermore, IVC terminology is widely used in the CVC literature, for example, Chemmanur, Loutskina, and Tian (2014) and Ma (2020).

idea of the instrument is that, if an independent VC firm j (IVC j) receives a positive fund inflow shock today, and meanwhile, the CVC Firm i is in the IVC's past syndicate network, then the IVC j is very likely to initiate new deals and invite CVC Firm i, its old partner, to join in its new investments.[9] Alternatively, IVC j could simply recommend some deals to the connected CVC.[10]

Notably, the fund inflow shocks are idiosyncratic across IVCs, orthogonal to any VC industry investment opportunities and any technology shocks. I construct my instrument following the recent Granular IV (GIV) approach developed in Gabaix and Koijen (2020). More precisely, the GIV is the sum of the idiosyncratic fund inflow shocks of those IVCs in the past 5-year syndicate network of each CVC program.

The instrument works on a small sample of US listed firms already starting the foray of CVC investments in the past and thus enjoying some IVCs network today. In both the first and second stage regressions involving these CVC firms, I control for their past IVC network size and IVC characteristics (average age and past IPO performance) in the network, as well as the past three-year CVC investments. I document that my instrument strongly predicts the continuation of investments by a CVC program. In the second stage, I find that the number of CVC deals positively predicts the firm scope change measured by both the emerging phrases and Compustat segment dummies.

The paper is related to three broad strands of the literature. First, this paper is related to canonical literature regarding the firm scope, dating back to Teece (1980) and Panzar and Willig (1981) in economics and Lang and Stulz (1994), Berger and Ofek (1995), and Lamont (1997) in finance. Recently, Hoberg and Phillips (2021) documents that the 21st century US firms usually expand their businesses across related industries and thus are immune from the

[9] Here is an example of the invitation. Between 1994 and 2000, Cisco Systems (a large industrial firm) was invited into 13 syndications led by Sequoia Capital (a pure VC firm), as documented in Ferrary (2010).

[10] The idea is based on previous findings in VC and CVC literature: (1) syndicating network is crucial in the VC world, and many IVCs invite their old partners in the previous syndicate network to join in their new deals (Hochberg, Ljungqvist, and Lu, 2007, 2010; Keil, Maula, and Wilson, 2010); (2) IVC is the largest deal source of most CVC programs. IVC usually recommends deals to CVC or offers "deal flow" to CVC (MacMillan, Roberts, Livada, and Wang, 2008).

well-known diversification discount. In this paper, I find a new mechanism of scope change by US-listed firms through CVC experimentation.

Second, the paper contributes to the experimentation literature in entrepreneurial finance, such as Manso (2016) and Ewens et al. (2018). Specifically, Ewens et al. (2018) find that recent VC firms adopt a new experimentation strategy in their investments, so-called the "spray and pray" strategy, especially after the cost of starting software and internet-related ventures drops significantly. There is one key difference between their experimentation and my experimentation – the goal. While VC firms do experimentation to search "unicorns", CVC firms aim to figure out the optimal growth direction for company's future.

Third, the paper contributes to the VC literature, and more precisely, the CVC literature pioneered by Gompers and Lerner (2000) and Hellmann (2002). Previous literature documents that established firms in more competitive industries (Fulghieri and Sevilir, 2009; Kim, Gopal, and Hoberg, 2016), in industries with higher technology uncertainty (Basu, Phelps, and Kotha, 2011) and low intellectual property protection (Dushnitsky and Lenox, 2005), with lower institutional ownership (Tian and Ye, 2018), and firms experiencing deterioration of internal innovation (Ma, 2020) are more likely to conduct CVC investments.[11] I complement the aforementioned studies by relating CVC to the firm scope change of its parent corporation.

The rest of the paper proceeds as follows. Section II introduces the background of CVC and develops the hypotheses; Section III describes the data and summary statistics; Section IV provides the basic facts between CVC and firm scope changes; Section V studies CVC as an experimentation process; Section VI provides the identification between CVC and firm scope changes. The last section concludes.

[11] Recent new papers on CVC include Hamm, Jung, and Park (2018) and Shan (2019).

II. Background and Hypothesis Development

This section starts with the institutional background of corporate venture capital (CVC). A CVC deal is formally defined as a minority equity investment by an established corporation in a privately held entrepreneurial company (Dushnitsky, 2012). Alternatively, one could interpret a CVC program as the venture capital arm affiliated with the established corporation (such as Google Venture affiliated with Google).

CVC departs from traditional VC firms mainly in three key respects. First, whereas the traditional venture capital firm solicits funding from prospective limited partners, the funding for the investment of a CVC program mostly comes from its unique corporate parent. Second, around two-third of CVC programs do not have a dedicated fund structure (fixed fund period length); instead, they are more akin to "discretionary" or "evergreen" funds: they invest when the investment opportunities arrive. (MacMillan et al., 2008).

Third, the most important one, although seeking financial returns remains an essential objective, in most cases, CVC also seeks strategic goals for its corporate parent, such as identifying new technology, seeking new growth opportunities, and importing innovation with existing business units (Siegel, Siegel, and MacMillan, 1988; MacMillan et al., 2008). Regarding the strategic goals, the CVC literature has reached a consensus about its importance. Chesbrough (2002) argues that if CVC investments are uncoupled from the corporate strategy and its operating capabilities and are only justified by the prospect of financial gains, then the shareholders of CVC parent could invest in VC by themselves, and the CVC investments lose the meaning.

Strategic management scholars have conducted various interviews and surveys toward CVC managers worldwide to understand those strategic goals.[12] There are two strategic objects frequently appearing in those survey reports. The first one is to offer a window for new technology (open innovation); the other one, which is more likely to be neglected, is that CVC investments could help identify new business opportunities for its corporate parent.

[12]For the summarization, see Dushnitsky (2012).

For example, in a recent survey of 48 large CVC programs conducted by the National Venture Capital Association (NVCA), more than half of CVC managers report that identifying the new market and new business direction are critical strategic aims of the programs (MacMillan et al., 2008). Other survey evidence supporting the CVC objective in finding new business opportunities lies in Winters and Murfin (1988), Sykes (1990), McNally (1997), and Ernst and Young (2009). This paper's hypothesis is thus developed on this second important strategic objective.

As illustrated in Figure 1, I argue that CVC programs and CVC investments could help firms identify new business opportunities. After identifying an opportunity (say an emerging business), a CVC parent firm will naturally integrate the emerging business into its current business, thus changing the firm scope. I use two distinct but complementary methods to capture those changes in firm scope: on the one hand, I use the textual analysis to identify emerging businesses in the US economy and further gauge the business integration of those emerging businesses by CVC parents through the SEC annual 10-K filings; on the other hand, I apply the traditional Compustat Segment dataset in measuring the firm scope change (see Figure 1).

A natural follow-up question is about how CVC helps to identify new business opportunities (see the question mark in Figure 1). In this paper, I argue that it is a learning-through-experimentation story. More generally, this story is in line with the "long-shot bets" feature of VC investments documented in the literature with very few "unicorn" startups reaching big success (Bergemann and Hege, 2005).

The experimentation story postulates that CVC allows the corporate to experiment various business opportunities in which case the manager is uncertain about their final results. Each CVC deal thus could be regarded as an experiment that creates a real option for a potential new line of products or activities (Keil et al., 2008). Through interacting with the startup managers and participating in the startup's operation in the deal, a CVC parent firm could receive valuable information (called it a signal in this paper) about the future potential of this new business. Crucially, the signal could contain both soft and hard information, which is not available without investments and interactions with the CVC-backed startups.[13] A CVC parent finally pins down the best business option through signals from multiple experiments.

In the strategic management literature, this view is supported by Keil et al. (2008) who conducted several interviews toward CVC programs senior managers and argue that CVC is a process of "disembodied experimentation" in learning the knowledge from CVC-backed startups. In Keil et al. (2008)'s interview, a CVC manager recalled that

> If the [venture] turns out to be something important, you have to put in your own machines (page 1485). Sometimes we just speak up and say: 'That will never work. I have seen it! Guys, that's complete nonsense, I have seen the total opposite [faliure] here in a start-up.' (page 1490)

This view is also a good application of Ewens et al. (2018)'s experimentation theory. Ewens et al. (2018) document that many VCs start to conduct a "spray and pray" strategy in response to the reduced cost of initiating businesses in the software and internet-related industry. More specifically, VCs spray their deals to more ventures in the early investment stage and also abandon more when they receive bad signals from startups. Interestingly, the software and internet-related industry is the sector with the most intensive CVC deals.

The view is also supported by Lerner (2012) who argues that CVC has the function of leveraging limited resources to pursue or test a variety of technology options. Its cost-saving function is crucial when an established firm needs to test a large number of technology options. CVC also helps to quickly pull the plug of unpromising initiatives in the experiments, while the inside project will never stop optimally as the internal R&D manager has a strong incentive to hide unfavorable signals (Seru, 2014).

[13] Keil et al. (2008) argues that the knowledge regarding emerging business from CVC-backed startups are usually non-codified or colloquial information. Accessing the knowledge is possible only if the firm accesses to the community, the VC industry.

Two empirical predictions could be derived from the experimentation story. First, given the purpose of experimentation to solve the future uncertainty and figure out the best growth opportunity, a CVC program should necessarily diversify its deals across technology fields and industries. Therefore, we should expect that those CVC programs with higher diversification strategies are more likely to pin down a good business opportunity and ultimately lead to business changes (firm scope changes). Second, firms should conduct "winner picking" in choosing the industry in which they start a new business. Specifically, they should only establish a new division in an industry where they receive positive signals from CVC experiments. Similarly, when a negative signal arrives, they should quickly pull the plug and avoid the business traps.

III. Data and Sample Selection

A. *CVC Sample*

The raw data of my CVC sample is extracted from the Thomson Reuters SDC VentureXpert Database. Following Chemmanur et al. (2014) and Ma (2020), I start with a list of 1,248 US corporate-affiliated venture capital firms as reported by VentureXpert.[14] I then manually link these CVC program names with historical names of CRSP and Compustat firms (provided by WRDS) by checking various sources from Google, Factiva, LexisNexis, and PitchBook. This step helps me to identify the corporate parent(s) of each CVC program. As VentureXpert sometimes mislabels the CVC program as IVC or other types, I conduct an extensive search among all VC types following Hellmann, Lindsey, and Puri (2008) and supplement the above beginning CVC list with extra 35 CVC firms. Taken together, I obtain 623 unique CVC firms (programs) affiliated with either CRSP firms or Compustat firms from

[14] In detail, these VCs are Non-Financial Corporate Affiliate or Subsidiary Partnership, Venture/PE Subsidiary of Non-Financial Corporation, Venture/PE Subsidiary of Other Companies NEC, Venture/PE Subsidiary of Service Providers, Direct Investor/Non-Financial Corporation, Direct Investor/Service Provider, SBIC Affiliate with Non-Financial Corporation, and Non-Financial Corporate Affiliate or Subsidiary. In addition, I require the VC firms located in the US.

1980 to 2017.

In the next step, I impose extra filters on these 623 CVC firms/programs by limiting that (1) the corporate parent of CVC is incorporated in the US and is not operated in financial industries (SIC code starting with 6); (2) only a single corporate parent is matched to the CVC; (3) the CVC program is not initiated by financial division(s) of company, such as GE Capital Equity Group or Exxon Pension Fund, as these CVCs only seek financial goals but no strategic goals in their investments.

The final CVC sample contains 497 CVC programs and 448 unique public corporations, investing in CVC at least once in the sample period, with 11,300 total deals. Finally, VentureXpert also provides the 4-digit primary SIC code for each startup. The SIC code allows me to match each CVC deal to the Compustat Historical Segment Database (by the SIC-3) and further sort each deal into an unrelated or related deal. The unrelated CVC deals, those not related to the corporate parent's existing divisions or businesses, account for about 52% of total CVC deals sample.

Figure 2 plots the annual aggregate CVC investments by US public (non-financial) corporations in Compustat database. Investments are measured by (1) the number of deals (left axis) and (2) the fraction of deals among all VC deals (right axis).

B. Sample for Firm Scope Change

B.1. Textual Data on Emerging Business

To obtain a textual measure capturing time-varying emerging businesses in the US economy, I combine two text sources from VentureXpert and from the SEC digital filing system (EDGAR).

First, I download the detailed business description for each US-based VC-backed startup from VentureXpert.[15] I group a set of startups' detailed business descriptions into a yearly

[15] One caveat of this approach is that those startups' business descriptions are not historical but are updated to the date of data downloading.

single corpus, where the set contains all active VC-backed startups receiving VC funding in the given year. I drop common words and stop words and form word pairs for any two adjacent words.[16] Next, I define each year's "emerging phrases" set as those word pairs which are popularly used by the VC-backed startup community during that year. More precisely, I select the top 5% most frequently-used word pairs from the yearly startup corpus. Around 30 word pairs representing too general business (for example, "business service" or "product service") are excluded manually.

Ideally, each emerging phrase represents an emerging business that is popular among the startup community. I call them "emerging phrases" under the implicit assumption that any popular businesses in the VC industry should be novel and emerging relative to the businesses of US listed firms. Figure 3 shows two examples of emerging phrases set in 2000 and 2017 using the word clouds. As shown in Panels A and B, emerging phrases significantly evolve over time. The emerging phrase often relates to "internet" and "e-commerce" during 2000 (the internet bubble period), while, in the most recent year of my sample, more "tech buzzwords" are included, such as artificial intelligence, virtual reality, online platform, and digital health. More words clouds of emerging phrases are plotted in the online appendix.

On the other hand, I obtain the US public listed firms' business descriptions from the annual 10-K filings following Hoberg and Phillips (2016). First, I download all the 10-K filings from the SEC EDGAR system with Python automation scripts. Following Hoberg and Phillips (2016), I extract the Item 1 (Business Description) as my text source of firms' business in my regression sample, including those 450 US firms with CVC investments and other firms without CVC deals.

In the main analysis, I search the emerging phrases in each 10-K filing in order to identify the emerging business that is newly integrated by CVC parents as well as by other public listed firms with no CVC investments. The detailed procedures are in Section IV.A.

[16] The online appendix lists those common words and stop words. Stop words are mainly from the NLTK.

B.2. Compustat Segment Data

To obtain traditional measures of firm scope changes, I begin with all firms and their segments listed in Compustat Historical Segment Database for 1980-2017.[17] For each segment observation, I require that the segment primary SIC code (item $SICS1$) is not missing, as well as the segment sales being non-negative. Next, within each firm, I aggregate sales and total assets of raw segments by SIC-3 industries. I refer to the resulting firm-industry-year observations as the divisions. In other words, a division is equivalent to an operating SIC-3 industry of the firm.

Next, I construct two dummies, namely dummies of creating a new division and of removing an obsolete division, for the firm scope change, using the divisions reporting information of each firm. Establishing a new division is identified if the firm reports a new division with its SIC-3 code appearing in the first time in the company history. In other words, a division creation means that the firm steps into a new industry for the first time. Similarly, removing an old division means that a firm stops reporting a division with a certain SIC-3 code and this division is never reported again.[18] Finally, the primary corporate industry is defined as the industry of the division with the largest sales in the year.[19]

Of course, I am fully aware of the potential drawback of using Compustat Segment Data to capture the business change. As recently argued by Hoberg and Phillips (2021), firms tend to under-report their true industry scope in the Compustat Segment Database. Moreover, the recent SFAS 131 regulation change in 1997 requires that managers report segments based on how managers themselves internally evaluate operating performance (management approach). Prior to this rule change, segment reporting was instead based on an industry

[17]Following the recent conglomerate study by Matvos, Seru, and Silva (2018), I include only business segments during the data retrieving and further keep only observations of unrestated segments by choosing the SRCDATE that exactly matches to DATADATE.

[18]In the creation and removal of the division, I do not consider the temporary change (turnover) to reduce the noise of the measure.

[19]I do not choose the Compustat company-level historical SIC code in gauging the industry change since, in the full Compustat sample (not CRSP/Compustat), the item $SICH$ is missing for about one-third of the firm-year sample.

approach. Therefore, the newly established division before and after 1997 might not be comparable, and we should expect more new reported divisions after the 1997 fiscal year.

To tackle these different challenges, I split my sample before and after 1997 and conduct robustness checks for the two subsamples. Furthermore, I replicate and obtain Hoberg and Phillips (2021)'s textual segments (divisions) based on the overlap between 10-K Item 1 business description and SIC (NAIC) industry description. The textual segments are immune from the criticism as mentioned above. In the online appendix, I show that my results do not change if I use the textual-based segments to construct dummies of the firm scope change.

C. Other Data Sources and Summary Statistics

As discussed in the introduction, this paper uses two different identification strategies to estimate CVC's impact on the firm scope change: the fund inflow shocks of independent venture capital firms and the US non-stop airline routes. First, I obtain information regarding independent venture capital firms again from SDC VentureXpert. I measure and proxy the fund inflow shocks using data from SDC VentureXpert and PitchBook. Second, the US non-stop airline routes are downloaded from the T-100 Domestic Segment Database in the United States Department of Transportation. Furthermore, I get the distribution of US metropolitan statistical areas (MSA) from the US Census Bureau.

Table I Panel A presents the summary statistics of the firm-year sample, the sample used in most regression analysis. Following CVC literature (Ma (2020)), industries (3-digit SIC) with no CVC activities during the whole sample period are excluded entirely. All variables (except dummies) are winsorized at the 1% and 99% levels. Key variable constructions are illustrated in the Appendix A. As shown in Table I, firms with CVC investments are larger in firm size, are more profitable measured in ROA, and more likely to be a conglomerate.

IV. CVC and Change of Firm Scope

As a first step, I provide some strong suggestive evidences about CVC and the firm scope change. Section IV.A gauges the firm scope change using the text-based "emerging phrases", while Section IV.B captures it with the traditional Compustat Segment measures. Lastly, I leave the identification for causality in the next section.

A. *Evidence on Emerging Business*

Given the hypothesis development in Section II, it is ultimately an empirical question whether CVC leads to the firm scope change. One way to capture firm scope change is to gauge the emerging business integration by CVC parents as well as their industry peers using textual analysis. I estimate the following regression,

$$EmergingPhrases_{i,t+1} = \beta D(CVC)_{i,t} + \gamma \mathbf{X}_{i,t} + \upsilon_t \times \iota_j + (\tau_i) + \varepsilon_{i,t} \tag{1}$$

where $EmergingPhrases_{i,t+1}$ denotes the number of "emerging phrases" – those top 5% business word pairs popular among the startup community (see Figure 3) – that are newly added into the Firm i's 10-K Item 1 (business description) in Year t+1. (By saying "newly added", I mean that the phrases are found in Year t+1's 10-K but not in Year t's 10-K.) Intuitively, this measure captures newly added businesses that are new relative to the CVC parent's old business and also new to the whole US economy.[20] D(CVC) is a dummy equal to 1 if the firm conducts at least one CVC deal in Year t. Firm-level controls ($\mathbf{X}$) include Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate), Firm Age, as well as two mechanical textual measures: the number of any new word pairs appearing in the 10-K Item 1 and the total length of 10-K Item 1.

Before delving into regression results, I illustrate the regression design in Figure 4. Take

[20]Those popular businesses in the VC-backed startups' community should be new and emerging. Otherwise, there is no reason that VC will invest in those companies. The emerging characteristic could be a new technology, a new business model, or a new industry or product.

Google as an example. Suppose the Google CVC program (Google Venture) invests in startups in 2016, and during that year, the set of emerging phrases includes virtual reality, digital health, and smart home. Then I search in Google's 2017 10-K these three emerging phrases. The dependent variable thus counts the number of 2016 emerging phrases newly added in 2017's 10-K. The intuition is that, when Google invests in CVC in 2016, it helps Google identify new business opportunities such as digital health, and one year after investment (2017), Google should be more likely to add it into its own business.

Table II corroborates that CVC investments are strongly correlated with adding new emerging phrases. As shown in Column (1) (Column (4)), on average, a CVC parent will add 0.78 (0.68) more emerging phrases compared with its industry-year peers in the next year (second year) after investment. This amount of increase translates into 100% of the sample average (0.75). In Columns (2) and (3), I split D(CVC) into two dummies, D(CVC Unrelated) and D(CVC Related), according to whether the startup in the deal is related to the parent firm's current business. The regressions suggest that both related and unrelated deals could lead to the emerging business integration, with the effect being stronger for related ones in Year t+1. Finally, the result is very robust with firm fixed effects.

Figure 5 turns to scrutinizing the new "emerging phrases" usage in the years around each CVC deal. The point estimates (from OLS) and confidence intervals are taken from the following regression specification,

$$EmergingPhrases_{i,t} = \sum_{k=-3}^{+5} \gamma_k D(CVC\ Unr;k)_{i,t} + \sum_{k=-3}^{+5} \alpha_k D(CVC\ Rel;k)_{i,t} + \beta\mathbf{X} + \tau_i + \upsilon_t + \varepsilon_{i,t} \tag{2}$$

$EmergingPhrases_{i,t}$ simply counts the number of newly added emerging phrases in Year t's 10-K. $\{D(CVC\ Unr;k)\}_{k=-3}^{+5}$ denotes a set of nine dummies in the [-3, +5] year window around each unrelated CVC deal. As an example, $D(CVC\ Unr;+3)$ is equal to 1 if the firm-year observation is the third year after an unrelated CVC deal. A similar set-up applies to $\{D(CVC\ Rel;k)\}_{k=-3}^{+5}$ for CVC related deals. τ_i and υ_t are firm and year fixed effects, respectively.

Figure 5 shows that the "treatment" effect of CVC deals mainly lies within the two years after investments, whereas there is no significant effect before and three years after the investment. This is intuitive since old CVC deals (say the deals in 2010) could not help identify any current emerging business opportunities in 2014, and neither do future deals in 2018. In other words, CVC deals perform the best in helping identify contemporaneous business opportunities.

Next, do CVC parents add new emerging phrases that are directly corresponded with the CVC deals? For example, CVC deals related to artificial intelligence should predict the adding of "artificial intelligence" into the 10-K. Table III tends to answer this type of question. First, I sort each emerging phrase into one of the eight VentureXpert VEIC industries according to the industry of those startups that use the emerging phrases to describe businesses.[21] As shown in the main diagonal, CVC deals in Industry j usually predict the Industry j specific emerging phrases newly added into the firm's 10-K annual report.

Finally, a natural concern is how long a CVC parent retains the emerging phrases in its subsequent annual 10-Ks after adding them. Panel C and D of Figure 3 answer the question, where Panel C plots the words cloud of those top 50 frequent emerging words newly added into the CVC parents' business descriptions within the three years after investments.[22] Consistent with the intuition, more general phrases (such as information technology) are more likely to be added than those tech buzzwords across the CVC firms sample.[23] Furthermore, Panel D plots the distribution of years of surviving for each emerging phrase after being added into the business of a CVC parent. On average, each phrase survives in the next 2.5 years, with more than 75% at least surviving in the next annual 10-K report.

[21]This sorting is non-exclusive. Take an example: Artificial Intelligence is sorted into both Computer Software and Internet Specific industries.

[22]There are 2,081 emerging phrases added by CVC parents, with 512 unique phrases.

[23]Many words in Panel C also relate to software and the internet, consistent with the fact that around 50% of CVC-backed startups operate in the SIC-3 737 industry, the software service industry.

B. Evidence on Dynamics of Firm Segments

An alternative method to measure the firm scope change is to explore the Compustat Segment data. The following logit model thus examines whether CVC leads to future firm scope changes using division (segment) measures. The empirical model takes the following form, on the firm-year panel with all US public firms in non-financial industries,

$$D(Scope\ Change)_{i,t+1:t+k} = \beta D(CVC)_{i,t} + \gamma \mathbf{X}_{i,t} + \upsilon_t \times \iota_j + (\tau_i) + \varepsilon_{i,t} \tag{3}$$

where $D(Scope\ Change)$ corresponds to three different dummies for firm scope changes as illustrated in Section III.B. Regarding establishing new divisions and removing obsolete divisions, I examine them within the next two years after each CVC deal, whereas I identify the change of primary corporate industry (SIC-3) in the next three to five years. Later, I will provide evidence about why I choose those specific intervals. The regression sample is further adjusted to alleviate the potential survivorship bias.[24]

Table IV, Panel A (Columns 1 – 3) investigates the post-CVC division creation, where the dependent variable is a dummy equal to 1 if the firm creates at least one new division (in a new industry) within the next two years.[25] In Columns (2) and (3), I split the D(CVC) into two dummies, D(CVC Unrelated) and D(CVC Related), according to whether the startup in the deal is matchable to the CVC parent's business using SIC-3 codes.[26]

In Panel A, the coefficients of both D(CVC) and D(CVC Unrelated) are positive and statistically significant, but not the coefficient of D(CVC Related). It implies that firms are more likely to create a new business (in a new industry) within the next two years following unrelated CVC investments, consistent with the prediction in Section II. Unrelated CVC deals help its corporate parent identify the new business opportunities outside its current

[24] Specifically, in Panel A of Table IV, I require that each firm observation survives, at least, in the next two years. For panel B, I require each firm observation to survive, at least, in the next five years.

[25] See Section III.B for details.

[26] Furthermore, in Columns (1) and (2), I add the high dimensional industry by year fixed effect to absorb any industry shocks driving the firm scope change, as studied in Harford (2005) and Maksimovic and Phillips (2008), while the firm fixed effect is controlled in Column (3).

business domain and further prompt the firm to integrate the new business.

The marginal effect is very significant: the probability increased by conducting CVC unrelated deals is about 4.91%, equivalent to 57% of the unconditional probability of creating new divisions in the sample.

The reason why related deals have no impact on the division creation is simple: the dummy-version firm scope change is a pretty coarse measure, without capturing the granular change of the business within a SIC-3 code. Therefore, even though the related deals are strongly associated with business changes in the firm's current domain measured by "emerging phrases", it does not show in Table IV.

Turn to Columns 4-6 in Panel A, I obtain a similar pattern on post-CVC divisions removal: firms are 2.73% – 5.02% more likely to remove existing divisions within the next two years following the CVC deals. Again, the impact is only confined to unrelated deals.

In Panel B, the dependent variable is instead a dummy of changing a firm's primary corporate industry in the next three to five years (in the next four to six years in Columns (4) to (6)). As shown by the coefficients, CVC investments (only unrelated deals) significantly lead to the future change of the primary corporate industry. If I instead identify industry change within the next two years, there is no effect. It suggests that the industry change takes longer time than the division creation or removal. This timing makes sense since it takes time for the newly established business to grow and become the core business. Indeed, as shown in Panel C of Table I, there are 104 industry changes taking place within 3-5 years following unrelated CVC deals, and 43 of them are attributable to the continuing growth of the newly established division which finally turns to the new primary business.

To provide reasons for selecting those specific intervals of the scope change dummies, Figure 6 studies the scope change activities in the [-3 Year, +5 Year] window around the CVC investments. The regression specification is the same as Figure 5. For simplicity, only the coefficients of unrelated CVC deals are plotted since I do document that only unrelated deals impose significant impacts in Table IV.[27]

The result of Figure 6 closely mirrors Table IV: CVC investments lead to the division creation and termination in the next three years and further push the changes of the industry in the fourth to the fifth year. Nevertheless, there is no pre-CVC scope change activity in the three years before. The joint tests of the difference between the coefficients of three years before and three years after in Panel A and B are significant: $p = 0.0279$ and $p = 0.0166$, respectively. Interestingly, the division removal seems to be concomitant with the division creation, which might be due to the financing of establishing a new business.

All told, the launch of a CVC program is associated with the firm scope change of its parent firm using both measures. However, the potential endogeneity still could contaminate the baseline results. There are three different sources of endogeneity. First, firms with some obsolete businesses or technologies are more likely to leverage on CVC to obtain new ideas, which finally leads to firm scope changes (Ma, 2020). Second, in contrast to the story that CVC facilitates to identify new business opportunities, the emergence of new technology or new business opportunities might incentivize the manager to invest in CVC (reverse causality in the first arrow of Figure 1). Third, after deciding to enter the CVC foray, choosing between CVC-related and unrelated deals is endogenous.

Unfortunately, the CVC literature is silent on tackling this kind of complicated endogeneity. Therefore, I introduce a new IV strategy using the independent VC firm's fund inflow as the source of exogenous variation in Section VI. But before elaborating this IV, I will first study how CVC helps identify new business opportunities. In fact, understanding the underlying mechanism also helps to rule out some parts of the endogeneity.

V. Experimentation and Firm Scope

This section explores how CVC could help identify new business opportunities and ultimately spur firm scope changes. To summarize the story in one sentence, CVC is a learning-

[27]However, the complete estimate results are shown in the online appendix.

through-experimentation process. In this process, a CVC parent firm usually sprays VC deals across various technology or business options (this is in the same spirit of the "spraying" strategy in Ewens et al. (2018)), then waits for the signals revealing the potential of the options, and finally responds to the signals by conducting the firm scope changes.

A. *Diversifying CVC Investment Strategy*

As developed in the hypothesis section (Section II), two empirical predictions could be derived from the experimentation hypothesis. First, given the purpose of experimentation to solve the future uncertainty and pin down the best growth opportunity, a CVC program should necessarily diversify its deals across technology fields and industries, which optimally increases the probability of figuring out great business opportunities. If the experimentation is indeed the underlying mechanism (shown in Figure 1), we should find in the data that this diversifying CVC strategy, a necessary step of experimentation and tackling with the uncertainty, is strongly linked to the firm scope changes. Furthermore, an underlying prediction of the experimentation story is that, before the CVC investments (and of course before CVC program launch), the CVC parent does not clearly know which direction is the best for its future business expansion.[28]

Table V aims to test this diversification hypothesis. First, I define a new dummy variable of CVC investments, D(CVC Past 3yr), by grouping all past 3-year CVC deals by a firm. Precisely, the dummy is equal to 1 if the firm conducted at least one CVC deal within the past three years. Choosing three years follows Ma (2020)'s finding that a typical CVC program lasts three to four years. Then I interact this new dummy with two new variables, measuring to which extent a CVC parent firm diversifies its CVC deals across ten detailed VentureXpert industries (VEIC Industries).[29] Inverse HHI(VEIC) calculates the one minus the HHI of the past three-year CVC deals across the VEIC industries, while Num(VEIC) counts the number

[28]The alternative story is that, before the CVC investments, the firm clearly knows the future direction of the business growth, such as AI, and therefore the firm chooses to only invest in AI.

[29]VentureXpert assigns startups to its own industry classification called VEIC. All results are robust if I instead choose a more granular VEIC industry classification.

of VEIC industries covered by the firm's past 3-year investments. Furthermore, I also control the number of deals within the past three years.

Table V shows that the interaction term between D(CVC Past 3yr) and CVC diversification measure (Inverse HHI (VEIC) or Num(VEIC)) is always positive and significant, implying that firms with higher diversifying strategy are more likely to conduct division creation and industry change (Column (1) – (4)) and add new emerging phrases (Column (5) – (6)). In contrast, conducting more CVC deals does not guarantee success in identifying new businesses since the interaction term between CVC dummy and Num(CVC Deals) is not always positive.

In conclusion, Table V supports the experimentation story (with the diversifying strategy). Furthermore, it implies that, before a CVC program launch or during the CVC investments, the parent firm is unknown about the growth direction. Perhaps the strongest result lies in Columns (3) and (4) of Table V. If the firm already knows to which industry it will shift the course, there is no need to diversify deals across industries ex-ante but instead to focus deals in the predetermined industry. This evidence helps partially rule out the reverse causality that it is not CVC that identifies business opportunities but that firms first observe business opportunities and then decide to invest in CVC.

B. *CVC Signals and Firm Scope Change*

Having documented the diversifying investment strategy, now I turn to the CVC signal and the industry choice for a new division creation. Intuitively, a CVC parent firm will not establish a new division in every industry where it sprays CVC deals; instead, only the industry with positive post-investment signals (from startups) is considered. And this response to the signal is the key feature of experimentation. In other words, if the experimentation is the accurate underlying mechanism to the firm scope change, I should find in the data that firms learn from the experimentation the signals about the future potential of business options and make responses by creating new divisions following those signals.

B.1. Introduction to a New Discrete Choice Model

Table VI and VII develop and estimate a McFadden discrete choice model regarding the industry choice for division creation, therefore providing some insights into the response to CVC signals. First, I introduce the empirical model setup and then turn to the testing of CVC-signal responses.

The observation unit is at the firm-year-industry level, where each observation represents an alternative (Industry j) in which Firm i in Year t could choose to create a new division. The set of alternatives (choice set) consists of 404 non-financial SIC-3 industries documented at least once in the Compustat Historical Segment database from 1980 to 2017.

I only include firms having invested in at least one CVC deal during the sample period. For each decision-maker (a firm-year pair), I drop those alternatives (industries) that already exist as a division of the firm in Year t-1 or those that have already been created before Year t. In the model, the dependent variable of interest is a dummy equal to 1 if the alternative industry is chosen by the Firm i in Year t to establish a new division.

B.2. Some Basic Facts

Table VI starts to explore some basic features of the division creation process, while Table VII focuses on signal responses. In Column (1) of VI, I start to introduce an crucial control variable, D(CVC 3yr), a dummy equal to 1 if, within the past three years, the Firm i has invested CVC deals in that industry (with the invested startup in Industry j). The positive and strongly significant coefficient shows that a CVC parent often creates a new division in the industry where it has sprayed CVC deals in the past.

In Column (2), I interact D(CVC 3yr) with two industry proximity measures.[30] The coefficients of these two proximity measures are both positive and significant, which implies

[30]D(Ind. Proxy SIC1) and D(Ind. Proxy SIC2) both capture the industry proximity between the alternative and the industries of existing divisions of Firm i in Year t-1. D(Ind. Proxy SIC2) is a dummy equal to 1 if the alternative has the same 2-digit SIC with one of the existing divisions of Firm i. Similarly, D(Ind. Proxy SIC1) is a dummy equal to 1 if the alternative has the same 1-digit SIC with one of the existing divisions of Firm i but does not have the same 2-SIC with them.

that, in general, firms are more likely to establish a new division close to their existing business domains. This is possibly due to higher asset redeployability or closer product language usage (Hoberg and Phillips, 2018). In contrast, the interaction term is negative and highly significant, showing that CVC usually helps its parent create a new division far away from its current business domains.

B.3. Evidence on CVC Signals

Next, I turn to Table VII to test the CVC signal hypothesis, where I interact D(CVC 3yr) with each signal variable iteratively. However, it is challenging to proxy the signal variable since researchers are not able to observe the information transmitting from startups to the CVC parent firm. For example, to which extent the potential new business will fit with the parent's old business is a kind of soft information only observed by the insiders.

As a result, in Panel A, I use the startup's measurable performance as the proxy of signals. Importantly, each signal variable is observable to the CVC parent after the investment but not before. For example, in Column (1), the signal is measured by the number of startups in the parent's CVC portfolio which finally exit through IPO.[31] In other words, the signal variable is based on information from the past three-year investments in Industry j. Consistent with my hypothesis, firms are more likely to create a new division when they receive a positive signal from their past three-year investments in the relevant industry.

Moreover, I control the number of startups invested in the past three years in Industry j and denote it as $Num(Startups\ Invested)$. It is essential since, naturally, the more you invest, the more IPO startups you will have. In Panel B, I define the signal variables as the fraction, such as the fraction of IPO startups, but do not control $Num(Startups\ Invested)$.

[31] As an example, if Google Venture has invested five startups in the past three years of Year t in Industry j, and finally, three of them go IPO, then the signal variable is equal to 3. The number of deals is 5. One important assumption I have to make here is that, since the IPO date of these three startups will be naturally after the Year t (the decision-making year of division creation), Google could not directly observe that the signal is equal to 3 at the time of division creation decision. But Google should be able to draw valuable information about the potential of the IPO through CVC investments and taking board seats, which is supported by Dushnitsky (2012).

One might also argue that the IPO signal variable might proxy the industry-year general IPO trend or clustering, and thus I control the industry-year IPO trend as a separate control, as well as the interaction between the CVC dummy and IPO trend. The negative and significant sign of the interaction term shows that CVC parents usually do not over-react to the IPO industry cluster like non-CVC peer firms (captured by the non-interaction, the IPO Cluster), but only respond to their deal-specific signals.

Similarly, I construct signal variables from the startup's acquisition, bankruptcy, and patent information. In summary, CVC parent firms react to positive signals from IPO and patent growth, the negative signal of bankruptcy, but not the acquisition of the startups by third-party. Column (3) of both Panel A and B also documents that acquisition of portfolio companies in the CVC investments could positively predict the division creation. It is consistent with the intuition that the CVC firm understands the capability needs for adding new businesses and directly builds the capability through acquiring startups which it invests within the past three years (Keil et al., 2008).

C. CVC Signals and Emerging Phrases

This section turns to studying the CVC signals and subsequent emerging business integration. By similar reasoning, CVC parent firms will not integrate every emerging business they have invested in through CVC programs; instead, they pick winners with positive signals.

Table VIII thus estimates a similar discrete choice model for the industry choice of emerging phrases added into annual 10-K reports, where each emerging phrase is sorted into eight VentureXpert Industries (VEIC). However, strictly speaking, it is not a discrete choice model since a firm could add emerging phrases in multiple VEIC industries simultaneously, which is very rare in the case of division creation.

In the choice model, the unit of observations is at the firm-year-VEIC level. I sort each emerging phrase defined in Section III.B into 8 VEIC industries following the procedure of Table III. The dependent variable of interest is the number of VEIC-j specific emerging phrases newly added into the firm's 10-K in Year t. Panel A studies the basic discrete choice model, while Panel B focuses on CVC signals and responses. In Panel A, the key control variable, D(CVC VEIC j), is defined as a dummy equal to 1 if the firm has invested startups in the VEIC-j industry within the past three years. The regressions show that firms usually add industry-specific emerging phrases following their industry-specific investments. More specifically, CVC parents do not add phrases in each industry. They absorb emerging words only from Biotechnology, Communication, Computer Software, Internet Specific industries, and others.

Panel B repeats the same exercise as in Table VII by interacting the CVC dummy with the signal variables mentioned above. Interestingly, all results are exactly the same as those in Table VII, suggesting that, in the decision of establishing new divisions and adding emerging businesses, firms react to the same set of signals.

VI. Exogenous Variation on CVC Experimentation

It is very challenging to find any exogenous variations on the CVC program initiation. This is implied by the fact that there is no attempt to provide any identification strategies for the CVC program launch in current finance and strategic literature.

To have the first attempt, I introduce an exogenous variation on the continuation of CVC investments conditional on that the CVC program has already been started. In Section VI.A, I introduce this instrument and report the estimates, Section VI.B offers more discussion about its validity.

A. Identification Strategy with IVC Fund Inflow Shock

I exploit the fund (capital) inflow shocks of independent venture capital firms (IVC) in each CVC program's past syndicate network. Notably, those fund inflow shocks are idiosyncratic ones, being orthogonal to aggregate shocks in the VC industry. An example can be a pension fund that injects a large amount of capital into a non-star VC during the non-bubble period.

The instrument works on a small sample of US public firms already starting the foray of CVC investments in the past. It relies on the VC literature about syndicating investments. First, the syndicating investment and its network formation are common in the venture capital world, and many VC firms commonly invite their past syndicating partners to join in their new investments (Hochberg et al., 2007). Second, the IVC is the most crucial channel of deal sourcing for CVC firms, as documented in Sykes (1990) and MacMillan et al. (2008), among others.

Based on the two premises, the idea of the instrument is that, if an independent VC firm j (IVC j) receives a positive fund inflow shock today, and meanwhile, the CVC Firm i is in its past syndicate network, then the IVC j is very likely to initiate new deals and invite CVC Firm i, its old partner, to join in its new investments. Alternatively, IVCs can recommend new deals to CVCs when IVCs start new funds and seek deals. As a result, the new investment of CVC Firm i is driven by IVC's idiosyncratic fund inflow shocks instead of the CVC firm's product life cycle and other unobserved corporate strategies.

To facilitate the understanding, consider the example illustrated in Figure 7, showing how its IVC partners drive the CVC investment decisions of Apple Inc. In the past five years before 1990, Apple Inc has built three connections with three distinct IVCs through syndicate investments. Among the three IVCs, two received positive inflow shocks in 1990. One of these two IVCs, Mayfield Fund LLC, then spent its new money on investing in a seed-stage startup called BioCAD Corp in 1990, followed by Apple Inc's joining due to Mayfield's invitation.[32]

To construct this instrument, I proceed with two steps. First, for each CVC Firm i in Year t, I obtain its past five-year syndicate network by searching all IVCs that have co-

[32]Another example of the invitation is that, between 1994 and 2000, Cisco Systems (a large industrial firm) was invited into 13 syndications led by Sequoia Capital (an independent VC firm), as documented in Ferrary (2010).

invested with Firm i within the past five years. The co-investment (syndication) is defined as the scenario in which CVC Firm i and IVC Firm j invest in the same round of the same startup k (Hochberg et al., 2007). In the second step, for each IVC in the network, I check whether it receives a positive fund inflow shock in Year t.

Here, the main challenge is obtaining the IVC's fund inflow shock exogenous to any VC investment opportunities and any technology shocks. I construct it following the recent Granular IV approach developed by Gabaix and Koijen (2020). First, I proxy an IVC's raw capital inflow by its raising of new follow-on funds since (i) fundraising is usually accompanied by the largest capital inflow, and (ii) when an IVC starts a new follow-on (sequential) fund, it is more likely to invite CVC firms to join its new deals.

Next step, I estimate Gompers and Lerner (1998)'s fundraising model with plenty of VC funding factors and VC organization controls, along with high dimensional fixed effects. I obtain the idiosyncratic fund inflow shock from the error term of the fundraising model. Appendix B shows the detailed procedures, estimated results, and error terms' properties. In the last step, I sum up the error term (the idiosyncratic shock) across IVCs in each CVC program's network and define it as my Granular IV.

The intuition behind my Granular IV is similar as in Gabaix and Koijen (2020). Gabaix and Koijen (2020) argues that the Granular IV heavily relies on the "unexpected" change in the loading on a common shock. If OPEC decided to cut down the oil production, but Saudi Arabia cuts down more than anticipated, that is an idiosyncratic shock. The same argument applies to the idiosyncratic capital inflow shock of IVCs.

Table IX reports the first stage regression where I use the Granular IV (sum of the idiosyncratic fund inflow shocks) to instrument the continuation of CVC investments by each CVC program. I restrict my analysis to a small sub-sample of CVC firms having already initiated a CVC program in the past five years before and thus enjoy some VC networks today. In the regressions, I control the size and quality of the past IVC network, given that the network (past investments) is endogenous. Finally, Table IX shows that the sum of IVC's idiosyncratic fund inflow shocks highly predicts the new CVC investments for both general deals and initial deals (no follow-on deals).[33]

Regarding the exclusive condition, a potential concern is that some specific industry (technology)-year shocks might drive both the fund inflow shocks (new VC fundraising) and firm scope changes. For example, the introduction of cloud computing services by Amazon, studied in Ewens et al. (2018), might push many past-connected VC firms to launch new funds and to invest in e-commerce startups. Many established firms in the retail sales industry might follow the technology shock and start creating a new division regarding e-commerce.

To mitigate this concern, I always include the industry (SIC-3) by year fixed effects in both the first and second stage regressions. Furthermore, when estimating Gompers and Lerner (1998)'s fundraising model (where I get the error term and thus the shock), I add both the VC industry specialization by year and VC location by year fixed effects.[34]

Equipped with the instrument, I conduct 2SLS regression by instrumenting the number of CVC initial investments (with natural logarithm) with the Granular IV and report the results in Table X. Columns (1) to (3) analyze the text-based scope measures, whereas the segment measures are used as the dependent variable in Columns (4) to (6).

In Column (2), I introduce a new textual measure, the Business Change, which is another granular measure of the CVC parent's business change. Following Hoberg, Phillips, and Prabhala (2014), it equals one minus the cosine-similarity between the firm's year t and year $t+1$'s business descriptions.[35] I document a strong and positive effect of CVC on firm scope change (3.77% change of Cosine similarity) and adding 0.85 more emerging phrases.

[33]Initial deals are those deals in which the CVC firm invests in a specific startup for the first time, i.e., not the follow-on investments. The number of initial deals better measures the impact of GIV on the deal sourcing availability of CVC firms.

[34]To further provide the deal-level evidence of my instrument (the evidence that IVCs do invite CVC), I estimate a discrete choice model (McFadden (1973)) (in the online appendix) regarding the choice of portfolio companies by CVC programs. The empirical model shows that CVC does follow the choice of picking startups by its past-connected IVC partners, especially when the latter receives positive fund inflow shocks.

[35]In unreported results, I find that the segment dummies are all strongly and positively correlated with the new textual measure.

As shown in Columns (4) to (6), CVC investments impose a positive and significant impact on division creation and industry change but not division removal. For example, one standard deviation increase of Num(CVC Initial Deals) leads to about 6% of probability increase of establishing a new division in the next two years.

B. *Further Discussion*

Two upshots deserve further clarification. First, as my instrument relies on the argument that IVC invites CVC or at least recommends deals to CVC after receiving fund inflow shock, one might worry that the counter-hypothesis that CVC invites IVC might also happen, which could dampen my instrument.

Nevertheless, this hypothesis is not supported in both the data and survey evidence. In most syndicating cases between CVC and IVC, the IVC usually leads the deal, while the CVC does not lead. And only the leading investors invite others to join the deal. Moreover, MacMillan et al. (2008) summarizes the CVC-IVC syndicate as follows: "CVCs and independent venture capital often co-invest in companies through syndicated investments. The independent venture capital investor usually takes the role of lead investor. CVCs benefit from access to the investment 'deal flow' of independent venture capital, while independent venture capital benefits from strategic insight and technology expertise provided by CVCs."

Second, one might argue that the invitation from IVCs to CVCs is endogenous (depending on CVC's technology expertise), as well as the decision regarding whether CVC accepts the invitation. However, notice that the instrument purely relies on the fund inflow shock and does not hang on the two aforementioned decisions. Thereby the instrument is valid as long as the idiosyncratic fund inflow shock is truly exogenous.

VII. Additional Analysis

A. *Alternative Identification: Evidence from US Airline Route*

One major caveat of my Granular IV is that it fails in distinguishing between related and unrelated CVC deals. Therefore, this section provides an alternative identification exercise by exploring the US airline route in Bernstein, Giroud, and Townsend (2016), thus focusing exclusively on the unrelated CVC deals.

To start this analysis, I first gather all *unrelated* CVC deals with both the startup and the CVC firm located in the US.[36] The deals range from 1990 to 2017, the interval in which I can obtain both the US airline route data and Compustat Segment data. Next, I classify all unrelated deals into two groups according to whether there is a direct (non-stop) flight route, during the year right after the deal year, between the metropolitan statistic area (MSA) of the CVC firm and of the startup.[37]

Following Bernstein et al. (2016), the idea is that a higher frequency of non-stop flights between CVC and the startup's location provides more chances for the CVC manager to visit the startup and subsequently acquire knowledge regarding the startup and its emerging business opportunities behind it.[38] Therefore, the CVC parent should be more likely to conduct firm scope changes after a more frequent interaction with the invested startup.

The identification assumption is that the number of non-stop flights is quasi-exogenous to any firm or industry characteristics driving the scope change decisions after controlling both the startup and CVC location by year fixed effects. This assumption is quite reasonable as the first-order effects driving the number of non-stop flights between two cities are usually the number of travelers between them and airports' hub-and-spoke connection.[39]

Figure 8 plots the two groups of CVC deals, separately, in Panel A and B, on the US

[36] Furthermore, I require that the startup's SIC-3 code is not missing, as well as no-missing MSA information for both the startup and CVC firm in the deal.

[37] I obtain very similar results if I instead measure the direct flight during the deal year.

[38] It is in line with the fact that much of the emerging business knowledge consists of tacit and narrative knowledge, in which case visiting the startup is the only way to access it (Keil et al., 2008).

[39] For example, see the article: https://www.afar.com/magazine/how-airlines-get-new-routes.

states map. The blue point denotes the CVC firm location, while the red point represents the startup's location. Moreover, around 50 blue-red pairs appear in both Panel A and B due to either introducing a new non-stop airline or stopping an old airline route in my sample period.[40]

Table XI, Panel A provides the deal sample's summary statistics, breaking them down by deals with and without any direct flights.[41] As shown in the panel, startups in those CVC deals with direct flights are more likely to be located in the hot areas for VC activities – California, Massachusetts, and New York.

In Panel B and C of Table XI, I run OLS regressions on the CVC deals sample and study the relation between direct flight and scope changes. In Panel B, the dependent variable is a dummy equal to 1 if, within the next three years after the deal, the CVC parent establishes a new division in the same industry of the startup of the deal. As shown in the panel, the number of non-stop flights (measured in the year right after the deal year) is positively correlated with creating a new division by the CVC parent (and the division is exactly located in the startup industry). The results are robust across different controls and fixed effect specifications. Importantly, I add both the CVC firm and startup's location by year fixed effect to control local shocks in Column (4), along with the CVC firm fixed effect in Column (5). In Panel C, I obtain very similar results on changing the primary industry: more frequent non-stop flights lead to a higher likelihood of turning its primary industry close to the startup's business.

Lastly, one might worry that the deal-selection (between deals with and without direct flight) might bias the result. If it does, I argue that theoretically it could only generate opposite results and dampens my findings. According to the deal selection view, the deals without direct flight should offers the CVC parent manager more insights about future business opportunities since the (timing) cost of monitoring and interaction is higher than

[40] Since this subsample is too small, I could not conduct any diff-in-diff analysis as in Bernstein et al. (2016).

[41] Those deals with the startups and the CVC firm located in the same MSA are not included in this panel.

those deals with direct flights. As a result, I expect that the results of Table XI should be even stronger without the deal selection effect.

B. Post-CVC Value Creation

Given that CVC creates value for shareholders of its corporate parent (Dushnitsky and Lenox, 2006; Ma, 2020), my final analysis is about whether post-CVC scope changes also creates value for its parent in Table XII. The left-hand side variable is the difference of Tobin's Q of Firm i between Year $t+h$ and Year t, where h usually takes a value of three or four. In Column (1) and (2), I conduct the horse-racing test between CVC related and unrelated dummy. As shown in the two columns, only the coefficients of D(CVC Unrelated) dummy are positive and significant, showing that only unrelated CVC deals create significant values.

From Column (3) to Column (8), I iteratively interact the D(CVC Unrelated) dummy with three scope changes dummies. The scope changes dummies are again forward-looking: in the next two years for division creation and removal and within three to five years for changing of primary corporate industry. As a result, in Columns (7) and (8), the Δ of Tobin's Q takes five or six years to ensure that the industry changes have already been completed before measuring the value improvement. I find that only D(CVC Unrelated) $\times$ D(scope changes) is significant, while D(CVC Unrelated) $\times$ (1$-$D(scope changes)) is not. This shows that CVC investments' value creation mostly derives from post-CVC scope changes, as CVC without any scope changes does not bring significant value improvement.

It is a bit arbitrary regarding whether I should include the firm fixed effects in Table XII. On the one hand, it might be redundant since the left-hand side variable already takes the difference, eliminating any time-invariant firm characteristics. On the other hand, if I add them, the joint F test for those firm fixed effects is significant. However, my results are quite robust with and without firm fixed effects. In the online appendix, I show the alternative regressions without any firm fixed effects for Column (3) to (8) of Table XII.

VIII. Conclusion

This paper investigates the effects of corporate venture capital (CVC) on the scope changes and product innovation of the CVC parent corporation. To deal with the potential endogeneity, I develop two different identification strategies.

First, I introduce a new instrument for CVC investments using the fund inflow shocks of independent venture capital firms (IVC) in each CVC program's past syndicate network. The idea of the instrument is that, if an independent VC firm j (IVC j) receives a positive fund inflow shock today, and meanwhile, the CVC Firm i is in its past syndicate network, then the IVC j is very likely to launch a new sequential fund, initiate new deals, and invite CVC Firm i, its old partner, to join in its new investments. Second, I introduce the US non-stop airline routes as a quasi-natural experiment. I consistently corroborate the causal relationships.

In the various extension analysis, I document that a CVC parent is more likely to establish a new division in the industry where it has sprayed CVC deals before. Furthermore, the post-CVC value enhancement of the CVC parent derives mostly from post-CVC scope changes. Overall, the evidence is consistent with the idea that CVC helps to identify new business opportunities.

REFERENCES

Basu, Sandip, Corey Phelps, and Suresh Kotha, 2011, Towards Understanding Who Makes Corporate Venture Capital Investments and Why, *Journal of Business Venturing* 26, 153–171.

Bergemann, Dirk, and Ulrich Hege, 2005, The Financing of Innovation: Learning and Stopping, *RAND Journal of Economics* 719–752.

Berger, Philip G., and Eli Ofek, 1995, Diversification's Effect on Firm Value, *Journal of Financial Economics* 37, 39 – 65.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, 2016, The Impact of Venture Capital Monitoring, *The Journal of Finance* 71, 1591–1622.

Chemmanur, Thomas J., Elena Loutskina, and Xuan Tian, 2014, Corporate Venture Capital, Value Creation, and Innovation, *The Review of Financial Studies* 27, 2434–2473.

Chesbrough, Henry W., 2002, Making Sense of Corporate Venture Capital, *Harvard Business Review* 80, 90–99.

Coase, Ronald H, 1937, The Nature of the Firm, *The Nature of the Firm. Origins, Evolution, and Development. New York, Oxford* 18, 33.

Dushnitsky, Gary, 2012, Corporate Venture Capital in the Twenty-First Century: an Integral Part of Firms' Innovation Toolkit, *The Oxford Handbook of Venture Capital* .

Dushnitsky, Gary, and Michael J. Lenox, 2005, When do Firms Undertake R&D by Investing in New Ventures?, *Strategic Management Journal* 26, 947–965.

Dushnitsky, Gary, and Michael J. Lenox, 2006, When does Corporate Venture Capital Investment Create Firm Value?, *Journal of Business Venturing* 21, 753 – 772.

Ernst, and Young, 2009, Global Corporate Venture Capital Report .

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of Experimentation and the Evolution of Venture Capital, *Journal of Financial Economics* 128, 422–442.

Ferrary, Michel, 2010, Syndication of Venture Capital Investment: The Art of Resource Pooling, *Entrepreneurship Theory and Practice* 34, 885–908.

Fulghieri, Paolo, and Merih Sevilir, 2009, Organization and Financing of Innovation, and the Choice between Corporate and Independent Venture Capital, *Journal of Financial and Quantitative Analysis* 44, 1291–1321.

Gabaix, Xavier, and Ralph S.J. Koijen, 2020, Granular Instrumental Variables, Technical report, National Bureau of Economic Research.

Gompers, Paul, and Josh Lerner, 2000, Money Chasing Deals? The Impact of Fund Inflows on Private Equity Valuation, *Journal of financial economics* 55, 281–325.

Gompers, Paul A., and Josh Lerner, 1998, What Drives Venture Capital Fundraising?, Technical report, National bureau of economic research.

Grossman, Sanford J, and Oliver D. Hart, 1986, The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, *Journal of Political Economy* 94, 691–719.

Hamm, Sophia J.W., Michael J. Jung, and Min Park, 2018, How Transparent are Firms about Their Corporate Venture Capital Investments?, *Available at SSRN* .

Harford, Jarrad, 2005, What Drives Merger Waves?, *Journal of Financial Economics* 77, 529 – 560.

Hart, Oliver, and John Moore, 1990, Property Rights and the Nature of the Firm, *Journal of Political Economy* 98, 1119–1158.

Hellmann, Thomas, 2002, A Theory of Strategic Venture Investing, *Journal of Financial Economics* 64, 285 – 314.

Hellmann, Thomas, Laura Lindsey, and Manju Puri, 2008, Building Relationships Early: Banks in Venture Capital, *The Review of Financial Studies* 21, 513–541.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-Based Network Industries and Endogenous Product Differentiation, *Journal of Political Economy* 124, 1423–1465.

Hoberg, Gerard, and Gordon Phillips, 2018, Conglomerate Industry Choice and Product Language, *Management Science* 64, 3735–3755.

Hoberg, Gerard, Gordon Phillips, and Nagpurnanand Prabhala, 2014, Product Market Threats, Payouts, and Financial Flexibility, *The Journal of Finance* 69, 293–324.

Hoberg, Gerard, and Gordon M. Phillips, 2021, Scope, Scale and Competition: The 21st Century Firm, *Available at SSRN* .

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu, 2007, Whom You Know Matters: Venture Capital Networks and Investment Performance, *The Journal of Finance* 62, 251–301.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu, 2010, Networking as a Barrier to Entry and the Competitive Supply of Venture Capital, *The Journal of Finance* 65, 829–859.

Keil, Thomas, Erkko Autio, and Gerard George, 2008, Corporate Venture Capital, Disembodied Experimentation and Capability Development, *Journal of Management Studies* 45, 1475–1505.

Keil, Thomas, Markku V.J. Maula, and Cameron Wilson, 2010, Unique Resources of Corporate Venture Capitalists as a Key to Entry into Rigid Venture Capital Syndication Networks, *Entrepreneurship Theory and Practice* 34, 83–103.

Kim, Keongtae, Anandasivam Gopal, and Gerard Hoberg, 2016, Does Product Market Competition Drive CVC Investment? Evidence from the U.S. IT Industry, *Information Systems Research* 27, 259–281.

Lamont, Owen, 1997, Cash Flow and Investment: Evidence from Internal Capital Markets, *The Journal of Finance* 52, 83–109.

Lang, Larry H.P., and Rene M. Stulz, 1994, Tobin's Q, Corporate Diversification, and Firm Performance, *Journal of political economy* 102, 1248–1280.

Lerner, Joshua, 2012, *The Architecture of Innovation: The Economics of Creative Organizations* (Harvard Business Press).

Ma, Song, 2020, The Life Cycle of Corporate Venture Capital, *The Review of Financial Studies* .

MacMillan, I., E. Roberts, Val Livada, and A. Wang, 2008, *Corporate Venture Capital (CVC) Seeking Innovation and Strategic Growth* (National Institute of Standards and Technology).

Maksimovic, Vojislav, and Gordon Phillips, 2007, Chapter 8 - Conglomerate Firms and Internal Capital Markets, in B. Espen Eckbo, ed., *Handbook of Empirical Corporate Finance*, Handbooks in Finance, 423 – 479 (Elsevier, San Diego).

Maksimovic, Vojislav, and Gordon Phillips, 2008, The Industry Life Cycle, Acquisitions and Investment: Does Firm Organization Matter?, *The Journal of Finance* 63, 673–708.

Manso, Gustavo, 2016, Experimentation and the Returns to Entrepreneurship, *The Review of Financial Studies* 29, 2319–2340.

Matvos, Gregor, Amit Seru, and Rui C. Silva, 2018, Financial Market Frictions and Diversification, *Journal of Financial Economics* 127, 21–50.

McFadden, Daniel, 1973, Conditional Logit Analysis of Qualitative Choice Behavior .

McNally, Kevin, 1997, *Corporate Venture Capital: Bridging the Equity Gap in the Small Business Sector* (Routledge).

Panzar, John C., and Robert D. Willig, 1981, Economies of Scope, *The American Economic Review* 71, 268–272.

Seru, Amit, 2014, Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity, *Journal of Financial Economics* 111, 381 – 405.

Shan, Hongyu, 2019, Corporate Venture Capital and the Boundaries of the Firm, in *2nd Emerging Trends in Entrepreneurial Finance Conference*.

Siegel, Robin, Eric Siegel, and Ian C. MacMillan, 1988, Corporate Venture Capitalists: Autonomy, Obstacles, and Performance, *Journal of Business Venturing* 3, 233 – 247.

Sykes, Hollister B., 1990, Corporate Venture Capital: Strategies for Success, *Journal of Business Venturing* 5, 37–47.

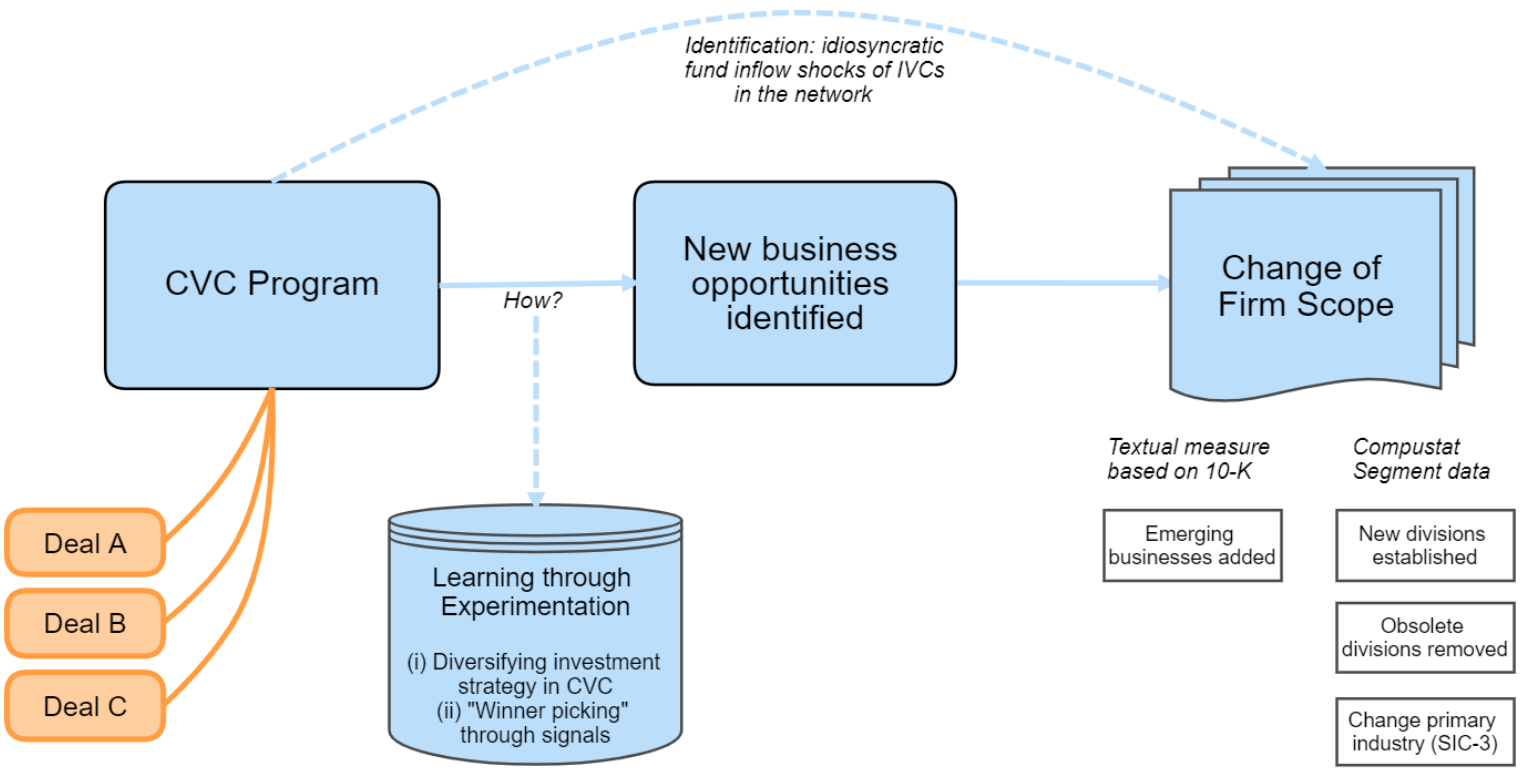
Teece, David J., 1980, Economies of Scope and the Scope of the Enterprise, *Journal of Economic Behavior & Organization* 1, 223–247.

Tian, Xuan, and Kailei Ye, 2018, A Dark Side of Corporate Venture Capital, *Available at SSRN* .

Williamson, Oliver E., 1985, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting* (The Free Press).

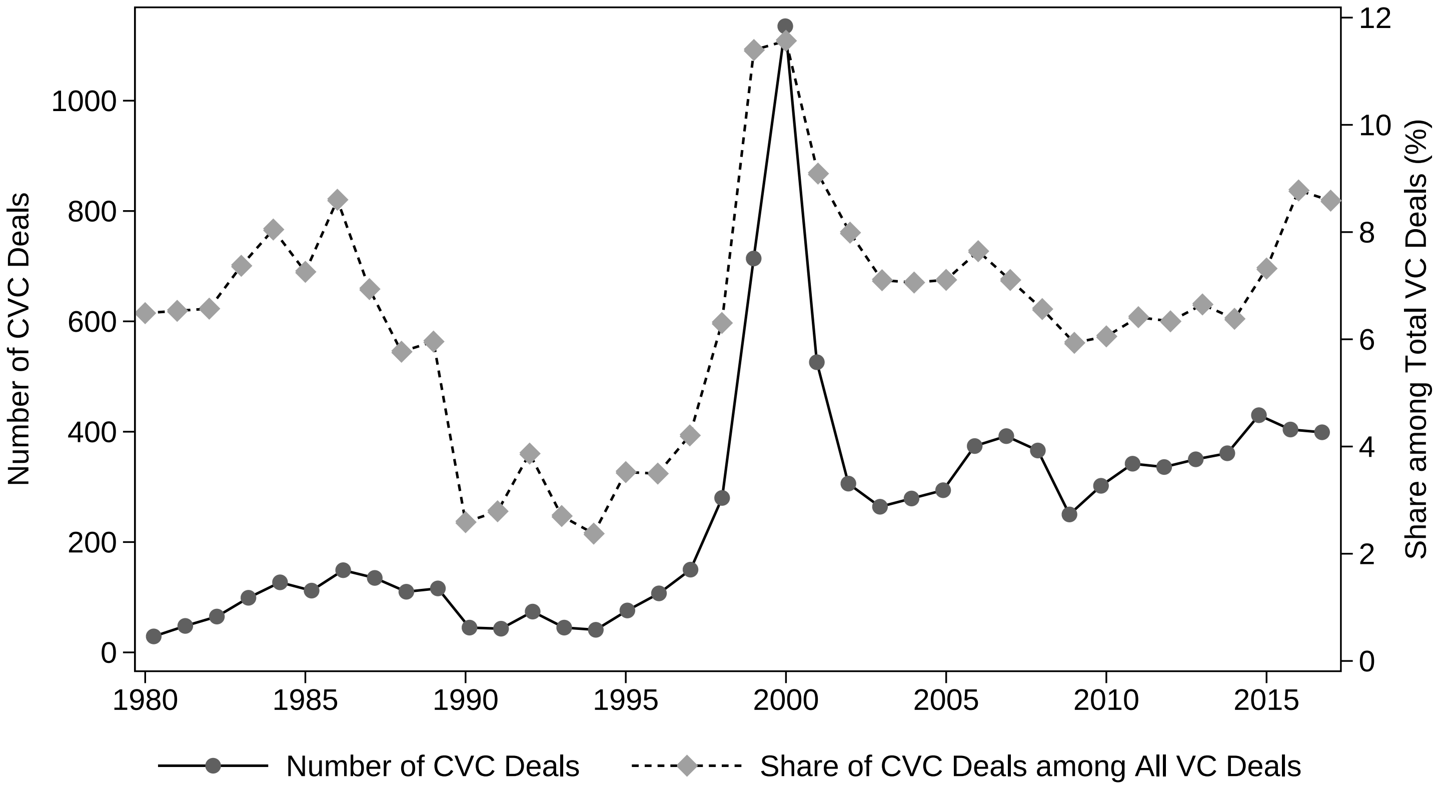
Winters, Terry E, and Donald L. Murfin, 1988, Venture Capital Investing for Corporate Development Objectives, *Journal of Business Venturing* 3, 207–222.

Figure 1. The Overview of the Main Idea and Findings



The figure provides an overview of the main idea and findings in the paper. A CVC program could help identify new business opportunities for its parent firm and further spur the firm to change the firm scope. Two types of measures are used in capturing the firm scope changes: a textual measure based on the annual 10-K filings and Compustat Segment measures. Furthermore, I document that the mechanism through which CVC could identify new business opportunities is the story of experimentation.

Figure 2. Corporate Venture Capital Deals by Calendar Year



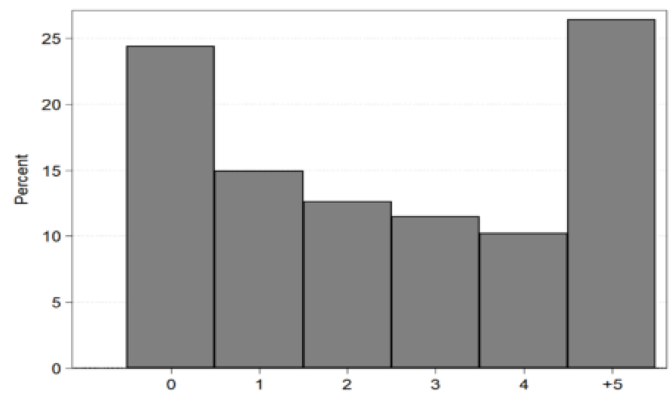
The figure plots the annual CVC investments initiated by US public (non-financial) corporations in Compustat database. The left axis is the number of CVC deals in each year, and the right axis is the share (in percentage) of the CVC deals among all VC deals. The data are mainly obtained from SDC VentureXpert. The data range is from 1980 to 2017.

Figure 3. "Emerging Phrases" and Emerging Business Integration

(a) CVC-backed startups' "emerging phrases" in 2000

(b) CVC-backed startups' "emerging phrases" in 2017

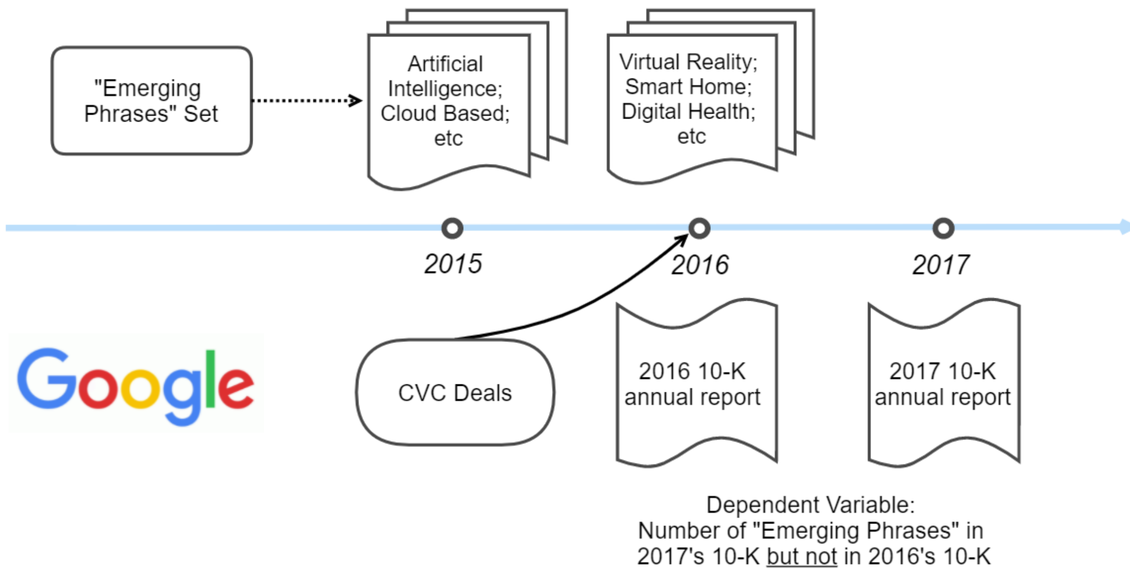
(c) Top 50 "emerging phrases" mostly integrated by CVC parents



(d) Years of surviving of "emerging phrases" in the subsequent 10-Ks

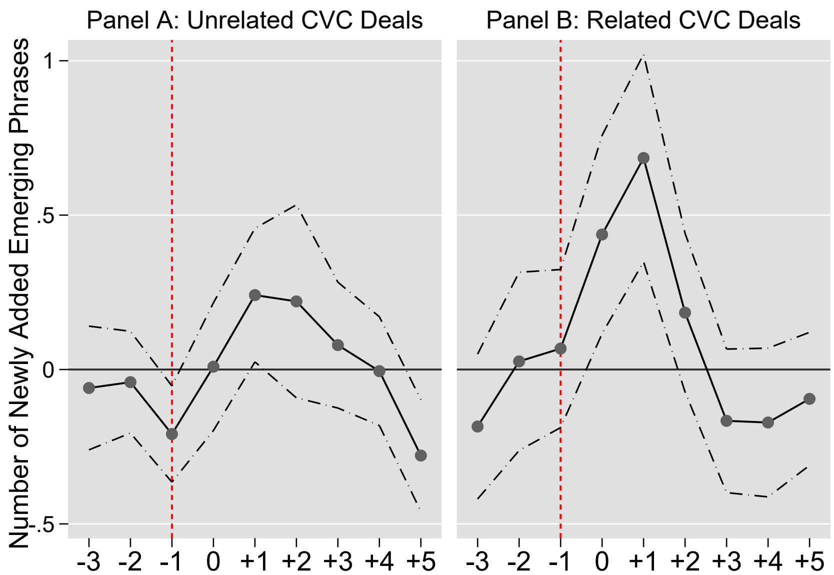
Panel A and B present the words clouds of "emerging phrases" in 2000 and 2017. Emerging phrases are the top 5% most popular word pairs (excluding stopwords and common words) in the detailed business descriptions of all VC-backed startups receiving VC funding in a given year. Panel C plots the top 50 most frequent emerging phrases newly added by CVC parents into 10-K Item 1 (business description) within two years after CVC deals. Panel D plots the distribution of years of surviving of all 2,081 emerging phrases added by CVC parents after investments.

Figure 4. Regression Design in Table II: An Example



The figure explains the regression design in Table II. Take Google as an example. Suppose the Google CVC program (Google Venture) invests in startups in 2016, and during that year, the set of emerging phrases includes virtual reality, digital health, and smart home. Then I search in Google's 2017 10-K (Item 1 Business Description) these three emerging phrases. The dependent variable thus counts the number of 2016 emerging phrases newly added in 2017's 10-K business description. The intuition is that when Google invests in CVC in 2016, it helps Google identify new business opportunities such as digital health, and one year after investment (2017), Google should be more likely to integrate it into its own business.

Figure 5. Firm Scope Change around the CVC Deals (Measured by "Emerging Phrases")

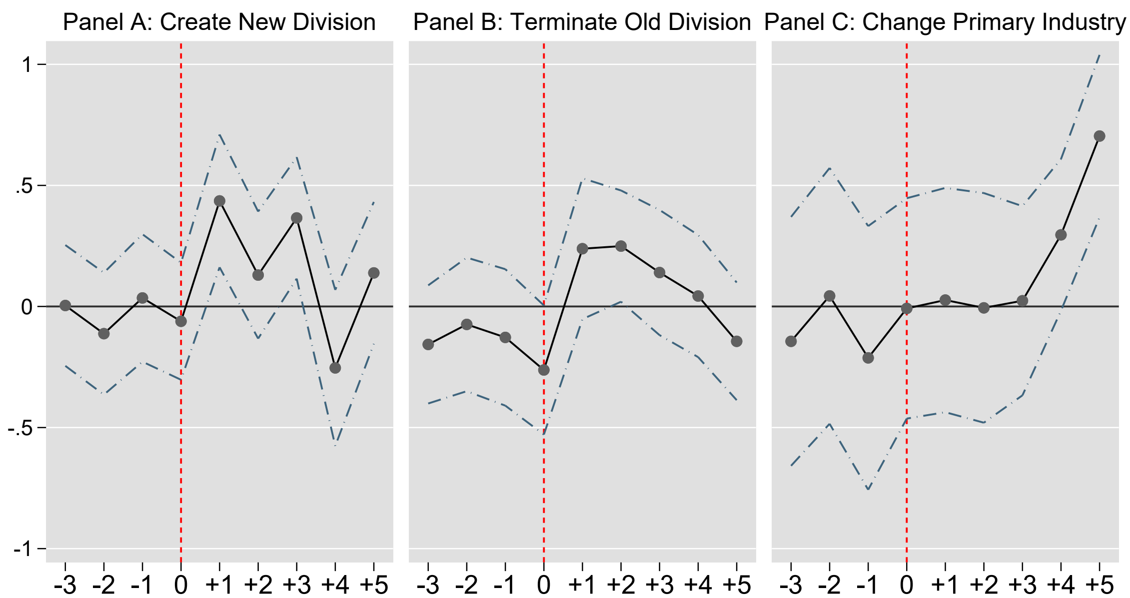


The figure examines the emerging phrases usage in the years around the CVC deals. The estimates (OLS) and confidence intervals are taken from the following regression specification,

$$EmergingPhrases_{i,t} = \sum_{k=-3}^{+5} \gamma_k D(CVC\ Unr; k)_{i,t} + \sum_{k=-3}^{+5} \alpha_k D(CVC\ Rel; k)_{i,t} + \beta \mathbf{X} + \tau_i + \upsilon_t + \varepsilon_{i,t}$$

where the left-hand side variable counts the number of new emerging phrases newly added into the firm's 10-K Item 1, i.e. appears in Year t but not in Year t-1. Emerging phrases (plotted in the online appendix) are the top 5% most popular word pairs (excluding common words and stopwords) taken from all VC-backed startup's business in a given year. $\{D(CVC\ Unr; k)\}_{k=-3}^{+5}$ is a bunch of dummies equal to 1 if the year is k years before or after each CVC unrelated deal. A similar setup applies to $\{D(CVC\ Rel; k)\}_{k=-3}^{+5}$ for CVC related deals. The firm and year fixed effects are included in all regressions. Standard errors are clustered at the firm level. The confidence intervals are calculated at the 90% confidence level. Coefficients of $\{D(CVC\ Unr; k)\}_{k=-3}^{+5}$ and of $\{D(CVC\ Rel; k)\}_{k=-3}^{+5}$ are plotted in Panel A and B separately. $\mathbf{X}$ includes Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., Cash, Sales Growth, HHI, Firm Age, and D(Conglomerate)(lagged).

Figure 6. Firm Scope Change around the CVC Deals (Measured by Segment Variables)

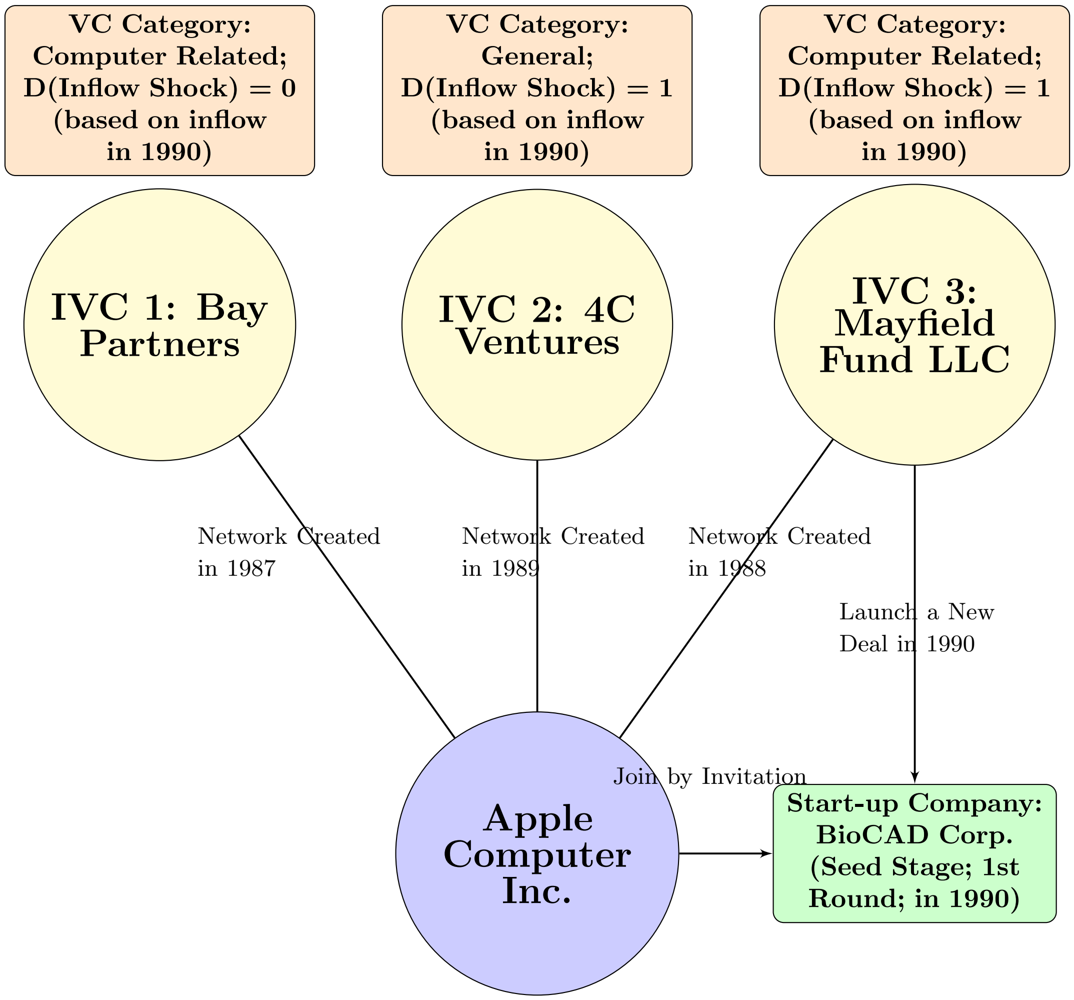


The figure examines the firm scope change in the years around the CVC deals. The estimates (from logit) and confidence intervals are taken from the following regression specification,

$$D[Scope\ Change]_{i,t} = \sum_{k=-3}^{+5} \gamma_k D(CVC\ Unr; k)_{i,t} + \sum_{k=-3}^{+5} \alpha_k D(CVC\ Rel; k)_{i,t} + \beta \mathbf{X} + \tau_i + \upsilon_t + \varepsilon_{i,t}$$

where $D[Scope\ Change]$ denotes three firm scope change dummies regarding creating a new division, removing an old division, and changing the corporate primary industry, respectively, measured in Year t. $\left\{D(CVC\ Unr; k)\right\}_{k=-3}^{+5}$ is a bunch of dummies equal to 1 if the year is k years before or after each CVC unrelated deal. A similar setup applies to $\left\{D(CVC\ Rel; k)\right\}_{k=-3}^{+5}$ for CVC related deals. Firm and year fixed effects are included in all regressions. Standard errors are clustered at the firm level. The confidence intervals are calculated at 90% confidence level. For simplicity, only coefficients of $\left\{D(CVC\ Unr; k)\right\}_{k=-3}^{+5}$ are plotted in the figure. $\mathbf{X}$ includes Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., Cash, Sales Growth, HHI, Firm Age, and D(Conglomerate)(lagged).

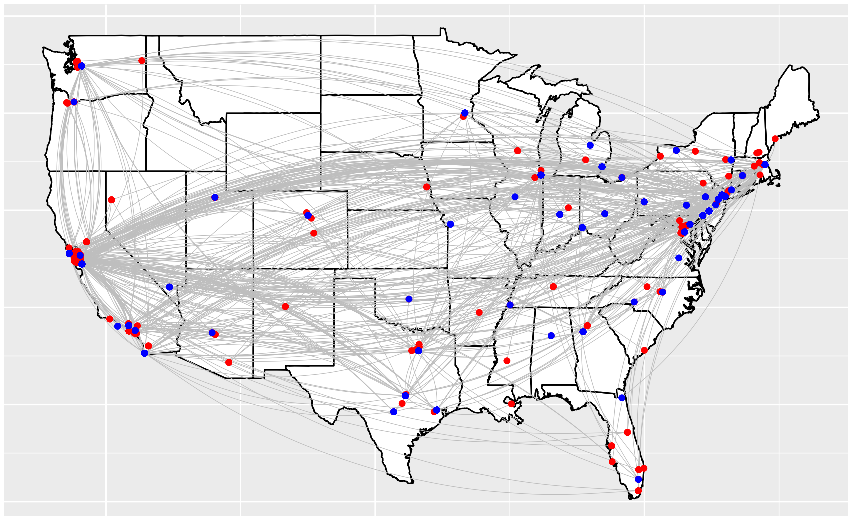
Figure 7. An Example of Instrument Variable of CVC Investments



The figure shows a simple example of the instrument variable of CVC investments. The idea of the instrument is that, if an independent VC firm j (IVC j) receives a positive fund inflow shock today, and meanwhile, the CVC Firm i is in its past syndicate network, then, the IVC j is very likely to initiate new deals and invite CVC Firm i, its old partner, to join in its new investments. Alternatively, IVCs can recommend new deals to CVCs when IVCs start new funds and seek deals. Consider the case illustrated in the above figure. This figure illustrates how its IVC partners drive the CVC investment decision of Apple Computer Inc. In the past five years of 1990, Apple Inc has built three connections with three distinct IVCs through syndicate investments. Among the three IVCs, two received positive inflow shocks in 1990. One of these two IVCs, Mayfield Fund LLC, then spent its new money on investing a seed-stage startup called BioCAD Corp in 1990, followed by the joining of Apple Inc due to the invitation of Mayfield Fund. The idiosyncratic fund inflow shock is constructed following the granular IV approach (Gabaix and Koijen, 2020). The construction of the past 5-year syndication network is illustrated in the text.

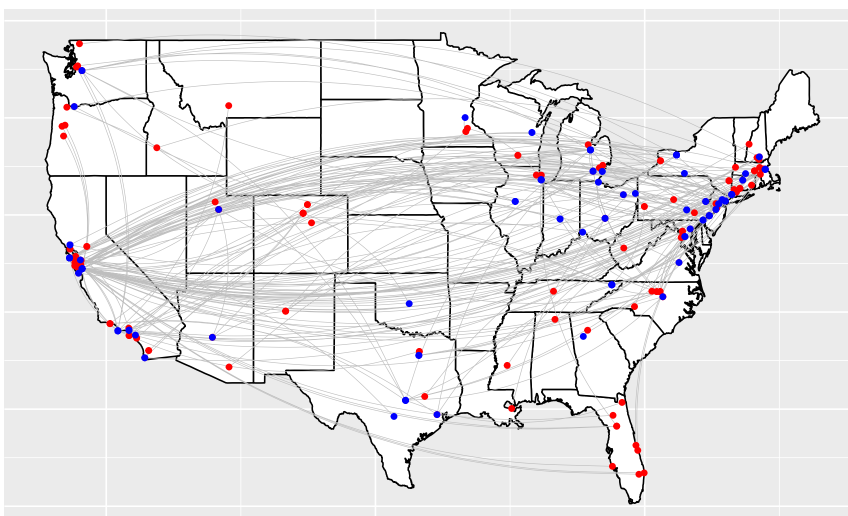
Figure 8. CVC Deals with and without Direct Flights

Panel A: CVC Deals with Direct Flights



Blue Dot: CVC Firm's location; Red Dot: startup's location

Panel B: CVC Deals without Direct Flights



Blue Dot: CVC Firm's location; Red Dot: startup's location

This figure presents the geographic location of CVC deals, where the blue point denotes the CVC firm's location, and the red point denotes the location of the startup. Panel A includes those deals with direct flights, and Panel B draws those deals without direct flights. A CVC deal with direct flight is defined as a deal with a non-stop airline route between the MSA of CVC firm's headquarter and MSA of the startup during the deal year.

Table I: Summary Statistics

This table presents the summary statistics of the firm-year sample used in most regressions. The firm-year sample consists of all observations recorded in both the Compustat and Compustat Historical Segment database from 1980 to 2017. I exclude foreign firms (firms incorporated outside of the US) and firms in financial industries (SIC industry codes starting with 6) from the sample. Furthermore, industries (3-digit SIC) with no CVC activity during the whole sample period are excluded entirely. All variables (except dummies) are winsorized at 1% and 99% levels. In Panel A, the sample is split by D(CVC), the dummy of CVC deals. D(CVC) is equal to 1 if the Firm i conducts at least one CVC deal in Year t. D(New Div.) is a dummy equal to 1 if the firm creates at least one new division in a new industry within the next two years of Year t. D(Div. Rem.) is a dummy equal to 1 if the firm removes at least one old division within the next two years of Year t. D(Chg. Ind.) (3-5) is a dummy equal to 1 if the firm changes the primary corporate industry in the next three to five years. D(Chg. Ind.) (4-6) is defined similarly but based on the next four to six years. Definitions of all other variables are in Appendix A.

Panel A: Firm-Year sample

Variables	D(CVC) = 1			D(CVC) = 0			Test of Mean
	Mean	S.D.	N.	Mean	S.D.	N.	p value
D(New Div.)	0.139	0.346	2,129	0.086	0.280	152,169	0.000
D(Div. Rem.)	0.171	0.377	2,129	0.099	0.298	152,169	0.000
D(Chg. Ind.) (3-5)	0.088	0.336	2,129	0.053	0.280	152,169	0.000
D(Chg. Ind.) (4-6)	0.083	0.339	2,129	0.045	0.276	152,169	0.000
Firm Size	7.979	1.790	2,096	4.430	2.313	129,622	0.000
Tobin's Q	2.592	3.314	1,894	3.472	17.622	124,357	0.030
R&D Exp.	0.088	0.233	2,125	0.203	1.598	150,678	0.000
ROA	0.130	0.286	2,080	-0.082	1.154	136,788	0.000
Book Leverage	0.322	0.297	2,106	0.338	0.549	148,477	0.169
Capx.	0.071	0.089	2,082	0.081	0.116	136,812	0.000
HHI	0.083	0.077	2,129	0.084	0.083	152,169	0.669
Cash	0.195	0.191	2,127	0.189	0.227	151,034	0.236
D(Conglomerate)	0.469	0.499	2,129	0.232	0.422	152,169	0.000

Panel B: CVC related and unrelated deals

CVC Deal Type	Number	Percentage
Related Deals	4,159	38.17%
Unrelated Deals	5,744	52.72%
The startup's SIC-3 code is missing	992	9.11%

Panel C: Change of the firm scope after CVC

	Num. Events
Within the next 2 years following CVC unrelated deals:	
Establish new divisions in new industries	243
Remove obsolete divisions	255
Within the next 3-5 years following CVC unrelated deals:	
Change the corporate primary industry	104
The new division becomes the business of the primary industry	43

Table II: CVC Investments and Firm Scope Change: Emerging Phrases

This table presents the regressions about CVC and the firm scope change measured by emerging phrases. The regression sample consists of all Compustat firms incorporated in the US, with 10-K fillings of Year t and t-1 searchable in SEC, and are not in financial industries. Industries (defined as 3-digit SIC) with no CVC activity during the whole sample period are excluded entirely. The dependent variable is defined as the number of "Emerging Phrases" newly added in the next year (or in the second year)'s 10-K business description. The Emerging Phrases are those top 5% most frequently-used word pairs (excluding stopwords and common words) in the detailed business descriptions of all VC-backed startups receiving VC funding in a given year. Column (1) - (3) count those "Emerging Phrases" appearing in Year t+1's 10-K Item 1 but not in the Year t. T-statistics are shown in parentheses, and standard errors are clustered by firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
OLS	Num. of "Emerging Phrases" Newly Added in 10-K Item 1					
	in Year t+1's business *but not* in Year t's business			in Year t+2's business *but not* in Year t+1's business		
D(CVC)	0.776***			0.676***		
	(7.36)			(6.53)		
D(CVC Unrelated)		0.442***	0.300***		0.562***	0.331**
		(3.82)	(2.82)		(4.41)	(2.40)
D(CVC Related)		0.782***	0.873***		0.433***	0.301**
		(5.23)	(5.45)		(3.48)	(2.31)
Num. New Word Pairs Added in t+i (÷1000)	0.863***	0.862***	0.937***	0.865***	0.865***	0.935***
	(33.49)	(33.58)	(30.50)	(32.56)	(32.58)	(31.14)
Firm-level Controls	Firm Size, Tobin's Q, ROA, R&D, Leverage, Cash, Sales Growth, Capx., HHI, D(Conglomerate), Firm Age, 10-K (Item 1) Text Length					
Year × Industry F.E.	✓	✓		✓	✓	
Year F.E.			✓			✓
Firm F.E.			✓			✓
Num. Obs.	50,931	50,931	49,916	46,749	46,749	45,856
Adj. R^2	0.394	0.394	0.425	0.379	0.379	0.419

Table III: Industry-specific CVC Investments and Industry-specific Emerging Phrases

This table presents the regressions about industry-specific CVC investments and industry-specific emerging phrases newly used by US public firms. The left-hand side variable is the number of industry-specific emerging phrases that are newly added into the firm's annual 10-K Item 1. The dependent variable takes the natural logarithm transformation (ln(1+variable)). Each emerging phrase is sorted into eight VEIC industries. The control variables are dummies of industry-specific CVC investments in the past three years. CVC investments are again sorted into eight VEIC industries. T-statistics are shown in parentheses, and standard errors are clustered by firm and year level. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of Newly Added Emerging Phrases (VEIC Industry Specific)							
VEIC Industry	Biotech- -nology	Communic- -ation	Computer Hardware	Computer Software	Internet Specific	Medical Health	Non-High- -Tech	Others
D(CVC in Biotechnology)	0.057**	0.006	0.021	-0.000	-0.027	0.014	0.012	0.009
	(2.315)	(0.388)	(1.517)	(-0.013)	(-0.803)	(0.762)	(0.818)	(0.442)
D(CVC in Communication)	-0.005	0.075**	-0.002	0.019	0.052	-0.005	-0.001	0.026*
	(-1.592)	(2.761)	(-0.176)	(0.512)	(1.283)	(-1.007)	(-0.088)	(1.864)
D(CVC in Computer Hardware)	0.010	0.004	-0.002	-0.045	0.028	-0.004	-0.004	-0.021
	(1.233)	(0.118)	(-0.154)	(-1.294)	(0.825)	(-0.349)	(-0.258)	(-1.260)
D(CVC in Computer Software)	-0.006	0.016	0.003	0.070**	0.022	0.008	0.002	-0.000
	(-1.069)	(0.936)	(0.385)	(2.573)	(0.753)	(1.333)	(0.160)	(-0.005)
D(CVC in Internet Specific)	0.008	0.008	0.009	0.084***	0.113**	-0.002	0.004	0.012
	(1.546)	(0.471)	(1.224)	(3.137)	(2.380)	(-0.606)	(0.436)	(1.243)
D(CVC in Medical Health)	0.016	-0.023	-0.019	0.049	0.055*	0.010	0.008	0.011
	(0.931)	(-1.016)	(-1.577)	(1.541)	(1.801)	(0.592)	(0.508)	(0.835)
D(CVC in Non-high-tech)	-0.009	-0.005	0.007	0.034	0.007	-0.012	-0.024**	0.005
	(-1.487)	(-0.260)	(0.628)	(1.179)	(0.270)	(-1.540)	(-2.196)	(0.447)
D(CVC in Others)	-0.011**	0.002	0.006	0.001	-0.010	-0.001	0.005	0.047***
	(-2.519)	(0.071)	(0.471)	(0.026)	(-0.323)	(-0.114)	(0.659)	(3.913)
Firm F.E.	✓	✓	✓	✓	✓	✓	✓	✓
Industry × Year F.E.	✓	✓	✓	✓	✓	✓	✓	✓
Num. Obs.	50,931	50,931	50,931	50,931	50,931	50,931	50,931	50,931
Adj. R^2	0.156	0.227	0.049	0.281	0.320	0.055	0.055	0.087

Table IV: CVC Investments and Firm Scope Change: Segment Measures

This table provides the estimate of logistic regressions about CVC investments and the subsequent firm scope change by CVC corporate parents. The regression sample consists of all Compustat firms which are incorporated in the US and are not in financial industries. Industries (defined as 3-digit SIC) with no CVC activity during the whole sample period are excluded entirely. Panel A (Columns 1 – 3) investigates the scenario of creating new divisions. The dependent variable is a dummy equal to 1 if the firm creates at least one new division within the next two years (Year t+1 and Year t+2). Establishing a new division is identified if the firm reports a new division with its SIC-3 code appearing in the first time in the company history. Panel A (Columns 4 – 6) studies the situation of removing old divisions. The dependent variable is a dummy equal to 1 if the firm removes at least one old division within the next two years. Panel B investigates the change of the primary corporate business. The dependent variable is a dummy equal to 1 if the firm's primary industry has changed in the next 3 to 5 years. About control variables, D(CVC) is a dummy equal to 1 if the firm invests in CVC deals in Year t. The D(CVC) variable is further divided into two variables in Columns (2) and (3) of each panel. D(CVC Related) is a dummy equal to 1 if the firm conducts at least one related CVC deal in Year t. The related CVC deal is the CVC deal related to the existing business of the corporate parent. D(CVC Unrelated) is a dummy equal to 1 if the firm conducts at least one unrelated CVC deal in Year t. The regression sample is further adjusted to alleviate the survivorship bias within the next two years for Panel A and B and within the next 3–5 years for Panel C. Industry fixed effects are defined in SIC-2 Industries. T-statistics are shown in parentheses, and standard errors are clustered by firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Panel A: Creating new divisions and removing old divisions						
Conditional Logit	(1)	(2)	(3)	(4)	(5)	(6)
	D(Create New Division)			D(Remove Old Division)		
Period	[t+1, t+2]			[t+1, t+2]		
D(CVC)	0.350***			0.323**		
	(2.68)			(2.57)		
D(CVC Unrelated)		0.531***	0.434***		0.464***	0.195
		(3.74)	(2.91)		(3.10)	(1.21)
D(CVC Related)		-0.294	-0.00921		-0.195	-0.231
		(-1.42)	(-0.04)		(-1.00)	(-1.06)
Division Creation/Removal in the Past 2 Years	0.189***	0.189***		0.283***	0.284***	
	(3.82)	(3.81)		(6.57)	(6.59)	
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., Cash, HHI, D(Conglomerate), Firm Age					
Year × Industry F.E.	✓	✓		✓	✓	
Year F.E.			✓			✓
Firm F.E.			✓			✓
Num. Obs.	86,030	86,030	42,584	87,066	87,066	39,191
Pseudo R^2	0.026	0.027	0.069	0.166	0.166	0.099
Prob. Increased						
by D(CVC) = 1	+3.45%	–	–	+3.23%	–	–
by D(CVC Unrelated) = 1	–	+5.86%	4.91%	–	5.02%	2.73%

Panel B: Change corporate primary industry						
Conditional Logit	(1)	(2)	(3)	(4)	(5)	(6)
Period	D(Change Industry) [t+3, t+5]			D(Change Industry) [t+4, t+6]		
D(CVC)	0.479**			0.501**		
	(2.45)			(2.40)		
D(CVC Unrelated)		0.524***	0.446**		0.608***	0.535**
		(2.60)	(2.10)		(2.85)	(2.57)
D(CVC Related)		-0.0128	-0.0161		-0.226	-0.329
		(-0.04)	(-0.04)		(-0.72)	(-0.89)
Change Primary Industry in the Past 2 Years	0.762***	0.762***		0.740***	0.740***	
	(12.43)	(12.42)		(11.20)	(11.19)	
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., Cash, HHI, D(Conglomerate), Firm Age					
Year × Industry F.E.	✓	✓		✓	✓	
Year F.E.			✓			✓
Firm F.E.			✓			✓
Num. Obs.	82,339	82,339	22,751	80,056	80,056	21,202
Pseudo R^2	0.071	0.070	0.062	0.062	0.062	0.076
Prob. Increased						
by D(CVC) = 1	+3.14%	–	–	+3.12%	–	–
by D(CVC Unrelated) = 1	–	+3.56%	3.08%	–	4.04%	3.58%

Table V: CVC Investments and Firm Scope Change: CVC Diversification Strategies

This table presents the diversification strategy in CVC investments and the firm scope change. The regression sample and definitions of dependent variables follow Table IV. D[New Div.] is a dummy equal to 1 if the firm establishes a new division within the next two years; D[Chg.Ind.] is equal to 1 if the firm changes the corporate primary business (industry) in the next 3-5 years. Num. Emerging Phrases is the number of "Emerging Phrases" newly added by the firm into its annual 10-K business description in Year t+1. D[CVC Past 3yr] is a dummy equal to 1 if the firm conducts at least one CVC deal in the past three years. Inverse HHI(VEIC) is the inverse of the HHI measure regarding the past three-year CVC deals across 10 VEIC industries. Num(VEIC) is the number of VEIC industries in which the firm has CVC investments during the past three years. Industry×Year fixed effects are defined in SIC-2 industries. T-statistics are shown in parentheses, and standard errors are clustered by firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Period	D[New Division] [t+1, t+2]		D[Change Industry] [t+3, t+5]		Num. Emerging Phrases in t+1 But not in t	
D[CVC Past 3yr]	0.158	-0.0101	0.333***	0.210*	0.394***	0.269**
	(1.63)	(-0.09)	(3.65)	(1.84)	(4.20)	(2.04)
D[CVC Past 3yr] × Inverse HHI(VEIC) (Past 3yr)	0.507**		0.725***		0.518*	
	(2.35)		(3.54)		(1.88)	
D[CVC Past 3yr] × Num(VEIC) (Past 3yr)		0.157***		0.168***		0.139**
		(3.62)		(3.68)		(2.09)
D[CVC Past 3yr] × Num Deals (Past 3yr)	0.000414	-0.00485	-0.00517***	-0.0113***	0.0102***	0.00656***
	(0.20)	(-1.50)	(-2.63)	(-2.98)	(4.13)	(2.66)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., CASH, sale_grt, HHI, D(Conglomerate), Age					
Industry*Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. Obs.	86,310	86,310	84,460	84,460	45,437	45,437
Pseudo R^2 / Adj. R^2	0.028	0.028	0.076	0.076	0.386	0.386

Table VI: Discrete Choice Model of Division Creation

This table presents the estimate of a discrete choice model (McFadden (1973)). The observations are at the firm-year-industry level. Each observation represents the alternative (Industry j) where Firm i in Year t could choose to create a new division. The set of alternatives (choice set) consists of 404 non-financial SIC-3 industries that haven been documented at least once in the Compustat Historical Segment database from 1980 to 2017. The set of alternatives varies across each firm-year pair (case). In the choice model, I only include firms investing at least one CVC deal from 1980 to 2017. For each firm-year pair (case), I drop those industries that already exist as divisions of the firm in Year t-1 and those that have already been created before Year t. The dependent variable is a dummy equal to 1 if the Firm i in Year t creates a new division in Industry j. D(CVC 3yr) is a dummy equal to 1 if, within the last three years, the Firm i has invested in CVC deals in Industry j. D(Ind. Proxy SIC1) and D(Ind. Proxy SIC2) capture industry proximity between the alternative and the industries of existing divisions of Firm i in Year t-1. D(Ind. Proxy SIC2) is a dummy equal to 1 if the alternative has the same 2-digit SIC with one of the existing divisions of Firm i. D(Ind. Proxy SIC1) is a dummy equal to 1 if the alternative has the same 1-digit SIC with one of the existing divisions of Firm i, but does not have the same 2-SIC with them. D(Ind. Services) are those industries starting with 7 in SIC-3. The conditional logit regression is grouped in the firm-year level. Standard errors are clustered by firm-year. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)
Conditional Logit	D(Create New Division)		
D(CVC 3yr)	3.249***	4.378***	4.004***
	(13.31)	(20.33)	(14.70)
D(Ind. Proxy SIC2)	2.830***	2.997***	2.988***
	(25.33)	(27.30)	(27.28)
D(Ind. Proxy SIC1)	0.751***	0.834***	0.823***
	(6.83)	(7.37)	(7.29)
D(CVC 3yr) × D(Ind. Proxy SIC2)		-2.854***	-2.724***
		(-6.46)	(-5.96)
D(CVC 3yr) × D(Ind. Proxy SIC1)		-1.594***	-1.431***
		(-3.12)	(-2.65)
D(CVC 3yr) × D(Ind. Business Services)			0.662**
			(1.99)
CLOGIT Grouped by Firm-Year	✓	✓	✓
Industry F.E.	✓	✓	✓
Num. Obs.	234,539	234,539	234,539
Pseudo R^2	0.131	0.138	0.138

Table VII: CVC Signal Response and Division Creation

This table studies the CVC signal and division creation following the signal. The setup and estimate of the discrete choice model follow Table VI. The dependent variable is a dummy equal to 1 if the Industry j is chosen by the Firm i in Year t to establish a new division. Each signal variable is constructed and based on the past three-year CVC investments in the given industry and is interacted with the D(CVC 3yr) dummy, which is equal to 1 if the firm has made CVC investments in that industry within the past three years. Panel A uses the main setup to construct signal variables, and Panel B uses the fraction as the definition. Regarding CVC signal variables, Num(Startups IPO) is the number of Industry-j startups (invested within three years before) that finally exit through IPO (IPO date after Year t is allowed). Num(Startups Acquired) is the number of Industry-j startups acquired by the third-party (not acquired by the CVC parent firm itself). The conditional logit regression is grouped at the firm-year level. Standard errors are clustered by firm and year. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Panel A: Main Setup

Conditional Logit	(1)	(2)	(3)	(4)	(5)	(6)
			D(Create New Division)			
D(CVC 3yr)	3.486***	3.729***	3.665***	3.606***	3.660***	3.492***
	(11.55)	(13.34)	(12.79)	(11.61)	(12.82)	(11.04)
D(CVC 3yr) × D(Ind. Proxy SIC2)	-2.363***	-2.437***	-2.441***	-2.480***	-2.448***	-2.520***
	(-5.06)	(-5.12)	(-5.13)	(-5.21)	(-5.11)	(-5.41)
D(CVC 3yr) × D(Ind. Proxy SIC1)	-1.561**	-1.665**	-1.556**	-1.547**	-1.809***	-1.836***
	(-2.48)	(-2.42)	(-2.40)	(-2.27)	(-2.68)	(-2.68)
D(CVC 3yr) × Num(Startups IPO)	0.701***					
	(2.90)					
D(CVC 3yr) × Num(Startups Acquired by Third Party)		0.0940				
		(0.59)				
D(CVC 3yr) × Num(Startups Acquired by CVC Parent Itself)			1.578***			
			(3.30)			
D(CVC 3yr) × Num(Startups Bankruptcy)				-3.811		
				(-1.40)		
D(CVC 3yr) × Patent Growth Signal					0.0301**	
					(2.52)	
D(CVC 3yr) × Positive Patent Signal						0.796**
						(2.11)
D(CVC 3yr) × Negative Patent Signal						-0.253
						(0.46)
D(CVC 3yr) × Num(Startups Invested)	-0.0292	0.0120	0.0353	0.144**	0.0525	0.0531
	(-0.56)	(0.14)	(0.90)	(2.52)	(1.27)	(1.28)
Industry Cluster Controls						
IPO Cluster	0.0226***	0.0226***	0.0227***	0.0226***	0.0226***	0.0226***
	(9.56)	(9.55)	(9.56)	(9.55)	(9.56)	(9.55)
Acquisition Cluster	0.0123***	0.0123***	0.0123***	0.0123***	0.0123***	0.0122***
	(5.79)	(5.78)	(5.80)	(5.79)	(5.79)	(5.77)
Patent Cluster	0.000405*	0.000407**	0.000407**	0.000406**	0.000408**	0.000410**
	(1.96)	(1.97)	(1.97)	(1.96)	(1.97)	(1.98)
D(CVC 3yr) × IPO Cluster	-0.0173***	-0.0128**	-0.0126**	-0.0125**	-0.0141***	-0.0128**
	(-3.26)	(-2.47)	(-2.43)	(-2.44)	(-2.60)	(-2.40)
D(CVC 3yr) × Acq. Cluster	-0.00493	-0.00747**	-0.00684**	-0.00790**	-0.00621*	-0.00759**
	(-1.37)	(-2.27)	(-2.06)	(-2.39)	(-1.86)	(-2.29)
D(CVC 3yr) × Patent Cluster	-0.0000594	-0.0000714	-0.000113	-0.0000563	-0.000180	-0.000167
	(-0.20)	(-0.24)	(-0.37)	(-0.19)	(-0.57)	(-0.54)
CLOGIT grouped by Firm-Year	✓	✓	✓	✓	✓	✓
Industry F.E.	✓	✓	✓	✓	✓	✓
Num. Obs.	234,539	234,539	234,539	234,539	234,539	234,539
Pseudo R^2	0.181	0.180	0.181	0.181	0.181	0.181

Panel B: Alternative Construction of Signals

Conditional Logit	(1)	(2)	(3) D(Create New Division)	(4)	(5)
D(CVC 3yr)	3.416***	3.759***	3.695***	3.751***	3.691***
	(10.45)	(11.57)	(12.92)	(13.42)	(13.01)
D(CVC 3yr) × D(Ind. Proxy SIC2)	-2.352***	-2.394***	-2.390***	-2.411***	-2.372***
	(-4.93)	(-4.95)	(-4.98)	(-4.98)	(-4.90)
D(CVC 3yr) × D(Ind. Proxy SIC1)	-1.560**	-1.485**	-1.460**	-1.434**	-1.555**
	(-2.50)	(-2.43)	(-2.41)	(-2.38)	(-2.49)
D(CVC 3yr) × Frac(Startups IPO)	0.858**				
	(1.98)				
D(CVC 3yr) × Frac(Startups Acquired) by Third Party)		-0.0400			
		(-0.10)			
D(CVC 3yr) × Frac(Startups Acquired) by CVC Parent Itself)			1.552*		
			(1.69)		
D(CVC 3yr) × Frac(Startups Bankruptcy)				-8.688	
				(-0.48)	
D(CVC 3yr) × Frac(Positive Patent Signal)					0.0208
					(1.48)
Industry Cluster Controls					
IPO Cluster	0.0227***	0.0226***	0.0226***	0.0226***	0.0226***
	(9.56)	(9.55)	(9.55)	(9.55)	(9.55)
Acquisition Cluster	0.0122***	0.0123***	0.0123***	0.0122***	0.0122***
	(5.76)	(5.77)	(5.77)	(5.76)	(5.76)
Patent Cluster	0.000409**	0.000410**	0.000410**	0.000410**	0.000411**
	(1.98)	(1.98)	(1.98)	(1.98)	(1.98)
D(CVC 3yr) × IPO Cluster	-0.0149***	-0.0133**	-0.0131**	-0.0134**	-0.0133**
	(-2.77)	(-2.52)	(-2.48)	(-2.54)	(-2.48)
D(CVC 3yr) × Acquisition Cluster	-0.00468	-0.00593*	-0.00587*	-0.00591*	-0.00566*
	(-1.40)	(-1.89)	(-1.87)	(-1.90)	(-1.79)
D(CVC 3yr) × Patent Cluster	-0.0000563	-0.0000576	-0.0000738	-0.0000527	-0.0000723
	(-0.19)	(-0.19)	(-0.24)	(-0.18)	(-0.24)
CLOGIT grouped by Firm-Year	✓	✓	✓	✓	✓
Industry F.E.	✓	✓	✓	✓	✓
Num. Obs.	234,539	234,539	234,539	234,539	234,539
Pseudo R^2	0.181	0.180	0.180	0.180	0.180

Table VIII: CVC Signal Response and Adding Emerging Phrases

This table presents the analysis of CVC investments and the integration of emerging business across VEIC industries. The sample in Column (1) is at the Firm-Year-VEIC level, while the remaining columns use the Firm-Year level data. The dependent variable is the number of VEIC-j specific emerging phrases newly added into the 10-K. Each emerging phrase is sorted into 8 VEIC industries. 8 VEIC industries are Biotechnology; Communication and Media; Computer Hardware; Computer Software; Internet Specific; Medical and Health; Non-High-Tech; and Others. D(CVC VEIC j) is equal to 1 if the firm invests at least one CVC-backed startup in the VEIC Industry j in the past three years. T-statistics are shown in parentheses, and standard errors are clustered by firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Panel A: By VEIC industry	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Number of Newly Added Emerging Phrases (VEIC j)								
	All VEIC	Biotech- -nology	Communic- -ation	Computer Hardware	Computer Software	Internet Specific	Medical Health	Non-High- -Tech	Others
D(CVC VEIC j)	0.142***	0.0589**	0.0809***	0.00213	0.103***	0.131**	0.0107	-0.0204*	0.0541***
	(10.89)	(2.45)	(3.01)	(0.16)	(3.04)	(2.64)	(0.68)	(-1.75)	(4.39)
Firm F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
VEIC*Year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Num. Obs.	616,544	75,578	75,578	75,578	75,578	75,578	75,578	75,578	75,578

Panel B: Interact with the signal variable	(1)	(2)	(3)	(4)	(5)
	Number of Newly Added Emerging Phrases (VEIC j)				
D(CVC VEIC j)	0.120***	0.122***	0.119***	0.122***	0.122***
	(8.17)	(7.08)	(6.99)	(7.07)	(7.06)
D(CVC VEIC j) × Num(Startups IPO)	0.0189**				
	(1.99)				
D(CVC VEIC j) × Num(Startups Acquired by Third Party)		0.00000			
		(0.00)			
D(CVC VEIC j) × Num(Startups Acquired by Parent Itself)			0.112***		
			(3.51)		
D(CVC VEIC j) × Num(Startups Bankruptcy)				-0.0363	
				(-0.74)	
D(CVC VEIC j) × Patent Signal					0.00614***
					(3.31)
D(CVC VEIC j) × Num(Startups Invested)	0.00818***	0.00999***	0.00935***	0.0102***	0.00997***
	(4.90)	(3.45)	(3.98)	(3.49)	(3.71)
Firm F.E.	✓	✓	✓	✓	✓
VEIC*Year F.E.	✓	✓	✓	✓	✓
Num. Obs.	616,544	616,544	616,544	616,544	616,544
Adj. R^2	0.183	0.183	0.183	0.183	0.183

Table IX: First Stage Regression regarding CVC Instrument

This table presents the first stage regression regarding the instrument variable of CVC investments. I use the VC fund inflow shock of those independent VC firms in the past 5-year syndicate network of CVC Firm i as the instrument of CVC investments by the CVC Firm i. Figure 7 provides an example about how the instrument works. The regression sample follows Table IV and further requires that the firm has invested at least one CVC deal in the past five years (and thus enjoys some networks with IVCs). The instrument variable, Granular IV, is defined as the sum of the idiosyncratic fund inflow shocks of those IVCs in the past 5-year syndicating network. Num(IVC in the Network) is the natural logarithm of one plus the number of IVCs in the past 5-year syndication network of CVC Firm i. Industry VC Deal Flow is measured by the total amount of VC deals in the SIC-2 industry in Year t. The dependent variable Num(CVC Deal) (Num(CVC Initial Deal) for Column (4) to (6)) is the natural logarithm of one plus the number of CVC deals (CVC initial deals) conducted by the Firm i in Year t. The CVC initial deal is defined as the deal in which case the CVC firm invests in an entrepreneurial Start-up j for the first time, that is, not the follow-on investments. The standard errors are clustered at the CVC firm level. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Num(CVC Deal)			Num(CVC Initial Deal)		
Granular IV (IVC Fund Inflow Shock)	0.100***	0.0987***	0.0937***	0.0866***	0.0858***	0.0824***
	(14.59)	(13.73)	(11.48)	(13.62)	(13.90)	(11.17)
Num(IVC in the Network)	0.182***	0.189***	0.218***	0.123***	0.128***	0.148***
	(5.00)	(5.03)	(4.69)	(4.30)	(4.33)	(3.99)
IVC's Average Age In the Network		-0.0146***	-0.0225***		-0.0144***	-0.0224***
		(-3.84)	(-4.00)		(-3.83)	(-4.04)
IVC's Average Past IPO In the Network		0.00870	-0.00176		0.0132	0.00279
		(1.00)	(-0.15)		(1.53)	(0.23)
Industry VC Deal Flow	0.00204	0.00207		0.00197	0.00203	
	(1.42)	(1.45)		(1.28)	(1.32)	
D(CVC Past 1yr)	0.311***	0.297***	0.302***	0.173***	0.160***	0.155***
	(12.42)	(12.12)	(9.17)	(7.74)	(7.15)	(4.75)
D(CVC Past 2yr)	0.0815***	0.0742***	0.0921***	0.0250	0.0179	0.0373
	(3.71)	(3.28)	(3.15)	(1.26)	(0.87)	(1.37)
D(CVC Past 3yr)	0.0166	0.0100	0.0257	-0.00219	-0.00840	-0.000671
	(0.63)	(0.38)	(0.84)	(-0.09)	(-0.33)	(-0.02)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglo), Age					
Year Fixed Effect	Yes	Yes	No	Yes	Yes	No
Industry Fixed Effect	Yes	Yes	No	Yes	Yes	No
Industry*Year Fixed Effect	No	No	Yes	No	No	Yes
Num. Obs.	3,236	3,236	3,236	3,236	3,236	3,236
Adj. R^2	0.539	0.548	0.560	0.481	0.487	0.497

Table X: CVC Investments and Firm Scope Change: 2SLS Estimator

This table presents the 2SLS regression regarding CVC investments and the subsequent firm scope change. The regression sample consists of all Compustat firms which are incorporated in the US and conduct at least one CVC deal in the past five years (and thus enjoy the IVC network formed by the past investments). In Column (2), the left-hand side variable, Business Change, captures the general business change of a CVC parent firm. It is defined as one minus the cosine similarity between the firm's textual business description in Year t and Year t+1. The variable construction follows Hoberg et al. (2014). The instrument variable, Granular IV, is defined as the sum of the idiosyncratic fund inflow shocks of those IVCs in the past 5-year syndicating network. Num(CVC Initial Deal) is the natural logarithm of one plus the number of CVC deals (excluding follow-on investments) conducted by the Firm i in Year t. Industry fixed effects are defined in SIC-3 Industries. T-statistics are shown in parentheses, and standard errors are clustered at the CVC Firm level. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

2SLS	(1)	(2)	(3)	(4)	(5)	(6)
	Textual Measure			Segment Dummies		
	Emerging Phrases	Business Changes	New Products	New Division	Remove Divisions	Change Industry
Time Period	[t+1]	[t+1]	[t+1]	[t+1, t+2]	[t+1, t+2]	[t+3, t+5]
Num(CVC Initial Deals)	0.851***	3.766**	0.347**	0.079**	-0.028	0.065**
(Instrumented by GIV)	(2.728)	(2.501)	(2.364)	(2.022)	(-0.559)	(2.158)
Num(IVC in the Network)	-0.033	-0.180	-0.033	0.012	0.008	0.015
	(-0.241)	(-0.228)	(-0.528)	(0.665)	(0.366)	(0.749)
IVC's Average Age	0.048**	0.128	0.020*	0.001	-0.001	0.000
in Network	(2.303)	(1.015)	(1.793)	(0.394)	(-0.234)	(0.145)
IVC's Average Past IPO	-0.207**	-0.287	-0.026	-0.001	0.008	-0.006
in Network	(-2.069)	(-1.069)	(-0.984)	(-0.301)	(1.519)	(-1.047)
D[CVC Past 1yr]	0.072	0.540	0.114	0.007	0.000	0.007
	(0.474)	(0.517)	(1.182)	(0.355)	(0.016)	(0.367)
D[CVC Past 2yr]	-0.135	-1.329	-0.034	-0.009	0.010	0.013
	(-1.031)	(-1.526)	(-0.411)	(-0.702)	(0.638)	(1.156)
D[CVC Past 3yr]	-0.068	-0.407	0.137	0.008	0.012	0.004
	(-0.491)	(-0.439)	(1.550)	(0.585)	(0.760)	(0.323)
Kleibergen-Paap F statistic	192.46	107.99	63.21	127.06	127.06	127.06
Other Firm Controls	✓	✓	✓	✓	✓	✓
Industry*Year F.E.	✓	✓	✓	✓	✓	✓
Num. Obs.	1450	1569	567	2474	2474	2474
R^2	0.065	0.030	0.419	0.026	0.083	0.051

Table XI: CVC Investments and Firm Scope Change: Evidence with Airline Route

This table presents the post-CVC scope change analysis using the US airline route as a quasi-natural experiment. I match the CVC deals sample with the US T-100 Airline Domestic Segment database from 1990 to 2017. CVC deals sample only includes CVC unrelated deals and deals in which the start-up is located in the US. A CVC deal is identified as *a deal with direct flights* if there are direct airline flights, during the year right after the investment (deal) year, between the metropolitan statistical area (MSA) of the CVC firm and the MSA of the start-up's headquarter. Panel A provides summary statistics about the CVC deal sample. Deals are broken down by those with and without direct flights. I exclude deals in which the start-up and the CVC firm are located in the same MSA. Panel B and Panel C provide OLS regressions estimated with the CVC deal level sample. The dependent variable in Panel B is a dummy equal to 1 if the CVC firm creates a new division within the next two years after the deal and the newly created division is in the same SIC-3 industry of the start-up in the deal. In Panel C, the dependent variable is a dummy equal to 1 if the CVC firm changes its primary industry within 3-5 years after the deal and the industry the firm changes to is the same as the start-up's industry in the deal. CVC Parent controls are Firm Size, ROA, Book Leverage, Capx, HHI, and D(Conglomerate). T-statistics are shown in parentheses, and standard errors are clustered by CVC firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Panel A: Summary statistics

CVC Deals Sample	With Direct Flights			Without Direct Flights			Test of Mean
	Mean	S.D.	N	Mean	S.D.	N	p value
Location Variables							
% Start-ups in CA	46.91%	0.499	2262	38.69%	0.487	765	0.000
% Start-ups in NY	9.55%	0.294	2262	3.40%	0.181	765	0.000
% Start-ups in MA	14.32%	0.350	2262	10.07%	0.301	765	0.002
% CVC Firms in CA	47.75%	0.500	2262	50.72%	0.500	765	0.155
% CVC Firms in NY	8.71%	0.282	2262	6.67%	0.250	765	0.075
% CVC Firms in MA	3.09%	0.173	2262	3.27%	0.178	765	0.812
Distance (miles)	1347.42	988.10	2222	1263.34	995.51	749	0.045
Start-up Variables							
Start-up's Age	6.625	5.826	2002	7.181	6.120	689	0.033
Num. Co-investors	5.722	3.762	2262	6.022	3.679	765	0.055
CVC Paret Variables							
Firm Size (Total Sales)	9.579	1.773	2259	9.397	1.700	765	0.013
ROA	0.156	0.266	2247	0.159	0.179	762	0.810
Book Leverage	0.079	0.213	2259	0.069	0.067	764	0.196
R&D Exp.	0.306	0.264	2237	0.334	0.248	762	0.289
HHI	0.083	0.073	2262	0.081	0.072	765	0.576

Panel B: Regression analysis – Creating New Divisions					
	(1)	(2)	(3)	(4)	(5)
OLS	D(Create New Div. *in Ind. of Startup*)[t+1,t+3]				
Num(Non-Stop Flights)	0.00345***	0.00342***	0.00469***	0.00356**	0.00353***
	(2.98)	(2.67)	(2.80)	(2.13)	(2.69)
Start-up's Age	-0.00143	-0.00175	0.000466	0.00221	0.000653
	(-0.63)	(-0.77)	(0.18)	(0.83)	(0.38)
Num. Co-investors	-0.00112	-0.00102	-0.00109	0.00116	0.00000837
	(-1.07)	(-0.96)	(-0.81)	(1.06)	(0.01)
D(Seed or Early Stage)	-0.00176	-0.00372	0.00208	0.0140	0.0105
	(-0.12)	(-0.28)	(0.15)	(0.88)	(0.76)
D(Same MSA Area)	-0.00749	-0.00217	0.00443	0.000515	0.00663
	(-0.66)	(-0.20)	(0.36)	(0.04)	(0.59)
Distance	-0.00314	-0.00341	-0.00435	0.000191	0.00200
	(-0.81)	(-0.86)	(-0.77)	(0.04)	(0.49)
CVC Parent Controls		✓	✓	✓	✓
Year F.E.	✓	✓			
Start-up MSA F.E.	✓	✓			
CVC Firm MSA F.E.	✓	✓	✓		
Start-up MSA × Year F.E.			✓	✓	✓
CVC Firm MSA × Year F.E.				✓	✓
CVC Firm F.E.					✓
Num. Obs.	3275	3212	2923	2764	2705
Adj. R^2	0.074	0.094	0.111	0.272	0.450

Panel C: Regression analysis – Changing Primary Industry					
	(1)	(2)	(3)	(4)	(5)
OLS	D(Change Industry: *Shift to Start-up*)[t+3,t+5]				
Num(Non-Stop Flights)	0.00240**	0.00244**	0.00341**	0.00218*	0.00217*
	(2.10)	(2.10)	(2.12)	(1.72)	(1.66)
Start-up's Age	-0.00162	-0.00177	-0.00146	-0.000933	-0.000663
	(-1.10)	(-1.10)	(-0.77)	(-0.67)	(-0.54)
Num. Co-investors	0.000541	0.000449	0.000476	0.00128*	0.000457
	(0.79)	(0.70)	(0.76)	(1.82)	(0.85)
D(Seed or Early Stage)	0.00105	-0.000726	-0.00419	0.000856	0.000631
	(0.13)	(-0.10)	(-0.81)	(0.11)	(0.10)
D(Same MSA Area)	-0.00697	-0.00680	-0.000688	0.00220	0.000625
	(-0.81)	(-0.90)	(-0.09)	(0.32)	(0.11)
Distance	0.000556	0.000551	0.00215	0.00222	-0.000925
	(0.20)	(0.19)	(0.57)	(0.58)	(-0.33)
CVC Parent Controls		✓	✓	✓	✓
Year F.E.	✓	✓			
Start-up MSA F.E.	✓	✓			
CVC Firm MSA F.E.	✓	✓	✓		
Start-up MSA × Year F.E.			✓	✓	✓
CVC Firm MSA × Year F.E.				✓	✓
CVC Firm F.E.					✓
Num. Obs.	3275	3212	2923	2764	2705
Adj. R^2	0.055	0.061	0.022	0.113	0.377

Table XII: Post CVC Firm Value Creation

This table studies the post-CVC value creation of CVC parents. The dependent variable is the difference of Tobin's Q between Year $t{+}h$ and Year t, where h is shown in the table. All dependent variables are winsorized at 1% and 99% level before being brought into regressions. Industry fixed effects are defined in SIC-2 industries. T-statistics are shown in parentheses, and standard errors are clustered by firm. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Change of Tobin's Q of the CVC Parent							
$\Delta =$	(t+3)-t	(t+4)-t	(t+3)-t	(t+4)-t	(t+3)-t	(t+4)-t	(t+5)-t	(t+6)-t
D(CVC Unrelated)	0.304***	0.307***						
	(3.42)	(3.21)						
D(CVC Related)	-0.186	-0.256	-0.133	-0.155	-0.142	-0.166	-0.201	-0.0983
	(-1.04)	(-1.27)	(-0.67)	(-0.79)	(-0.71)	(-0.84)	(-0.98)	(-0.40)
D(CVC Unrelated) × D(New Div.)[t+1,t+2]			0.363***	0.376***				
			(2.69)	(3.17)				
D(CVC Unrelated) × (1−D(New Div.)[t+1,t+2])			0.0766	-0.00393				
			(0.73)	(-0.03)				
D(CVC Unrelated) × D(Div. Rem.)[t+1,t+2]					0.538**	0.592**		
					(2.11)	(2.01)		
D(CVC Unrelated) × (1−D(Div. Rem.)[t+1,t+2])					0.0857	0.0139		
					(0.81)	(0.12)		
D(CVC Unrelated) × D(Chg. Ind.)[t+3,t+5]							0.299*	0.321*
							(1.81)	(1.70)
D(CVC Unrelated) × (1−D(Chg. Ind.)[t+3,t+5])							0.0271	0.0660
							(0.23)	(0.58)
Firm Controls	Firm Size; ROA; Cash; R&D; Leverage; Capital Exp.; HHI; D(Conglomerate)							
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm F.E.	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Num. Obs.	74,128	65,292	74,128	65,292	74,128	65,292	57,747	51,249
Adj. R^2	0.080	0.075	0.252	0.257	0.252	0.257	0.287	0.291

Appendix A. Variable Definition

Variable Name	Definition and Construction of Variable	Data Source
D(CVC)	Dummy variable equal to 1 if the firm conducts at least one CVC investment in year t	VentureXpert
D(CVC Unrelated)	Dummy variable equal to 1 if the firm conducts at least one unrelated CVC deal in year t. An unrelated CVC deal is defined as a deal with the entrepreneurial company's SIC-3 code not matching with any SIC-3 codes of CVC corporate parent. A conglomerate firm has multiple SIC-3 codes, whereas a stand-alone firm has a single SIC-3 code.	VentureXpert & Compustat Historical Segment
D(CVC Related)	Dummy variable equal to 1 if the firm conducts at least one related CVC deal in year t. A related CVC deal is defined as a deal with the entrepreneurial company's SIC-3 code matching with one of SIC-3 codes of CVC corporate parent.	VentureXpert & Compustat Historical Segment
Num(CVC Deal)	Number of CVC deals conducted by Firm i in year t	VentureXpert
Num(CVC Initial Deal)	Number of CVC initial deals conducted by Firm i in year t. CVC initial deal is defined as the deal in which case the CVC firm invests in an entrepreneurial start-up for the first time, that is, not the follow-on investments.	VentureXpert
Granular IV (Fund Inflow Shock)	Defined as the sum of the (positive) idiosyncratic fund inflow shocks of those IVCs in the past 5-year syndicating network. The idiosyncratic fund inflow shocks are obtained as the error term of the Gompers and Lerner (1998)'s fundraising model, as illustrated in Online Appendix, Section B.	VentureXpert
Num(IVC in the Network)	The natural logarithm of one plus the number of IVCs in the past 5-year syndication network of CVC Firm i.	VentureXpert
D(New Div.)[t+1,t+2]	A dummy equal to 1 if the firm creates at least one new division within the next two years (Year t+1 and Year t+2). Divisions are aggregated and defined in SIC-3 industries. Establishing a new division is identified if the firm reports a new division with its SIC-3 code appearing in the first time in the company history.	Compustat Historical Segment
D(Div. Rem.)[t+1,t+2]	A dummy equal to 1 if the firm removes at least one old division within the next two years. Removing an old division means that a firm stops reporting a division in the future forever.	Compustat Historical Segment
D(Chg. Ind.)[t+3,t+5]	A dummy equal to 1 if, in the next 3 to 5 years, the firm's primary industry has changed.	Compustat Historical Segment
D(Conglomerate)	Dummy equal to 1 if the firm is a conglomerate in year t. A conglomerate is defined as a firm reporting multiple segments in at least 2 different SIC-3 industries in Compustat Historical Segment database.	Compustat Historical Segment
Firm Size	Firm size, measured as natural logarithm of the firm's market capitalization (Compustat item $CSHO_t$ × item $PRCC\_F_t$)	Compustat

Continued on next page

Appendix A continued from previous page

Variable name	Definition and construction of variable	Data Source
Tobin's Q	Market-to-book ratio in assets. Market value of assets equals the book value of assets (item AT_t) + the market value of common equity at fiscal year-end (item $CSHO_t$ $\times$ item $PRCC\_F_t$) $-$ the book value of common equity (item CEQ_t) $-$ balance sheet deferred taxes (item $TXDB_t$)	Compustat
R&D Exp.	R&D expenditure, measured as item XRD_t scaled by lagged book assets (item AT_{t-1}). If the item XRD_t is missing, I replace it with the industry-year median XRD_t.	Compustat
ROA	Return on assets, defined as EBITDA scaled by lagged book assets	Compustat
Leverage	Book leverage, defined as debt including long-term debt (item $DLTT_t$) plus debt in current liabilities (item DLC_t) divided by the sum of debt and book value of common equity (item CEQ_t)	Compustat
Capx.	Capital expenditure, measured as item $CAPX_t$ scaled by lagged book assets	Compustat
HHI	The Hirschman-Herfindahl index of sales (item $SALE_t$) in the industry where the firm is located	Compustat
Cash	Defined as cash and cash equivalents (item CHE_t) scaled by lagged book assets	Compustat

Appendix B. The Construction of Granular IV

As discussed in the paper, I use the idiosyncratic fund inflow shock of IVC firms in the past 5-year network of each CVC firm as the instrument of the CVC investment. To obtain the idiosyncratic fund inflow shock, I follow Gabaix and Koijen (2020)'s Granular IV (GIV) approach.

To illustrate the link between my instrument and the GIV approach, I apply Gabaix and Koijen (2020)'s non-loop model (see Section 2.1.4 Model with an enriched factor structure). The model is as follows. Suppose CVC Firm i's investment decision is influenced by the raw fund inflow of IVCs in its network (this is due to the fact that IVC is usually the largest deal source of CVC firms and IVC frequently invites CVC to join in their new deals (MacMillan et al., 2008)). Then, CVC i's investment amount in year t follows,

$$Num\_CVC_{i,t} = \alpha \bar{S}_{i,t} + \beta_1 X_{i,t} + \varepsilon_{i,t} \tag{B1}$$

where $Num\_CVC_{i,t}$ gauges the number of CVC deals initiated by CVC Firm i in Year t; while $\bar{S}_{i,t}$ is the sum of raw fund inflow of k IVC firms in the past 5-year network of the CVC firm, where k is equal to 3 in the Figure 7's example, the Apple Inc's example. So,

$$\bar{S}_{i,t} = \sum_{j \in Network_i} S_{j,t} \tag{B2}$$

And $S_{j,t}$ is the raw fund inflow of IVC firm j in Year t. Next, the raw fund inflow is a function of IVC's firm characteristics $\bar{X}_{j,t}$ ($\bar{X}$ includes large sets of fixed effects) and time factors λ_t.

$$S_{j,t} = \gamma_{j,t}\lambda_t + \beta_2 \bar{X}_{j,t} + \mu_{j,t} \tag{B3}$$

$\mu_{j,t}$ is assumed to be the idiosyncratic fund inflow shock. The crucial assumption to validate the GIV is then $E(\mu_{j,t}\varepsilon_{i,t}) = 0$ for any i and j. Then the formula of the GIV is,

$$GIV_{i,t} = \sum_{j \in Network_i} \mu_{j,t} \tag{B4}$$

Following Gabaix and Koijen (2020)'s main setting, I consider the parametric factor exposures,

$$\gamma_{j,t} = \gamma_0 + \gamma_1 \tilde{X}_{j,t} \tag{B5}$$

To obtain the $\mu_{j,t}$, I implement an empirical fundraising model from Gompers and Lerner (1998). Moreover, I proxy the fund inflow ($S_{j,t}$) of IVC firms with the dummy of raising a new follow-on fund. This proxy has two practical reasons: (1) new fundraising is always accompanied by the largest fund inflow; (2) when the IVC launches a new fund, it is most likely that the IVC conducts new deals and invites CVCs. The distribution of those new follow-on fundraising is plotted in Figure B.2.

In Gompers and Lerner (1998)'s two-stage Heckman selection model, the time factors λ_t include the Number of startups brought public last year by all VCs, T-bill return (10-year), Real GDP growth, and CRSP value-weighted return. $\tilde{X}_{j,t}$ contains Years since raising last fund, the square of Years since raising last fund, Age of the venture organization (years), Number of startups brought public this year, Number of startups brought public last year, and finally the Number of funds launched before.

For simplicity, I assume that $\tilde{X}_{j,t} = \bar{X}_{j,t}$. In other words, the interaction terms between each IVC firm characteristics and time factors are included in equation (3). I estimate the equation (3) with OLS, adding VC industry specialization (VEIC) by year fixed effect and the location (State) by year fixed effect. Following Hochberg et al. (2007), I take a VC firm's industry specialization to be the broad Venture Economics industry group (VEIC) that accounts for most of its invested capital. The error term from the above regression is thus the $\mu_{j,t}$. Finally, I only take the positive idiosyncratic fund inflow shock (in my case, the negative shocks and positive shocks cannot cancel out since what matters finally is how many IVC receives the positive inflow shocks),

$$\widehat{GIV}_{i,t} = \sum_{j \in Network_i} \max\{\hat{\mu}_{j,t}, 0\} \tag{B6}$$

This follows the threshold GIV as discussed in Section 2.5 of Gabaix and Koijen (2020). Table B.1 reports the OLS estimate of equation (3), where I use the error term of the Column (3) to construct my GIV.

Table B.1: Gompers and Lerner (1998)'s fundraising model

OLS	(1)	(2)	(3)
		D(Launch New Fund)	
Individual IVC characteristics			
Years since raising last fund	-0.00828***	0.00402***	0.00463***
	(-10.87)	(2.67)	(3.00)
(Years since raising last fund)^2	0.0000827**	-0.0000563*	-0.0000808**
	(2.56)	(-1.79)	(-2.46)
Age of the venture organization	0.00264***	-0.00286**	-0.00294**
	(10.47)	(-2.50)	(-2.51)
Number of startups brought public this year	0.0304***	-0.0135**	-0.0117**
	(19.88)	(-2.47)	(-2.09)
Number of startups brought public last year	0.0240***	-0.00163	-0.00164
	(15.35)	(-1.00)	(-0.98)
Number of past funds launched		0.00679**	0.00685**
		(2.09)	(2.07)
VC funding factors			
Number of startups brought public last year by all VCs	0.000279***		
	(12.06)		
T-bill return	0.0259		
	(1.40)		
Real GDP Growth	0.00480***		
	(5.05)		
CRSP value weighted return	0.112		
	(0.95)		
Interaction terms λ_2			
Years since raising last fund*		0.0000424***	0.0000445***
Number of startups brought public last year by all VCs		(4.99)	(5.13)

Years since raising last fund* T-bill return		0.0710*** (8.90)	0.0664*** (8.08)
Years since raising last fund* Real GDP Growth		0.00181*** (5.03)	0.00175*** (4.75)
Years since raising last fund* CRSP value weighted return		0.200*** (4.38)	0.190*** (4.06)
Age of the venture organization* Number of startups brought public last year by all VCs		-0.0000605*** (-8.17)	-0.0000617*** (-8.19)
Age of the venture organization* T-bill return		-0.0685*** (-9.30)	-0.0643*** (-8.51)
Age of the venture organization* Real GDP Growth		-0.00203*** (-6.37)	-0.00196*** (-6.02)
Age of the venture organization* CRSP value weighted return		-0.214*** (-5.18)	-0.209*** (-4.92)
Number of startups brought public this year* Number of startups brought public last year by all VCs		-0.0000253 (-1.12)	-0.0000285 (-1.22)
Number of startups brought public this year* T-bill return		-0.0290* (-1.73)	-0.0345** (-2.00)
Number of startups brought public this year* Real GDP Growth		0.00789*** (6.33)	0.00776*** (6.03)
Number of startups brought public this year* CRSP value weighted return		0.235 (1.55)	0.204 (1.30)
Number of past funds launched* Number of startups brought public last year by all VCs		0.000355*** (16.69)	0.000358*** (16.55)
Number of past funds launched* T-bill return		0.228*** (11.57)	0.215*** (10.65)
Number of past funds launched* Real GDP Growth		0.00642*** (7.29)	0.00620*** (6.91)
Number of past funds launched* CRSP value weighted return		0.934*** (8.25)	0.918*** (7.93)
VEIC × Year F.E.		Yes	Yes
Location × Year F.E.			Yes
Num. Obs.	33,163	33,163	33,163
Adj. R^2	0.076	0.205	0.203

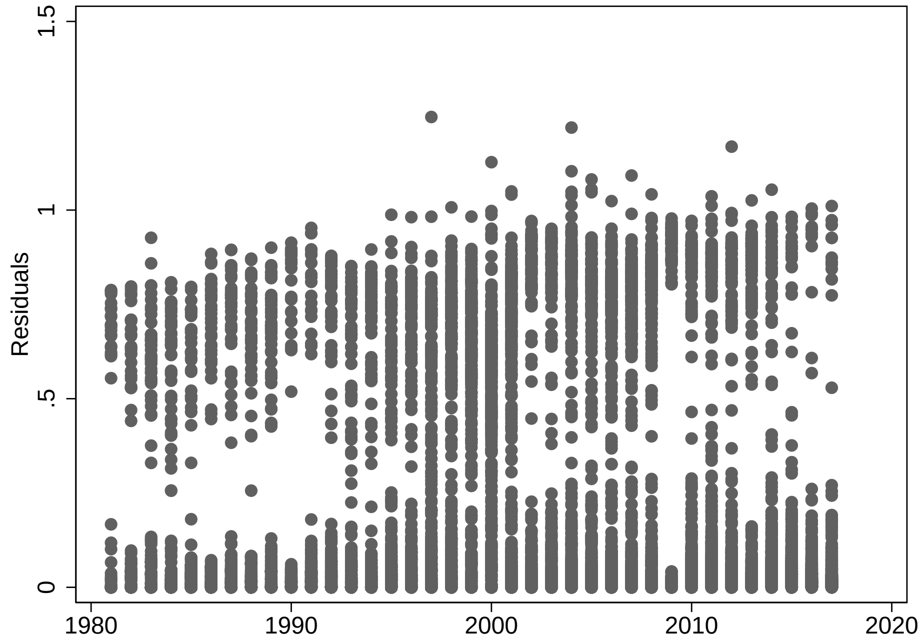


Figure B.1: The distribution of the residues by years

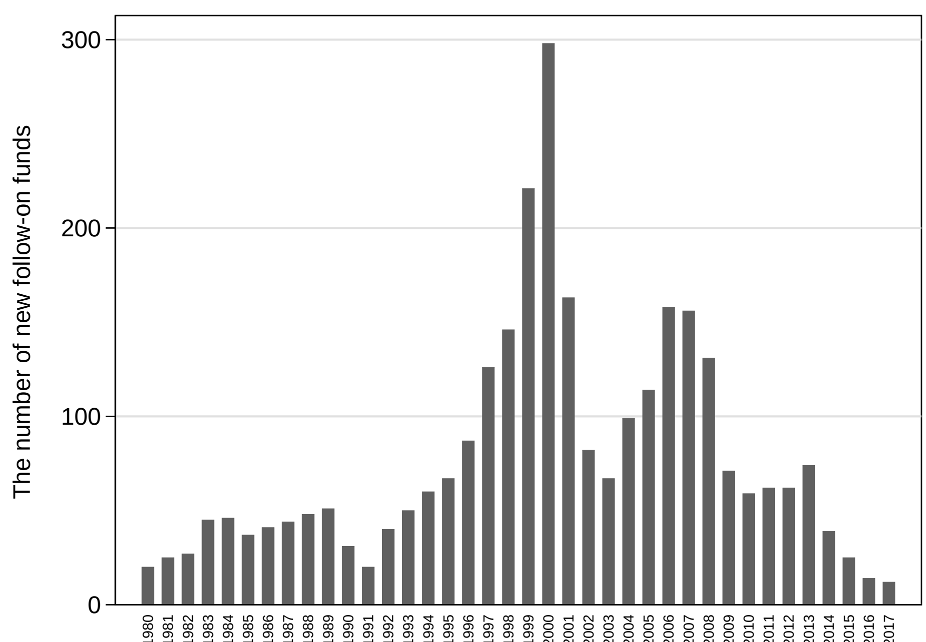


Figure B.2: The distribution of new launching of IVC follow-on funds

Appendix C. Technique Details of Textual Analysis

This appendix section contains technique details in my textual analysis. It includes downloading textual data, forming word pairs, searching emerging phrases, and stopping words and common words used in the textual cleaning procedures.

C.1 Textual Data Downloading

The startups' business descriptions are from VentureXpert. First, I download all VC-backed startups' business descriptions and save them into an excel file. In the above procedure, I only choose to save those VC-backed startups which at least receive once the funding from a CVC program (CVC programs from foreign companies and private firms are allowed). It facilitates in picking technology-focused VC-backed startups, and their business should be more attractive to CVC firms as well as their industry peers. The business description sample is at the startup by year level, which means, if the Startup j receives funding in Year 2008 and 2010, its business will have two unique observations in excel: one in 2008 and one in 2010. Next, I group the startups' business text into the yearly corpus, i.e., in each year, there is a text file containing the business text of startups receiving funding in that year.

Regarding the 10-K business descriptions, I download them using Python. The original code is written by Tzu-Hsiang Lin at Amazon and is revised by myself. The major difference is that (1) my code aims to download, parse, and extract the Item 1 business description, not the Item 7; (2) I download 10-K forms, including 10-K, 10-K405, 10KSB, 10KSB40 forms; (3) I download the 10-Ks if the CIK is listed in the Compustat Historical Segment Database; (4) being able to detect the double indices in the 10-K fillings.

C.2 Forming Word Pairs

I form word pairs for two adjacent words in the same sentence in each yearly corpus txt file. As an example, the 2016 yearly corpus contains the following short sentence:

> Domo, Inc. is a provider of cloud-based platform for business optimization.

After tokenizing the sentence and lemmatizing each token, I obtain several word pairs as "domo provider", "provider cloud", "cloud based", "based platform", "platform business", and "business optimization". The "Inc", "is", "for", and "a" are stop words dropped before generating word pairs. The same procedure is repeated for each sentence in the 2016 yearly corpus text. Next, I select the top 5% most popular word pairs in each yearly corpus as the "emerging phrases" in that year.

C.3 Stop Words

Apart from the built-in stop words in NLTK, I read more than 100 startups business descriptions and manually identify stop words as follows:

stop_words2 = ['provides', 'manufactures', 'distributes', 'makes', 'offers', 'engages', 'establishes', 'produces', 'conducts', 'operates', 'supplies', 'owns', 'markets', 'designs', 'specializes', 'sells', 'maintains', 'publishes', 'focuses', 'develops', 'delivers', 'provide', 'manufacture', 'distribute', 'make', 'offer', 'engage', 'establish', 'produce', 'conduct', 'operate', 'supply', 'own', 'market', 'design', 'specialize', 'sell', 'maintain', 'publish', 'focus', 'develop', 'development' 'providing', 'manufacturing', 'distributing', 'making', 'offering', 'engaging', 'establishing', 'producing', 'conducting', 'operating', 'supplying', 'owning', 'marketing', 'designing', 'specializing', 'selling', 'maintaining', 'publishing', 'focusing', 'developing', 'focused', 'formed', 'related', 'united', 'state', 'ny', 'ca', 'ma', 'fund', 'firm', 'north', 'america', 'england', 'seattle', 'startup', 'mnfrs', 'dvlps', 'mfrs', 'manages', 'inc', 'corporation', 'corp', 'llc', 'company', 'holding', 'using', 'manufacturer']

C.4 Dropping Too Generic Emerging Phrases

Since some emerging phrases are too generic, i.e. adding those words into the 10-K doesn't mean that the firm is integrating some new businesses, I drop the following common word pairs from the emerging phrases set,

stop_words_2gram = ['venture capital', 'small business', 'web site', 'product service', 'private equity', 'next generation', 'service provider', 'public private', 'capital private', 'science technology', 'commercial product', 'service via', 'medical non', 'financial service', 'service based', 'privately held', 'customer relation', 'customer relationship', 'management solution', 'business service', 'service solution', 'solution business', 'product based', 'solution service', 'business solution', 'service management', 'system service', 'management service', 'product designed', 'product use', 'service business', 'analysis solution', 'analytics solution', 'commerce business', 'commerce service', 'engaged building', 'engaged creating', 'engaged information', 'managed service', 'new used', 'intellectual property', 'product technology', 'service commercial', 'service featuring', 'solution commercial', 'solution enable', 'service industry', 'solution product', 'solution provider', 'world wide', 'engaged providing', 'venture backed']