Corporate Venture Capital

An Empirical Investigation of Antecedents to the Choice between Internal and External Corporate Venture Capital Units

Date: 01. August 2018

Authors: Laura Fischer and Jesper Brodersen

Student N°: 106066 and 68109

Supervisor: Francesco Di Lorenzo

Master Thesis

M.Sc. Finance and Strategic Management

Copenhagen Business School

Number of characters | standard pages | physical pages: 241,306 | 106 | 114

--- This page intentionally left blank ---

ABSTRACT

This thesis provides a *first step* towards an understanding of the antecedents to the choice of setting up corporate venture capital (CVC) units as internal or external units. Drawing on existing literature on CVC and dominant theories of strategic management, we propose theoretical links between the following organizational dimensions of parent companies and the organizational structure of CVC units: (i) value of innovations, (ii) firm specificity of innovations and (iii) technological diversification. Our findings confirm the theorized links. Specifically, value of innovations is significantly negatively related, whereas firm specificity and technological diversification are significantly positively related to the likelihood of setting up an external rather than an internal CVC unit. We find no evidence of differences between industries, and technological diversification does not significantly moderate the relationship between the choice of an internal or external CVC unit and the value of innovations and firm specificity, respectively.

The empirical analysis performed in this thesis is based on a sample of data on internal and external CVC units' activities in the years 1985 to 2015 and the patenting activities of their parent organizations in the years 1976 to 2017. The data, which was retrieved from Compustat, Thomson One Banker and PatentsView, has been manually enriched through several rounds of clerical review. The final sample on which the analysis is based comprises 161 US-based CVC units in the pharmaceutical, semiconductor and IT software industries.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. INTRODUCTION

1.1. Motivation

The phenomenon Corporate Venture Capital (CVC) is growing. In the first quarter of 2018, CVC participated in 28% of all deals to VC-backed companies globally (PwC & CB Insights, 2018). CVC deals are on average, compared to traditional venture capital (VC), bigger, and in 2015, annual global CVC financing amounted to USD 28.6 billion (CB Insights, 2017a).

There is no doubt that the sheer magnitude of the CVC activities makes it an interesting field for researchers, and the field has indeed attracted a rising number of research articles that delve into the motives behind CVC (see e.g. Rind, 1981), characteristics of CVC investments (see e.g. Dushnitsky & Lenox, 2006), and many other essential aspects (for a comprehensive overview of relevant literature, see Dushnitsky, 2008).

Evidently, the CVC field in general is an interesting field. However, the specific motivation behind this thesis can best be conveyed via an illustrative example: In the period 1985 to 2015, the Microsoft Corporation made more than one hundred investments through its *internal* CVC unit. In the same period, Intel, the most prolific CVC investor in the world, made more than 1,500 investments through its *external*[1] CVC unit, Intel Capital Corp. Obviously, Microsoft and Intel have chosen different ways of organizing their CVC activities. Hence, a natural question arises: *what can explain why* firms have different organizational setups for their CVC activities? Attempting to find the answer to this simple question is the main motivation behind this thesis.

Some research has been done within the field of *organizational structure* of CVC (in some cases corporate venturing[2]) units. To name a few; Rind's (1981) work on investment mediation, Sykes' (1986) work on decision-making autonomy, Hill and Birkinshaw's (2008) work on *organizational*

[1] The specific definition applied in this thesis for *internal* and *external*, respectively, will be defined in a later section.
[2] The nuance is essential, and will later be dealt with more in detail.

profiles and many more (including, but not limited to Lee, Park, & Kang, 2018; McNally, 1997; Miles & Covin, 2002; Sykes, 1986; Winters & Murfin, 1988).[3]

What most past research in the area of organizational structure of CVC units and activities has in common is that it is focused on the *performance* of different structural configurations (defined and measured in a range of different ways), with the exception of Souitaris, Zerbinati, and Liu (2012).

Consequently, research has offered little insight into the nature of the organizational structure of CVC units that is not performance-oriented or anecdotal. To the best of our knowledge, no research has directly investigated the *antecedents* to the choice of setting up CVC units internally or externally. This apparent research gap forms the basis of this thesis. Since prior research has neglected to identify antecedents to the setup of CVC units as internal or external units, this thesis aims at *opening the debate* with regards to antecedents, through a rigorous and stylized empirical analysis.

1.2. Research Question

The purpose of this thesis is to examine the *antecedents* to the choice between internal and external CVC units. Specifically, we wish to investigate which characteristics or dimensions of a company that engages in CVC activities, can be effective predictors of whether the company invests via an independent external unit or via an internal unit. To operationalize this aim, we explore the following research question:

What are the antecedents to the choice of setting up corporate venture capital units internally or externally?

We choose to focus on the organizational dimensions of the parent companies to investigate *why* companies set up their CVC units differently. Hence, this thesis contributes to the existing literature on CVC by bridging a, quite narrow, research gap regarding antecedents of the setup of CVC units.

In this thesis, two different set-ups will be distinguished, namely an *internal* and an *external* CVC unit, which are defined by the legal structure of the unit. Hereby, drawing on Dushnitsky and

[3] A comprehensive review of this and more research on organizational structure will be set out in the literature review.

Shaver (2009), an external unit is defined as a separate legal entity, wholly owned by the parent organization, whereas an internal unit is not legally separated from its parent.

1.3. Delimitation

In the following section, the delimitation of the thesis will be addressed. While there are many interesting fields of research within CVC, this thesis focuses solely on the above research question.

This thesis aims at investigating antecedents to the setup of CVC units as internal or external units through an empirical investigation that is apt at identifying tendencies across a larger sample of companies, i.e. investigating the "broader picture". This thesis does not aim at developing a *fine-grain nuanced* image of why companies set up internal and external units, respectively, in specific cases. This was decided due to the size and scope of this thesis.

Moreover, as the link between innovation and organization is well-established in the literature (including in the CVC literature, see e.g. Lee et al., 2018), and since this thesis is a first step towards an in-depth understanding of the antecedents to the choice of setup, this thesis will mainly (though not exclusively) focus on innovation-related predictors. Moreover, this thesis will focus on the knowledge-related organizational dimensions of the *parent* companies of the CVC units as antecedents.

Our raw data extract, which was retrieved from the Thomson One Banker Database (Thomson Reuters, 2018), consists of 1,433 distinct corporate investors from 49 nations.[4] These investors invested in the period between 1985 and 2015. Therefore, our analysis is limited to CVC investments performed within this time frame. Similarly, the patent data, which is used for investigating innovation-related variables, is available for the years 1976 to 2017. Innovative activities that fall outside this period are therefore not included in the analysis. Moreover, the CVC field is dynamic and changes over time. This thesis does not address the risk that any conclusions inferred from analysing data from the years 1985 to 2015 might be inaccurate in the future.

While we cover non-US CVC investors in the descriptive statistics part of the thesis, the innovation-related parts and the empirical analysis will only cover US investors. The argumentation

[4] This list will be subject to a thorough clerical review, as explained in the methodology section.

for this decision can be found in the descriptive statistics section (6.1). As such, empirical findings do not extend to firms outside the US.

The same is the case for industries. While our initial descriptive statistics cover all industries, the main analysis will only cover three industries; IT software, semiconductors and pharmaceuticals. As such, empirical findings of the main analysis do not necessarily extend to other industries.

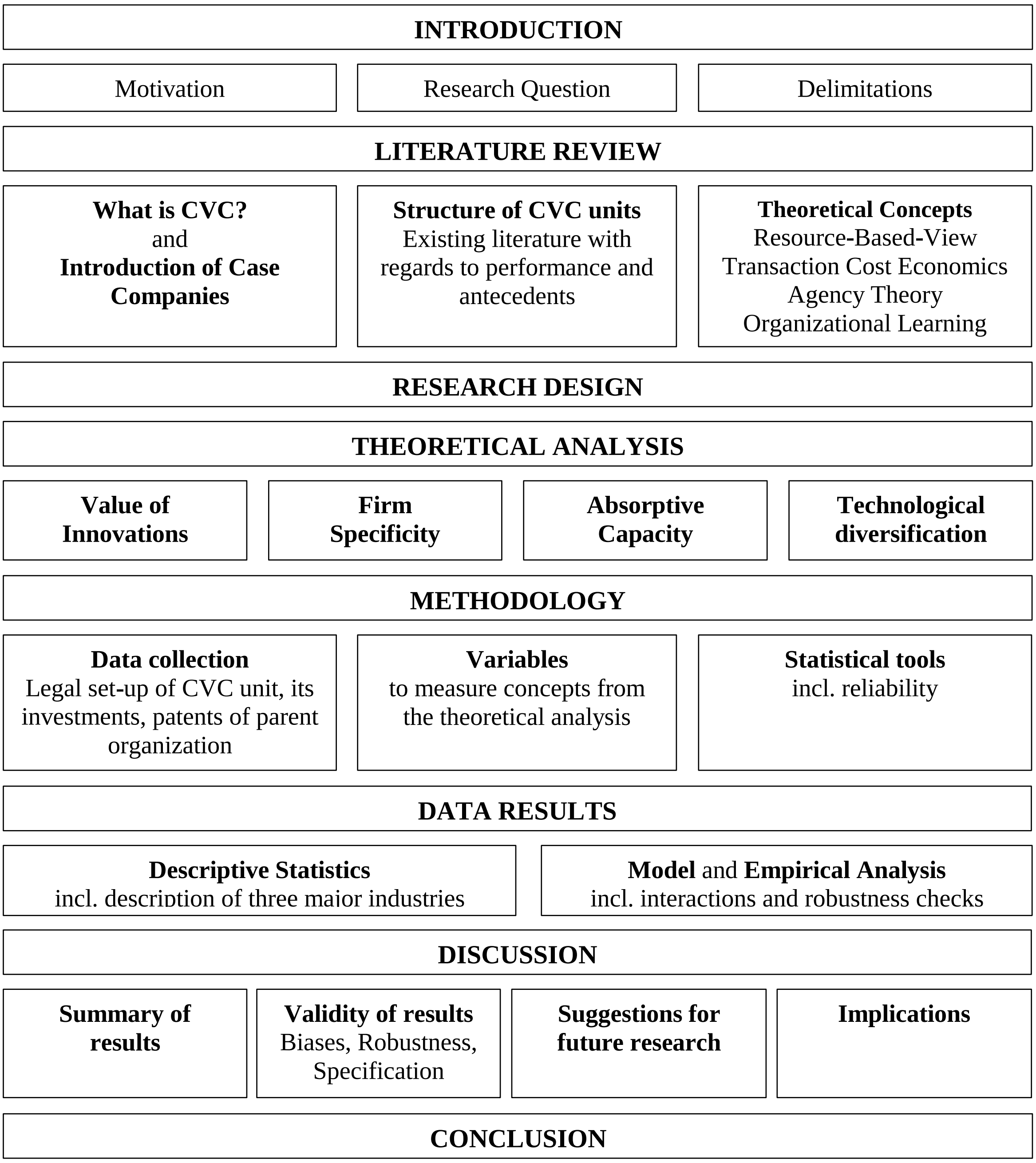
1.4. Structure of the thesis

The structure of the thesis is illustrated in *Figure 1*. After the introduction, a literature review will be performed. The literature review will revolve around four main parts: (i) a broad and brief introduction to the CVC field, (ii) a brief introduction to three case companies that exclusively will serve illustrative purposes, (iii) an in-depth review of papers related to *organizational structure* (due to the lack of research within the narrow scope of this thesis, the literature review will focus on prior research within *organizational structure*, in a broad sense and defined in numerous ways, within the CVC field), and (iv) a brief review of the main theories applied in this thesis. Following the literature review, we will introduce the research design applied, which will guide the analytical section of the paper.

Subsequently, a theoretical analysis that will serve as the background for developing the model will be performed. It will consist of four theoretical concepts, that will be operationalized in the analysis: (i) value of innovations, (ii) firm specificity, (iii) absorptive capacity (iv) technological diversification.

Thereafter, the methodology applied in the empirical analysis of this thesis will be described. The method of data collection, applying variables as proxies for the theorized concepts and statistical tools will be discussed in this section. After the methodology section, the results section will follow. Firstly, the data will be described in the *descriptive statistics* section. This will reduce the sample to only US-based CVC units and the three above-mentioned industries, which will require a qualitative overview of the three main industries. Afterwards, the main regression models (primarily employing the Stata *probit* function) will be developed. These will be subject to robustness checks and exemplified with the case companies, which were introduced earlier. Finally, a discussion of the validity of the results and more will be performed. The discussion will, amongst other things, address assumptions, limitations and biases. Moreover, the discussion will be used to suggest further research within the field as well as describe for whom this thesis is relevant. Lastly, we will conclude on the thesis.

Figure 1: Structure of the thesis



2. LITERATURE REVIEW

This literature review will include (i) a very brief overview of some parts of the existing literature on CVC and CVC activities, (ii) an introduction of case companies to exemplify the phenomenon CVC, (iii) an in-depth review of academic papers related to CVC unit structure, which will lead to the research gap this thesis aims to bridge (as was briefly summarized in the introduction) and (iv) a short summary of the most important theoretical concepts that are applied in this thesis. This review will not include a comprehensive review of all literature on CVC, as this is deemed outside the scope of this thesis.

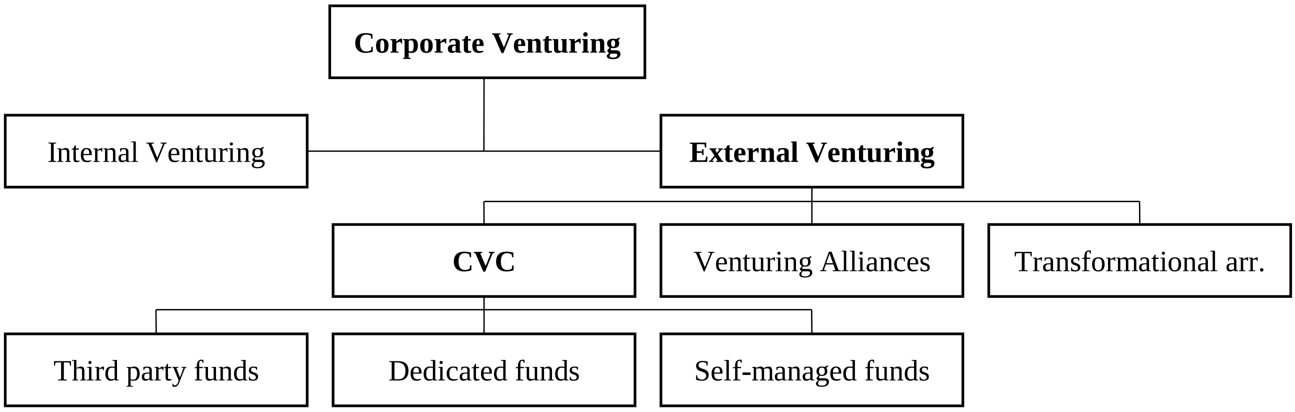
2.1. What is Corporate Venture Capital?

While there are many ways to define CVC and some degree of term confusion, this paper will apply the definition proposed by Dushnitsky (2008): "Minority equity investment by an established corporation in a privately-held entrepreneurial venture" (p. 2). More specifically, Dushnitsky defines three characteristics that are shared amongst CVC investments: (i) investments are often based on strategic objectives and not solely focused on financial returns, (ii) the ventures are independent and privately owned and (iii) the investments are minority equity stakes (Dushnitsky, 2008).

As several researchers point out, it is important to distinguish between CVC and other types of corporate activities, hereunder corporate venturing, which is more broadly defined and also comprises investments in internal entrepreneurial initiatives (Dushnitsky, 2008; Dushnitsky & Lenox, 2006).

Figure 2 shows a hierarchy of corporate venturing activities (Keil, 2002). As shown, CVC is a subcategory to external venturing. Only CVC activities will be covered in this thesis.

Figure 2: Modes used for external corporate venturing



Note. Illustration of own making based on Keil (2002)

The objectives of CVC units have been discussed extensively. Common for the research on CVC is that strategic objectives are emphasized as important in most CVC investments. One of the early papers on CVC cites the following eight reasons that companies engage in CVC investments: (i) identifying technologies and products to leverage, (ii) gaining insights about management before a possible complete acquisition, (iii) cheaper/faster production than internally, (iv) an “early window” into new developments (e.g. technology), (v) assuring supply, (vi) a research opportunity into new markets/methods, (vii) a way to redeem value from failed internal projects and (viii) a value-add for suppliers and customers (Rind, 1981). Newer research offers more nuances, for instance, Dushnitsky and Lenox (2006) find that CVC investments create more firm value when firms explicitly engage in CVC to harness new technology. There are a number of related areas of research within the field of CVC, that lie outside the scope of this thesis.

To offer a brief introduction to CVC activities over the years, the *three waves* of CVC activities will be described. For a more in-depth review of these three waves, see Dushnitsky (2008). One can argue that the recent high levels of CVC activity is a fourth wave.

The first wave of CVC activities started in the middle of the 1960s. This wave is called Conglomerate Venture Capital (CB Insights, 2017b). According to Dushnitsky (2008), the rise of CVC was driven by three factors; (i) a diversification trend, (ii) excess cash and (iii) inspiration from the success of independent VCs. About 25% of Fortune 500 firms engaged in CVC at the time. Already from the first wave, CVC units were structured both externally and internally

(Dushnitsky, 2008). This first wave ultimately ended when the IPO markets collapsed in 1973 (Dushnitsky, 2008).

The second wave was initiated in the 1980s. This wave is also called the Silicon Valley wave (CB Insights, 2017b). This time, the rising level of CVC activities was driven by changes in legislation, new commercial opportunities (especially in technology) and favourable market conditions. (Dushnitsky, 2008). When the market crashed in 1987, a sharp decline of CVC activities followed.

The third wave occurred in the 1990s. Access to new technologies were the primary driver of the corporations' wish to set up CVC funds. The number of CVC funds rose to new heights. By 2000, 15% of all VC investments (approximately USD 16 billion) were made by CVC funds (Dushnitsky, 2008). Even after the 2001 IT bubble crash, many corporations continued their CVC activities (Chesbrough, 2002).

While it is evident that CVC activities come in waves, CVC today is more prolific than ever. As stated in the introduction, global CVC financing reached USD 28.6 billion in 2015 (CB Insights, 2017a).

In 2017 alone, 166 new CVC investors entered the global CVC market. This represents a 66% growth in new entrants compared to 2016. Today, a number of well-known firms are making a large number of investments. Google Ventures, Intel Capital, Novartis Venture Fund, Roche Venture Fund, Novo Holdings (investing through Novo Ventures) and Pfizer Venture Investments are among the most prolific CVC investors today. Similarly, a large number of prominent start-ups have accepted CVC investments, including, but not limited to: Dollar Shave Club, Flatiron Health, Corvus Pharmaceuticals, Dropbox and Uber (CB Insights, 2018a).

The internet sector accounts for more than 40% of all CVC activities. Mobile and telecommunications account for around 15%, closely followed by Healthcare, Software and Computer and Hardware. Other sectors account for approximately 15% of all CVC investments (CB Insights, 2017a). Evidently, high-tech industries attract the most investments.

To bridge and illustrate the academic field and the state of CVC today, three case companies will be introduced in the following section. The aim is to link the practical side of CVC (e.g. their investment activities) with the aim of this thesis, i.e. investigating antecedents to the choice of setting up CVC units internally or externally. Hence, the case companies only serve illustrative purposes. The illustrative case companies are Intel Corp., Netscape Communications Corp. and Microsoft Corp.

2.2. Case companies

We exemplify the phenomenon of CVC as well as the organizational set-up as internal or external units by introducing three case companies that engage in CVC, to which we will refer at later points in this thesis. In the following, we will introduce each of them in separate.

2.2.1. Intel Corp. (external CVC unit)

The by far biggest corporate investor engaging in Venture Capital is the Intel Corporation with 1551 investments from 1992 until 2015 recorded in our sample. Intel invests through their CVC unit Intel Capital Corp., which is set-up as a wholly-owned subsidiary and hence an external unit according to the definition employed in this thesis. Intel is one of the companies that has most consistently engaged in CVC, and renewed their commitment to CVC investment activity throughout the downs in CVC activity worldwide (Chesbrough, 2002).

According to its website (Intel, 2018a), Intel Capital has been active since 1991, having invested more than USD 12.3 billion in 1,530 portfolio companies and 57 countries worldwide as of June 2018. Out of the portfolio companies, 660 have been acquired or have gone public. In 2017 alone, USD 690 million were invested through almost 90 investments, of which half were new targets and the remaining half follow-on investments (Intel, 2018a). In over 60% of their deals, they have been the lead investor in 2017 (Intel, 2018a). Successful investments include, for example, V-Cube, a video conferencing service in the Asia-Pacific region, Gudeng, a semiconductor supply manufacturer that was able to go public two years after the Intel investment, and Performance Lab, active in sports wearables (Intel, 2018b).

The Intel Corp. is active in the industry classified as SIC code 3764, namely “Semiconductors and Related Devices”. The corporation describes its emphasis of CVC as “building technology ecosystems”, with investments in, amongst others, Artificial Intelligence, Internet of Things and Autonomous Driving, Drones and Robotics, Software and Security, and Sports, Health and Entertainment (Intel, 2018a). Chesbrough (2002) classifies Intel’s investment as both enabling and emergent, investing in companies with products that are complementary to Intel's strategy, but also in targets that are active in technologies which could be future substitutes of Intel's technologies, and hence hedging against their current technology direction. For example, Intel's investment in Berkeley Networks in 1997, who did not use the prevailing Ethernet standard as Intel but a competing approach, "helped Intel identify a promising opportunity more quickly than it might have otherwise" (Chesbrough, 2002, p. 9).

The Intel Corp. states that its venture capital arm supports its strategic objectives (Intel Corp., 2018), and Intel Capital is an often-named example of successful CVC activity (e.g. Mawson, 2017). Intel Capital attributes its success to the possibility of target companies to draw from its technological expertise, brand capital and their business development programs and various events, which provide investees access to a global network (Intel, 2018c). This is confirmed by the CEO of one of the portfolio companies, Maana, who stated that the corporate events held by Intel Capital were crucial in Maana's early history, much more important than money or marketing (Mawson, 2018).

However, there has also been criticism with regards to Intel's investment activity, amongst it the concern that Intel cannot possibly actively coordinate and share its own resources and operations with targets due to the large number of investments (Chesbrough, 2002). However, Chesbrough (2002) argues that this was originally not the goal of Intel's CVC activity, but rather to increase demand for its own products through the activities of the portfolio companies. Intel Capital's former president, Arvind Sodhani, confirmed this by stating in an interview that by the CVC activity, Intel helped "create a new industry that in turn will need a lot of our [Intel's] products" (Beyers, 2016).

2.2.2. Netscape Communications Corp. (internal CVC unit)

Netscape Communications Corp., founded as Mosaic Communications Corp. in 1994, was mainly known for its internet browser "Navigator" (Norr, 2017). The company was active in the industry with the SIC code 7372, namely Prepackaged Software. It went public in 1995 through a private placement, and investors included, amongst others, well-known companies such as The Hearst Corporation and Adobe Systems Inc (Hearst Communications Inc, 1995). While going public, it doubled the value of its shares during the first day, which signalled the beginning of the dot com boom, leading to a peak in CVC activity (CB Insights, 2017b). The market capitalization of the company, which was less than two years old at that point, amounted to USD 2.2 billion (Norr, 2017). These proceeds enabled access to external knowledge, through small acquisitions, joint ventures, and also CVC investments (Norr, 2017).

In 1996, Netscape made its first CVC investment of USD 1 million in the company Voxware Inc, a voice-processing company (CNET, 1996; Lohr, 1997). The investment was part of its plan to integrate voice, sound and video in its web browser, and Netscape simultaneously signed a licensing agreement to use the Voxware's technology within Netscape products (CNET, 1996; Netscape Communications Corp, 1996). In return, Voxware profited from the endorsement by the, at this time, well-known investing company, especially with regards to its initial public offering (Lohr, 1997). Within the next two years, three more CVC investments in different companies followed, amounting to four total investments in the period from 1996 to 1998, with an estimated equity investment of around USD 9 million. All of the investments were made through an internal CVC unit according to the definition of this thesis.

America Online (AOL) announced to acquire Netscape in November 1998 for USD 4.3 billion (Corcoran, 1998) before the burst of the dot com bubble, which marked the end of Netscape Communications Inc and its CVC activity.

2.2.3. Microsoft Corp. (internal CVC unit)

The Microsoft Corporation, active in the industry with the SIC code 7372, namely Prepackaged Software, has made CVC investments as early as 1987 in our sample. From 1987 to 2011, the company engaged in CVC internally according to our definition, and investments were made directly by the parent company – 118 investments throughout the investment period, amounting to a total estimated equity of over USD 1 billion. From 2014 to 2015, three investments were made through the CVC unit Microsoft Ventures, which, however, did (at that point) not operate as a separate legal entity and hence still can be categorized as internal.

Chesbrough (2002) classifies Microsoft's investments as "driving", meaning advancing the current business strategy by establishing close links between the investee and the investor: For example, Microsoft invested in start-ups using its Internet services architecture ".Net" and by that aimed to establish a new standard. As Microsoft provides resources to target companies that enable them to develop their products (such as software and related tools), their usage makes target companies “tightly linked to Microsoft’s operation’s” (Chesbrough, 2002, p. 6), implying a certain degree of dependence.

Microsoft, as one of the early players in CVC, experienced some of the downsides of CVC as well. In the third quarter of 2000, it had to write off almost USD 1 billion of its CVC portfolio (Chesbrough, 2002), and more than USD 5.7 billion in 2001 as a consequence of the burst of the internet bubble (CB Insights, 2017b). The number of investments decreased significantly after that year, with only 14 investments from 2002 to 2011. There was no formal legally separate CVC unit at that time – Microsoft Ventures, which had three investments from 2014 to 2015, operated internally as an accelerator program (Kashyap, 2018).

A formal, independent CVC unit, was not founded until 2016, lying outside the investment period recorded in our dataset, which includes investments up until 2015. The unit initially operated under the same name (“Microsoft Ventures”), while the former Microsoft Ventures was renamed to “Microsoft Accelerator” and is today known as “Microsoft ScaleUp” (Foley, 2018; Kashyap, 2018). In 2018, the CVC unit changed its name to “M12” to avoid confusion when approaching entrepreneurs (Kashyap, 2018). In 2016 and 2017, after the creation of a formal CVC unit, Microsoft ranked amongst the ten most active CVC investors (CB Insights, 2018b), with over 50 investments within that period (Kashyap, 2018). M12 describes its benefits as acting as autonomously and fast as traditional VCs, but at the same time providing patient capital, as it does not need to raise or pay back money in a set timeframe (Kashyap, 2018; M12, 2018). Furthermore, target companies gain strategic access to customers, resources, e.g. its sales team, partners and relations (M12, 2018).

In summary, CVC is a phenomenon that has been pursued to different degrees and within different organizational set-ups by the case companies. Intel Corp. engages in CVC through an external unit, and Netscape and Microsoft through an internal unit. As can be seen, companies’ motives and their investment strategy differ, but advantages for target companies are often described similarly.

2.3. <u>Structure of CVC units: performance</u>

While literature has largely neglected to investigate antecedents of the set-up of CVC units as either internal or external units, there have been different contributions relating the structural aspects to organizational performance and learning. This literature review aims to give an overview of structure-related work in the field of CVC and is meant to provide the reader with a contextual understanding of the topic. Furthermore, it will become clear that the arguments used have implications for theorizing possible antecedents in section 4, i.e. *why* CVC units are set-up differently as internal or external units.

It is important to note that arguments from related areas often are applied within the context of CVC, mainly from the broader field of corporate venturing, including the organization of internally developed entrepreneurial initiatives (e.g. Birkinshaw, van Basten Batenburg, & Murray, 2002; Burgelman, 1983, 1984; Chesbrough, 2000). In some cases, we deem the transfer of these arguments valid, but it is important to be aware of the definitional distinction as explained in the section 2.1. This literature review primarily aims to provide a summary of work related to CVC specifically, and referrals to literature using a broader definition (e.g. corporate venturing) will be explicitly marked.

An overview of the structure-related studies in the context of CVC can be found in *Table 1*. Evidently, there has been some inconsistences in the definition of structure, and the emphasis on structure differs depending on the study.

Only a few studies consider the legal set-up as applied in this thesis as part of the definition of the structure of the CVC unit (Lee et al., 2018; McNally, 1997; Winters & Murfin, 1988; Yang, Chen, & Zhang, 2016). Furthermore, due to the nature of the organizational structure-related data, most studies are based on self-reported surveys, sometimes accompanied by third-party or archival data.

The structure or programme governance of CVC has been investigated differently in different studies, namely as (i) the presence of investment intermediation, meaning if the CVC unit invest *directly* or *indirectly* through a Limited Partner, (ii) a combination of investment intermediation and legal set-up and (iii) venture unit autonomy, mostly with regards to decision making. As scholars differ in nuance and measurement of structure-related factors, the literature review will follow the logic of the respective argumentation of the author and will be organised according to the three broad definitional groups.

Table 1: Existing studies concerned with the structure of CVC units

N°	Author(s), year	Perspective	Dimension(s) / Definition of structure	Role of structure	Data used with regards to structure	Focus
1	Rind (1981)	Performance (Corp. development)	Investment intermediation	Determining factor	Theory	CVC
2	Sykes (1986)	Performance (technical, financial)	Decision-making autonomy	Determining factor	Case study	CV, incl. CVC
3	Siegel, Siegel &, MacMillan (1988)	Performance (strategic, financial)	Venture unit autonomy (decision-making, financial commitment)	Determining factor	Survey	CVC
4	Winters & Murfin (1988)	Performance (strategic)	Investment intermediation, legal structure, objectives	Determining factor	Anecdotal evidence, theory	CVC
5	Sykes (1990)	Performance (strategic)	Investment intermediation	Determining factor	Survey	CVC
6	McNally (1997)	Characteristics / Investment process	Investment intermediation, legal structure	Determining factor	Survey	CVC
7	Miles & Covin (2002)	Decision framework	Investment intermediation	Configu-rational element	Survey, anecdotal evidence	CV, incl. CVC
8/9	Keil (2002; 2004)	Performance (learning)	Structural autonomy	Boundary condition	Survey	Ext. CV, incl. CVC
10	Weber & Weber (2005)	Performance (strategic, financial)	Venture unit autonomy (decision-making, financial commitment)	Determining factor	Survey	CVC
11	Hill & Birkinshaw (2008)	Performance (strategic, financial, survival)	"Organizational profile"	Configu-rational element	Survey	CV, incl. CVC
12	Souitaris & Liu (2012)	Antecedent	Specialization, centrali-zation, standardization, communication	Determined factor	Survey	CVC
13	Yang, Chen & Zhang (2016)	Performance (portfolio diversification)	Legal structure	Determining factor	VentureXpert, survey	CVC
14	Lee, Park & Kang (2018)	Performance (innovation)	Legal structure	Determining factor	Online databases, e.g. Lexis-Nexis	CVC

Note. The papers have been numbered for future reference throughout this thesis, as to guide the reader through the section.

2.3.1. Presence of investment intermediation

One stream of literature differentiates structural types depending on the presence of investment intermediation, meaning if the CVC unit invests *directly* in a target company or *indirectly* as a limited partner (LP) of an outside VC fund. The key takeaway of this section is that investment intermediation has been found to be related to both financial and strategic performance, as well as to the degree of control by the parent, along with its willingness to commit and share resources. While the definition is different, this alludes to the fact that setting up a CVC unit either internally or externally can have an impact on how it is perceived in the market, and affect the relationship with the target company, and hence the degree of resource-sharing.

Sykes (1990) (paper 5) finds investment intermediation, and consequently organizational structure of the CVC unit, to have a significant influence on strategic performance. Specifically, this is due to a different degree of (i) relationship-building in direct and indirect investments and (ii) reputational effects based on corporate compensation schemes and motives compared to VC.

Firstly, direct investments allow the corporation to build unique, high-quality business relationships with target companies (Rind, 1981; Sykes, 1990). Meanwhile, investing indirectly implies a greater effort to build a specific relationship with entrepreneurs, which can be time-consuming (Rind, 1981) (paper 1). Secondly, direct investments can entail multiple disadvantages in relation to the *reputation* of CVC in the VC community. For example, CVC units encounter difficulty to recruit skilled staff, as corporations cannot always offer an incentivized compensation scheme to attract experienced people from the VC environment similar to independent funds (Rind, 1981). In addition, it is difficult for a CVC unit that invests directly to establish a sufficient deal flow because motives, strategies and time commitment often differ from independent VC funds (Rind, 1981). These well-known differences manifest themselves in reputational effects, which often obstruct CVC actors from being accepted as equals by the VC community (Rind, 1981). In line with this argument, Sykes (1990) observes a relatively better deal flow and contact to the VC community for indirect compared to direct investments.

Whereas Sykes (1990) focuses on the strategic performance of differently structured CVC units, Miles and Covin (2002) (paper 7) shift the focus to a managerial decision framework to *select* an organizational form of corporate venturing (including CVC). With regards to CVC specifically, the choice between direct and indirect investments differs with regards to corporate management needs, including (i) the desired degree of control, (ii) willingness to commit resources and (iii) the acceptance of entrepreneurial risk. Hereby, the definition of indirect investments is broader than in previous studies, as it includes funds that are wholly-owned by the parent company and managed by its employees (which are in this thesis defined as “external”, making implications transferrable to the context of definition and hence, research question). The framework proposes investing *directly* if the need to control the target company, the ability of the parent company to commit resources or the acceptance of entrepreneurial risk is high.

Firstly, *direct* investments come with a high level of control over the target. They are undertaken in order to ease transferring technologies, resources and capabilities between the parent and the target company. Indirect investments, on the other hand, are not oriented towards sharing parent-owned technologies, capabilities or other resources, even though they can be used to access new markets and technologies. Consequently, the structural set-up of CVC investment has implications for the degree of resource-sharing between the investor and the investee. Secondly, the form of investment is related to the degree of resources commitment by the parent. Investing directly might provoke conflicts of interests between (often internal) stakeholders, who are concerned about the allocation of scarce resources within the company. Thirdly, venturing activity can put the corporation at risk. For example, it can be damaging for the company’s reputation, its brand or intellectual capital, depending on the operating culture of the target company. In indirect investments, downside risks are lower than in direct investments, and mostly of financial nature (Miles & Covin, 2002).

In short, the decision of investing directly or indirectly in external entrepreneurial initiatives should be taken based on an assessment of three parameters (Miles & Covin, 2002): degree of control, commitment of resources and risk acceptance, all of which are parameters that have direct or indirect consequences for the main dependent variable of this thesis. For example, the degree of resource sharing is affected by the set-up of a unit, and hence differs in external and internal CVC units.

To summarize the work on the presence of investment intermediation, two main implications are of relevance for this thesis: Firstly, it proxies the *distance* to the target company, which affects the degree of relationship-building, resource-sharing and control between the entrepreneur, the VC community and the parent company. Secondly, the types of investment require a different degree of *involvement* of the parent corporation, which has implications with regards to the internal allocation of time, resources, information and risk.

2.3.2. Combination of investment intermediation and legal structure

A second stream of literature broadens the definition of structure by taking the legal-set up of CVC units into account in addition to investment intermediation, and thus constitutes a direct link to the definition used in this thesis. Most importantly, the findings show that the organizational structure of a CVC unit is related to the organization's objectives, as it can influence relations to the VC community, mitigate concerns by entrepreneurs and increase internal support for the CVC activity.

Winters and Murfin (1988) (paper 4) expand the structural distinction of investing directly or indirectly by additionally introducing two forms of CVC subsidiary, which are both legally separated units but differ in their objectives. Namely, those are (i) a subsidiary operating similar to independent VC with primarily financial objectives, and (ii) a "venture development subsidiary", which exercises an extended strategic scope. In the context of this thesis, the latter is most relevant (although the former is not completely irrelevant). The venture development subsidiary is considered an optimal structure to maximize strategic gains due to three major benefits. Firstly, it signals commitment of the organization to CVC activity internally. Secondly, a formal subsidiary facilitates relations with the target company, as it can mitigate some of the concerns faced by entrepreneurs when collaborating with an established corporation (e.g. appropriation, asset stripping and the like). Thirdly, it increases approval by the VC community, as it signals commitment and a low degree of bureaucracy. However, Winters and Murfin (1988) recommend corporate involvement by both corporate executives as well as business units representatives in the decision-making process, even if the subsidiary is otherwise structurally separated from the parent organization. Hereby, it is ensured that the attention does not lie exclusively on existing business developments. In summary, Winters and Murfin (1988) identify the key success factors for strategic gains in CVC as high deal exposure, the combination of people that manage the CVC unit, contact to the VC community, long-term commitment, co-investors and internal management support, which are best supported by a formal subsidiary with strategic objectives.

Adding a level of granularity, McNally (1997) (paper 6) combines the two structural dimensions, investment intermediation and legal set-up (in-house operating division or separate subsidiary). Based on a study of the CVC activity in the UK, it is found that both in-house CVC units as well as formal subsidiaries make direct and indirect investments. With regards to goal achievement, direct investments are better suited to obtain strategic objectives, while indirect investments primarily pursue financial goals. The organizational structure is related to concepts of decision-making and funding authority in the context of CVC programme governance: CVC units that invest directly often encounter a strict corporate approval process. However, a separate subsidiary is generally associated with a higher degree of autonomy in decision-making (McNally, 1997), confirming the notion that structural autonomy is closely tied to separation in practice. In line with this observation, recent studies have used the legal set-up of the CVC unit (internal programme vs. wholly-owned subsidiary) as a proxy of venture unit autonomy (Lee et al., 2018; Yang et al., 2016). This confirms the validity of the measure used in this thesis, where we define an internal or external unit according to its legal structure.

In summary, investing through a legally separate CVC unit has similar implications to indirect investments while still being able to draw on advantages of investing directly. Establishing a subsidiary, which invests strategically for its parent company, signals a high degree of venture unit autonomy through the creation of formal *distance*, which is an important concept when theorizing about antecedents of setting up a CVC unit either internally or externally.

2.3.3. Venture unit autonomy

The largest stream of literature in the context of structure is concerned with venture unit autonomy, largely associated with authority in decision-making. Compared to the previously described definitions, venture unit autonomy is defined less consistently, as it is derived through survey data, and different studies use slightly different measures to proxy the concept. In the following, we will briefly summarize relevant findings of the literature on corporate venturing, followed by an in-depth review of studies related to CVC specifically. The findings are, in short, that venture unit autonomy has been largely associated with enhanced performance, but structural separation entails downsides as well, such as increased difficulty of sharing resources. This could, in turn, influence the decision to set-up a CVC unit externally or internally, which makes it relevant in the context of the research question.

The effect of the venturing unit's autonomy in relation to performance has been extensively discussed by scholars with regards to corporate venturing and entrepreneurship (e.g. Birkinshaw et al., 2002; Burgelman, 1983, 1984; Chesbrough, 2000). As goals, time-horizon, risk-aversion and speed of the corporate venturing activity differ from the parent organization, a separate venture unit outside the established organizational structure is often recommended (Birkinshaw & Hill, 2005; Birkinshaw et al., 2002). This allows for decision-making autonomy, strong links to the VC community and an incentive-based compensation (Birkinshaw et al., 2002). This has been investigated empirically, and venture unit autonomy has been found to be positively related to both financial and strategic performance (Hill, Maula, Birkinshaw, & Murray, 2009). On the downside, autonomy can create barriers to innovative success, as the venture unit is essentially dependent on the parent company's resources (Garrett & Neubaum, 2013), from which autonomy creates distance; and a certain amount of control is required to ensure alignment (Crockett, McGee, & Payne, 2013), which can be of relevance for the decision to set up an external or an internal CVC unit.

In the context of CVC specifically, venture unit autonomy is analysed in relation to performance, defined as (i) the achievement of the CVC unit's financial or strategic objectives, (ii) organizational learning and (iii) portfolio diversification. The key implication for the research question of this thesis is that the structure of the CVC unit is related to the learning processes undertaken through the CVC activity, its ability to leverage the resources of the parent, and the focus of attention of the CVC unit.

Firstly, the structural autonomy of CVC units has been investigated with regards to financial or strategic goal achievement. This review includes three contributions, namely the work of Sykes (1986) (paper 2), Siegel, Siegel, and MacMillan (1988) (paper 3) and Weber and Weber (2005) (paper 9).

Sykes (1986) (paper 2), who studies both internal and external corporate venturing activities, finds that structural factors such as decision-making autonomy are related to venture success. He identifies both advantages and disadvantages of structural autonomy. The former includes (i) the motivation of employees (for example because it allows for a faster decision-making processes) and (ii) a reduction of potential conflicts of interest, as objectives of the CV unit might differ from the parent corporation. However, the degree of separation is dependent on the characteristics of the business, as a greater distance makes the venture more reliant on developing own processes, resources and capabilities (considerations which can be of relevance when deciding to set up an external or internal CVC unit). In total, structural factors are found to be related to venture performance, but less significantly than intrinsic factors related to the venture itself, such as manager experience (Sykes, 1986).

In contrast to Sykes (1986), Siegel et al. (1988) (paper 3) focus exclusively on organizational structure. They find organizational independence with regards to funding and decision-making authority to enhance financial performance, but evidence is less conclusive for strategic performance. Specifically, the financial performance of CVC units that operate with a high degree of authority and receive sustainable financial commitment by their parent (so-called “pilots”) is higher than for CVC units that are highly dependent on their parent corporation (so-called “co-pilots”). With regards to strategic performance, the effect is less clear: concerning the exposure to new technologies and markets, there is no significant difference between pilots and co-pilots, but co-pilots show higher performance regarding acquisition candidates. However, pilots rate strategic obstacles as less damaging than co-pilots, which include a lack of clear mission by the parent company, a lack of patience and flexibility by corporate management, and an inadequate deal flow. The same holds for obstacles related to the entrepreneur, namely the fear of appropriation of their idea and corporate control, and the general incompatibility between the culture of the target and the corporation. In total, pilots and co-pilots do not perform significantly different with regards to strategic performance, making a largely independent CVC unit the overall superior model (Siegel et al., 1988). However, the nuance that is added with regards to strategic performance highlights that co-pilots also seem to encounter certain strategic advantages through their structural set-up, which could be of relevance in connection with antecedents to set up the CVC activity in an internal or an external unit.

Unlike Siegel et al. (1988), Weber and Weber (2005) (paper 9) find differing results for the parameters decision-making autonomy and financial commitment of the parent company in separate. Specifically, CVC units with a larger decision-making autonomy to a larger extent achieve both financial as well as strategic objectives, but results are less consistent with regards to financial commitment of the parent company. Hereby, strategic goal achievement seems to be higher for units with freely available funds, whereas financial goal fulfilment was higher for CVC units that could not freely access a pool of money.

As shown, studies that set CVC unit structure in relation to financial and strategic performance have yielded slightly differing results, and no structural set-up is consistently seen as superior in all aspects, which highlights the relevance of our research question with regards to *why* companies would set up their CVC unit internally or externally. Further insights can be offered by relating organizational structure of a CVC unit, specifically venture unit autonomy and structural independence, to organizational learning (Keil, 2002, 2004) (papers 8/9).

In the context of organizational learning, Keil (2002) argues that external corporate venturing is essential to both explorative and exploitative activities of the corporation, whereby the organizational set-up results in different learning mechanisms. Through the example of two case companies (which are called ALPHA and BETA), learning mechanisms of differently structured CVC units are compared (Keil, 2004). Even though both companies create a separate division for corporate venturing, ALPHA's unit combines CVC and other corporate venturing activities, is closely integrated in the corporation's core business and dependent on its expertise and funding. BETA's CVC unit is, on the other hand, completely separated from the core business, physically as well as with regards to funding and resources. It operates under high degree of autonomy, similar to the "pilots" defined by Siegel et al. (1988).

Specifically, organizational structure is viewed as an initial boundary condition which determines the learning process and path undertaken by the organization (Keil, 2004). Namely, it manifests itself through (i) the ability to create initial knowledge and (ii) the transfer of knowledge. Firstly, BETA successfully implemented operations similar to an independent VC with a separate, organizationally independent unit, which allowed comparable compensation structure to VC and thus attracting experienced personnel. This resulted in the creation of initial knowledge and consequently a faster learning process compared to ALPHA, which tried to imitate the investment process of independent VC but did not succeed because their organizational structure was incompatible. The independent structure of BETA further enabled the CVC unit to make investments independent of corporate, short-term interests. Secondly, knowledge transfer is essential for the learning process and can take place through networks (formal and informal) and through processes of codification, which is in turn influenced by the organizational structure. For ALPHA, both time and resources were restrained, leading to a less efficient codification. On the contrary, codification and social contacts were stronger in BETA, which facilitated the exchange of knowledge.[5] In the context of the research question, this implies that the use of knowledge differs in external and internal CVC units, which is of relevance for antecedents to the set-up.

Taking into account the notion of organizational learning, Lee et al. (2018) (paper 14) add a level of granularity to performance-based studies by analysing the relation of structural autonomy to both explorative and exploitative innovation performance of the parent company in separate, while defining internal and external CVC units in the same manner as this thesis. They find that an external structure is positively related to *explorative* innovation performance, but negatively related to *exploitative* innovation performance. The reason is that a completely separate CVC unit is structurally disconnected from its parent's resources, including skilled personnel and tacit knowledge, making it more difficult to access and leverage them – an argument which will be applied in this thesis when theorizing about antecedents of setting up a CVC unit externally or internally. According to Lee et al. (2018), the structural disconnection is especially problematic if the goal of the unit is of exploitative nature and if the target companies rely on compatible resources in the same field, as the distance could impede effective collaboration. In line with this argumentation, they find a significant, negative relationship between CVC unit autonomy and the exploitative performance of the parent.

Closely tying into the concept of organizational learning, Hill and Birkinshaw (2008) (paper 11) argue that the *fit* between the organizational profile and venture type impacts venture performance, as the organizational profile can influence the degree of resource-sharing, for example. Specifically, they find that an alignment between organizational structure and strategic logic results in higher performance. Four distinct venture types are distinguished, based on the locus of opportunity, meaning if the idea is internal or external to the organization (in this context, external is of interest)[6] and the strategic logic of the venture unit, i.e. if the focus lies primarily on exploring novel business areas or exploiting existing assets or capabilities. The organizational profile hereby includes (i) the network of relationships, (ii) the activities and (iii) the management systems of a CVC unit. Venture unit autonomy plays a role primarily in (i): the relationship between the CVC unit and corporate executives of the organization and operational autonomy influence the degree of resource-

[5] While this is not completely in line with later findings, as will be shown, it should be noted that this paper is case-based and as such does not offer statistically generalizable findings.
[6] Their definition of external and internal locus of opportunity relates to where the entrepreneurial idea resides (similarly to the notion of internal and external corporate venturing, as explained in section 2.1). CVC hereby falls under "external", which is why only this part is reviewed in this context.

leveraging and learning, which is relevant for internal or external CVC units in the context of our research question, respectively. Hill and Birkinshaw (2008) find that an alignment between organizational structure and strategic logic results in a higher performance. However, it becomes less important with regards to venture unit survival in general – this is determined solely by the strategic logic itself.

The literature on CVC unit structure and organizational learning has shown that differences arise with regards to the learning process undertaken by the organization depending on the organizational structure, which is related to the degree of resource- and knowledge-sharing.

Introducing a different dimension of performance, Yang et al. (2016) (paper 13) study the effect of CVC unit structure (measured as in this thesis) on portfolio diversification from an attention-based perspective. They find CVC unit autonomy to be positively related to the industry diversification of the CVC portfolio. This is due to two effects: Firstly, a tightly controlled, internal CVC unit might retain a narrow focus bound by the corporation's overall strategy, and hence result in a less diversified, mostly exploitation-oriented CVC portfolio, than a highly autonomous programme. Secondly, managers of highly independent CVC units are less likely to be subject to myopic tendencies residing within the organization and can focus their attention outside the current business (Yang et al., 2016). In the context of the research question, this notion of myopia is important with regards to why organizations set up a CVC unit internally or externally.

In short, the following three effects are most relevant in the context of this thesis: Firstly, and consistent with the idea of distance in studies concerned with investment intermediation, structural autonomy allows a higher degree of flexibility, but does not perform strategically better in all aspects. Secondly, the creation of formal distance can both enable *and* restrict the process of learning depending on other organizational characteristics and objectives, which implies that the organizational structure of the CVC unit has to be adjusted according to the abilities of the parent, which justifies the perspective on the research question taken by this thesis. Thirdly, this alludes to the fact that external units are less likely to retain a narrow focus of attention due to a sense of myopia and can mitigate conflicts of interest between the parent organization and the CVC unit. These findings are important when considering antecedents to the choice of setting up a CVC unit internally or externally.

2.4. Structure of CVC units: antecedents

There has been little systematic evidence for possible antecedents to the organizational set-up of CVC units, with the exception of Souitaris et al. (2012) (paper 12), whose definition of structure is much more broad than the definition used in this thesis. They view the organizational structure of the CVC programme as an outcome of "competing forces from two different institutional environments" (Souitaris et al., 2012, p. 477). Specifically, an institutional environment, whose primary goal is legitimacy with the parent corporation likely results in a mechanistic structure; whereas an institutional environment, which sets an external focus and primarily seeks legitimacy with the VC community and entrepreneurs, leads to an organic structure. Organic and mechanistic structures are differentiated through four dimensions[7], of which the definition applied in this thesis falls into the degree of centralization[8], which is lower for organic than for mechanistic structures. Consequently, the *orientation* of the CVC unit is seen as an antecedent to its structural set-up. However, Souitaris et al.'s, (2012) definition of structure is rather broad, and characteristics of the parent organization are not investigated in their paper, but suggested as a future direction of research.

As becomes evident, literature has provided little guidance on *why* CVC units are structured differently (as internal or external units) across firms. Scholars agree that there is no unique dominant way of structuring a CVC unit, and that the organizational set-up is contingent on goals, needs, capabilities and characteristics of the corporation (Winters & Murfin, 1988). This implies that no organizational form is *by definition* superior. However, there is little to no systematic evidence on CVC structure which is not performance- or anecdotally based. This thesis aims to address this research gap by examining which organizational dimensions have an influence on the set-up of internal or external CVC units. Hereby, as already described, we follow the definition first introduced by Dushnitsky and Shaver (2009) and differentiate between two set-ups, namely an (i) internal CVC unit, which is not legally separated from its parent (this includes dedicated, internal CVC divisions), and (ii) an external CVC unit, which is a stand-alone, separate legal entity, wholly-owned by the parent company. This approach is a simplified measure for autonomy, but allows for a

[7] For completeness, the four dimensions are: (i) specialization, (ii) centralization, (iii) standardization and formalization and (iv) the direction and content of communication. An organic structure, as opposed to a mechanistic structure, is characterized through a low degree of specialization, centralization and standardization, as well as a multidirectional, consultation-based communication (Souitaris et al., 2012).

[8] defined as the concentration of authority

systematic, stylized and quantitative study, which does not rely on self-reported surveys by corporate venture capitalists and thus takes a more objective perspective.

This concludes the literature review intended to justify the research question as set out in section 1.2.

2.5. <u>Theoretical Concepts</u>

This section will provide the reader with a brief introduction to the main theories that are applied in this paper. Specifically, these theories include the Resource-Based View, Transaction Cost Economics, Agency Theory and Organizational Learning. Rather than an exhaustive review of all contributions within these theoretical directions, this section will focus on the streams of literature that are pertinent to the applications in this thesis, in order to provide context.

2.5.1. Resource-Based View

Before the introduction of the Resource-Based View (RBV), the firm-specific resources and capabilities and their impact on firm performance played a negligent (though not completely absent) role in the strategic management literature (Barney, 1991). Rather, the dominant literature focused on the external environment, with Porter's five forces being among the most important contributions (Porter, 1980).

Barney (1991) defines (drawing on Daft, 1983) firm resources as "all assets, capabilities, organizational processes, firm attributes, information, knowledge, etc. controlled by a firm that enable the firm to conceive of and implement strategies that improve its efficiency and effectiveness" (p. 101).

The main insight of RBV (as developed by Barney, 1991; influenced by Wernerfelt, 1984) is that certain resources can be a source of Sustained Competitive Advantage (SCA), i.e. "implementing a value creating strategy not simultaneously being implemented by… competitors *and* when these other firms are unable to duplicate the benefits…" (Barney, 1991, p. 102). This is the case when resources display the following four characteristics: (i) valuable (exploitative of opportunities or threat-neutralizing), (ii) rare, (iii) perfectly inimitable and (iv) not substitutable (Barney, 1991). Collectively, these traits are referred to as VRIN. Consequently (and simply put), to achieve SCA, companies must strive to identify their resources, evaluate them based on VRIN-criteria and develop and protect these resources (Barney, 1991).

Resources can, broadly speaking, be categorized in three different categories: (i) tangible (physical assets, financial), (ii) intangible (technology, e.g. patents, reputation, culture) and (iii) human (skills, communication capacity, motivation etc.) (Grant, 2016).

The RBV is related to the literature on knowledge, knowledge transfer and -mobilization and innovation literature, as knowledge (in its many forms) and innovation processes themselves can be viewed as resources, and thus can be a source of SCA (see e.g. Grant, 2016).

2.5.2. Transaction Cost Economics

While Transaction Cost Economics (TCE) was not coined by a single person, Coase's work is considered seminal in this regard (see e.g. Coase, 1937). Originally, TCE was a theoretical explanation for the scope of the firm (i.e. if there are no transaction costs, why do economic activities need to be organized in the form of firms?). Today, TCE has expanded conceptually, and is used to explain a range of behaviours (Shelanski & Klein, 1995; Williamson, 1979), including choices of organizational form (e.g. licensing, vertical and lateral integration), contracting and more.

TCE offers insights into how parties protect themselves when entering into a transactional relationship, which is necessary as the underlying assumption is that *contracts are incomplete*. A key insight of TCE is that certain governance structures are better suited for certain situations (Shelanski & Klein, 1995). Transactions can, according to Williamson (1979), vary on the following parameters: (i) asset specificity, (ii) uncertainty, (iii) complexity and (iv) frequency. The suitable choice of governance structure depends on these parameters for any transaction, and trading partners will choose the governance structures that offers the lowest total costs; from spot market transactions to a fully-integrated firm (Shelanski & Klein, 1995) and various "hybrid" models in between, including minority equity stakes.

TCE describes various costs that are incurred for different types of transactions, which are governed in different ways. These include, but are not limited to, coordination costs (i.e. coordinating activities and transferring necessary information), including bargaining cost, search costs (e.g.

finding a suitable product and assessing its condition), a number of bureaucratic costs (hereunder influence cost), monitoring costs and more (Artz & Brush, 2000; Shelanski & Klein, 1995; Williamson, 1979).

A large amount of research has corroborated the significance of the findings of TCE; empirical evidence consistently supports the model (Shelanski & Klein, 1995).

2.5.3. Agency Theory

Whereas TCE is concerned with organizational forms and boundaries, agency theory analyses the relationship between different parties, independent from formal organizational boundaries (Eisenhardt, 1989). Namely, the relationship between a principal and an agent is of focus, where the former delegates tasks to the latter (Jensen & Meckling, 1976). This relationship is subject to a contractual agreement under incomplete information (Ross, 1973). Two streams of literature have evolved within agency theory; the positivist agency approach and the principal-agent approach, which can be seen as complementary to each other (Eisenhardt, 1989). As the latter covers a broader set of relationships, it will be in focus in the following review.

The simple agency theory is largely concerned with identifying the most efficient contract to resolve issues which arise through differing goals and attitudes towards risk by principal and agent (Eisenhardt, 1989). As the agent performs work on behalf of the principal, two forms of *agency problems* occur: Firstly, there is a potential conflict of interest as the agent's goal might deviate from the principal's, and secondly, the agent's actions are to a large extent unobservable for the principal and hence difficult to verify (Eisenhardt, 1989). These problems caused by information asymmetry can be differentiated into *adverse selection,* which is caused by unobservability of an agent's ability or skill, and *moral hazard*, which results from unobservability of the agent's effort (Demski & Feltham, 1978)[9]. Furthermore, next to agency problems, a problem of *risk-sharing* arises as the agent and the principal are assumed to have different attitudes toward risk, namely that the agent is risk-averse and the principal is risk-neutral (Shapiro, 2005). As the described problems entail *agency costs,* e.g. through monitoring and risk-shifting, the main concern of agency theory is to find the optimal contract to reduce agency problems and consequently costs (Demski & Feltham,

[9] An important condition for moral hazard to occur is that output is dependent on environmental factors *other* than the agent's effort. Hence, even though output is observable, the principal cannot distinguish to which extent this output is a function of the agent's effort (Grossman & Hart, 1983).

1978). This can be achieved through either controlling for behaviour or for outcome, for example by implementing incentivizing compensation structure (Eisenhardt, 1985). Hereby, the focus of principal-agent theories lies in finding a balance between the costs of measuring behaviour or measuring outcomes, respectively.

There have been various extensions of the simple agency model, for example through the modification of certain assumptions or the introduction of different types of agency relationships, which lie outside the scope of this review. A large body of literature on agency theory has been applied to and empirically tested in multiple areas of research, including economics, accounting, organizational theory and more (for a comprehensive overview, see Eisenhardt, 1989).

2.5.4. Organizational Learning

Organizational learning is a phenomenon widely explored by literature (e.g. Levitt & March, 1988), and examining all its facets would go beyond the scope of this thesis. A commonly accepted definition is that organizational learning is the modification of current knowledge, dependent on an organization's past experience (Argote, 2013). From a behavioural perspective, Levitt and March (1988) conceptualize organizational learning as capturing lessons from past experiences in routines to guide future behaviour, and thus as *history-dependent*, *routine-based* and *target-oriented*. Organizational learning is hereby dependent on internal factors, such as organizational structure and culture, as well as external factors, for example the relationship with other firms (Argote, 2013).

Two important mechanisms of organizational learning are exploration and exploitation. As noted by Gupta, Smith and Shalley (2006), the definition of these concepts is used ambiguously by literature. First introduced by March (1991), exploration is described as "things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation" (March, 1991, p. 71), whereas exploitation includes "things as refinement, choice, production, efficiency, selection, implementation, execution" (March, 1991, p. 71). An often-used definition is consequently that exploration is the search of novel, unknown alternatives outside the current organizational knowledge, whereas exploitation is the modification and leverage of existing knowledge or capabilities (March, 1991; Quintana-García & Benavides-Velasco, 2008). Consequently, the former is associated with a higher degree of uncertainty, whereas the latter is considered less variable in terms of outcome (March, 1991). Even though both learning activities compete for organizational resources (March, 1991), literature consistently suggests that exploration and exploitation need to be carefully balanced to improve organizational and innovation performance, as an exclusive focus on one or the other can be value-impeding in the long-term (e.g. Gupta et al., 2006; He & Wong, 2004; March, 1991). Katila and Ahuja (2002) further argue that both mechanisms of learning are related, as exploitation is not only important to refine *existing* technologies, but also necessary to create *new* knowledge, and consequently a pre-requisite for successful exploration.

This interplay between existing and new knowledge is also central within the notion of absorptive capacity, defined by Cohen and Levinthal (1990) as the "ability to recognize the value of new, external information, assimilate it and apply it to commercial ends" (p. 128). Absorptive capacity is dependent on the level of prior knowledge of the organization, making innovation capabilities path-dependent (Cohen & Levinthal, 1990). Zahra and George (2002) reconceptualise absorptive capacity to four dimensions, namely acquisition, assimilation, transformation and exploitation, which are based on a set of organizational routines and strategic processes. It is defined as a *dynamic* capability to which an organization's prior experience is an important, but not the sole influencing factor – knowledge complementarity and diversity exercise an influence as well (Zahra & George, 2002). Furthermore, a distinction between potential and realized absorptive capacity is introduced, emphasizing that value creation depends both on the ability to create a competitive advantage and sustain that advantage, composing an iterative process (Zahra & George, 2002).

The concept of organizational learning is vastly applied in different streams of research. The above introduced notions of organizational learning have been empirically investigated, for example by Fabrizio (2009) and Uotila, Maula, Keil, and Zahra (2009).

3. RESEARCH DESIGN

The following section will provide a brief explanation of how the research question will be approached, and why the chosen approach is deemed most suitable. Specifically, the approach to theory development, the objective of the research (descriptive vs. explanatory), the time horizon and the methodological approach will be addressed. While research design is a topic which by itself has been widely discussed in literature from different angles (for an overview, see for example Abbott

& McKinney, 2013; de Vaus, 2001; Matthews & Ross, 2010), this section does not aim to explore all nuances of research design.

3.1. **Approach to theory development**

Due to the unexplored nature of the topic addressed by our research question, this thesis aims to open a debate, rather than to provide an exhaustive study of possible antecedents to setting up a CVC unit internally or externally. Due to the limited scope of this thesis, we focus the search of influencing factors to innovation-related attributes, as laid out in the introduction. To address the research question, deductive and inductive research approaches are combined in this thesis. We apply a deductive approach, as we begin by theorizing potential links between knowledge-related dimensions of the parent organization and the choice of an internal or external CVC unit. By developing measures and collecting data, we then test the *ex-ante* proposed theories (Saunders, Lewis, & Thornhill, 2015). After having analysed obtained correlation and regression results, we apply an inductive approach, reassess the developed theories and theorize the strengths of the effects *ex post* (de Vaus, 2001; Saunders et al., 2015). Combining both approaches is deemed most suitable, mostly due to the novelty of the research, which makes theorizing an explorative exercise, stimulating an iterative process of theory development.

3.2. **Descriptive vs. explanatory research**

Research can be categorized into descriptive research, answering the question of *what*, and explanatory research, addressing the question of *why* (de Vaus, 2001). Descriptive research does not make causal inferences, whereas explanatory research aims to establish causal explanations about how a phenomenon (in our case, setting up CVC unit internally or externally) is affected by different factors (de Vaus, 2001). Causation can be either deterministic or probabilistic; the former assumes that a factor variable invariably produces a response, while the latter is concerned about how an independent variable affects the probability of a particular outcome (de Vaus, 2001). As the topic addressed in the research question is largely unexplored by literature, this thesis conducts mostly descriptive research. In section 6.1, we will present a descriptive overview of *what* is observed in the data with regards to the set-up of internal or external CVC units. This serves the purpose to provide the reader with a contextual understanding of the phenomenon and gives first insights on interesting aspects subject to further investigation. Moreover, the descriptive statistics part will yield findings that are academically interesting in their own right. Based on a theoretical analysis of *why* certain effects might influence the likelihood of setting up a CVC unit externally, this thesis explores potential explanations and linkages, however makes no causal claims. We do not develop clear, one-sided hypotheses, but rather explore different theoretical effects. However, the main purpose of this thesis is to test if the theoretical concepts are significantly (and, in which direction) *related* to the set-up of internal or external CVC units, as not all criteria linked to causation[10] are fulfilled in this context due to its limited scope and objective. It is very probable that other factors influence the choice of setting up a CVC unit internally or externally, which remain unexplored by this thesis.

3.3. Time horizon

The study is performed on the level of analysis of the CVC unit with regards to the response variable *subsidiary*, which remains unchanged throughout time (as will become evident when describing data collection). Consequently, the research design most suitable to derive a relationship between variables is a cross-sectional design, treating observations as time indifferent (Matthews & Ross, 2010). As factor variables, namely data related to the parent organization of the CVC unit, are retrieved *consistently* according to the CVC unit's maximum year of investment, eliminating the time dimension is deemed unproblematic.

3.4. Methodological choice

In general, the research design should be regarded as independent from the applied method, specifically a quantitative or qualitative method, or a combination of the two (de Vaus, 2001). However, certain designs are more likely to be suited for a quantitative or qualitative method (Matthews & Ross, 2010). Cross-sectional studies frequently use quantitative methods (Saunders et al., 2015), which is deemed most suitable in this thesis as well. Different from most prior studies related to the structure of CVC, qualitative data obtained through interviews or questionnaires will not be included. There are two main reasons for this, namely (i) that this thesis aims to address the ambiguity of definitions of structure in other papers by applying a systematic, replicable measure,

---

[10] The criteria for causation are defined by Abbott and McKinney (2013) as *time* (meaning if the dependent variable occurs after the independent variable(s)), *correlation* (independent and dependent variable vary with each other), and *non-spuriousness* (there is no unnoticed variable that could cause both dependent and independent variable).

and (ii) that this thesis was constrained in time and resources to nuance the applied definition by survey data. To explore a relationship between variables, this thesis employs an empirical method, through applying concepts of correlation and regression (Punch, 2014). The methodology applied will be explained in section 5.

4. THEORETICAL ANALYSIS

This paper takes on a pragmatic and, to a high extent, open-ended theorizing approach. We discuss possible theoretical links between the factors that are theorized to exert an influence on our variable of interest (i.e. whether the CVC unit is set-up internally or externally). As the area of research is largely unexplored, this thesis as an explorative exercise aims to keep an open-ended approach, and thus we *do not hypothesize* as to the direction of the overall effect.

Many of the arguments that will be applied will rely on the assumption of actors (in many cases managers) being rational. While this may not be consistent with reality, the following sections will attempt to bring several perspectives in play as to develop nuanced theoretical explanations of what is expected to be seen in the data.

4.1. Value of innovations

No previous papers directly relate the value of innovations to the setup of CVC units as internal or external. However, as previously shown, innovation and structure are related, and alignment between innovation strategy and organizational structure is vital (e.g. Hill & Birkinshaw, 2008). The value of a companies' patents is expected to exert an influence on the decision to engage in CVC with an external or internal unit.

Specifically, we theorize that three effects play a role in this relationship: (i) a *protection effect*, (ii) a *parenting advantage* and (iii) a *decentralization effect*.

Firstly, with valuable innovations, protection and control of the innovations are expected to be relatively more important - companies are expected to be more likely to want to protect their innovations if they have more to protect. As per TCE, an external unit is more vulnerable than an internal unit, as a *less integrated* governance structure offers less protection. Specifically, an external unit could be subject to misuse or appropriation of IP rights, which could hamper the value the firm itself can capture from its innovations/patents, and the general loss of control could make barriers to imitation lower. Theoretically, agency theory can nuance this further: As an external unit would formally use the technology (in form of patents) in a way similar to licensing, agency issues could arise between the parent company (the principal) and the subsidiary manager (the agent). These can manifest themselves in a range of different ways, including opportunistic behaviour (supporting the above points) or simply not being incentivized to protect the IP properly.[11] This *protection effect* is expected to have an effect on setting up the CVC unit internally or externally. Specifically, per the above theoretical arguments, we theorize that this effect leads to a negative relationship between a high value of patents the likelihood to set up the CVC unit externally.

Secondly, the value of patents is a sign of a successful innovation process and consequently, valuable innovation resources, which come in many forms (research facilities, subject matter experts to draw on and more). Specifically, corporate know-how and skills (Fast, 1979; Garrett & Neubaum, 2013), existing product facilities (Sorrentino & Williams, 1995; Garrett & Neubaum, 2013), high-quality technical knowledge and skilled experts (Lee et al., 2018) have been cited as resources that the investees can draw on after a CVC investment. Garrett and Neubaum (2013) have dubbed this effect *parenting advantage*. As per these arguments and the RBV (and complementary assets, specifically), such resources could be of high value to the target companies invested in when leveraged. Hence, it is established in the literature that leveraging parent resources is a primary driver when performing CVC (and on a related note, not just for innovation resources, but also cross-selling, access to sales channels and more)[12].

TCE predicts the coordination costs related to leveraging such resources to be higher for an external unit than an internal unit. This argument is supported by Lee et al. (2018), who state that there will be a “structural disconnection” from the valuable resources, if the CVC fund is operated as a completely external unit. Specifically, this manifests itself in a lack of cooperation with experts

[11] While many external CVC units are operated as VC funds in terms of incentive schemes (Hill, Maula, Birkinshaw, & Murray, 2009), which are designed to minimize agency issues between limited partners and general partners, this might help explain why many CVC units also decide *not* to fully implement such compensation mechanisms; they could induce opportunistic behaviour in terms of not properly protecting proprietary IP rights.

[12] The iFund (CVC unit of Apple) serves as an interesting example in this context, as Apple set up an external fund to finance game developers to build a critical mass of applications for the AppStore, which had a positive impact on the parent company, Apple (Lerner, 2013).

from the parent company (Lee et al., 2018) (one can also imagine that business unit managers will be more reluctant to coordinate, as the benefits for their business unit may be less clear). Logically, a company with a high value of innovations will, on average, have more innovation-dependent resources to leverage. Since external units are structurally disconnected from these resources, they will have more difficulty accessing and leveraging them than internal units. This indicates that internal units will realize more benefits from the *parenting advantage*. This indicates that high value of innovations is negatively related to setting up a CVC unit externally. However, as Garrett and Neubaum (2013), point out, sometimes the parenting advantages are “hard to harvest” and, in some cases, can be detrimental, which could dampen this effect.

Thirdly, a *decentralization effect* might be in play. Prior research has shown that decentralized (centralization is here proxied by share of patent assignment to parent or affiliate) firms are, on average, better suited for deriving value from (knowledge-related) acquisitions since (i) decentralized firms can better manage and transfer external patents when necessary[13] (as the quote in the footnote points out, usage of patents by affiliates happens through assignment of patents, which induces autonomy) and (ii) integration of acquisition is costly, but becomes unnecessary for decentralized firms, as they are better able to accommodate targets as autonomous units (Arora, Belenzon, & Rios, 2014).

While Arora et al.'s (2014) paper focuses on a different definition of decentralization and a different context (i.e. mainly leveraging external patents), the above concepts do yield important findings for this thesis. As documented by Arora et al. (2014), patent assignment to affiliates creates a higher degree of autonomy in usage of patents. Logically, this implies that an external unit as per the definition applied in this thesis (i.e. an *affiliate*) will enjoy a higher degree of autonomy in usage of patents (if they are assigned). This, in turn, will make patent usage, i.e. making patents available for the target company, more efficient for external units. More simply; it will be easier to leverage the patents of the parent company (given assignment) because the parent company to a higher degree will be “hands-off” (and in consequence, leveraging external knowledge together with the target).

[13] “… assignment of patent rights may be associated with a credible delegation of informal authority, since assignment allows the affiliate to directly contract with outside licensees, without formally requiring headquarters to sign off on deals. More simply, assignment may reflect a broader hands-off orientation. We are agnostic as to which mechanism might be at play, since all evidence points to assignment as associated with increase autonomy” (Arora, Belenzon, & Rios, 2014, p. 321).

The assumption that makes this effect relevant for the value of innovations is the following: With more valuable innovations, the need to manage and transfer patents efficiently becomes more important. The above argument is quite speculative, and the parenting advantage might offset this to some degree (perhaps patents will not be assigned for usage), but nevertheless a nuance that is worth mentioning.

Consequently, this *decentralization effect* is expected to have the opposite effect of the two prior-mentioned effects; with a higher patent value, this effect suggests a positive relationship with the likelihood to set up an external unit.

Summarizing, we theorize that the decision to engage in CVC via an internal or external unit is affected by the value of innovation. Specifically, we link them via three theoretical effects; a *protection effect* and a *parenting advantage*, which both seem to suggest a negative relation between value of innovations and the likelihood to set up an external unit, and a *decentralization effect*, which suggests the opposite.

4.2. Firm specificity

As shown, no papers have, to the best of our knowledge, investigated innovation-related *firm specificity* as an antecedent to the decision to put a CVC unit externally or internally, respectively. Firm specificity, in this context, refers to the degree to which a company's innovation process is specific to that firm. Firm specificity is expected to exert an influence on the likelihood of setting up the CVC unit externally.

Specifically, we theorize that two effects play a role in this relationship: (i) the *Not-Invented-Here syndrome* and (ii) a *difficulty to absorb.*

Firstly, the Not-Invented-Here syndrome (NIH) is a term used to describe a bias towards valuing internally developed technology higher than externally developed technology and, consequently, investing less in external technology, as described by several papers (Katz & Allen, 1982; Arora & Gambardella, 2010). According to Arora and Gambardella (2010), the causes for the NIH syndrome have not been thoroughly investigated and documented. However, they suggest the following causes: External technologies may put existing synergies (e.g. communication between in-house departments) at risk and NIH work as a "commitment device" to overcome incomplete contracting (i.e. creating effective and credible incentives for internal development)[14]. Furthermore, firms may wish to reward their own inventors and engineers for innovating (Arora & Gambardella, 2010; Rotemberg & Saloner, 1994).

It is expected that the NIH syndrome is more present when firm specificity is high, as the need to incentivize internal innovation becomes relatively more important when innovations are specific to the firm – external technology cannot, to the same degree, substitute internal innovation in this case. Since the NIH syndrome creates resistance against external technology, managers are (assuming rationality) expected to be more likely to set up a CVC unit externally than internally. This would allow managers to achieve maximum benefit from the unit without encountering internal resistance.

To nuance this further, we will dive into how the NIH syndrome specifically manifests itself to create such internal resistance. According to Arora and Gambardella (2010), the competition posed by external developments might dissuade employees from being innovative. However, as also pointed out, the effect could potentially be the opposite; external competition might spur a greater internal effort. Another way resistance might manifest itself is through other department managers' political meddling, i.e. leveraging their positions to exercise a degree of control over resources, decision-making etc. in a sub-optimizing manner (Arora & Gambardella, 2010).

This argument is supported by Cirillo, Brusoni and Valentini (2013), who show that spin-outs can act as a "rejuvenation strategy" for companies' innovation efforts, i.e. balance exploration and exploitation. They argue that inventors are socialized when they are a part of an organization, and that this socialization induces inertia. Joining a spinout can break up such inertia, and Cirillo et al. (2013) also theorize that the *external stimuli* induces more explorative behaviour. Specifically, they show that inventors actually become more explorative in the case of a spinout.[15] While the context is slightly different from an external CVC unit, this research supports the above arguments that there may in fact be "resistance" against external ideas, here explained as a socialization effect, which can be applicable in the context of CVC as well. Other scholars find similar effects to

---

[14] Moreover, they suggest that companies with good abilities in evaluating quality of external technology (by their definition, this is a part of absorptive capacity) will display NIH, since they invest in less but more valuable technology.
[15] They test for a range of potential biases, most importantly endogeneity of the decision to join the spinout, and the results are robust.

influence external knowledge (see e.g. Levinthal & March's, 1993, work on "the myopia of learning").

Secondly, a simpler effect might be in play. Per the theory on absorptive capacity, it will be more difficult to leverage or efficiently absorb external innovations when innovations are firm-specific, and thus to a lower degree related to novel external knowledge. This could, in turn, prompt managers to be to set CVC unit externally (again, assuming rationality).

Summarizing, we theorize that the decision to engage in CVC via an internal or external unit is affected by the degree of firm specificity. Specifically, we link them via two effects, which both are theorized to exert a positive influence on the likelihood of setting up a CVC unit externally; a *Not-Invented Here* syndrome (which is strengthened by a "rejuvenation"-like effect) and a simple *difficulty to absorb*.

4.3. <u>Absorptive capacity</u>

No previous papers have linked absorptive capacity to the decision to engage in CVC via an internal or external unit. We theorize that absorptive capacity exerts an influence on the likelihood of setting up an external CVC unit.

Specifically, absorptive capacity is based on a "set of organizational routines and strategic processes" (Zahra & George, 2002, p. 186), and as shown in section 2.5.4, the definition involves leveraging *external* information. We can deduce that firms with a high absorptive capacity have demonstrated that they are able to leverage routines, structure and existing knowledge to develop capabilities from novel, external information. As the ability to leverage external knowledge is dependent on the firm's existing knowledge stock, this means that firms with higher absorptive capacity will be better suited to integrate new ventures into the parent company (to the degree necessary). Consequently, we theorize that companies that display a higher degree of absorptive capacity will have a lower likelihood of setting up an external CVC unit. To some degree, the ability to use existing capabilities might be offset by other factors, such as *internal stickiness*, defined as barriers to internal transfer of knowledge (Szulanski, 1996) and consequently weaken the relationship between absorptive capacity and the likelihood of setting up a CVC unit internally. However, as we are looking for the effect on average, this should not be a theoretical obstacle.

To summarize, we theorize that absorptive capacity has an impact on engaging in CVC via an internal or external unit. Specifically, there is an indication that companies with a higher absorptive capacity on average have a lower likelihood of setting up an external unit.[16]

4.4. Technological diversification

4.4.1. Technological diversification as an influencing variable

The relationship between technological diversification and the set-up of the CVC unit internally or externally remains unexplored by literature to date. In the following, we theorize this relationship while drawing from some of the literature on technological diversification itself (e.g. Ahuja & Lampert, 2001; Breschi, Lissoni, & Malerba, 2003; Patel & Pavitt, 1997). Specifically, technological diversification is theorized to have an influence on the choice of setting up an internal or external CVC unit through (i) path dependency, (ii) search, coordination and bureaucratic costs and (iii) replication.

Firstly, in line with the notion of absorptive capacity (Cohen & Levinthal, 1990), organizational learning often takes places *locally* (Breschi et al., 2003): This implies that firms are bound by existing technologies when searching for new ones, and managers are constrained by current knowledge and technologies with regards to the *direction* of search (Breschi et al., 2003; Patel & Pavitt, 1997). Consequently, technological diversification can be viewed as *path dependent*, and partly determined by knowledge-relatedness (Breschi et al., 2003), imposing limits to the possible exploration of technologies. Furthermore, technological diversification is to a large degree an outcome of combining and re-combining existing technologies, eventually resulting in new inventions: A technologically diversified firm is thus subject to a greater set of opportunities (Granstrand, 1998).

As CVC is often viewed as a mean to access technologies *novel* to the organization, setting up the CVC unit externally can be a tool to reduce the impact of the limitations imposed by path dependency. Hereby, an external CVC unit can increase the *breadth* of search for technological

[16] As will be argued for an shown in the results section, the variable that will proxy absorptive capacity will be omitted due to correlation issues. Therefore, this thesis will not offer any conclusive evidence for absorptive capacity.

innovation to fields outside of the organization’s current technological expertise, exposing the organization to a wider range of opportunities.

This effect is theorized to be especially present in firms with a *low* degree of technological diversification, as there is less breadth of knowledge to be leveraged when searching new technologies. Since engineers might develop a myopic view with regards to technological expansion, a highly specialized organization, only active in few technological fields, is limited in its capability to search for new opportunities, which can be addressed by setting up the CVC unit externally. Highly technologically diversified companies, on the other hand, already possess a wide range of knowledge and thus are internally less constrained to recognize technological opportunities. Consequently, the higher the technological diversification, the more the CVC unit can leverage existing knowledge internally to ensure sufficient exposure to technological opportunities. This effect implies a *negative* relationship between technological diversification and the likelihood of an external CVC unit. It is important to note that this argument is based on the assumption that, at least some, companies with a low technological diversification are actually interested in technologies that lie *outside* the scope of their current technological portfolio (exploration). However, this might not be the case for all organizations; some companies might be doing well exploiting opportunities in their narrow scope (exploitation). Taking this into account, the negative relationship between technological diversification and the likelihood of an external CVC unit could be weakened.

However, there is a second effect, closely related to the breadth of search: An increasing number of technological opportunities requires an increasing amount of time to differentiate and rank those opportunities according to their value. Because of this increased level of complexity, it is potentially more difficult to recognize valuable opportunities. As per TCE, it entails three different types of costs: Firstly, the effort to process and evaluate the available options implies *search costs*. Secondly, since the sources of knowledge about the diverse technologies are spread throughout the firm, information needs to be coordinated and transferred internally, leading to *coordination costs*. Thirdly, organizational interdependencies result in *bureaucratic costs* (Jones & Hill, 1988). With a larger technological diversification, these costs increase. Structurally separating the CVC unit from its parent company reduces the complexity and thus the transaction costs stemming from the need to coordinate (i.e. coordination and bureaucratic costs) and thereby paves the way for a more targeted, focused search. Due to the costs related to the breadth of search, technological diversification is expected be *positively* related to the likelihood of an external CVC unit.

Lastly, literature suggests that diversified firms (in terms of products or industries) are more likely to be organized by units that operate under a high degree of organizational autonomy (Chandler, 1962). Technological diversification is typically higher than product diversification (one product can require multiple technologies) within a firm, but the two concepts are related: As technological diversification often precedes product diversification (Pavitt, 1998), the argument can be transferred in the context of technological diversification. Thus, the *replication* of existing structural characteristics suggests a positive relationship between technological diversification and the likelihood to set up a CVC unit externally.

As shown, technological diversification can be related to the structure of a CVC unit in multiple ways. Path dependency of the parent could indicate a negative relationship of technological diversification and an external unit, while search, coordination and bureaucratic costs as well as replication effects imply the opposite. The overall theorized effect is hence not completely clear.

4.4.2. Technological diversification as a moderating factor

Technological diversification is not only theorized as an influencing factor by itself but further as a moderating factor in the relationship between other theorized effects and the likelihood of setting up a CVC unit externally. Specifically, it is expected to moderate the relationship between the likelihood of setting up an external CVC unit and (i) the value of innovations and (ii) firm specificity. To explain the moderating influence, we relate technological diversification to the specific sub-effects of the two concepts. As the overall effect is determined by the strength of the individual effects, moderating factors can indirectly influence the overall nature of the relationship between value of innovations and firm specificity, respectively, and the likelihood of setting up the CVC unit externally.

Firstly, we theorize the relationship between the value of innovations and the likelihood of setting up an external CVC unit to be moderated by technological diversification. Specifically, we propose that the relationship is *weakened* if technological diversification is high. The reasoning is as follows: The effect of a *parenting advantage* was theorized to be negatively related to the likelihood of setting up an external CVC unit. In the case of a high technological diversification, the parenting advantage is dampened. As the breadth of knowledge increases (i.e. more technologically diverse), TCE suggests that coordination becomes costlier; it simply becomes less easy to find, assess and use the right technology and associated resources when there are more to pick from. The increased complexity makes resource-sharing more difficult, which would be essential to leverage benefits from a parenting advantage in an internal CVC unit. Hence, through dampening the parenting advantage, technological diversification is suggested to moderate the relationship between value of innovations and the likelihood of an external CVC unit.

Secondly, we theorized that the degree of firm specificity is related to the likelihood of setting up a CVC unit internally or externally through the presence of the *NIH-syndrome* and *difficulty to absorb*. Both of the effects suggest a positive relationship with the likelihood of an external CVC unit. We theorize that technological diversification could be a moderating factor in this relationship. Specifically, if the firm is highly specialized (technological diversification is low), the effect of firm specificity on the likelihood of an external CVC unit is theorized to be *stronger*. This can be explained by the underlying effects: Firstly, if knowledge is concentrated to few technologies, the NIH-syndrome is expected to be more strongly present, as the bias toward internal innovations is even larger. Secondly, with little diversification and a high degree of firm specificity, absorbing external knowledge might be even more difficult, as the knowledge base to draw from is more narrow: External innovations outside the prevalent technologies are even harder to efficiently leverage. Vice versa, with a high technological diversification, firm specificity might exert a *weaker* influence in to set up CVC unit externally, as technological diversification might mitigate the prevailing sense of myopia.

In summary, technological diversification is expected, next to being a complementary factor, to have a *moderating* influence on both the relationship between value of innovations and firm specificity, respectively, and the choice of an external or internal CVC unit. This is because of two main reasons: Firstly, a high degree of technological diversification can hamper internal resource-sharing and hence affect the influence of the value of innovations on the setup of the CVC unit. Secondly, with regards to firm specificity, a high degree of technological diversification potentially weakens a prevailing sense of myopia, which has an effect on the likelihood of setting up a CVC unit externally.

5. METHODOLOGY

In the following section, we will describe the methodology applied to address our research question. Firstly, a description about the process of data collection will be provided. Secondly, we will explain how concepts from the theoretical analysis were proxied (i.e. what variables are applied) and tested with the collected data, specifically, which variables were chosen for this purpose. Thirdly, we will explain the statistical tools used in the empirical analysis.

5.1. Data collection

To address our research question, we established a database consisting of information about the structure of the CVC unit (i.e. the subsidiary variable), the patents held by its parent organization (patent data) and investments conducted by the corporate investor (investment data). In a last step, we merged the available data into a single database for the purpose of analysis.

5.1.1. Subsidiary data and transcode table

A list of investments from the Thomson One Banker database (Thomson Reuters, 2018) constituted the basis of our data sample. The names of all investors which were classified as “corporate” were extracted, accompanied by information about the corporate investor’s nation and the minimum and maximum year of investment. The raw list amounted to 1,433 distinct corporate investors from 49 different nations, investing in the period between 1985 and 2015. The **first iteration** of manual sample construction included the following steps: eliminating non-strategic investors from the sample, assigning parent company information for each corporate investor and determining the structure of the CVC unit (internal or external) according to the definition used in this paper.

In the first step, by performing a Google search about the companies’ activities, we eliminated non-strategic investors which were nonetheless classified as “corporate” by Thomson One Banker. This enhances the quality of the data, as the nature of their investments differs from strategic corporate investors. Excluding private equity firms, non-corporate venture capital firms and asset management companies eliminated 325 of the 1,433 corporate investors, reducing the sample to 1,108 distinct corporate investors. The fact that 325 investors were eliminated in this step highlights the need to enhance the quality of Thomson One Banker’s data through a manual review of the data.

Secondly, we manually assigned a parent organization to each corporate investor to derive both the organization's name, its industry as defined by the Standard Industry Classification (SIC) code, as well as the legal structure of the CVC unit. We consulted the Compustat database (Standard & Poor's, 2018) for a list of publicly traded companies (global and North America). If the corporate investor from Thomson One Banker could be matched directly with input from Compustat, we were able to copy its name and SIC code directly into our sample. In the case of a non-match, a Google search (later revisited for higher level of reliability, as will be explained) was performed to identify the parent company, which was again copied directly if it could be matched with the list from Compustat. If the parent could not be found in Compustat, such as non-listed or family-owned firms, both name and industry were inserted manually. To later control for potential issues regarding this manual insertion of the parent organization's name and industry (i.e. the data source), it was documented in the dummy variable *organization_source*, where 1 indicates that the information was taken from Compustat and 0 indicates manual insertion.

The third step consisted of collecting information about choice of the CVC unit's setup, which serves as the dependent variable of the analysis. We differentiate between internal and external CVC units, which are defined by the legal structure the unit is set up within. For this purpose, we replicated the approach of Dushnitsky and Shaver (2009) and introduced the variable *subsidiary*. This is a binary variable which was set to 1 if the CVC unit is a separate legal entity, wholly owned by its parent organization; and set to 0 if the CVC unit is not structurally independent from the parent organization. If the corporate investor and parent organization derived in the previous step are identical, *subsidiary* was set to 0 (e.g. Abbott Laboratories). The same holds for operating subsidiaries: As the separate legal entity does not represent a dedicated CVC unit, the investment was classified as internal (e.g. SoftBank China & India Holdings Ltd). Only for dedicated CVC units, *subsidiary* was set to 1 (e.g. AOL Ventures). We classify CVC units with a *subsidiary* value of 1 as an external, and of 0 as internal unit. Furthermore, the investment period was adjusted in cases where the CVC unit became independent from the parent company and operated as a Venture Capital firm afterwards while keeping its name, as was the case in a few incidents.

The first iteration resulted in 1,108 corporate investors, of which 661 were classified as internal (59.66%) and 447 as external (40.34%).

In the **second iteration** of sample construction, we verified the collected information with regards to the *subsidiary* variable, hence cross-checking if the CVC unit is indeed a separate legal entity if it was classified as such in the first, more superficial iteration for all 1,108 CVC units. For this purpose, we introduced three different validation levels captured in the variable *validated*: Level 3 is considered a self-validation, in cases where the corporate investor is the parent organization (they have the same name). If the corporate investor was listed as a subsidiary in an official company filing, such as the Exhibit 21 of the 10-k SEC filing, national equivalents, annual reports or the Orbis database, the validation level was set to 2. If the subsidiary information could be verified by online sources such as company websites or newspaper articles, the validation level was set to 1. A validation level of 0 indicates that the judgement of the first iteration could not be verified by a second source. Practically, this step involved a manual search in the above-mentioned databases and sources to validate each of the 1,108 separate observations. During the validation round, we also identified a small number of investors as non-corporate and eliminated those, which had been kept in the first iteration kept in the sample

The second iteration resulted in a list of 1,089 corporate investors, of which 739 (67.9%) are classified as an internal CVC unit and 350 operated as a separate legal entity, which accounts for 32.1% of all investors. Hence, compared to the first iteration, the share of internal CVC units increased: It becomes evident that many units are actually not legally independent, even if the name would imply so (e.g. Commerce One Ventures), and even if third-party websites state they are. With regards to validation levels, a total of 77.7% are level 3 or 2 validations, whereas only 6.2% could not be formally validated (*validated* = 0).

To later merge this sample to both patent and investment data, we created a transcode table containing the most important information, namely:

- The name of the corporate investor / CVC unit (*firmname*)
- The unique id for the corporate investor / CVC unit (*idinvestor*)
- The nation of the corporate investor / CVC unit (*firmnation*)
- The minimum year of investment (*year_inv_min*) and the maximum year of investment (*year_inv_max*) of the corporate investor / CVC unit

- The name of the parent organization (*organization*)
- The unique ID for each parent organization (*idorganization*)
- The industry classification of the parent organization (*sic_4*)
- The source of the parent organization (*organization_source*)
- The structure of the corporate investor / CVC unit (*subsidiary*).

From this data, *firmname, firmnation, year_inv_min* and *year_inv_max* were sourced from the Thomson One Banker database, *organization* and *sic_4* were searched manually and matched with Compustat if possible, which is indicated in *organization_source*. The variables *idinvestor, idorganization* and *subsidiary* were assigned by the authors.

5.1.2. Patent data

In a next step, we retrieved patent and related information (e.g. citations) of the parent organizations of the CVC units. For this purpose, we used input files from the publicly available PatentsView database (PatentsView, 2017). The PatentsView database is sourced from the United States Patent and Trademark Office (USPTO) and hence includes official information on patents granted or applied for in the US. We retrieved information on granted patents from 1976 to 2017, including, amongst other, data on their current (for patents up to May 2015) technology classification (United States Patent Classification, abbr. USPC).[17] Moreover, citations made to US patents by US patents were also included. As multiple input files from PatentsView had to be linked through common identifiers (i.e. variables that exist in more than one data file, such as patent ID), standard merging was performed in Stata. A list of used files from PatentsView can be found in *Appendix A* and how they were linked is summarized in *Appendix B.*

For each patent, the database contains information on the number of claims, its application and granted dates, the current technology classification (USPC) and the unique assignee ID, which is the unique identifier for the parent organization. Firstly, as we are solely interested in utility patents

[17] Since new USPC classes are introduced when new technologies are developed, there is a need to reclassify older USPC classes to allow for comparability across time (i.e. USPC classes of older patents are reclassified from official side to accommodate comparability with newer USPC classes).

(there are also design patents, for example), only those patents were kept. Secondly, we merged in the citation data by collapsing[18] data on the level of the patent ID. The technical explanations will follow:

- Count of backward and forward citations: Based on citation information of PatentsView, we counted citing (forward citations) as well as cited (backward citations) patents per patent ID. More simply: for each patent, the number of other patents that it cites and the number of other patents that cite it were counted.

- Count of backward and forward self-citations: If the assignee ID of the citing patent equals the assignee ID of the cited patent, it was counted as a self-citation for the cited patent (and vice versa). More simply: For any given patent of an organization, the number of citations to this patent (and from this patent, respectively) made by patents of the organization that created the patent were counted.

In the **third iteration** of the sample construction, the sample resulting from the second iteration was linked to the patent database through the name of the parent organization. In general, the USPTO does not assign a unique organization ID for each individual firm in patent filings: As organizations use different names or abbreviations, and names frequently contain spelling errors in patent filings, it is difficult to retrieve all patent information belonging to a specific firm – a problem widely recognized in patent-related research (e.g. Hall, Jaffe, & Trajtenberg, 2005). To mitigate this problem, PatentsView uses a disambiguation algorithm to assign unique IDs for each organization (the assignee ID, as described above). As our data sample contains organization names only, we needed to link the sample to the PatentsView database via the organizations’ names in order to derive the unique assignee ID used by PatentsView. This would then enable us to retrieve an exhaustive list of patents assigned to the organization, including those where the assignee’s name was spelled differently.

As there is no unambiguous common identifier between the two datasets resulting from the potential spelling differences of organization names, standard merging using *merge* in Stata is impossible. Therefore, a probabilistic record linkage as performed by the Stata command *reclink2*

[18] The database with regards to citations shows the cited patent, the citing patent, and a citation date. The Stata command *collapse* allows us to count citations, by reporting frequencies of observations per patent ID. This was employed both for cited and citing patents.

was employed (Wasi & Flaaen, 2015). Generally, and thus in probabilistic linkage as well, Stata matches pairs correctly only if the formatting in both datasets is consistent (i.e. the names must resemble each other to some degree): In a first step, we hence capitalized all organization names in our sample as well as the established patent database.

In a second step, the Stata command *reclink2* was used to derive the best matches for each parent company (*organization*) of the 1,089 different CVC units. The reclink2 command computes a number from zero to one based on the degree of similarity between the two values. However, the highest scored name is not always the correct name. To capture correct matches which do not yield the highest score whilst keeping the number of incorrect matches to a minimum, we set the number of matches to three as recommended by Wasi and Flaaen (2015). This means that for each name, a list of three possible “real matches” were presented, based on their degree of similarity with each other.

In a third step, a clerical review of the reported matches was performed. This manual review was used to address and correct four identified issues:

1) Pair-similarity employed by *reclink2* is an imperfect metric, as the highest score does not necessarily equal the correct match (Wasi & Flaaen, 2015). Out of the three matches, we manually chose and retained the match that indeed equalled the parent organization. This included, for example, cases in which the organization name contains a common ending such as HOLDING. For instance, "AB Holding" is more likely to be matched with "XY Holding" than "Alpha Beta", which could be the real parent organization.

2) The PatentsView disambiguation algorithm did not capture all versions of the organization names’ spelling. In case of name ambiguity, i.e. multiple assignee IDs per organization, all matches were kept.

3) We corrected the merged, acquired or renamed companies and adjusted the name to the parent organization within the investment period. Multiple lines of observations where created in case of overlaps.

4) In a few cases, the same assignee ID was used for different organizations (due to minor flaws in the disambiguation algorithm). To eliminate this error source, these were removed from the sample.

After linking both datasets, we replaced the assignee ID by PatentsView with the unique organization ID in our sample (*idorganization*) to clearly identify different organizations in the dataset, including those with multiple assignee IDs. On an organizational level, we created the following variables with regards to granted patents, both as of application and granted date (marked in variable names through the extension suffixes *_app* resp. *_g*):

- Cumulative number of granted patents in any given year (*cum_patents*)[19]

- Cumulative sum of forward citations (*cum_fc*), forward self-citations (*cum_fsc*), backward citations (*cum_bc*) and backward self-citations (*cum_bsc*) of granted patents in any given year

- Number of distinct USPC classes of patents in any given year (*cum_dist_upsc*)

- The standard deviation of the dispersion of patents on different USPC classes up to any given year (*sd_tot_uspc*). This measure hence takes patent dispersion (i.e. how many patents were filed in each USPC class) into account. This is calculated by counting the total number of patents per distinct USPC main class at any given year and then calculating the standard deviation from the mean of that patent count.[20]

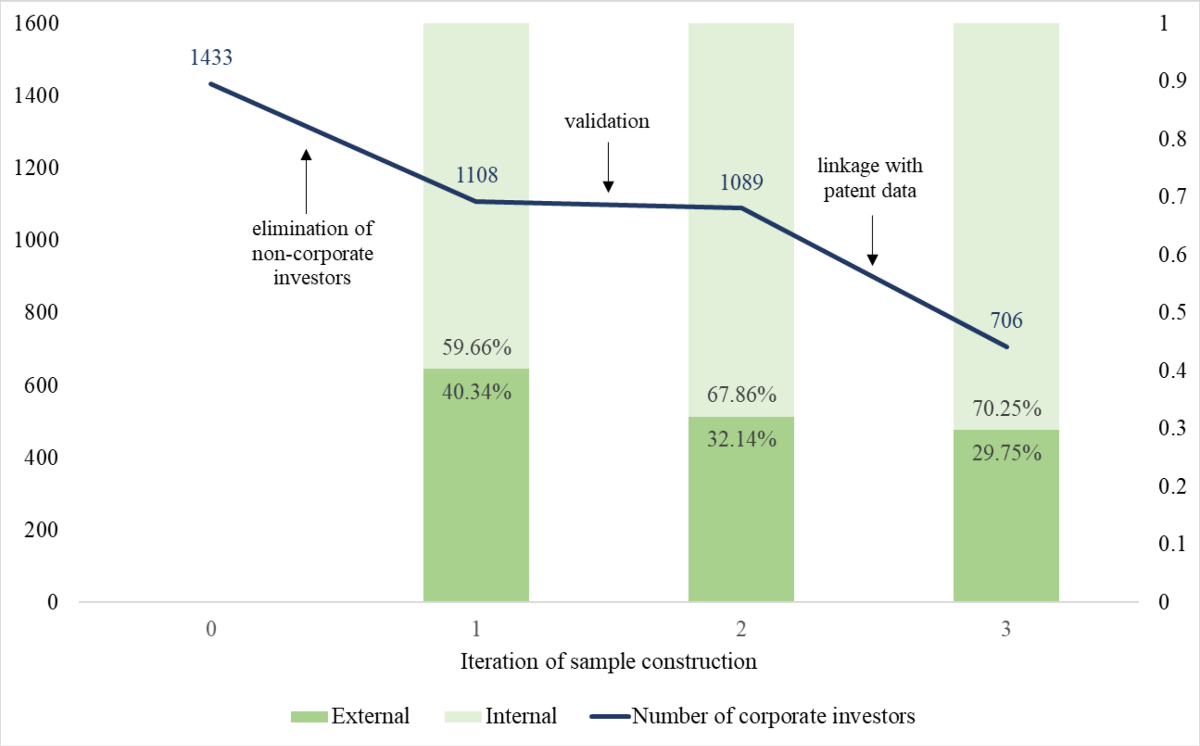
This third iteration reduced our sample by the CVC units with parent organizations to which we could not assign patent information and resulted in our final sample. In total, the parent organizations of 706 corporate investors could be matched with the patent database, out of which 34 observations (4.82%) were identified manually (i.e. in the case of merged, acquired or renamed companies, and no correct reclink matches at all). Out of the 706 investors, 496 (70.25%) are

[19] One might argue that this is a total number up until any given year. However, we use the word cumulative (which is not incorrect) to enhance the understanding of the difference between this variable and the variable used in the analysis later.
[20] Suppose an organization has ($x_1, x_2, \ldots, x_n$) assigned patents in N distinct main USPC classes at a certain point in time, resulting in a mean number of patents per USPC class $\bar{x}$. Then, the *sd_tot_upsc* was calculated as $s = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}}$.

internal CVC units and 210 (29.25%) operate as a subsidiary. With regards to validation levels, only 37 observations (5.24%) could not be second-source validated (level 0). A summary of the three iterations to deduce the final data sample can be seen in *Figure 3*.

Evidently, the share of internal units rose for each iterative step. This can be explained largely by the following aspect: Many names suggested external units, which, upon validation, proved to be wrong (i.e. companies choose "externally-sounding" names even though the unit is, in fact, internal).

Figure 3: Summary of data sample iterations



5.1.3. Investment data

Data on investments of the CVC units was retrieved from the Thomson One Banker database (Thomson Reuters, 2018). The data comprised, on a target company level, information on each round that a corporate investor participated in, as well as information on the other rounds for the same target company. Firstly, we adjusted the definition of "corporate" by Thomson One Banker based on our findings in the first iteration of the sample construction. Secondly, we eliminated investments in the sample that were conducted by the CVC unit outside the manually adjusted investment period (for example, if the CVC unit became independent). For each investment of the corporate investor, information about the date of the investment, the estimated equity investment of the investor and each of the co-investors as well as the number of co-investors and corporate co-investors was kept. On the level of the target company, we kept the age of the target company at the time of the investment, its nation, and number of total (corporate) investors as well as estimated equity investments over all investments rounds.

As the *subsidiary* variable remains unchanged throughout time for each corporate investor, we reshape our data to an investor level to obtain cross-sectional data, resulting in the following investment-related information over the period of investment per CVC unit:

- The maximum year of investment (*year_inv_max*)
- The total number of investments (*num_investments_tot*)
- The total estimated equity investment in USD million (*equity_est_firmname_tot*)
- The average age of the target company at the time of the investment (*comp_age_avg_mean*)
- The mean number of co-investors per round (*num_coinvestors_round_mean*)
- The mean number of corporate co-investors per round (*num_corpinv_round_mean*)
- The proportion of investments in which the target company operates in the same industry, based on the comparison of SIC codes, a number between 0 and 1, whereas 1 indicates that the SIC code of corporate investor and target are equal (*same_sic_proportion_mean*)
- The proportion of investments in which the target company operates in the same nation, a number between 0 and 1, whereas 1 indicates that the nation of corporate investor and target are equal (*same_nation_proportion_mean*).

5.1.4. Final dataset for analysis

As a basis for the analysis and further variable construction, we connected the databases with the Stata *merge* command using the unique investor ID respectively organization ID as the common identifier. The patent data was retrieved as of the maximum year of investment. A summary of variables (as previously defined) and the merging logic is portrayed in *Figure 4*.

Figure 4: Merging the datasets through the transcode table

Investment Data	Transcode Table	Patent Data
idinvestor ↔	↔ *idorganization* *year_inv_max*	
firmname	*firmname*	*idorganization*
idinvestor	*idinvestor*	*year_inv_max*
year_inv_max	*firmnation*	*cum_patents*
num_investments_tot	*year_inv_min*	*cum_fc*
equity_est_firmname_tot	*year_inv_max*	*cum_fsc*
comp_age_avg_mean	*organization*	*cum_bc*
num_coinvestors_round_mean	*idorganization*	*cum_bsc*
num_corpinv_round_mean	*sic_4*	*cum_distinct_upsc*
same_sic_proportion_mean	*organization_source*	*sd_tot_uspc*
same_nation_proportion_mean	*subsidiary*	

5.2. <u>Variables</u>

In the following section, we will explain which variables we use for our analysis. The variables below are deemed to be good measures or proxies of the theoretical concepts. We will argue for the aptness of the variables employed by bridging them with the theory, which was set out in section 4. The variables are summarized in *Table 2* (inspiration drawn from the configuration of Quintana-García & Benavides-Velasco, 2008).

5.2.1. Dependent variable

<u>Internal vs. external CVC unit</u>

Our dependent variable is a dummy variable, where 0 signifies an internal CVC unit and 1 signifies an external CVC unit (i.e. a wholly-owned subsidiary, as previously described). This way of operationalizing this specific variable has first been introduced as a control variable by Dushnitsky and Shaver (2009) and was employed by recent studies in the field (e.g. Lee et al., 2018).

The variable is called *subsidiary* in the analysis.

5.2.2. Independent variables

Value of innovations

To measure the value of innovations, we use the cumulative number of forward citations assigned to the patents of the parent company at the maximum year of investment. Forward citations is an often-used proxy for value of innovations since the number of times other patents cite a patent is a good observable and quantifiable metric for how important the patent is to firms, and hence, how valuable the patent is (Hall et al., 2005). Specifically, we employ the total number of forward citations assigned to the entire granted patent stock of the parent company up until the end of the investment period. We use the granted patents (instead of applied-for patents) since the measure is *forward-looking*, and since there is no certainty that applied-for patents will be granted (we have that information as the database only includes applied-for patents that are later granted).

The variable is called *cum_fc_g* in the analysis.

Firm specificity

To measure firm specificity, we employ the share of backward self-citations in total backward citations of the patents of the parent company at the maximum year of investment. Self-citations reflect "the cumulative nature of innovation", i.e. how much a company relies on its own prior innovations when developing new innovations (Hall et al., 2005, p. 32). According to Hall et al. (2005), self-citations offer an insight into how much of the knowledge spillovers the firm can internalize (rather than spilling over to other companies), i.e. how specific the patents are to that company. Specifically, we employ a share measure (as also seen in e.g. Hall et al., 2005), which is the total number of backward self-citations divided by the total number of backward citations. As Hall et al. (2005) also point out, it is necessary to control for the size of the patent portfolio when using this variable, as the share of self-citations might increase when the portfolio is larger. The reason is that the larger the portfolio, the larger the chance that the company will cite one of its own patents (as there are more patents to cite from, i.e. a "mechanical" effect) but not necessarily because they are firm-specific. Still, the share of backwards self-citations is deemed a meaningful measure of firm specificity. For the self-citations, we have used applied-for patent data, as the variable is backward-looking, and thus the applied-for patents (which were at a later point granted) offer the best possible representation of the available data. The applied-for patents will also be used for the other patent-related measures (except forward citations, as discussed above) for the same reason.

This variable is called *share_bsc_app_cum* in the analysis.

Absorptive capacity

To measure absorptive capacity, we employ the company's patent stock. According to Hall, Jaffe and Trajtenberg (2001), patents proxy knowledge capital and the success of the innovation process and, consequently, the company's ability to absorb new knowledge. Patent stock is widely used as a proxy for absorptive capacity (Cohen & Levinthal, 1990; Dushnitsky & Lenox, 2006; Hall et al., 2001), which supports its aptness. Specifically, we use the cumulative number of applied-for patents (that are later granted) in the maximum year of investment.

This variable is called *cum_patents_app* in the analysis.

Technological diversifications

To measure technological diversification, we use the standard deviation of the cumulative number of patents per USPC main class of patents assigned to a given company at the end of the investment period. To put it more plainly, this measure displays the dispersion of patents in different technology classes. Some other papers use a simply count of technology classes (see e.g. Breschi et al., 2003). However, this simplified approach does not take the magnitude of patents in each technology class into account, which is why the approach undertaken here adds nuance: it gives a measure of the magnitude and spread in different technology classes of the patents of a given company. The cumulative number of distinct USPC classes of applied-for (and later granted) patents (*cum_dist_uspc_app*) in the maximum year of investment is instead employed as a control variable.

This variable is called *sd_tot_uspc_app* in the analysis.

5.2.3. Control variables

We have employed a number of control variables in order to account for other factors influencing our dependent variable. Specifically, we have employed the following variables: a) the cumulative number of distinct USPC classes in the maximum year of investment as mentioned above, b) the total number of investments performed by the CVC unit, c) an equity estimate of the investments performed by the CVC unit, d) a proportion of the target investment companies that have the same industry (SIC) code as the parent company, e) a proportion of the target investment companies that are from the same nation as the parent company, f) the average age of the target investment companies, g) the average number of co-investors in any given round of investment, h) the average number of corporate co-investors (other CVC investors) per investment round.

Table 2: Summary of variables

Variables	Type	Measurement method	Source of the data
subsidiary	Dependent	Dummy variable, 0 signifies internal, 1 signifies external	Manual search by authors
cum_fc_g	Independent	Cumulative number of forward citations of the company's granted patents in the maximum year of investment	PatentsView
share_bsc_app_cum	Independent	The share of backward self-citations to total number of backward citations of the company's applied-for (later granted) patents in the maximum year of investment	PatentsView
cum_patents_app	Independent	The cumulative number of applied-for (later granted) patents of the company in the maximum year of investment	PatentsView
sd_tot_uspc_app	Independent	The standard deviation of the dispersion of all applied-for (later granted) patents in cumulative distinct USPC main classes of the company in the maximum year of investment	PatentsView
cum_dist_uspc_app	Control	The cumulative number of distinct USPC classes of total applied-for (later granted) patents in the maximum year of investment	PatentsView
num_investments_tot	Control	The total number of investments performed by the CVC unit	Thomson One
equity_est_firmname_tot	Control	The estimated total equity investment of the CVC unit in USD million	Thomson One
same_sic_proportion_mean	Control	The proportion of the target investment companies with the same SIC code as the parent company	Thomson One
same_nation_proportion_mean	Control	The proportion of the target investment companies that are from the same nation as the parent company	Thomson One
comp_age_avg_mean	Control	The average age of target investment companies	Thomson One
num_coinvestors_round_mean	Control	The average number of co-investors per round of investment, in which the unit participates	Thomson One
num_corpinv_round_mean	Control	The average number of corporate co-investors per round of investment, in which the unit participates	Thomson One

5.3. Statistical tools

This section will explain the econometric tools, which are used in this thesis to address our research question. Firstly, we will explain the regression model applied, its interpretation and some of its underlying assumptions. Secondly, interaction terms in non-linear models will be addressed, as these will be applied as well. Thirdly, we will critically examine if the used measures are internally and externally valid, and hence reliable.

5.3.1. Non-linear regression model: probit

This thesis employs the non-linear probit regression model, as it, as will be argued, is deemed most suitable in this context. However, the most common and widespread regression model is the linear (Ordinary Least Squares, OLS) model. As will be explained, this is not apt for the purposes of this thesis. However, as its properties are well-known, we will illustrate the use of the probit model by comparing it to the OLS model.

As described, our dependent variable is the dichotomous variable *subsidiary*, restricted to the possible values zero and one. The linear (OLS) regression model, on the other hand, assumes a continuous dependent variable (Aldrich & Nelson, 1984). For this reason, instead of the predicting the binary value, the *probability* that the response variable is equal to one is modelled, allowing for any value between zero and one, given the different factor variables, mathematically denoted $\Pr(Y = 1|X_1, X_2, \dots, X_k)$. If Y denotes the dependent variable in the model, $X_1, X_2, \dots, X_k$ denote the independent variables, and $\beta_1, \beta_2, \dots, \beta_k$ their respective coefficients, probabilities can be estimated using a linear probability model (Stock & Watson, 2015):

$$(1) \quad \Pr(Y = 1|X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

The linear probability model has multiple flaws. Firstly, an OLS regression estimates sampling variances incorrectly (not the smallest variance possible) for dichotomous dependent variables (Aldrich & Nelson, 1984). Secondly, a linear probability regression model assumes constant effects of given changes factor variables on the response variable (the slope is linear), allowing the probability to exceed the value of one and fall below zero (Stock & Watson, 2015). However, probabilities are constrained to values between zero and one, hence changes of independent variables must have *non-linear* effects (Stock & Watson, 2015). This is shown graphically based on an example by Stock and Watson (2015) in *Appendix C*. This is why we employ the non-linear probit regression model in this context.[21]

The probit regression model uses a standard normal cumulative probability distribution function, and can be expressed as (Stock & Watson, 2015):

$$(2) \quad \Pr(Y = 1|X_1, X_2, \ldots, X_k) = \phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

The probability $\Pr(Y = 1|X_1, X_2, \ldots, X_k)$ for any given input can be computed by calculating the value of the term $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$ and employing it as the "z-value" in the standard normal distribution table, to look up the respective probability (Stock & Watson, 2015). To operationalize the probit regression, we make use of the Stata command *probit*.

A unit change of any regressor is associated with the change of the z-value while holding the remaining regressors constant (Stock & Watson, 2015). In this thesis, in line with common practice, the coefficients are estimated using maximum likelihood estimation (Stock & Watson, 2015).[22] A maximum likelihood estimator chooses the values of the coefficients so that they maximize the likelihood to obtain the observed data (Stock & Watson, 2015), which is estimated through an iterative process (Cameron & Trivedi, 2010).

The interpretation of coefficients in the non-linear probit regression differs from a linear regression: The coefficient of a given variable is conditional on the other input variables in the model (Norton, Wang, & Ai, 2004). The effect of a unit change of a given factor variable it is not constant but instead dependent on the starting value of the factor variable: the slope of the curve differs along the curve (Stock & Watson, 2015). To illustrate; a change from $X_1 = 0.1$ to $X_1 = 0.2$ ($\Delta X_1 = 0.1$) might have a different effect on the probability that Y equals 1 than a change from $X_1 = 0.5$ to $X_1 = 0.6$ ($\Delta X_1 = 0.1$). This makes the coefficients more difficult to interpret. However, the statistical significance level of the coefficients from the regression is clear in interpretation; and the same holds for the sign of the coefficient; a positive coefficient implies a positive effect of the

[21] The logit model, which applies a standard logistic instead of the standard normal cumulative distribution, poses an alternative model in this context (Stock & Watson, 2015). Both models often lead to similar results; to check for errors in model specification, we will later perform a logit regression as well.

[22] Alternatively, it could be estimated using Berkson's minimum chi-square method (Berkson, 1980) or Gibbs sampling (Albert & Chib, 1993). Going into detail of the statistical estimation methods lies outside of the scope of this thesis.

respective independent variable on the likelihood of the dependent variable taking the value one and vice versa (Stock & Watson, 2015).

With regards to statistical inference of the overall model, Stata reports the so-called Wald statistic for probit models with robust errors.[23] The Wald statistic tests that at least one of the coefficients is not equal to zero (Greene, 2014). The reported probability > chi^2 is consequently the likelihood of obtaining the observed Wald statistic if the Null Hypothesis holds, i.e. all regressors are simultaneously equal to zero.

To determine goodness-of-fit for the model, a linear regression uses the R^2 measure, a value between zero and one, defined as the (by the model) explained sum of squares divided by the total sum of squares (Berenson, Levine, & Krehbiel, 2012). In non-linear models, R^2 can decrease when regressors are added and fall below zero or above one (Cameron & Windmeijer, 1997). It is not considered a good measure of the goodness-of-fit in non-linear models because estimated probabilities by the model cannot be compared to "true" probabilities, since the latter is unknown (Windmeijer, 1995). Therefore, the probit regression employs "Pseudo-R^2" as a measure of fit instead, which compares values of the maximized likelihood function including all independent variables to the value of the likelihood function without any regressors, i.e. the intercept only (Stock & Watson, 2015). Formally, McFadden Pseudo-R^2 is hence defined as (Cameron & Trivedi, 2010):[24]

$$(3) \quad \hat{R}^2 = 1 - \frac{L_N(\hat{\beta})}{L_N(\bar{y})}$$

Hereby, $L_N(\hat{\beta})$ denotes the log-likelihood value of the fitted model, and $L_N(\bar{y})$ the log-likelihood value of the model with intercepts only (Cameron & Trivedi, 2010). The Pseudo-R^2 tends to be considerably below the value of the standard R^2 measure, thus goodness of fit should not be

[23] Alternative tests, specifically the Likelihood-Ratio test, is less appropriate in case of robust standard errors (Cameron & Trivedi, 2010) and hence lies outside the scope of this thesis. Furthermore, it would require estimating both restricted and unrestricted parameters, whereas the Wald statistic only requires the latter (Greene, 2014).

[24] As Stata employs McFadden's Pseudo-R^2, alternative Pseudo-R^2 measures will not be discussed in this thesis.

evaluated by the same standard. Instead, Pseudo-R^2 values between 0.2 and 0.4 are considered an "excellent fit" (McFadden, 1977, p. 35).

5.3.2. Interaction terms in probit

Interaction terms are often computed to "infer how the effect of one independent variable on the dependent variable depends on the magnitude of another independent variable" (Ai & Norton, 2003, p. 123). In linear models, interactions can be easily interpreted. This can be illustrated by taking the expected value of y (which equals $\Pr(Y = 1|X_1, X_2, \dots, X_k)$ for binary dependent variables), conditional of the independent variables (equation 4), and computing its marginal effects (equation 5) (Ai & Norton, 2003; Norton et al., 2004):

$$(4) \quad \mathrm{E}[Y|X_1, X_2, \dots, X_k] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \cdots + \beta_k X_k$$

$$(5) \quad \frac{\partial^2\, \mathrm{E}[Y|X_1, X_2, \dots, X_k]}{\partial X_1 \partial X_2} = \beta_{12}$$

Consequently, if X_1 and X_2 are independent from the remaining regressors in the model, then the marginal effect of the interaction is β_{12} and a single t test can test its statistical significance (Norton et al., 2004). However, in non-linear models, X_1 and X_2 are not independent from the remaining regressors in the model (Norton et al., 2004):

$$(6) \quad \mathrm{E}[Y|X_1, X_2, \dots, X_k] = \phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \cdots + \beta_k X_k) = \phi(\cdot)$$

Most applied economists incorrectly define the interaction effect as the marginal effect of just the interaction itself, i.e. $\frac{\partial^2\, \mathrm{E}[Y|X_1, X_2, \dots, X_k]}{\partial X_1 X_2}$ (Ai & Norton, 2003). The full interaction effect, however, is instead defined as (Norton et al., 2004):

$$(7) \quad \frac{\partial^2\, \mathrm{E}[Y|X_1, X_2, \dots, X_k]}{\partial X_1 \partial X_2} = \frac{\partial^2 \phi(\cdot)}{\partial X_1 \partial X_2} = \beta_{12}\phi'(\cdot) + (\beta_1 + \beta_{12} x_1)\phi''(\cdot)$$

This has four important implications for the interpretation of interaction terms in non-linear models as derived by (Norton et al., 2004):

1) Even if the reported coefficient β_{12} is zero, the interaction could be non-zero.

2) The statistical significance has to be computed for the entire cross derivative, resulting in more than one z-score. Hence, the reported z-statistics is not necessarily valid for the entire curve and needs to be investigated further.

3) The interaction effect depends on the remaining factor variables in the model.

4) As the interaction effect is comprised of two terms which are added, the sign might differ along the curve dependent on the values of the independent variables. Consequently, the sign of the interaction effect might not equal the reported sign of the interaction coefficient.

Based on this explanation, this thesis will apply the *inteff* Stata command to graphically investigate interactions along the curve, proposed by (Norton et al., 2004). This methodology has been applied by recent studies (e.g. Di Lorenzo & Almeida, 2017).

5.3.3. Reliability

To assess the reliability of our regression model, we follow the proposed framework of Stock and Watson (2015) and examine the concept of validity, specifically internal and external validity. Internal validity is established if the statistical inferences of the model hold for the population or setting of focus, and external validity is concerned with the generalisability of results to other populations or settings. The sample studied in this thesis is essentially non-random (solely consisting of CVC units in specific industries, as will be explained in section 6.1). Essentially, it is a sub-group of a population (all CVC investors) – as a consequence, statistical inferences in the analysis are made to this specific sub-group only (which will be referred to as the population studied).

<u>External validity</u>

External validity holds if results can be generalisable to *other* populations. In the following, primary threats to external validity will be discussed in the context of this study.

a) *Differences in populations* can imply that the relationship or causal effects in the studied population is not the same in different populations (Stock & Watson, 2015). The population studied in this thesis constitutes CVC units in three industries based in the United States. A test will be performed to check if there are significant differences in between the industries of focus, specifically by adding industry dummies into the model. This can be seen as a tool to check for external validity: if results do not differ in between industries, they may perhaps be transferred to US-American CVC investors in other industries as well, which constitute a different population. We suggest that the results can be inferred to other US industries. However, we suggest that this is explored more in-depth in further studies.

b) Identical populations can still differ in *setting*, i.e. the environment (institutional, legal, social, economic) that the population is embedded in (Stock & Watson, 2015). In this thesis, CVC investors based in the United States are studied. As the descriptive statistics part in section 6.1 will show, there are significant differences across country lines, suggesting that the findings are not externally valid. This may be caused by a range of factors, e.g. different investment environments, legal systems etc. Therefore, external validity is not given in this context.

In summary, we cast doubt with regards to the external validity of our study. While the results may be externally valid for the entire US CVC population, the findings are likely not generalizable to populations outside the US. Further research with regards to the antecedents to the choice of setting up a CVC unit internally or externally, especially studying non-US investors, is necessary in order to deem results externally valid to other settings.

<u>Internal validity</u>

Internal validity can be decomposed into two factors: Firstly, the estimators of causal effects must be consistent and unbiased; and secondly, hypothesis tests and confidence intervals should have the required significance resp. confidence level (Stock & Watson, 2015). We will discuss primary threats to internal validity in separate below.

a) An *Omitted Variable Bias* (OBV) occurs if a variable, which is correlated with the remaining regressors and a determinant of the response variable, is not included in the regression model (Stock & Watson, 2015). This can consequently bias the results. This phenomenon is also called *under-specification*. We address this bias by adding additional, adequate regressors in form of control variables.

b) If the regression function differs from the true underlying population regression function, the *functional form is mis-specified* (Stock & Watson, 2015). We apply a probit model to derive results. To check for potential errors of misspecification, we will run a robustness check with both a logit and a linear regression model, in order to investigate if results differ substantially.

c) *Measurement errors* and *errors in variables* can introduce bias (Stock & Watson, 2015). With regards to measurement error in our regressors, we rely on the databases applied (Thomson One Banker and PatentsView). We checked for measurement errors in our dependent variable, *subsidiary*, by having introduced validation levels in multiple iterations of sample construction. As shown, most of our data stems from official records or similar, and we did not observe that CVC units change their legal status throughout time. However, due to limited access to data and the limited scope of this thesis, we cannot exclude the possibility of a change in legal status of the CVC unit for every year of investment, hence cannot completely rule out measurement errors in this regard. To illustrate, if the minimum and maximum year of the investment is 1990 and 2014, respectively, for the same CVC unit we cannot validate the legal structure of *subsidiary* in all of those 24 years. Further, we observe that corporations often open new CVC units with a different name in case of a different legal structure.[25]. Moreover, with regards to errors in variables, we correct the minor errors stemming from the disambiguation algorithm from the PatentsView databases, as previously explained.

d) A potential threat related to the sample selection is the issue of *missing data*. We treat the issue of missing data regarding the regressors (namely, patent data) by reducing the sample size as proposed by Stock and Watson (2015) to not introduce bias. If the process of sample selection is dependent on the availability of the data and is related to both dependent variable and regressors, there is a threat of a *sample selection bias* (Stock & Watson, 2015). In this study, we established a non-random sample of CVC investors, hence are concerned

[25] This is for example the case for the corporation Glaxosmithkline, which has multiple CVC units of different legal statuses: “GSK Ventures” is an internal unit and “Action Potential Venture Capital, Ltd.” is set-up externally.

with corporations which engage in CVC already. However, this does not constitute a selection bias, as we do not make inferences about other populations than the one of focus.

e) If there is not only a causal relationship from the independent variables to the dependent variables (from X to Y), but also from the dependent variable to the independent variables (from Y to X), *simultaneous causality* occurs (Stock & Watson, 2015). As laid out in section 3, this thesis is meant to open a debate about potential causal relationships and does not make causal claims by itself, as it is not an exhaustive study of possible antecedents to setting up an internal or external CVC unit. Thus, we are merely suggesting causal links, and do not live up to all criteria for conclusively drawing causal relationships.

f) All model parameters have to be *identifiable* (Greene, 2014). This means that there is no non-zero parameter unequal to the estimator applied which leads to the exact same result (Greene, 2014). Simply put, no other variable would have the exact same impact in the model. This comprises the assumption used in linear models of absence of perfect multicollinearity, which arises when one of the independent variables is a perfect linear combination of other factor variables (Stock & Watson, 2015). Furthermore, imperfect multicollinearity, even though it does not prevent an estimation of the regression, could lead to imprecise estimation of the coefficients (Stock & Watson, 2015). We ensure internal validity in this regards by checking for correlations between independent variables and eliminating or modifying highly correlated variables to reduce estimation errors.

g) Non-linear regressions assume homoscedasticity and non-autocorrelation (Greene, 2014). These principles are concerned with the error term μ_i. As opposed to heteroscedasticity, *homoscedasticity* implies a *constant* variance of the conditional distribution of μ_i for $i=1,\ldots,n$ and independence from X_i (Stock & Watson, 2015). In economic theory, error terms are rarely homoscedastic (Stock & Watson, 2015). To make our model robust in this regard, heteroscedasticity-robust standard errors are applied, which make statistical inference valid for both cases, as homoscedasticity is a special, narrow case of heteroscedasticity (Stock & Watson, 2015). *Autocorrelation* in regression analysis refers to the independence of error terms. As we did not use simple random sampling, but included all CVC investors of the three chosen industries in the sample, error terms could potentially be correlated (Stock & Watson, 2015). However, as we use cross sectional data, and autocorrelation is mostly regarded as problematic when using time series data (Berenson et al., 2012), the possibility of autocorrelation is rather small. Autocorrelation is consequently not deemed to pose a major threat to internal validity.

In summary, we deem our study internally valid because we take into account potential issues regarding the OBV through the introduction of control variables, check for errors that could result from a model mis-specification, deem the parameters as identifiable and apply heteroscedasticity-robust standard errors. However, measurement errors cannot be completely ruled out, and this thesis does not make causal claims.

6. RESULTS

6.1. Descriptive statistics

6.1.1. Subsidiary data

In this section, our data on CVC units will be described. As we shall see, the industries with SIC codes 737, 283 and 367 are amongst the most prevalent with regards to CVC. As previously set out, these are the main focus of our investigation, and shall therefore also receive special attention in this section. The argument for focusing on these industries will be set out in this section. However, we will start out more broadly by investigating our sample of 706 CVC units. Specifically, the following will be analysed: distribution of internal and external CVC units, industry distribution of CVC units, geographic distribution of CVC units and a number of matrices that combine these variables. Furthermore, patent data for the organizations which own the CVC units will be examined in a similar manner: mean number of patents, mean number of average patent citations per patent, share of self-citations (backward and forward). Lastly, the industries with SIC codes 737, 283 and 367 will be investigated separately.

For the sake of clarity: for the variable *subsidiary*, 0 denotes an internal unit and 1 denotes an external unit (separate legal entity, wholly-owned subsidiary of parent organization).

Table 3: Subsidiary distribution

subsidiary	Freq.	Percent	Cum.
0	496	70.25	70.25
1	210	29.75	100.00
Total	706	100.00	

As shown in *Table 3*, approximately 70% of the worldwide sample investors have internal units.

--- insert *Appendix D* about here ---

In *Appendix D*, a distribution of CVC units in the broad industry categories is laid out (i.e. two-ciphered SIC codes)[26]. Evidently, a few industries are responsible for a large share of the CVC units in our sample: Only four broad industries make up more than 50% of all units, and 80% of CVC units are found in only 13 industries.

With 121 investors, the most prominent industry is 73, Business Services (this is how the broad code is defined) (OSHA, 2018), of which 109 investors are categorized within 737 (Computer Programming, Data Processing, And Other Computer Related Services). Chemicals and Allied Products (broad code 28) make up 91 investors, with 65 of those within the pharma and medicinal industries (Drugs, code 283). 'Electronic And Other Electrical Equipment And Components, Except Computer Equipment' (broad code 36) is the third largest industry, with a total of 84 investors – of which 34 are within category 367 (semiconductors, officially 'Electronic Components And Accessories') and 28 within category 366 (communications equipment). Other significant industries (defined as number of investors, in this case +20 investors) include 'Communications' (cat. 48), 'Industrial And Commercial Machinery And Computer Equipment' (cat. 35 – to a large degree computers, cat. 357), instruments and related products (cat. 38, officially 'Measuring, Analyzing, And Controlling Instruments; Photographic, Medical And Optical Goods; Watches And Clocks') and 'Wholesale Trade-durable Goods' (cat. 50, includes the automobile industry).

[26] The reason that we apply two-ciphered rather than three-ciphered SIC codes here is for a better overview. We will still comment on the most important points with regards to the three-ciphered SIC codes.

The key implication is that the industries with SIC codes 737, 283 and 367 are among the three three-ciphered SIC codes that, from our sample, display the most significant CVC activity. This is a key argument for choosing to focus the analysis on these.

--- insert *Appendix E* about here ---

In *Appendix E*, a matrix of two-ciphered SIC industry codes and the binary subsidiary variable is reported.

Interestingly, industries display different characteristics with regards to the subsidiary variable. For the eight industries with more than 20 observations/investors, the percentage of internal units range from approx. 35.00 % (Wholesale Trade-durable Goods, cat. 50) to 85.71% in category 36. However, most of the industries fall in the range 65-84%. This is also the case for SIC codes 737, 283 and 367, that have 85%, 67% and 85% internal units, respectively. While this paper is, due to limited scope, reduced to a few industries, examining cross-industry differences and examining the other industries in-depth with regards to the antecedents to setting up the CVC unit internally or externally could be a direction of further research.

--- insert *Appendix F* about here ---

In *Appendix F*, a matrix of geography of the parent company and the binary subsidiary variable is reported. Evidently, the share of internal vs. external units differs greatly between countries. The US pulls the average internal (the average is approx. 70%) up due to its large weight (68% of investors are US-based – and 78% of these are internal). Half of the of the 30 countries (notably Germany, Korea, China, France, Sweden, which, interestingly is external in all 8 Swedish observations and more) have more external units than internal, suggesting that a country-by-country approach to researching the phenomenon might yield the most accurate results. A number of factors might cause CVC activity to differ from country to country, including, but not limited to legislation, business culture and tradition, as well as reporting biases (database-wise, that is).

As already stated, our econometric analysis is limited to cover US American CVC activities. This is the case for three reasons: (i) evidently, US activities are by far the most predominant in our sample, (ii) the wish to focus the analysis and (iii) the fact that the patent data is from the US

American database, which creates a natural bias (as e.g. Chinese companies might not register their patents with the US American patent authorities).

Table 4: Subsidiary distribution per industry, US only (≥10 CVC units)

	subsidiary		
SIC_broad	0	1	Total
73	89	11	100
	89.00	11.00	100.00
36	55	6	61
	90.16	9.84	100.00
28	43	17	60
	71.67	28.33	100.00
48	31	8	39
	79.49	20.51	100.00
35	31	6	37
	83.78	16.22	100.00
38	18	9	27
	66.67	33.33	100.00
37	13	5	18
	72.22	27.78	100.00
27	8	3	11
	72.73	27.27	100.00

In *Table 4*, US industries are reported in a matrix with the subsidiary variable. To give a simple overview, only industries with more than 10 active CVC units (in our sample) are included. The whole picture can be found in *Appendix G*. Compared to the international, bigger sample, the cross-industry differences in terms of the subsidiary variable vary considerably less in the US-only part of the sample. Except for category 49 through 51, the different industries are more similar. Specifically, the industries with more than 10 CVC units all fall in the range of 67-90% with internal units.

Table 5: Source of organization data

	subsidiary		
organization_source	0	1	Total
1	428	177	605
	70.74	29.26	100.00
0	68	33	101
	67.33	32.67	100.00
Total	496	210	706
	70.25	29.75	100.00

Before delving into the descriptive statistics for the patent data, we check for any potential issues with organization data origin. Specifically, we check the subsidiary variable in connection with the *organization_source* variable, which is a dummy variable where 1 denotes that the parent organization was taken from the Compustat list of listed companies, and 0 denotes that it was found via clerical review (SEC filings etc.). *Table 5* shows a simple count and a share of internal and external units (0 and 1) for the two data sources. Evidently, the difference is not large, with merely a few percentage points separating the two. This test reduces the risk of potential bias stemming from the clerical review or the database, and thus adds validity to the analysis.

6.1.2. Patent data

As set out, a patent database was constructed using different data sets from official patent data gathered by PatentsView (PatentsView, 2017). The data available includes, but is not limited to, information about the sequence, number of claims, application and granted date as well as forward and backward (self-) citations and the current (as of May 2015) United States Patent Classification (USPC). We have only kept information about *utility* patents, as these were deemed relevant (other patents include e.g. *design* patents). The database solely includes data on US patents. For this reason (and other reasons previously explained), we exclude non-US investment units. Including foreign investment units could potentially lead to a significant bias in the data, as the parent organizations of foreign investment units are likely to have the bulk of their patents registered elsewhere, which would render the data inaccurate for these units and thus bias the results. This is also the reason for only including US investment units in the descriptive statistics part relating to patents. Firstly, we include all industries, where after we investigate our main industries separately. All of the following tables represent data as of the maximum year of investment of the CVC units.

Table 6: Mean number of patents and distinct USPC classes, US only (across all industries)

subsidiary	Mean, number of applied-for patents	Mean, number of distinct USPC classes of applied-for patents	Mean, standard deviation of applied-for patents per USPC class
0	1565.669	43.00875	31.89475
1	1371.99	41.71429	44.34915

As shown in *Table 6*, both the average number of patents and number of different USPC is higher for parent organizations with internal CVC units than for parent organizations with external CVC units. However, the difference between the mean number of distinct USPC classes is very small. As is shown, the standard deviation of total patents per USPC class is higher for external than internal units. This highlights the difference between the two measures, as standard deviation takes the dispersion of patents into account, whereas the mean number of USPC classes is a simple count. More interestingly, there is a large difference between the cross-industry mean total number of patents for parent organizations with internal CVC units and for those with external CVC units.

Table 7: Mean number of forward citations, share of backward self-citations, US only (across all industries)

subsidiary	Mean, number of forward citations of granted patents	Mean, share of backward self-citations of applied-for patents
0	30535.84	.0386923
1	21412.69	.0500639

As shown in *Table 7*, the total number of forward citations of granted patents is higher for parent organizations that decide to set up internal units rather than external units. The share of backward self-citations of applied-for patents is lower for these. While there is a large difference between the total number of forward citations (the total number of forward citations is almost 50% larger, on average, for parent organizations with subsidiary value 0), the difference in the share of backward self-citations is less pronounced.

The descriptive statistics have now been performed for the overall data sample. Now, we investigate the main industries separately. To provide an overview of the reduced sample containing only US American CVC units from three industries, which constitutes the basis of analysis, *Table 8* shows the distribution of the variable subsidiary.

Table 8: Subsidiary distribution in industries 283, 367 and 737, US only

subsidiary	Freq.	Percent	Cum.
0	136	84.47	84.47
1	25	15.53	100.00
Total	161	100.00	

As becomes evident, the share of internal units in the sample amounts to almost 85%, which is much higher than the complete sample including all industries and countries with a share of approximately 70% internal CVC units.

Table 9: Mean number of patents and distinct USPC classes, US only (SIC 737)

subsidiary	Mean, number of applied-for patents	Mean, number of distinct USPC classes of applied-for patents	Mean, standard deviation of applied-for patents per USPC class
0	1395.086	22.27273	25.33797
1	326.2222	27.77778	15.97045

As can be seen in *Table 9*, the total number of applied-for patents is a lot higher for parent organizations that decide to set up internal units than external units (1395 and 326, respectively) in the industry 737. This value is more distinctly different than in the total dataset. The total number of distinct USPC classes of applied-for patents is, contrary to the dataset including the other industries, higher for organizations with a *subsidiary* value of 1 than 0. On the contrary, the standard deviation of total patents per USPC class is higher for organizations with internal than external units on average. However, as shall be seen, the industry with the SIC code 737 is the only industry where this is the case (on average – we will look into how variables are distributed and how conclusive the average number is in section 6.2.1).

Table 10: Mean number of forward citations, share of backward self-citations, US only (SIC 737)

subsidiary	Mean, number of forward citations of granted patents	Mean, share of backward self-citations of applied-for patents
0	32258.11	.0230217
1	4343.556	.0393907

As shown in *Table 10*, which includes only the industry with SIC code 737, the difference between the number of forward citations of granted patents is much larger than in the overall sample. While organizations with internal units have a total of, on average, 32,258 forward citations to their patents, those with external units have only 4,343. While both the values for the share of backward self-citations of applied-for patents are lower for both subsidiary values 0 and 1 in this industry, the proportion between the values is approximately the same as for the overall sample.

Table 11: Mean number of patents and distinct USPC classes, US only (SIC 283)

subsidiary	Mean, number of applied-for patents	Mean, number of distinct USPC classes of applied-for patents	Mean, standard deviation of applied-for patents per USPC class
0	912.2903	35.3	58.0149
1	1369.308	41.15385	84.29617

As shown in *Table 11*, which includes only the industry with SIC code 283, the values are different than in the overall sample. The mean number of applied-for patents for organizations with an internal unit are lower than those with an external unit, differing from both the overall sample and from industry 737. While this might seem counter-intuitive, note that these are mean values and hence do not convey a perfect image of the data. The number of distinct USPC classes, however, shows a similar relative proportion, but higher values, than industry 737. As in the overall sample, the standard deviation of total patents per USPC class is higher for corporations with an external CVC unit.

Table 12: Mean number of forward citations, share of backward self-citations, US only (SIC 283)

subsidiary	Mean, number of forward citations of granted patents	Mean, share of backward self-citations of applied-for patents
0	17726.92	.0837177
1	12870.69	.1131655

Table 12 shows a higher mean number of forward citations, and again in this case, the organizations with internal units have more forward citations to their applied-for patents. However, the difference between organizations with subsidiary value 0 and 1 are less pronounced in this industry. The mean share of backward self-citations is a lot higher than both the overall sample and industry 737, but the organizations with internal units, similar to both the overall sample and industry 737, have lower shares compared to external units.

Table 13: Mean number of patents and distinct USPC classes, US only (SIC 367)

subsidiary	Mean, number of applied-for patents	Mean, number of distinct USPC classes of applied-for patents	Mean, standard deviation of applied-for patents per USPC class
0	1733.25	52.33333	38.62084
1	3611.333	59.33333	96.97372

Table 13 shows the number of patents and distinct USPC classes for the industry 367. As in industry 283, organizations with an internal unit have fewer patents, on average, than those with external units, which differs from both industry 737 and the overall sample. The number of distinct USPC classes of applied for patents is higher for organizations with external than with internal units, with a difference of about 7 classes on average. Furthermore, consistent with the overall sample, the standard deviation of patents per USPC class is higher for corporations with an external unit.

Table 14: Mean number of forward citations, share of backward self-citations, US only (SIC 367)

subsidiary	Mean, number of forward citations of granted patents	Mean, share of backward self-citations of applied-for patents
0	28084.87	.0560614
1	63710.33	.0615108

Table 14 shows the mean number of forward citations to granted patents and the mean share of backward self-citations of applied-for patents for organizations in the industry with SIC code 367. Differing from the overall sample and the other industries of interest, the organizations that decide to set up internal units have a lower number of total forward citations than those that engage in CVC via external units. However, consistent with the overall sample, the mean share of backward self-citations is higher for those with external than internal units, respectively.

The differences between industries are evident, which could have a number of implications for this study, such as the importance of studying industry-by-industry and the need to conduct tests for significance when analysing the data further. We will conduct a more in-depth analysis how variables are distributed (as this can make the reported mean value less conclusive) and how they should be modified for modelling purposes in section 6.2.1.

6.1.3. Qualitative description of three major industries

As described, this thesis will focus on three industries, namely Drugs (*sic_3*=283), Electronic Components And Accessories (*sic_3*=367) and Computer Programming, Data Processing, And Other Computer Related Services (*sic_3*=737). In the following, each of these industries will be described qualitatively, in order to provide the reader with a better contextual understanding of the CVC activity in those industries.

Drugs

The following short overview is focused on Pharmaceutical and Biotechnology,[27] as these two subindustries constitute the vast majority of corporations in that industry in our dataset. Some of the main characteristics of this industry are (i) central role of innovation and R&D productivity, (ii) long product cycles, (iii) heavy regulation and (iv) relatively high growth levels.

The pharmaceutical industry (SIC code 283) is characterized by many regulatory and legal requirements, which results in lengthy lead times for new products, taking up to ten to fifteen years (MarketLine, 2018e). Consequently, the industry is subject to long product cycles as well as a high risk of failure of new products, and many corporations depend on a small number of products (Reaume, 2003). This makes innovation and consequently R&D development central to the industry (Reaume, 2003). Technologies for drug discovery, which are subject to constant change and improvements, play a major role in increasing R&D productivity (Reaume, 2003). To insure a company against the high risk, patents are of special importance in order to reap benefits from its long-term R&D efforts; and lately, patent expirations have resulted in large revenue losses for many corporations (MarketLine, 2018e). Leading players in the pharmaceutical industry in the United States include Johnson & Johnson, Merck & Co., Inc. and Pfizer Inc. (MarketLine, 2018e).

The biotechnology industry comprises the "development, manufacturing and marketing of products based on advanced biotechnology research" (MarketLine, 2018b, p. 7), and is thus very technology-heavy. Leading companies in the United States currently include Amgen, Inc., Baxter International

[27]The industry with the three-digit SIC code 283 is a subordinate of the major category Chemicals And Allied Products (two-digit SIC code 28) and includes four major sub-groups, namely Medical Chemicals and Botanical Products (SIC 2833), Pharmaceutical Preparations (SIC 2834), In Vitro and In Vivo Diagnostic Substances (SIC 2835) and Biological Products, Except Diagnostic Substances (SIC 2836).

Inc. and Biogen, Inc. (MarketLine, 2018a). In the United States, Biotech grew with a compound annual growth rate (CAGR) of 14.4% between 2013 and 2017 (MarketLine, 2018a), and hence faster than the pharmaceutical industry with a CAGR of 8% in the same time period (MarketLine, 2018e). However, the industries are similar, as intellectual property in the Biotechnology industry is also essential, and along with costly R&D and high regulations constitutes high barriers of entrance for new players (MarketLine, 2018a).

According to David, Mehta, Norris, Singh and Tramontin (2010), the declining R&D productivity in the industry has prompted companies to externalize more R&D, mostly through traditional models. However, as the R&D process strongly depends on changing technologies (Reaume, 2003), CVC can provide opportunities in this context. CVC investments further constitute a bridge between the Pharmaceutical and the Biotechnology industry to access innovative assets and provide capital (Booth, 2011; Reaume, 2003). In our dataset, the industry with the three-digit SIC code 283 accounts for a total of 1376 investments, with an equity estimate of close to USD 5 billion.

<u>Electronic Components and Accessories</u>

For Electronic Components and Accessories industry, some of the main characteristics are: (i) capital intensity (expensive R&D), (ii) reliance on new technologies, (iii) high M&A activity level.

In the Electronic Components and Accessories industry (SIC code 367),[28] semiconductors (production/sale of integrated circuits present in electronic devices) is the largest segment, accounting for over 70% of the total market value in 2011 (MarketLine, 2012). The semiconductor market has experienced strong growth globally, with a CAGR of 7.2% between 2013 to 2017 (MarketLine, 2018c). The global semiconductors and electronic components market is led by Intel with a market share of more than 10% in 2011, followed by Samsung Electronics with a share of slightly over 6% (MarketLine, 2012).

The semiconductor industry is characterized as capital intensive and technology-heavy due to sophisticated R&D processes, which leads to market entry barriers (MarketLine, 2018f). For leading industry players, patent rights to certain products are essential (MarketLine, 2018f). Even

[28] The industry with the three-digit SIC code 367 comprises different electronic components (such as capacitors, connectors, etc.) and semiconductors and related devices.

though sales in the semiconductor industry reported a record high in 2016, the landscape is evolving, which imposes challenges (Bauer, Kenevan, Patel, & Santhanman, 2017). Major challenges include novel manufacturing technologies, changes in market demand and increased price pressures (Bauer, Kenevan, et al., 2017). For the future, demand in the semiconductor industry is expected to be driven by growth in artificial intelligence (AI) and virtual reality products (MarketLine, 2018c). As competition has increased, many industry players are engaging in M&A activity, "hoping to capture the wave of productivity improvements" (de Backer, Mancini, & Sharma, 2017, p. 59). Furthermore, there is an increasing focus from hardware to software, an area in which many start-ups are active (Bauer, Burkacky, Kupferschmidt, & Rocha, 2017), which constitutes a motive for engaging in CVC.

The Intel Corp. has not only been dominating the industry, but also its CVC activity, as already described when introducing the company as an exemplifying case. The industry with the three-digit SIC code 367 has accounted for 2064 CVC investments in our dataset, out of which Intel performed 1551 investments alone. The estimated equity invested from that industry amounts to over USD 10 billion.

Computer Programming, Data Processing, And Other Computer Related Services

This industry (SIC code 737)[29] is characterized, amongst other things, by: (i) high level of competition, (ii) technology heavy, (iii) newly imposed regulation and (iv) high levels of M&A activities.

The global software market grew with a CAGR of 2.4% between 2013 and 2017, while the United States constitutes the largest market (MarketLine, 2018d). However, with an increasing digitalization of different industries (for example finance and healthcare) and the introduction of complex (and hence expensive) software based on AI and the IoT (Internet of Things), revenues are expected to grow in the US with a CAGR of 4.3% during the next years (MarketLine, 2018d).

[29] The industry with the three-digit SIC code 737 includes different computer services, incl. programming, maintenance, and leasing, as well as prepackaged software (SIC 7372), which makes up around half of the investors in the three-digit SIC industry.

Leading players in the US market include International Business Machines Corp. (IBM), Microsoft Corp. and Oracle Corp. (MarketLine, 2018g).

The US market is characterized by high competition and constant technological progress, requiring specific and modifiable knowledge resources (MarketLine, 2018g). Patents are important in the industry, but very complex: Patent infringement is often a problem for new market entrants and copyright wars as well as anti-trust lawsuits are common (MarketLine, 2018d). Furthermore, the industry is characterized by a high disruptive potential and hence an elevated degree of risk (Chitkara, Gloger, & McCaffrey, 2018). This is demonstrated, for example, by recently imposed regulations, such as the EU's General Data Protection Regulation, which addresses unforeseen effects of the industry's products (Chitkara et al., 2018). To keep track with technological changes and new standards, large corporations frequently engage in M&A-related activities to gain access to the technological capabilities of innovative smaller companies (MarketLine, 2018g).

In our dataset, organizations active in the industry with the three-digit SIC code 737 are most prevalent, amounting to 90 out of 161 investors with a total of 1523 investments and an estimated equity of over USD 7 billion. Google is the most active investor with 305 investments, followed by Microsoft Corp. and the CVC unit of SAP SE.

In summary, the industry overview shows that *technology* plays a major role in all of the three industries, and that these industries constantly have to adapt to change. As CVC is often regarded as a mean to access new technologies, it appears reasonable that those three industries are the most active investors in CVC. Entrepreneurs in these three segments, next to Telecommunication & Networking, have also been most targeted by CVC activity (Cumming, 2012). Furthermore, the importance of *patenting* activities in these sectors is highlighted, which supports the application of the chosen predictors in the model based on patent information.

6.2. Model

6.2.1. Descriptive statistics for model building

In the following section, main statistics of the variables relevant for the development of our model will be described. Specifically, we will look at how the different variables are distributed to enhance the understanding of the underlying data and justify certain necessary transformations. Secondly, we will look at how the variables are correlated, and make necessary changes to address potential multicollinearity problems, which could impede internal validity of the model as described in section 5.3.3.

a) Distributions and summary statistics

Table 15 shows a summary of the main statistics of each of the variables in our analysis. As previously explained, only investors based in the United States are included (161 distinct CVC units). A detailed summary as well as the graphed distribution of each of the variables can be found in *Appendix H*.

Table 15: Summary statistics of variables

Variable	Number of observations	Mean	Std. Dev.	Median	Min	Max
cum_fc_g	136	25874.47	133659	1770.5	0	1485383
share_bsc_app_cum	156	.0489736	.0689145	.019913	0	.3261895
cum_patents_app	161	1332	6388.593	58	0	75546
sd_tot_uspc_app	144	39.68134	81.65334	7.59869	0	680.4418
cum_dist_uspc_app	156	32.00641	45.09338	14.5	0	298
num_investments_tot	161	30.82609	130.7358	5	1	1551
equity_est_firmname_tot	161	138.0093	690.5847	21.24	.066	8471.891
same_sic_proportion_mean	159	.5994788	.3508371	.685185	0	1
same_nation_proportion_mean	161	.8927167	.2207646	1	0	1
comp_age_avg_mean	157	4.461598	3.42369	4.2	0	37.25
num_coinvestors_round_mean	161	4.12845	2.031323	4.0625	0	13
num_corpinv_round_mean	161	.569109	.5240646	.5	0	2.666667

As previously explained, forward citations were counted by granted date, whereas the number of patents was retrieved by application date. This is the main reason why the number of observations is lower for forward citations (*cum_fc_g*) than for the number of patents (*cum_patents_app*): A few organizations do not have any *granted* patents by the date of the last investments (and hence, forward citations are essentially missing values in these cases), but have *applied-for* patents (which were later granted). The distribution of both variables, however, looks very similar. For both total forward citations and total number of patents, the distribution is highly skewed to the right. The maximum number of forward citations is 1,485,383 – these are forward citations of granted patents by International Business Machines Corp (IBM), a company active in the industry Computer Integrated Systems Design (*sic_4*=7373). This is an extreme, however, as 50% of organizations received 1,771 or fewer (median value), and 99% of organizations received 359,559 or fewer forward citations to their granted patents. Similarly, the highest number of applied-for (and later granted) patents, namely 75,546, is also assigned to IBM. The median observation, however, is much lower with only 58 patents, and the 99^{th} percentile amounts to 19,555 patents, which is still much below the maximum value. Only five organizations do not count any granted patents which were applied for by the maximum year of investment of their CVC unit. Four of these are also active in Computer Programming, Data Processing, And Other Computer Related Services (*sic_3*=737), all of which invested prior to 2003.

Evidently, the distribution of both patent and forward citation count is skewed. For this reason, the variables will be logged. To no surprise, the data on the number of patents and forward citations further shows that firms, but also industries, differ in their patenting activity. Whether these differences are significantly related to the structure of the CVC unit will be examined in our model.

The share of backward self-citations in total backward citations (*share_bsc_app_cum*) is also positively skewed, which is the reason why its natural logarithm will be employed in the model. While the mean lies around 5%, the organization Merck & Co, active in Pharmaceutical Preparations (*sic_4*=2834), has the highest share of backward self-citations for 4,757 applied for (and later granted) patents, amounting to more than 32%. In plain words, this means that 32% of Merck & Co's backward citations come from their own patents. However, it is worth noting that in the Drugs industry (SIC code 283), the mean share (9.3%) is higher than in both 367 and 737 with a mean of 5.7% and 2.5% respectively (see ***Appendix I*** for overview based on 3-digit SIC codes), i.e. the CVC units active in the Drugs industry (from our sample) have a relatively high share of self-citations.

With regards to the standard deviation of the number of patents in different technology classes (*sd_tot_uspc_app*), the mean of approximately 39.9 is much higher than the median of 7.6, and the distribution is skewed to the right. The organizations with a standard deviation in the top 5^{th} percentile hold a minimum of 3,868 patents and include IBM, Texas Instruments Inc, Advanced Micro Devices, Merck & Co and Pfizer - coming from all three industries. IBM is also the company with the highest number of distinct USPC classes (*cum_dist_uspc_app*). There are nine organizations which only patented in one USPC main class, mostly from software and programming. However, most of them have a very low number of patents (maximum of 5), indicating a correlation between the two variables, which will be examined in sub-section b).

The control variables show a lower degree of skewness (see *Appendix H*), with the exception of the number of investments (*num_investments_tot*), total estimated equity invested (*equity_est_firmname_tot*) and the number of distinct USPC classes (*cum_dist_uspc_app*), which are all skewed to the right. With a total number of 1551 investments and an equity estimate of USD 8,472 million, the most active corporate investor is by far Intel Corp, which invests through its external CVC unit Intel Capital Corp. The second biggest investor is Johnson & Johnson, a pharmaceutical company investing through an external CVC unit, with 350 and hence much less total investments. Half of the investors in the sample have made 5 or fewer investments. The median for invested equity is also much lower than its mean, namely USD 21.24 million as opposed to USD 138 million. Interestingly, this implies that most of the observed CVC activity is performed by a few actors. Specifically, almost 75% of all CVC activity in terms of number of investments (and 77% in terms of equity invested) is performed by merely 10% of the investment units.

In summary, it becomes evident that in the model, the natural logarithm should be employed for all independent variables, namely *cum_fc_g, cum_patents_app*, *share_bsc_app_cum* and *sd_tot_uspc_app*, due to positive skewness. With regards to control variables, taking the natural logarithm of *num_investments_tot*, *equity_est_firmname_tot* and *cum_dist_uspc_app* is meaningful. However, for the remaining variables, a mean has already been taken while collapsing the dataset as described in section 5.1.4. Hence, even though slightly skewed, transforming those variables would eliminate too much of the variance, and we thus choose not to transform them to maintain the integrity of these variables.

b) Correlations

In a next step, we will look at the how the variables, which we want to insert in our model (transformed as previously described) are related. High correlations indicate potential multicollinearity, which could lead to an imprecise estimation of the partial effects of the regression coefficients in form of a large sampling variance (Stock & Watson, 2015). This implies that it is difficult to detangle the different predictors. Changing the set of predictors is a possible solutions to multicollinearity problems (Stock & Watson, 2015).

Correlations between all of the independent and control variables are shown in *Table 16*.

Table 16: Full correlations matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	1.0000											
(2)	0.6608	1.0000										
(3)	0.8194	0.7457	1.0000									
(4)	0.7880	0.7169	0.9322	1.0000								
(5)	0.7349	0.5880	0.7753	0.6364	1.0000							
(6)	0.1787	0.2331	0.3547	0.2769	0.2105	1.0000						
(7)	0.1225	0.1992	0.3688	0.2534	0.1920	0.8835	1.0000					
(8)	-0.0284	-0.1478	0.0174	-0.1245	-0.1948	0.0189	-0.0128	1.0000				
(9)	-0.0152	0.0152	0.0418	0.0019	-0.0244	0.0748	-0.0917	0.1856	1.0000			
(10)	-0.0060	0.0268	0.0214	0.0713	0.0264	0.0709	0.0880	-0.1161	0.0077	1.0000		
(11)	-0.0232	-0.1215	0.0376	-0.0843	0.0413	0.1455	0.0940	-0.1278	0.1219	-0.1159	1.0000	
(12)	-0.0665	-0.0308	0.0468	-0.0330	0.0503	0.0454	0.0837	-0.1325	0.1036	-0.0134	0.5118	1.000

Note. Variables are denoted as follows: (1) cum_fc_g_ln, (2) share_bsc_app_cum_ln, (3) cum_patents_app_ln, (4) sd_tot_uspc_app_ln, (5) cum_dist_uspc_app_ln, (6) num_investments_tot_ln, (7) equity_est_firmname_tot_ln, (8) same_sic_proportion_mean, (9) same_nation_proportion_mean, (10) comp_age_avg_mean, (11) num_coinvestors_round_mean, (12) num_corpinv_round_mean

As shown, there are several very highly correlated variables. Each of those pairs will be discussed in separate below.

Firstly, it is evident that (3) *cum_patents_app_ln*, the cumulative number of applied-for patents, is highly correlated with many variables, namely (1) *cum_fc_g_ln* ($\rho = 0.8194$), (4) *sd_tot_uspc_app_ln* ($\rho = 0.9322$), and (2) *share_bsc_app_cum_ln* ($\rho = 0.7457$) and (5) *cum_dist_uspc_app_ln* ($\rho = 0.7753$). We used that variable to measure absorptive capacity, as explained in section 0. As the model does not work with such high correlations, we must omit one or more variables. In this case, we choose to omit variable (3), *cum_patents_app_ln*, as it has high correlations with all other independent variables. This means that we will be unable to conclude on the theorized effects for absorptive capacity. [30]

[30] While this thesis will not conclude on findings for absorptive capacity, it should be mentioned that the very high correlation its proxy variable has with the proxy for value of innovations (forward citations) essentially is evidence that the two variables measure very similar concepts.

Secondly, (1) *cum_fc_g_ln* and (4) *sd_tot_uspc_app_ln* are highly correlated ($\rho = 0.7880$). To offset this issue, we transform this variable into a binary variable instead (called *sd_tot_uspc_app_bin*), taking the value of 0 for a low standard deviation and the value 1 for a high standard deviation of USPC classes. We base this distinction on the median instead of the mean value of the un-transformed variable *sd_tot_uspc_app* (7.59869) to account for the previously described positive skewness. This transformation is also used to accommodate the issue of high correlation between (2) *share_bsc_app_cum_ln* and (4) *sd_tot_uspc_app_ln* ($\rho = 0.7169$).

Thirdly, there is a high correlation between the independent variable (1) *cum_fc_g_ln* and the control variable (5) *cum_dist_uspc_app_ln* ($\rho = 0.7349$). As a consequence, we decide to exclude the number of distinct USPC classes as a control variable. While previously argued that the standard deviation of USPC dispersion is only meaningful in conjunction with this variable, the high correlation with *cum_fc_g_ln* implies that the model effectively still is specified sufficiently.

Fourthly, the variables (6) *num_investments_tot_ln* and (7) *equity_est_firmname_tot_ln* show a high correlation ($\rho = 0.8835$). Essentially, they both control to which extend the magnitude and scope of the CVC activity is related to the set-up as an internal or external unit. Based on running a maximum-likelihood regression with each of the two as the sole independent variable separately (see *Appendix J*), we deem (5) *num_investments_tot_ln* the most apt control variable of the two and eliminate (6) *equity_est_firmname_tot_ln*.

Lastly, even though the correlation between (11) *num_coinvestors_round_mean* and (12) *num_corpinv_round_mean* is acceptable ($\rho = 0.5118$), we transformed *num_corpinv_round_mean* into a binary variable as well. We consider this meaningful as the decision to co-invest with another corporate investor (who also has strategic interest) itself matters more than the count of actual corporate co-investors (for the purposes of this analysis). Hereby, the variable *num_coinvestors_round_mean* includes both corporate and other investors, and hence is sufficient to control for the number of other investors participating in the investments. Consequently, a binary variable is introduced, *corp_co_invest*, which takes the value 1 if the number of corporate co-investors is at least one and takes the value 0 if the investors do not invest alongside other corporate investors.

The final correlations matrix of model input variables with transformed and adjusted variables is shown in *Table 17*.

Table 17: Final correlations matrix of model input variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)	1.0000								
(2)	0.6608	1.0000							
(3)	0.6642	0.6442	1.0000						
(4)	0.1787	0.2331	0.2803	1.0000					
(5)	-0.0284	-0.1478	-0.0924	0.0189	1.0000				
(6)	-0.0152	0.0152	0.0077	0.0748	0.1856	1.0000			
(7)	-0.0060	0.0268	0.0243	0.0709	-0.1161	0.0077	1.0000		
(8)	-0.0232	-0.1215	-0.0864	0.1455	-0.1278	0.1219	-0.1159	1.0000	
(9)	0.0636	0.1204	0.1404	0.4895	-0.0413	0.1332	0.0639	0.3729	1.0000

Note. Variables are denoted as follows: (1) cum_fc_g_ln, (2) share_bsc_app_cum_ln, (3) sd_tot_uspc_app_bin, (4) num_investments_tot_ln, (5) same_sic_proportion_mean, (6) same_nation_proportion_mean, (7) comp_age_avg_mean, (8) num_coinvestors_round_mean, (9) corp_co_invest

As can be seen, some variables still have correlation coefficients greater than 60%. As we do not encounter perfect multicollinearity in the model, and since the correlations are not too high for the separate variables to be meaningful, the correlations are deemed acceptable.[31]

6.2.2. Empirical model

After having established the input variables for our model, results from the empirical analysis will be presented in the following section. In sub-section a), we will build our final model. Next, we will check for industry differences in b) and introduce interaction terms in c). Lastly, in sub-section d), a robustness check will be performed.

a) Final model

To predict our dichotomous dependent variable *subsidiary*, we conduct a maximum likelihood estimation. We employ the *probit* function in Stata with robust standard errors for this purpose. To build our final model, we follow the approach of Di Lorenzo and Almeida (2017) and specify

[31] To determine if multicollinearity poses a problem, the variance inflation factor (VIF) is a commonly employed tool (Greene, 2014). However, as often-used rules of thumb are to some extent arbitrary, and the VIF does not constitute a conclusive measure on its own (O’Brien, 2007), it will not be discussed in this thesis.

different stages of our model: Model 1 is specified with our control variables and Model 2 our independent variables only, and Model 3 includes both control and independent variables (our final model). The output of the regressions is reported in *Table 18* (please see *Appendix K* for full regression models as in Stata). The final model, including independent and control variables, is hence specified as:

$$(8) \quad \Pr(Y = 1 | X_1, X_2, \dots, X_k)$$
$$= \phi(\beta_0 + \beta_1 cum\_fc\_g\_ln + \beta_2 share\_bsc\_app\_cum\_ln$$
$$+ \beta_3 sd\_tot\_uspc\_app\_bin + \beta_4 num\_investments\_tot\_ln$$
$$+ \beta_5 same\_sic\_proportion\_mean$$
$$+ \beta_6 same\_nation\_proportion\_mean$$
$$+ \beta_7 comp\_age\_avg\_mean + \beta_8 num\_coinvestors\_round\_mean$$
$$+ \beta_9 corp\_co\_invest)$$

Table 18: Regression results

	Model 1 (control) subsidiary	Model 2 (independent) subsidiary	Model 3 (final) subsidiary
cum_fc_g_ln		-.2009895**	-.2369588***
		(.0801362)	(.0895531)
share_bsc_app_cum_ln		.5308412*	.7033397**
		(.2850126)	(.3105812)
sd_tot_uspc_app_bin		.8360644*	.7935149*
		(.4397314)	(.4625462)
num_investments_tot_ln	.4143148***		.4250752***
	(.1019094)		(.135089)
same_sic_proportion_mean	.17129		.5394992
	(.4553366)		(.5779719)
same_nation_proportion_mean	.3479895		.5084136
	(.7292707)		(1.167192)
comp_age_avg_mean	.0253039		.0400396
	(.0301437)		(.0313146)
num_coinvestors_round_mean	-.0074228		.05557
	(.0765965)		(.0930378)
corp_co_invest	-.2688364		-.2894907
	(.358583)		(.4493534)
_cons	-2.290837***	1.525771	.2069709
	(.6926807)	(1.149041)	(1.778032)
Observations	155	132	128
Robust errors	yes	yes	yes
Wald chi²	18.44	10.62	24.72
Prob. > chi²	0.0052	0.0140	0.0033
Pseudo R²	0.1492	0.1074	0.2345

Note. Standard errors in parenthesis, ***p<0.01, **p<0.05, *p<0.1

Model 1 shows that our control variable *num_investments_tot_ln* is statistically significant and positively related to subsidiary ($p<0.01$), while the remaining control variables do not report significant coefficients. Model 2 reports that all our independent variables are significantly related to the structure of the CVC unit, however the model fit (Prob. > chi² as well as Pseudo R^2) is rather poor. It can be shown that overall model fit as well as the significance level of our independent variables is significantly improved in Model 3, which combines control and independent variables. Specifically, with a Pseudo-R^2 of 0.2345, our model fit falls in the range deemed as an "excellent fit" by (McFadden, 1977). All independent variables are significantly related to subsidiary at least on a 10% level, and *cum_fc_g_ln* and *share_bsc_app_cum_ln* become more significant compared to the underspecified Model 2. We will discuss each of the coefficients in the final model below.

The number of forward citations (*cum_fc_g_ln*) is negatively and significantly ($p<0.01$) related to the *subsidiary* variable. This variable was applied as a proxy for the value of innovations. Based on this output, we conclude that there is a negative relationship between the value of innovations and the set-up of a CVC unit as internal or external. As described in the theory section, we suggest that the following two effects may explain this result: a protection effect and a parenting advantage. Specifically, when innovations are valuable, the results indicate that organizations make an effort to mitigate value erosion and protect the company's IP. They are more prone to facilitate sharing their valuable internal innovation resources with the investee.[32] Organizations are more likely to set up the CVC unit internally the more valuable their innovations, potentially based on these two effects. The result implies that the decentralization effect, i.e. managing knowledge more efficiently in a decentralized manner, plays a smaller role and is outweighed by the former two effects in the context of this variable, at least. Summarizing, the negative and significant coefficient for *cum_fc_g_ln* confirms a negative relationship between the value of innovations and the likelihood of setting up an external CVC unit. This also implies that we can confirm value of innovations as an antecedent to the the choice of an internal or external CVC unit.

The relationship between the share of backward self-citations (*share_bsc_app_cum_ln*) and the likelihood of an external CVC unit is positive and significant on a 5% level. The variable was employed as a proxy of firm specificity, which was theorized to have a positive influence on setting up the CVC unit externally through both the NIH-syndrome and difficulty to absorb. Based on the results, we propose that the effects hold: To mitigate behavioural biases which lead to internal resistance by the NIH-syndrome, managers seem to set up the CVC unit outside the existing organizational boundaries. When firm specificity is high, managers are more likely to set up the CVC unit externally, indicating that the advantages of an internal CVC unit could be dampened as

---

[32] Importantly, since the *parenting advantage* is used to theoretically explain the relationship, which essentially states that internal units are likely to more easily leverage the parent company's resources, we test whether companies with higher forward citations values actually display such use of own resources. For this purpose, we look at the correlation between forward citations and forward self-citations (i.e. a measure that essentially shows to what degrees company uses own prior patents), which is 0.96. This supports the above explanation.

external innovations are more difficult to absorb. Based on the results, we can confirm that firm specificity is positively and significantly related to the likelihood of an external CVC unit. This also implies that we can confirm firm specificity as an antecedent to the setup of an internal or external CVC unit.

The coefficient of the standard deviation of patent dispersion in different USPC classes (*sd_tot_uspc_app_bin*) is positive and significant on a 10% level. We employed this variable to proxy technological diversification of the CVC unit's parent organization. Based on the results, we conclude that technological diversification is positively and significantly related to the likelihood of an external CVC unit. This can be explained by the theorized effects: Firstly, organizations could be more likely to set up an external unit as search, coordination and bureaucratic costs related to realizing novel technological opportunities can be reduced. Secondly, technologically diversified organizations might replicate existing organizational structures and hence are more likely to set up a CVC unit externally, i.e. the replication effect. The effect of path dependency, which suggests the opposite relationship, exerts a weaker influence than the sum of the first two effects, as indicated by the results. Overall, technological diversification is positively related to the likelihood of an external CVC unit, and can be confirmed as an antecedent to the choice of setting up an internal or external CVC unit.

Interestingly, one of our control variables, specifically the number of investments, shows a positive and highly significant relationship with the likelihood of setting up a CVC unit externally ($p<0.01$). We will discuss this in section 7.

To conclude: value of innovations, firm specificity and technological diversification are all found to be significantly related to the setup of CVC units either internally or externally, and can thus be confirmed as antecedents. Specifically, value of innovations, proxied by total forward citations of the patents of the parent organization of the CVC unit, is negatively and significantly ($p<0.01$) related to the likelihood of setting up an external CVC unit. Firm specificity, proxied by the share of backward self-citations in total backward citations of the parent organization's patents, is positively and significantly significant ($p<0.05$) related to the likelihood of setting up an external CVC unit. Technological diversification, proxied by the standard deviation of patent dispersion in different USPC classes, is positively related to the likelihood of setting up an external CVC unit. The relation is significant ($p<0.1$).

b) Industry differences

As outlined in the descriptive statistics and *Appendix I*, industries differ slightly in their patenting activity, suggesting that it could be meaningful to check for industry differences in the model. This is done through transforming the categorical variable *sic_3* to dummy variables in Stata and performing a maximum likelihood regression with an interaction expansion. The results can be seen (in comparison to Model 3) in *Table 19*. Please refer to *Appendix L* for the model output as in Stata.

Table 19: Industry differences regression output

	(Model 3) subsidiary	(Model 4) subsidiary incl. industry dummies (relative to s*ic_3*=737)
cum_fc_g_ln	-.2369588***	-.2241777**
	(.0895531)	(.0887568)
share_bsc_app_cum_ln	.7033397**	.6589684**
	(.3105812)	(.3122546)
sd_tot_uspc_app_bin	.7935149*	.7590212
	(.4625462)	(.4690956)
num_investments_tot_ln	.4250752***	.4238445***
	(.135089)	(.1344288)
same_sic_proportion_mean	.5394992	.4799851
	(.5779719)	(.574017)
same_nation_proportion_mean	.5084136	.5304377
	(1.167192)	(1.106186)
comp_age_avg_mean	.0400396	.0398522
	(.0313146)	(.030954)
num_coinvestors_round_mean	.05557	.0582879
	(.0930378)	(.0874849)
corp_co_invest	-.2894907	-.3235496
	(.4493534)	(.4466599)
_Isic_3_283		.0689648
		(.3440994)
_Isic_3_367		-.1109067
		(.3970956)
_cons	.2069709	.0527311
	(1.778032)	(1.65846)
Observations	128	128
Robust errors	yes	yes
Wald chi²	24.72	27.54
Prob. > chi²	0.0033	0.0038
Pseudo R²	0.2345	0.2355

Note. Standard errors in parenthesis, ***p<0.01, **p<0.05, *p<0.1

As is shown, both industry dummies for the industries with the SIC code 283 and 367 are insignificant, indicating that there are no significant intercept differences between the industries. In this context, it is important to note that investors active in Computer Programming, Data Processing, And Other Computer Related Services (sic_3=737) make up more than half of the observations included in the regression (68 observations), whereas Drugs (sic_3=283) and Electronic Components and Accessories (sic_3=367) only account for 36 and 24 observations, respectively. The small size of the observations in the different industries could hence partly account for the lack of significance, which is why we cannot rule out the possibility of cross-industry differences completely. Future research could address this issue.

Both forward citations and backward self-citations stay significant on a 5% level. Interestingly, when adding industry dummies, the standard deviation of USPC classes becomes insignificant (even though only slightly, as can be seen in model in *Appendix L*), potentially indicating that it is not as consistent as an explanatory variable. The control variable for number of investments, however, behaves very consistently and is still significant on a 1% level.

c) Interactions with technological diversification

We theorized technological diversification, next to being an independent variable, to exercise a moderating influence on the relationship between value of innovations (proxied by the number of forward citations) resp. firm specificity (proxied by the share of backward self-citations) and the structure of a CVC unit. We check for this moderating influence in our model by introducing interaction terms.

For the sake of interpretation of the interaction, we first created a binary variable for both forward citations (*cum_fc_g*) and the share of backward self-citations (*share_bsc_app_cum*). We used the respective median of the untransformed variable to determine "high" (dummy variable takes value 1), and "low" (dummy variable takes value 0). The binary variables are called *cum_fc_g_bin* and *share_bsc_app_cum_bin*.

However, in order to employ those newly created dummy variables, we have to check if they behave the same way as the continuous variables (in terms of direction). Hence, we run a probit regression as performed in Model 3, while replacing forward citations resp. backward self-citations with the binary variables. As can be seen in *Appendix M*, forward citations become insignificant, as too much of the variance is eliminated. Consequently, the interaction between forward citations and standard deviation of USPC classes as dummy variables will not be performed. Even though backward citations behave consistently and do not loose significance, only two observations are significant in the interaction (see *Appendix N*).

As much of the variance is eliminated through binary variables, we performed both interactions with one continuous (*cum_fc_g_ln* resp. *share_bsc_app_cum_ln*) and one dummy variable (*sd_tot_uspc_app_bin*). For this purpose, we created interaction variables (namely, *int_fc_uspc_app* and *int_sbsc_uspc_app*) by multiplying *share_bsc_app_cum_ln* resp. *cum_fc_g_ln* with *sd_tot_uspc_app_bin*. The results from both interaction models (in comparison to Model 3) are shown in *Table 20*. Please refer to *Appendix O* for the full regression output as in Stata.

Table 20: Regression output for interaction between dummy and continuous variable

	(Model 3) subsidiary	(Model 5) subsidiary incl. interaction with forward citations	(Model 6) subsidiary incl. interaction with backward self-citations
cum_fc_g_ln	-.2369588***	-.2107021*	-.2564012***
	(.0895531)	(.1107342)	(.0937601)
share_bsc_app_cum_ln	.7033397**	. 728843**	1.714671***
	(.3105812)	(.3268735)	(.4920035)
sd_tot_uspc_app_bin	.7935149*	1.292	-2.625323*
	(.4625462)	(1.165856)	(1.558117)
int_fc_uspc_app		-.0677132	
		(.1558644)	
int_sbsc_uspc_app			-1.222746**
			(.5271655)
num_investments_tot_ln	.4250752***	.418277***	.414441***
	(.135089)	(.1309421)	(.1425111)
same_sic_proportion_mean	.5394992	.4694017	.247017
	(.5779719)	(.5273748)	(.5569467)
same_nation_proportion_mean	.5084136	.6219729	.6758169
	(1.167192)	(1.177829)	(1.171887)
comp_age_avg_mean	.0400396	.0400783	.0390005
	(.0313146)	(.0312821)	(.0311927)
num_coinvestors_round_mean	.05557	.0565661	.0571714
	(.0930378)	(.0911843)	(.0996921)
corp_co_invest	-.2894907	-.2709885	-.3698114
	(.4493534)	(.4456059)	(.4596346)
_cons	.2069709	.0794011	3.451685*
	(1.778032)	(1.803404)	(1.88179)
Observations	128	128	128
Robust errors	yes	yes	yes
Wald chi²	24.72	24.93	30.92
Prob. > chi²	0.0033	0.0055	0.0006
Pseudo R²	0.2345	0.2359	0.2567

Note. Standard errors in parenthesis, ***p<0.01, **p<0.05, *p<0.1

As described in section 5.3.2, the reported significance level for interactions in maximum-likelihood models has to be examined more closely in order to see if significance changes along the curve – even if the z-statistics is reported as insignificant, certain observations or parts of the curve might be significantly influenced by the interaction. For this purpose, we use the inteff command in Stata as proposed by (Norton et al., 2004).

Figure 5: Interaction effect with forward citations

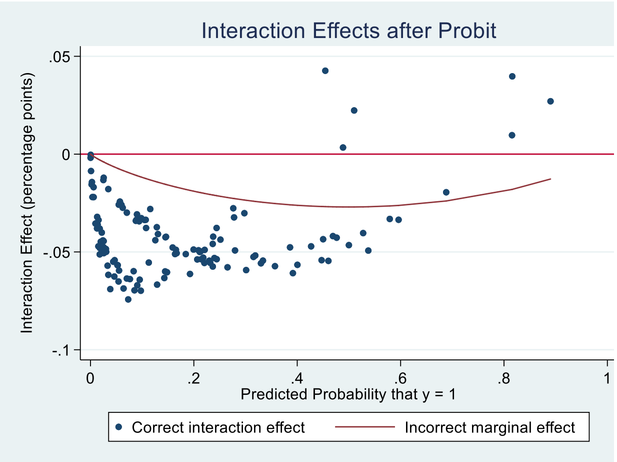


Figure 6: Significance of interaction effect with forward citations

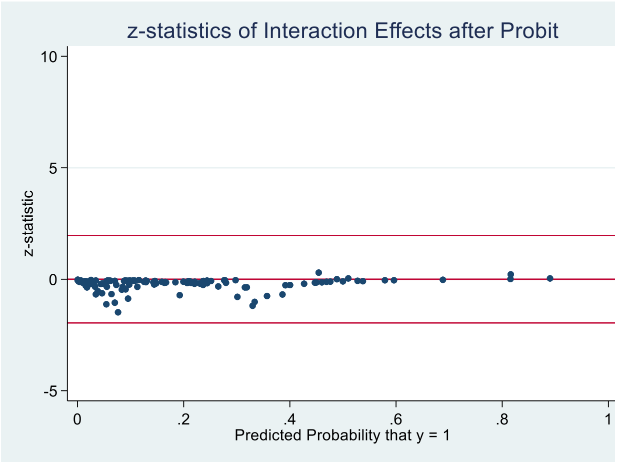


Figure 7: Interaction effect with backward self-citations

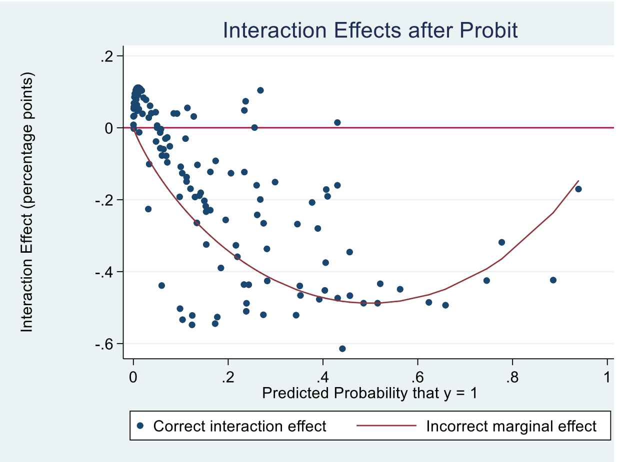
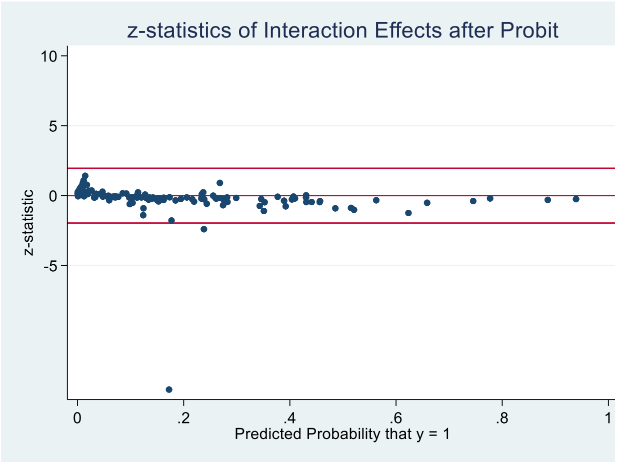


Figure 8: Significance of interaction effect with backward self-citations



As can be shown, none of the observations are significant with regards to the interaction between the value of innovations (proxied by *cum_fc_g_ln*) and the technological diversification of the organization (proxied by *sd_tot_uspc_app_bin*). This implies that technological diversification does not exert a significant moderating influence in the relationship between the value of innovations and the likelihood of an external CVC unit. The overall effect of value of innovations on the structure of a CVC unit was theorized by three sub-effects, namely a protection effect, a parenting advantage, and a decentralization effect. We theorized the moderating influence of technological diversification mainly through an interaction with the parenting advantage, but not the other two sub-effects. We cannot proxy the effects separately with the available data, but only the overall effect. Based on the result, we conclude that the theorized moderating effect on the parenting advantage does not influence the overall relationship sufficiently to show significant differences.

Interestingly, the interaction of firm specificity (proxied by *share_bsc_app_cum_ln*) and the technological diversification of the organization (proxied by *sd_tot_uspc_app_bin*) was indicated to be significant ($p<0.05$) in the regression model. As previously described, researchers often mistakenly conclude marginal effects on the basis of the reported coefficients and significance levels, but due to nature of non-linear models, interactions have to be examined more precisely (Norton et al., 2004). When taking a closer look, only three observations are actually significant on a 10% level. These hover around a likelihood of an external unit of 20%. Namely, these are CVC units by three highly diversified parent organizations: Pfizer Inc., Bristol-Myers Squibb Co. and Texas Instruments Inc., of which the former two are active in Drugs (sic_3=283) and the latter in Electronical Components and Accessories (sic_3=367). Due to the low number of significant observations, no other common patterns can be identified, and the results thus remain inconclusive.

To check if the lack of significance is caused by the creation of our binary variable, which is based on the median, we altered the dichotomous variable *sd_tot_uspc_app_bin* (as well as both independent variables) based on different percentiles (lower and upper 95^{th} and 90^{th} percentile). However, this did not lead to different findings. An overview of the performed interactions can be found in *Appendix P*.

In summary, we can conclude that technological diversification does not significantly moderate the relationship between the value of innovations resp. firm specificity and the likelihood of an external CVC unit. This indicates that technological diversification does not strongly dampen the parenting advantage, and firms can still leverage valuable internal resources well if the CVC is set up internally, independently of their level of technological diversification. Furthermore, these results indicate that an existing sense of myopia caused by high firm specificity is not sufficiently moderated by a high technological diversification, resulting in a to a large extent unchanged relationship between firm specificity and the likelihood of an external unit.

d) Robustness checks

Based on the available data, we conducted three robustness checks with regards to our final model (Model 3). Firstly, as explained in section 5.3.3, we will check if alternative model specifications (logit and linear regression) will significantly change the results (Model 7 and 8). Secondly, we construct a model including only CVC units with more than one investment, as this indicates commitment to the CVC activity (Model 9). Thirdly, as we showed that the distribution of the number of patents is highly positively skewed, we performed a test excluding the extreme tail and hence omit the upper 10th percentile of firms in the sample (Model 10). The results are reported in *Table 21* (full regression models as in Stata can be found in *Appendix Q*.

Table 21: Robustness checks

	Model 3 (final) subsidiary	Model 7 (linear reg) subsidiary	Model 8 (logit) subsidiary	Model 9 (num_investments_tot>1) subsidiary	Model 10 (cum_patents_app<2527) subsidiary
cum_fc_g_ln	-.2369588***	-.0580367**	-.4374742***	-.2300644**	-.2810393***
	(.0895531)	(.0225354)	(.1602011)	(.0893569)	(.1023336)
share_bsc_app_cum_ln	.7033397**	.1681937**	1.243449**	.6165299**	.5972301
	(.3105812)	(.0815141)	(.562518)	(.308978)	(.3767153)
sd_tot_uspc_app_bin	.7935149*	.1362785	1.378605	.7831892*	.9281639*
	(.4625462)	(.110233)	(.8801051)	(.46756)	(.5230826)
num_investments_tot_ln	.4250752***	.0990678***	.7740989***	.4371651***	.5433885***
	(.135089)	(.0304006)	(.2615508)	(.139266)	(.141337)
same_sic_proportion_mean	.5394992	.06799	1.08685	.4315674	1.794968***
	(.5779719)	(.0907457)	(1.106391)	(.6190961)	(.5859612)
same_nation_proportion_mean	.5084136	-.0714757	.9151251	.4581229	.0130999
	(1.167192)	(.1166022)	(2.313135)	(1.298668)	(1.170693)
comp_age_avg_mean	.0400396	.0075173	.0743002	.0485997	.0569046*
	(.0313146)	(.0093355)	(.0530909)	(.0347326)	(.0308646)
num_coinvestors_round_mean	.05557	.012028	.0776628	.0826557	.0348689
	(.0930378)	(.0142115)	(.1983941)	(.0970781)	(.1103527)
corp_co_invest	-.2894907	-.0824079	-.4281178	-.0592458	-.3629421
	(.4493534)	(.0771873)	(.8897281)	(.581344)	(.4887772)
_cons	.2069709	.7942025**	.3963124	-.3715493	-.5351951
	(1.778032)	(.3577836)	(3.366784)	(1.931017)	(2.040261)
Observations	128	128	128	116	111
Robust errors	yes	yes	yes	yes	yes
Wald chi²	24.72	3.60[1]	19.21	21.67	27.83
Prob. > chi²	0.0033	0.0005[1]	0.0234	0.0100	0.0010
Pseudo R²	0.2345	0.2115[1]	0.2358	0.2238	0.2952

Note. Standard errors in parenthesis, ***p<0.01, **p<0.05, *p<0.1; [1]The linear regression uses the F-statistic as a test statistic, reports Prob>F, and employs R^2 instead of Pseudo R^2

Model 7 or 8 check if there is the possibility of major error due to model misspecification. As shown, for both the linear and the logit regressions, *cum_fc_g_ln* and *share_bsc_app_cum_ln* stay significant on at least a 5% level, and the sign of the coefficient is in the same direction as in Model 3. However, the binary variable employed as a proxy for technological diversification loses significance in both models, even though in the logit only slightly (P>|z| = 0.117). This could be partly due to a weaker model fit: As described in section 5.3.1, the linear approximation is not well suited for dichotomous dependent variables. Nevertheless, considering that we could observe similar behaviour concerning the coefficient of *sd_tot_uspc_app_bin* when adding industry dummies, we can conclude that *sd_tot_uspc_app_bin* is not as robust as the remaining independent variables. This might partly be due to its binary nature, and future research could be conducted to investigate the influence of technological diversification using alternative measures.

Even though the significance level of variable *cum_fc_g_ln* decreases from 1% to 5% in Model 9, the coefficients of all independent variables remain significant and move in the same direction as in Model 3. When including only firms that engaged in CVC more than once, results are still robust.

In Model 10, when eliminating the CVC units whose parent companies have applied for the most patents (the upper 10^{th} percentile of the curve), *cum_fc_g_ln* and *sd_tot_uspc_app_bin* do not behave differently from the final model (Model 3), but *share_bsc_app_cum_ln* loses significance ($P>|z| = 0.113$). This can potentially indicate that observations with a high number of patents are necessary for firm specificity to be significant. Further research, potentially including non-US investors and their patenting activity, could investigate if this is valid for a larger sample.

Interestingly, and consistent with prior observations, the control variable *num_investments_tot_ln* behaves robustly in all specified models. Investigating this as an independent variable consequently is an interesting field for future research (and will be addressed in the discussion).

In summary, the performed robustness checks do not alter results with regards to *cum_fc_g_ln* and *share_bsc_app_cum_ln* significantly. However, for *sd_tot_uspc_app_bin*, the coefficient seems to be less robust based on the computed models. We will discuss further potential robustness checks in section 7.2.3.

6.2.3. Exemplify results with case companies

The following section aims to exemplify the found results by the three case companies, which were introduced in section 2.2 as illustrative examples of companies who engage in CVC, namely Intel Corp., Netscape Communications Corp. and Microsoft Corp. Intel is active in Semiconductors and Related Devices, and both Netscape and Microsoft are active in Prepackaged Software, which is the most prevalent industry of our data sample. Intel has an external CVC unit, and both Netscape as well as Microsoft invest in CVC through an internal CVC unit. As will be shown, Netscape illustrates our results with regards to all theorized aspects, while Intel and Microsoft deviate to some extent. In the following, each of the companies will be explained in separate.

Intel's total number of forward citations of granted patents (927 compared to a median of approx. 1771) at the maximum year of investment is relatively low. As this shows a relatively low value of innovations, it exemplifies well that the value of innovations is *negatively* related to the likelihood of an external CVC unit; as investments are made through the external CVC unit Intel Capital. Hereby, advantages of an internal unit are potentially dampened, as both the parenting advantage and the protection effects are less important if innovations are less valuable. When looking at firm specificity, the case of Intel is not in line with the obtained results in this thesis, as the share of backward self-citations is clearly below average, which implies that firm specificity is low. The anticipated effects of the NIH syndrome and the difficulty to absorb are not reflected in the data for the case of Intel. The same holds for the technological diversification of the company, which is classified as "low". Based on our results, technological diversification overall is positively related to the likelihood of an external unit. However, different theorized effects pointed in opposite directions. In this specific case, the effect of path dependency might be especially important, which is theorized to increase the likelihood of an external unit in cases where technological diversification is low. The other theorized effects do not hold in this case specifically. In short, Intel as the biggest corporate investor does not exemplify the results well.

Netscape, on the other hand, exemplifies the results of this thesis well. It counts 2565 total forward citations to its granted patents until the maximum year of investment, which classifies its innovations as highly valuable. In the analysis, we find a negative relationship between the value of innovations and the likelihood of an external unit, which is confirmed in the example of Netscape, investing through an internal CVC unit. This could imply that Netscape indeed aims to leverage benefits through the parenting advantage and to protect its innovations. Similarly, firm specificity is low, as the share of backward self-citations amounts to only 1.70% (compared to a median value of 2.00%, rounded to 2-digits). This is in line with the finding of a positive relationship between firm specificity and an external CVC unit in this thesis, as the data suggests that the NIH-syndrome, in accordance with our theorizing, might not be present. Moreover, the data suggests that Netscape does not have a *difficulty to absorb*. Together, we suggest that this could be causing a low need to have an external unit. Furthermore, the company shows a low degree of technological diversification, which dampens the effects of search, coordination and bureaucratic costs, decentralization and replication. The example of Netscape consequently exemplifies the relationship between all the independent variables in the model and the likelihood of an external unit well.

Microsoft shows both a high value of innovations as well as a very low firm specificity, while investing through an internal CVC unit. Specifically, forward citations amount to 4877 by 2015 and the share of backward self-citations revolves around 0.5% throughout the investment period. This is in line with the result of our analysis. The theorizing suggests that this could be caused by the high value of innovations enabling an internal CVC unit to leverage benefits from the parenting advantage while protecting its innovations. As firm specificity is low, there is no need to set-up the CVC unit externally due to the NIH-syndrome and a difficulty to absorb, in line with the suggested explanations of this thesis. On the contrary, Microsoft is classified as highly technologically diversified, which is positively related to the likelihood of an external CVC unit according to our findings.

In summary, while Netscape Communications Corp. exemplifies the results of this thesis well in all aspects, Intel and Microsoft deviate to some extent. However, it is important to note that the probit model reports coefficients conditional on the other factor variables in the model, and that the strength of the coefficients differ along the curve, as the relationship is essentially non-linear. This implies that effects cannot be regarded (and exemplified) completely isolated from one another. Furthermore, as results were derived for a larger sample and hence represent an average, it is only natural that the three case companies deviate to some extent. However, they still enhance the understanding of the results.

7. DISCUSSION

In the following section, the results and potential areas of future research will be discussed. Firstly, the most important results of the analysis will be summarized briefly. Then, the validity of the results will be discussed. This will include potential biases that might affect the results (e.g. data issues such as truncation of the forward citations) and other data-related issues. Furthermore, identified potential problems with the model developed in this thesis will be addressed, namely robustness and specification. With regards to the latter, alternative antecedents will be discussed, which can potentially predict whether organizations that engage in CVC will invest through an internal or external unit (i.e. other variables that can enhance the quality of the model). Lastly, suggestions for further research and the implications of this thesis (i.e. for whom is it relevant) will be set out.

7.1. Summary of results

As shown in the results section, we find a significant relationship between forward citations (proxy for value of innovations) and the subsidiary variable. Specifically, the number of forward citations is significantly ($p<0.01$) and negatively related to the subsidiary variable. We suggest to explain this with a *protection effect* and a *parenting advantage,* as set out in previous sections. Based on the theoretical explanation and the observed significant relationship, the value of innovations can be seen as an antecedent to the set-up of an internal or external CVC unit.

Moreover, we find a significant relationship between the share of backward self-citations (proxy for firm specificity) and the subsidiary variable. Specifically, the share of backward self-citations is significantly ($p<0.05$) and positively related to the subsidiary variable. We explain this with the *NIH-syndrome* and a *difficulty to absorb*. On the basis of these theoretical explanations and the observed significant correlation, we propose that organizations with a higher degree of firm specificity of innovations are more likely to set up an external CVC unit when deciding to engage in CVC.

Furthermore, we find a significant relationship between the standard deviation of patent dispersion in different USPC classes of parent organizations (proxy for technological diversification) and the subsidiary variable. Specifically, the coefficient is positive and significant ($p<0.1$). We explain this relationship with *search, coordination and bureaucratic costs* and a *replication effect*. Consequently, organizations that are more technologically diverse seem more likely to set up an external CVC unit when engaging in CVC.

For the model developed, we find no significant differences between our main industries, Computer Programming, Data Processing, And Other Computer Related Services (sic_3=737, 68 observations), Drugs (sic_3=283, 36 observations) and Electronic Components and Accessories (sic_3=367, 24 observations). As laid out in the results section, the relatively few observations in industries 283 and 367 could potentially be responsible for the lack of industry-difference, which compels further more in-depth industry-by-industry studies.

Moreover, we find that organizations' degree of technological diversification does not significantly moderate the relationship between the value of innovations or firm specificity, respectively, and the likelihood for organizations that engage in CVC to set up an internal or an external unit.

7.2. Validity of results

The data collection, application and usage were described in detail in the methodology and results sections. In the following, certain issues with regards to data will be discussed, namely patent and subsidiary data. With regards to the patent data, truncation of forward citations, transformations and general issues will be discussed. Following, other data-related issues concerning the information on the CVC unit and its connection to the parent organization will be addressed.

7.2.1. Biases related to patent data

Forward citations are employed as a proxy for the value of innovations. Forward citations naturally occur after a patent is created and applied for (and many of them after the granted date). This means that the number of forward citations for older patents are likely to fairly reflect the total number these patents will receive and thus is a good proxy of the value of innovations. However, since the data suffers from a natural truncation (patent data only available until 2017), newer patents will not accurately reflect the number of forward citations these patents are likely to receive (which would more accurately proxy the *value*). This truncation issue poses an issue to the integrity of the forward citations variable. The magnitude of the issue is hard to assess, as many of the observations have patents throughout much of the period (i.e. the issue might be distributed across observations). While this is a natural problem when dealing with patent data, it is one that is important to be aware of. Some other papers (see e.g. Hall et al., 2005) apply correction methods, but these are also to some degree problematic. One way is to estimate a normal distribution for the forward citations and thus "forecast" future forward citations (which might not fairly reflect the actual number of forward citations the patents will receive). However, as Hall et al. (2005) point out, even after 25 years, new forward citations keep coming at a "non-declining rate", which makes such corrections difficult to perform and conjectural, at best. We have not utilized correction methods in this thesis for the forward citations variable. Consequently, the results with regards to the forward citations variable might therefore suffer from a *truncation bias*. It should be noted that similar corrections could be applied to e.g. backwards citations in the other end of the period.

A similar issue arises with all patent variables for which we use “applied-for” patents. Since the database only includes applied-for patents that are later granted, some patents are likely left out in the last years of available data, as they have not been granted yet. Also here a correction could be applied, if all applied-for patents had been used (e.g. an application-grant empirical distribution, which could help compute a “weight factor” of the applied-for patents, that would likely be granted, as suggested by Hall et al. (2005). We have not corrected such truncation issues, as they have been deemed minor, and since the correction methods are also flawed and would thus introduce similar bias. The reason that they are deemed minor is that we have more available patent data than firm-level data (by two years), and most patents take 2-3 years to be granted (Hall et al., 2005). Therefore, only patents that took relatively long to be granted could pose an issue, and only in the very last years of the analysed data. The arguments for choosing applied-for rather than granted patents can be found in the methodology section.

With regards to transformations, we have logged variables where necessary due to skewness (as previously argued). No variables have been lagged, as this has not been deemed necessary. In general, as much of the “natural” variance has been kept as possible. The most radical transformation performed is transforming the standard deviation of patent dispersion on USPC classes (proxy for technological diversification) into a binary variable. The transformation into a binary variable was made on the basis of the median value of the variable. While it could be argued that the mean would be more appropriate, the median offers a benefit: it splits the sample at a point that ensures that the natural “high” and “low” values are respected, and further splits the sample in two equal parts. An argument can be made that some variance is lost when performing a binary transformation. However, in this case a binary variable made sense from a theoretical perspective (as argued), and there were correlation issues hindering a correct model specification. Consequently, a binary transformation was deemed appropriate. As also shown, this predictor is the weakest, and lost significance in one of the robustness checks, which might be caused by this transformation, as it results in some loss of variance. In the interactions, we tested creating the binary variables at different percentiles than the median, including the upper and lower 90th and 95th percentile, but no significant differences were found. For this reason, variables for the interaction were not modified compared to the basic model, i.e. we used a binary variable for the standard deviation, and continuous variables for both forward citations and backward self-citations.

Moreover, the technological diversification proxy, standard deviation of patent dispersion on USPC classes can be discussed; is it a fair measure of technological diversification on its own? In an extreme "thought-up" case, a company could display a high variable value by being active in only two classes; one with a lot of patents and one with only one patent. This would certainly not be a technologically diversified company. While this extreme (and unlikely) case highlights a potential logical fallacy of the variable, the variable is functional in practice. As already described, we controlled for the number of distinct USPC classes (which was later omitted due to correlation issues). Furthermore, we checked the 90^{th} and the 95^{th} percentile of the standard deviation variable values: Companies with a high value for the technological diversification proxy variable (standard deviation) indeed have the highest number of distinct USPC classes as well. Hence, we assess that the variable indeed is a good proxy for technological diversification, while being more nuanced than a simple count of distinct USPC classes, as dispersion is taken into account.

For patent data in general, there are certain issues. As a number of researchers point out (see e.g. Patel & Pavitt, 1997), patent data only captures codified knowledge. Since we are using patent data to proxy firm specificity, technological diversification, value of innovations and absorptive capacity, this issue is general for this thesis. Some companies might possess more tacit knowledge (Patel & Pavitt, 1997) while having the same patent values as another company. Moreover, some companies might not want to disclose their innovations in patents due to a risk of imitation (Lee et al., 2018). However, patents are still an invaluable source of innovation and knowledge-related information, which is supported by the magnitude of research that relies on patent data (e.g. Arora et al., 2014; Dushnitsky & Lenox, 2005; Lee et al., 2018).

7.2.2. Biases related to subsidiary data

As described, a portion of the data was retrieved from Compustat, Thomson One Banker and through several rounds of clerical review. There are certain potential issues with these parts of the data as well.

Firstly, we decided to weigh each CVC unit equally. This means that each investment entity has a weight of one in the econometric analysis, even when several units belong to the same parent organization. Consequently, organizations with more than one CVC unit are “counted” several times. Specifically, the patent characteristics of an organization with two internal units will constitute two equally important analytical elements in the probit model. The question is: is this reasonable? There are a number of reasons behind the decision to treat the analysis this way; (i) the CVC units are the main targets of our analysis (and not the organizations) (ii) you can argue the parent organization essentially made the decision more than once (some have external and internal units), and this method therefore accurately treats the CVC units separately and (iii) most organizations only have one unit, and the issue is therefore not grave – specifically, for 161 CVC units in the three industries in the US, 144 unique organizations are identified. We contemplated weighting the units per organization, which would offset some of the issues. However, this would produce a number of other issues (e.g. a 0.5 subsidiary value). Alternatively, this research could be designed in form of panel data if not the CVC unit, but the parent organization is the level of analysis – and changing values of *subsidiary* could be accounted for. This could be addressed by future research of the topic.

Secondly, while the subsidiary variable provides an intuitive and manageable unit for analysis that is great for investigating the *antecedents* to the parent organizations behaviour in terms of granting autonomy to the CVC unit, it is also fairly coarse. As Lee et al. (2018) point out, two identical *subsidiary* values (e.g. 1 and 1) might reflect different levels of autonomy, since autonomy is complex; it consists of a number of spectrums (decision-making, financial resources etc., see the literature review in section 2.3.3). An organization can exercise influence over a wholly-owned subsidiary, as it can exercise control over an internal unit. Therefore, the subsidiary variable is an apt, yet simplified measure. On the other hand, this measure has the virtue of being comparable and objective, as it is independent from, mostly self-reported, questionnaires and their inherent bias. While previous studies have used more fine-grained measurements of autonomy (through e.g. surveys), they have not been used to investigate the antecedents to autonomy. Future research could delve deeper into this aspect.

7.2.3. Model robustness and specification

Two general issues with econometric models, that are relevant in this context, are (i) robustness and (ii) under-specification. Moreover, the specific specification of the variable for *number of investments* will be discussed.

Firstly, as laid out in the results section, three robustness checks were performed, (i) alternative model specifications, (ii) including only CVC units with more than one investment and (iii) excluding the extreme tail. While these robustness checks add validity to the analysis, other robustness checks could add significant value. Specifically, using alternative variables that proxy the same theoretical concepts is a widespread technique that ensures that the model is robust. In this analysis, the value of innovations could, for instance, be proxied by R&D stock (Hall et al., 2005). Similarly, other proxies could be used for robustness checks for the other independent variables. While this would add further validity to the analysis, it was not deemed essential in the scope of this thesis. Future studies could be conducted using alternative measures in order to confirm the findings of this thesis.

Moreover, while our model is testing theorised relations between the subsidiary variable and specific theoretical concepts, the specific effects that are theorized to be in play (e.g. *protection effect*, *parenting advantage* and *decentralization effect*) cannot be disentangled with this analysis' level of granularity. This thesis aims at opening the discussion on the antecedents to the set-up of an internal or external CVC unit with a specific focus on innovation through empirical analysis. However, case-oriented or survey-based approaches would be better suited for adding depth in terms of granular understanding of these rather "broad", theorized effects.

Secondly, there could be alternative antecedents to the setting up CVC units internally or externally, risking an under-specification of the model. While a number of control variables were used, a number of other factors might also play roles as antecedents to the organizational setup of CVC units as internal or external units. These could include e.g. firm size (measured in headcount or financial metrics such as revenue, Free Cash flow etc.), manager experience, leadership, culture, regional differences. Moreover, risk profile might play a role. As a subsidiary effectively decreases the downside risk involved with any transaction (e.g. in case of bankruptcy) and since some companies might use CVC investments as a financial diversification mechanism, different risk profiles across parent companies might induce different behaviour in terms of weighing the benefits/disadvantages of an internal or an external unit, respectively.

Other papers have highlighted the link between the aim of the corporate venturing unit and the organizational setup (see e.g. Hill & Birkinshaw, 2014). While the context is slightly different, it seems reasonable that CVC activities can be characterized similarly. This means that the *goal* of the management team that decides to engage in CVC can be an *antecedent* to its setup as an internal or external unit. While a goal is difficult to observe, this could potentially be operationalized in an analysis via e.g. survey data.

As shown in the results section, the variable that behaved most consistently was the *number of investments*. It is positively related and highly significant ($p<0.01$). While the number of investments was specified as a control variable rather than an independent variable, no in-depth theorizing has been performed for this variable.

However, there are a number of logical explanations for this behaviour: Firstly, when a firm anticipates a high investment activity, it is more likely to want to have a dedicated unit. If you anticipate making a single investment, it might not be worth setting up a unit for it – you can handle it internally more easily. Hence, a high level of CVC activity might indicate a continuous dedication of resources, which might manifest itself in setting up a subsidiary. Secondly, dedicated external units might be able to attract more investments, as they are better rooted in the VC and start-up communities and can reap reputational benefits (as has been previously described). Note that there is a risk of a chicken-and-egg logic here; do they set up the external unit to invest more, or do they invest more because they set up an external unit?

Another explanation (that is not mutually exclusive with the first explanation) might be that the number of investments to a high degree is related to firm size – larger firms (in terms of financial metrics such as Free Cash Flow, market cap et cetera) are likely to make more investments. Larger firms are more likely to be experienced in setting up subsidiaries – and doing so might be "business as usual".[33] To state it more clearly; the number of investments variable might both be proxying size and resource-dedication. Both effects are likely to have a positive effect on the likelihood of setting up an external unit, which might explain the high degree of significance and the positive coefficient.

[33] The authors of this thesis attended a conference where a CVC practitioner from a large, Danish bank was describing the work of his department. On the question of why they set the unit up internally, he was puzzled – there had been no discussion of setting up an external unit. They were only making a few investments, so why would they? While anecdotal, this situation highlights an interesting dynamic: firms do not necessarily make an active decision.

7.3. Suggestions for future research

A number of suggestions for further research have already been made. As set out in the introduction, this paper aims at opening a debate within the field of CVC on the antecedents to the choice of setting up a CVC unit internally or externally. A number of factors in this paper need further corroboration, nuancing and a finer-grain lens to gain a thorough understanding (e.g. disentangling effects as previously mentioned). Survey data, case studies and other variables used in quantitative studies could help provide such enhanced understanding.

More importantly, this thesis did not provide conclusive causal explanations, but rather suggestions in terms of theoretical explanations. Further studies could aim to establish conclusive causal links.

As laid out in the discussion, further robustness checks would enhance the validity of the results. Specifically, using alternative measures for the independent variables might provide an enhanced understanding of the quality of the antecedents investigated. Seeing that technological diversification was less robust than the other measures, this concept could fruitfully be investigated more in-depth.

Moreover, the findings are the fruits of our efforts to investigate innovation-related data as antecedents to the decision to engage in CVC via an internal or an external unit. A number of other avenues of interest could be explored. These include, but are not limited to, investigating firm size, free cash flow, number of employees, experience of specific employees, motives behind engaging in CVC (although hard to observe) and level of risk aversion as antecedents. A more nuanced theoretical explanation behind the significant relation between the number of investments and the likelihood of setting up an external CVC unit could also enhance the understanding of the dynamics.

Moreover, as the descriptive statistics suggested, there are great differences between countries. For instance, none of the Swedish units in our sample were internal. This warrants further study: is investment culture also playing a role, or is local legislation driving country differences in organizational setups of CVC units? The same goes for industry. While this study does not find any significant differences across our three main industries, the evidence is not conclusive, and more in-depth research is warranted.

As the data has also suggested, a few players are responsible for a very large portion of the collective US CVC activities. This suggests that further studies can benefit relatively much from case-based research methods, as these can capture a large share of the investment activities.

Lastly, research that connects the antecedents of the an internal or external CVC setup to performance would be of great value to practitioners and enhance the understanding of the field.

7.4. Implications

As described, this thesis aims at contributing to the existing literature on CVC by bridging a research gap regarding antecedents to the choice of setting up an external or internal CVC unit. However, the conclusions of this thesis will have a number of other non-academic implications as well. This section will briefly describe for whom this thesis may be relevant.

While this thesis does not investigate the *performance* of internal and external CVC units, the thesis still holds managerial implications: An increased understanding of the fact that certain organizational dimensions influence the setup of CVC units as internal or external will benefit managers of corporations engaging in or contemplating engaging in CVC activities. Specifically, decision-making managers and investment professionals can look to this thesis for guidance as to *understanding the dynamics* of internal and external units, respectively.

Moreover, startups contemplating accepting funding from CVC investors may also benefit from a deeper understanding of the dynamics of the setup as internal or external CVC units of potential investors, as it is, to some degree and on average, related to the investment relationship (e.g. resource sharing with parent organization may be less easily facilitated when the investment is executed through an external unit, as suggested in this thesis). The same argument can be applied to other VCs and CVC units contemplating investing in syndication.

It is important to emphasize that this thesis does not offer conclusive evidence on causal relations but rather suggests theoretical explanations to the observed significant relations. This is why the above arguments related to *understanding the dynamics* of internal and external CVC units should be considered as a *first step* in this direction, which, hopefully, will be further investigated in the future.

8. CONCLUSION

This thesis set out to bridge a research gap by opening the debate on the antecedents to the choice of setting up of corporate venture capital (CVC) units as internal or external units. Specifically, the aim is to investigate organizational, knowledge-related dimensions of the parent companies, which may function as predictors as to whether CVC units are set up internally or externally, respectively.

Drawing on existing CVC literature, innovation literature and dominant theories in strategic management, the following four organizational dimensions were theorized to have an influence on the setup of internal or external CVC units: the value of innovations, firm specificity of innovations, absorptive capacity and technological diversification. Moreover, technological diversification was theorized to have a moderating effect on the relationship between the setup of CVC units and both value of innovations and firm specificity, respectively.

Through an empirical analysis based on a sample of data on CVC unit's investment activities in the years 1985 to 2015 and the parent organization's patenting activities in the years 1976 to 2017, a number of findings were made. The data was retrieved from Compustat, Thomson One Banker and PatentsView, manually merged and enriched through several rounds of clerical review for more than 1400 investment unit-level observations, of which 161 US-based investment units in the pharmaceutical, semiconductor and IT software industries comprised the final sample.

The value of innovations, firm specificity and technological diversification are all found to be significantly related to the main dependent variable of this thesis, the internal or external setup of CVC units (dichotomous variable, internal or external) and can thus be confirmed as antecedents. Specifically, the value of innovations, proxied by the number of forward citations of the patents of the parent organization of the CVC unit, is negatively related to the likelihood of setting up an external CVC unit. The relation is highly significant ($p<0.01$). This finding is suggested to be influenced by a *protection effect* and a *parenting advantage.* Firm specificity, proxied by the share of backward self-citations in total citations of the parent organization's patents, is positively related to the likelihood of setting up an external CVC unit. The relation is significant ($p<0.05$). We propose this relationship to be characterized through the *not-invented-here*-syndrome and a *difficulty to absorb*. Technological diversification, proxied by the standard deviation of patent dispersion in different USPC classes, is positively and significantly ($p<0.1$) related to the likelihood of setting up an external CVC unit. We propose the *replication* effect and *transaction costs* effect as explanations.

Industry differences between the three industries were found to be insignificant, and the theorized moderating effect of technological diversification was also found to be inconclusive. The results for the relationship between the setup of CVC units as internal or external and both value of innovations and firm specificity withstood robustness tests well. However, technological diversification became insignificant under certain circumstances, suggesting a less clear result. Interestingly, the control variable for number of investments of the CVC unit was significantly positively related to the main dependent variable under all circumstances.

This thesis concludes with suggestions for further research. Most importantly, this *first step* towards an understanding of the antecedents to the choice of setting up a CVC unit internally or externally should be investigated more in-depth – both with regards to alternative antecedents and finer-grain analysis of the theorized effects.

BIBLIOGRAPHY

Abbott, M. L., & McKinney, J. (2013). *Understanding and Applying Research Design*. Somerset, United States: John Wiley & Sons, Incorporated. Retrieved from http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=1120726

Ahuja, G., & Lampert, C. M. (2001). Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions. *Strategic Management Journal*, *22*(6/7), 521–543. https://doi.org/10.1002/smj.176

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Albert, J. H., & Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, *88*(422), 669. https://doi.org/10.2307/2290350

Aldrich, J., & Nelson, F. (1984). *Linear Probability, Logit, and Probit Models*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc. https://doi.org/10.4135/9781412984744

Argote, L. (2013). *Organizational learning: creating, retaining, and transferring knowledge* (Second edition). New York: Springer.

Arora, A., Belenzon, S., & Rios, L. A. (2014). Make, buy, organize: The interplay between research, external knowledge, and firm structure. *Strategic Management Journal*, *35*(3), 317–337. https://doi.org/10.1002/smj.2098

Arora, A., & Gambardella, A. (2010). Ideas for rent: an overview of markets for technology. *Industrial & Corporate Change*, *19*(3), 775–803. https://doi.org/10.1093/icc/dtq022

Artz, K. W., & Brush, T. H. (2000). Asset specificity, uncertainty and relational norms: an examination of coordination costs in collaborative strategic alliances. *Journal of Economic Behavior & Organization*, *41*(4), 337–362. https://doi.org/10.1016/S0167-2681(99)00080-3

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1093/icc/dtq022

Bauer, H., Burkacky, O., Kupferschmidt, J., & Rocha, A. (2017). From hardware to software: How semiconductor companies can lead a successful transformation. *McKinsey on Semiconductors*, *6*, 65–72.

Bauer, H., Kenevan, P., Patel, M., & Santhanman, N. (2017). Introduction. *McKinsey on Semiconductors*, *6*, 59–64.

Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2012). *Basic Business Statistics: Concepts and Applications* (12th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Berkson, J. (1980). MINIMUM CHI-SQUARE, NOT MAXIMUM LIKELIHOOD! *The Annals of Statistics*, *8*(3), 457–487. https://doi.org/10.1214/aos/1176345003

Beyers, T. (2016, November 3). Intel Capital: A Study In Long-Term Success. Retrieved 1 July 2018, from https://www.rocketspace.com/corporate-innovation/intel-capital-a-study-in-long-term-success

Birkinshaw, J., & Hill, S. A. (2005). Corporate Venturing Units: Vehicles for Strategic Success in the New Europe. *Organizational Dynamics*, *34*(3), 247–257. https://doi.org/10.1016/j.orgdyn.2005.06.009

Birkinshaw, J., van Basten Batenburg, R., & Murray, G. (2002). Venturing to succeed. *London Business School Review*, *13*(4), 10–17. https://doi.org/10.1111/1467-8616.00229

Booth, B. (2011, September 25). Debunking Corporate Venture Capital in Biotech. *Forbes*. Retrieved from https://www.forbes.com/sites/brucebooth/2011/09/25/debunking-corporate-venture-capital-in-biotech/

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, *32*(1), 69–87. https://doi.org/10.1016/S0048-7333(02)00004-5

Burgelman, R. A. (1983). A Process Model of Internal Corporate Venturing in the Diversified Major Firm. *Administrative Science Quarterly*, *28*(2), 223–244. https://doi.org/10.2307/2392619

Burgelman, R. A. (1984). Designs for corporate entrepreneurship in established firms. *California Management Review*, *26*(3), 154–166. https://doi.org/10.2307/41165086

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics Using Stata*. College Station, Texas: Stata Press. Retrieved from https://www-dawsonera-com.esc-web.lib.cbs.dk:8443/abstract/9786000038984

Cameron, A. C., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, *77*(2), 329–342. https://doi.org/10.1016/S0304-4076(96)01818-0

CB Insights. (2017a). *The H1'17 Global CVC Report*. CB Insights. Retrieved from https://www.cbinsights.com/research/report/corporate-venture-capital-trends-2017-h1/

CB Insights. (2017b). *The History Of CVC: From Exxon And DuPont To Xerox And Microsoft, How Corporates Began Chasing 'The Future'*. Retrieved from https://www.cbinsights.com/research/report/corporate-venture-capital-history/

CB Insights. (2018a). *The Top 20 Corporate Venture Capital Firms*. Retrieved from https://www.cbinsights.com/research/top-corporate-venture-capital-investors/

CB Insights. (2018b, February 28). The Most Active Corporate VC Firms Globally. *CB Insights Research*. Retrieved from https://www.cbinsights.com/research/corporate-venture-capital-active-2014/

Chandler, A. D. (1962). *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Washington D.C.: Massachusetts Institute of Technology.

Chesbrough, H. (2000). Designing Corporate Ventures in the Shadow of Private Venture Capital. *California Management Review*, *42*(3), 31–49. https://doi.org/10.2307/41166041

Chesbrough, H. W. (2002). Making Sense of Corporate Venture Capital. *Harvard Business Review*, *80*(3), 4–11.

Chitkara, R., Gloger, M., & McCaffrey, M. (2018). *Technology Trends 2018-19* (PwC Strategy&). PwC. Retrieved from https://www.strategyand.pwc.com/trend/2018-technology

Cirillo, B., Brusoni, S., & Valentini, G. (2013). The Rejuvenation of Inventors Through Corporate Spinouts. *Organization Science*, *25*(6), 1764–1784. https://doi.org/10.1287/orsc.2013.0868

CNET. (1996, March 21). Netscape invests in voice processing. *CNET*. Retrieved from https://www.cnet.com/news/netscape-invests-in-voice-processing/

Coase, R. H. (1937). The Nature of the Firm. *Economica*, *4*(16), 386–405. https://doi.org/10.2307/2626876

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152. https://doi.org/10.2307/2393553

Corcoran, E. (1998, November 25). AOL CLOSES DEAL TO BUY NETSCAPE FOR $4.2 BILLION. *Washington Post*. Retrieved from https://www.washingtonpost.com/archive/politics/1998/11/25/aol-closes-deal-to-buy-netscape-for-42-billion/0b258c99-d4e7-4f65-8b36-46aa2ce97ca6/

Crockett, D. R., McGee, J. E., & Payne, G. T. (2013). Employing New Business Divisions to Exploit Disruptive Innovations: The Interplay between Characteristics of the Corporation and Those of the Venture Management Team. *Journal of Product Innovation Management*, *30*(5), 856–879. https://doi.org/10.1111/jpim.12034

Cumming, D. (2012). *The Oxford Handbook of Venture Capital*. USA: Oxford University Press.

Daft, R. L. (1983). *Organization theory and design*. St. Paul: West Publishing Company.

David, E., Mehta, A., Norris, T., Singh, N., & Tramontin, T. (2010, February). New frontiers in pharma R&D investment. *McKinsey & Company: Pharmaceuticals & Medical Products*. Retrieved from https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/new-frontiers-in-pharma-r-and-38d-investment

de Backer, K., Mancini, M., & Sharma, A. (2017). Optimizing back-end semiconductor manufacturing through Industry 4.0. *McKinsey on Semiconductors*, *6*. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our%20Insights/McKinsey%20on%20Semiconductors%20Issue%206%20-%20Spring%202017/McK%20on%20Semiconductors_Issue%206_2017.ashx

de Vaus, D. A. (2001). What is Research Design? In *Research Design in Social Research*. London: SAGE Publications. Retrieved from http://dx.doi.org/10.4135/9781446263259

Demski, J. S., & Feltham, G. A. (1978). Economic Incentives in Budgetary Control Systems. *The Accounting Review*, *53*(2), 336–359.

Di Lorenzo, F., & Almeida, P. (2017). The role of relative performance in inter-firm mobility of inventors. *Research Policy*, *46*(6), 1162–1174. https://doi.org/10.1016/j.respol.2017.05.002

Dushnitsky, G. (2008). Corporate venture capital: Past evidence and future directions. In A. Basu, M. Casson, N. Wadeson, & B. Yeung (Eds.), *The Oxford Handbook of Entrepreneurship*. Oxford, UK: Oxford University Press.

Dushnitsky, G., & Lenox, M. J. (2005). When do incumbents learn from entrepreneurial ventures?: Corporate venture capital and investing firm innovation rates. *Research Policy*, *34*(5), 615–639. https://doi.org/10.1016/j.respol.2005.01.017

Dushnitsky, G., & Lenox, M. J. (2006). When does corporate venture capital investment create firm value? *Journal of Business Venturing*, *21*(6), 753–772. https://doi.org/10.1016/j.jbusvent.2005.04.012

Dushnitsky, G., & Shaver, J. M. (2009). Limitations to interorganizational knowledge acquisition: the paradox of corporate venture capital. *Strategic Management Journal*, *30*(10), 1045–1064. https://doi.org/10.1002/smj.781

Eisenhardt, K. M. (1985). Control: Organizational and Economic Approaches. *Management Science*, *31*(2), 134–149. https://doi.org/10.1287/mnsc.31.2.134

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, *14*(1), 57–74. https://doi.org/10.2307/258191

Fabrizio, K. R. (2009). Absorptive capacity and the search for innovation. *Research Policy*, *38*(2), 255–267. https://doi.org/10.1016/j.respol.2008.10.023

Fast, N. D. (1979). The future of industrial new venture departments. *Industrial Marketing Management*, *8*(4), 264–273. https://doi.org/10.1016/0019-8501(79)90032-4

Foley, M. J. (2018, April 30). Microsoft renames its Ventures unit 'M12'. Retrieved 3 July 2018, from https://www.zdnet.com/article/microsoft-renames-its-ventures-unit-m12/

Garrett, R. P., & Neubaum, D. O. (2013). Top Management Support and Initial Strategic Assets: A Dependency Model for Internal Corporate Venture Performance. *Journal of Product Innovation Management*, *30*(5), 896–915. https://doi.org/10.1111/jpim.12036

Granstrand, O. (1998). Towards a theory of the technology-based firm. *Research Policy*, *27*(5), 465–489. https://doi.org/10.1016/S0048-7333(98)00067-5

Grant, R. M. (2016). *Contemporary Strategy Analysis: Text and Cases Edition* (7th ed.). West Sussex: John Wiley & Sons.

Greene, W. H. (2014). *Econometric Analysis: International Edition* (7th ed.). London: Pearson Education UK. Retrieved from http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=5137900

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*(4), 693–706. https://doi.org/10.5465/amj.2006.22083026

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools* (Working Paper No. 8498). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w8498

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, *36*(1), 16–38.

He, Z.-L., & Wong, P.-K. (2004). Exploration vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis. *Organization Science*, *15*(4), 481–494. https://doi.org/10.1287/orsc.1040.0078

Hearst Communications Inc. (1995, April 6). Netscape Communications Announces Investment by Publishing and Technology Companies. *Hearst Communications Inc*. Retrieved from http://www.hearst.com/newsroom/netscape-communications-announces-investment-by-publishing-and-technology-companies

Hill, S. A., & Birkinshaw, J. (2008). Strategy–organization configurations in corporate venture units: Impact on performance and survival. *Journal of Business Venturing*, *23*(4), 423–444. https://doi.org/10.1016/j.jbusvent.2007.04.001

Hill, S. A., & Birkinshaw, J. (2014). Ambidexterity and Survival in Corporate Venture Units. *Journal of Management*, *40*(7), 1899–1931. https://doi.org/10.1177/0149206312445925

Hill, S. A., Maula, M. V. J., Birkinshaw, J. M., & Murray, G. C. (2009). Transferability of the venture capital model to the corporate context: Implications for the performance of corporate venture units. *Strategic Entrepreneurship Journal*, *3*(1), 3–27. https://doi.org/10.1002/sej.54

Intel. (2018a). Our Focus. Retrieved 1 July 2018, from http://www.intelcapital.com/advantage/index.html

Intel. (2018b). Success Stories. Retrieved 3 July 2018, from http://www.intelcapital.com/success/

Intel. (2018c). The Power of Partnership. Retrieved 1 July 2018, from http://www.intelcapital.com/advantage/power-of-partnership.html

Intel Corp. (2018). *2017 Annual Report*. Retrieved from https://s21.q4cdn.com/600692695/files/doc_financials/2017/annual/Intel_Annual_Report_Final-3.20.pdf

Jensen, M. C., & Meckling, W. H. (1976). THEORY OF THE FIRM: MANAGERIAL BEHAVIOR, AGENCY COSTS AND OWNERSHIP STRUCTURE. *Journal of Financial Economics*, *3*, 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Jones, G. R., & Hill, C. W. L. (1988). Transaction Cost Analysis of Strategy-Structure Choice. *Strategic Management Journal*, *9*(2), 159–172. https://doi.org/10.1002/smj.4250090206

Kashyap, N. (2018, April 30). Introducing M12. *M12*. Retrieved from https://m12.vc/blogs/introducing-m12/

Katila, R., & Ahuja, G. (2002). SOMETHING OLD, SOMETHING NEW: A LONGITUDINAL STUDY OF SEARCH BEHAVIOR AND NEW PRODUCT INTRODUCTION. *Academy of Management Journal*, *45*(6), 1183–1194. https://doi.org/10.2307/3069433

Katz, R., & Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups. *R&D Management*, *12*(1), 7–20. https://doi.org/10.1111/j.1467-9310.1982.tb00478.x

Keil, T. (2002). *External Corporate Venturing: Strategic Renewal in Rapidly Changing Industries*. Westport, CT: Quorum Books.

Keil, T. (2004). Building external corporate venturing capability. *Journal of Management Studies*, *41*(5), 799–825. https://doi.org/10.1111/j.1467-6486.2004.00454.x

Lee, S. U., Park, G., & Kang, J. (2018). The double-edged effects of the corporate venture capital unit's structural autonomy on corporate investors' explorative and exploitative innovation. *Journal of Business Research*, *88*, 141–149. https://doi.org/10.1016/j.jbusres.2018.01.049

Lerner, J. (2013). Corporate Venturing. *Harvard Business Review*, (October). Retrieved from https://hbr.org/2013/10/corporate-venturing)

Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. *Strategic Management Journal*, *14*(Special Issue), 95–112. https://doi.org/10.1002/smj.4250141009

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *1*, 319–340. https://doi.org/10.1146/annurev.so.14.080188.001535

Lohr, S. (1997, June 9). If at First You Don't Succeed, Turn to Netscape or Microsoft for Help. *The New York Times*. Retrieved from https://www.nytimes.com/1997/06/09/business/if-at-first-you-don-t-succeed-turn-to-netscape-or-microsoft-for-help.html

M12. (2018). About. Retrieved 3 July 2018, from https://m12.vc/about/

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

MarketLine. (2012). *Global Semiconductors & Electronic Components* (MarketLine Industry Profile No. 0199–2503). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP0426-0001

MarketLine. (2018a). *Biotechnology in the United States* (MarketLine Industry Profile No. 0072–0695). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2736-0010

MarketLine. (2018b). *Global Biotechnology* (MarketLine Industry Profile No. 0199–0695). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2736-0006

MarketLine. (2018c). *Global Semiconductors* (MarketLine Industry Profile No. 0199–0682). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2645-0003

MarketLine. (2018d). *Global Software* (MarketLine Industry Profile No. 0199–0381). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2631-0009

MarketLine. (2018e). *Pharmaceuticals in the United States* (MarketLine Industry Profile No. 0072–0372). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2714-0025

MarketLine. (2018f). *Semiconductors in the United States* (MarketLine Industry Profile No. 0072–0682). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2645-0006

MarketLine. (2018g). *Software in the United States* (MarketLine Industry Profile No. 0072–0381). Retrieved from http://advantage.marketline.com.esc-web.lib.cbs.dk/Product?ptype=Industries&pid=MLIP2631-0025

Matthews, B., & Ross, L. (2010). *Research Methods*. Harlow, United Kingdom: Pearson Education UK. Retrieved from http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=5138361

Mawson, J. (2017, January 24). Global Corporate Venturing Rising Stars Awards 2017. *Global Corporate Venturing*. Retrieved from http://www.globalcorporateventuring.com/article.php/16039/global-corporate-venturing-rising-stars-awards-2017-contents

Mawson, J. (2018, May 21). GCV Powerlist 2018: #1 Wendell Brooks. *Global Corporate Venturing*. Retrieved from http://www.globalcorporateventuring.com/article.php/20411/gcv-powerlist-2018-1-wendell-brooks#/type=QWxs/page=0/term=/tags=

McFadden, D. (1977). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments. *Cowles Foundation Discussion Paper No. 474*, 1–47.

McNally, K. (1997). *Corporate Venture Capital: Bridging the Equity Gap in the Small Business Sector*. London; New York: Routledge Chapman & Hall.

Miles, M. P., & Covin, J. G. (2002). Exploring the Practice of Corporate Venturing: Some Common Forms and Their Organizational Implications. *Entrepreneurship Theory and Practice*, *26*(3), 21–40. https://doi.org/10.1177/104225870202600302

Netscape Communications Corp. (1996, January 31). Netscape announces new real-time audio and video framework for internet applications. Retrieved 4 July 2018, from https://www.cs.columbia.edu/~hgs/rtp/newsrelease81.html

Norr, H. R. (2017). Netscape Communications Corp. In *Encyclopedia Britannica*. Encyclopaedia Britannica, Inc. Retrieved from https://www.britannica.com/topic/Netscape-Communications-Corp

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, *4*(2), 154–167.

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

OSHA. (2018). *SIC division structure*. Occupational Safety and Health Administration, United States Department of Labor. Retrieved from https://www.osha.gov/pls/imis/sic_manual.html

Patel, P., & Pavitt, K. (1997). The technological competencies of the world's largest firms: Complex and path-dependent, but not much variety. *Research Policy*, *26*(2), 141–156. https://doi.org/10.1016/S0048-7333(97)00005-X

PatentsView. (2017). PatentsView Database. Retrieved 12 April 2018, from http://www.patentsview.org/

Pavitt, K. (1998). Technologies, Products and Organization in the Innovating Firm: What Adam Smith Tells Us and Joseph Schumpeter Doesn't. *Industrial and Corporate Change*, *7*(3), 433–452. https://doi.org/10.1093/icc/7.3.433

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Punch, K. F. (2014). *Introduction to Social Research: Quantitative and Qualitative Approaches* (3rd ed.). London: SAGE.

PwC, & CB Insights. (2018). *MoneyTree^{TM} Report - Q1 2018*. Retrieved from https://www.pwc.com/us/en/technology/assets/MoneyTree_Report_2018_Q1_FINAL.pdf

Quintana-García, C., & Benavides-Velasco, C. A. (2008). Innovative competence, exploration and exploitation: The influence of technological diversification. *Research Policy*, *37*(3), 492–507. https://doi.org/10.1016/j.respol.2007.12.002

Reaume, A. (2003). Is Corporate Venture Capital a Prescription for Success in the Pharmaceutical Industry? *The Journal of Private Equity*, *6*(4), 77–87. https://doi.org/10.3905/jpe.2003.320058

Rind, K. W. (1981). The role of venture capital in corporate development. *Strategic Management Journal*, *2*(2), 169–180. https://doi.org/10.1002/smj.4250020206

Ross, S. A. (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review*, *63*(2), 134–139.

Rotemberg, J. J., & Saloner, G. (1994). Benefits of Narrow Business Strategies. *The American Economic Review*, *84*(5), 1330–1349.

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2015). *Research Methods for Business Students* (7th ed.). London: Pearson Education UK. Retrieved from http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=5138717

Shapiro, S. P. (2005). Agency Theory. *Annual Review of Sociology*, *3,* 263–284. https://doi.org/10.1146/annurev.soc.31.041304.122159

Shelanski, H. A., & Klein, P. G. (1995). Empirical Research in Transaction Cost Economics: A Review and Assessment. *Journal of Law, Economics, & Organization*, *11*(2), 335–361.

Siegel, R., Siegel, E., & MacMillan, I. C. (1988). Corporate venture capitalists: Autonomy, obstacles, and performance. *Journal of Business Venturing*, *3*(3), 233–247. https://doi.org/10.1016/0883-9026(88)90017-1

Sorrentino, M., & Williams, M. L. (1995). Relatedness and corporate venturing: Does it really matter? *Journal of Business Venturing*, *10*(1), 59–73. https://doi.org/10.1016/0883-9026(94)00007-H

Souitaris, V., Zerbinati, S., & Liu, G. (2012). Which Iron Cage? Endo- and exoisomorphism in Corporate Venture Capital Programs. *Academy of Management Journal*, *55*(2), 477–505. https://doi.org/10.5465/amj.2009.0709

Standard & Poor's. (2018). Compustat Database. Retrieved 20 March 2018, from https://www.spglobal.com/marketintelligence/en/?product=compustat-research-insight

Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics, Update, Global Edition* (3rd ed.). London: Pearson Education UK. Retrieved from http://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=5138846

Sykes, H. B. (1986). The anatomy of a corporate venturing program: factors influencing success. *Journal of Business Venturing*, *1*, 275–293.

Sykes, H. B. (1990). Corporate Venture Capital: Strategies for Success. *Journal of Business Venturing*, *5*, 37–47. https://doi.org/10.1016/0883-9026(90)90025-O

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, *17*(S2), 27–43. https://doi.org/10.1002/smj.4250171105

Thomson Reuters. (2018). Thomson ONE Banker. Retrieved 20 March 2018, from http://banker.thomsonib.com/

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, Exploitation, and Financial Performance: Analysis of S&P 500 Corporations. *Strategic Management Journal*, *30*(2), 221–231. https://doi.org/10.1002/smj.738

Wasi, N., & Flaaen, A. (2015). Record Linkage using STATA: Pre-processing, Linking and Reviewing Utilities. *The Stata Journal*, *15*(3), 672–697.

Weber, C., & Weber, B. (2005). Corporate Venture Capital Organizations in Germany. *Venture Capital*, *7*(1), 51–73. https://doi.org/10.1080/1369106042000316350

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180. https://doi.org/10.1002/smj.4250050207

Williamson, O. E. (1979). Transaction-Cost Economics: The Governance of Contractual Relations. *The Journal of Law and Economics*, *22*(2), 233–261. https://doi.org/10.1086/466942

Windmeijer, F. A. G. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews*, *14*(1), 101–116. https://doi.org/10.1080/07474939508800306

Winters, T. E., & Murfin, D. L. (1988). Venture capital investing for corporate development objectives. *Journal of Business Venturing*, *3*(3), 207–222. https://doi.org/10.1016/0883-9026(88)90015-8

Yang, Y., Chen, T., & Zhang, L. (2016). Corporate venture capital program autonomy, corporate investors' attention and portfolio diversification. *Journal of Strategy and Management*, *9*(3), 302–321. https://doi.org/10.1108/JSMA-03-2015-0023

Zahra, S. A., & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. *The Academy of Management Review*, *27*(2), 185–203. https://doi.org/10.2307/4134351

APPENDICES

Appendix A: Input files from PatentsView

Table	Table Description	Data Element Name	Definition	Example	Years Present	Type
application	**Information on the applications for granted patent.**	id	application id assigned by USPTO	02/002761	all	varchar(36)
		patent_id	patent number	D345393	all	varchar(20)
		series_code	application series; "D" for some designs; (http://www.uspto.gov/web/offices/ac/ido/oeip/taf/filingyr.htm)	2	all	varchar(20)
		number	unique applicaiton identifying number	2002761	all	varchar(64)
		country	country this application was filed in	US	all	varchar(20)
		date	date of application filing	33959	all	date

Table	Table Description	Data Element Name	Definition	Example	Years Present	Type
patent	**Data concerning granted patents**	id	patent this record corresponds to	3930271	all	varchar(20)
		type	category of patent. Usually "design", "reissue", etc.	utility	all	varchar(100)
		number	patent number	3930271	all	varchar(64)
		country	country in which patent was granted (always US)	US	all	varchar(20)
		date	date when patent was granted	27765	all	date
		abstract	abstract text of patent	A golf glove is disclosed h…	all	text
		title	title of patent	Golf glove	all	text
		kind	WIPO document kind codes (http://www.uspto.gov/learning-and-resources/support-centers/electronic-business-center/kind-codes-included-uspto-patent)	A	all	varchar(10)
		num_claims	number of claims	4	all	int(11)
		filename	name of the raw data file where patent information is parsed from	pftaps19760106_wk01.zip	all	varchar(120)

Table	Table Description	Data Element Name	Definition	Example	Years Present	Type
rawassignee	**Raw assignee information as it appears in the source XML files**	uuid	unique id	0000p94wkezw94s8cz7dbxlvz	all	varchar(36)
		patent_id	patent number	5856666	all	varchar(20)
		assignee_id	unique assignee ID generated by the disambiguation algorithm	eaa92f175be7bfb71011f17eafb1e71f	all	varchar(36)
		rawlocation_id	assignee's location. If the assigne's sequence number is 1, then his/her location is the "location" of the related patent	orskbf54s58e97lkmw8na5rpx	all	varchar(128)
		type	assignee type (2 - US Company or Corporation, 3 - Foreign Company or Corporation, 4 - US Individual, 5 - Foreign Individual, 6 - US Government, 7 - Foreign Government, 8 - Country Government, 9 - State Government (US). Note: A "1" appearing before any of these codes signifies part interest)	2	2002 and After	int(4)
		name_first	first name, if assignee is individual	Thomas	all	varchar(64)
		name_last	last name, if assignee is individual	Bushey	all	varchar(64)
		organization	organization name if assignee is organization	U.S. Philips Corporation	all	varchar(256)
		sequence	order in which assignee appears in patent file	0	all	int(11)

Table	Table Description	Data Element Name	Definition	Example	Years Present	Type
uspatentcitation	**Citations made to US granted patents by US patents**	uuid	unique id	000007b7c0x3n9iy1othb9hz7	all	varchar(36)
		patent_id	patent number	9009250	all	varchar(20)
		citation_id	identifying number of patent to which select patent cites	8127342	all	varchar(20)
		date	date select patent (patent_id) cites patent (citation_id)	40940	all	date
		name	name of cited record	Boynton et al.	all	varchar(64)
		kind	WIPO document kind codes (http://www.uspto.gov/learning-and-resources/support-centers/electronic-business-center/kind-codes-included-uspto-patent)	B2	2002 and After	varchar(10)
		country	country cited patent was granted (always US)	US	all	varchar(10)
		category	who cited the patent (examiner, applicant, other etc)	cited by patent	2002 and After	varchar(20)
		sequence	order in which this reference is cited by select patent	622	all	int(11)

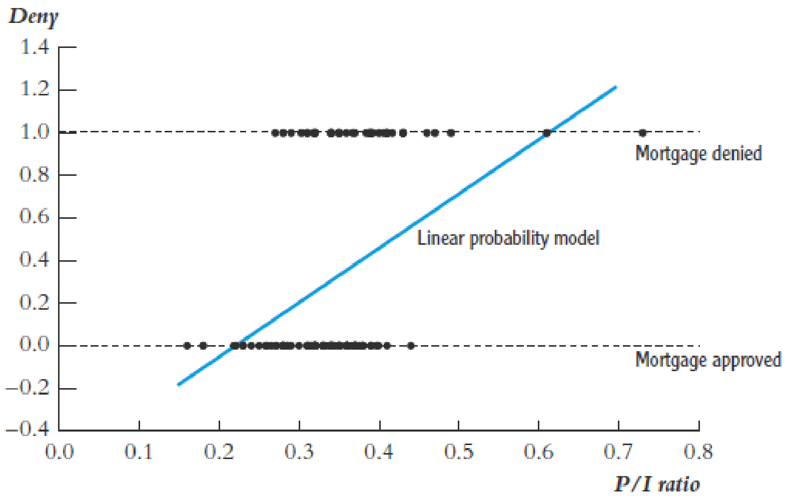
Table	Table Description	Data Element Name	Definition	Example	Years Present	Type
uspc_current	**USPTO current patent classification**	uuid	unique id	q0t5pp52pzpd41mauxa16eo	all	varchar(36)
		patent_id	patent number	3930271	all	varchar(20)
		mainclass_id	uspc mainclass current	2	all	varchar(20)
		subclass_id	uspc subclass current	2/161.4	all	varchar(20)
		sequence	order in which uspc class appears in patent file	0	all	int(11)

Appendix B: Linking elements for the patent database

variable PatentsView filename	patent_id	type	appdate	gdate	assignee_id	organization	num_claims	uspc (main class)	citation_id
patent	x	x		x			x		
application	x		x						
rawassignee	x				x	x			
uspc_current	x							x	
uspatentcitation	x								x

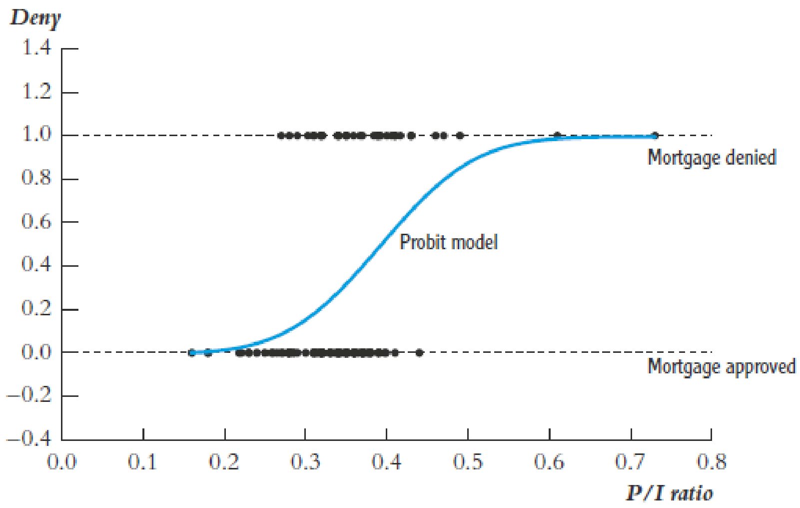
Appendix C: Difference between linear and probit regression with a dichotomous variable

Linear regression:



Source: (Stock & Watson, 2015, p. 433)

Probit regression:



Source: (Stock & Watson, 2015, p. 438)

Appendix D: CVC unit distribution on industries

SIC_broad	Freq.	Percent	Cum.
73	121	17.14	17.14
28	91	12.89	30.03
36	84	11.90	41.93
48	70	9.92	51.84
35	45	6.37	58.22
38	35	4.96	63.17
37	27	3.82	67.00
50	20	2.83	69.83
99	19	2.69	72.52
49	18	2.55	75.07
27	14	1.98	77.05
87	13	1.84	78.90
33	12	1.70	80.59
67	12	1.70	82.29
60	11	1.56	83.85
29	10	1.42	85.27
51	10	1.42	86.69
61	8	1.13	87.82
59	7	0.99	88.81
13	6	0.85	89.66
20	6	0.85	90.51
63	6	0.85	91.36
32	5	0.71	92.07
26	4	0.57	92.63
42	4	0.57	93.20
57	4	0.57	93.77
80	4	0.57	94.33
82	4	0.57	94.90
16	3	0.42	95.33
39	3	0.42	95.75
53	3	0.42	96.18
1	2	0.28	96.46
10	2	0.28	96.74
23	2	0.28	97.03
34	2	0.28	97.31
58	2	0.28	97.59
62	2	0.28	97.88
70	2	0.28	98.16
78	2	0.28	98.44
81	2	0.28	98.73
12	1	0.14	98.87
14	1	0.14	99.01

21	1	0.14	99.15
30	1	0.14	99.29
44	1	0.14	99.43
45	1	0.14	99.58
54	1	0.14	99.72
56	1	0.14	99.86
83	1	0.14	100.00
Total	706	100.00	

Appendix E: Subsidiary and industry distribution

	subsidiary		
SIC_broad	0	1	Total
73	84.30	15.70	100.00
28	64.84	35.16	100.00
36	85.71	14.29	100.00
48	68.57	31.43	100.00
35	82.22	17.78	100.00
38	65.71	34.29	100.00
37	51.85	48.15	100.00
50	35.00	65.00	100.00
99	63.16	36.84	100.00
49	33.33	66.67	100.00
27	64.29	35.71	100.00
87	46.15	53.85	100.00
33	58.33	41.67	100.00
67	66.67	33.33	100.00
60	54.55	45.45	100.00
29	50.00	50.00	100.00
51	60.00	40.00	100.00
61	87.50	12.50	100.00
59	100.00	0.00	100.00
13	66.67	33.33	100.00
20	83.33	16.67	100.00
63	83.33	16.67	100.00
32	80.00	20.00	100.00
26	25.00	75.00	100.00
42	75.00	25.00	100.00
57	75.00	25.00	100.00
80	100.00	0.00	100.00
82	50.00	50.00	100.00
16	33.33	66.67	100.00
39	100.00	0.00	100.00
53	66.67	33.33	100.00
1	100.00	0.00	100.00
10	100.00	0.00	100.00
23	0.00	100.00	100.00
34	100.00	0.00	100.00
58	100.00	0.00	100.00
62	0.00	100.00	100.00
70	50.00	50.00	100.00
78	100.00	0.00	100.00
81	50.00	50.00	100.00
12	100.00	0.00	100.00

14	100.00	0.00	100.00
21	0.00	100.00	100.00
30	100.00	0.00	100.00
44	100.00	0.00	100.00
45	100.00	0.00	100.00
54	0.00	100.00	100.00
56	100.00	0.00	100.00
83	0.00	100.00	100.00
Total	70.25	29.75	100.00

Note. Distribution of the variable *subsidiary* in per cent

Appendix F: Subsidiary and geographic distribution

Below are frequencies/percentages of subsidiary values per industry

	Subsidiary		
firmnation	0	1	Total
United States	378	102	480
	78.75	21.25	100.00
Japan	35	9	44
	79.55	20.45	100.00
Germany	8	17	25
	32.00	68.00	100.00
Canada	19	4	23
	82.61	17.39	100.00
United Kingdom	10	8	18
	55.56	44.44	100.00
France	5	10	15
	33.33	66.67	100.00
South Korea	3	11	14
	21.43	78.57	100.00
China	3	9	12
	25.00	75.00	100.00
Sweden	0	8	8
	0.00	100.00	100.00
Denmark	3	4	7
	42.86	57.14	100.00
Italy	3	4	7
	42.86	57.14	100.00
Switzerland	4	3	7
	57.14	42.86	100.00
Australia	4	2	6
	66.67	33.33	100.00
Israel	2	4	6
	33.33	66.67	100.00
Singapore	2	3	5
	40.00	60.00	100.00
Spain	5	0	5
	100.00	0.00	100.00
India	4	0	4
	100.00	0.00	100.00
Norway	2	2	4
	50.00	50.00	100.00
Ireland	2	1	3
	66.67	33.33	100.00
Hong Kong	0	2	2

	0.00	100.00	100.00
Netherlands	0	2	2
	0.00	100.00	100.00
Saudi Arabia	1	1	2
	50.00	50.00	100.00
Austria	0	1	1
	0.00	100.00	100.00
Cayman Islands	0	1	1
	0.00	100.00	100.00
Czech Republic	1	0	1
	100.00	0.00	100.00
Finland	0	1	1
	0.00	100.00	100.00
Luxembourg	1	0	1
	100.00	0.00	100.00
Taiwan	1	0	1
	100.00	0.00	100.00
Thailand	0	1	1
	0.00	100.00	100.00
Total	496	210	706
	70.25	29.75	100.00

Appendix G: Subsidiary and industry, US only

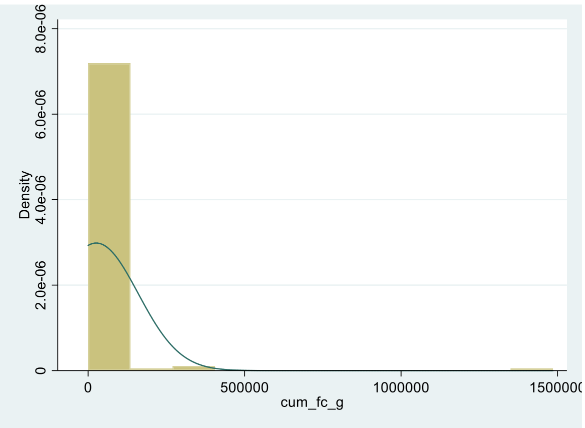
Below are frequencies/percentages of subsidiary values per industry

	subsidiary		
SIC_broad	0	1	Total
73	89	11	100
	89.00	11.00	100.00
36	55	6	61
	90.16	9.84	100.00
28	43	17	60
	71.67	28.33	100.00
48	31	8	39
	79.49	20.51	100.00
35	31	6	37
	83.78	16.22	100.00
38	18	9	27
	66.67	33.33	100.00
37	13	5	18
	72.22	27.78	100.00
27	8	3	11
	72.73	27.27	100.00
49	2	7	9
	22.22	77.78	100.00
51	5	4	9
	55.56	44.44	100.00
99	7	2	9
	77.78	22.22	100.00
50	2	6	8
	25.00	75.00	100.00
61	7	1	8
	87.50	12.50	100.00
87	5	3	8
	62.50	37.50	100.00
29	4	2	6
	66.67	33.33	100.00
59	6	0	6
	100.00	0.00	100.00
63	5	1	6
	83.33	16.67	100.00
13	4	1	5
	80.00	20.00	100.00
60	5	0	5
	100.00	0.00	100.00
33	2	2	4
	50.00	50.00	100.00
57	3	1	4

	75.00	25.00	100.00
67	4	0	4
	100.00	0.00	100.00
82	2	2	4
	50.00	50.00	100.00
42	2	1	3
	66.67	33.33	100.00
80	3	0	3
	100.00	0.00	100.00
1	2	0	2
	100.00	0.00	100.00
20	2	0	2
	100.00	0.00	100.00
26	1	1	2
	50.00	50.00	100.00
34	2	0	2
	100.00	0.00	100.00
39	2	0	2
	100.00	0.00	100.00
53	2	0	2
	100.00	0.00	100.00
58	2	0	2
	100.00	0.00	100.00
81	1	1	2
	50.00	50.00	100.00
12	1	0	1
	100.00	0.00	100.00
30	1	0	1
	100.00	0.00	100.00
32	1	0	1
	100.00	0.00	100.00
44	1	0	1
	100.00	0.00	100.00
45	1	0	1
	100.00	0.00	100.00
54	0	1	1
	0.00	100.00	100.00
56	1	0	1
	100.00	0.00	100.00
62	0	1	1
	0.00	100.00	100.00
70	1	0	1
	100.00	0.00	100.00
78	1	0	1
	100.00	0.00	100.00
Total	378	102	480
	78.75	21.25	100.00

Appendix H: Detailed summary of independent and control variables

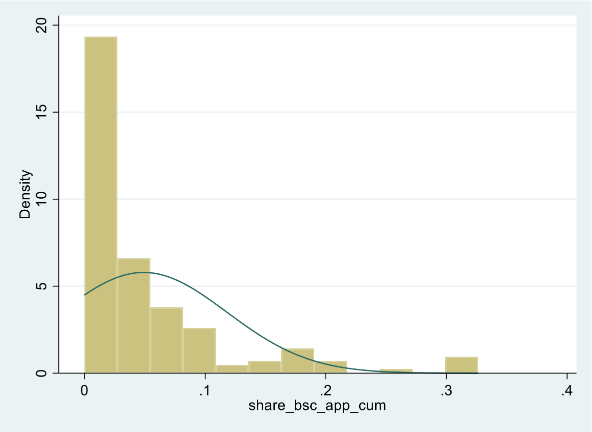
Total number of forward citations of the company's granted patents



cum_fc_g

	Percentiles	Smallest		
1%	1	0		
5%	29	1		
10%	86	6	Obs	136
25%	292.5	7	Sum of Wgt.	136
50%	1770.5		Mean	25874.47
		Largest	Std. Dev.	133659
75%	10554	189943		
90%	35562	284047	Variance	1.79e+10
95%	70093	359559	Skewness	9.861013
99%	359559	1485383	Kurtosis	106.5098

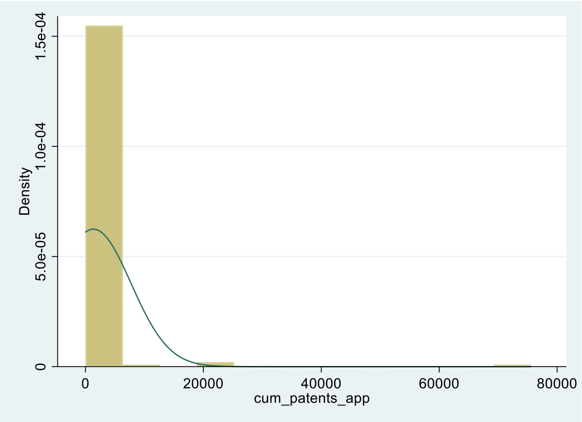
Share of backward self-citations of the company's applied-for (later granted) patents



share_bsc_app_cum

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	156
25%	0	0	Sum of Wgt.	156
50%	.0199131		Mean	.0489736
		Largest	Std. Dev.	.0689145
75%	.0645398	.303693		
90%	.1418297	.3261895	Variance	.0047492
95%	.1994506	.3261895	Skewness	2.184734
99%	.3261895	.3261895	Kurtosis	7.974492

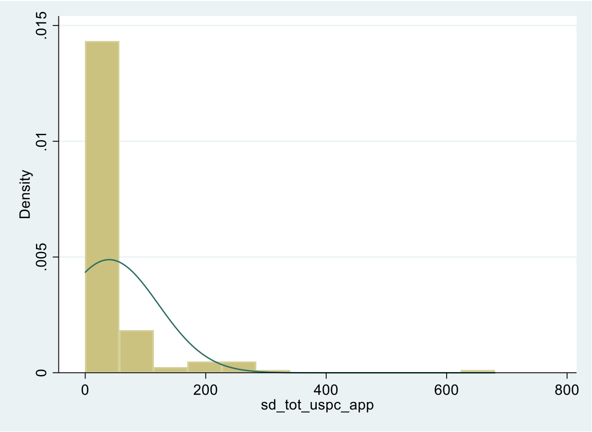
Total number of applied-for (later granted) patents of the company



cum_patents_app

	Percentiles	Smallest		
1%	0	0		
5%	1	0		
10%	2	0	Obs	161
25%	9	0	Sum of Wgt.	161
50%	58		Mean	1332
		Largest	Std. Dev.	6388.593
75%	669	10778		
90%	2527	19426	Variance	4.08e+07
95%	3868	19555	Skewness	10.13626
99%	19555	75546	Kurtosis	115.3819

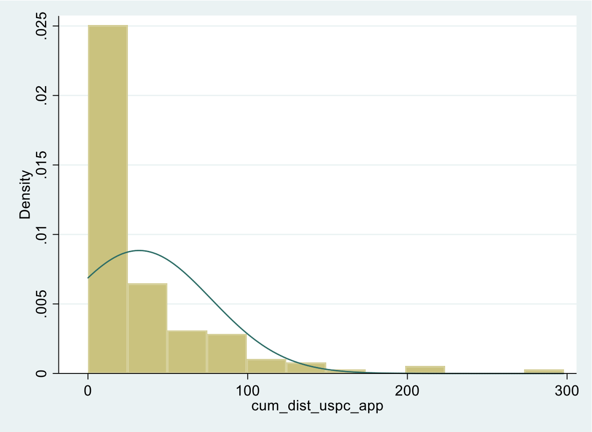
Standard deviation of the dispersion of all patents in distinct USPC main classes of the company



sd_tot_uspc_app

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	.5773503	0	Obs	144
25%	1.527525	0	Sum of Wgt.	144
50%	7.598686		Mean	39.68134
		Largest	Std. Dev.	81.65334
75%	48.40668	262.6467		
90%	108.0683	265.4429	Variance	6667.267
95%	215.2937	288.7386	Skewness	4.390463
99%	288.7386	680.4418	Kurtosis	29.44207

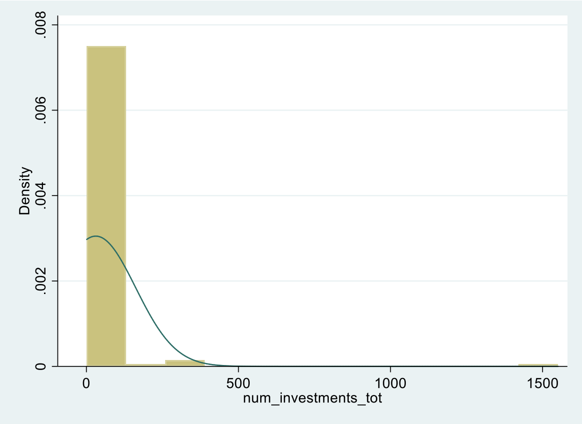
Number of distinct USPC classes of total applied-for (later granted) patents



cum_dist_uspc_app

	Percentiles	Smallest		
1%	0	0		
5%	1	0		
10%	2	1	Obs	156
25%	5	1	Sum of Wgt.	156
50%	14.5		Mean	32.00641
		Largest	Std. Dev.	45.09338
75%	35	166		
90%	91	212	Variance	2033.413
95%	121	222	Skewness	2.782155
99%	222	298	Kurtosis	13.03962

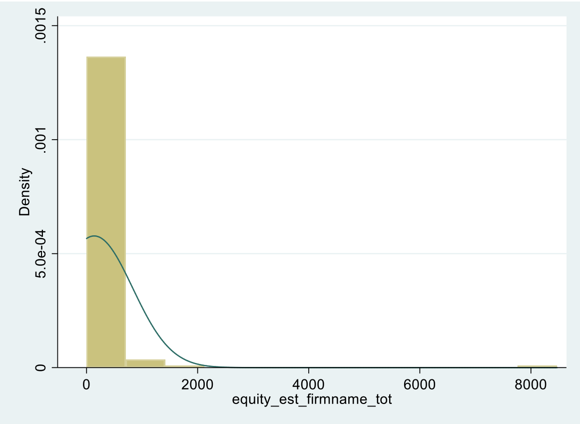
Total number of investments performed by the CVC unit



num_investments_tot

	Percentiles	Smallest		
1%	1	1		
5%	1	1		
10%	1	1	Obs	161
25%	2	1	Sum of Wgt.	161
50%	5		Mean	30.82609
		Largest	Std. Dev.	130.7358
75%	20	305		
90%	44	329	Variance	17091.86
95%	87	350	Skewness	10.10772
99%	350	1551	Kurtosis	115.535

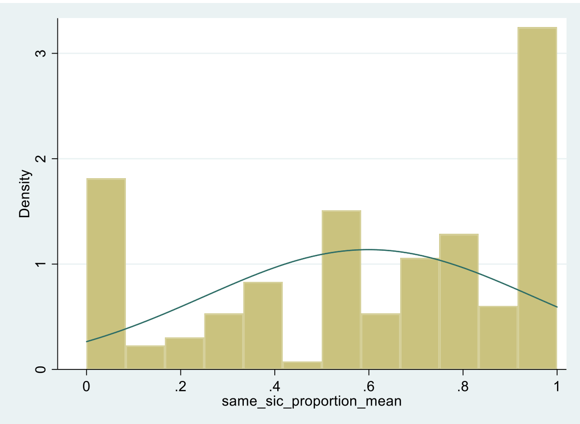
Estimated total equity investments of the CVC unit in USD million



equity_est_firmname_tot

	Percentiles	Smallest		
1%	.1	.066		
5%	.804	.1		
10%	1.7611	.228	Obs	161
25%	6.5135	.25	Sum of Wgt.	161
50%	21.24		Mean	138.0093
		Largest	Std. Dev.	690.5847
75%	67.9512	1078.142		
90%	191.1582	1102.392	Variance	476907.2
95%	375.905	1509.919	Skewness	11.1114
99%	1509.919	8471.891	Kurtosis	133.5438

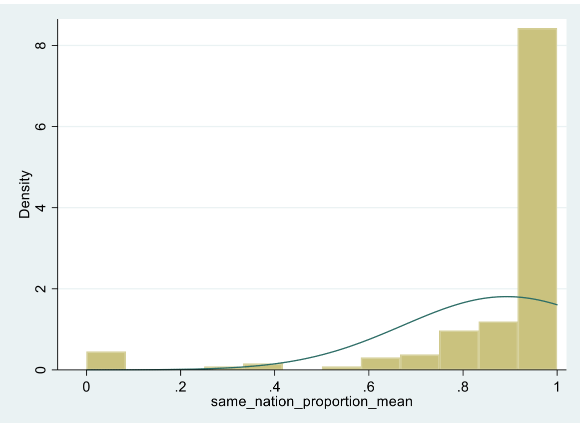
Proportion of the target investment companies with the same SIC code as the parent company



same_sic_proportion_mean

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	159
25%	.3333333	0	Sum of Wgt.	159
50%	.6851852		Mean	.5994788
		Largest	Std. Dev.	.3508371
75%	.9807692	1		
90%	1	1	Variance	.1230866
95%	1	1	Skewness	-.463288
99%	1	1	Kurtosis	1.924861

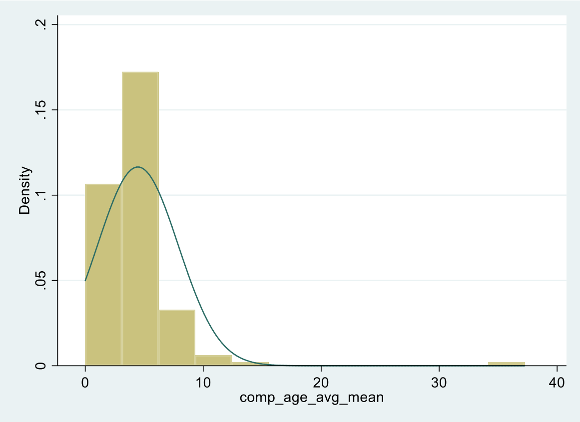
Proportion of the target investment companies that are from the same nation as the parent company



same_nation_proportion_mean

	Percentiles	Smallest		
1%	0	0		
5%	.3333333	0		
10%	.6666667	0	Obs	161
25%	.8888889	0	Sum of Wgt.	161
50%	1		Mean	.8927167
		Largest	Std. Dev.	.2207646
75%	1	1		
90%	1	1	Variance	.048737
95%	1	1	Skewness	-2.897839
99%	1	1	Kurtosis	11.25201

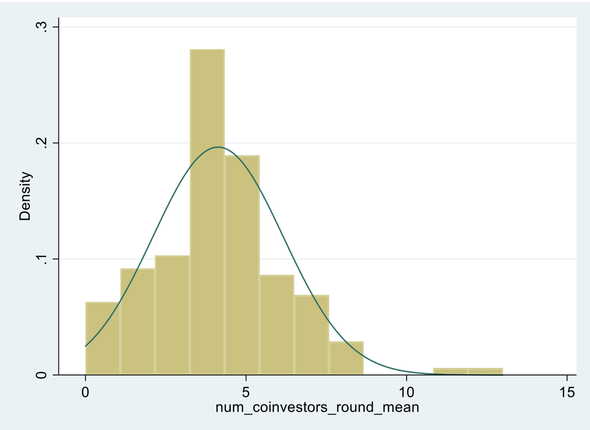
Average age of the target company at the time of investment



comp_age_avg_mean

	Percentiles	Smallest		
1%	.5	0		
5%	1	.5		
10%	1.75	.5	Obs	157
25%	2.888889	1	Sum of Wgt.	157
50%	4.2		Mean	4.461598
		Largest	Std. Dev.	3.42369
75%	5.202645	11		
90%	7.222222	12	Variance	11.72165
95%	8.375	12.5	Skewness	5.83345
99%	12.5	37.25	Kurtosis	55.02922

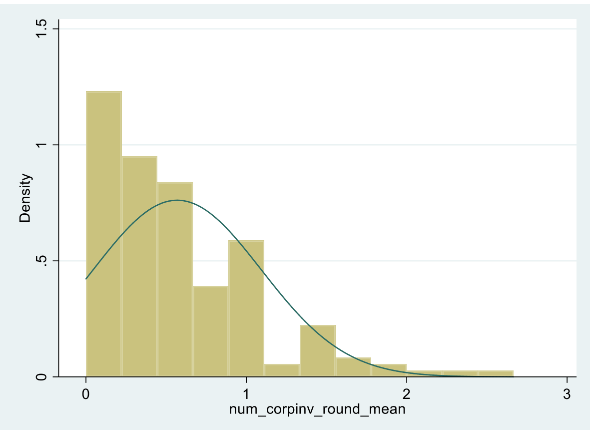
Average number of co-investors per round of investment, in which the unit participates



num_coinvestors_round_mean

	Percentiles	Smallest		
1%	0	0		
5%	1	0		
10%	1.666667	0	Obs	161
25%	3	0	Sum of Wgt.	161
50%	4.0625		Mean	4.12845
		Largest	Std. Dev.	2.031323
75%	5.041667	8		
90%	6.87963	8.5	Variance	4.126274
95%	7.5	11	Skewness	.6350653
99%	11	13	Kurtosis	5.031602

Average number of corporate co-investors per round of investment, in which the unit participates



num_corpinv_round_mean

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	161
25%	.1856061	0	Sum of Wgt.	161
50%	.5		Mean	.569109
		Largest	Std. Dev.	.5240646
75%	.8325	2		
90%	1.277778	2.166667	Variance	.2746437
95%	1.547619	2.333333	Skewness	1.281727
99%	2.333333	2.666667	Kurtosis	4.886056

Appendix I: Industry differences in mean value

Observations per industry:

sic_3	Freq.	Percent	Cum.
283	44	27.33	27.33
367	27	16.77	44.10
737	90	55.90	100.00
Total	161	100.00	

Independent variables

sic_3	mean(cum_fc_g)	mean(share_bsc_app_cum)	mean(cum_patents_app)	mean(sd_tot_uspc_app)
283	16065.58	.0926205	1047.318	66.77532
367	32195.5	.0566669	1941.926	45.10449
737	28768.79	.0247348	1288.2	24.25711

Control variables

sic_3	mean (num_investments_tot)	mean (equity_est_firmname_tot)	mean (same_sic_proportion _mean)	mean (same_nation_proportion_ mean)
283	31.27273	108.9517	.5364094	.9340948
367	76.44444	376.8986	.3912486	.859347
737	16.92222	80.5485	.6949022	.8824984

sic_3	mean(cum_dist_uspc_ap p)	mean(comp_age_avg_m ean)	mean(num_coinvestors_ round_mean)	mean(num_corpinv_rou nd_mean)
283	37.06977	4.798474	3.85849	.4629101
367	53.11111	4.858923	4.309241	.57676
737	22.84884	4.187229	4.206192	.6187332

Appendix J: Maximum-likelihood regression of control variables

Maximum-likelihood regression with estimated equity investment as an independent variable

```
Probit regression                               Number of obs     =        161
                                                Wald chi2(1)      =       9.47
                                                Prob > chi2       =     0.0021
Log pseudolikelihood = -62.223017               Pseudo R2         =     0.1049

----------------------------------------------------------------------------------------------
                           |               Robust
                subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+------------------------------------------------------------------
equity_est_firmname_tot_ln |    .258948   .0841471     3.08   0.002     .0940227    .4238733
                     _cons |  -1.910917   .3485443    -5.48   0.000    -2.594052   -1.227783
----------------------------------------------------------------------------------------------
```

Maximum-likelihood regression with number of investments as an independent variable

```
Probit regression                               Number of obs     =        161
                                                Wald chi2(1)      =      15.56
                                                Prob > chi2       =     0.0001
Log pseudolikelihood = -60.471752               Pseudo R2         =     0.1301

------------------------------------------------------------------------------------------
                       |               Robust
            subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+------------------------------------------------------------------
num_investments_tot_ln |   .3551978   .0900448     3.94   0.000     .1787132    .5316824
                 _cons |  -1.835988   .2650263    -6.93   0.000     -2.35543   -1.316546
------------------------------------------------------------------------------------------
```

Appendix K: Full regression models

Model 1 (only control variables)

```
Probit regression                               Number of obs     =        155
                                                Wald chi2(6)      =      18.44
                                                Prob > chi2       =     0.0052
Log pseudolikelihood = -56.838181               Pseudo R2         =     0.1492

------------------------------------------------------------------------------
                 |               Robust
      subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
num_investments~n |   .4143148   .1019094     4.07   0.000      .214576    .6140536
same_sic_propor~n |     .17129   .4553366     0.38   0.707    -.7211532    1.063733
same_nation_pro~n |   .3479895   .7292707     0.48   0.633    -1.081355    1.777334
comp_age_avg_mean |   .0253039   .0301437     0.84   0.401    -.0337766    .0843845
num_coinve~d_mean |  -.0074228   .0765965    -0.10   0.923    -.1575492    .1427036
   corp_co_invest |  -.2688364    .358583    -0.75   0.453    -.9716462    .4339733
            _cons |  -2.290837   .6926807    -3.31   0.001    -3.648466   -.9332076
------------------------------------------------------------------------------
```

Model 2 (only independent variables)

```
Probit regression                               Number of obs     =        132
                                                Wald chi2(3)      =      10.62
                                                Prob > chi2       =     0.0140
Log pseudolikelihood = -57.183294               Pseudo R2         =     0.1074

------------------------------------------------------------------------------
                 |               Robust
      subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
      cum_fc_g_ln |  -.2009895   .0801362    -2.51   0.012    -.3580535   -.0439255
share_bsc_app_~ln |   .5308412   .2850126     1.86   0.063    -.0277732    1.089456
sd_tot_uspc_ap~in |   .8360644   .4397314     1.90   0.057    -.0257934    1.697922
            _cons |   1.525771   1.149041     1.33   0.184    -.7263071    3.777849
------------------------------------------------------------------------------
```

Model 3 (final model)

```
Probit regression                               Number of obs     =        128
                                                Wald chi2(9)      =      24.72
                                                Prob > chi2       =     0.0033
Log pseudolikelihood = -47.285274               Pseudo R2         =     0.2345

----------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+-----------------------------------------------------------------
                cum_fc_g_ln |  -.2369588   .0895531    -2.65   0.008    -.4124796   -.0614379
       share_bsc_app_cum_ln |   .7033397   .3105812     2.26   0.024     .0946116    1.312068
        sd_tot_uspc_app_bin |   .7935149   .4625462     1.72   0.086     -.113059    1.700089
     num_investments_tot_ln |   .4250752    .135089     3.15   0.002     .1603056    .6898448
   same_sic_proportion_mean |   .5394992   .5779719     0.93   0.351    -.5933049    1.672303
same_nation_proportion_mean |   .5084136   1.167192     0.44   0.663    -1.779242    2.796069
          comp_age_avg_mean |   .0400396   .0313146     1.28   0.201    -.0213358    .1014151
  num_coinvestors_round_mean |     .05557   .0930378     0.60   0.550    -.1267808    .2379207
             corp_co_invest |  -.2894907   .4493534    -0.64   0.519    -1.170207    .5912258
                      _cons |   .2069709   1.778032     0.12   0.907    -3.277908     3.69185
----------------------------------------------------------------------------------------------
```

Appendix L: Regression model with industry differences

```
i.sic_3           _Isic_3_283-737     (naturally coded; _Isic_3_737 omitted)

Iteration 0:   log pseudolikelihood = -61.769928
Iteration 1:   log pseudolikelihood = -47.847489
Iteration 2:   log pseudolikelihood = -47.237164
Iteration 3:   log pseudolikelihood = -47.220312
Iteration 4:   log pseudolikelihood =  -47.22028
Iteration 5:   log pseudolikelihood =  -47.22028

Probit regression                                 Number of obs     =        128
                                                  Wald chi2(11)     =      27.54
                                                  Prob > chi2       =     0.0038
Log pseudolikelihood =  -47.22028                 Pseudo R2         =     0.2355

----------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+-----------------------------------------------------------------
               cum_fc_g_ln |  -.2241777   .0887568    -2.53   0.012    -.3981378   -.0502176
       share_bsc_app_cum_ln |   .6589684   .3122546     2.11   0.035     .0469606    1.270976
        sd_tot_uspc_app_bin |   .7590212   .4690956     1.62   0.106    -.1603892    1.678432
     num_investments_tot_ln |   .4238445   .1344288     3.15   0.002     .1603689    .6873202
   same_sic_proportion_mean |   .4799851    .574017     0.84   0.403    -.6450675    1.605038
same_nation_proportion_mean |   .5304377   1.106186     0.48   0.632    -1.637647    2.698522
          comp_age_avg_mean |   .0398522    .030954     1.29   0.198    -.0208166    .1005209
 num_coinvestors_round_mean |   .0582879   .0874849     0.67   0.505    -.1131793    .2297551
             corp_co_invest |  -.3235496   .4466599    -0.72   0.469    -1.198987    .5518877
                _Isic_3_283 |   .0689648   .3440994     0.20   0.841    -.6054575    .7433872
                _Isic_3_367 |  -.1109067   .3970956    -0.28   0.780    -.8891997    .6673864
                      _cons |   .0527311    1.65846     0.03   0.975    -3.197792    3.303254
----------------------------------------------------------------------------------------------
```

Appendix M: Regression model with forward citations as dummy variable

```
Probit regression                               Number of obs     =        128
                                                Wald chi2(9)      =      21.90
                                                Prob > chi2       =     0.0092
Log pseudolikelihood = -51.467612               Pseudo R2         =     0.1668

----------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+-----------------------------------------------------------------
               cum_fc_g_bin |  -.2686466   .4447914    -0.60   0.546    -1.140422    .6031286
       share_bsc_app_cum_ln |   .3557878   .2885397     1.23   0.218    -.2097397    .9213153
        sd_tot_uspc_app_bin |    .361469   .4788918     0.75   0.450    -.5771416     1.30008
     num_investments_tot_ln |   .3891872   .1228521     3.17   0.002     .1484016    .6299728
   same_sic_proportion_mean |   .4877683   .5583449     0.87   0.382    -.6065677    1.582104
same_nation_proportion_mean |   .3808873   .8504027     0.45   0.654    -1.285871    2.047646
          comp_age_avg_mean |   .0371675   .0301585     1.23   0.218    -.0219422    .0962771
 num_coinvestors_round_mean |   .0155033   .0854408     0.18   0.856    -.1519576    .1829642
             corp_co_invest |  -.2552296   .4404105    -0.58   0.562    -1.118418    .6079591
                      _cons |  -1.684414   1.303791    -1.29   0.196    -4.239797    .8709691
----------------------------------------------------------------------------------------------
```

Appendix N: Regression model with share of backward self-citations as dummy variable

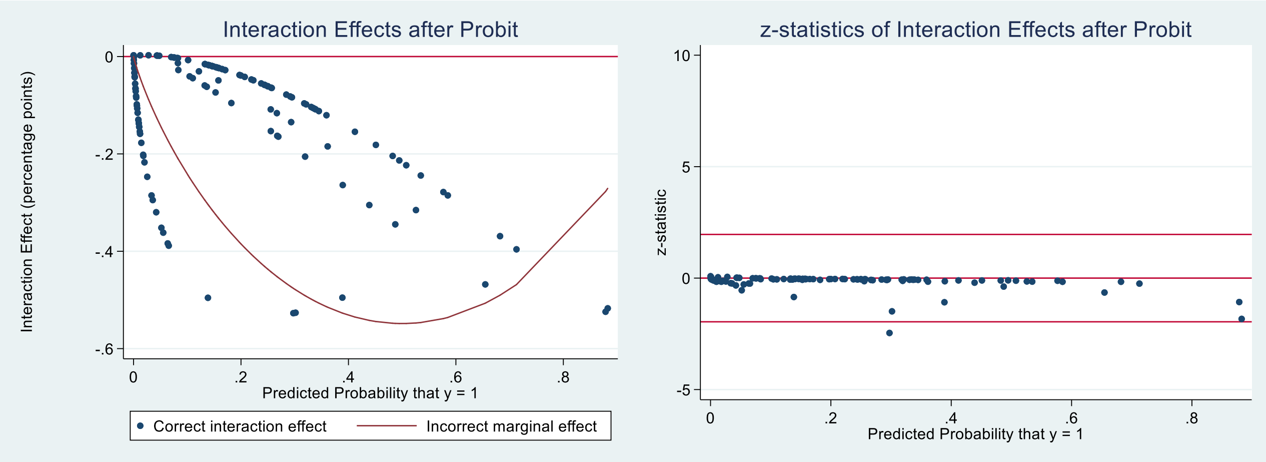
```
Probit regression                                 Number of obs     =        128
                                                  Wald chi2(9)      =      26.92
                                                  Prob > chi2       =     0.0014
Log pseudolikelihood =  -47.43076                 Pseudo R2         =     0.2321

-----------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+------------------------------------------------------------------
                cum_fc_g_ln |  -.1991094   .0815554    -2.44   0.015     -.358955   -.0392638
       share_bsc_app_cum_bin |   .8874995   .4481534     1.98   0.048     .0091351    1.765864
         sd_tot_uspc_app_bin |   .7783817   .4450564     1.75   0.080    -.0939127    1.650676
      num_investments_tot_ln |    .378242      .1323     2.86   0.004     .1189388    .6375452
    same_sic_proportion_mean |   .2405057   .5775619     0.42   0.677    -.8914949    1.372506
same_nation_proportion_mean |   .8546139    1.18787     0.72   0.472    -1.473568    3.182796
           comp_age_avg_mean |    .044966   .0313317     1.44   0.151     -.016443    .1063749
 num_coinvestors_round_mean |   .0074522   .0956815     0.08   0.938    -.1800802    .1949846
             corp_co_invest |  -.0971142    .428396    -0.23   0.821    -.9367548    .7425265
                      _cons |   -2.49367   1.226659    -2.03   0.042    -4.897878    -.089463
-----------------------------------------------------------------------------------------------
```

Regression model including interaction term (multiplication of both binary variables)

```
Probit regression                                 Number of obs     =        128
                                                  Wald chi2(10)     =      31.64
                                                  Prob > chi2       =     0.0005
Log pseudolikelihood = -45.989087                 Pseudo R2         =     0.2555

-----------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+------------------------------------------------------------------
                cum_fc_g_ln |  -.2162257   .0853846    -2.53   0.011    -.3835765   -.0488749
       share_bsc_app_cum_bin |     1.6176    .440156     3.68   0.000     .7549105     2.48029
         sd_tot_uspc_app_bin |   1.804469   .6501755     2.78   0.006     .5301489     3.07879
       int_sbsc_uspc_app_bin |  -1.374709   .6982772    -1.97   0.049    -2.743307   -.0061111
      num_investments_tot_ln |   .3817244   .1413042     2.70   0.007     .1047733    .6586756
    same_sic_proportion_mean |   .0284583   .5565476     0.05   0.959    -1.062355    1.119272
same_nation_proportion_mean |   .9451554    1.23393     0.77   0.444    -1.473303    3.363614
           comp_age_avg_mean |   .0514651    .032937     1.56   0.118    -.0130902    .1160204
 num_coinvestors_round_mean |  -.0024069   .1053274    -0.02   0.982    -.2088448    .2040311
             corp_co_invest |  -.1670059   .4654638    -0.36   0.720    -1.079298    .7452864
                      _cons |   -2.68844   1.277297    -2.10   0.035    -5.191896   -.1849832
-----------------------------------------------------------------------------------------------
```



➔ Only 2 observations are significant on a 10% level

Appendix O: Regression output for interactions with one continuous and one dummy variable

Interaction of *sd_tot_uspc_app_bin* with *cum_fc_g_ln*:

```
Probit regression                               Number of obs     =        128
                                                Wald chi2(10)     =      24.93
                                                Prob > chi2       =     0.0055
Log pseudolikelihood = -47.200584               Pseudo R2         =     0.2359

----------------------------------------------------------------------------------------------
                             |               Robust
                  subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+----------------------------------------------------------------
                 cum_fc_g_ln |  -.2107021   .1107342    -1.90   0.057    -.4277371    .0063328
        share_bsc_app_cum_ln |    .728843   .3268735     2.23   0.026     .0881827    1.369503
         sd_tot_uspc_app_bin |      1.292   1.165856     1.11   0.268    -.9930362    3.577036
             int_fc_uspc_app |  -.0677132   .1558644    -0.43   0.664    -.3732017    .2377753
      num_investments_tot_ln |    .418277   .1309421     3.19   0.001     .1616352    .6749187
    same_sic_proportion_mean |   .4694017   .5273748     0.89   0.373     -.564234    1.503037
 same_nation_proportion_mean |   .6219729   1.177829     0.53   0.597     -1.68653    2.930476
           comp_age_avg_mean |   .0400783   .0312821     1.28   0.200    -.0212335      .10139
  num_coinvestors_round_mean |   .0565661   .0911843     0.62   0.535    -.1221519    .2352842
             corp_co_invest |  -.2709885   .4456059    -0.61   0.543     -1.14436    .6023831
                       _cons |   .0794011   1.803404     0.04   0.965    -3.455207    3.614009
----------------------------------------------------------------------------------------------
```

Interaction of *sd_tot_uspc_app_bin* with *share_bsc_app_cum_ln*:

```
Probit regression                               Number of obs     =        128
                                                Wald chi2(10)     =      30.92
                                                Prob > chi2       =     0.0006
Log pseudolikelihood = -45.912599               Pseudo R2         =     0.2567

----------------------------------------------------------------------------------------------
                             |               Robust
                  subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+----------------------------------------------------------------
                 cum_fc_g_ln |  -.2564012   .0937601    -2.73   0.006    -.4401675   -.0726348
        share_bsc_app_cum_ln |   1.714671   .4920035     3.49   0.000     .7503622    2.678981
         sd_tot_uspc_app_bin |  -2.625323   1.558117    -1.68   0.092    -5.679175    .4285299
            int_sbsc_uspc_app |  -1.222746   .5271655    -2.32   0.020    -2.255972    -.189521
      num_investments_tot_ln |    .414441   .1425111     2.91   0.004     .1351243    .6937577
    same_sic_proportion_mean |    .247017   .5569467     0.44   0.657    -.8445784    1.338612
 same_nation_proportion_mean |   .6758169   1.171887     0.58   0.564     -1.62104    2.972674
           comp_age_avg_mean |   .0390005   .0311927     1.25   0.211    -.0221361    .1001371
  num_coinvestors_round_mean |   .0571714   .0996921     0.57   0.566    -.1382214    .2525643
              corp_co_invest |  -.3698114   .4596346    -0.80   0.421    -1.270679    .5310558
                       _cons |   3.451685    1.88179     1.83   0.067    -.2365561    7.139925
----------------------------------------------------------------------------------------------
```

Appendix P: Exhaustive overview of performed interactions

Variable definition	Result
cum_fc_g – binary (median) *share_bsc_app_cum* – binary (median) *sd_tot_uspc_app* – binary (median)	Model failed for *cum_fc_g* 2 significant observations for interaction with *share_bsc_app_cum*
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (median)	No significant observations for interaction with *cum_fc_g* 3 significant observations for interaction with *share_bsc_app_cum*
cum_fc_g – binary (upper 90th percentile) *share_bsc_app_cum* – binary (upper 90th percentile) *sd_tot_uspc_app* – binary (median)	Model failed
cum_fc_g – binary (lower 90th percentile) *share_bsc_app_cum* – binary (lower 90th percentile) *sd_tot_uspc_app* – binary (median)	Model failed
cum_fc_g – binary (upper 95th percentile) *share_bsc_app_cum* – binary (upper 95th percentile) *sd_tot_uspc_app* – binary (median)	Model failed
cum_fc_g – binary (lower 95th percentile) *share_bsc_app_cum* – binary (lower 95th percentile) *sd_tot_uspc_app* – binary (median)	Model failed
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (upper 90th percentile)	Model failed
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (lower 90th percentile)	Model failed
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (upper 95th percentile)	4 significant observations for interaction with *cum_fc_g* 3 significant observations for interaction with *share_bsc_app_cum* (and inverse relationship)
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (lower 95th percentile)	Model failed
cum_fc_g – continuous, logged *share_bsc_app_cum* – continuous, logged *sd_tot_uspc_app* – binary (mean)	Model failed
cum_fc_g – binary (mean) *share_bsc_app_cum* – binary (mean) *sd_tot_uspc_app* – binary (median)	Model failed for *cum_fc_g* 2 significant observations for interaction with *share_bsc_app_cum*

Appendix Q: Robustness check regression output

Regression output for a linear regression model (Model 7)

```
Linear regression                               Number of obs     =        128
                                                F(9, 118)         =       3.60
                                                Prob > F          =     0.0005
                                                R-squared         =     0.2115
                                                Root MSE          =     .36098

---------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------------+----------------------------------------------------------------
               cum_fc_g_ln |  -.0580367   .0225354    -2.58   0.011    -.1026628   -.0134105
       share_bsc_app_cum_ln |   .1681937   .0815141     2.06   0.041     .0067737    .3296137
         sd_tot_uspc_app_bin |   .1362785    .110233     1.24   0.219    -.0820128    .3545698
      num_investments_tot_ln |   .0990678   .0304006     3.26   0.001     .0388664    .1592692
    same_sic_proportion_mean |     .06799   .0907457     0.75   0.455    -.1117111    .2476912
same_nation_proportion_mean |  -.0714757   .1166022    -0.61   0.541    -.3023798    .1594284
            comp_age_avg_mean |   .0075173   .0093355     0.81   0.422    -.0109695    .0260041
 num_coinvestors_round_mean |    .012028   .0142115     0.85   0.399    -.0161147    .0401707
             corp_co_invest |  -.0824079   .0771873    -1.07   0.288    -.2352597    .0704439
                      _cons |   .7942025   .3577836     2.22   0.028     .0856937    1.502711
---------------------------------------------------------------------------------------------
```

Regression output for logit regression model (Model 8)

```
Logistic regression                             Number of obs     =        128
                                                Wald chi2(9)      =      19.21
                                                Prob > chi2       =     0.0234
Log pseudolikelihood = -47.203592               Pseudo R2         =     0.2358

---------------------------------------------------------------------------------------------
                            |               Robust
                 subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------+----------------------------------------------------------------
               cum_fc_g_ln |  -.4374742   .1602011    -2.73   0.006    -.7514626   -.1234859
       share_bsc_app_cum_ln |   1.243449    .562518     2.21   0.027     .1409336    2.345963
         sd_tot_uspc_app_bin |   1.378605   .8801051     1.57   0.117    -.3463691    3.103579
      num_investments_tot_ln |   .7740989   .2615508     2.96   0.003     .2614687    1.286729
    same_sic_proportion_mean |    1.08685   1.106391     0.98   0.326    -1.081637    3.255336
same_nation_proportion_mean |   .9151251   2.313135     0.40   0.692    -3.618536    5.448786
            comp_age_avg_mean |   .0743002   .0530909     1.40   0.162     -.029756    .1783564
 num_coinvestors_round_mean |   .0776628   .1983941     0.39   0.695    -.3111826    .4665081
             corp_co_invest |  -.4281178   .8897281    -0.48   0.630    -2.171953    1.315717
                      _cons |   .3963124   3.366784     0.12   0.906    -6.202464    6.995088
---------------------------------------------------------------------------------------------
```

Regression output if sample only includes CVC units with more than one investment (Model 9)

```
Probit regression                                 Number of obs     =        116
                                                  Wald chi2(9)      =      21.67
                                                  Prob > chi2       =     0.0100
Log pseudolikelihood = -44.839298                 Pseudo R2         =     0.2238

------------------------------------------------------------------------------------------------
                             |               Robust
                  subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                cum_fc_g_ln |  -.2300644   .0893569    -2.57   0.010    -.4052007   -.0549281
        share_bsc_app_cum_ln |   .6165299    .308978     2.00   0.046     .0109441    1.222116
          sd_tot_uspc_app_bin |   .7831892     .46756     1.68   0.094    -.1332115     1.69959
      num_investments_tot_ln |   .4371651    .139266     3.14   0.002     .1642087    .7101215
     same_sic_proportion_mean |   .4315674   .6190961     0.70   0.486    -.7818387    1.644973
  same_nation_proportion_mean |   .4581229   1.298668     0.35   0.724    -2.087219    3.003465
            comp_age_avg_mean |   .0485997   .0347326     1.40   0.162     -.019475    .1166743
   num_coinvestors_round_mean |   .0826557   .0970781     0.85   0.395    -.1076138    .2729253
               corp_co_invest |  -.0592458    .581344    -0.10   0.919    -1.198659    1.080168
                        _cons |  -.3715493   1.931017    -0.19   0.847    -4.156273    3.413175
------------------------------------------------------------------------------------------------
```

Regression output if sample only includes CVC units within 90th percentile of *cum_patents_app* (Model 10)

```
Probit regression                                 Number of obs     =        111
                                                  Wald chi2(9)      =      27.83
                                                  Prob > chi2       =     0.0010
Log pseudolikelihood = -36.900087                 Pseudo R2         =     0.2952

------------------------------------------------------------------------------------------------
                             |               Robust
                  subsidiary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                cum_fc_g_ln |  -.2810393   .1023336    -2.75   0.006    -.4816095   -.0804692
        share_bsc_app_cum_ln |   .5972301   .3767153     1.59   0.113    -.1411184    1.335579
          sd_tot_uspc_app_bin |   .9281639   .5230826     1.77   0.076    -.0970591    1.953387
      num_investments_tot_ln |   .5433885    .141337     3.84   0.000      .266373     .820404
     same_sic_proportion_mean |   1.794968   .5859612     3.06   0.002     .6465054    2.943431
  same_nation_proportion_mean |   .0130999   1.170693     0.01   0.991    -2.281416    2.307615
            comp_age_avg_mean |   .0569046   .0308646     1.84   0.065    -.0035889    .1173981
   num_coinvestors_round_mean |   .0348689   .1103527     0.32   0.752    -.1814185    .2511562
               corp_co_invest |  -.3629421   .4887772    -0.74   0.458    -1.320928    .5950436
                        _cons |  -.5351951   2.040261    -0.26   0.793    -4.534034    3.463643
------------------------------------------------------------------------------------------------
```