Internet Appendix for
Corporate Venture Capital and Firm Scope

A Technique Details about Textual Analysis

This appendix section contains technique details in my textual analysis. It involves collecting textual data, creating word pairings (bigram phrases), downloading 10-K data, and identifying common terms and words that are employed in textual cleaning methods.

A.1 Textual Data Downloading

VentureXpert provides business summaries for startups. I start by downloading the business summary for each VC-backed entrepreneurial company and saving all business descriptions in a single excel file. My sample of business description only contains those startups at least receive one round funding from a CVC investor, and the CVC investor can be a foreign investor or a firm not listed on any stock markets. 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, i.e., if startup j receives VC funding (in two different rounds) 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 a yearly corpus, that is, in each year, there is a text file containing the business text of startups receiving VC 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; and (4) my code is able to detect the double indices in the 10-K fillings.

A.2 Forming Word Pairs

I form word pairs for two consecutive 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.

A.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']

A.4 Dropping Too Generic Emerging Phrases

I remove the following common word pairs from the emerging phrases set because they are too generic, i.e., adding such words to the 10-K will not really suggest that the firm is integrating some new businesses.

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']

B The Alternative Identification

In this alternative identification, I exploit 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, 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 a 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, Ljungqvist, and Lu, 2007). Second, the IVC is the most crucial channel of deal sourcing for CVC firms, as documented in Sykes (1990) and MacMillan, Roberts, Livada, and Wang (2008), among others.

Based on the two premises, an idea of the instrument is that, if an independent VC firm j (IVC j) receives a positive fund inflow shock today, and meanwhile, a 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 an example illustrated in Figure A3, 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.[1]

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-invested with Firm i within the past five years. The co-investment (syndication) is defined as a 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 **??** 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 oil productions, 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 A5 reports the first stage regression where I use the Granular IV (sum of the

[1] 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).

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 A5 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).[2]

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, Nanda, and Rhodes-Kropf (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.[3]

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 A6. 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

[2]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.

[3]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.

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.[4] 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.

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.

C Additional Figures and Tables

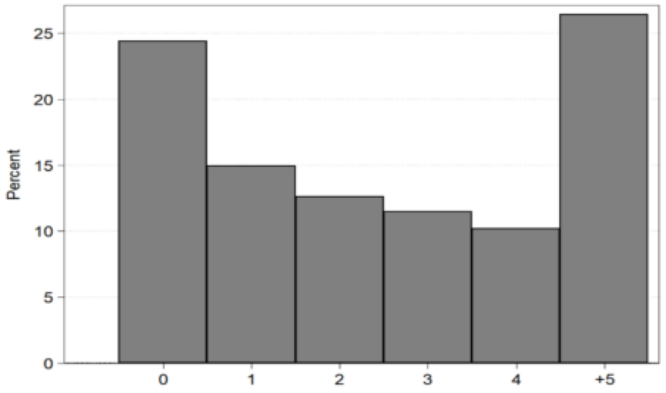
[4] In unreported results, I find that the segment dummies are all strongly and positively correlated with the new textual measure.

Figure A1. "Emerging Phrases" and Emerging Business Integration

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

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

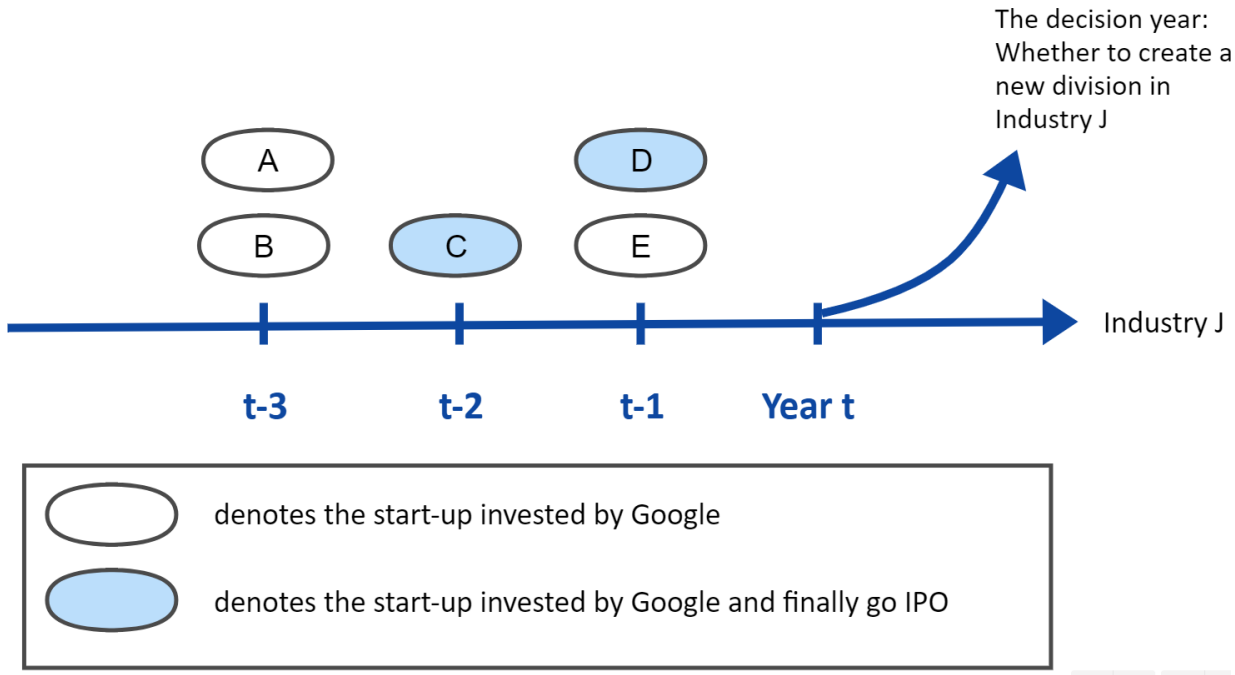
(c) Top 50 "emerging phrases" mostly integrated by CVC parents



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

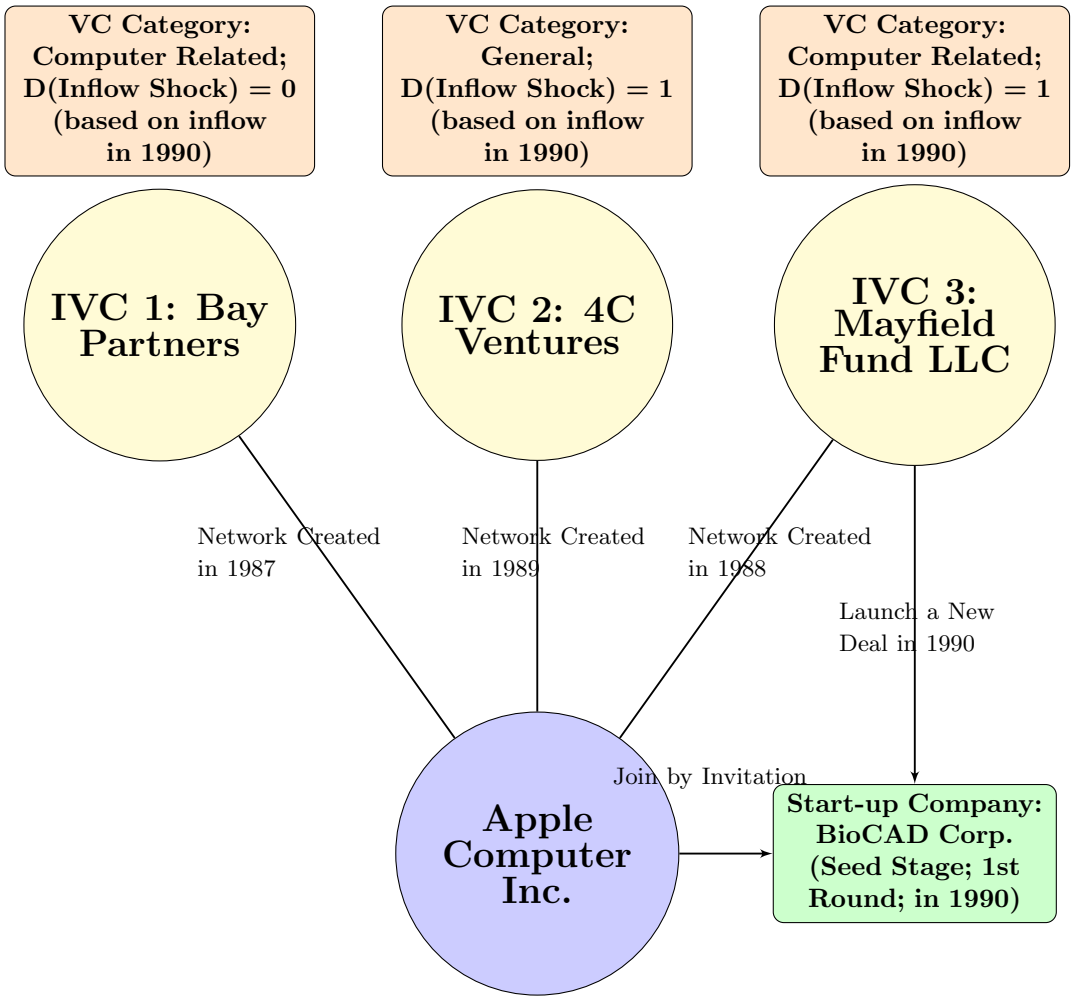
Panel A and B present the words clouds of "emerging phrases" in 2000 and 2017. Emerging phrases are the top 5% most popular short phrases (excluding stopwords and common words) in the detailed business descriptions of all VC-backed startups receiving VC fundings in a given year. Panel C plots the top 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 A2. An Example of Constructing CVC Investment Signals



This figure provides a simple example about the construction of CVC signal variables. In this figure, the CVC signal is measured by startup's IPO. The detailed procedures are explained in the text.

Figure A3. 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.

Table A1: 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)
Period	D(Create New Division) [t+1, t+2]			D(Remove Old Division) [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 A2: CVC Investments and Corporate Restructuring: Sample Before 1997

This table provides the robustness check of Table A1 by using the sample before 1997 in the regressions. 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 approach. The regression design and sample construction follows Table A1.

Panel A: Creating new divisions	(1) Logit	(2) Logit	(3) CLogit	(4) CLogit
	D(Create New Division)[t+1,t+2]			
D(CVC)	0.290*		0.280	
	(1.77)		(1.54)	
D(CVC Unrelated)		0.460***		0.443**
		(2.70)		(2.38)
D(CVC Related)		-0.341		-0.280
		(-0.91)		(-0.72)
D(New Div.)[t-2,t-1]	0.249***	0.248***	0.278***	0.277***
	(4.41)	(4.38)	(4.54)	(4.52)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Year Fixed Effect	Yes	Yes	No	No
Industry Fixed Effect	Yes	Yes	No	No
Industry*Year Fixed Effect	No	No	Yes	Yes
Num. Obs.	42,340	42,340	41,078	41,078
Pseudo R^2	0.098	0.098	0.043	0.043

Panel B: Removing old divisions	(1) Logit	(2) Logit	(3) CLogit	(4) CLogit
	D(Remove Old Division)[t+1,t+2]			
D(CVC)	0.353**		0.332*	
	(2.10)		(1.80)	
D(CVC Unrelated)		0.507***		0.473**
		(2.61)		(2.26)
D(CVC Related)		-0.403		-0.338
		(-1.05)		(-0.91)
D(Div. Rem.)[t-2,t-1]	0.249***	0.248***	0.281***	0.280***
	(4.90)	(4.87)	(5.12)	(5.11)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Year Fixed Effect	Yes	Yes	No	No
Industry Fixed Effect	Yes	Yes	No	No
Industry*Year Fixed Effect	No	No	Yes	Yes
Num. Obs.	42,340	42,340	41,540	41,540
Pseudo R^2	0.200	0.200	0.188	0.188

Panel C: Changing the primary business (industry)

	(1) CLogit	(2) CLogit	(3) CLogit	(4) CLogit
	D(Chg.Ind.)[t+3,t+5]		D(Chg.Ind.)[t+4,t+6]	
D(CVC)	0.194		0.212	
	(0.79)		(0.82)	
D(CVC Unrelated)		0.373*		0.399*
		(1.94)		(1.96)
D(CVC Related)		-0.214		-0.265
		(-1.10)		(-0.99)
D(Chg.Ind.)[t-2,t-1]	0.842***	0.843***	0.824***	0.825***
	(10.68)	(10.69)	(9.94)	(9.96)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Industry*Year Fixed Effect	Yes	Yes	Yes	Yes
Num. Obs.	32,888	32,888	30,527	30,527
Pseudo R^2	0.082	0.083	0.080	0.080

Table A3: CVC Investments and Corporate Restructuring: Post-1997 Sample

This table provides the robustness check of Table A1 by using the post-1997 sample in the regressions. 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 approach. The regression design and sample construction follows Table A1.

Panel A: Creating new divisions

	(1) Logit	(2) Logit	(3) CLogit	(4) CLogit
	D(Create New Division)[t+1,t+2]			
D(CVC)	0.432***		0.430**	
	(2.88)		(2.46)	
D(CVC Unrelated)		0.620***		0.629***
		(3.48)		(3.17)
D(CVC Related)		-0.249		-0.296
		(-0.90)		(-1.01)
D(New Div.)[t-2,t-1]	0.126*	0.126*	0.0990	0.0995
	(1.72)	(1.73)	(1.20)	(1.21)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Year Fixed Effect	Yes	Yes	No	No
Industry Fixed Effect	Yes	Yes	No	No
Industry*Year Fixed Effect	No	No	Yes	Yes
Num. Obs.	43,903	43,903	41,658	41,658
Pseudo R^2	0.058	0.058	0.017	0.018

Panel B: Removing old divisions

	(1) Logit	(2) Logit	(3) CLogit	(4) CLogit
	D(Remove Old Division)[t+1,t+2]			
D(CVC)	0.347***		0.324**	
	(2.70)		(2.10)	
D(CVC Unrelated)		0.521***		0.494***
		(3.12)		(2.71)
D(CVC Related)		-0.159		-0.161
		(-0.66)		(-0.66)
D(Div. Rem.)[t-2,t-1]	0.203***	0.205***	0.197***	0.198***
	(3.15)	(3.19)	(2.90)	(2.93)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Year Fixed Effect	Yes	Yes	No	No
Industry Fixed Effect	Yes	Yes	No	No
Industry*Year Fixed Effect	No	No	Yes	Yes
Num. Obs.	43,903	43,903	42,407	42,407
Pseudo R^2	0.187	0.188	0.165	0.165

Panel C: Changing the primary business (industry)

	(1) CLogit	(2) CLogit	(3) CLogit	(4) CLogit
	D(Chg.Ind.)[t+3,t+5]		D(Chg.Ind.)[t+4,t+6]	
D(CVC)	0.547**		0.483*	
	(2.13)		(1.82)	
D(CVC Unrelated)		0.541*		0.546**
		(1.96)		(1.98)
D(CVC Related)		-0.143		-0.231
		(-0.40)		(-0.61)
D(Chg.Ind.)[t-2,t-1]	1.076***	1.077***	1.074***	1.075***
	(12.13)	(12.14)	(11.38)	(11.40)
Firm Controls:	Firm Size, Tobin's Q, ROA, R&D, Leverage, Capx., HHI, D(Conglomerate)			
Industry*Year Fixed Effect	Yes	Yes	Yes	Yes
Num. Obs.	27,303	27,303	23,818	23,818
Pseudo R^2	0.106	0.106	0.097	0.097

Table A4. Introduction of New (Non-Stop) Airline Routes and Emerging Phrases Adding (with Placebo Test)

This table provides analyses of introductions of new (non-stop) airlines and changes in firm scope for CVC parent firms. The sample is at the CVC deal level and includes all deals between 1995 and 2017 with non-missing location information for both the CVC firm's and the invested start-up's headquarter in a deal. In Panel A, the dependent variable is the number of emerging phrases newly added into the CVC firm's 10-K in year $t+2$ (two years after the CVC deal year). The emerging phrases are required to be previously used by the start-up in the deal in year t (deal year). `I(Intro. New Airline) [t+1]` is equal to 1 if there is a new non-stop airline added between the location of the start-up and CVC firm in the deal. New airlines are identified at the airline company level. `Num. Existing Airline [t] (deal year)` and `Num. Passengers [t] (deal year)` are all measured in the deal year (year t). Figure **??** offers a detailed illustration of timing. In Panel B, `I(Intro. New Airline) [t+1]` is further interacted with a distance measure between the locations of the CVC firm and start-up. Panel C and D provide two placebo tests. In Panel C, the dependent variable captures new emerging phrases added in year t instead of the year $t+2$. In Panel D, introductions of new airways are measured in year $t+3$ instead of $t+1$. T-statistics are shown in parentheses, and standard errors are double clustered by firm and start-up levels. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels respectively.

Panel A: Main Regressions

Dependent Var.	(1)	(2)	(3)	(4)	(5)
Measured in	Dummy of Adding Emerging Phrases in 10-K $t+2$				
I(Intro. New Airline) [t+1]	0.00883**	0.00903**	0.00989***	0.0101***	0.00931**
	(2.57)	(2.58)	(2.65)	(2.66)	(2.30)
Num. Existing Airline [t] (deal year)	-0.000307	-0.0000237	-0.000283	0.00000776	-0.00117*
	(-0.45)	(-0.03)	(-0.41)	(0.01)	(-1.66)
Num. Passengers [t] (deal year)	0.0104	0.00982	0.00980	0.00914	0.00840*
	(1.30)	(1.19)	(1.16)	(1.05)	(1.77)
Start-up's Age	-0.000997	-0.00165	-0.00108	-0.00174	-0.000966*
	(-0.50)	(-0.81)	(-0.52)	(-0.83)	(-1.82)
I(Seed or Early Stage)	-0.00315	-0.00347	-0.00331	-0.00366	-0.000455
	(-0.47)	(-0.50)	(-0.46)	(-0.49)	(-0.09)
Num. Co-investors	0.000786	0.000812	0.000807	0.000831	0.000795
	(0.99)	(0.98)	(0.98)	(0.97)	(1.36)
Round Amount (1000 USD)	-1.28e-08	-1.07e-08	-1.27e-08	-1.07e-08	-2.09e-08
	(-0.61)	(-0.55)	(-0.60)	(-0.53)	(-1.34)
CVC Patent Firm F.E.	Y	Y	Y	Y	Y
Year F.E.	Y	Y	Y	Y	
Start-up F.E.	Y	Y	Y	Y	
CVC Location F.E.		Y		Y	
Startup Location F.E.		Y		Y	
Start-up Industry F.E.			Y	Y	Y
CVC Location × Year F.E.					Y
Start-up Location × Year F.E.					Y
Num. Obs.	4,263	4,263	4,091	4,091	5,182

Panel B: Interact with the Distance

	(1)	(2)	(3)	(4)	(5)
Dependent Var.	Dummy of Adding Emerging Phrases in 10-K				
Measured in	t+2				
I(Intro. New Airline) [t+1]	0.0151*	0.0154*	0.0159*	0.0172**	0.0159**
× Long Distance	(1.70)	(1.72)	(1.76)	(1.98)	(2.22)
I(Intro. New Airline) [t+1]	0.00381	0.00384	0.00438	0.00441	0.00533
× Short Distance	(1.07)	(1.06)	(1.08)	(1.06)	(1.20)
CVC Patent Firm F.E.	Y	Y	Y	Y	Y
Year F.E.	Y	Y	Y	Y	
Start-up F.E.	Y	Y	Y	Y	
CVC Location F.E.		Y		Y	
Startup Location F.E.		Y		Y	
Start-up Industry F.E.			Y	Y	Y
CVC Location × Year F.E.					Y
Start-up Location × Year F.E.					Y
Num. Obs.	4,263	4,263	4,091	4,091	5,182

Panel C: Placebo Test 1

	(1)	(2)	(3)	(4)	(5)
Dependent Var.	Dummy of Adding Emerging Phrases in 10-K				
Measured in	t Instead of Originally t+2				
I(Intro. New Airline) [t+1]	-0.00807	-0.00780	-0.00728	-0.00697	-0.00686
	(-1.02)	(-0.98)	(-0.90)	(-0.86)	(-0.78)
CVC Patent Firm F.E.	Y	Y	Y	Y	Y
Year F.E.	Y	Y	Y	Y	
Start-up F.E.	Y	Y	Y	Y	
CVC Location F.E.		Y		Y	
Startup Location F.E.		Y		Y	
Start-up Industry F.E.			Y	Y	Y
CVC Location × Year F.E.					Y
Start-up Location × Year F.E.					Y
Num. Obs.	4,263	4,263	4,091	4,091	5,182

Panel D: Placebo Test 2

	(1)	(2)	(3)	(4)	(5)
Dependent Var.	Dummy of Adding Emerging Phrases in 10-K				
Measured in	t+2				
I(Intro. New Airline) [t+3]	-0.00563*	-0.00572	-0.00591	-0.00600	-0.00541
	(-1.66)	(-1.65)	(-1.64)	(-1.62)	(-1.42)
CVC Patent Firm F.E.	Y	Y	Y	Y	Y
Year F.E.	Y	Y	Y	Y	
Start-up F.E.	Y	Y	Y	Y	
CVC Location F.E.		Y		Y	
Startup Location F.E.		Y		Y	
Start-up Industry F.E.			Y	Y	Y
CVC Location × Year F.E.					Y
Start-up Location × Year F.E.					Y
Num. Obs.	4,263	4,263	4,091	4,091	5,182

Table A5: 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 A3 provides an example about how the instrument works. The regression sample follows Table A3 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 A6: 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	
Time Period	Emerging Phrases [t+1]	Business Changes [t+1]	New Products [t+1]	New Division [t+1, t+2]	Remove Divisions [t+1, t+2]	Change Industry [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

References

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.

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

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

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

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.

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).

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

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