NOVA SCHOOL OF BUSINESS & ECONOMICS

Master Thesis

The effect of corporate venture capital portfolio diversification on firm innovation: A replication and extension study

---

Sabrina Steeb

31796

Nova School of Business & Economics

International MSc Management

Double Degree with Maastricht University

Maastricht, 04.01.2019

Supervisor: Luis Filipe Lages, PhD

Abstract

This replication and extension study fails to replicate the main finding of the 2016 study "Corporate venture capital portfolios and firm innovation" (2016) which reported that the relationship between corporate venture capital portfolio diversity and firm innovation performance is inverse U-shaped. Instead, with a fixed effects negative binominal regression and panel data consisting of 99 corporate venture capital investors and ranging from 2010 to 2015, the present study finds a significant linear relationship which is positively moderated by geographic diversity and negatively moderated by investor absorptive capacity. These study findings suggest that a firm with a relatively diverse CVC portfolio has on average ~12.5% more patents than a firm with a relatively less diverse portfolio.

KEYWORDS:

corporate venture capital, exploratory innovation, inter-organizational portfolios, replication study

Table of Contents

List of Tables and Figures

1 Introduction

Understandably, management research is as obsessed with publishing new and original content as any other discipline (Block & Kuckertz, 2018). Incremental research is not very popular among researchers even though this type of research plays a crucial role in verifying previous results (Kiri, Lacetera, & Zirulia, 2018) instead of merely assuming that every new study finding validates as ‘proven theory’ to be built upon (Bettis, 2012). Scholars have observed that there is even a “bias against publication of replication studies” (Bettis, 2012, p. 111). This observation is alarming since “The academic community is increasingly concerned that many of these novel findings might be nonreplicable artifacts” (Block & Kuckertz, 2018. p. 355) and as such, threats to the credibility and applicability of the entire research field (Bettis, Helfat, & Shaver, 2016).

Indeed, trustworthy, empirical evidence is crucial to secure the empirical evidence base of every research field, yet social sciences, are particularly susceptible to overestimated effect sizes (Camerer et al., 2018). Consequently, journals like *Strategic Management Journal*, *Research Policy* and *Management Review Quarterly* advocate explicitly for an enriched scientific debate about reproducibility (Bettis, 2012; Bettis, Ethiraj, Gambardella, Helfat, & Mitchell, 2016; Block & Kuckertz, 2018; Ethiraj, Gambardella, & Helfat, 2016; Kiri et al., 2018). To support this development, more and more journals - including the ones just mentioned - are changing their editorial policies to animate replication. Supporters of the debate argue that only with such additional evidence, social sciences can build up truly “repeatable cumulative knowledge” (Bettis, Ethiraj, Gambardella, Helfat, & Mitchell, 2016, p. 260).

Moreover, robust, high-quality research is not only valued in academia but also by firms which turn to scientific findings to support decision-making processes (Bettis, Helfat, et al., 2016; Kiri et al., 2018). Replication studies are especially important in such cases because they "establish the range of applicability of prior studies and better support what implications can be drawn for managerial practice" (Bettis, Helfat, et al., 2016, p. 2193). In summary, there is an urge for more replication studies and the present paper intends to make a valuable contribution to this ongoing scientific debate. To address these concerns both from an academic and managerial perspective, the present thesis replicates a study about a trending topic in strategic management research and practice. Before beginning the replication exercise, the topic will be introduced shortly below.

Unarguably, globalization and rapid technological change are making the world spin ever faster. Nearly no company can avoid being affected by the consequences of this change. The more dramatic the shift in a company's external environment, the bigger the urge for strategic renewal and innovation. (Agarwal & Helfat, 2009) To secure sustained success, companies are pressured to adapt and pursue alternative opportunities for growth and profitability (Zahra & Hayton, 2008). Strategic renewal can be accomplished through internal change or the aid of external mechanisms (Agarwal & Helfat, 2009). Either way, knowledge creation and organizational learning play an essential role whenever pursuing innovation. This is true for incremental, exploitative as well as disruptive, explorative innovation (March, 1991). The knowledge-based view of the firm, which can be considered an extension of the resource-based view, even acknowledges "knowledge as the most strategically important of the firm's resources" (Grant, 1996, p. 110).

But how do firms grow their knowledge base and become more innovative? One strategic option to answer this question is to form inter-organizational knowledge-sharing relationships. These interfirm ties fall into the category of external mechanisms of knowledge sourcing. Next to alliances or joint ventures, corporate venture capital also falls into this category.

Corporate venture capital (CVC) is a particular form of venture capital and can be defined as "direct minority equity investments made by established firms in privately held entrepreneurial ventures" (Wadhwa, Phelps, & Kotha, 2016, p. 97). The phenomenon of corporate venture capital emerged for the first time in the 1960s and has since experienced three major waves of growth which, due to the relatedness of the sectors, are closely linked to the cycles of growth of the independent venture capital (IVC) industry. The most recent growth wave of the overall VC industry and the CVC sector in particular started in the late 1990s with the rise of information and communication technologies. (Gompers, 2002) Although the wave ebbed away after the bubble of the internet economy boom burst around the turn of the millennium, CVC investments are experiencing steady growth again in this century. The U.S. National Venture Capital Association (2018) reports that the total number of U.S. VC deals with CVC participation has nearly tripled between 2004 (508) and 2017 (1,355). Although corporate venture arms were only involved in roughly one out of six U.S. venture capital deals in 2017, the aggregated deal size of these deals amounted to US$ 39 billion. This figure represents a relative share of 45% compared to the overall venture capital raised of US$ 85 billion in the same year (National Venture Capital Association, 2017). The numbers demonstrate how significant the share and consequently the influence of CVC is. In conclusion, CVC has established itself as an important element of the start-up ecosystem. (Weiblen & Chesbrough, 2015).

The influence and popularity of CVC have not only triggered more and higher investments, but also more academic research in this field. Nevertheless, the literature on CVC remains rather fragmented due to the multifaceted nature of the phenomenon. (Röhm, 2018) The aim of the present study is to contribute simultaneously to the ongoing research streams about the popular topic of CVC and to the ongoing debate about the importance of replication studies.

The present paper is therefore a (partial) replication and extension of a study by Wadhwa, Phelps and Kotha published in the *Journal of Business Venturing:* "Corporate venture capital portfolios and firm innovation" (2018). The goals of this study are twofold. First, this study aims to assess the generalizability of the previously found inverted U-shaped relationship between CVC portfolio diversity and innovation performance in a different research context using the same research design. Because of this intention, the present study collected new data from a different population; concretely, data from a broader industry setting and a more recent time period. This type of *quasi-replication* studies "help[s] us to understand whether the original results were idiosyncratic to a particular setting or not." (Bettis, Helfat, et al., 2016, p. 2195) Although narrow replication studies that assess "the reliability and representativeness of the data for a specific population" (Bettis et al., 2016, p. 2195) are equally important, this type of quasi-replication is especially interesting (Ethiraj et al., 2016) and appropriate in the field of strategic management to evaluate the generalizability of managerial implications of findings so far limited to very specific settings (Bettis et al., 2016). Important boundary conditions to previous findings relevant for practice might be identified through these incremental studies. Second, this study augments the theoretical and managerial scope of CVC portfolio diversification theory through the extension of two additional concepts to the model. This approach is in line with top-tier journal suggestions on how to conduct replication studies (Bettis et al., 2016; Ethiraj et al., 2016).

Based on the article by Wadhwa, Phelps and Kotha (2016), the present study questions if the authors' main finding - an inverse U-shaped relationship between CVC portfolio diversity and investor innovation performance - holds true in a different sample across industries. In other words, this thesis questions the study's external validity. This approach addresses the research gap of generalizability of the original study. Various studies in this field face the same limitation (e.g. Belderbos, Jacob, & Lokshin, 2018; Wadhwa & Kotha, 2006) The original study further analyzed a moderation effect focused on characteristics of the portfolio ventures. However, an additional research gap remains whether and which focal firm characteristics moderate the effect. Linking into the prior research, this study will therefore first investigate whether the benefits gained from CVC activity can be conditioned through the corporate investor's level of absorptive capacity. Next, to extend the original study even further, the present study will also analyze one more characteristic of the portfolio firms, namely the portfolio's geographic diversity. To further stimulate replication, all materials including the empirical data and code (R and STATA) used for this analysis are accessible via www.github.com/sabrinasteeb/CVC-portfolio-diversity. Additionally, the syntax is displayed in the appendix of this paper.

2 Empirical context

While a firm's knowledge base can grow through a series of knowledge-enhancing investments over time, firms can also grow knowledge through acquiring or 'grafting' external knowledge bases (Cohen and Levinthal, 1989; Huber, 1991). Corporate venture capital is a particularly appropriate setting to study the effects of knowledge-sharing interfirm relationships for the investing firm. The underlying reasons for this notion are explained by the strengths and weaknesses of the parties involved. Large corporations, on the one hand, have access to resources and power, can run and scale proven business models. Yet, they struggle to identify and exploit new sources of innovation and growth. Start-ups, on the other hand, typically have aspiring ideas, are agile and benefit from the visionary mindsets of their people; however, these new ventures lack what established corporations bring to the table. (Weiblen & Chesbrough, 2015) Corporate venture capital bridges the gap between these two worlds by combining "entrepreneurial activity with corporate ability" (Weiblen & Chesbrough, 2015, p. 66). In most cases, the investing company pursues a strategic goal by seeking "to identify and exploit synergies between itself and a new venture" (Chesbrough, 2002, p. 92). Both parties benefit mutually from the relationship in the best-case scenario. Nonetheless, some CVC funds also pursue purely financial goals. (Chesbrough, 2002)

Röhm (2018) summarizes the ongoing research streams, like drivers of CVC adoption (industry and firm level), governance aspects or investment procedure, in his recent CVC review. One of the ongoing research streams analyses the implications CVC activity has for the corporate investor and in particular if and how the investor benefits from strategic learning. The influence of corporate venture capital on knowledge creation and hence firm innovation for corporate investors has been discussed in the literature, yet evidence remains mixed and inconclusive (Wadhwa, Phelps, & Kotha, 2016).

Dushnitsky and Lenox (2005) found evidence that engagement in CVC positively and directly influences the innovation rates of the corporate mother company. In the author's research, the effect was strongest when intellectual property protection in the industry of the investee was weak. Furthermore, the researchers demonstrated that absorptive capacity is necessary for the investing firm to benefit from the relationship through being able to leverage accessible knowledge. Wadhwa and Kotha (2006) in contrast introduced limitations to the main effect and "an optimum point beyond which the contribution of CVC investments to investor knowledge creation declines." (Wadhwa & Kotha, 2006, p. 820) Hence, they argued for an inverted U-shaped effect moderated by the level of involvement of the investor with the portfolio firms. The study by Wadhwa et al. (2016), which serves as a foundation for this replication, has again found evidence for an inverted U-shaped relationship. However, the researchers have not analyzed the number of CVC investments but the CVC portfolio diversity and how this diversity influences the corporate mother's innovation performance. Further, the researchers showed that this effect is moderated by the depth of knowledge resources accessible throughout the portfolio companies. Their research is limited to one high-tech industry and thus, the generalizability of the findings is a concern. Indeed, this concern is a wide-ranging one as many studies limit their sample setting to one specific industry (e.g., Belderbos, Jacob, & Lokshin, 2018; Wadhwa & Kotha, 2006; Wadhwa et al., 2016) which makes it difficult to assess the findings regarding generalizability both for academia and practice. To conclude, past research argues for a contingency perspective to examine if the relationship of CVC engagement on the focal firm's innovation performance is conditioned by other factors (Wadhwa et al., 2016) and for more generalizability of results. Both of these issues will be addressed in the present paper.

3 Theory and hypotheses

The present study aims to re-analyze the baseline hypothesis suggested by Wadhwa et al., (2016) and extend their research with two additional hypotheses. In the following, the theoretical arguments used by the original paper to develop the baseline hypothesis will first be reviewed and summarized. Next, the two additional hypotheses to extend the research will be developed.

Wadhwa et al. (2016) based their research on the recombinatory search literature. This research stream argues that innovation is a complex search process that only results in the desired problem-solving exercise when existing knowledge components are recombined or reconfigured into something new (Katila & Ahuja, 2002). The most common search strategy is to search for knowledge within a company's boundaries (Katila & Ahuja, 2002). The knowledge available within these boundaries is typically called local knowledge. To further increase a company's chances of innovation, some firms however also extend their search radius to distant knowledge which resides in other firms (Nahapiet & Ghoshal, 1998). To access this potentially useful knowledge, interorganizational ties are needed and additionally, certain capabilities to successfully recombine old and new, local and distant knowledge (Cohen & Levinthal, 1990; Katila & Ahuja, 2002). In conclusion, interfirm relationships can be part of knowledge acquisition and as such, of innovation strategies.

During a recombinatory search process, the intermediate aim is to increase and diversify knowledge flows with which to fuel recombinatory activities (Wadhwa & Kotha, 2006). This occurs when one or both partners provide access to their respective resources (Dushnitsky & Shaver, 2009). Therefore, CVC is an especially interesting setting to study distant knowledge recombination efforts because it is often strategically motivated. Moreover, the startup ecosystem is known for its innovativeness and visionary thinking which makes it an attractive search space for incumbents looking for 'the next big thing'. (Weiblen & Chesbrough, 2015) Wadhwa et al. further argue that corporate investors usually form close bonds with their startup targets and gain a board seat or similar controlling rights in the target venture which makes it easier to access and understand their knowledge bases and to steer the transfer of knowledge back to the investor. The perceived advantages for corporations to engage in strategically motivated CVC investments are easily understood. But why would innovative startups take the risk of potential imitation and opt for CVC?

Entrepreneurs seeking financial capital to fuel potentially high-growth startup ideas can turn to a variety of sources typically distinguished in debt and equity financing. On the equity financing landscape, the venture capital market is a popular choice to turn to for new ventures. (Drover et al., 2017) Traditional players in this market are entrepreneurs, independent venture capitalists, corporate venture capitalists and angel investors (Dushnitsky & Shaver, 2009). Most recently, crowdfunding and accelerators have also gained in popularity. (Drover et al., 2017) CVC funds are a valuable choice for ventures that do not only seek for "working capital, but also strategic guidance, sales and marketing channels, business development opportunities, and specific domain expertise." (National Venture Capital Association, 2018, p. 27) The relationship between the corporate investor and the target venture can be strategically beneficial for both partners. To conclude, the growing numbers of CVC deals and deal volume further suggest that both partners value the specific characteristics of this kind of knowledge-sharing partnership (National Venture Capital Association, 2018). Research has indeed demonstrated the effectiveness of CVC investments for the focal firm's innovation performance (Dushnitsky & Lenox, 2005; Keil, 2004; Wadhwa & Kotha, 2006; Wadhwa et al., 2016).

CVC funds commonly invest in more than one startup over time. The different investments form the investor's portfolio. (Wadhwa et al., 2016) Larger portfolios represent a potentially superior source of distant knowledge if the knowledge spread across portfolio firms is diverse and non-redundant. Highly diverse knowledge, however, is difficult to the corporate investor to absorb and recombine into something meaningful because the costs of searching increase. Wadhwa et al. (2016) therefore hypothesize that "increasing portfolio diversity presents both increasing benefits and challenges to firm innovation" (p.98). The authors argue that it is important to understand the conditions under which the gains compensate the costs.

To address these concerns, Wadhwa et al. (2016) extend their original research model with two moderators that are both based on the portfolio ventures' characteristics. The two conditioning characteristics are (1) "depth of the knowledge base of the ventures" (hypothesized and supported to have a positive effect) and (2) "the knowledge to which the ventures have access through their other interfirm partnerships" (p. 98) (hypothesized and partially supported to have a positive effect). These two moderating conditions will not be replicated in the present study. Instead, to add to the literature under which conditions corporate investors' costs of diversification overshadow the benefits, the present study introduces two additional concepts. The first, absorptive capacity, is an important characteristic of the focal firm. The second, geographic diversity, is an important characteristic of the portfolio setup.

3.1 Portfolio diversity

Diversity is an ambiguous, yet key concept to this analysis. Wadhwa et al. (2016) base their understanding of diversity on an interdisciplinary framework developed by Stirling (2007) to ensure a heuristic and systematic approach comparable to different contexts. In line with this framework, the authors conceptualize portfolio diversity "as the extent to which the resources and capabilities, including technological knowledge, of startups in a portfolio differ from each other and from those of the investor" (p. 98). The focus will be on technological knowledge when constructing the variable later. The definition is in line with prior research about technological diversity (Phelps, 2010) and knowledge diversity (Rodan & Galunic, 2004). Concretely, Wadhwa et al. (2016) hypothesize that portfolio diversity affects firm innovation in four ways which will be summarized in the following.

The first argument used by Wadhwa et al. (2016) builds on the recombinatory search literature (e.g., Fleming, 2001; Gavetti & Levinthal, 2000; Katila & Ahuja, 2002) The authors make the argument that as diversity increases, so does the relative novelty of knowledge available among the members of the portfolio. Diverse portfolios are thus high in search scope meaning that a high number of dissimilar knowledge elements are available within them (Schilling & Green, 2011). The higher this number of dissimilar knowledge elements, the greater the combinatorial possibilities and thus the potential for invention (Fleming, 2001) because exploitation in organizational learning (March, 1991) happens especially in cases "when inventors try completely new components or combinations" (Fleming, 2001, p. 119).

The second argument is based on the 'paradox of corporate venture capital', a phenomenon analyzed by Dushnitsky and Shaver (2009). The underlying problem of the phenomenon is that if two firms "occupy structurally equivalent positions in the technological network [they] do so because they perform similar roles as innovators." (Stuart & Podolny, 1996, p. 35). Is this the case among members of a CVC network, then young innovative ventures are exposed to higher risk of potential imitation through either the corporate investor or other portfolio firms (Dushnitsky & Shaver, 2009). Dushnitsky and Shaver's study suggests that when the corporate investor and target startup are active in the same industry, the target startup is likely to prefer IVC over CVC backing in the first place. If part of a CVC portfolio, startups that fear that their intellectual property could be imitated by other members of the network demonstrate diminishing willingness to share knowledge and increased protectiveness; hence, essentially counteracting with many corporate investors' strategic intention of gaining a 'window on technology' (Benson & Ziedonis, 2009). On the contrary, diverse portfolios lower partner protectiveness across portfolio ties (Dushnitsky & Shaver, 2009) and can even trigger knowledge sharing (Khanna, Gulati, & Nohria, 1998).

The third argument builds on the theory of absorptive capacity. The present paper hypothesizes in the extension part that a focal firm's absorptive capacity conditions the relationship between CVC portfolio diversity and firm innovation. Therefore, the role of absorptive capacity will be discussed more comprehensively in the following section.

Ultimately, corporate investors are indeed limited in the resources they allocate to each collaboration effort. The authors make the argument that as diversity increases, so does the "potential for conflicts and congestion among the resources an investor commits to these relationships" (Wadhwa et al., 2016, p. 99). This argument is tightly linked to the concept of absorptive capacity because it, not only but also, aims at human capital constraints related to investor personnel. The investing firm needs savvy personnel to achieve meaningful, value-adding recombination in the distant search space, yet cognitive resources are primarily reserved for internal activities. The authors hence suggest that there is a turning point beyond which portfolio-level synergistic benefits turn negative because the knowledge accessible in the ventures becomes too diverse and thus, the resources needed for and costs of recombination become too high.

Considering the advantages and disadvantages of diverse portfolios, Wadhwa et al. (2016) hypothesize an inverted U-shaped relationship between a corporate investor's innovation performance and its CVC portfolio diversity. The researchers thus assume that moderate levels of portfolio diversity are optimal to realize innovation performance gains for the corporate investor, while extreme levels diversity, both low and high, should be avoided.

Hypothesis 1. "The diversity of a firm's portfolio of new ventures will have an inverted U-shaped relationship with its innovation performance."
(replicated from Wadhwa et al. (2016, p. 99))

3.2 Investor absorptive capacity

When firms engage in interfirm knowledge sourcing, one key aspect to make the investments worthwhile is the firm's ability to effectively utilize the knowledge to which they are exposed to during the search process. A prominent motive for corporations to engage in corporate venture capital is to gain a 'window on new technologies' (Benson & Ziedonis, 2009). As such, CVC forms part of a company's innovation strategy. However, to realize the benefits from these potential external sources of innovation, the focal firm needs to be able to integrate and exploit dissimilar knowledge. From an organizational learning perspective, various organizational abilities that come into play in this process are joint in the construct of a firm's absorptive capacity.

The importance of absorptive capacity has been widely acknowledged in the strategic management and organizational learning literature (e.g., Lane & Lubatkin, 1998; Nahapiet & Ghoshal, 1998). Absorptive capacity is commonly understood as supporting an organization in managing knowledge and the construct is often used to explain different organizational phenomena; yet, its definition and operationalization remain ambiguous (Zahra & George, 2002). The most prominent concept of absorptive capacity was coined and defined by Cohen and Levinthal's (1990) seminal article as a firm's ability "to recognize the value of new, external information, assimilate it, and apply it to commercial ends" (p. 128). The researchers tightly link absorptive capacity to the exploitation of external sources of innovation; thereby explaining why some companies succeed in exploiting and leveraging external knowledge for their own means - and some companies fail to. As such, absorptive capacity is a salient factor in building up and enduring innovation capabilities and successful organizational learning (Cohen & Levinthal, 1990) It follows that a firm’s absorptive capacity will also condition the learning gains realized through CVC (Dushnitsky & Lenox, 2005). In summary, the greater a firm’s absorptive capacity, the greater the anticipated benefits from external knowledge sourcing strategies such as CVC.

Absorptive capacity has been studied in similar contexts before. Vasudeva and Anand (2011) have studied its implications in alliance portfolios. Moreover, prior research in the context of CVC by Dushnitsky and Lenox (2005) has shown that sufficient absorptive capacity positively influences a firm’s innovation rates. The present study will add to the existing knowledge and analyze absorptive capacity under different levels of portfolio diversity.

Hypothesis 2. A firm’s absorptive capacity will moderate the relationship between CVC portfolio diversity and investor innovation performance. That is, the higher the corporate investor’s level of absorptive capacity, the stronger the positive effect.

3.3 Portfolio geographic diversity

The effect of CVC portfolio diversity on firm innovation performance will also depend in part on the composition of the porfolio in term of the geografic spread of the ventures. Like the technological heterogenity across the network, geographic diversity adds another dimension of complexity to the extramural search space and thus, should also be taken into consideration. It is commonly acknowledged nowadays that national backgrounds play an important role in understanding the various trajectories that countries have taken in similar or even in the same technological domains. National environments exert a significant influence on the process of innovation and technological development in various ways, for example, through a nation’s formal and informal culture, infrastructure and regulatory context. (Ahuja & Katila, 2004; Porter, 1990) Locational identities are embodied in startups and their employees. Thus, entrepreneurial ventures from distant geographic background are an interesting opportunity to turn to when assessing CVC investment options because their addition can significantly broaden the search space. In consequence, this might lead to more successful problem-solving, hence, innovation for the corporate investor. (Belderbos et al., 2018)

Drawing on the recombinatory search literature, prior research on interorganizational ties has studied the effect of geographic diversity before (e.g., Belderbos et al., 2018; Lavie & Miller, 2008), yet the linkage with overall portfolio diversity is novel.

Hypothesis 3. Geographic diversity of the CVC portfolio will moderate the relationship between CVC portfolio diversity and investor innovation performance. That is, the higher the geographic diversity, the stronger the positive effect.

4 Data and methodology

The present thesis is a replication and extension of Wadhwa, Phelps and Kotha's (2016) paper. The original study is set in the 'Communications Equipment' industry (SIC 366) and analyzes "40 publicly traded telecom-equipment manufacturers headquartered in 11 countries" (p. 101) between 1989 and 2000. To assess the external validity of the study's main findings, the present study replicates the authors' analytical strategy as closely as possible in a broad industry setting. This study includes companies from 44 different industries therefore[1].Additionally, this study is set during a different observation period.

Wadhwa et al. chose the third wave of the corporate venture capital industry (Gompers, 2002) as their observation period. Because of the previously discussed emerging fourth wave

[1] Based on the three-digit primary SIC code classification

of CVC, a more recent period was chosen for this study using the following approach. First, the decision whether to grant a patent is time-lagged to the patent application date. To reduce the risk that the study's patent data is censored to the right, 2015 was marked as the final observation year. Second, to exclude potential bias due to the latest global financial crisis, 2010 was marked as the initial year. Thus, this study is set during the period 2010-2015. In line with Wadhwa et al. (2016), all dependent variables are lagged one year to the independent variable. Hence, the last CVC investments were observed in 2014.

The starting point for the sample construction was a sample composition list disclosed by Hamm, Jung, & Park (2018). The list was extended through online research. In total, 340 CVC funds of 291 corporate investors were identified. The following four criteria were applied to narrow the list. First, all private companies were excluded to ensure available and reliable financial data. Second, CVC funds of financial investors were excluded to comply with the common definition of corporate venture capitalists (e.g., Wadhwa et al., 2016). Third, corporate investors with zero investments during the sample period were excluded. Forth, investors who did not apply for at least one patent at the USPTO during the sample period. The final sample consist of 98 publicly traded corporate investors across various industries headquartered in 21 countries. Together they form an unbalanced panel with 466 firm-year observations[2].

4.1 Data and measures

To partially replicate Wadhwa et al.'s (2016) findings, the first part of the analysis will follow the authors' methodology as closely as possible to study the original main effect. In the second part, two new moderators will be introduced and used instead of the moderators suggested by Wadhwa et al. (2016)[3]. In coherence with the authors, this study uses U.S. patent data to measure investor innovation. This approach is common in the literature. Several scholars

[2] In the full regression model (table 3, model 5)
[3] Portfolio depth and portfolio firms' partners

like Griliches (1990) argue that patent data used in an analysis should only be retrieved from one country to avoid unreliable, inconsistent or incomparable data across firms. In the present case, patent data was obtained from the USPTO like in the original study. Similar to the 2016 sample setting, many firms analyzed in this study are headquartered outside the US. Wadhwa et al. (2016) argue that this fact does not bias the study outcomes because in today's globalized world firms have "strong incentives to gain patent protection in the world's largest market" (p. 101). The patent application date was used by this and the original study. Again, this is in line with standard practice because it captures the moments when firms themselves perceive that novel knowledge is created (Griliches, 1990; Jaffe, 1986). All financial data and industry classification information for the corporate investors was collected via Compustat. Industry classification information for the ventures was collected from Zephyr. CVC, alliance and joint venture deal data was also obtained via Zephyr.

The present replication study has two limitations in contrast to the original study regarding data. First, absolute patent counts were used instead of forward citation-weighted counts. Second, is was out of scope for this thesis to collect patent data from focal firm's current and historical divisions and subsidiaries. For this study, only patent data applied for by the corporate mother of the CVC funds was assessed. Third, the original study used the VentureXpert database. This database is a very common source across CVC research, however it was not accessible in this case. The Zephyr database was used instead.

4.1.1 Dependent variable

Wadhwa et al. (2016) follow prior research (Griliches, 1990) and operationalize investor innovation performance "as the total number of forward citations that a firm *i*'s patents applied for in year *t* received in the next 7 years" (p. 102). Using forward citation-weighted counts is standard practice in this research field (Lahiri, 2010; Trajtenberg, 1990; Yang et al., 2010; Ziedonis, 2007) because this measure adds another layer of qualitative analysis to mere patent counts. However, for ease of analysis, this paper operationalizes innovation performance as the total number of patents applied for by a firm *i* in year *t*. This is a clearly a limitation regarding the replication exercise.

4.1.2 Explanatory variables

Following Wadhwa et al.'s (2016) analysis, this study also assumes that the CVC portfolio of an investor *i* is constituted of "all startups in which a firm invested during the 4 years prior to and including the focal year" (p. 102). This approach is in line with standard practice in this research field. Wadhwa et al.'s (2016) do not specify how they handled corporate investors with multiple CVC funds. This paper assumes that the ventures invested in by either one of the funds together jointly form the corporate mother's CVC portfolio. The completion date of each CVC investment deal was identified using Zephyr. In case a second round of funding to the same startup was identified, this paper assumes a portfolio affiliation extension of another four years from the date of the second founding onwards. The deal completion year is considered as portfolio year zero. Like in Wadhwa et al.'s (2016) analysis, the explanatory and control variables are lagged by one year relative to innovation performance to avoid concerns regarding reverse causality.

Portfolio diversity. Following Wadhwa et al.'s (2016) analytical strategy, the authors' measure of portfolio diversity was replicated as closely as possible. In case assumptions had to be made about the interpretation of the original approach, these will be stated clearly.

The underlying idea of the measure is that heterogenic ventures represent heterogenic knowledge bases. Access to and exploitation of a varied, dissimilar innovation search space through interorganizational relations can result in potential organizational advantages for the corporate investor if diverse knowledge elements are successfully recombined. Essentially, CVC portfolios are viewed as social networks within which social capital resides (Wadhwa et al., 2016). In the original and present study, these networks are studied with regards to the content available amongst the nodes of the network. Together with the network structure approach, this is common in social capital research (Nahapiet & Ghoshal, 1998). The content approach is particularly appropriate for this research context since it studies the qualitative nature of portfolio affiliations. As standard practice in network structure research, the relationships between all possible combinations of nods are analyzed. This means that the investor is not only compared to the portfolio ventures, but all ventures are also compared to each other. The basis of comparison is the degree to which knowledge between two nods is dissimilar. The original study adapts a measure developed by Rodan and Galunic (2004). In essence, this measure "incorporates information on the knowledge distance between a focal actor and each of its partners and the distance among the partners" (Wadhwa et al., 2016, p. 102).

Knowledge differences can ideally be assessed using classification of patent data. However, all research in this field faces the same constrain, namely that many new ventures do not yet hold patents at the time they join knowledge-based interorganizational networks (Dushnitsky & Lenox, 2005; Jiang, Toa, & Santoro, 2010; Keil, Maula, Schildt, & Zahra, 2008; Lin & Lee, 2011). Following the approach of these and other researchers, standard industry classification (SIC) codes are used in this analysis to asses knowledge similarities. The underlying reason is the assumption that activity in different SIC codes represents different underlying technologic activity (Bryce & Winter, 2009). As such, Wadhwa et al. (2016) argue that "the distribution of a firm's four-digit SICs reflects the distribution of its technical knowledge" (p. 102).

In line with this approach, all primary and secondary SIC codes were collected for the corporate investors for each year *t* of observation via Compustat. Private companies are not allocated SIC codes, therefore, this approach was not possible for the startup. Databases like Zephyr or VentureXpert however offer a concordance mapping. The original study used VentureXpert. This was not possible for this thesis due to database access restrictions.

Accordingly, the SIC codes of the ventures were collected via Zephyr. In contrast to Compustat, Zephyr does not contain information on historical SIC codes. Therefore, the assumption that the SIC codes in which the venture is present at the time of CVC investment do not change over the course of its 4-year CVC portfolio affiliation was made. This assumption might be a limitation compared to the original study, but it is not clear from the original paper if Wadhwa et al. (2016) had access to historical SIC codes for the ventures.

Following Wadhwa et al.'s (2016) analytical strategy, portfolio diversity is assessed in three steps. First, knowledge distance is computed at the dyad-level between all nods of a portfolio in year *t-1*. An adapted index delveoped by Jaffe (1986) is used and the following formula applied:

$$d_{ijt} = 1 - \left[\sum_{k=1}^{K} f_{ik} \, f_{jk} \Bigg/ \left(\sum_{k=1}^{K} {f_{ik}}^{2} \right)^{1/2} \left(\sum_{k=1}^{K} {f_{jk}}^{2} \right)^{1/2} \right]$$

(Wadhwa et al., 2016, p. 102)

As the interpretation of the methodology is not entirely unambiguous from the original paper, the present paper makes the following assumptions: (1) In accordance with Jaffe's (1986) original approach of grouping patent classes into categories, all SIC codes were re-classified from the four-digit to the three-digit level. The three-digit levels thus represent the industry categories k. (2) Hence, f_{ik} = fraction of all of investor i's primary and secondary four-digit SIC codes in industry category k. (3) f_{jk} = fraction of all of venture j's primary and secondary four-digit SIC codes in industry categories k. (4) $K =$ total number of industry classes k that the investor and venture were present in. The resulting distance between two firms, d_{ijt}, ranges between 0 (complete similarity) and 1 (complete dissimilarity). Next, a distance matrix, D_t, including every combination of pairs in a portfolio and all respective pairwise distances d_{ijt}, was set up for every year of observation. The resulting matrix is square and symmetric.

In a second step, portfolio diversity is computed by calculating a "value of the uniqueness of each portfolio firm *j* in corporate investor *i*'s portfolio" (Wadhwa et al., 2016, p. 102). The uniqueness of a portfolio firm is interdependent to every pairs' respective knowledge distances and thus, to all other nods' uniqueness values. The following formula is applied to measure the uniqueness of a portfolio firm j, u_j:

$\lambda u_j = \sum_k d_{jk}$ (Rodan & Galunic, 2004; Wadhwa et al., 2016, p. 102)

The solution is found in the eigen equation $\lambda U = DU$ "where *D* is the matrix of pairwise distances between the investor *i* and each of its portfolio ventures, *U* is an eigenvector of *D* and λ is its eigenvalue." (Wadhwa et al., 2016, p. 103) The largest eigenvalue (λ) is used. Lastly, portfolio diversity of an investor *i* throughout its CVC network of ventures *j* in year *t-1* is computed as:

$Portfolio\ Diversity_{it-1} = \frac{1}{N}\sum_{j=1}^{N} d_{ij}\ \lambda u_j$ (Wadhwa et al., 2016, p. 103)

"where d_{ij} is the SIC-based distance of venture *j* from investor *i* and λu_j is *j*'s uniqueness score computed for all of *i*'s portfolio firms." (Wadhwa et al., 2016, p. 103) The original study notes that this measure captures knowledge distances (based on SIC codes) between all portfolio nods for a given corporate investor *i* in a given year *t* and is therefore an appropriate measure of portfolio diversity.

The replication exercise ends here. Two new moderators will be introduced in the following to extend the original study.

Investor absorptive capacity. In line with Wadhwa et al.'s (2016) argumentation and prior studies (e.g., Dushnitsky & Lenox, 2005), this study does also not expect that the marginal effects of CVC portfolio diversity on firm innovation performance will be uniform across all corporate investors. Drawing on Dushnitsky and Lenox's (2005) research, "the degree to which a firm may learn from its CVC investments will depend in part on the absorptive capacity of the firm" (p. 620). Absorptive capacity is captured in various ways across the literature. While is is often measured as contemporaneous research and development costs (e.g. Cohen & Levinthal, 1990), Dushnitsky and Lenox's (2005) have proposed that following this approach is not appropriate for the CVC setting because CVC funds and R&D departments are likely to compete for the same resources. Past R&D expenditure levels will be considered to avoid these concerns (Cohen & Levinthal, 1990; Dushnitsky & Lenox, 2005). Further, following Cohen and Levinthal's (1990) approach, absorptive capacity will be captured as R&D intensity to control for firm size. In summary, an investor *i*'s absorptive capacity in year *t-1* is operationalized as the past three-year average of R&D intensity, which equals to the ratio of R&D expenditures to sales.

Portfolio geographic diversity. In addition to an investor's level of absorptive capacity, the marginal effects of CVC portfolio diversity on firm innovation performance will also depend in part on the composition of the porfolio in term of the geografic spread of the ventures. Like the technological heterogenity across the network, geographic diversity adds another dimension of complexity to the extramural search space and thus, should also be taken into consideration. The present study uses Belderbos et al.'s measure of geographic diversity. The following formula is applied to measure geographic diversity of an investor *i*'s portfolio of CVC activity in year *t-1*:

$$Geo.diversity_{CV_{t-1}} = 1 \,/ \sum_{j \in L(t)} \left[\frac{\sum_{t-3}^{t-1} cv_{l,t}}{CV_t} \right]^2 \qquad \text{(Belderbos et al., 2018, p. 26)}$$

where "$cv_{l,t}$ refers to the number of CVC investments of the focal firm in country l at time *t-1*, *L* is the total number of CVC target countries, and *CV* is the total number of CVC investments" (Belderbos et al., 2018, p. 26). The resulting geographic distance within a portfolio "varies between 1 (concentration of activities in one country) and the theoretical maximum equal to the total number of countries represented in the portfolio" (Belderbos et al., 2018, p. 26). The measure is an inversed Herfindahl index.

4.1.3 Control variables

With the aim of replicating Wadhwa et al.'s (2016) research strategy, the present study controls as closely as possible for the variables used in the original study. These variables are *Investor CVC experience, Investor size, Investor current ratio, Investor age, Portfolio size, Investor alliances* and *Investor acquisitions*. This replication presents three limitations in contrast to the original study regarding this aspect: (1) The present study does not control for R&D intensity, but instead uses R&D intensity as a measure of absorptive capacity and thus, as an interaction effect. (2) The present study does not control for 'Investor Technological Diversity' because it was out of scope for this study to assess the patents applied for by firm *i* regarding their technology classes. (3) The present study does also not control for 'Investor Total Patents' because the patent count of firm *i* in year *t-1* is used as the dependent variable in this study. (4) Only two of the six controls could be narrowly replicated. For the other two controls, assumptions about their computation had to be made because the described approach was unclear. The following four variables were replicated following the exact same approaches as in the original analysis:

Investor size. This measure controls for firm size and is operationalized as the "natural log of sales in $US million" (Wadhwa et al., 2016, p. 103) year *t-1*.

Investor current ratio. Current ratio is a liquidity ratio and controls for slack. The measure is standardly computed as current assets divided by current liabilities in year *t-1*..

Investor age. The number of years since the firm's incorporation.

Portfolio size. All ventures that are part of a portfolio in year *t-1*.

The present study tried to replicate the next two variables as closely as possible. However, assumptions had to be made.

Investor CVC experience. Wadhwa et al. (2016) description of the variable computation is ambiguous. The article states that investor CVC experience is measured as the weighted, cumulative count of CVC investments in every year *t-1* since the investor's first investment. The present study assumes that the following approach is appropriate: First, the complete CVC deal history of every investor was obtained via Zephyr. Second, for every year of observation, all CVC deals of an investor *i* up until that year were counted. Third, the deal completion dates served as reference to identify the year of every CVC fund's first investment activity. Fourth, the time distance between the completion date of each deal and the initial investment date was calculated. Lastly, these durations were used to weight the historical cumulative counts of CVC investments per investor per year. Investor CVC experience was log transformed.

Investor alliances and *Investor acquisitions.* In the author's understanding, Wadhwa et al. (2016) construct alliance and acquisition portfolios similar to CVC portfolios. For every alliance or acquisition deal, a four-year time horizon of affiliation of the target firm to the respective portfolio is assumed. Portfolio sizes are computed. Wadhwa et al. (2016) state that they use "straight-line depreciation to account for the declining influence of acquisitions and alliances over time" (p. 103). This statement is ambiguous. The authors further state that they control for alliances and acquisitions because these are alternative sources of distant knowledge. Drawing on this argument, the present study assumes that each firm in the alliance and acquisition portfolios is weighted according to its duration of affiliation to the portfolio before the portfolio size is computed to account for the mentioned decrease in influence. As such, alliances and acquisitions made by investor *i* in year *t-1* were weighted as follows: In the first year of affiliation to the portfolio, the portfolio firm was weighted with 100%; in den second year with 75%; in the third year with 50% and in the fourth and last year with 25%. Minority stake acquisitions were ignored to avoid overlaps with CVC activity. Due to data restriction in the Zephyr database, alliance ties are limited to joint ventures.

4.2 Model specification and estimation

In contrast to the original study, this replication measures innovation performance using patent counts in absolute terms and not weighted by forward citations. Nevertheless, the dependent variable remains a non-negative count variable. Like in the original paper, innovation performance is significantly over-dispersed meaning that the variance is significantly larger than the mean as seen in table 1. Therefore, the main assumption of a Poisson distribution is violated and a negative binominal model required. Replicating Wadhwa et al. (2016) analytical strategy, year fixed effects were included for two reasons: (1) to control for unobserved, systematic time period effect; (2) to control for unmeasured differences in the setup of CVC arms. A Hausmann test was conducted to confirm the fixed effects approach. As previously mentioned, all dependent variables are lagged one year relative to the independent variable. *Portfolio diversity, Investor sales* and *Investor CVC experience* were log transformed to replicate the original analytical strategy narrowly.

In order to compare the fit of different models, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used. The model with the lowest BIC / AIC value is the best according to the respective criterion. While both BIC and AIC are similar, they can lead to different conclusions since BIC penalizes additional independent variables and thus preferers a less complex model over a more complex model with all else being equal.

5 Results

Table 1 shows the descriptive statistics. Table 2 shows the correlation matrix. Table 3 shows the results of five different models fitted to the data using a negative binominal panel regression with year fixed effects. The same table but with the original coefficients can be found in the appendix. In table 3, model 1 is the baseline model with only controls. Model 2 introduces portfolio diversity. Model 3 introduces the square term of portfolio diversity. With the introduction of the square term of portfolio diversity, both the linear and its square become insignificant. Therefore, the following models continue with the linear term only. Model 4 introduces both moderation variables, but so far without interaction. The interaction terms are introduced in model 5 which represents the full model.

According to AIC, model 5 fits the data best. However, according to BIC, model 4 should be preferred because it is sparser than model 5. Since the extension part of this analysis is especially interested in the interaction terms, this study will rely on the AIC result and concentrate on model 5 for the further analysis.

The yearly time dummies show that the overall level of innovation performance decreases in 2015 – the last year for which patent data was collected. Effectively, the overall level of innovation performance decreases by 20% compared to 2010. This can most probably be explained with the lag between patent application and publication dates. This study collected the data at the end of 2018. The author assumed that a three years lag is sufficient to reduce the risk of the patent data being right-censored. Nevertheless, this might still be a concern. Apart from this effect in 2015, the dummies accounting for the yearly fixed effects show no significant effects.

Next, incidence rate ratios (IRR) will be used to quantify the size of the estimated effects. IRRs can be directly calculated based on the estimated coefficients and are more intuitive to interpret. The term IRR refers to the ratio between incident rates for different levels of independent variables. In this case, the incident ratio is the patent count per year. In this analysis, the IRR for portfolio diversity (ln) is estimated to be 1.051. Therefore, all else equal, a one unit increase in portfolio diversity (ln) leads on average to a 5.1% increase in innovation performance. Since a one unit increase in portfolio diversity (ln) is difficult to interpret, one can instead calculate the difference between the 25th percentile and the 75th percentile of portfolio diversity (ln) and thereby, calculate how many more patents on average a firm with a relatively diverse portfolio (75th percentile, Portfolio diversity (ln) =0.39) has compared to a firm with relatively less diverse portfolio (25th percentile, Portfolio diversity (ln) = -1.99). Accordingly, model 5 suggests that while controlling for other variables a firm with a relatively diverse CVC portfolio has on average ~12.5% more patents than a firm with a relatively less diverse portfolio. This effect is highly significant on the 1% level.

A non-linear effect of portfolio diversity on innovation performance was expected; however, this study fails to replicate the inverse U-shaped relationship reported in the original study. Model 3 illustrates that the square term of portfolio diversity (ln) has no significant effect on investor innovation performance and that indeed, its introduction makes the linear term insignificant as well. This also holds true if the squared diversity term is included in any of the other models presented and also in additional models with other combinations of control variables. The relationship between portfolio diversity and patent count and therefore investor innovation performance seems to be linear and positive in all models of the present analysis. Thus, this study fails to confirm the existence of an inverse-u shaped relationship between the two beforementioned variables in a generalized industry sample setting. Hypothesis 1 is not supported.

Based on calculations similar to those performed above, the effect size of absorptive capacity suggests that a relatively high absorptive capacity (75th percentile) leads on average to ~32.5% more patents than a relatively low absorptive capacity (25th percentile). This effect is as expected and in line with literature. The coefficient is significant on the 5% level. The interaction term however suggests that with all else equal, a relatively high absorptive capacity in combination with a relatively diverse portfolio on average has a negative effect on investor innovation performance. In contrast, a high geographic diversity in combination with a high portfolio diversity has a positive effect.

In conclusion, first, the result show that the moderation effect of investor absorptive capacity on the relationship between portfolio diversity and patent count, and therefore investor innovation performance, is negative. Second, the moderation effect of portfolio geographic diversity on the relationship between portfolio diversity and patent count, and therefore investor innovation performance, is positive. Both effects are significant on the 1% and 5% level respectively. Consequently, hypothesis 2 id not supported, whereas hypothesis 3 is supported.

The margins graphs in figure 1 and 2 illustrate the two moderation effects based on model 5. Both graphs show the effect of different levels of portfolio diversity of an average firm on the number of patents per year. Figure 1 shows this relationship conditional on investor absorptive capacity, whereas figure 2 shows the relationship conditional on portfolio geographic diversity. The 95% confidence intervals are also displayed. The different levels of both absorptive capacity and geographic diversity are selected based on the 10th, the 50th and the 90th percentile in the sample. Note that the 25th percentile of portfolio diversity (ln) in the sample is -1.99 whereas the 75th percentile is 0.39. The graphs empathize the previously reported findings. On the one hand, for firms with high levels of absorptive capacity, the relationship between portfolio diversity and innovation performance is unexpectedly less steep than for firms with low levels of absorptive capacity. This result contradicts the expected outcome. One possible explanation could be that investor firms with high levels of absorptive capacity, and therefore relatively strong R&D departments, rely more on their internal research and development programs than on external sources of knowledge creation. These firms might therefore not invest sufficient resources into CVC relationships to maintain close relationships with partner firms. However, close collaboration might be a boundary condition to benefit from the full potential of CVC networks. However, this is only a presumption and more research is needed to clarify these aspects. On the other hand, for firms with high geographically diverse portfolios, the relationship between portfolio diversity and innovation performance is steeper than for firms with low geographic portfolio diversity. This result supports the expected outcome.

5.1 Robustness analysis

In line with Wadhwa et al. (2016), a robustness analysis was conducted using a subsample of observations for which portfolio size was greater than zero. This eliminated firm-years in which the corporate investors did not have any ventures in its CVC portfolio. Additional subsamples were constructed to control for further potential selection biases. One additional subsample is constructed with high-tech firms only. The second one is constructed only with firms headquartered in the US. The regressions were rerun for all subsamples. None of the three subsamples shows evidence of the existence of a non-linear relationship between portfolio diversity and innovation performance, hence, supporting the present paper's approach to continue the analysis with linear models.

6 Contributions, limitations and future research

6.1 Contribution

The present paper makes serval contribution to the ongoing research streams about interfirm knowledge-sharing relationships by examining under which condition CVC partnerships are beneficial for the corporate investor. Moreover, this study contributes to the scientific debate about the importance of replication by actually replicating a paper. Ultimately, this paper contributes to the discussion about which managerial implications can be drawn from CVC research by showing that generalizability of many studies is a concern. Managers are advised to review carefully the study settings before relying on such findings for important decision making. Academics are advised to conduct more such quasi-replication studies.

6.2 Limitations and future research

Naturally, the presented study has several limitations which can be broadly divided into limitations regarding a preferable narrow replication of the original paper and general limitations regarding data and methodology. Some of these limitations have already been discussed and will therefore only be mentioned shortly in the following. Following limitations regarding a narrow replication of the original paper exist. First, the exact methodology of the original paper is not always unambiguous and therefore, assumptions had to be made. This is clearly a limitation; however, the assumptions that were made should not alter the results substantially due to their limited scope. Second, due to different data sources, a different timeframe and a broader industry setting, the replication is naturally limited at least in the narrow sense of replication. However, the difference in timeframe and industry is a desired extension of the scope of the original paper and therefore only partially a limitation. Following limitations regrading data and methodology are as follows. First, since the sample of CVC funds used in this study is not exhaustive, a certain selection bias could be present. The author has tried to mitigate this risk by choosing CVC funds based on specific criteria, but the risk of selection bias has to be considered.

Second, the use of SIC codes as a proxy for technological knowledge and capabilities is a limitation in itself. Not only because technological capabilities and knowledge are not directly dividable into different industries, but also because of the limitations of SIC codes in general. Those limitations are that SIC codes do not perfectly describe the true industry of a firm and that if a firm has more than one industry classification, those SCI codes are not weighted. Furthermore, historical SIC classification could not be retrieved for all sample firms and therefore changes in the industry classification for some firms are not considered. However, the effect is assumed to be minor.

Third, given the precaution to take into consideration a patent publication lag of at least three years, the present study chose 2015 as the last year to observe CVC investments. As a consequence, many active corporate venture capital firms had to be excluded from the present study because they were only founded or only made their first investment after 2015. Nonetheless, a 14-year high in aggregated deal size of U.S. CVC activity as reported by the National Venture Capital Association (2018) demonstrates that corporations are relying more and more on CVC as a vehicle for innovation. The growing numbers also suggest that startups in turn value more and more what corporate investors bring to the table. The VC community expects that CVC activity will not slow down (National Venture Capital Association, 2018) in the coming year. Therefore, an interesting avenue for future research is to analyze the relationship between portfolio diversity and firm innovation performance again in the future considering the latest CVC activities.

Considering these limitations and the findings presented above, future research is needed to further clarify the way CVC influences innovation performance. Promising areas for future research are manifold. One the one hand, the question about the inverse U-shaped relationship between portfolio diversity and investor innovation performance remains to be definitively answered and therefore open for future research. Different definitions of portfolio diversity such as inter-industry and intra-industry portfolio diversity could potentially help in this regard. Additionally, a broad study across different industries with forward citation-weighted patent counts, historical SIC codes, and potentially a longer timeframe would be ideal. On the other hand, broader research into the different channels through which portfolio diversity influences innovation performance is needed and could help to better understand the role of the different moderation effects. Especially the significantly negative moderation effect of absorptive capacity is against intuition and should be further analyzed.

7 Conclusion

All in all, this study fails to replicate and thus verify the generalizability of the original paper's main finding of an inverted U-shaped relationship between portfolio diversity and innovation performance. In contrast to the original analysis, a generalize industry setting was chosen instead of a specific high-tech industry. Nevertheless, the original results could also not be replicated in a sub-sample of high-tech firms. Across industries, the present study does however find support for a positive linear relationship. The original two moderators were not replicated. Instead, two new moderators were introduced to the model. Both extension moderators are significant in the final model, however, investor absorptive capacity shows an unexpected negative interaction effect. Portfolio geographic diversity on the other hand shows a positive effect as expected.

These findings do not mean that the previous findings should be rejected. On the one hand, this replication is subject to various limitations. On the other hand, scholars suggest that such failed replication exercises only "mean that the balance of evidence regarding the existing results moves toward questioning the original result" (Bettis, et al., 2016). In any case, the generalizability of the results found by Wadhwa, Phelps and Kotha (2016) is under question. Future studies of the issue are needed.

Table 1

Descriptive statistics
N = 466 in Model 5 (full model)

	Variables	Mean	S.D.	Min	Max	Skewness	Kurtosis	Percentile 25	Percentile 50	Percentile 75
(1)	Innovation performance	404.83	909.27	0.00	7,685.00	4.64	28.49	33.00	117.50	322.00
(2)	Portfolio diversity, t – 1	0.96	1.13	0.00[4]	9.00	2.28	12.26	0.14	0.52	1.48
(3)	Portfolio diversity (ln), t – 1	-1.68	3.01	-7.77	2.20	-1.26	3.11	-1.99	-0.65	0.39
(4)	Portfolio diversity (ln) squared, t – 1	11.87	22.51	0.00	60.45	1.65	3.80	0.25	1.02	4.36
(5)	Investor absorptive capacity, t – 1	0.27	0.22	0.00	0.88	0.56	2.25	0.06	0.20	0.43
(6)	Portfolio geo. diversity, t – 1	1.51	1.03	0.00	5.88	0.85	4.45	1.00	1.38	2.00
(7)	Investor size, t – 1	43,753.55	66,278.67	400.95	470,171.00	3.76	19.97	8,696.00	22,467.85	50,406.88
(8)	Investor size (ln), t – 1	9.92	1.32	5.99	13.06	-0.28	2.91	9.07	10.02	10.83
(9)	Investor current ratio, t – 1	1.91	1.04	0.58	7.15	1.98	7.96	1.20	1.65	2.21
(10)	Investor age, t – 1	62.37	50.75	2.00	233.00	0.99	3.33	24.00	41.00	103.00
(11)	Investor acquisitions, t – 1	4.87	6.02	0.00	51.25	3.39	19.99	1.00	3.25	6.25
(12)	Investor joint ventures, t – 1	0.75	1.12	0.00	8.25	2.47	11.86	0.00	0.25	1.00
(13)	Portfolio size, t – 1	12.40	27.60	0.00	239.00	6.02	44.72	1.00	5.00	12.00
(14)	Investor CVC experience, t – 1	27.93	106.63	1.00[5]	1,134.92	8.89	85.55	1.00	6.09	20.80
(15)	Investor CVC experience (ln), t – 1	1.77	1.66	0.00	7.03	0.47	2.36	0.00	1.81	3.04

[4] To avoid problems arising in case of ln(0), zero values were replaced by the second lowest value in the distribution (0.00042).
[5] To avoid problems arising in case of ln(0), zero values were replaced by the second lowest value in the distribution (1).

Unfortunately, Wadhwa et al. (2016) do not specify how they have adressed this issue and therefore, the author of the present study had to make an assumption about how to treat this problem. However, this might not be in line with the original approach.

Table 2

Correlations
N = 466 in Model 5 (full model)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Innovation performance	1														
(2) Portfolio diversity, t – 1	0.210***	1													
(3) Portfolio diversity (ln), t – 1	0.0853	0.627***	1												
(4) Portfolio diversity (ln) squared, t – 1	-0.0501	-0.410***	-0.958***	1											
(5) Investor absorptive capacity, t – 1	0.0663	-0.0948*	0.0742	-0.169***	1										
(6) Portfolio geo. diversity, t – 1	0.194***	0.474***	0.656***	-0.617***	0.230***	1									
(7) Investor size , t – 1	0.185***	0.200***	0.133**	-0.0731	-0.303***	0.0109	1								
(8) Investor size (ln), t – 1	0.287***	0.353***	0.268***	-0.184***	-0.282***	0.171***	0.720***	1							
(9) Investor current ratio, t – 1	0.0690	-0.113*	-0.0191	-0.0381	0.416***	0.0447	-0.226***	-0.367***	1						
(10) Investor age, t – 1	0.0597	0.194***	0.198***	-0.146**	-0.0868	0.152***	0.282***	0.382***	-0.179***	1					
(11) Investor acquisitions, t – 1	0.585***	0.105*	0.0568	-0.0403	0.0803	0.152***	0.0713	0.210***	0.116*	-0.0176	1				
(12) Investor joint ventures, t – 1	0.0395	0.206***	0.0651	-0.00336	-0.267***	0.0215	0.338***	0.351***	-0.137**	0.0754	0.0217	1			
(13) Portfolio size, t – 1	0.315***	0.621***	0.315***	-0.204***	0.206***	0.283***	0.0414	0.186***	0.121**	-0.0561	0.380***	0.0429	1		
(14) Investor CVC experience, t – 1	0.237***	0.582***	0.210***	-0.110*	0.156***	0.212***	0.0290	0.132**	0.0474	-0.0178	0.212***	0.0712	0.889***	1	
(15) Investor CVC experience (ln), t – 1	0.175***	0.660***	0.589***	-0.482***	0.250***	0.555***	0.0784	0.286***	0.0161	0.0925*	0.219***	0.104*	0.634***	0.504***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Most of the variables show moderate levels of correlations. High correlation is naturally observed between the linear and squared term of portfolio diversity. In addition, high levels are observed between portfolio diversity (ln) and portfolio geo. diversity (ln), between portfolio diversity (ln) and investor size (ln) and between portfolio diversity (ln) and CVC experience. These high levels raise concerns about possible multicollinearity issues.

Table 3

Negative binominal panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	0.657	0.931	0.930	0.818	0.899
	(0.428)	(0.629)	(0.632)	(0.572)	(0.641)
Portfolio diversity (ln), t – 1		1.029***	1.030	1.034***	1.051***
		(0.010)	(0.047)	(0.013)	(0.018)
Portfolio diversity (ln) squared, t – 1			1.000		
			(0.005)		
Investor absorptive capacity, t – 1				2.223**	2.143**
				(0.738)	(0.709)
Portfolio geo. diversity, t – 1				0.967	0.992
				(0.046)	(0.047)
Portfolio diversity (ln) × Investor absorptive capacity					0.874***
					(0.045)
Portfolio diversity (ln) × Portfolio geo. diversity					1.032**
				1	(0.015)
Investor size (ln), t – 1	1.307***	1.273***	1.273***	1.284***	1.261***
	(0.091)	(0.092)	(0.093)	(0.095)	(0.095)
Investor current ratio, t – 1	1.013	1.023	1.023	0.996	1.000
	(0.043)	(0.043)	(0.043)	(0.042)	(0.041)
Investor age, t – 1	1.000	1.000	1.000	0.999	0.999
	1.307***	1.273***	1.273***	1.284***	1.261***
Investor acquisitions, t – 1	1.009	1.011	1.011	1.010	1.014*
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
Investor joint ventures, t – 1	0.990	0.999	0.999	1.001	1.008
	(0.023)	(0.023)	(0.023)	(0.025)	(0.025)
Portfolio size, t – 1	1.000	1.000	1.000	0.999	0.999
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Investor CVC experience (ln), t – 1	0.995	0.973	0.972	0.972	0.966
	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)
Observations	484	484	484	466	466
Number of groups	99	99	99	98	98
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-1865	-1860	-1860	-1782	-1778
Wald Chi 2	64.57	74.16	74.17	78.64	89.98
AIC	3753	3746	3748	3594	3589
BIC	3803	3800	3807	3656	3660

Irr se in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1

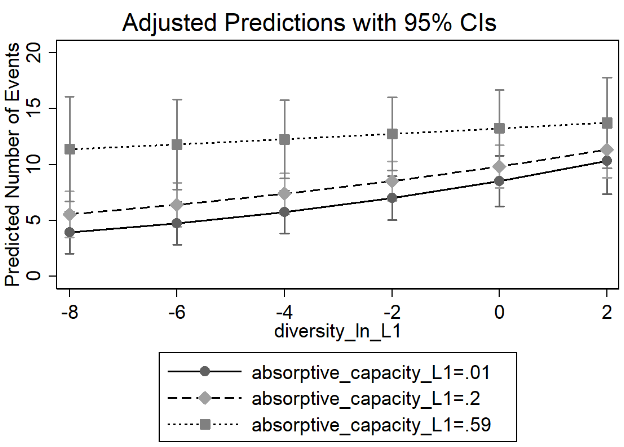
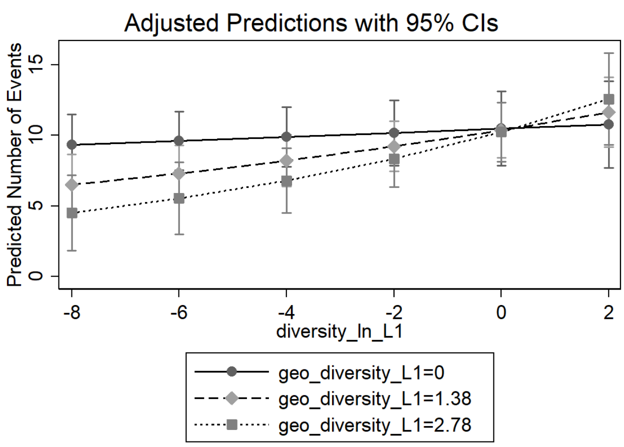


Figure 2



References

Agarwal, R., & Helfat, C. E. (2009). Strategic Renewal of Organizations. *Organization Science*, *20*(2), 281–293.

Ahuja, G., & Katila, R. (2004). Where Do Resources Come from? The Role of Idiosyncratic Situations. *Strategic Management Journal*, *25*(8/9), 887–907.

Belderbos, R., Jacob, J., & Lokshin, B. (2018). Corporate venture capital (CVC) investments and technological performance: Geographic diversity and the interplay with technology alliances. *Journal of Business Venturing*, *33*(1), 20–34. https://doi.org/10.1016/j.jbusvent.2017.10.003

Benson, D., & Ziedonis, R. H. (2009). Corporate Venture Capital as a Window on New Technologies: Implications for the Performance of Corporate Investors When Acquiring Startups. *Organization Science*, *20*(2), 329–351. https://doi.org/10.1287/orsc.1080.0386

Bettis, R. A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal*, *33*(1), 108–113.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, *37*(2), 257–261. https://doi.org/10.1002/smj.2477

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, *37*(11), 2193–2203. https://doi.org/10.1002/smj.2580

Block, J., & Kuckertz, A. (2018). Seven principles of effective replication studies: strengthening the evidence base of management research. *Management Review Quarterly*, *68*(4), 355–359. https://doi.org/10.1007/s11301-018-0149-3

Bryce, D. J., & Winter, S. G. (2009). A General Interindustry Relatedness Index. *Management Science*, *55*(9), 1570–1585.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Chesbrough, H. (2002, March 1). Making Sense of Corporate Venture Capital. *Harvard Business Review*, (March 2002), 90–99.

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

Drover, W., Busenitz, L., Matusik, S., Townsend, D., Anglin, A., & Dushnitsky, G. (2017). A review and road map of entrepreneurial equity financing research: Venture capital, corporate venture capital, angel investment, crowdfunding, and accelerators. *Journal of Management*, *43*(6), 1820–1853. https://doi.org/10.1177/0149206317690584

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., & 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

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2016). Replication in strategic management. *Strategic Management Journal*, *37*(11), 2191–2192. https://doi.org/10.1002/smj.2581

Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science*, *47*(1), 117–132.

Gavetti, G., & Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative Science Quarterly*, *45*(1), 113–137. https://doi.org/10.2307/2666981

Gompers, P. A. (2002). Corporations and the Financing of Innovation: The Corporate Venturing Experience. *Federal Reserve Bank of Atlanta Economic Review*, *87*(4), 1–17.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*(S2), 109–122. https://doi.org/10.1002/smj.4250171110

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, *28*(4), 1661–1707.

Hamm, S. J. W., Jung, M. J., & Park, M. (2018). *How Transparent are Firms About Their Corporate Venture Capital Investments?* (SSRN Scholarly Paper No. ID 3090797). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=3090797

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review*, (5), 984.

Jiang, R. J., Toa, Q. T., & Santoro, M. D. (2010). Alliance portfolio diversity and firm performance. *Strategic Management Journal*, *31*(10), 1136–1144.

Katila, R., & Ahuja, G. (2002). Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. *The Academy of Management Journal*, *45*(6), 1183–1194. https://doi.org/10.2307/3069433

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

Keil, T., Maula, M., Schildt, H., & Zahra, S. A. (2008). The Effect of Governance Modes and Relatedness of External Business Development Activities on Innovative Performance. *Strategic Management Journal*, *29*(8), 895–907.

Khanna, T., Gulati, R., & Nohria, N. (1998). The Dynamics of Learning Alliances: Competition, Cooperation, and Relative Scope. *Strategic Management Journal*, *19*(3), 193–210.

Kiri, B., Lacetera, N., & Zirulia, L. (2018). Above a swamp: A theory of high-quality scientific production. *Research Policy*, *47*(5), 827–839. https://doi.org/10.1016/j.respol.2018.01.011

Lahiri, N. (2010). Geographic Distribution of R&D Activity: How Does it Affect Innovation Quality? *Academy of Management Journal*, *53*(5), 1194–1209. https://doi.org/10.5465/amj.2010.54533233

Lane, P. J., & Lubatkin, M. (1998). Relative Absorptive Capacity and Interorganizational Learning. *Strategic Management Journal*, *19*(5), 461–477.

Lavie, D., & Miller, S. R. (2008). Alliance Portfolio Internationalization and Firm Performance. *Organization Science*, *19*(4), 623–646.

Lin, S.-J., & Lee, J.-R. (2011). Configuring a corporate venturing portfolio to create growth value: Within-portfolio diversity and strategic linkage. *Journal of Business Venturing*, *26*(4), 489–503. https://doi.org/10.1016/j.jbusvent.2009.10.005

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Nahapiet, J., & Ghoshal, S. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, *23*(2), 242–266. https://doi.org/10.2307/259373

National Venture Capital Association. (2018). NVCA 2018 Yearbook. National Venture Capital Association by PitchBook Data, Inc.

Phelps, C. C. (2010). A longitudinal study of the influence of alliance network structure and composition on firm exploratory innovation. *The Academy of Management Journal*, *53*(4), 890–913.

Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, *March-April*, 21.

Rodan, S., & Galunic, C. (2004). More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal*, *25*(6), 541–562. https://doi.org/10.1002/smj.398

Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, *40*(10), 1321–1331. https://doi.org/10.1016/j.respol.2011.06.009

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Stuart, T. E., & Podolny, J. M. (1996). Local Search and the Evolution of Technological Capabilities. *Strategic Management Journal*, *17*, 21–38.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics (RAND Journal of Economics)*, *21*(1), 172–187. https://doi.org/10.2307/2555502

Vasudeva, G., & Anand, J. (2011). Unpacking Absorptive Capacity: A Study of Knowledge Utilization from Alliance Portfolios. *Academy of Management Journal*, *54*(3), 611–623. https://doi.org/10.5465/AMJ.2011.61968108

Wadhwa, A., & Kotha, S. (2006). Knowledge Creation through External Venturing: Evidence from the Telecommunications Equipment Manufacturing Industry. *The Academy of Management Journal*, *49*(4), 819–835. https://doi.org/10.2307/20159800

Wadhwa, A., Phelps, C., & Kotha, S. (2016). Corporate venture capital portfolios and firm innovation. *Journal of Business Venturing*, *31*(1), 95–112. https://doi.org/10.1016/j.jbusvent.2015.04.006

Weiblen, T., & Chesbrough, H. W. (2015). Engaging with Startups to Enhance Corporate Innovation. *California Management Review*, *57*(2), 66–90. https://doi.org/10.1525/cmr.2015.57.2.66

Yang, H., Phelps, C., & Steensma, H. K. (2010). Learning from What Others Have Learned from You: The Effects of Knowledge Spillovers on Originating Firms. *Academy of Management Journal*, *53*(2), 371–389. https://doi.org/10.5465/amj.2010.49389018

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

Zahra, S. A., & Hayton, J. C. (2008). The effect of international venturing on firm performance: The moderating influence of absorptive capacity. *Journal of Business Venturing*, *23*(2), 195–220. https://doi.org/10.1016/j.jbusvent.2007.01.001

Ziedonis, A. A. (2007). Real Options in Technology Licensing. *Management Science*, *53*(10), 1618–1633. https://doi.org/10.1287/mnsc.1070.0705

Appendix

Negative binominal panel regression with original coefficients

Negative binominal panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.421	-0.072	-0.073	-0.201	-0.106
	(0.652)	(0.676)	(0.680)	(0.700)	(0.713)
Portfolio diversity (ln), t – 1		0.029***	0.029	0.034***	0.050***
		(0.010)	(0.045)	(0.013)	(0.017)
Portfolio diversity (ln) squared, t – 1			0.000		
			(0.005)		
Investor absorptive capacity, t – 1				0.799**	0.762**
				(0.332)	(0.331)
Portfolio geo. diversity, t – 1				-0.034	-0.008
				(0.047)	(0.048)
Portfolio diversity (ln) × Investor absorptive capacity					-0.134***
					(0.051)
Portfolio diversity (ln) × Portfolio geo. diversity					0.032**
					(0.014)
Investor size (ln), t – 1	0.268***	0.241***	0.241***	0.250***	0.232***
	(0.070)	(0.072)	(0.073)	(0.074)	(0.075)
Investor current ratio, t – 1	0.013	0.023	0.023	-0.004	0.000
	(0.043)	(0.042)	(0.042)	(0.042)	(0.041)
Investor age, t – 1	0.000	0.000	0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Investor acquisitions, t – 1	0.009	0.011	0.011	0.010	0.013*
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Investor joint ventures, t – 1	-0.010	-0.001	-0.001	0.001	0.008
	(0.023)	(0.023)	(0.023)	(0.025)	(0.025)
Portfolio size, t – 1	-0.000	-0.000	-0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Investor CVC experience (ln), t – 1	-0.005	-0.028	-0.028	-0.029	-0.034
	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
Observations	484	484	484	466	466
Number of groups	99	99	99	98	98
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-1865	-1860	-1860	-1782	-1778
Wald Chi 2	64.57	74.16	74.17	78.64	89.98
AIC	3753	3746	3748	3594	3589
BIC	3803	3800	3807	3656	3660

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Poisson panel regression results

Poisson panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Portfolio diversity (ln), t – 1		0.024***	0.152***	0.025***	0.025***
		(0.002)	(0.007)	(0.002)	(0.004)
Portfolio diversity (ln) squared, t – 1			0.015***		
			(0.001)		
Investor absorptive capacity, t – 1				0.206	0.267**
				(0.134)	(0.135)
Portfolio geo. diversity, t – 1				-0.008	0.011
				(0.007)	(0.007)
Portfolio diversity (ln) × Investor absorptive capacity					-0.059***
					(0.014)
Portfolio diversity (ln) × Portfolio geo. diversity					0.011***
					(0.002)
Investor size (ln), t – 1	0.524***	0.519***	0.540***	0.469***	0.485***
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
Investor current ratio, t – 1	-0.004	0.004	-0.001	-0.032***	-0.036***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Investor age, t – 1	-0.182	-0.141	-0.140	-0.156	-0.164
	(0.425)	(0.422)	(0.424)	(0.423)	(0.423)
Investor acquisitions, t – 1	0.004***	0.007***	0.008***	0.002**	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Investor joint ventures, t – 1	-0.004	0.006	0.001	0.007*	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Portfolio size, t – 1	-0.004***	-0.003***	-0.004***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Investor CVC experience (ln), t – 1	0.047***	0.021***	0.009**	0.026***	0.020***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	484	484	484	466	466
Number of groups	99	99	99	98	98
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-6003	-5874	-5720	-5537	-5497
Wald Chi 2	2229	2473	2748	2460	2537
AIC	12028	11771	11467	11103	11027
BIC	12074	11822	11521	11161	11093

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Subsample observations with portfolio size >0 only

Negative binominal panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.258	-0.147	-0.078	-0.280	0.025
	(0.712)	(0.728)	(0.739)	(0.762)	(0.805)
Portfolio diversity (ln), t – 1		0.042**	0.013	0.045**	0.093
		(0.018)	(0.047)	(0.018)	(0.061)
Portfolio diversity (ln) squared, t – 1			-0.004		
			(0.006)		
Investor absorptive capacity, t – 1				0.483	0.528
				(0.358)	(0.361)
Portfolio geo. diversity, t – 1				-0.013	-0.032
				(0.051)	(0.052)
Portfolio diversity (ln) × Investor absorptive capacity					-0.195**
					(0.098)
Portfolio diversity (ln) × Portfolio geo. diversity					0.019
					(0.033)
Investor size (ln), t – 1	0.258***	0.256***	0.248***	0.273***	0.239***
	(0.078)	(0.079)	(0.080)	(0.081)	(0.085)
Investor current ratio, t – 1	0.061	0.064	0.062	0.038	0.041
	(0.041)	(0.040)	(0.040)	(0.042)	(0.041)
Investor age, t – 1	-0.001	-0.002	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Investor acquisitions, t – 1	0.012*	0.016**	0.015**	0.015*	0.018**
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Investor joint ventures, t – 1	-0.020	-0.013	-0.015	-0.017	-0.008
	(0.026)	(0.026)	(0.026)	(0.028)	(0.028)
Portfolio size, t – 1	-0.001	-0.001	-0.001	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Investor CVC experience (ln), t – 1	-0.010	-0.027	-0.023	-0.028	-0.031
	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
Observations	403	403	403	386	386
Number of groups	90	90	90	89	89
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-1550	-1547	-1547	-1476	-1473
Wald Chi 2	53.62	59.68	59.81	62.25	67.68
AIC	3124	3121	3122	2981	2981
BIC	3172	3173	3178	3040	3048

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Subsample high-tech industries only

Negative binominal panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.896	-0.735	-0.764	-0.528	0.309
	(0.794)	(0.794)	(0.803)	(0.855)	(0.838)
Portfolio diversity (ln), t – 1		0.036***	0.046	0.044***	0.133***
		(0.014)	(0.049)	(0.016)	(0.033)
Portfolio diversity (ln) squared, t – 1			0.001		
			(0.006)		
Investor absorptive capacity, t – 1				-0.001	-0.342
				(0.507)	(0.506)
Portfolio geo. diversity, t – 1				-0.032	0.010
				(0.053)	(0.056)
Portfolio diversity (ln) × Investor absorptive capacity					-0.308***
					(0.075)
Portfolio diversity (ln) × Portfolio geo. diversity					0.038**
					(0.017)
Investor size (ln), t – 1	0.400***	0.402***	0.405***	0.391***	0.306***
	(0.090)	(0.090)	(0.091)	(0.092)	(0.090)
Investor current ratio, t – 1	-0.036	-0.017	-0.017	-0.035	-0.015
	(0.045)	(0.043)	(0.043)	(0.045)	(0.042)
Investor age, t – 1	-0.009***	-0.010***	-0.010***	-0.010***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Investor acquisitions, t – 1	0.018**	0.020**	0.021**	0.020**	0.022***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)
Investor joint ventures, t – 1	-0.015	-0.005	-0.006	-0.004	0.003
	(0.042)	(0.041)	(0.041)	(0.041)	(0.041)
Portfolio size, t – 1	-0.003*	-0.003**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Investor CVC experience (ln), t – 1	0.009	-0.016	-0.017	-0.018	-0.023
	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)
Observations	257	257	257	253	253
Number of groups	54	54	54	54	54
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-1044	-1041	-1041	-1015	-1006
Wald Chi 2	48.04	56.23	56.64	55.87	80.15
AIC	2113	2108	2110	2060	2047
BIC	2155	2154	2159	2113	2107

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Subsample US-based firms only

Negative binominal panel regression with fixed effects
Dependent variable = Investor innovation performance

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	2.041***	2.309***	2.475***	2.100***	2.936***
	(0.685)	(0.699)	(0.719)	(0.737)	(0.803)
Portfolio diversity (ln), t – 1		0.024**	-0.015	0.015	0.089***
		(0.010)	(0.043)	(0.013)	(0.025)
Portfolio diversity (ln) squared, t – 1			-0.005		
			(0.005)		
Investor absorptive capacity, t – 1				0.649	0.066
				(0.450)	(0.473)
Portfolio geo. diversity, t – 1				0.050	0.016
				(0.051)	(0.063)
Portfolio diversity (ln) × Investor absorptive capacity					-0.228***
					(0.053)
Portfolio diversity (ln) × Portfolio geo. diversity					0.010
					(0.014)
Investor size (ln), t – 1	0.158**	0.138*	0.120	0.136*	0.084
	(0.069)	(0.071)	(0.073)	(0.072)	(0.075)
Investor current ratio, t – 1	-0.019	-0.009	-0.012	-0.028	-0.025
	(0.036)	(0.034)	(0.034)	(0.035)	(0.033)
Investor age, t – 1	-0.005**	-0.005**	-0.005**	-0.005*	-0.006**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Investor acquisitions, t – 1	-0.003	-0.001	-0.003	-0.007	0.002
	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
Investor joint ventures, t – 1	-0.065	-0.049	-0.045	-0.049	-0.045
	(0.044)	(0.044)	(0.045)	(0.043)	(0.043)
Portfolio size, t – 1	-0.001	-0.001	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Investor CVC experience (ln), t – 1	-0.007	-0.028	-0.022	-0.043*	-0.033
	(0.023)	(0.024)	(0.025)	(0.025)	(0.024)
Observations	208	208	208	201	201
Number of groups	43	43	43	42	42
Firm dummies	Fixed	Fixed	Fixed	Fixed	Fixed
Time dummies	Yes	Yes	Yes	Yes	Yes
Degrees of freedom	11	12	13	14	16
Log likelihood	-825.7	-822.6	-822.2	-801.6	-792.7
Wald Chi 2	44.23	50.66	51.73	55.20	76.42
AIC	1675	1671	1672	1633	1620
BIC	1715	1715	1719	1683	1676

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

R SYNTAX

```
############# PACKAGES NEEDED ############
library(readxl)
library(tidyr)
library(ggplot2)
library(RPostgres)
library(tidyverse)
library(tictoc)
library(dplyr)
library(zoo)
library(stringr)
library(lubridate)
############# FUNCTIONS ##################
# name:        fun_distance.rmd
# description:  calculate distance between two sic vectors
# input:       two vectors
# output:      single number -> distance
fun_distance <- function(firm_i,firm_j) {
  firm_i <- unlist(firm_i)
  firm_j <- unlist(firm_j)
  firm_i <- firm_i[!is.na(firm_i)]
  firm_j <- firm_j[!is.na(firm_j)]
  # supporting variables
  combined_sic <- c(firm_i,firm_j)
  unique_sic <- unique(combined_sic)
  max_unique_sic_num <- length(unique_sic)
  sic_num <- c(length(firm_i),
               length(firm_j))
  unique_sic_num <- c(length(unique(firm_i)),
               length(unique(firm_j)))
  max_sic_num <- max(sic_num)
  # 2nd step
  sic_recalc <- matrix(nrow = 2,ncol = max_unique_sic_num)
  row.names(sic_recalc) <- c("Firm i","Firm j")
  # -1 wird als nicht vorhanden code verwendet
  temp_sic <- matrix(NA,nrow = 2,ncol = max_sic_num)
  temp_sic[1,1:sic_num[1]] <- firm_i[1:sic_num[1]]
  temp_sic[2,1:sic_num[2]] <- firm_j[1:sic_num[2]]
  for (firm in 1:2) {
    for (i in 1:max_unique_sic_num) {
      temp_fraq <- sum(temp_sic[firm,] == unique_sic[i],na.rm=T) / sic_num[firm]
      sic_recalc[firm,i] <- temp_fraq
    }
  }
  # 3rd step
  arg_1 <- sic_recalc[1,] %*% sic_recalc[2,]
  arg_2 <- sic_recalc[1,] %*% sic_recalc[1,]
  arg_3 <- sic_recalc[2,] %*% sic_recalc[2,]
  distance <- 1 - (arg_1 / (sqrt(arg_2) * sqrt(arg_3) ))
  return(distance)
}
firm_investor <- c(1,2)
firm_portfolio <- c(2,3)
fun_distance(firm_i = firm_investor,firm_j = firm_portfolio)
save(fun_distance,file = "../R_data/fun_distance.Rda")
# name:        fun_diversity.Rmd
# description:  calculate portfolio diversity
# input:       all sic codes from portfolio firms
# output:      single number -> portfolio piversity
load("../R_data/fun_distance.Rda")
fun_diversity <- function(input_matrix) {
  number_firms <- nrow(input_matrix)
  distance_matrix <- matrix(nrow = number_firms, ncol = number_firms)
  colnames(distance_matrix) <- row.names(input_matrix)
  row.names(distance_matrix) <- row.names(input_matrix)
  for (i in seq_len(number_firms)) {
    for (j in i:number_firms) {
      distance_matrix[i,j] <- fun_distance(input_matrix[i,],input_matrix[j,])
      distance_matrix[j,i] <- distance_matrix[i,j]
    }
    distance_matrix[i,i] <- 0
  }
  # to avoid error if matrix is empty just return 0 and skip the rest
  if (nrow(distance_matrix) == 0) {
    portfoilio_diversity <- 0
    return(portfoilio_diversity)
  }
  e <- eigen(distance_matrix, symmetric = TRUE)
  # largest eigenvalue
  eigenvalue_lambda <- max(e$values)
  # corresponding eigenvector
  eigenvector_U <- e$vectors[,1]
  N <- length(eigenvector_U)
  fraq <- 1/N
  temp_vec <- c()
  for (j in seq_len(N)) {
    temp_vec[j] <- distance_matrix[1,j] *
      eigenvalue_lambda * eigenvector_U[j]
  }
  arg_1 <- sum(temp_vec)
  portfoilio_diversity <- abs(fraq * arg_1)
```

```
  return(portfoilio_diversity)
}
save(fun_diversity,file = "../R_data/fun_diversity.Rda")

---
title: "longlist: data enrichment w/ compustat"
---
# interface to WRDS database has previously been established
# connect to WRDS databse
wrds <- dbConnect(Postgres(),
                  host='wrds-pgdata.wharton.upenn.edu',
                  port=9737,
                  user='user',
                  password='password',
                  sslmode='require',
                  dbname='wrds')
save(wrds, file = "../R_data/wrds.Rda")
# import CVC funds longlist
longlist_raw <- as_tibble(read_excel("../Data_tables/CVC_longlist.xlsx"))
longlist_raw$cvc_fund <- toupper(longlist_raw$cvc_fund)
# TIDY DATA TO ARRIVE AT SHORTLIST
# exclude private companies so that only public companies remain
# exclude companies that did not make a cvc investment in observation period
# result: shorlist used for sample
shortlist_raw <- longlist_raw %>% filter(is.na(private) & is.na(no_invest_period) & is.na(no_patent_activity)) %>%
  select(-private, -no_invest_period, -no_patent_activity, -cvc_fund_2, -cvc_fund_3)
save(shortlist_raw, file = "../R_data/shortlist_raw.Rda")
# DATA ENRICHMENT W/ COMPUSTAT
# query 'Compustat North America - Fundamentals Annual' for financial and market information on parent companies using gvkey as identifier
# observation period 2010-2015
id_shortlist <- shortlist_raw$gvkey
shortlist_US <- shortlist_raw %>% filter(country == "US")
id_US <- shortlist_US$gvkey
sqlcmd <- paste("SELECT gvkey, conm, datadate, fyear, datafmt, indfmt, fic, curcd, act, lct, sale, sich, xrd
                  FROM compd.funda
                  WHERE fyear BETWEEN '2008'
                  AND '2015'
                  AND datafmt = 'STD'
                  AND gvkey = '", id_US,"'", sep="")
# LIST OF DATAFRAMES
funda_list_US <- lapply(sqlcmd, function(x) dbGetQuery(wrds, x))
# FINAL DATAFRAME
funda_US <- do.call(rbind, funda_list_US)
# query 'Compustat Global - Fundamentals Annual' using gvkey as identifier
# observation period 2010-2015 / include 2008 & 2009 for absorptive capacity measure
shortlist_global <- shortlist_raw %>% filter(country != "US")
id_global <- shortlist_global$gvkey
sqlcmd <- paste("SELECT gvkey, conm, datadate, fyear, datafmt, indfmt, fic, curcd, act, lct, sale, sich, xrd
                  FROM compgd.g_funda
                  WHERE fyear BETWEEN '2008'
                  AND '2015'
                  AND gvkey = '", id_global,"'", sep="")
# LIST OF DATAFRAMES
funda_list_global <- lapply(sqlcmd, function(x) dbGetQuery(wrds, x))
# FINAL DATAFRAME
funda_global <- do.call(rbind, funda_list_global)
# join US and global Compustat outputs
shortlist <- bind_rows(funda_US, funda_global) %>% rename(parent_company = conm, country = fic) %>% arrange(parent_company)
# ENRICH SHORTLIST WITH SIC INFORMATION
# query 'Segment NAICS North America' for SIC code industry information on parent companies using gvkey as identifier
id_shortlist <- shortlist$gvkey
sqlcmd <- paste("SELECT gvkey, sics, datadate
                FROM comp.seg_naics
                WHERE datadate BETWEEN '2010-01-01'
                AND '2015-12-31'
                AND stype = 'BUSSEG'
                AND gvkey = '", id_shortlist,"'", sep="")
# LIST OF DATAFRAMES
seg_list <- lapply(sqlcmd, function(x) dbGetQuery(wrds, x))
# FINAL DATAFRAME
seg_df <- do.call(rbind, seg_list)
# disconnect wrds connection
dbDisconnect(wrds)
# RESHAPING SIC OUTPUT DATA
# re-classification of sic codes to 3-digit level
seg_df <- seg_df %>% group_by(gvkey, datadate) %>%
  unique() %>%
  mutate(fyear = as.numeric(substr(datadate, 1, 4))) %>%
  mutate(sic = substr(sics, start = 1, stop = 3))
# summarize class characteristics for shortlist -> warning: no differentation made between primary and secondary sic codes
shortlist_sum <- seg_df %>%
  group_by(sic) %>%
  summarise(parent_companies = n()) %>%
  mutate(rank = dense_rank(desc(parent_companies))) %>%
  arrange(rank) %>%
  mutate(cum_count = cumsum(parent_companies))
# change format of seg_df from long to wide
# change order of columns
seg_df <- seg_df[,c(1,3,4,5)]
# group by gvkey und date
seg_df_grouped <- seg_df %>%
  dplyr::group_by(gvkey) %>%
 plyr::count(c("gvkey", "datadate"))
```

```
# how many SICs do the companies each have?
qplot(seg_df_grouped$freq, geom="histogram", binwidth=1)
## check
# order by n descending
arrange(seg_df_grouped,desc(freq))
# max number of SICs
max(seg_df_grouped$freq)
# reshape
seg_df <- ddply(seg_df, c("gvkey","datadate"), transform,
        sic_number = paste0("sic", seq(length(sic))))
detach(package:plyr, unload=TRUE)
distinct(seg_df,sic_number)
# reshape
seg_wide <- seg_df %>% spread(sic_number, sic)
# JOIN INFO FROM SIC CODES WITH FUNDAMENTALS TO ARRIVE AT FINAL SHORTLIST
shortlist <- shortlist %>% left_join(seg_wide, by = c("gvkey", "fyear", "datadate")) %>% unique()
# tidy df
# assumption: if no primary/secondary sic data is available from Segment database, then historical sic from Fundamentals applies (only primary one)
shortlist$sic1 <- ifelse(is.na(shortlist$sic1), str_sub(shortlist$sich, start = 1, end = 3), shortlist$sic1)
# save outout
save(shortlist, file = "../R_data/shortlist.Rda")
---
title: "Finalise sample"
---
load(file = "../R_data/shortlist.Rda")
CVC_longlist <- as_tibble(read_excel("../Data_tables/CVC_longlist.xlsx"))
# join with CVC longlist again to join CVC fund names again & tidy
CVC_longlist$cvc_fund <- toupper(CVC_longlist$cvc_fund)
CVC_longlist <- CVC_longlist %>% select(cvc_fund, parent_company)
shortlist <- shortlist %>% left_join(CVC_longlist,  by = "parent_company") %>% select(fyear, parent_company, cvc_fund, everything(), -datadate, -datafmt)
# exclude financial corporate investors identified by Computstat variable indfmt = FS
#unique FS companies in sample
fs <- shortlist %>% filter(indfmt == "FS") %>% distinct(parent_company)
#unique industrial corporate investors identified by Computstat variable indfmt = FS
indl <- shortlist %>% filter(indfmt == "INDL") %>% distinct(parent_company)
#exclude & tidy
shortlist <- shortlist %>% anti_join(fs, by = "parent_company") %>% select(-indfmt)
# compute R&D intensity & absorptive capacity (past 3-year sum of R&D intensity)
width <- 3
shortlist <- shortlist %>% mutate(rxd_intensity = xrd/sale, absorptive_capacity = rollapply(rxd_intensity, width, sum, align = "right", partial=T)) %>% select(-
starts_with("sic"), everything())
# full data coverage of all variables must be given in at least 2 years of observation period, otherwise exclude from sample
# fyear 2009, 2010 exluded automatically because all absorptive_capacity = 0
sample <- shortlist %>% group_by(parent_company) %>% filter(sum(!is.na(absorptive_capacity))>=2 & sum(!is.na(!act))>=2 & sum(!is.na(lct))>=2) %>% filter(fyear >=
2010 & fyear <= 2015)
n_distinct(sample$parent_company)
n_distinct(sample$country)
# convert all currencies to USD (conversion factor based on historical mean change rates)
currencies <- read_excel(path = "../Data_tables/currency_conversion.xlsx", col_types = c("numeric", "text", "numeric"))
sample <- left_join(sample, currencies, by = c("fyear", "curcd"))
sample$current_assets = sample$act * sample$conversionFactor
sample$current_liabilies = sample$lct * sample$conversionFactor
sample$sales = sample$sale * sample$conversionFactor
sample$rxd = sample$xrd * sample$conversionFactor
# tidy
sample <- sample %>% select(-act, -lct, -sale, -xrd, -rxd_intensity, -curcd, -conversionFactor) %>% select(-starts_with("sic"), everything())
# join with patent & year of incorporation data
patents <- read.csv2(file = "../Data_tables/patents.csv", stringsAsFactors=F)
sample <- full_join(sample, patents, by = c("fyear", "parent_company")) %>% select(-starts_with("sic"), everything())
# tidy
sample$sic1 <- ifelse(is.na(sample$sic1), sample$sich_x, sample$sic1)
sample$cvc_fund <- ifelse(is.na(sample$cvc_fund), sample$cvc_x, sample$cvc_fund)
sample <- sample %>% select(-sich, -sich_x, -cvc_x)
# save sample to compute diversity variable
save(sample, file = "../R_data/sample.Rda")
NEXT
# 1st: RUN PORTFOLIO SCRIPT
# 2nd: RUN DIVERSITY SCRIPT
# 3rd: RUN SCRIPTS TO COMPUTE OTHER VARIABLES
---
title: "Portfolios"
---
# load sample & function
load(file = "../R_data/sample.Rda")
# import all cvc deals data
file.list <- list.files(path="../CVC_raw_data",
                  pattern='*.csv', full.names = TRUE)
df.list <- lapply(file.list, read_csv2)
cvc_data_raw <- rbind.fill(df.list) %>% dplyr::select(-1,-2, - 8) %>%
  dplyr::rename(cvc_fund = 'Acquiror name', target_name = "Target name", target_country = "Target country code")
detach(package:plyr, unload=TRUE)
save(cvc_data_raw, file = "../R_data/cvc_data_raw.Rda")
uniquecvcfunds <- cvc_data_raw %>% distinct(cvc_fund)
# create fyear & separate sic codes into separate columns
# dplyr::filter deals data for observation period (2008-2016)
cvc_data <- cvc_data_raw %>%
  mutate(fyear = str_sub(`Completed date`, start = -4)) %>%
  select(-"Completed date") %>%
  select(fyear, everything()) %>%
  dplyr::filter(fyear >= 2007 & fyear <= 2015) %>%
  separate(`Target US SIC code(s)`, c("sic1", "sic2", "sic3", "sic4", "sic5", "sic6", "sic7", "sic8", "sic9", "sic10", "sic11", "sic12", "sic13", "sic14", "sic15", "sic16"), sep = " /
", remove = FALSE, extra = "warn")  %>%
  select(- "Target US SIC code(s)")
```

```
# convert fyear to numeric
cvc_data$fyear <- as.numeric(cvc_data$fyear)
# re-classification of sic codes to 3-digit level
cvc_data$sic1 <- str_sub(cvc_data$sic1, 1, end = -2)
cvc_data$sic2 <- str_sub(cvc_data$sic2, 1, end = -2)
cvc_data$sic3 <- str_sub(cvc_data$sic3, 1, end = -2)
cvc_data$sic4 <- str_sub(cvc_data$sic4, 1, end = -2)
cvc_data$sic5 <- str_sub(cvc_data$sic5, 1, end = -2)
cvc_data$sic6 <- str_sub(cvc_data$sic6, 1, end = -2)
cvc_data$sic7 <- str_sub(cvc_data$sic7, 1, end = -2)
cvc_data$sic8 <- str_sub(cvc_data$sic8, 1, end = -2)
cvc_data$sic9 <- str_sub(cvc_data$sic9, 1, end = -2)
cvc_data$sic10 <- str_sub(cvc_data$sic10, 1, end = -2)
cvc_data$sic11 <- str_sub(cvc_data$sic11, 1, end = -2)
cvc_data$sic12 <- str_sub(cvc_data$sic12, 1, end = -2)
cvc_data$sic13 <- str_sub(cvc_data$sic13, 1, end = -2)
cvc_data$sic14 <- str_sub(cvc_data$sic14, 1, end = -2)
cvc_data$sic15 <- str_sub(cvc_data$sic15, 1, end = -2)
cvc_data$sic16 <- str_sub(cvc_data$sic16, 1, end = -2)
# CREATE PORTFOLIOS
# determine to which portfolio years cvc investment belongs
# assumption: each target remains in portfolio for 4y on average, if no further round of funding made
# if further funding, then assumption that target remains in portfolio from first invest. year onwards until nth invest. year +4
cvc_data$portfolio_year <- NA
cvc_data <- cvc_data %>% select(fyear, portfolio_year, everything())
deals_number <- dim(cvc_data)[1]
cvc_data_copy <- cvc_data
cvc_data <- cvc_data_copy[1,]
cvc_data[] <-  cvc_data[FALSE, ]
tracker <- 1
# run through each deal
for (i in 1:deals_number) {
  cvc_temp <- cvc_data_copy[i, ]
  cvc_temp$portfolio_year <- cvc_temp$fyear
  cvc_data[dim(cvc_data)[1]+1,] <- cvc_temp
  cvc_temp$portfolio_year <- cvc_temp$fyear+1
  cvc_data[dim(cvc_data)[1]+1,] <- cvc_temp
  cvc_temp$portfolio_year <- cvc_temp$fyear+2
  cvc_data[dim(cvc_data)[1]+1,] <- cvc_temp
  cvc_temp$portfolio_year <- cvc_temp$fyear+3
  cvc_data[dim(cvc_data)[1]+1,] <- cvc_temp
}
# exlucde duplicated rows which were created in case of multiple rounds of investment
cvc_data <- cvc_data %>%
  dplyr::filter(!is.na(cvc_fund)) %>%
  group_by_at(vars(cvc_fund, target_name)) %>%
  distinct(portfolio_year, .keep_all = TRUE)
# exlude portfolio years outside observation period
# group by portfolios in each year per investor
portfolios <- cvc_data %>% filter(portfolio_year >= 2007 & portfolio_year <= 2015) %>% group_by_at(vars(portfolio_year, cvc_fund, target_name))
pct <- function(x) {x/lag(x)}
# summary statistics for observation period
portfolios_copy <- portfolios %>%
  ungroup() %>%
  group_by_at(vars(cvc_fund, portfolio_year)) %>%
  mutate(n = row_number(target_name))
# portfolio size per investor per year
p_size <- portfolios_copy %>%
  dplyr::summarize(p_size = n())
# mean portfolio size per investor
avg_p_size <- p_size %>%
  summarize(avg = mean(p_size))
# mean portfolio size per year
avg_p_size2 <- p_size %>% group_by(portfolio_year) %>%
  summarize(avg = mean(p_size)) %>%
  mutate(perc_change = pct((avg)))
# mean portfolio size for whole sample
avg_p_size3 <- avg_p_size %>% ungroup() %>% summarize(avg = mean(avg))
# save output
save(portfolios, file = "../R_data/portfolios.Rda")
---
title: "Diversity measure"
---
# load portfolio data & functions
load(file = "../R_data/portfolios.Rda")
load(file = "../R_data/sample.Rda")
load(file = "../R_data/fun_distance.Rda")
load(file = "../R_data/fun_diversity.Rda")
fund_names <- unique(portfolios$cvc_fund)
fund_names <- unique(fund_names)
fund_years <- c(2010:2015)
result_matrix <- as.tibble(matrix(0,nrow = length(fund_names) *
                                  length(fund_years),ncol = 5))
result_matrix[,1] <- fund_names
result_matrix[,2] <- fund_years
names(result_matrix) <- c("cvc_fund","fyear","diversity","portfolio_size","distinct_sics")
counter <- 1
total_counter <- length(fund_names) * length(fund_years) +1
for (fund in fund_names) {
  for (year in fund_years) {
    # First row investor sics aufnehmen
    temp_parent_matrix <- sample[sample$fyear == year &
                                 sample$cvc_fund == fund, grepl("sic\\d", names(sample))]
```

```
    temp_portfolio_matrix <- portfolios[portfolios$portfolio_year == year &
                            portfolios$cvc_fund == fund, grepl("sic\\d", names(portfolios))]
    temp_matrix <- bind_rows(temp_parent_matrix, temp_portfolio_matrix)
    result_matrix[counter,1] <- fund
    result_matrix[counter,2] <- year
    result_matrix[counter,3] <- fun_diversity(temp_matrix)
    # substract fund itself
    result_matrix[counter,4] <- nrow(temp_matrix)-1
    unique_sics <- unique(as.vector(as.matrix(temp_matrix)))
    unique_sics <- unique_sics[!is.na(unique_sics)]
    result_matrix[counter,5] <- length(unique_sics)
    counter <- counter + 1
    fraq <- round(counter/total_counter,4) * 100
      print(paste0(fraq,"% (" ,fund," - ",year,")"))
  }
}
# manipulate diversity=0 because of log transformation --> work around: replace 0 with 2nd lowest value of distribution
result_matrix$diversity <- ifelse(result_matrix$diversity == 0, min(result_matrix$diversity[result_matrix$diversity != min(result_matrix$diversity)]),
result_matrix$diversity)
# mutate log & squared of diversity result
result_matrix <- result_matrix %>% mutate(diversity_ln = log(diversity), diversity_ln_sqrt = diversity_ln^2)
save(result_matrix, file = "../R_data/result_matrix.Rda")
# Visual analysis of data
ggplot(result_matrix, aes(diversity)) +
  geom_histogram(bins= 100)
ggplot(result_matrix, aes(portfolio_size)) +
  geom_histogram(binwidth = +1)
ggplot(result_matrix, aes(distinct_sics)) +
  geom_histogram(binwidth = 1)
ggplot(result_matrix, aes(x=diversity, y=distinct_sics, colour = as.factor(fyear))) +
  geom_point(size=2, shape=23)
ggplot(result_matrix, aes(x=diversity, y=portfolio_size, colour = as.factor(fyear))) +
  geom_point(size=2, shape=23) +
  geom_smooth(method=lm, se=FALSE)
pairs(~diversity+portfolio_size+distinct_sics,data=result_matrix,
   main="Scatterplot Matrix")
---
title: "Geo diversity"
---
# load sample & function
load(file = "../R_data/portfolios.Rda")
# filter for observation period
portfolios_geo <- portfolios %>% select(portfolio_year, cvc_fund, target_name, target_country) %>%
  filter(portfolio_year >= 2010)
# number of ventures in portfolio per year --> CV(t)
ventures_peryear <- portfolios_geo %>% group_by_at(vars(portfolio_year, cvc_fund)) %>%
  tally() %>%
  rename(ventures.peryear = n)
# number of ventures in portfolio per year / per country / per investor --> cv(j,t)
ventures_peryear_percountry <- portfolios_geo %>% group_by_at(vars(portfolio_year, cvc_fund, target_country)) %>%
  tally() %>%
  rename(ventures.peryear.percountry = n)
# number of distinct countries per year / per investor
countries_peryear <- ventures_peryear_percountry %>%  select(-ventures.peryear.percountry) %>% group_by_at(vars(portfolio_year, cvc_fund)) %>%
  summarize(countries_peryear = n())
# join results
ventures_peryear_percountry <- ventures_peryear_percountry %>%  left_join(countries_peryear, by = c("portfolio_year", "cvc_fund")) %>% left_join(ventures_peryear, by
= c("portfolio_year", "cvc_fund"))
# compute arg1
ventures_peryear_percountry$arg1 <- ((ventures_peryear_percountry$ventures.peryear.percountry / ventures_peryear_percountry$ventures.peryear)^2)
# compute HHI & geografic diversity as reverse HHI
geo_diversity <- ventures_peryear_percountry %>% group_by_at(vars(portfolio_year, cvc_fund)) %>% select(portfolio_year, cvc_fund, arg1) %>% summarise(HHI =
sum(arg1), geo_diversity = 1/HHI)
# tidy
geo_diversity <- geo_diversity %>% select(portfolio_year, cvc_fund, geo_diversity) %>% rename(fyear = portfolio_year)
# save output
save(geo_diversity, file = "../R_data/geo_diversity.Rda")
---
title: "Experience"
---
load("../R_data/cvc_data_raw.Rda")
# tidy & transform
# date as date column
cvc_data <- cvc_data_raw %>% rename(date = `Completed date`) %>% filter(!is.na(date)) %>% mutate(date= dmy(date)) %>% select(-target_country, - `Target US SIC
code(s)`)
# mutate new variable with date of oldest investment per investor
cvc_data <- cvc_data %>% group_by(cvc_fund) %>% mutate(first_invest = min(date))
# mutate new variable with duration in days between today! (first investment?) and each investment --> (This is the weight)
date_of_today <- today()
# compute variable with weighted cvc experience per investor in each year
# number of investments per date per investor
investments_perdate <- cvc_data %>% group_by_at(vars(date, cvc_fund)) %>%
  tally() %>% rename(investments.perdate = n) %>%
  mutate(duration = as.numeric(difftime(date_of_today, date) / 365.242), investmentsXduration = investments.perdate*duration) %>%
  group_by(cvc_fund)
investments_perdate <- investments_perdate %>% left_join(cvc_data[,-3], by = c("date", "cvc_fund")) %>% mutate(total_weight = as.numeric(difftime(date_of_today,
first_invest)  / 365.242), weighted = investmentsXduration/total_weight)
# experience = cumulative count of all weighted investments per year
cvc_experience <- investments_perdate %>% mutate(fyear = year(date)) %>% select(fyear, cvc_fund, weighted) %>% group_by_at(vars(fyear, cvc_fund)) %>%
summarize(sum = sum(weighted)) %>% group_by(cvc_fund) %>% mutate(cvc_experience = cumsum(sum))
# tidy
cvc_experience <- cvc_experience %>%  filter(fyear >= 2010 & fyear <= 2015) %>% select(fyear, cvc_fund, cvc_experience)
save(cvc_experience, file = "../R_data/cvc_experience.Rda")
```

```
# check distribution of variable before and after log transformation
hist(cvc_experience$cvc_experience)
hist(log(cvc_experience$cvc_experience))
---
title: "Acquisitions & Joint Ventures"
---
# import deals data
ma_jv_raw <- read_excel("../Data_tables/MA&JV.xlsx")
ma_jv <- ma_jv_raw
# transform & tidy
ma_jv$completed_date <- if_else(is.na(ma_jv$completed_date) == T, ma_jv$assumed_completion_date, ma_jv$completed_date)
ma_jv <- ma_jv %>% mutate(fyear = year(completed_date)) %>% select(-"Deal Number", -completed_date, -assumed_completion_date, -"Acquiror name")
head(ma_jv)
# CREATE PORTFOLIOS
# determine to which portfolio years M&A und Joint Venture activities belong
# assumption: each target can be considered an external source of knowledge for 4y on average
ma_jv$portfolio_year <- NA
ma_jv <- ma_jv %>% select(fyear, portfolio_year, everything())
deals_number <- dim(ma_jv)[1]
ma_jv_copy <- ma_jv
ma_jv <- ma_jv_copy[1,]
ma_jv[] <-  ma_jv[FALSE, ]
tracker <- 1
# run through each deal
for (i in 1:deals_number) {
  ma_jv_temp <- ma_jv_copy[i, ]
  ma_jv_temp$portfolio_year <- ma_jv_temp$fyear
  ma_jv[dim(ma_jv)[1]+1,] <- ma_jv_temp
  ma_jv_temp$portfolio_year <- ma_jv_temp$fyear+1
  ma_jv[dim(ma_jv)[1]+1,] <- ma_jv_temp
  ma_jv_temp$portfolio_year <- ma_jv_temp$fyear+2
  ma_jv[dim(ma_jv)[1]+1,] <- ma_jv_temp
  ma_jv_temp$portfolio_year <- ma_jv_temp$fyear+3
  ma_jv[dim(ma_jv)[1]+1,] <- ma_jv_temp
}
# tidy
ma_jv <- ma_jv %>% filter(!is.na(parent_company) & portfolio_year >= 2010 & portfolio_year <= 2015) %>% mutate(duration = portfolio_year - fyear)
# mutate weightning variable
# assumption: straight-line depreciation because of declining influence as knowledge source / influence in 1st year=100%, in 5th year=0%
ma_jv$weight <- NA
ma_jv$weight <- ifelse(ma_jv$duration==0, 1, ifelse(ma_jv$duration==1, 0.75, ifelse(ma_jv$duration==2, 0.5, ifelse(ma_jv$duration==3, 0.25, NA))))
# compute depreciated acquisitions count per investor / per year
ma <- ma_jv %>% filter(str_detect(deal_type, "Acquisition")) %>% group_by_at(vars("portfolio_year", "parent_company"))
acquisitions <- ma %>% select(portfolio_year, parent_company, weight) %>% summarize(acquisitions = sum(weight)) %>% rename(fyear=portfolio_year)
save(acquisitions, file = "../R_data/acquisitions.Rda")
# compute depreciated joint ventures count per investor / per year
jv <- ma_jv %>% filter(str_detect(deal_type, "Joint")) %>% group_by_at(vars("portfolio_year", "parent_company"))
jointventures <- jv %>% select(portfolio_year, parent_company, weight) %>%  summarize(jointventures = sum(weight)) %>% rename(fyear=portfolio_year)
save(jointventures, file = "../R_data/jointventures.Rda")
---
title: "Final sample"
---
options("scipen" = 10)
load(file = "../R_data/sample.Rda")
# unique industries represented in the sample
n_distinct(substr(sample$sic1, start=1, stop=2))
# join diversity results
load(file = "../R_data/result_matrix.Rda")
sample_final <- sample %>% left_join(result_matrix) %>% dplyr::select(-starts_with("sic"), sic1)
# check
anti_join(sample, result_matrix, by = c("fyear", "cvc_fund")) %>% filter(fyear == 2014)
# join geo diversity results
load(file = "../R_data/geo_diversity.Rda")
sample_final <- sample_final %>% full_join(geo_diversity)
# tidy (because geo diversity=NA when portfolio size=0, in that case logically geo diversity should also be 0)
sample_final$geo_diversity <- ifelse(is.na(sample_final$geo_diversity), 0, sample_final$geo_diversity)
# join cvc experience
load(file = "../R_data/cvc_experience.Rda")
sample_final <- sample_final %>% full_join(cvc_experience)
#tidy
sample_final$cvc_experience <- ifelse(is.na(sample_final$cvc_experience), 0, sample_final$cvc_experience)
# manipulate experience=0 because of log --> work around: replace 0 by 2nd min value of distribution
sample_final$cvc_experience <- ifelse(sample_final$cvc_experience == 0, min(sample_final$cvc_experience[sample_final$cvc_experience !=
min(sample_final$cvc_experience)]), sample_final$cvc_experience)
# mutate log
sample_final <- sample_final %>% mutate(cvc_experience_ln = log(cvc_experience))
# join acquisitions & joint ventures
load(file = "../R_data/acquisitions.Rda")
load(file = "../R_data/jointventures.Rda")
sample_final <- sample_final %>% full_join(acquisitions) %>% full_join(jointventures)
# tidy
sample_final$acquisitions <- ifelse(is.na(sample_final$acquisitions), 0, sample_final$acquisitions)
sample_final$jointventures <- ifelse(is.na(sample_final$jointventures), 0, sample_final$jointventures)
# compute missing & lag variables
sample_final <- sample_final %>% arrange(fyear, .by_group=T) %>%
                              mutate(sales_ln = log(sales),
                              current_ratio = current_assets/current_liabilies,
                              age = fyear - first_year,
                              #time lag -1
                              diversity_L1 = lag(diversity, n = 1, default = NA),
                              diversity_ln_L1 = lag(diversity_ln, n = 1, default = NA),
                              diversity_ln_sqrt_L1 = lag(diversity_ln_sqrt, n = 1, default = NA),
                              absorptive_capacity_L1 = lag(absorptive_capacity, n = 1, default = NA),
```

```
                         geo_diversity_L1 = lag(geo_diversity, n = 1, default = NA),
                         sales_L1 = lag(sales, n = 1, default = NA),
                         sales_ln_L1 = lag(sales_ln, n = 1, default = NA),
                         current_ratio_L1 = lag(current_ratio, n = 1, default = NA),
                         age_L1 = lag(age, n = 1, default = NA),
                         acquisitions_L1 = lag(acquisitions, n = 1, default = NA),
                         jointventures_L1 = lag(jointventures, n = 1, default = NA),
                         portfolio_size_L1 = lag(portfolio_size, n = 1, default = NA),
                         cvc_experience_L1 = lag(cvc_experience, n = 1, default = NA),
                         cvc_experience_ln_L1 = lag(cvc_experience_ln, n = 1, default = NA))
# tidy & create subsamples & write STATA output
sample_final$sic1 <- as.numeric(sample_final$sic1)
library(foreign)
sample_stata <- sample_final %>% dplyr::select(fyear, parent_company, patent_count, diversity_L1, diversity_ln_L1, diversity_ln_sqrt_L1, absorptive_capacity_L1,
geo_diversity_L1, sales_L1, sales_ln_L1, current_ratio_L1, age_L1, acquisitions_L1, jointventures_L1, portfolio_size_L1, cvc_experience_L1, cvc_experience_ln_L1, sic1,
country)
save(sample_stata, file = "../R_data/sample_stata.Rda")
write.dta(sample_stata, file = "../Data_tables/STATA/sample_stata.dta")
hightech <- c(283, 357, 361, 365, 366, 367, 372, 376, 381, 382, 384, 737, 873)
subsample_hightech <- sample_stata %>% filter(sic1 %in% hightech)
write.dta(subsample_hightech, file = "../Data_tables/STATA/subsample_hightech.dta")
subsample_US <- sample_stata %>% filter(country == "USA")
write.dta(subsample_US, file = "../Data_tables/STATA/subsample_US.dta")
subsample_uncensored <- sample_stata %>% filter(portfolio_size_L1 > 0)
write.dta(subsample_uncensored, file = "../Data_tables/STATA/subsample_uncensored.dta")
```

SELECTED STATA SYNTAX

```
clear
* import packages
ssc install outreg2
ssc install estout
set more off
use "sample_stata.dta", clear
############# DESCRIPTIVE & CORRELATION ############
summarize patent_count ///
      diversity_L1 diversity_ln_L1 diversity_ln_sqrt_L1 absorptive_capacity_L1 geo_diversity_L1 sales_L1 sales_ln_L1  ///
      current_ratio_L1 age_L1 acquisitions_L1 jointventures_L1 ///
      portfolio_size_L1  cvc_experience_L1 cvc_experience_ln_L1, detail
outreg2 using DESCRIPTIVES.doc, replace sum(detail) ///
      dec(2) keep(patent_count  ///
      diversity_L1 diversity_ln_L1 diversity_ln_sqrt_L1 absorptive_capacity_L1 geo_diversity_L1 sales_L1 sales_ln_L1  ///
      current_ratio_L1 age_L1 acquisitions_L1 jointventures_L1 ///
      portfolio_size_L1  cvc_experience_L1 cvc_experience_ln_L1)
*correlations
estpost correlate patent_count  ///
      diversity_L1 diversity_ln_L1 diversity_ln_sqrt_L1 absorptive_capacity_L1 geo_diversity_L1 sales_L1 sales_ln_L1  ///
      current_ratio_L1 age_L1 acquisitions_L1 jointventures_L1 ///
      portfolio_size_L1  cvc_experience_L1 cvc_experience_ln_L1, matrix listwise
est store c1
esttab * using CORRELATION.rtf, unstack not noobs compress
############# MODELS ############
*compute identifier for panel analysis
egen sample = group(parent_company)
*set cross-section & time sequence ID for panel analysis
xtset sample fyear, yearly
*Model 1 baseline
xtnbreg patent_count sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1  ///
      acquisitions_L1 jointventures_L1 portfolio_size_L1 i.fyear,fe irr
estat ic
mat es_ic = r(S)
local AIC: display %4.1f es_ic[1,5]
local BIC: display %4.1f es_ic[1,6]
estimate store re
hausman fe re
outreg2 using NBREG-FE.doc,  ///
dec(3) replace eform ctitle(Model 1) keep(sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1) ///
adds(Degrees of freedom, e(df_m), Log likelihood, e(ll), Wald chi2, e(chi2), AIC, `AIC', BIC, `BIC')  ///
addtext(Firm dummies, Fixed, Time dummies, Yes) title("Negative binominal panel regression with fixed effects Dependent variable = Investor innovation performance")
*Model 2 diversity only
xtnbreg patent_count diversity_ln_L1 sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1 i.fyear,fe irr
estat ic
mat es_ic = r(S)
local AIC: display %4.1f es_ic[1,5]
local BIC: display %4.1f es_ic[1,6]
outreg2 using NBREG-FE.doc,  ///
dec(3) eform addstat(Degrees of freedom, e(df_m), Log likelihood, e(ll), Wald chi2, e(chi2), AIC, `AIC', BIC, `BIC')  ///
append ctitle(Model 2)  ///
keep(diversity_ln_L1 sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1) ///
addtext(Firm dummies, Fixed, Time dummies, Yes)
*Model 3 diversity + diversity sqrt
xtnbreg patent_count diversity_ln_L1 diversity_ln_sqrt_L1   ///
sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1 i.fyear,fe  irr
estat ic
mat es_ic = r(S)
local AIC: display %4.1f es_ic[1,5]
local BIC: display %4.1f es_ic[1,6]
outreg2 using NBREG-FE.doc,  ///
dec(3) eform addstat(Degrees of freedom, e(df_m), Log likelihood, e(ll), Wald chi2, e(chi2), AIC, `AIC', BIC, `BIC')  ///
```

```
append ctitle(Model 3)   ///
keep(diversity_ln_L1 diversity_ln_sqrt_L1   ///
sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1) ///
addtext(Firm dummies, Fixed, Time dummies, Yes)
*Model 4 moderation variables without interaction
xtnbreg patent_count diversity_ln_L1 absorptive_capacity_L1 geo_diversity_L1 sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1
jointventures_L1 portfolio_size_L1 i.fyear,fe  irr
estat ic
mat es_ic = r(S)
local AIC: display %4.1f es_ic[1,5]
local BIC: display %4.1f es_ic[1,6]
outreg2 using NBREG-FE.doc,  ///
dec(3) eform addstat(Degrees of freedom, e(df_m), Log likelihood, e(ll), Wald chi2, e(chi2), AIC, `AIC', BIC, `BIC')  ///
append ctitle(Model 4)   ///
keep(diversity_ln_L1 absorptive_capacity_L1 geo_diversity_L1 sales_ln_L1 current_ratio_L1 age_L1 cvc_experience_ln_L1 acquisitions_L1 jointventures_L1
portfolio_size_L1) ///
addtext(Firm dummies, Fixed, Time dummies, Yes)
*Model 5 moderation variables with interaction
*FULL MODEL
xtnbreg patent_count diversity_ln_L1 absorptive_capacity_L1 geo_diversity_L1 c.absorptive_capacity_L1#c.diversity_ln_L1  c.geo_diversity_L1#c.diversity_ln_L1
sales_ln_L1 current_ratio_L1 age_L1  cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1  i.fyear,fe  irr
estat ic
mat es_ic = r(S)
local AIC: display %4.1f es_ic[1,5]
local BIC: display %4.1f es_ic[1,6]
outreg2 using NBREG-FE.doc,  ///
dec(3) eform addstat(Degrees of freedom, e(df_m), Log likelihood, e(ll), Wald chi2, e(chi2), AIC, `AIC', BIC, `BIC')  ///
append ctitle(Model 5) ///
keep(diversity_ln_L1  absorptive_capacity_L1 geo_diversity_L1 c.absorptive_capacity_L1#c.diversity_ln_L1  c.geo_diversity_L1#c.diversity_ln_L1 sales_ln_L1
current_ratio_L1 age_L1  cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1) ///
addtext(Firm dummies, Fixed, Time dummies, Yes)




############ MARGINS ############
xtnbreg patent_count diversity_ln_L1 absorptive_capacity_L1 geo_diversity_L1 c.absorptive_capacity_L1#c.diversity_ln_L1  c.geo_diversity_L1#c.diversity_ln_L1
sales_ln_L1 current_ratio_L1 age_L1  cvc_experience_ln_L1 acquisitions_L1 jointventures_L1 portfolio_size_L1  i.fyear,fe
margins, at(diversity_ln_L1=(-8(2)2)) predict(iru0) atmeans
marginsplot, scheme(s1mono)
* absorptive_capacity
margins, at(diversity_ln_L1=(-8 -6 -4 -2 0 2) absorptive_capacity_L1=(0.01 0.2 0.59 )) predict(nu0) atmeans
marginsplot, xdimension(at(diversity_ln_L1)) scheme(s1mono) legend(cols(1)) ytitle(Predicted Number of Events) scale(1.2)
* geo_diversity
margins, at(diversity_ln_L1=(-8 -6 -4 -2 0 2) geo_diversity_L1=(0 1.38 2.78 )) predict(nu0) atmeans
marginsplot, xdimension(at(diversity_ln_L1)) scheme(s1mono) legend(cols(1)) ytitle(Predicted Number of Events) scale(1.2)
```

Official Statement of Originality of Maastricht University

By signing this statement, I hereby acknowledge the submitted master thesis titled “The effect of corporate venture capital portfolio diversification on firm innovation: A replication and extension study” to be produced independently by me, without external help.

Wherever I paraphrase or cite literally, a reference to the original source (journal, book, report, internet, etc.) is given.

By signing this statement, I explicitly declare that I am aware of the fraud sanctions as stated in the Education and Examination Regulations (EER) of the SBE.

Place: Maastricht

Date: 04.01.2019

First and Last Name: Sabrina Steeb

Study Program: International Business – Strategy & Innovation

Course/Skill: Master Thesis

I-Number: i6172598

Signature: S. Steeb