Stockholm School of Economics

MSc in Finance

Combined thesis title:

"And the similar shall lead: the role of the affinity bias in the Italian Venture Capital investment decisions"

Bocconi thesis title:

"And the similar shall lead: the role of the affinity bias in the Italian Asset Management Companies' investment decisions"

SSE thesis extension title:

"And the similar shall lead: the role of the affinity bias in the Italian Corporate Venture Capital investment decisions"

Autor: Marco Parente

Student number: 42014

a.a.: 2021-2022

Contents

Glossary

Word	Definition
Asset Management Company (AMC)	Each of the 19 professional firms managing Venture Capital funds mapped in the analysis.
Capitalization Table (Cap Table)	A chart used to show how ownership is distributed among the company's shareholders.
Corporate Venture Capital (CVC)	A corporate division fully dedicated to Venture Capital investments. This paper includes the 9 most active among those headquartered in Italy.
Failure rate	The proportion of write-offs on the total number of initial investments made by an AMC or a CVC.
Followers	In the context of syndicate financing, the followers are all investors different from the lead.
Follow-on investment	Any subsequent investment made by an AMC or a CVC after the initial investment in a portfolio company.
Initial investment	First investment made by an AMC or a CVC in a start-up.
Investment status	The current condition of an investment made by an AMC or a CVC, can be: - *active*, if the VC fund still owns the shareholding; - *transferred to new fund*, if the AMC has moved its shareholding from one of its funds to another (this applies only to AMCs, as it was not observed for CVCs); - *write-off*, in case the invested start-up has gone bankrupt; - *exited*, if the fund has already sold its shareholding.
Lead Investor	In the context of syndicate financing, the lead investor conducts the due diligence on the start-up and is responsible for direct negotiation of the round's terms and conditions.
Liquidity event	Transaction allowing the VC fund to exit from the investment. For the purpose of this paper, three liquidity events were considered: - *M&A*, if the investee was acquired by a financial or strategic buyer, or merged with another company; - *IPO*, in case the investee was listed on a public Stock Exchange; - *Secondary Purchase*, in case the AMC or the CVC sold its shares on the secondary market to either existing or new shareholders.

Post-money valuation	Value of a company after receiving an equity injection by a VC fund. It is computed by adding the equity injection to the company's pre-money valuation.
Pre-money valuation	Value of a company before receiving an equity injection by a VC fund.
Sector	Broad aggregation of business verticals sharing similar traits. This paper defines a total of 9 sectors based on the start-ups' business verticals.
Syndicate	Group of investors participating in a deal. Normally, a syndicated is made up of a lead investor and a series of followers.
Valuation step-up	A measure of a shareholding's appreciation. It is computed as the ratio between the latest available pre-money valuation of the company and its post-money valuation on the occasion of the VC fund's initial investment.
VC Fund	The vehicle through which an AMC or a CVC makes professional Venture Capital investments.
Vertical	A business vertical describes a group of companies that focus on a shared niche or specialized market spanning multiple industries. The business vertical of each start-up was retrieved from PitchBook.

1. Introduction

A substantial research literature has established that individual and collective decisions often diverge from the paradigm of rationality. This observation applies to a variety of contexts: for instance, people assess probabilities incorrectly (Kahneman & Tversky, 1973), they violate the axioms of utility theory (Kahneman & Tversky, 1978) and they interpret information in a way that confirms their prior beliefs or values (Klayman & Ha, 1987; Nickerson, 1998).

Some of the most impactful departures from normative decision-making are caused by the affinity bias, i.e. the unconscious tendency to gravitate towards people to whom we feel to be close for interests, background, ethnicity and other personal traits. The affinity bias induces preference for what is similar and distrust for what is diverse, which can ultimately lead to suboptimal outcomes in many areas, including hirings (Ross, 2008; Gompers & Wank, 2017), access to credit (Hunter & Walker, 1995) and even quality of medical treatments (Marcelin *et al.*, 2019).

This paper analyses the role of the affinity bias in the Italian Venture Capital (VC) ecosystem. The background idea is that VC funds' partners may be subject to psychological biases when deciding the start-ups to invest in. Specifically, they could unintentionally prioritise founders who share cultural and genetic features with them. Clearly, this approach is not guaranteed to produce the best investment decisions, as it is not coherent with rational economic theory and utility maximization.

To check for the affinity bias influence on VC partners, two different samples were collected: the former (so-called "Sample 1") includes 593 initial investments made by 19 Italian Asset Management Companies (AMC) from January 2000 to December 2021, while the latter (so-called "Sample 2") is made up of 95 investments closed by 7 Italian Corporate Venture Capital arms (CVC) over the same period.

For both Sample 1 and Sample 2, all the analyses (and relative conclusions) are referred to the period spanning from January 2000 to December 2021.

For each transaction, a percentage similarity score was built to capture the degree of proximity between the start-up's founding team and the partners of the fund that participated in the deal. The similarity score, computed as a weighted average of seven variables, is presented in three different specifications, which depend on the system of weights applied.

The results obtained are quite interesting. As far as Sample 1 is concerned, each specification of the similarity score presents realistic distributional features and intuitive links with selected sample variables. In particular, the affinity bias seems to affect all professional investors covered in the analysis, with average similarity scores well above 50%. Notably, the differences at AMC level can be at least partially linked to return performance, which suggests that those suffering the most from irrationality-induced decisions reach poorer financial results. Furthermore, the impact of the affinity bias, as measured by the score, is negatively correlated to partners' experience and gender diversity.

As for Sample 2, the limited number of observations prevents from making strong conclusions as those in Sample 1. Nevertheless, the similarity score stays, on average, above 0.50 under all specifications and preserves realistic distributional features at least under two out of three specifications. Additionally, under one specification it appears correlated to partners' experience, while under two it may be (partially) linked to CVCs' performance.

The remainder of the paper is organized as follows. Chapter 2 offers a brief overview of the existing literature contributions on the affinity bias and its implications for rational decision-making. Chapter 3 describes the data gathering process. Chapter 4 provides a general description of sample features. Chapter 5 describes the construction of the similarity score. Chapter 6 analyses the distribution of the similarity score across its three specifications, with a focus on how results change when segmenting data according to several criteria (investing entity, time, financing round, investment status, round structure and sector). Chapter 7 links the similarity score to certain VC partners' features (average age at deal date, gender diversity and number of investments closed) and VC funds (size and overall performance). Chapter 8 provides conclusions.

2. Literature review

Unconscious biases are an unavoidable component of human life: according to Wilson (2002), we are faced with approximately 11 million bits of information at any given moment, while our brain is able to process only 40 at a time. This makes it impossible to always analyse the reality through a rational paradigm, creating the need to use non-fully rational shortcuts.

In this regard, Stanovich & West (2000) make a useful distinction between System 1 and System 2 cognitive functioning: the former resorts to intuition and is typically fast, automatic, effortless and emotional; the latter uses rationality and, as a consequence, is slower, conscious and effortful. As a matter of fact, most decisions in life are made using System 1 thinking, and while this can be helpful in many cases[1], it can lead to serious mistakes in others[2].

In fact, cognitive biases are much more likely to happen under System 1 thinking than under System 2. One of the most discussed in literature is the affinity bias, i.e. the tendency to prefer people, things and situations with which we feel a certain degree of familiarity. The affinity bias influences many of the most important decisions we make and has a profound effect on others' lives.

For example, Hunter & Walker (1995) notice that, *ceteris paribus*, US white loan agents penalize the access to credit of minorities, and they argue that this discrimination could result from the lack of cultural affinity between the two ethnic groups. Indeed, since loan agents feel to know little about minorities, they prefer to rely more on objective loan application information in appraising their creditworthiness. As a consequence, hard metrics (e.g. credit history and the ratio of total monthly obligations-to-total monthly income) have a substantially greater impact on the probability to receive a loan for minorities than for whites.

[1] For instance, it would be impractical (and potentially confusing) to rationally ponder every choice we make when shopping groceries.

[2] For example, people usually lose much money when gambling because they badly assess probabilities, or they overstate their level of control on the events.

Also, affinity bias can induce significant distortions in hiring practices (Louis, 2019). When evaluating candidates, recruiters tend to favour those who are more similar to them. This can create a vicious cycle whereby the newly selected members of an organization will, in turn, choose people who are affine to them, and so on. It is easy to see, then, that the affinity bias can lead to suboptimal hirings and harm diversity, especially in small firms.

The impact of the affinity bias has been evaluated also in relation to the healthcare industry. For example, Marcelin *et al.* (2019) study the US medical system and find that minority groups suffer from cognitive-bias-induced discriminations when seeking treatments. This happens because the increasing diversity in the US population is reflected in patients, but it is often missing in healthcare professionals. Therefore, under-represented categories risk experiencing health inequities caused by cultural stereotypes.

Finally, Chhaochharia & Laeven (2009) collect data on approximately 30,000 equity investments by sovereign wealth funds and find that they concentrate most of their allocations in countries displaying common cultural traits. This suggests that sovereign wealth funds prefer to "invest in the familiar", which may depend on the exploitation of informational advantages, but also on the influence of irrational affinity considerations.

Within this framework, a growing attention has been given in the last years to the role of affinity bias in the dynamics of the VC industry. Gompers *et al.* (2016) investigate how personal traits affect VC partners' desire to collaborate and whether this attraction influences VC funds' performance. Specifically, they consider four characteristics: two (educational and professional background) are related to abilities and, as such, shall have a key role in venture capitalists' success; the other two (ethnicity and gender) are affinity-related features which do not depend on ability and, thus, shall not influence investment performance. Interestingly, the authors find that ethnicity and gender have a non-negligible impact on VC partners' desire to collaborate with other venture capitalists through syndicated investments. Notably, the authors show that this behaviour dramatically reduces returns: for instance, if two partners belonging to the same ethnic minority group invest together, performance[3] can drop by as much as 20%.

Another relevant contribution comes from Gompers & Wang (2017), to whom I partly owe the inspiration for the title of this paper. The authors analyse the impact of the affinity bias on new VC partners' hirings from an innovative perspective: specifically, they gather data on the gender of VC partners' children and find that, when existing partners have more daughters, they are more likely to hire a female investor partner, naturally increasing diversity within the organization. As a further step, they assess the consequences for the fund returns and show that greater gender diversity increases performance by a meaningful amount: on average, if existing partners have a daughter rather than a son, deal success[4] rises by almost 3% and net excess IRR[5] increases by 3.20%.

As it can be seen, the studies cited mainly focus on the internal dynamics of VC funds. Conversely, the role of the affinity bias in the interplay between VC partners and invested start-ups is still a relatively unexplored area. This paper tries to fill the gap by quantifying the degree of cultural, ethnic, educational and professional similarity between teams of founders and partners involved in VC transactions. Having confirmed the strong presence of the affinity bias in both Sample 1 and Sample 2 through the computation of a similarity score, the analysis assesses whether this latter can be related to specific (C)VC funds and partners' characteristics.

---

[3] Performance is measured by the probability of realizing a successful exit through IPO.

[4] Deal success is defined as a dummy variable taking value of 1 in case the VC fund realized an exit through IPO or M&A with acquisition value higher than the invested capital.

[5] Net excess IRR is defined as the difference between a fund's net IRR and the median fund return in the same region and year.

3. Data collection

As anticipated in the Introduction, Sample 1 is made up of 593 VC initial investments by 19 Italian AMCs from January 2000 to December 2021, while Sample 2 was built by aggregating 95 initial investments of 7 Italian CVCs over the same period. It is important to underline that neither Sample 1 nor Sample 2 coincide with the population of deals closed by the AMCs and CVCs over the period under analysis for two main reasons. Firstly, follow-on transactions (as defined in the Glossary) were not considered: this happens because the two samples of investments were tuned for the similarity score computation, which must depend only on initial investments, as considering subsequent financings would have implied double-counting. Secondly, information in the VC industry is traditionally opaque: *inter alia*, this implies that several transactions remain undisclosed and cannot be mapped (intuitively, this is true especially for Sample 1, since AMCs make much more investments than CVCs).

The 19 AMCs and the 7 CVCs considered are all headquartered in Italy and represent the most active investors among the legal entities in the Italian VC ecosystem. In detail, AMCs have a considerably larger deal flow than CVCs for two main reasons: firstly, a CVC is a non-core division, so that only a part of the corporate's budget, time and personnel are devoted to its functioning; secondly, a CVC is naturally limited in its activity, as it targets only those start-ups which are synergic to the corporate business. When an AMC operates both VC and Private Equity investments (as it happens, for instance, for Vertis), only the VC funds it manages were considered. An exhaustive list of the AMCs (and relating funds) covered in this paper is provided in Table I of Annex 1. The same is done in Table II of Annex 1 for the 7 CVCs mapped.

For each investment, four main categories of information were collected: i) company-specific data (e.g. date of incorporation and headquarters location); ii) deal-specific data (e.g. deal date and deal size); iii) data on investee's founders (e.g. gender and age); iv) data on the AMC or CVC's partners who participated in the deal (same information as the investee's founders). A more detailed explanation is furnished below.

3.1 Company-specific data

3.1.1 Date of incorporation

In almost all cases, the company's date of incorporation was obtained from Orbis. When not available, the foundation year was taken from PitchBook and the date of incorporation was assumed to coincide with the 1st of January. For instance, if a company's foundation year were 2010, then the date of incorporation would be set to January 1, 2010.

3.1.2 Headquarters city and country

To find geographic information on each investee, the proprietary websites and Orbis were used as primary sources. If data were not found in this way, then PitchBook and Crunchbase were checked, with priority given to the former because of its greater reliability.

3.1.3 Primary Business Vertical and Sector

Each company was assigned a sector based on its main business vertical, as provided by PitchBook. When information on primary vertical was not found, sector attribution was done by looking both at the company's business description (as provided by PitchBook) and at its website – when available.

Specifically, the following 9 sectors were defined:

- Digital;
- Education & HR;
- FinTech;
- Food & Agriculture;
- Healthcare & Biotech;
- Media;
- SaaS & Software;
- Smart City;
- Tech.

An exhaustive list of the verticals covered by each sector is provided in Table III of Annex 1.

3.2 Deal-specific data

3.2.1 Deal date

Deal date was retrieved from either PitchBook or the investment's press release. In case only the investment year was found, the deal date was assumed to coincide with the 1st of January, applying the same rationale followed for the start-ups' date of incorporation.

3.2.2 Transaction type and stage

The deals analysed were segmented by type (i.e. by round series) and by stage (acceleration, early stage VC and later stage VC).

Information on round series was obtained from either PitchBook or press releases. When not available, the series was assigned case by case by looking at the start-up's funding history: the first round was always regarded as "Seed", the second, if bigger, "Series A", otherwise "Seed" and so on. When it was not possible to unambiguously assign the series because of the lack of precise information on the company's equity story, the round type was labelled "Undisclosed".

Round stage attribution directly descends from round series (see Table IV of Annex 1 for more details).

3.2.3 Data on round structure

Each transaction was ranked based on the investors' number and geography.

As for the investors' number, the AMC (or the CVC) was considered as:

- *sole investor,* if it was the only investor financing the round;
- *syndicate member*, if it collaborated with other investors.

As for the investors' geography, deals with at least one foreign investor were distinguished from those with only national players.

Finally, specific attention was given to the AMC or the CVC's role in the transaction, separating the deals in which it acted as lead investor from those in which it was a follower.

All data were collected from PitchBook.

3.2.4 Round size and pre-money valuation

Information on round size and pre-money valuation was obtained from either PitchBook or press releases. While in the vast majority of cases it was possible to find the round size, pre-money valuation was disclosed for fewer deals.

As Chapter 7 will show, round size and pre-money valuation enter the computation of the company's valuation step-up, which can be used as an approximate measure of the investment performance.

3.2.5 Investment Status

The investment status was labelled as:

- *active,* if the VC fund is still on the company's capitalization table (Cap Table);
- *transferred to new fund,* if the AMC has moved its shareholding from one of its funds to another (this applies only to AMCs, as it was not observed for CVCs)[6];
- *write-off,* in case the company went bankrupt;
- *exited,* if the fund sold its shareholding on the occasion of a liquidity event. For the purpose of this paper, three liquidity events were taken into account, which leads to three potential exit clusters:
    - *M&A*, if the investee was acquired by a financial or strategic buyer, or merged with another company;
    - *IPO,* in case the investee was listed on a public Stock Exchange;
    - *Secondary Purchase,* in case the AMC (or the CVC) sold its shares on the secondary market to either existing or new shareholders.

[6] This scenario materializes if the fund that has originally invested in the company enters the divestment phase, but the AMC still wants to keep the shareholding.

Information on M&A and IPO activity was obtained from both Zephyr and PitchBook, while secondary market transactions were inferred by looking at changes of companies' Cap Tables on Orbis.

3.3 Data on start-ups' founders

The founders of each start-up were identified by looking at the company's profile on PitchBook and, when available, at its website and LinkedIn page.

In case either the founders were unidentifiable, or they had already left the company when the deal took place, they were replaced with C-level members.

By using this approach, a total of 979 unique profiles was found for Sample 1 and 214 for Sample 2.

Having identified founders (or C-level members), the following set of information was retrieved.

3.3.1 Gender

Data on gender was derived from founders' names and pictures found on the company's website and LinkedIn page. There were no cases in which identification was not possible.

3.3.2 Birth date

The birth date was either directly obtained by looking at the founder's profile on Orbis or indirectly inferred from information on graduation/high school completion date.

In case only the birth year was found, the birth date was assumed to coincide with the 1st of January of that year.

There were cases in which it was not possible to find the founders' birth date, but this was not detrimental to the similarity score computation shown in Chapter 5.

3.3.3 Nationality

Founders' nationality was retrieved by looking at their personal profiles on Orbis. As with the birth date, there were cases in which it was not possible to collect the data but, again, this lack of information did not severely impact the similarity score calculation.

3.3.4 Role start and end date

Logically, founders were assumed to begin their role on their start-up's date of incorporation. Conversely, C-level members' start date was found on either their personal LinkedIn pages or the company's website.

For active start-ups, role end date was obtained by looking at founders' (C-level members') LinkedIn pages, while for bankrupt companies it was assumed to coincide with the company's dissolution date, as given by Orbis.

3.3.5 Previous professional experience

Founders were categorized based on the prevalent professional experience they had before the deal date. In this respect, the following alternatives were identified:

- *academic,* in case a founder had at least one relevant academic experience (e.g. professorship);
- *financial,* in case a founder had at least one relevant professional experience in a financial institution (e.g. bank or asset management company);
- *entrepreneurial,* in case a founder had at least one relevant experience in a start-up or a corporate;
- *mixed – entrepreneurial/financial,* in case a founder had at least one relevant entrepreneurial experience and one financial experience, as previously defined;
- *mixed – entrepreneurial/academic,* in case a founder had at least one relevant entrepreneurial experience and one academic experience, as previously defined.

Information on professional experience was found by examining the founders' LinkedIn pages and *curricula* – when available.

There were cases in which it was not possible to retrieve founders' professional experience, but this had a limited impact on the similarity score computation.

3.3.6 Education level

The education level was defined by the qualification held by a founder. In this respect, the following qualifications were identified: High School Diploma, BSc, MSc, PhD, MBA, Post-Doctoral research.

Information on education level was collected by examining the founders' LinkedIn pages and *curricula* – when available.

There were cases in which it was not possible to find the data, but this was not detrimental to the similarity score computation.

3.3.7 Subject of study

Information on the subject of study was obtained by examining the founders' LinkedIn pages and *curricula* – when available.

To avoid excessive sample fragmentation, granular distinctions among subjects of the same type (e.g. mechanical engineering and electronic engineering) were not considered.

The cases in which it was not possible to find the data had a minor impact on the similarity score calculation.

3.3.8 Field of study

Field of study attribution directly descends from the subject of study (see Table V of Annex 1 for more details). With respect to the traditional classification proposed by the Italian Ministry of Education[7], it is worth mentioning that the following subjects have been reassigned to the Scientific field: i) Economics, ii) Finance and iii) Actuarial & Financial Science.

This choice avoids the creation of a large (and unrealistic) gap between those subjects and others belonging to the Scientific field (especially Engineering) when building the similarity score.

---

[7] Ministero Italiano dell'Istruzione, dell'Università e della Ricerca. "Raggruppamenti dei corsi di studio per Area disciplinare"

3.4 Data on VC partners

In order to compute the similarity score, the same information collected for start-up founders was retrieved for AMC and CVCs' partners. A total of 106 profiles was mapped for Sample 1 and 14 for Sample 2.

The only point of attention concerns the previous professional experience: indeed, some VC partners do not fall into the classification detailed in paragraph 3.3.5, as they have performed a mix of academic and financial roles. Those individuals were attributed the “Mixed – academic/financial” professional background.

4. Sample description

Before moving to the construction of the similarity score, it is worth providing more details on certain sample features. In this respect, paragraphs 4.1 and 4.2 focus on the descriptive analysis of founders and partners, while paragraph 4.3 gives an overview on the investments.

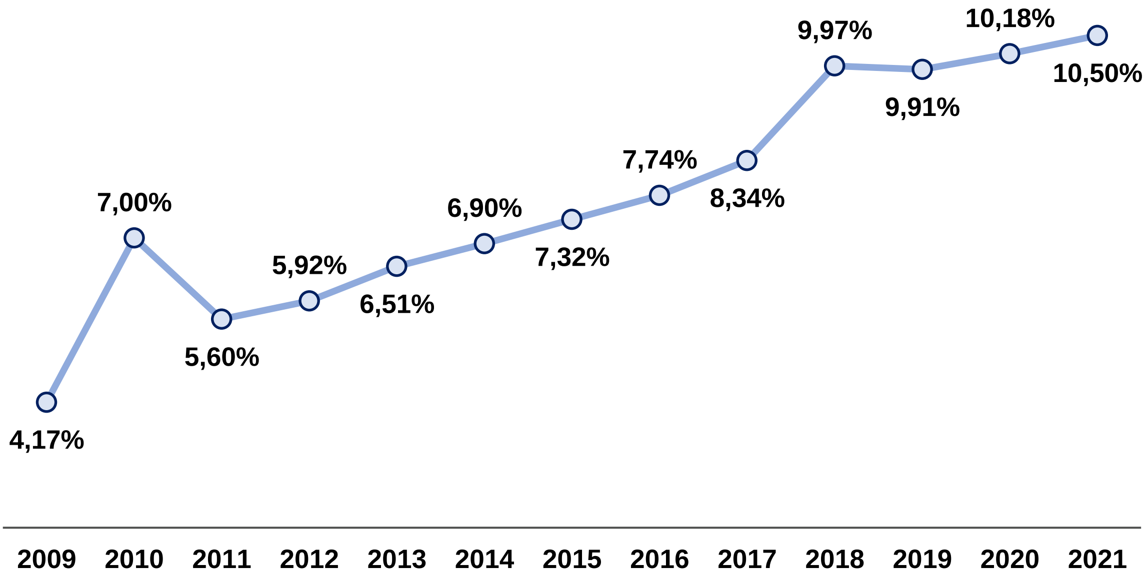
4.1 Founders

Sample 1

The sample of founders is made up of 979 individual profiles.

As far as the gender is concerned, male founders outnumber female founders by a factor of 9.5 (886 vs. 93). Since most of the firms included in Sample 1 are headquartered in Italy (see paragraph 4.3.3 for more details), this suggests that the Italian start-up ecosystem is heavily dominated by men, while women still struggle to emerge. It should be noticed, however, that the incidence of female founders has steadily increased over the last years. In this respect, *figure 1* shows the evolution of the percentage of women in the sample, which has more than doubled since late 2000s.

Figure 1. Cumulative incidence of Sample 1 female founders (%)



Note: *investments are cumulated over time. Years from 2000 to 2008 are not displayed due to the smallness of the sample.*

It is worth looking at founders' age when they launched their companies. The result is slightly higher than 36 years old, but this is likely to be an overestimation of the real datum, since many founders in the sample were not at their first entrepreneurial experience. This evidence is partially coherent with Azoulay *et al.* (2020), who used confidential administrative data sets from the U.S. Census Bureau covering the 2007-2014 period and found that, on average, US entrepreneurs start new ventures at the age of 42.

When it comes to the educational background, 82% of founders have successfully completed a university program (i.e. they are at least BSc graduates). Specifically, more than a third of them have a MSc (37%), while one fifth stopped at BSc level (21%). A negligible percentage (5%) did not earn a university degree.

Statistics on founders' educational background appear scarcely sensitive to gender, with the highest discrepancies registered for MSc (36% male founders vs. 43% female founders) and PhD (16% vs. 11%).

As for the previous professional experience, the vast majority of founders have an entrepreneurial background (65%), which is quite a predictable outcome. Notably, the financial component was part of founders' background only in 15% of cases. This result shall not be underestimated, as financial expertise becomes crucial during start-ups' fundraising and increases the opportunity to get better investment terms and conditions.

Data on previous professional background display only minor changes when the sample is segmented by gender.

A detailed set of Sample 1 founders' summary stats, divided by AMC, is provided in Table I of Annex 2.

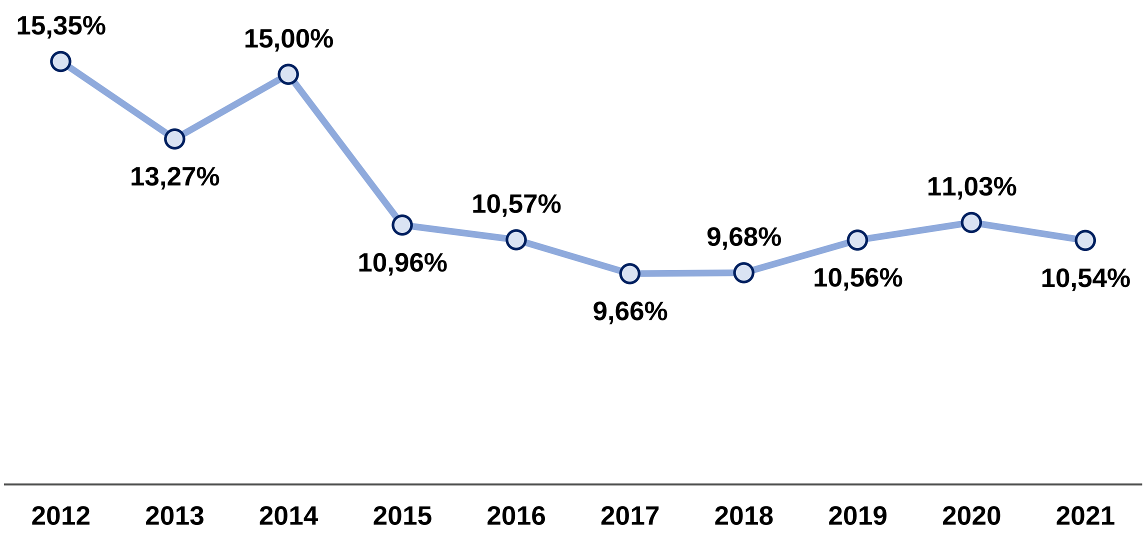
Sample 2

The sample of founders is made up of 214 unique profiles.

As with Sample 1, male founders are predominant, representing almost 90% of the total (187).

However, unlike Sample 1, the incidence of female founders has remained quite steady over the last years, oscillating around 10% from 2015 onwards. Interestingly, though, the two samples display very similar figures when the full sets of data are considered (10.50% Sample 1 vs. 10.54% Sample 2). Since also Sample 2 start-ups are mostly headquartered in Italy (see paragraph 4.3.7 for more details), this result appears coherent with that of Sample 1.

Figure 2. Cumulative incidence of Sample 2 female founders (%)



Note: *investments are cumulated over time. Years from 2000 to 2011 are not displayed due to the smallness of the sample.*

Data on Sample 2 founders are only partially influenced by gender. Male founders are, on average, slightly older than females (43 years vs. 41 years), but the two groups started their ventures approximately at the same age (33 years).

It is worth noting that, on average, companies in Sample 2 were started by younger founders than those in Sample 1[8].

For what concerns the educational background, only 5% of Sample 2 founders did not complete a university program, which is in line with the outcome of Sample 1. Moreover, as it happens in Sample 1, MSc (36%) and BSc (33%) are the most common qualifications.

[8] There is roughly a 3-year age difference between the two samples.

When it comes to the previous professional experience, the distribution appears even more extreme than that of Sample 1. In particular, 75% of Sample 2 founders have an entrepreneurial background (vs. 65% of Sample 1), while only 13% have a relevant financial experience in their resumes (vs. 15% of Sample 1).

Table II of Annex 2 provides a more detailed overview of Sample 2 founders' summary stats, sorted by CVC.

4.2 Partners

Sample 1

The sample of partners is made up of 106 individual profiles. 360 Capital Partners and Innogest Capital are the AMCs with most partners involved in deal execution over the period analysed (16 and 11 respectively). For 360 Capital Partners, this result can be motivated by the significant investment activity (see paragraph 4.3), while for Innogest Capital the sectoral specialization may have required a higher number of partners with strong technical expertise.

As for the gender, the Italian funds are heavily dominated by men. Nearly half of the AMCs never had a female partner over the sample period, and even when women are present, they are significantly outnumbered (CDP Venture Capital is the only AMC with more female than male partners). Moreover, female partners are generally older than male partners (53 vs. 50 years across the overall sample), albeit this shall not be interpreted as a proof that the time needed to reach the apical roles in Italian AMCs depends on the gender. In fact, when considering the age at which the 106 individuals in the sample became partners, there is no clear evidence that women are penalized.

It is also interesting to look at the average age that partners were at the time they participated in the different deals. The result for the whole Sample 1 is approximately 47 years, but this cannot be taken as a good generalization of the dynamics affecting each investment firm analysed. Indeed, when looking at data per AMC, a great variability arises, with the highest value (60 years for AVM Gestioni) and lowest value (35 years for Lumen Ventures) differing by 25 years.

As for the educational background, partners have, on average, higher level degrees than founders, which seems a reasonable outcome given the stricter requirements needed to achieve the role. Specifically, more than a half of Sample 1 partners have a MSc (vs. 37% of Sample 1 founders) and 29% of them have an MBA (vs. 8% of Sample 1 founders). Only 1 partner in Sample 1 did not earn an academic degree.

Differences between founders and partners are even deeper when it comes to previous professional experiences. Predictably, almost all partners (98%) have at least some relevant financial expertise: this seems an obvious result, given that AMCs operate in the financial sector.

A complete set of partners' summary stats is provided in Table III of Annex 2.

Sample 2

Sample 2 is made up of 14 partners, who are evenly distributed across the CVCs.

As with Sample 1, male gender is the most represented: 80% of partners are men, and 4 out of 7 CVCs have never had a female partner.

Sample 2 partners are roughly as old as their colleagues in Sample 1 (53 vs. 51 years old). However, in contrast with Sample 1, there is a small age difference between male and female partners in Sample 2 (53 vs. 52 years old).

Interesting observations can be made from the analysis of the educational and professional backgrounds. Specifically, all Sample 2 partners have completed an academic degree: 50% of them have a MSc, 36% a more prestigious title (MBA or PhD), 7% a BSc[9]. This result is coherent with that of Sample 1.

However, differently from Sample 1, a non-negligible portion of partners in Sample 2 have a non-financial background (21%). Partially, this may be explained by the peculiar nature of CVCs, i.e. divisions built inside corporates whose core business have often little in common with the financial industry and sometimes managed by partners who achieved their role after working in different areas of the same corporate (or for other corporates).

Table IV of Annex 2 offers a detailed set of Sample 2 partners' summary stats.

---

[9] Percentages do not sum to 100% because for 1 partner it was not possible to retrieve the datum on the educational background.

4.3 Investments

Sample 1 includes 593 VC initial investments made from January 2000 to December 2021 by 19 Italian AMCs, while Sample 2 was built by aggregating 95 investments closed over the same period by 7 Italian CVCs. As already explained in Chapter 3, follow-on transactions were not taken into account.

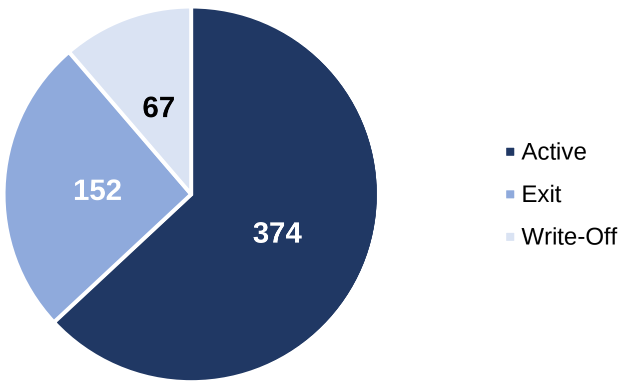
Sample 1

4.3.1 Investment status breakdown

360 Capital Partners is the most active AMC (108 deals). This comes as no surprise: it is the oldest AMC in the sample (its first fund was started in the early 2000s), the largest (8 funds, with more than € 650M raised) and the most international (more than three quarters of investments made outside Italy).

Figure 3 shows the distribution of the investments by status.

Figure 3. Sample 1 investments' breakdown by status (#)



As it can be noticed, most of the investments are still active (374), which depends on the combination of two main factors.

One the one hand, only 10 out of the 46 funds mapped are completely divested, while the others have been started not so long ago, so that are now either in the investment or in the portfolio management phase. This, in turn, mainly depends on the young age of the Italian VC ecosystem, to which the national AMCs are inevitably linked[10].

---

[10] As paragraph 4.3.3 shows, Italian AMCs make most of their investments in Italian start-ups.

On the other hand, public coverage on Italian funds' activity has considerably increased with time, so that it has been easier to retrieve information on the most recent investments. Not by chance, albeit the sample spreads over 21 years, 55% of deals analysed happened in the last 5 years (70% in the last 7).

The remaining 219 shareholdings were liquidated because of either the investee's failure (67) or a liquidity event as defined in the Glossary (152).

In detail, Xyence (formerly Principia) is the AMC with the highest number of write-offs (17, almost one third of the investments made), but this datum is likely to be influenced by the typical VC's lack of transparency on the least successful transactions. In other words, the number of write-offs in the sample and the implied failure rate (11%) are likely to be an underestimation of the real figures. This intuition is confirmed when looking at VC-backed companies write-off rates reported by the literature, which, according to the investment stage and the definition of failure, vary from 30% to 75% (Gage, 2012).

When it comes to the 152 exits, two thirds stem from M&A (100), while only a small part is due to public market listing (10). This is fully coherent with the overall maturity of the Italian VC market, far behind the top European ecosystems (e.g. UK and Germany), where IPOs are much more frequent[11].

Interestingly, although more than 70% of investments involved Italian start-ups (see paragraph 4.3.3 for more details), foreign companies represent 43% of exits, peaking around 50% for M&A and IPOs.

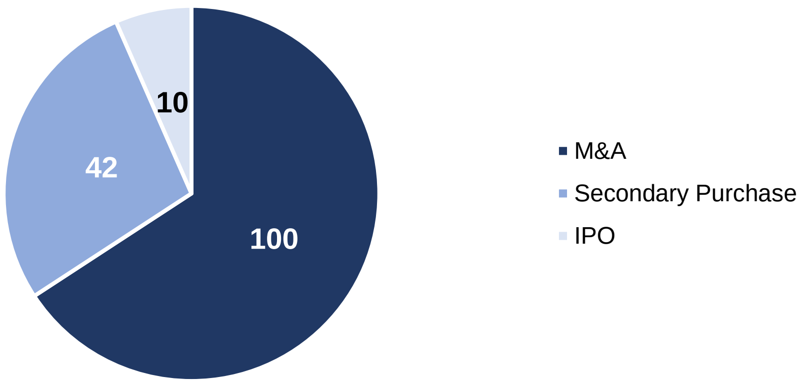
Moving to the analysis of data at AMC level, 360 Capital Partners is responsible for almost one third of the exits (47). This depends, at least partially, on the better performance of its investments when compared to the other AMCs in the sample.

The median time to exit across the overall sample is 44 months, which is coherent with the typical investment horizon of VC funds. Moreover, among the AMCs with at least

[11] Since January 2000 to December 2021, there have been 181 IPOs of UK companies backed by UK VC funds, and 79 IPOs of German start-ups backed by German VCs. Data have been extracted from PitchBook.

10 exits realized[12], 360 Capital Partners has the quickest median time to exit (44 months), while Vertis displays the worst result (70 months). Among the factors explaining this marked discrepancy, the difference in geography of investments (liquid foreign markets for 360 Capital Partners, mainly south of Italy for Vertis) may have played a key role.

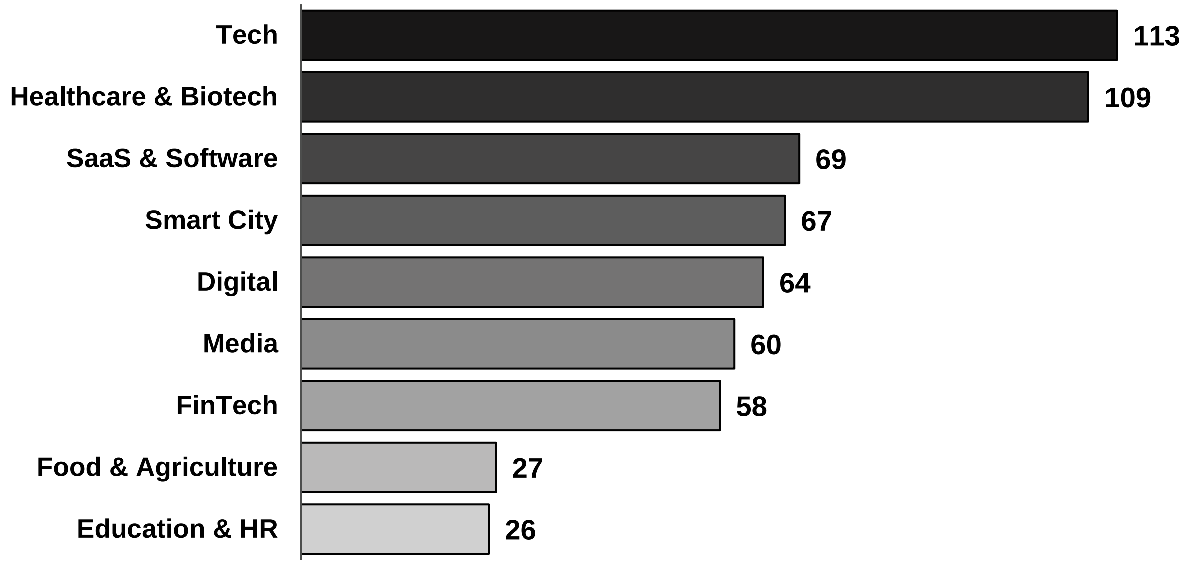
Figure 4. Sample 1 exits' breakdown by cluster (#)



4.3.2 Sector and primary vertical breakdown

Figure 5 offers a sectoral breakdown of the investments covered by the analysis.

Figure 5. Sample 1 investments' breakdown by sector (#)



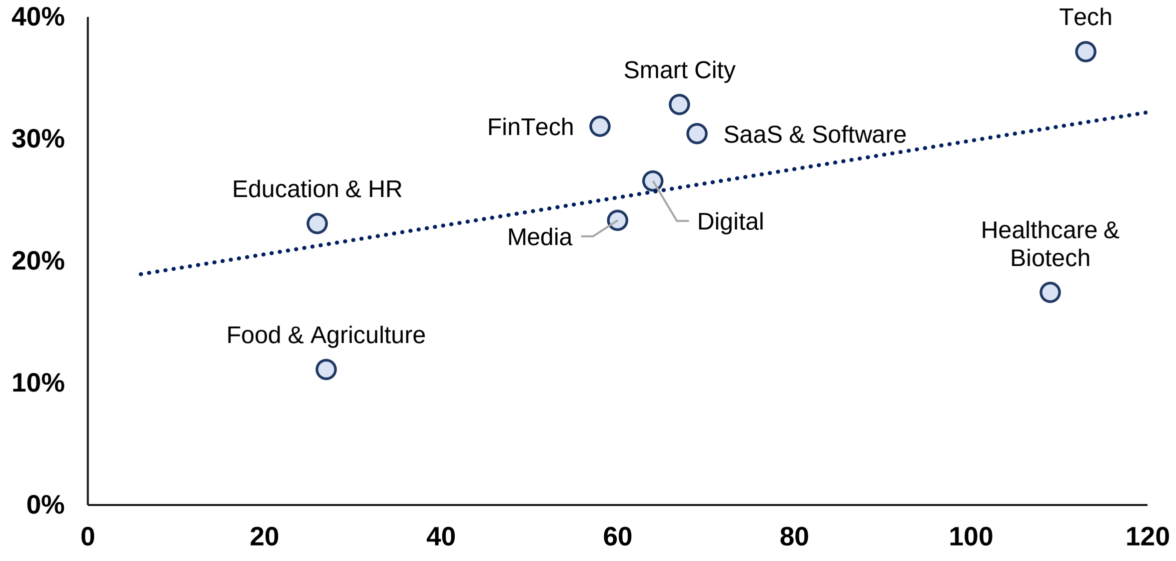
[12] The choice of imposing a cutoff of 10 investments when comparing averages and medians aims at reducing the bias in data. When samples are bigger (as it happens in subsequent chapters of the paper) the threshold is raised to 30.

The sample appears rather concentrated, with the top 3 sectors combined representing almost 50% of the investments. Specifically, Tech is the most popular sector (113), followed by Healthcare & Biotech (109) and SaaS & Software (69).

The large importance of Tech and SaaS & Software is quite expected, as they are the two sectors normally accommodating the most innovative ventures. On the other side, the eminent role of Healthcare & Biotech is explained by the presence of three AMCs strongly focused on this sector, i.e. Innogest Capital, Panakès Partners and Xyence. Not by chance, the three aforementioned AMCs account for almost half of the investments in Healthcare & Biotech.

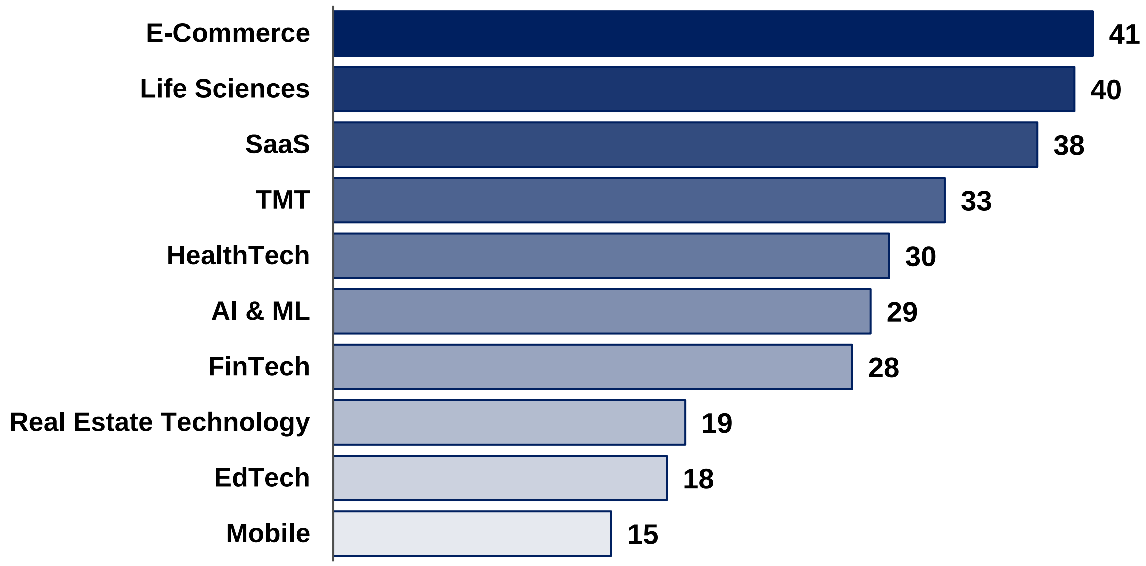
Another relevant observation is obtained by crossing the sectoral breakdown with the investees' headquarters country. Apparently, the incidence of non-Italian start-ups markedly varies based on the sector considered, ranging from 11% (for Food and Agriculture) to 37% (for Tech). Moreover, aside from Healthcare & Biotech (which, as said before, is characterized by peculiar dynamics), the presence of foreign companies is higher in the most invested sectors (see *figure 6*). This result is perfectly understandable: as the AMCs' appetite for certain sectors increases, their search for investment opportunities may encourage them to explore potential targets outside national boundaries, which ultimately leads to a higher incidence of foreign financings.

Figure 6. Incidence of non-Italian start-ups in Sample 1 (%) as a function of the number of deals, by sector



When it comes to business verticals, the analysis distinguishes among 78 unique items. However, data presents a certain homogeneity, with the 10 most invested verticals covering almost half of the deals. Specifically, E-Commerce is the most popular choice (41), followed by Life Science (40) and SaaS (38). *Figure 7* offers an overview of the top 10 verticals by number of deals associated (AI & ML stands for Artificial Intelligence & Machine Learning).

Figure 7. Sample 1 investments' breakdown - top 10 verticals (#)

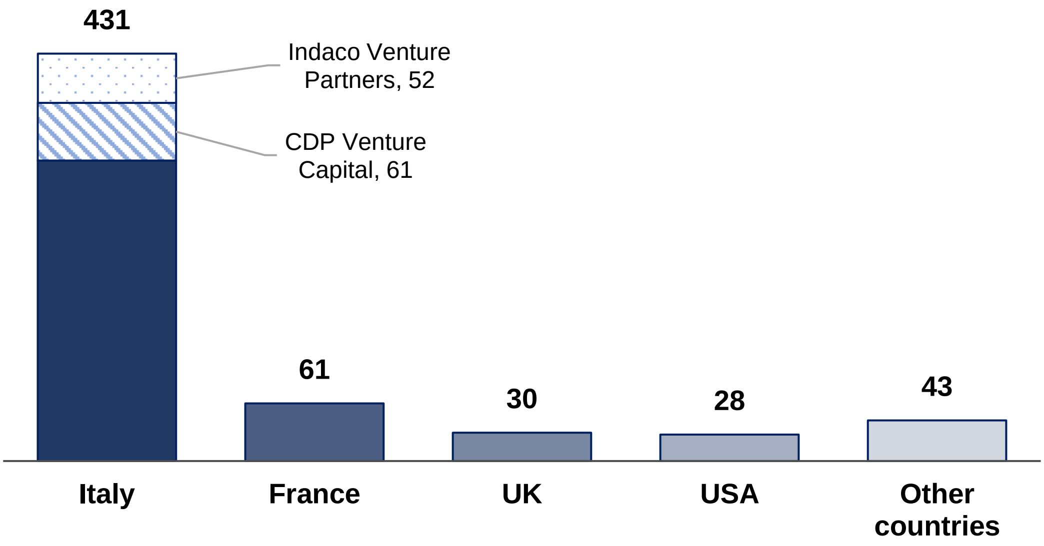


4.3.3 Geography, stage & series breakdown

The large part of the investments involved Italian start-ups (431): this observation suggests that investment decisions are somewhat biased by geographical affinity considerations.

Figure 8 shows the distribution of investments according to the investee's headquarter country.

Figure 8. Sample 1 investments' breakdown by geography (#)



Note: *other countries include Spain (13), Germany (7), Israel (6), Switzerland (6), Netherlands (5), Austria (1), Finland (1), Ireland (1), Singapore (1), Sweden (1) and UAE (1).*

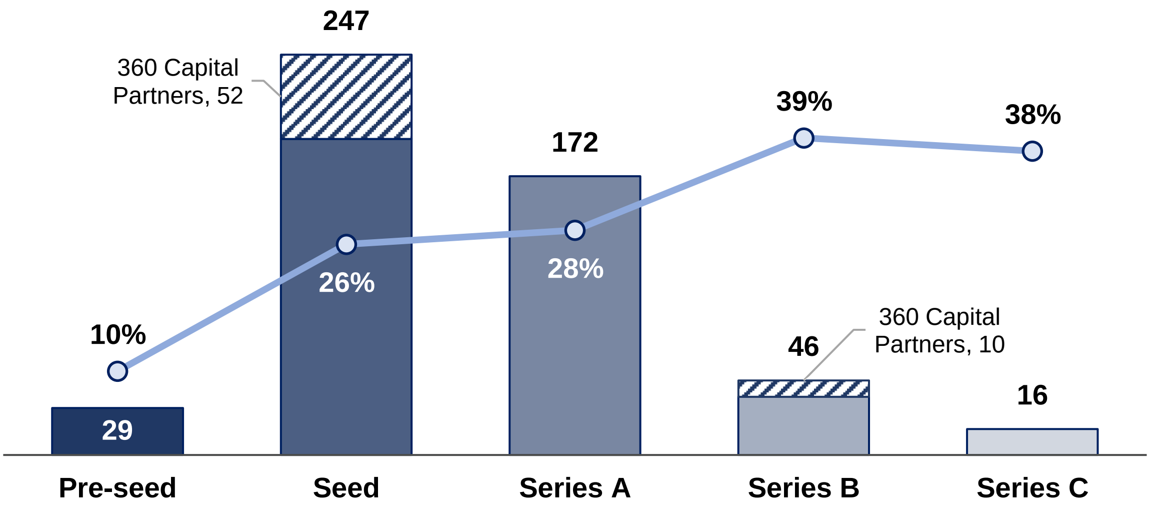
Predictably, 360 Capital Partners is the most internationally oriented AMC, with 82 deals made abroad (56 in France).

On the other hand, CDP Venture Capital and Vertis are the two AMCs with the lowest activity in foreign markets among those with at least 30 investments made. This is explained by government-imposed regulatory constraints (in the former case) and by the explicit choice to allocate considerable financial resources to the south of Italy (in the latter case).

As far as the investment stage is concerned, more than 80% of transactions involved early stage financing, with an overall median deal size of € 2.5M. These data furnish a further proof of the relative underdevelopment of the Italian VC market: indeed, more mature realities (e.g. UK and Germany) typically have higher median values because of the higher incidence of later stage financing.

With respect to the funding series, Seed (243) and Series A (173) are the sweet spot of Italian AMCs' investments. Notably, the incidence of 360 Capital Partners is significant across all deal types, reaching its peak in Seed (11%) and Series B (15%). Additionally, the data suggest a strong correlation between funding series and investee's geography: the relevance of non-Italian start-ups increases as later stage transactions are considered (see *figure 9*).

Figure 9. Sample 1 investments' breakdown (#) and incidence of deals in non-Italian start-ups (%), by funding series



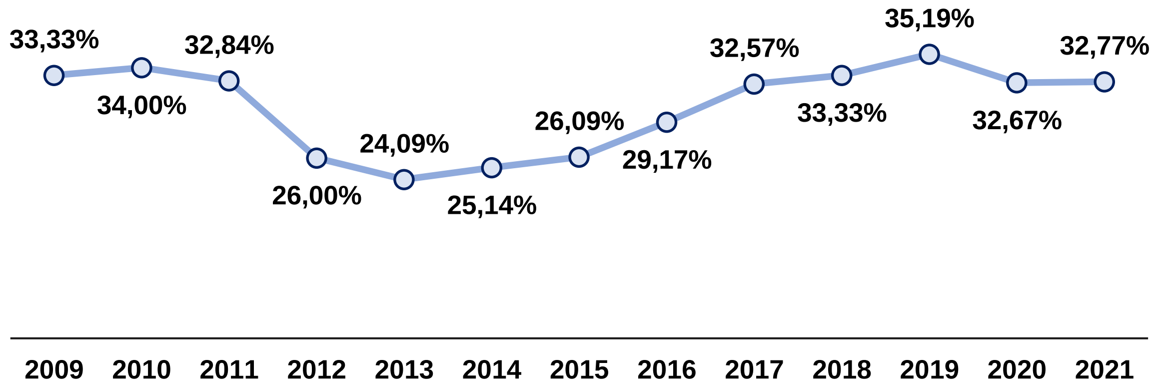
Note: *there are two reasons why the graph does not consider all the 593 investments. First, it was not always possible to identify the funding series. Secondly, it appeared more sensible to stop the analysis at Series C, as for superior series there were not enough investments to reach statistically meaningful conclusions.*

4.3.4 Round structure

Almost three quarters of deals were closed in syndicate (439). The cases in which AMCs acted as sole investors mainly involved initial funding series (Pre-seed and Seed): this result is totally expected, since the probability of collaboration among investors increases as the round amount and complexity rise.

As for the investors' geography, 194 out of the 593 transactions mapped involved the participation of at least one foreign player. From a merely numerical point of view, the incidence of rounds with international investors has remained quite stable over time, ranging from 24% to 35% of total deals (see *figure 10*).

Figure 10. Cumulative incidence of Sample 1 rounds with 1+ foreign investor (%)

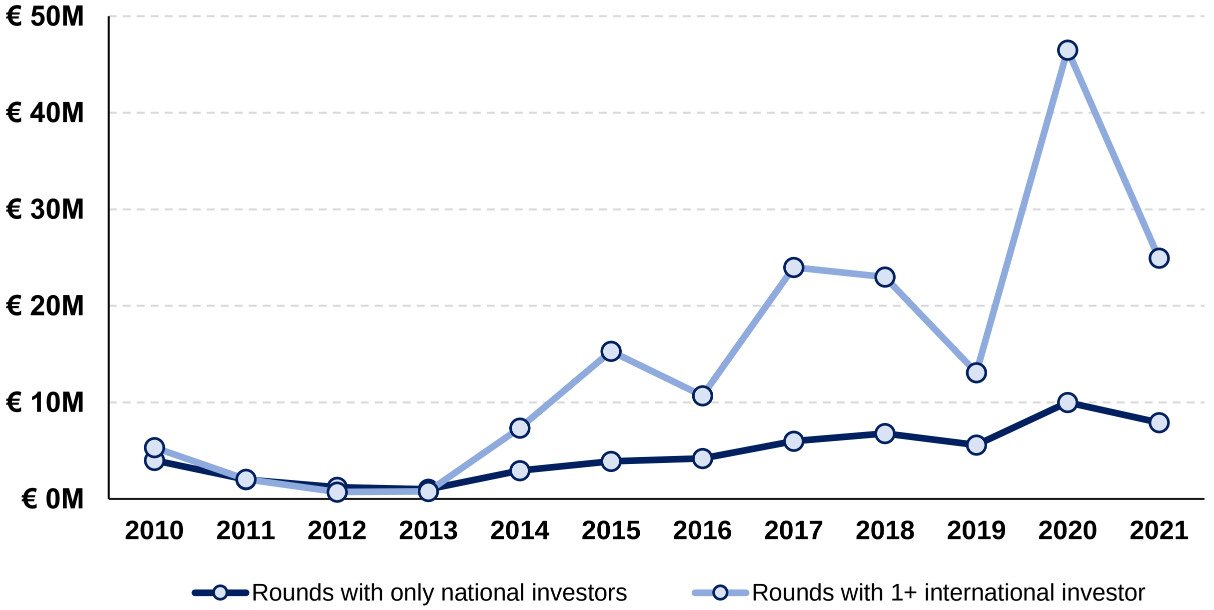


Note: *investments are cumulated over time. Years from 2000 to 2008 are not displayed due to the smallness of the sample.*

Notably, the comparison between median deal sizes reveals that rounds with foreign investors are typically larger than those with only national participants. Furthermore, as *figure 11* shows, this difference has deepened over time: while the median round size was approximately comparable in the early 2010s, the last years have seen transactions with international players being worth from 2 to 4 times the deals with sole national investors.

Within this framework, it is useful to recall that the sample covers only initial investments, disregarding follow-on transactions. Given that subsequent financings are typically larger than initial ones, they are more likely to attract investors from more mature international markets. Therefore, extending the analysis to all the investments made by the AMC in the sample would probably show an even higher divergence between the two medians.

Figure 11. Median deal size historical evolution, by type of investor



***Note:** years from 2000 to 2009 are not displayed due to the smallness of the sample.*

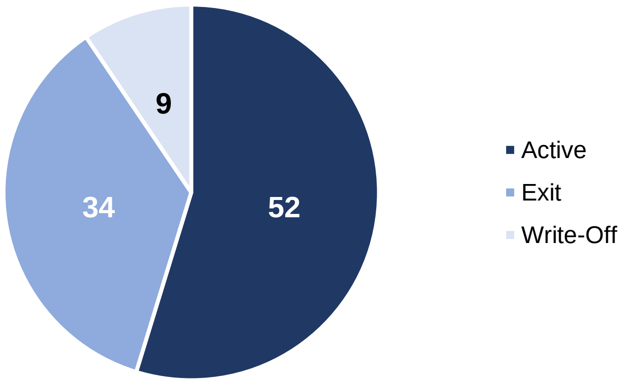
Sample 2

4.3.5 Investment status breakdown

ZernikeMeta Ventures is the most active CVC, with 29 deals closed (ca. 30% of the total in Sample 2). This result was quite predictable: it is the oldest CVC mapped (its first two regional funds were launched in 2004) and the only to manage multiple funds (5).

Figure 12 displays the distribution of Sample 2 investments by status.

Figure 12. Sample 2 investments' breakdown by status (#)



As with Sample 1, active shareholdings represent the majority of the investments (52 out of 95), but their incidence on the total number of deals is lower than in Sample 1

(55% vs. 63%). This happens, in part, because funds in Sample 2 have, on average, older vintages than those in Sample 1 (43% of funds in Sample 2 were started before 2010). Consequently, Sample 2 has a higher percentage of funds in the divestment phase than Sample 1.

The remaining 43 investments were either written off (9) or dismissed on the occasion of a liquidity event as defined in the Glossary (34).

For what concerns the write-offs, the implied failure rate registered in Sample 2 (9,5%) is not far from that of Sample 1 (11%). Clearly, the same considerations made for Sample 1 write-offs apply also for Sample 2: the volume of investments is likely to be underestimated, and this is especially true for the unsuccessful transactions, so that the real failure rate may be higher. This hypothesis is strengthened by looking at the breakdown of write-offs among the CVCs: only for 2 of them (ZernikeMeta Ventures and Tim) was it possible to find public information on write-offs, which is quite an unplausible outcome given the typically high-risk profile of VC investments.

As regards the 34 exits, M&A is the predominant cluster (20), followed by sales on the secondary market (13). The only case of IPO concerns Westwing, which had been originally invested by Mediaset. The low number of IPOs in Sample 2 could appear in contrast with the high incidence of investments made in liquid foreign VC markets (42%)[13]. However, this result can be explained by recalling that more than half of the investments in foreign Sample 2 start-ups are relatively recent (made from 2019 onwards) and involved companies in their very early stages. Consequently, time is still needed to see how non-Italian shareholdings will unfold.

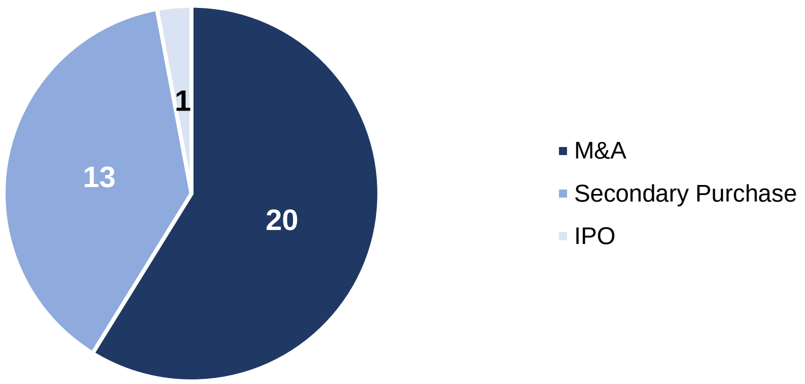
Looking at the segmentation by CVC, ZernikeMeta Ventures has the highest number of exits (14, representing 48% of the investments made). Conversely, Zanichelli has still not registered any liquidity event across its investees: however, this may well be explained by the recent timing of its transactions, which were all performed from 2019 to 2021.

---

[13] Compare this datum with the 27% registered in Sample 1.

The median time to exit across the overall Sample 2 is 37 months, half a year less than in Sample 1. There is wide variability across the CVCs, but the conclusions reached on this particular point appear scarcely reliable given the small sample size.

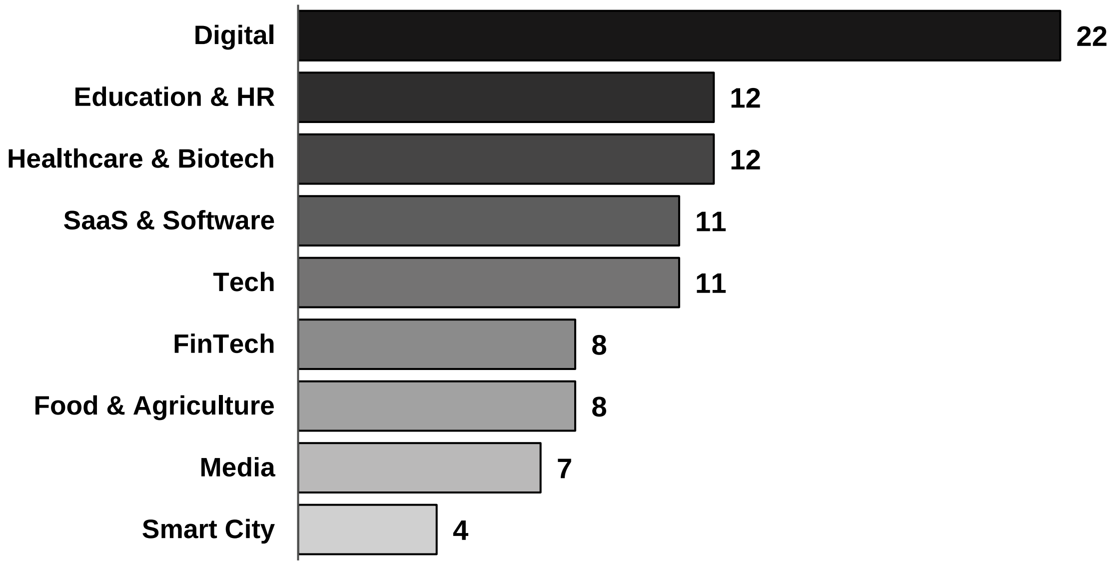
Figure 13. Sample 2 exits' breakdown by cluster (#)



4.3.6 Sector and primary vertical breakdown

Figure 14 shows the sectoral breakdown of Sample 2 investments.

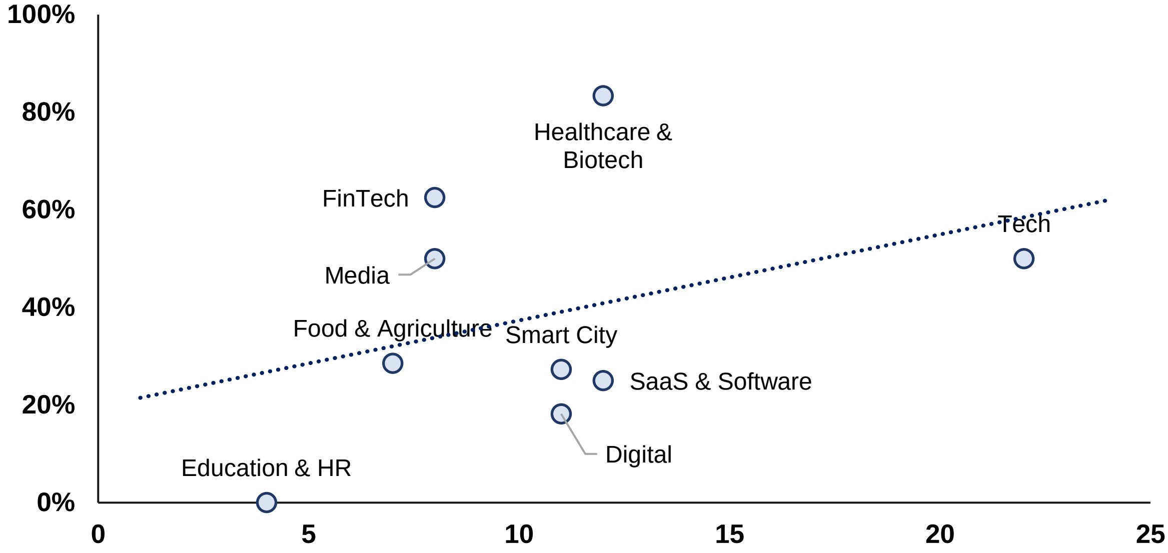
Figure 14. Sample 2 investments' breakdown by sector (#)



As with Sample 1, the distribution in Sample 2 is rather concentrated, with the top 3 sectors accounting for almost half of the investments. However, there are relevant differences in the most popular sectors, which this time are represented by Digital (22), Education & HR (12) and Healthcare & Biotech (12). This outcome largely depends on the strong sectoral specialization of the CVCs covered in the analysis: indeed, Mediaset accounts for more than 50% of the investments made in Digital, Zanichelli for 90% of those in Education & HR and ZernikeMeta Ventures for two thirds of those in Healthcare & Biotech.

Additionally, Sample 2 displays the same positive relationship between the number of investments made in each sector and the incidence of non-Italian start-ups noted in Sample 1, even if the sample fit is slightly weaker[14].

Figure 15. Incidence of non-Italian start-ups in Sample 2 (%) as a function of the number of deals, by sector

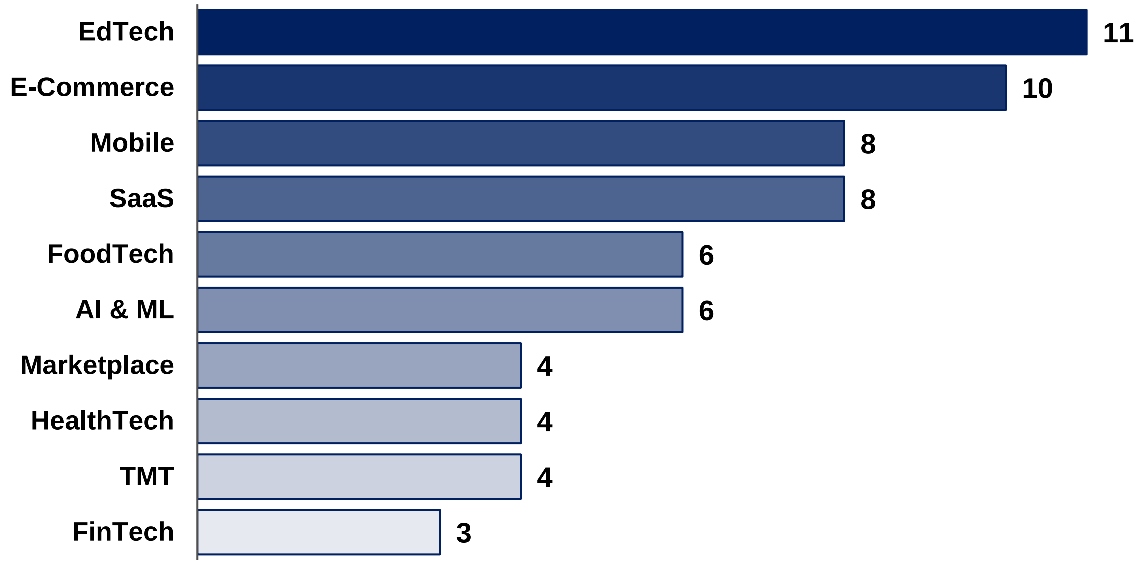


As for the breakdown by business vertical, a total of 30 unique specifications was registered in Sample 2. Nevertheless, as it happens in Sample 1, the distribution is quite homogenous, with the top 5 verticals covering 45% of the investments (the incidence rises to 67% when the top 10 verticals are considered). Similar to what observed for sectors, the concentration of data is explained by the strong specialization of the CVCs mapped: for instance, almost all the investments in EdTech (11) were made Zanichelli (9), while 80% of those in E-Commerce (10) involved Mediaset.

Figure 16 offers an overview of the top 10 verticals in Sample 2 for number of deals.

[14] Also, given the difference in sector ranking by number of investments, the order displayed in the figure is different from that of Sample 1.

Figure 16. Sample 2 investments' breakdown - top 10 verticals (#)

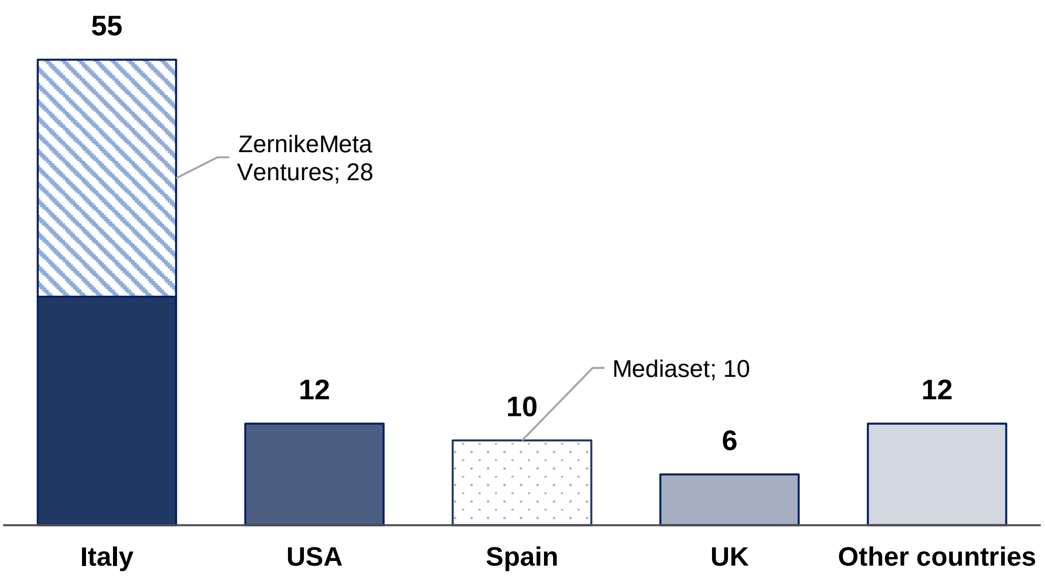


4.3.7 Geography, stage & series breakdown

In contrast to Sample 1, Sample 2 presents a lower share of Italian start-ups (58% vs. 73%). *De facto,* almost all the CVCs covered in this paper devoted a considerable proportion of their investments abroad (for Zanichelli and Barilla, this datum is as high as 100%).

Specifically, *figure 17* displays the distribution of deals according to the invested start-ups' headquarters.

Figure 17. Sample 2 investments' breakdown by geography (#)



Note: *other countries include Germany (4), Israel (2), Luxemburg (2), Finland (1), Singapore (1), Austria (1) and Norway (1).*

As it can be noticed, ZernikeMeta Ventures is responsible for more than half of the investments in Italian start-ups. This results from the combination of two main factors: from the one side, it is the most active CVC in the sample; on the other side, its geographical focus is quite narrow, as it manages regional funds intended to promote economic development in specific parts of Italy.

Moreover, it is worth mentioning that all the investments in Spain (10) were closed by Mediaset. This comes as no surprise, since its investment vehicle (Ad4Ventures) is based both in Italy and Spain and seeks synergistic opportunities to Mediaset's business in these two countries.

Predictably, rounds involving foreign start-ups tend to be bigger than those involving Italian start-ups (median size € 2.24M vs € 0.59M). This furnishes a further proof of the Italian VC market underdevelopment with respect to other ecosystems.

Anyway, it is possible to say that, in general terms, Sample 2 rounds are rather small in size. The median size across the overall Sample 2 is € 0.95M, even lower than that of Sample 1 (€ 2.5M): this supports the idea that the CVCs covered in the analysis prefer investing making minority investments in small companies at the first stages of their lifecycle[15].

This view is confirmed by the investments' breakdown by stage: Sample 2 displays similar results to Sample 1: 75% of transactions were labelled as "Early Stage VC" and only 11% as "Late Stage VC".[16]

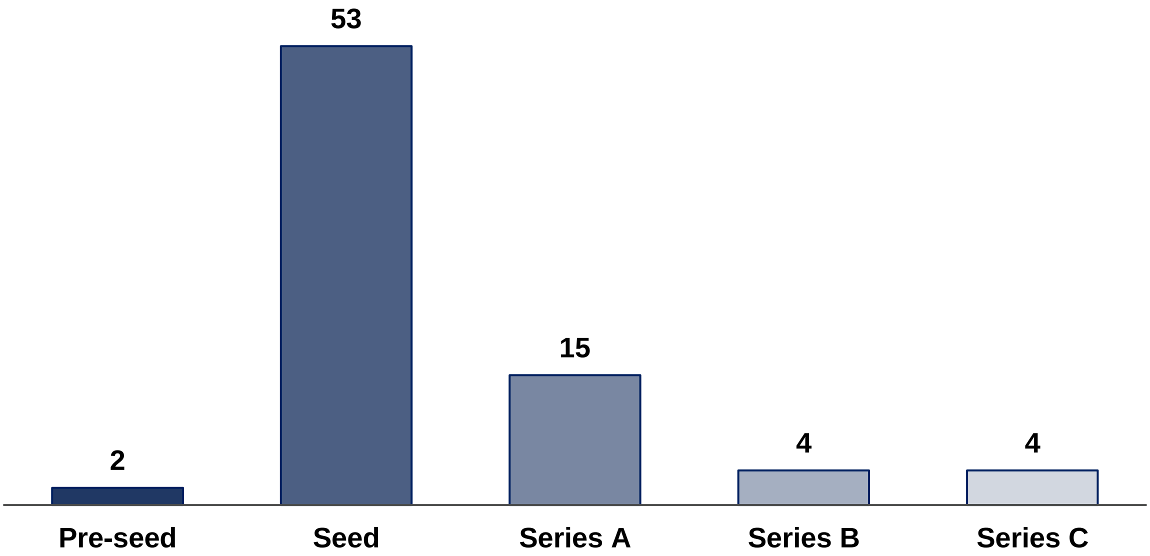
When it comes to the funding series, Sample 2 is aligned with Sample 1: Seed (53) is by far the most frequent case, followed by Series A (15). The other funding series reached negligible frequencies. At CVC level, almost all the corporates mapped have a strong bias towards Seed investments, except for Mediaset and Barilla, which display more balanced distributions.

Figure 18 shows the investments' breakdown by funding stage.

[15] However, CVCs' investments were rarely made at the *very first* stage of start-ups' lifecycle (in this regard, see data on Pre-seed transactions at the end of the paragraph).

[16] For 14 deals it was not possible to unambiguously assign the investment stage.

Figure 18. Sample 2 investments' breakdown by funding series (#)



Note*: there are two reasons why the graph does not consider all the 95 investments. First, it was not always possible to identify the funding series. Secondly, it appears more sensible to stop the analysis at Series C, as there were not enough investments of superior series to reach statistically meaningful conclusions.*

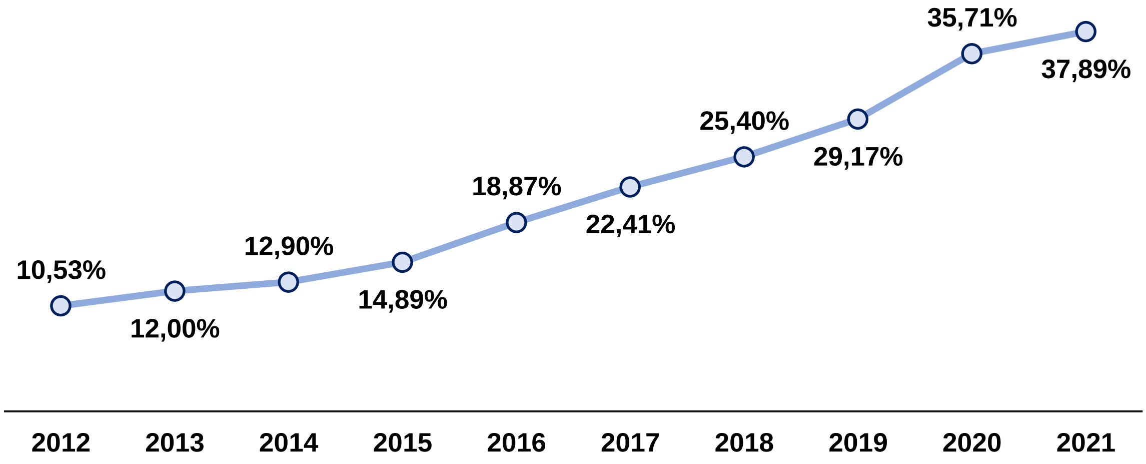
4.3.8 Round structure

As with Sample 1, the vast majority of Sample 2 transactions involved investors' syndicates (75). This result appears reasonable: corporates normally prefer collaborating with other VC investors (especially VC funds) to avoid being solely responsible for due diligence and negotiation.

This intuition if further strengthened by looking at the incidence of deals in which the CVCs acted as lead investors (only 26%).

For what concerns the investors' geography, 38% of the transactions mapped in Sample 2 involved the participation of at least one foreign investor. This outcome is in line with what observed in Sample 1. However, in contrast with Sample 1, the percentage of rounds with international investors has steadily increased, approaching 40% in 2021 (see *figure 19)*.

Figure 19. Cumulative incidence of Sample 2 rounds with 1+ foreign investor (%)



Note*: investments are cumulated over time. Years from 2000 to 2011 are not displayed due to the smallness of the sample.*

5. Model Construction

For each transaction mapped in Sample 1 and 2, a similarity score was built in order to capture the degree of affinity between the start-up's founding team and the partners of the fund that participated in the deal. The score was computed by taking the weighted average of seven variables measuring cultural, ethnic and personal differences between investors and founders:

- difference in gender (Δgen)
- difference in age (Δage)
- difference in nationality (Δnat)
- difference in previous professional experience (Δexp)
- difference in education level (Δed)
- difference in field of study (Δfst)
- difference in subject of study (Δsst)

Thus, the full expression of the similarity score is given by:

$$\begin{aligned} &Similarity\ score_i \\ &\qquad = \alpha_1 \cdot \Delta gen_i + \alpha_2 \cdot \Delta age_i + \alpha_3 \cdot \Delta nat_i + \alpha_4 \cdot \Delta exp_i + \alpha_5 \cdot \Delta ed_i \\ &\qquad + \alpha_6 \Delta fst_i + \alpha_7 \cdot \Delta sst_i \end{aligned}$$

where each difference is computed for the i-th investment.

Each variable was built to oscillate between 0 and 1, so that the similarity score takes minimum value of 0 (signalling absence of the affinity bias) and maximum value of 1 (signalling strong presence of the affinity bias).

Paragraphs 5.1 and 5.2 describe the two-step procedure which was followed to calculate each differential in the formula, while paragraph 5.3 focuses on the identification of the system of weights.

5.1 Step 1: Computation of the mode

First of all, individual data on founders and partners were aggregated at team level by computing either their statistical mode (for non-numerical variables) or their average (for numerical variables). Notably, age was the only term for which it was possible to apply the average, while the others required mode calculations.
The criteria applied to compute the average (mode) of the (non-)numerical terms vary.

5.1.1 Mode of Gender

When it was possible to compute the mode, gender was treated as a dummy variable taking value of 1 if most founders (partners) were male and 0 if they were female.
When the mode could not be calculated – consider, for instance, teams made up of 1 male founder (partner) and 1 female founder (partner) – gender was given the intermediate value of 0.5.

5.1.2 Average Age

As said before, age is the only numerical variable included in the similarity score computation. Therefore, the average age was computed for both the founding team and the group of VC partners participating in the investment.

5.1.3 Mode of Nationality

For each transaction, the most frequent nationality among the founders (partners) was taken. For instance, a team of 2 Italian founders (partners) was considered Italian.
The cases in which it was not possible to compute the mode were treated differently depending on whether the problem concerned founders or partners.

For founders, there were situations in which the mode was not retrievable, but nonetheless there was 1 Italian member in the team. In this scenario, the founding team was attributed the Italian nationality in order to capture the known tendency of Italian VC funds to invest in start-ups somewhat linked with Italy.
The other cases of mode unavailability were those in which two foreign nationalities had the same frequency: in this scenario, the two nationalities were given the same importance in the similarity score computation. For example, a team made up of 1 French founder and 1 Dutch founder was considered both Dutch and French. Clearly,

this choice impacts the value of the difference in nationality (Δnat): for instance, consider the case of two founding teams, one made up of 2 French individuals and the other comprising 1 French and 1 American individual. When calculating the difference in nationality (Δnat) between each of these two teams and a group of, say, Italian partners, the former was given a lower value than the latter because discrepancies between the French and the Italian nationalities are less evident than those between the American and the Italian ones.

There were no situations with 3 or more equally frequent nationalities.

As for partners, there was no need to check whether at least 1 member of the team was Italian, as all VC funds were predominantly composed of Italians. Thus, mode calculation was straightforward.

5.1.4 Mode of Previous professional experience

For each team of founders (partners), the most frequent previous professional experience was taken – when available.

In case the mode was not computable, the partial overlapping between different professional experiences was enhanced. For instance, a team consisting of 1 founder (partner) with “Mixed – entrepreneurial/financial” experience and 1 founder (partner) with pure “Financial” experience was assumed to have a *financial* background.

5.1.5 Mode of Education level

For each transaction, the most frequent education level among the founders (partners) was taken.

In case of lack of a statical mode, the member with the highest education level determined the value applied to the whole team. To this purpose, the following study path was considered:

- level 1: High School Diploma
- level 2: Bachelor of Science (BSc)
- level 3: Master of Science (MSc)
- level 4: MBA and PhD
- level 5: Postdoctoral

For example, a team made up of 1 founder (partner) with a BSc and 1 founder (partner) with a MSc was ascribed the latter's education level.

5.2.6 Mode of Field of study and Subject of study

For each team of founders (partners), the most frequent field and subject of study were considered – when available.

The cases where it was impossible to compute the mode were not solved at this stage but rather in the second step of the similarity score calculation procedure (see paragraph 5.3.6 and 5.3.7 for more details).

5.2 Step 2: Computation of the differences

The second step of the procedure consists in measuring how much the mode (average) of each variable in the similarity score differs between founders and partners. The approach followed varies based on the variable.

5.2.1. Difference in Gender

The difference in gender between founders and partners was computed simply by taking the absolute difference of the modes:

$$\Delta gen_i = \left|gen_{f,i} - gen_{p,i}\right|$$

where $gen_{f,i}$ is the mode of founders' gender and $gen_{p,i}$ is the mode of partners' gender for the i-th investment.

5.2.2 Difference in Age

For each investment in the sample, the difference in age was computed by taking the absolute standardized distance between the average founders' and partners' age.

Standardization was made with respect to the maximum absolute difference in age found in the sample and reflects the need to make the variable oscillate between 0 and 1.

The expression of the difference in age in the i-th investment can then be written as:

$$\Delta age_i = \frac{\left|age_{f,i} - age_{p,i}\right|}{\max\left|age_f - age_p\right|}$$

where $age_{f,i}$ is the average founders' age and $age_{f,i}$ is the average partners' age. Both measures relate to the i-th investment.

5.2.3 Difference in Nationality

Each pair of founders' and partners' nationality was assigned a distance taking into account geographical, linguistic and cultural differences. For obvious reasons, the distance attributed vary between 0 and 1, where the 0 (1) signals identical (very different) nationalities.

5.2.4 Difference in Previous professional experience

Similar to what was done for nationality, each pair of professional experiences received a distance based on their degree of affinity. For instance, the pair "Financial" & "Mixed-entrepreneurial/academic" was attributed the highest distance (1), as these two backgrounds have weak commonalities. Conversely, the pair "Financial" & "Mixed-entrepreneurial/financial" was assigned a smaller distance (0.5) because the two backgrounds share the financial component. A null distance was assigned in case of identical previous professional experiences.

5.2.5 Difference in Education level

Each pair of education levels was attributed a distance based on the study path illustrated in paragraph 5.1.5.
For instance, the pair "High School Diploma – Postdoctoral" was given the maximum distance (1), as these two education levels are at opposite ends of the study path. Conversely, the pair "BSc – MSc" was assigned a smaller distance (0.25), since the two education levels are contiguous in the study path. A null distance was attributed in case of identical education levels.

5.2.6 Difference in Field of study

When it was possible to compute the mode of the field of study both for founders and partners, the distance was set to 1 if the modes were equal, 0 if they were different.
In case of absence of one of the two modes, a distance of 0.5 was attributed if there was a match in the field of study between at least 1 founder and 1 partner, otherwise

the variable was excluded from the computation of the similarity score because of the lack of unambiguous data.

5.2.7 Difference in Subject of study

As for the subject of study, a similar approach to that applied for the area of study was followed. Therefore, when it was possible to compute the mode both for founders and partners, the distance was set to 1 if the modes were equal, 0 if they were different.
When the mode was not available, but pairs of founders and partners involved in a certain investment shared at least one common value, then the distance was linked to the ratio between the number of shared subjects and the maximum number of shared subjects in the sample:

$$distance_i = 1 - \frac{(number\ of\ shared\ subjects)_i}{(number\ of\ shared\ subjects)}$$

where the subscript is referred to the i-th investment in the sample.
In all the other cases, the distance was set to the maximum value (1) because of lack of commonalities between founders and partners on this particular aspect.

5.3 Weights identification

Three specifications of the similarity score were identified by varying the weights applied to each variable in the formula.

In the *base specification*, each differential was given the same weight, so that the similarity score was computed as a simple average.

In the *background-based specification,* a higher weight was given to the variables relating to founders' and partners' academic and professional background.

Finally, in the *ethnicity-based specification*, a higher weight was attributed to gender, age and nationality. Moreover, Δed, $\Delta fstu$ and $\Delta sstu$ received a lower weight so as to cushion the partial overlapping of information among them.
Notably, the third specification gives greater importance to variables which, according to the rational economic theory, shall not be taken into account when making an investment decision. Nevertheless, as Chapter 6 shows, this version of the similarity score produces interesting results in terms of distributional features (at least for

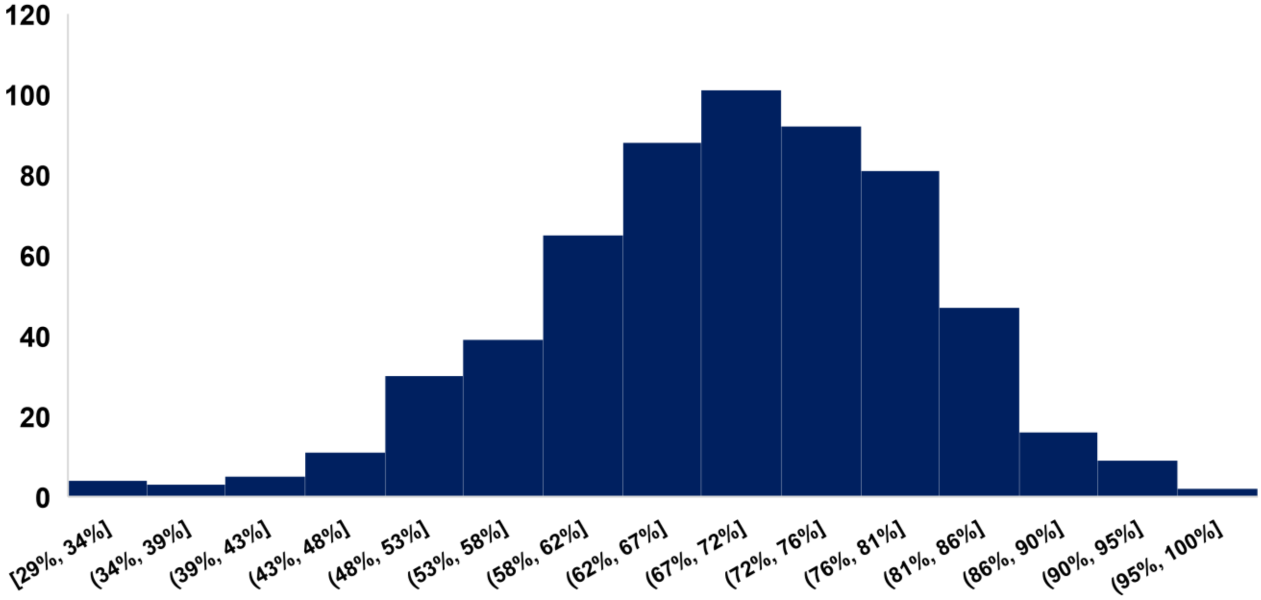
Sample 1). Moreover, as highlighted in Chapter 7, it presents a non-negligible correlation with AMCs' performance (as measured by average valuation step-up) and the experience of AMCs' partners (as measured by their age at deal date and by the number of investments they made). Finally, notwithstanding all the caveats related to small sample size, the score also appears somewhat correlated to the experience of CVCs' partners (this time, only if measured by their age at deal date).

6. Empirical results

6.1 Base specification – Sample 1

Figure 20 shows the probability distribution function (pdf) of the similarity score across the whole Sample 1 under the base specification.

Figure 20. Sample 1 Similarity Score distribution - base specification



The similarity score is bounded between 0 and 1, so that its pdf cannot be properly defined as Normal. However, the distribution depicted in the graph resembles a Gaussian, and this intuition is confirmed by the Normal Q-Q plot presented in Graph I of Annex 3.

The similarity score generally takes quite high values and is characterized by a limited variability: data range from 0.29 to almost 1.00, with an average of 0.69 and a standard deviation of 0.11. Moreover, almost 95% of observations is above the 0.50 threshold. The distribution presents a light negative asymmetry (-0.39) and is slightly concentrated around the mean, as proven by its small positive excess kurtosis (0.24).

Useful insights on the behaviour of the similarity score can be obtained by looking at the distribution for individual AMCs. A complete set of summary stats for each AMC under the base specification is provided in Table I of Annex 3.

As a first observation, the distribution of all AMCs is rather concentrated around the mean, with standard deviations ranging from 0.07 to 0.10.

More heterogeneous results are found for the difference between the maximum and the minimum: considering the AMCs with at least 30 investments[17], the value oscillates between 0.31 (for Vertis) and 0.56 (for Xyence). Additionally, the difference increases with the number of investments, which is a plausible outcome (as the sample size rises, so does the probability of finding outliers in the distribution).

Among the AMCs with at least 30 deals, Innogest Capital is the one with the highest average similarity score (0.75), while CDP Venture Capital is the one with the lowest (0.54, i.e. 0.15 less than the average across the whole Sample 1). To check whether this difference is statistically significant, the classical t-test for mean difference was applied[18]: the extremely small p-value (1.4E-18) signals that the impact of affinity bias on the two AMCs can be considered statistically different.

Interestingly, for the three funds with a strong specialization in Healthcare & Biotech (i.e. Innogest Capital, Panakès Partners and Xyence) the probability mass is shifted to the right when compared to the distribution of the overall Sample 1[19]. As this result persists also in the other two specifications of the similarity score, this could suggest that the affinity bias is stronger when investments are concentrated on specific sectors.

Furthermore, it is useful to focus on the historical evolution of the similarity score, which is shown in *figure 21*.

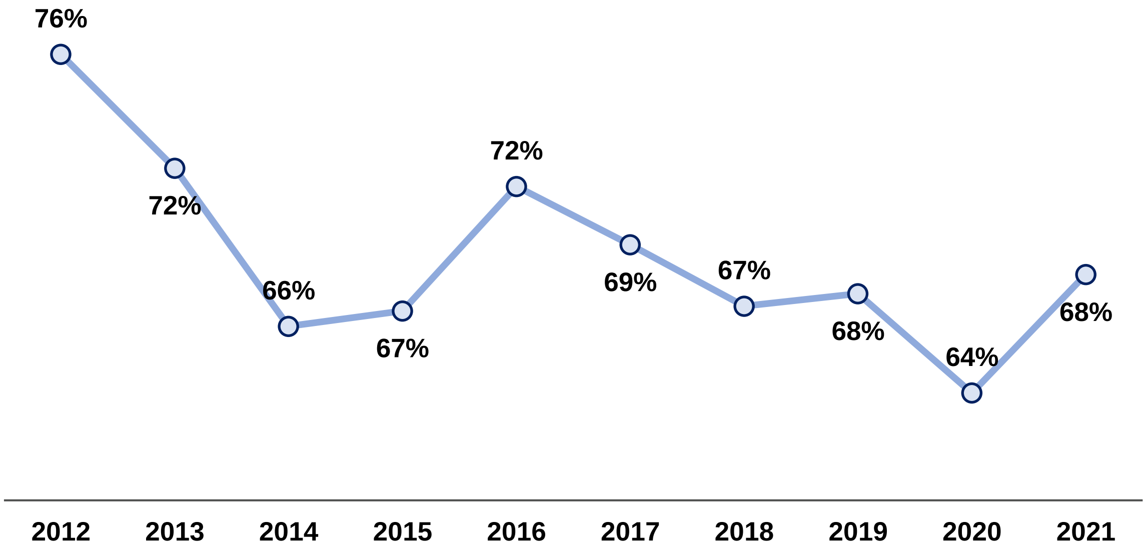
---

[17] As previously explained, the choice to impose a cutoff to the investments is made to avoid the risk of formulating statistically meaningless conclusions.

[18] As said before, the distribution of the similarity score under the base case is quite close to a Gaussian. Under these conditions, the use of the standard t-test to check for statistically relevant differences in means is a safe approach.

[19] For Panakès Partners, however, there are only 12 observations, so that this result is not particularly solid from a statistical standpoint and can well depend on the small sample size.

Figure 21. Average Sample 1 Similarity Score historical evolution - base specification



Note: *years from 2000 to 2012 are not displayed due to the smallness of the sample.*

As said before, the Italian market is still far from being a mature environment. Nonetheless, the last years have seen considerable growth in both the number of deals executed and the amounts financed. Within this framework, the VC funds' activity has intensified, which may have resulted in partners being more experienced and competent in investment decision-making. Therefore, a sensible measure of the similarity score shall capture this effect and show a decreasing tendency over time. *De facto*, this is exactly what happens: the average value switches from 0.76 in 2012 to 0.68 in 2021, and this change is statistically significant at any confidence (p-value 6E-4)[20].

The similarity score shall also decrease when comparing earlier rounds with later financings. Indeed, at Pre-seed and Seed stage, the founding team plays a key role in investment decisions because there is still a relative lack of hard metrics to analyse (e.g. revenue evolution, number of customers). However, as a company grows and its business expands, partners focus more on quantitative aspects before choosing to invest, so that they shall be less exposed to the affinity bias. The decreasing dynamics

[20] The reason why the analysis starts from 2012 is that this is the first year of Sample 1 with at least 25 investments made. In this way, the risk of making conclusions based on too few observations is cushioned.

of the similarity score across funding series is observed in the data: the average is 0.71 for Pre-seed and 0.66 for Series B, and this difference is statistically significant at 5% and 10% confidence levels (p-value 0.011).

Table II of Annex 3 provides an exhaustive set of summary stats for each funding series under the base specification.

Additionally, it worth looking at how the similarity score changes according to the three investment status defined in paragraph 3.2.5. The averages range from 0.68 (for write-offs) to 0.70 (for active investments), thereby resulting really close to each other: in confirmation of this, the use of the classical t-test reveals that pairwise mean differences are not statistically significant.

That said, a natural point of enquiry that follows from this discussion concerns the variations of the similarity score across the three exit clusters identified in paragraph 3.2.5. Specifically, because the dataset includes only 10 IPOs, it is safer to concentrate just on M&A and Secondary Purchases. The average score is higher for Secondary Purchases (0.74) than for M&A (0.69), and this difference is statistically significant at any confidence level (p-value 0.043).

Furthermore, the average similarity score under the base specification is scarcely sensitive to changes in round structure. As for the investors' geography, the rounds with at least one international investor were compared to those with only national players. With respect to the investors' structure, syndicated rounds were contrasted to those where the AMC was the sole investor. Lastly, for what concerns the AMC's role in the transaction, the deals where it acted as lead investor were contrasted to those where it was a follower. In all three cases, the null hypothesis of equal means was not rejected at any confidence level (p-values respectively 0.145, 0.914 and 0.214).

Nevertheless, as paragraphs 6.2 and 6.3, the results of the tests change when other specifications of the similarity score are used. This indicates, at least partially, the importance of specific variables entering the score calculation.

A final analysis concerns the distribution of the similarity score across the 9 sectors identified in paragraph 3.1.3, whose summary stats are presented in Table III of Annex 3. As a preliminary note, since all sectors in the sample have a satisfactory number of observations, there is no need to exclude some of them from the general conclusions. That said, the similarity score displays quite little variability when segmented according to this criterion: the biggest average (0.72 for FinTech) and the smallest one (0.63 for Education & HR) are separated by less than 10 percentage points, albeit this difference is statistically significant at any confidence level (p-value 0.002). Furthermore, the standard deviation displays quite low values for all sectors, oscillating between 0.10 (for 3 sectors) and 0.13 (for 2 sectors).

Moreover, in 7 out of the 9 cases the similarity score distribution is negatively asymmetric, with the index taking the smallest value for Digital (-1.20). Less unanimous evidence is found for the excess kurtosis: the pdf is platykurtic for 5 sectors (with a minimum of -0.51 for Healthcare & Biotech) and leptokurtic for the remaining 4 (with a maximum of more than 2.00 for Smart City).

In summary, the study of the similarity score pdf results more informative when segmenting data by AMC than when using sectors. This outcome is quite intuitive: discrepancies in similarity scores should arise from heterogeneity of personalities and approaches of different partners' groups, something which can be captured only dividing the sample by AMC.

Furthermore, interesting suggestions come from the analysis of data sorted by investment date and funding series, while the similarity score appears almost invariant to variations in the round structure.

Finally, the similarity score does not sensibly change across investment status, but when focusing only on exits, relevant differences among the various clusters emerge.

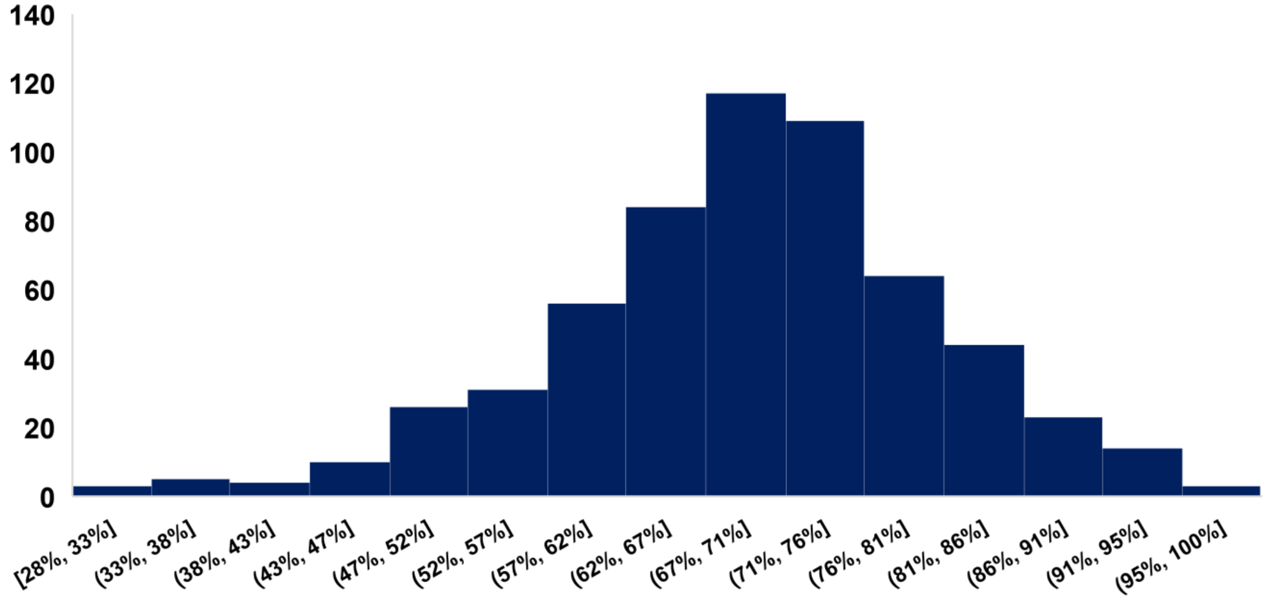
The next two paragraphs describe the other two specifications of the similarity score. Rather than replicating the same analysis made for the base version, it appears convenient to focus on the differences with this latter.

6.2 Background-based specification – Sample 1

Among the variables entering the computation of the similarity score, founders' education and professional background are the only two which shall be explicitly considered by partners when evaluating a potential investment. Therefore, an interesting extension of the analysis consists in studying how the similarity score distribution changes when a higher weight is given to differences in previous professional experience (Δexp) and field of study ($\Delta fstu$).

This gives rise to the background-based specification of the similarity score, whose pdf is shown in *figure 22*.

Figure 22. Sample 1 Similarity Score distribution - background-based specification



Moments up to the third are almost unaltered when compared to the base specification[21]. A more perceivable change concerns the excess kurtosis, which more than doubles (from 0.24 to 0.59). This suggests that the distribution departs from the Gaussian more than before, even if it still looks quite similar to it. These intuitions are confirmed by the Normal Q-Q plot shown in Graph II of Annex 3: the empirical and theoretical quantiles are almost identical in the central part of the pdf, while they diverge in the tails (especially in the left one).

Table IV of Annex 3 provides a complete set of summary stats at AMC level. The average similarity score remains extremely similar to the base case, while asymmetry and excess kurtosis display a more variable behaviour. In some cases (e.g. Eureka!

---

[21] For the first moment, the classical t-test for mean difference was performed. The related p-value (0.42) led not to reject the null of equal means between the two specifications at any confidence level.
For the second moment, the classical F-test was run. The resulting p-value (0.40) led not to reject the null hypothesis of equal variances at any confidence level.

Ventures), this can be linked to the restricted sample size, but in others it signals the impact of background-related variables on the similarity score. In this regard, an interesting example is provided by United Ventures: the similarity score distribution switches from being positively asymmetric (index 0.54) and slightly platykurtic (-0.20) to be left-skewed (-0.35) and leptokurtic (0.63).

On a general basis[22], the study at AMC level leads to conclude that, even if previous professional experience and field of study do not shift the distribution mean and standard deviation, they do alter the way in which the probability mass concentrates around the center and the extreme values.

As for the data breakdown by funding series, sector and year, the background-based specification produces the same logical outcomes as the base version, which supports the plausibility of its construction.

A last comment shall be made on the behaviour of the similarity score according to the round structure. Compared to the base specification, there are no substantial variations for what concerns the investors' structure (syndicated vs. non-syndicated rounds) and the AMC's role (lead vs. follower). However, the situation changes when investors' geography is analysed: the rounds with at least one international investor present a higher average similarity score than those with only national players, and the difference is statistically significant at 10% confidence level.

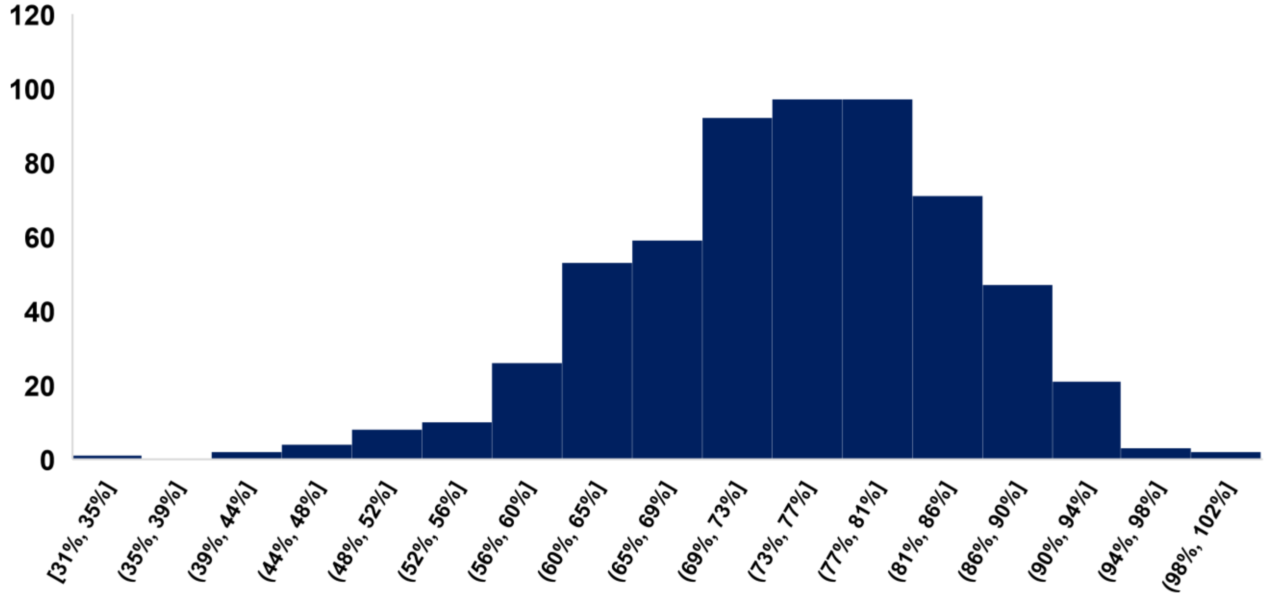
6.3 Ethnicity-based specification – Sample 1

The third and last specification of the similarity score was named *ethnicity-based* because a higher weight was attributed to the terms capturing gender and nationality differences among founders and partners. This responds to the need of verifying the degree to which the variables connected to the most irrational component of decision-making influence AMCs' investments.

[22] For obvious reasons, this conclusion applies only to the AMCs with statistically significant sample sizes.

However, before moving to the analysis at AMC level, it is useful to briefly describe the general distribution of data. To this purpose, *figure 23* displays the pdf of the similarity score under this specification across the whole sample.

Figure 23. Sample 1 Similarity Score distribution - ethnicity-based specification



As a first observation, apart from a slight negative asymmetry, the distribution appears quite close to a Normal. This intuition is confirmed by the Q-Q plot shown in Graph III of Annex 3.

Moreover, the general distribution is more concentrated around the mean and slightly shifted to the right than in the base case.

More in detail, the average similarity score increases by 4 percentage points when compared to the base specification, and this change is statistically significant, as confirmed by the t-test for mean difference (p-value 8.07E-19). Conversely, the standard deviation slightly decreases (from 0.13 to 0.10), and this negative shift is statistically significant[23].

[23] In order to check for changes of the second moment, the classical F-test was performed. The resulting p-value (0.001) led to reject the null of equal variances between the two specifications at any confidence level.

Similar to what happened in the second specification, asymmetry modestly reduces, while excess kurtosis rises considerably (even if less than in the background-based case).

A more interesting discussion concerns how the similarity score varies at AMC level with respect to the base specification. An exhaustive set of summary stats is provided in Table V of Annex 3.

Among the AMCs with at least 30 observations, CDP Venture Capital presents the highest change in the average similarity score (+0.11), but nevertheless it remains the AMC with the lowest value in the sample (0.65). However, data at AMC level result now more concentrated: the distance between the highest and the lowest average shrinks to 0.16, i.e. 5 percentage points less than in the other versions of the similarity score.

As for the other moments, standard deviations do not vary, while asymmetry and kurtosis change in different ways based on the AMC. Notably, United Ventures confirms to be the AMC with the most unstable third and fourth moments: with respect to the base case, the pdf becomes markedly left-skewed (-1.65) and extremely leptokurtic (4.20). This outcome may be in part influenced by the limited number of observations available for this AMC (34).

As a final notice, the similarity score produces interesting results when segmented by investors' geography. Indeed, rounds without international investors display a higher average score (the t-test for mean difference is rejected at 10% confidence level). Notably, this result is opposite to that observed in the background-based specification. Considering how the two variants of the score are computed, it could be inferred (at least partially) that foreign investors' affinity considerations are more sensitive to founders' ability-related variables (previous educational and professional experience) than to their ethnicity. However, this observation must be taken with extreme caution, since this paper only maps a very limited portion of the investments made by certain international players, without providing an exhaustive analysis of their full activity.

To summarize, apart from small-sample-induced effects, giving more importance to the irrational components of the similarity score leads to a general increase in similarity score levels, but the various distributions remain minimally dispersed.

As for the data breakdown by funding series, sector and year, the ethnicity-based specification produces the same intuitive results as the base version, which supports the plausibility of its construction.

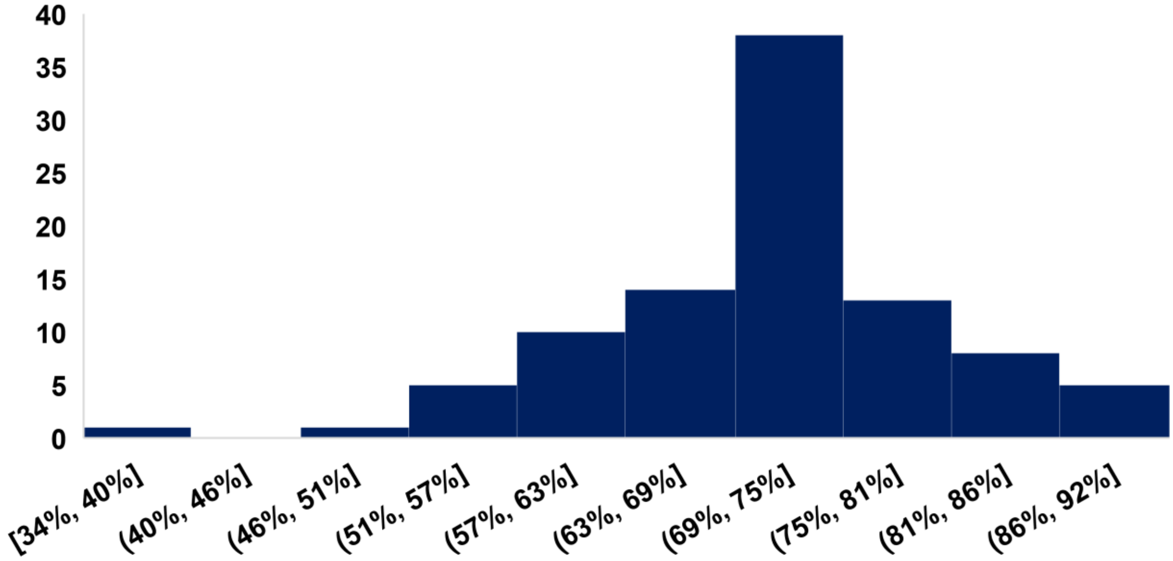
6.4 Base specification – Sample 2

The analysis of the similarity score distribution for Sample 2 reveals interesting insights behind CVCs' investment decisions.

As an introductory note (which applies to all three specifications of the score), Sample 2 has far less observations (95) than Sample 1 (493), so that it is difficult to reach statistically significant conclusions when segmenting data (e.g. by CVC or sector). Thus, the results presented in the next three paragraphs shall be taken with due attention.

Figure 24 shows the similarity score pdf across the full Sample 2 under the base specification.

Figure 24. Sample 2 Similarity Score distribution - base specification



As already pointed out in paragraph 6.1, the similarity score distribution cannot be properly defined as Normal. Nevertheless, *figure 24* suggests some affinities with the

Gaussian, an intuition confirmed by the Q-Q plot provided in Graph I of Annex 4 and by the analysis of the distribution moments.

Specifically, as with Sample 1, the score takes quite high values: the average is 0.71 and only 1 data point is below the 0.50 threshold. It is worth underlining that the mean difference between Sample 1 and 2 is not statistically significant at any confidence level (p-value 0.15): this outcome persists also in the other two score versions (p-values 0.92 and 0.63 respectively) and suggests that the affinity bias hits professional VC investors regardless of their type.

Moreover, under the base specification, Sample 2 distribution presents low variability (standard deviation 0.10) and a slightly negative asymmetry (-0.46).

The most relevant difference with Sample 1 concerns the excessive concentration around the mean, clearly visible in *figure 24.* This feature, which is also the main reason for departure from the Normal case, is captured by the positive excess kurtosis (1.66), far higher than the figures registered in Sample 1 for any score specification.

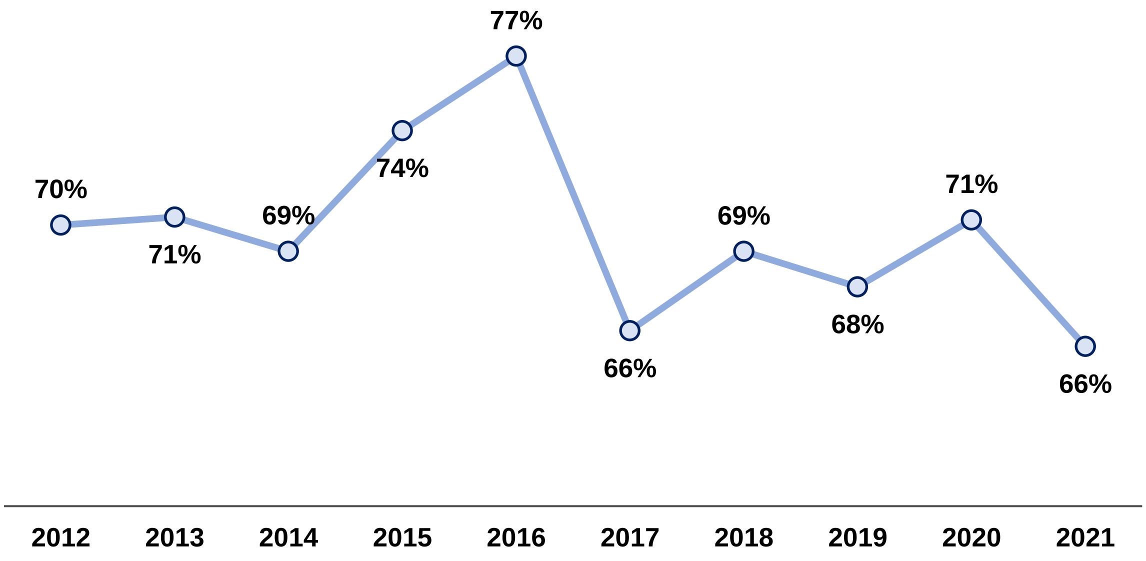
When it comes to data at CVC level, results have limited usefulness since the 7 sub-samples are quite small in size (only for ZernikeMeta Ventures there are more than 25 observations). A complete set of summary stats for each CVC in Sample 2 under the base specification is furnished in Table I of Annex 4.

As a general consideration, almost all distributions display little variability and a small range: standard deviations are quite homogeneous (spanning from 0.05 for ZernikeMeta Ventures to 0.14 for Zanichelli) and the difference between the minimum and the maximum is lower than 0.30 in 5 out of 7 cases. The result concerning data ranges may be explained by the small sample sizes, as it reasonable to expect that the probability to find outliers increases with the number of investments.

Healthware is the CVC with the highest average similarity score (0.80), while only Mediaset and Zanichelli have values below 0.70 (respectively, 0.64 and 0.67). The difference between the two CVCs at the opposite ends (0.16) is statistically significant at any confidence level (p-value 2.7E-04). As for the other distribution moments, asymmetry and kurtosis display wide differences among CVCs.

Looking at the historical evolution of the similarity score, Sample 2 do not confirm the results observed for Sample 1, even if the small sample size does not allow statistically meaning conclusions. In detail, the average score does display a slow tendency to decrease from 2012 to 2021, but the change is negligible (0.04) and not statistically significant (p-value 0.32).

Figure 25. Average Sample 2 Similarity Score historical evolution - base specification



***Note**: investments are cumulated over time. Years from 2000 to 2011 are not displayed due to the smallness of the sample.*

For what concerns the relationship between the similarity score and the funding series, the only sound comparison that can be made involves Seed and Series A (for the other funding series, sample size was too small). The difference between the average score for Seed (0.68) and Series A (0.74) is statistically relevant at 5% (p-value 0.02): this outcome is coherent with what emerged in Sample 1 and, as outlined in paragraph 6.1, makes perfect sense.

Table II of Annex 4 provides more details on the similarity score distributions for the two aforementioned funding series.

Furthermore, the similarity score does not seem to be influenced either by the investment status or by the exit cluster.

As for the former variable, the only sound t-test which can be made involves active and exited investments[24]: the difference between the two averages is slightly higher than 1% and this value is not statistically significant (p-value 0.50).

Focusing on the exited investments, it is worth recalling that Sample 2 includes only 1 IPO, so the t-test was performed between M&A and Secondary Purchases: the mean difference is less than 1.5% and is not statistically significant (p-value 0.52).

Interesting results are reached by looking at how the similarity score varies according to the round structure. As with Sample 1, the investors' structure and the CVC's role in the deal do not seem to influence the similarity score. As regards the former variable, the null hypothesis of equal means is not rejected at any confidence level when comparing syndicated rounds with those where the CVCs were the sole investors (p-value 0.56). For what concerns the latter variable, average similarity scores are not statistically different depending on whether the CVCs acted as lead investors or followers (p-value 0.14).

However, in contrast with Sample 1, investors' geography does impact the similarity score: rounds with only national investors have, on average, higher similarity scores than transactions with at least one foreign player, and this difference is statistically significant at 5% (p-value 0.04). Notably, this analysis is among the few in Sample 2 relying on sufficiently large samples (respectively, 55 and 36), which makes the outcome statistically sound.

The final section relates the variations of the similarity score according to the sectors. In this respect, a complete set of summary stats is provided in Table III of Annex 4. Unfortunately, samples are quite fragmented when sorting data according to this criterion (all below 25), so that it is rather difficult to reach statistically meaningful conclusions. As with Sample 1, distributions display quite little variability, with Fintech reaching the highest average score (0.76) and Food & Agriculture the lowest (0.66). The two data are less than 10 percentage points apart, and this difference is statistically relevant only at 10% (p-value 0.09).

[24] For write-offs, sample size is too small to formulate statistically significant conclusions.

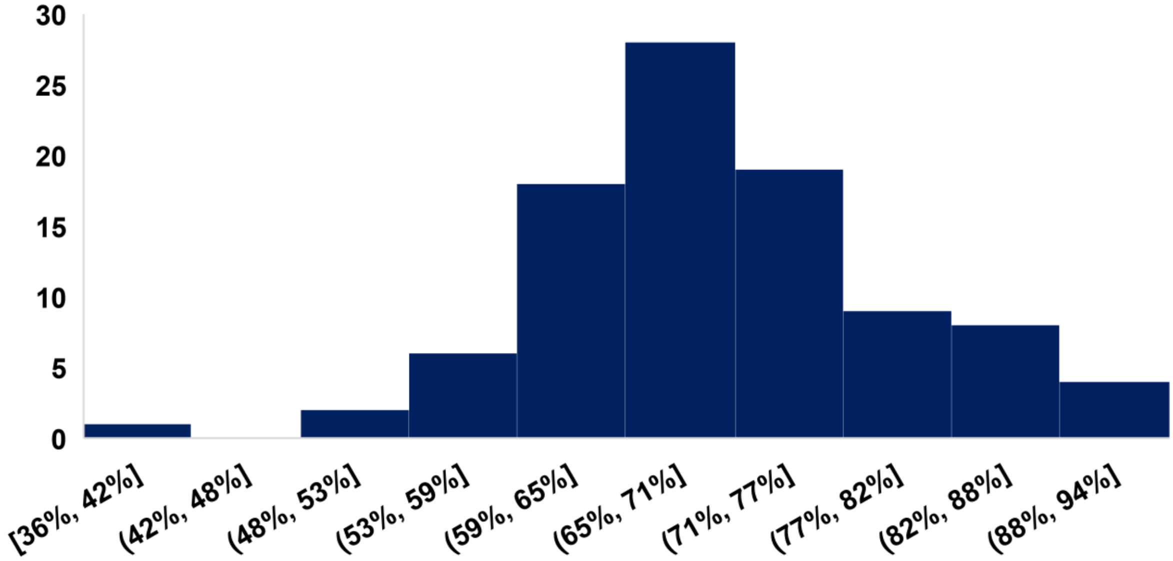
Additionally, the second moment takes small values for any sector, always remaining below 0.15. This suggests that all distributions present limited variability.
To sum up, Sample 2 confirms the (intuitive) idea the similarity score is more sensitive to changes in the investing subject that in sector.

In other respects, small sample sizes make it difficult to reach statistically meaningful conclusions. Nonetheless, the similarity score does seem to change according to the funding series and the investors' geography.
As it happened for Sample 1, the next two paragraphs are dedicated to the other two specifications of the similarity score, with emphasis given to the differences with the base case.

6.5 Background-based specification – Sample 2

Figure 26 displays the pdf of the Sample 2 similarity score under the background-based specification.

Figure 26. Sample 2 Similarity Score distribution - background-based specification



As it can be seen, the average is almost unaltered with respect to the base case. This happens also for the standard deviation, so that data preserve approximately the same level of variability.
Furthermore, the distribution is almost perfectly symmetric, with the third moment dropping from -0.46 (base specification) to -0.07. There still remains a probability

concentration around the mean, but it appears less pronounced than in the base case: in this regard, the excess kurtosis decreases to 1.17[25].

All in all, these observations lead to conclude that the pdf of the background-based similarity score presents reasonable features and is even closer to a Normal case than the base one. This is confirmed by the Q-Q plot shown in Graph 2 of Annex 4.

Moving to the analysis at CVC level, a complete set of summary stats is provided in Table IV of Annex 4. As a first consideration, the dynamics are strongly influenced by the sample fragmentation, which makes it difficult to formulate solid economic conclusions.

From a purely descriptive standpoint, the first two moments display minor variations, while asymmetry and kurtosis tend to change more. In particular, the third moment moves in different directions based on the CVC (e.g. it decreases for ZernikeMeta Ventures and rises for Tim), while the fourth one increases in 6 out of 7 cases.

Thus, as with Sample 1, the study at CVC level leads to conclude that giving more weight to previous professional experience and field of study does not alter the distribution mean and standard deviation, but modifies the probability concentration around the center and the extreme values.

Additionally, almost all distributions are characterized by small ranges: exception made for Zanichelli, the difference between the maximum and the minimum oscillates between 0.14 and 0.32, in line with what displayed in Sample 1.

Finally, switching to the background-based specification does not alter the hierarchy among CVCs, with Healthware still showing the highest average score (0.85) and Mediaset the lowest (0.66).

As regards the funding series, the background-based similarity score displays the same outcome as the base case: Seed rounds have a higher average similarity score than Series A rounds (0.73 vs. 0.66) and this difference is statistically significant at 5% (p-value 1.32E-02).

[25] Notwithstanding this drop, the value for the excessive kurtosis is still well above the figures registered in Sample 1. This may depend, at least partially, on the significant differences in samples sizes.

As for the sectoral segmentation, the same observations made for the base case apply. In particular, data preserve little variability (as shown by the low standard deviations), while the average score seem scarcely influenced by the sector: even worse than in base case, the mean difference between the sector with the highest score (Fintech, 0.75) and the one with the lowest (Media, 0.67) is not statistically significant even at 10% (p-value 0.13).

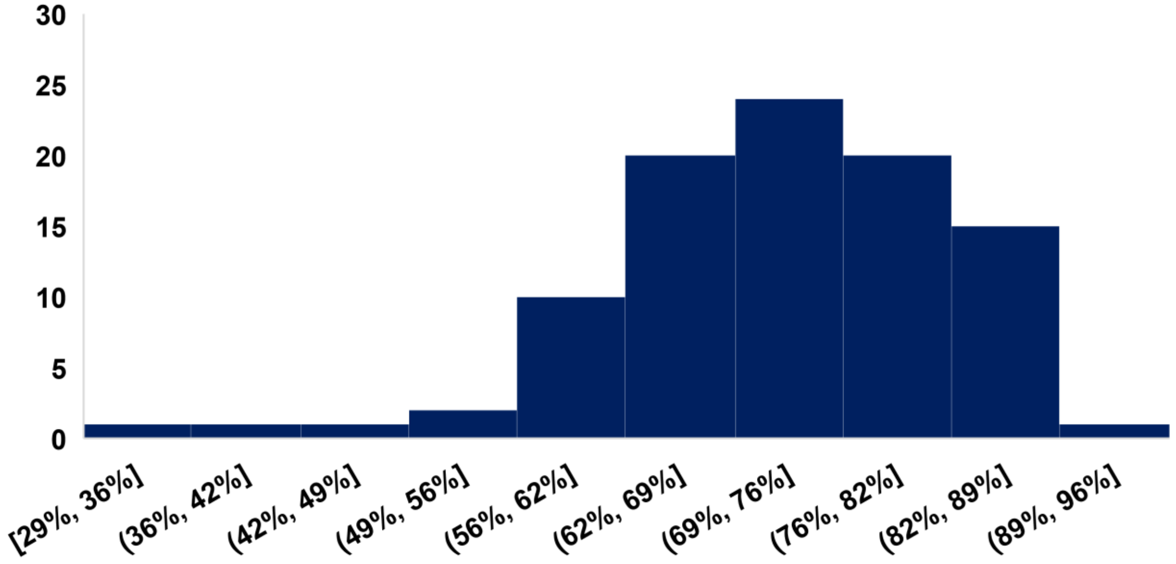
In contrast, investors' geography becomes not statistically relevant under the background-based assumption: from a statistical standpoint, the average similarity score of rounds with only Italian investors is not different from that of rounds with at least one foreign player (p-value 0.81).

The other variables of interest (time, investment status, exit cluster, round structure and CVC's role) are not statistically related to the similarity score under the background-based hypothesis.

6.6 Ethnicity-based specification – Sample 2

The pdf of the Sample 2 similarity score under the ethnicity-based case is shown in *figure 27*.

Figure 27. Sample 2 Similarity Score distribution - ethnicity-based specification



Visibly, the distribution departs more from the Gaussian, and this intuition is corroborated by both the Q-Q plot (see Graph III of Annex 4) and the analysis of moments. More specifically, the graphical analysis shows marked differences between empirical and theoretical quantiles in the distribution tails. Not by chance, excess kurtosis is well above the values registered for the other specifications (2.36). Furthermore, the asymmetry index becomes markedly negative (-1.09), so that the distribution appears rather skewed.

Nevertheless, giving more importance to the most irrational variables in the score does not produce a shift of the pdf average (which remains close to 0.70) and standard deviation (which changes by 0.01 when compared to the base case).

This leads to conclude that, similar to what happened in the second specification, a change in the score weights only alters the distribution symmetry and the way in which the probability mass concentrates around the center and the tails.

As far as the study at CVC level is concerned, the ethnicity-based specification causes an increase in the difference between the highest and lowest average score registered in the sample, which becomes statistically significant at any confidence level (p-value 1.88E-03)[26].

Moreover, the second moment remains almost unaltered across all CVCs, so that data preserve the same level of variability as in the base case. A more diversified behaviour is seen for asymmetry and excess kurtosis[27].

A comprehensive set of summary stats for each CVC under the ethnicity-based specification is furnished in Table V of Annex 4.

Interestingly, giving more emphasis to previous professional and educational experiences induces changes in the relationship between the similarity score and some of the variables selected in this paper.

Firstly, the rounds where the CVC was the sole investor present a significantly higher influence of the affinity bias than syndicated transactions (p-value 0.03).

This result appears reasonable: as the number of investors increase, so does the probability to have more diverse backgrounds and standpoints, which may ultimately

---

[26] Healthware remains the CVC with the highest average score in Sample 2 (0.80), while Barilla becomes the one with the lowest (0.55).

[27] Given the small sample sizes, it appears convenient to focus on the general dynamics instead of describing granular changes for each CVC.

cushion the impact of the affinity bias. It should be noticed, however, that Sample 2 includes few transactions with just one investor (17), so that the outcome may depend (at least in part) on small-sample-induced effects.

Secondly, the CVC's role in the deal becomes statistically relevant: the average similarity score is slightly higher for those rounds where the CVCs acted as lead investors (p-value 0.02).

Conversely, the funding series does not seem to influence the similarity score under the ethnicity-based specification: Seed rounds continue to display higher influence of the affinity bias than Series A rounds (average score 0.74 vs. 0.69), but the difference shrinks and turns into non-statistically relevant (p-value 0.12).

For all the other variables, the third version of the score produces the same outcomes as the base case.

7. Relations with partners' and funds' features

The last part of the analysis verifies whether the similarity score can be related to specific partners' and funds' features[28].

7.1 Similarity score and partners' features – Sample 1

The first step of the analysis verifies whether differences in similarity scores across AMCs can be related to selected partners' characteristics.

7.1.1 Partners' experience

As already noticed, sample data are quite heterogeneous when it comes to partners' age. Thus, an interesting question concerns the potential link between the similarity score of each investment and the average age of partners participating in it. Intuitively, older partners could be less influenced by the affinity bias because of their longer professional experience, which shall make them less likely to fall prey to irrational decisions. Therefore, a negative correlation between the two variables could be observed.

To check that, the similarity score was regressed on partners' average age at deal closing date. Table I of Annex 5 shows the summary output for the base similarity score[29]. As expected, there is a slightly negative correlation between the two variables, with the regression coefficient equal to -0.003 and statistically significant at any confidence level. However, the R square is quite low (less than 3 percent), which implies that the overall fit is poor.

Similar conclusions are reached when using the other two specifications of the similarity score. The regression coefficient remains statistically significant at any confidence level, while the R square increases up to a more interesting 0.056 for the ethnicity-based similarity score.

[28] As suggested by the sentence, the analysis in paragraph 7.2 was made at fund rather than at AMC level. This choice is motivated by the fact that information of return performance, treated in paragraph 7.2.2, was available only at fund level (and for a restricted number of funds).

[29] A comment on the difference of results when the other specifications of the score are used is provided in the subsequent section of the paragraph.

Another measure of partners' experience is given by the number of investments they took part in. Intuitively, there should be a negative relationship between the average similarity score of the AMCs and the number of transactions that each of them made. In fact, this is exactly what is observed: as Table II of Annex 5 displays, the regression coefficient for the base case is negative (-0.001) and statistically significant at 10% confidence level. Furthermore, the overall regression fit is quite high (R square 0.184).

The result slightly worsens as the background-based specification is considered, even if the regression coefficient continues to be statistically significant at 10% confidence level and the overall fit remains above 0.15.

Conversely, the situation improves with the ethnicity-based specification: the regression coefficient becomes statistically significant at 5% confidence level (but not at 1%) and the R square overcomes 20 percent.

Therefore, all in all data seem to confirm that more experienced partners tend to be less influenced by the affinity bias when making investment decisions.

7.1.2 Partners' diversity

Studying the impact of the affinity bias *inside* the AMCs, Gompers & Wang (2017) show that, when existing partners have more daughters, they are more likely to hire a female investor partner. Given the traditionally low presence of women in the VC industry, the bias can have a positive impact on the organization diversity. Ultimately, the increased heterogeneity of partners' groups enables them to attract a much wider deal flow and improve the average deal quality.

In other words, the affinity bias can exert a positive action on gender *inside* an AMC, and this may cushion the negative effect that it has on investment decisions (i.e. *outside* the organization). In order to verify this intuition, the average similarity score of each fund was regressed against the average ratio of female partners who took part in the fund's deals. The full regression output for the base specification is presented in Table III of Annex 5.

The results obtained confirm the theoretical intuition: the average similarity score decreases as the proportion of female partners in a VC fund increases. The overall regression fit is quite interesting (R square higher than 0.25), and the regression coefficient is statistically significant at any confidence level. Furthermore, the relation between the two variables is robust in absolute terms: a 10% raise in the percentage of female partners is associated with a 1.6% reduction of the average similarity score. This implies that switching from a male-centric organization (0% female ratio) to a more balance structure (50% female ratio) can cushion the negative influence of affinity bias on investments by almost 10 percentage points.

Similar observations can be made for the other two specifications of the similarity score. Notably, in the ethnicity-based case the overall regression fit drops to 0.12 and the regression coefficient halves, but this is a totally expected outcome: indeed, when compared to the other two specifications, the ethnicity-based version gives a lower weight to differences in gender, to which this part of the analysis is maximally sensitive.

7.2 Similarity score and funds' features – Sample 1

The following step of the analysis checks whether the similarity score depends on some relevant funds' features.

7.2.1 Fund size

The first relationship being tested is with funds' size, as measured by their target size. Ideally, bigger funds are more likely to participate in larger rounds, where more attention is devoted to hard metrics (e.g. revenue growth, client base expansion) rather than to "soft elements" (including the founding team). Moreover, they have a higher probability to act as lead investors, since they have the financial and human resources to oversee the round's progression: clearly, this means that they perform deeper analyses on potential investments. In light of these considerations, a negative relationship could be observed between fund size and average similarity score.

To corroborate this idea, the fund average similarity score was regressed on its target dimension. Table IV of Annex 5 displays the full summary output for the base specification. There does not seem to exist a meaningful linear relation between the two variables: the regression coefficient (which is actually positive) is not statistically significant (p-value 0.502) and the overall fit is very low (R square 0.01).

The situation does not improve with the other versions of the similarity score: in both cases, the regression coefficient is far from being of any relevance and the R square remains extremely small.

In conclusion, the theoretical intuition about the relationship between affinity bias and fund dimension is not confirmed by the evidence found in the data.

7.2.2 Fund performance

Chapter 6 has highlighted that Italian AMCs' investment decisions are somewhat influenced by affinity considerations. Indeed, the similarity score reaches high values (typically more than 0.60) in any sample breakdown considered, peaking at almost 1 in certain investments.

The question that naturally follows from this observation is whether and how the affinity bias impacts fund performance. Theoretically, it should induce suboptimal asset allocations, eventually leading to poorer returns. Therefore, the average similarity score of a fund should be negatively correlated with its IRR.

Unfortunately, data on funds' IRR were publicly available only in 5 cases, so that it was impossible to reach statistically solid conclusions. Anyway, it is worth mentioning that the worst performing funds in terms of IRR were those with the highest average similarity score, while the best performing ones showed the lowest sensitivity to the affinity bias.

An alternative (albeit less precise) approach to measure the performance of an investment consists in computing the investee's valuation step-up. This is obtained by taking the ratio between the latest available pre-money valuation of the company and its post-money valuation on the occasion of the fund's first investment. Clearly, if a start-up grows and its business expands, its value is likely to increase, which will cause the valuation step-up to rise; on the other side, poor market performance will lower the valuation, driving down the ratio.

It is important to recall that the valuation step-up is an approximate measure of the investment return, since it presents a number of deficiencies when compared to the IRR. Firstly, being a cash-on-cash multiple, it does not take into account the time value of money. Secondly, it just compares the latest valuation of the company with the post-money valuation when the fund invested for the first time. Therefore, it does not consider potential follow-on investments. Thirdly, it neglects the operating costs, which normally erode the fund performance. Finally, it is invariant to the company's percentage stake acquired by the fund.

That said, data on companies' pre- and post-money valuations were obtained from PitchBook and Zephyr – when publicly disclosed. Clearly, in case the fund entry coincided with the latest round made by the company, the valuation step-up was given the value of 1.

Then, the fund average valuation step-up was regressed on fund average similarity score. The results of the regression are displayed in Table V of Annex 5. The correlation between the two variables seems quite weak: as expected, the regression coefficient is negative (-12.69), but it is not statistically significant at any confidence level; moreover, the R square is slightly higher than 5 percent, which indicates an overall bad fit.

When the background-based similarity score is used, the results are even worse, with the regression coefficient remaining statistically non-significant (p-value 0.583) and the R square decreasing to a modest 0.024.

Interestingly, the scenario improves with the use of the ethnicity-based version of the score. Indeed, even if the regression coefficient is still not significantly different from 0 (p-value 0.166), the overall fit increases to 0.142.

Another (and even more approximate) way to check whether the similarity score is connected to fund performance consists in comparing the average scores of the best and the worst performing AMC[30]. If affinity bias negatively impacts performance, the average similarity score of the best performing AMC should be lower than the one of the worst performing AMC, and the difference between the two should be statistically significant.

[30] This is the only analysis of paragraph 7.2 that is done at AMC rather than at fund level. This choice is motivated by the willingness to avoid excessive sample fragmentation, which would have made the results of the test difficult to interpret.

This paper uses a “quantitative” and a “qualitative” criterion to identify the best and the worst AMC in the sample: the former defines the best (worst) performing AMC as the one with the highest (lowest) average valuation step-up, while the latter concentrates on the number of exits (for the best AMC) and write-offs (for the worst one).

Evidently, both methods present some limitations: the quantitative criterion is impacted by the lack of data on companies’ pre-money valuations, which makes it difficult to compute valuation step-ups; conversely, applying the qualitative criterion exposes to mistakes when identifying the best AMC (since not all exits are successful) and the worst one (since data on write-offs are often opaque).

That said, 360 Capital Partners results the best performing AMC according to both criteria. A divergent result is obtained for the worst performing AMC, with Indaco Venture Partners (Xyence) being selected by the quantitative (qualitative) criterion.

Firstly, let us consider the quantitative criterion. As expected, the average similarity score of 360 Capital Partners (best performing AMC) is lower than that of Indaco Venture Partners (worst performing AMC). When the base and the background-based specifications are used, the difference between the two means is not statistically significant. However, the situation improves with the ethnicity-based specification: the mean difference becomes statistically significant at any confidence level (p-value 0.003). This result confirms that the ethnicity-based specification is the most sensitive to performance metrics.

Secondly, let us apply the qualitative criterion. As for the previous case, the difference between the best performing AMC (360 Capital Partners) and the worst performing one (Xyence) is negative. This time, though, it is statistically significant regardless of the specification used.

To summarize, fund performance appears to be somewhat dependent on the affinity bias, even if the relationship is not particularly strong. Obviously, the results obtained are strongly influenced by the lack of detailed information on investments, which makes it necessary to adopt rough and approximate measures of fund and AMC performances.

7.3 Similarity score and partners' features – Sample 2

7.3.1 Partners' experience

The same analyses made for Sample 1 AMCs were performed for Sample 2 CVCs. However, since data collection was harder and sample sizes were smaller, the content of the current and next paragraph shall be treated with due attention.

Firstly, the link between the similarity score and CVC partners' experience was investigated by regressing the former on partners' average age at deal closing date. Table I of Annex 6 displays the summary output for the base case. Data show no clear relation between the two variables: the overall fit is poor (R square close to 0), while the regression coefficient is extremely small in absolute value (-4.23E-04) and not statistically significant at any confidence level. A similar outcome is reached in the background-based case.

The scenario changes when the ethnicity-based specification is used: the R square rises up to an interesting 0.13, while the regression beta becomes statistically relevant at any confidence level (p-value 3.94E-04).

Secondly, the average similarity score of each CVC was regressed against the respective number of investments. The relation between the two variables appears rather weak, with the beta coefficient being not statistically significant regardless of the score version. It should be mentioned, however, that the R square varies based on the specification chosen and, in the background-based case, it overcomes 0.38. Logically, such a heterogeneous behaviour can be explained by the exiguous sample size on which the analysis was conducted.

To summarize, the evidence found on CVCs is less strong than that on AMCs, but it is worth recalling the sensible difference in data points between the two samples.

7.3.2 Partners' diversity

There is no significant relationship between the similarity score and partners' diversity, as measured by partners' female ratio. The R square is quite low in any specification of the score, while the regression coefficient is never statistically significant. Nonetheless, given the way in which the variables (especially the partners' female ratio) are computed, the absence of fit appears strongly dependent on the low number of observations available.

7.4 Similarity score and funds' features – Sample 2

The last part of the analysis on CVCs consists in investigating the link between the similarity score and selected funds' characteristics. However, when compared to Sample 1, the scope of research was narrowed to adapt to the limited information available for Sample 2.

More specifically, as for fund size, data was never retrievable, so that it was impossible to perform regressions based on this variable.

For what concerns fund performance, CVCs' IRRs are not publicly disclosed. Moreover, data on Sample 2 pre-money valuations were available only for an exiguous number of investments, which made it rather difficult to compute valuation step-ups. Therefore, it appeared convenient to limit the analysis to the two (approximate) performance criteria cited in the part of paragraph 7.2.2.

From this perspective, Mediaset (Zanichelli) is the best (worst) CVC in terms of average valuation step-up (quantitative criterion). Conversely, when looking at the number of write-offs (qualitative criterion), ZernikeMeta Ventures achieved the poorest results in Sample 2.

The quantitative criterion does not show significant differences in the average score of the best (worst) CVC in Sample 2, regardless of the specification used.

Some progresses are seen when applying the qualitative method: according to both the base and the ethnicity-based version, Mediaset (best performing CVC) has a considerably lower average score than ZernikeMeta Ventures (worst performing CVC). In detail, the mean difference is statistically significant at any confidence level under the base hypothesis (p-value 2.10E-04) and at 5% with the ethnicity-based assumption (p-value 0.04).

In brief (and with all the caveats repeatedly mentioned) Sample 2 analysis mildly confirms the evidence found in Sample 1: the affinity bias seems to have an influence on fund performance, but the link is not particularly strong.

8. Conclusions and implications for future work

This paper studies the impact of the affinity bias on the investment decisions of Italian AMCs and CVCs operating in the VC industry.

The results obtained confirm that unconscious mental processes are a non-negligible component of human actions, even for highly professional individuals.

In Sample 1, the heterogeneity in the bias strength among the AMCs is partially explained by the different degree of partners' experience: those with more transactions performed tend to favour more diversity in investments.

Also, partners' teams with a more balanced gender structure tend to be less affected by the affinity bias. This suggests that increasing the proportion of females in Italian AMCs may have a positive impact of investment selection.

Moreover, the affinity bias seems to have a certain influence on funds' performance: the best performing funds are those with the lowest similarity scores, and this result is robust to the three specifications used in the analysis. Interestingly, the highest correlation is obtained when giving a higher weight to the least rational variables of the score (gender and nationality).

When it comes to CVCs, the small sample size does not allow to formulate the same strong observations made for Sample 1. Nevertheless, data suggest the influence of the affinity bias also for this type of investors and, under the right assumptions, its intuitive link to fund performance.

The conclusions reached by this paper are subject to extensions and improvements brought by future research on the topic.

Firstly, the analysis is limited to Italian AMCs and CVCs, which, as mentioned several times, still operate in a relatively underdeveloped VC market. In this regard, it would be interesting to extend the study to more mature European environments (e.g. the UK, Germany and France) to check whether more structured dynamics can cushion the impact of the affinity bias.

Finally, the similarity score built in this paper directly compares a number of characteristics of partners and founders. Elaborating on Gompers & Wang (2017), the score could be integrated with additional variables measuring the ethnic, educational and professional differences between partners' children and start-ups' founders. This would allow to verify the relevance of another channel through which the affinity bias could operate, namely venture capitalists' tendency to invest in founders who remind them of their children.

References

Azoulay, P. *et al*. (2020). "Age and high-growth entrepreneurship." *American Economic Review: Insights*, 2(1): 65-82

Bazerman, M. H. & Moore, D. A. (2012). "Introduction to Managerial Decision-making." In: M. H. Bazerman & D. A. Moore *Judgement in Managerial Decision-making, 8th ed.* John Wiley & Sons, pp. 1-13

Chhaochharia, V. & Laeven, L. (2009). "The investment allocation of sovereign wealth funds." *Available at SSRN 1262383*

Gage, D. (2012). "The Venture Capital Secret: 3 Out of 4 StartUps Fail". *Wall Street Journal.* < https://www.wsj.com/articles/SB10000872396390443720204578004980476429190> [last access: 28/02/2022]

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

Gompers, P. A. & Wang, S.A. (2017). *And the children shall lead: gender diversity and performance in Venture Capital*. Working Paper 23454. National Bureau of Economic Research

Hunter, W. C. & Walker, M.B. (1995). *The Cultural Affinity Hypothesis and Mortgage Lending Decisions*. Working Paper, n.95-8. Federal Reserve Bank of Chicago

Kahneman, D. & Tversky, A. (1973). "On the psychology of prediction". *Psychological Review*, 80: 237–251.

Kahneman, D. & Tversky, A. (1979). "Prospect theory: An analysis of decision under risk." *Econometrica*, 47(2): 263–291.

Klayman, J., Ha & Y.-w. (1987). "Confirmation, disconfirmation, and information in hypothesis testing". *Psychological Review*, 94(2): 211–228.

Louis, C. (2019). "Unconscious bias: The affinity bias". <https://podojo.com/unconscious-bias-the-affinity-bias/> [last access: 01/03/2022]

Marcelin, J. R. *et al.* (2019). "The Impact of Unconscious Bias in Healthcare: How to Recognize and Mitigate It." *The Journal of Infectious Diseases*, 220(2): S62–S73,

Ministero Italiano dell'Istruzione, dell'Università e della Ricerca. "Raggruppamenti dei corsi di studio per Area disciplinare". <http://attiministeriali.miur.it/media/275438/tabella1.pdf> [last access: 08/02/2022]

Nickerson, R. S. (1998). “Confirmation bias: A ubiquitous phenomenon in many guises.” *Review of General Psychology*, 2: 175–220.

Ross, H. (2008). “Exploring Unconscious Bias”. *Best Diversity Practices*, 2(5)

Stanovich, K. E. & West, R. F. (2008). “On the relative independence of thinking biases and cognitive ability”. *Journal of Personality and Social Psychology*, 94(4): 672.

Wilson, T. D. (2002). “Strangers to ourselves.” *Harvard University Press*.

ANNEX 1 – Additional information on collected data

Table I. List of AMCs (and relating funds) covered in the analysis

	AMC	Fund	Target (€M)[1]	Total raised UtD (€M)[1]	Kick-off date[2]	Expected closing date[3]
1	360 Capital Partners	Nestor 2000	130	130	01-2000	01-2011
2	360 Capital Partners	360 Capital One	100	100	02-2008	02-2019
3	360 Capital Partners	360 Capital 2011	75	75	10-2012	10-2023
4	360 Capital Partners	Robolution	80	80	01-2014	01-2025
5	360 Capital Partners	360 Square	35	35	12-2015	12-2026
6	360 Capital Partners	Poli360	54	54	01-2018	01-2029
7	360 Capital Partners	A+360 Fund	30	30	01-2020	01-2031
8	360 Capital Partners	360 Fund V	150	90	05-2020	05-2031
9	AVM Gestioni	Cysero	100	15	03-2021	03-2032
10	Azimut libera impresa	Italia 500	40	40	01-2020	01-2031
11	Azimut libera impresa	Azimut Digitech Fund	65	65	01-2021	01-2028
12	CDP Venture Capital	Fondo Italia Venture I	80	80	10-2015	10-2026
13	CDP Venture Capital	Fondo Italia Venture II	150	150	04-2018	04-2030
14	CDP Venture Capital	Fondo Evoluzione	200	100	03-2021	03-2032
15	Claris Ventures	Claris Biotech I	30	30	09-2020	09-2031
16	Eureka! Ventures	Eureka! Fund I	40	40	01-2020	01-2031
17	Indaco Venture Partners	Atlante Ventures	25	25	12-2007	12-2018
18	Indaco Venture Partners	TT Venture	65	65	01-2008	01-2023
19	Indaco Venture Partners	Atlante Ventures Mezzogiorno	25	25	04-2009	04-2020
20	Indaco Venture Partners	Atlante Seed	10	10	07-2011	07-2022
21	Indaco Venture Partners	Indaco Ventures Fund I	250	130	06-2018	06-2029
22	Innogest Capital	Innogest Capital I	80	80	05-2006	05-2017
23	Innogest Capital	Innogest Capital II	85	85	09-2015	09-2026
24	Liftt	Liftt	21	21	12-2019	12-2030
25	Lumen Ventures	Lumen Ventures Fund	25	25	07-2020	07-2031
26	Neva	Neva First Fund	250	180	08-2020	08-2032
27	Oltre Impact	Oltre I	8	8	01-2006	01-2017
28	Oltre Impact	Oltre II	43	43	03-2016	03-2027
29	P101	Programma101	67	67	11-2014	11-2025

30	P101	Programma102	103	103	05-2018	05-2029
31	Panakès Partners	Panakès Fund I	100	100	03-2016	03-2027
32	Primo Ventures	Digital Investments	6	6	10-2010	10-2021
33	Primo Ventures	Barcamper Venture	44	44	09-2016	09-2027
34	Primo Ventures	Barcamper Venture Lazio	8	8	08-2019	08-2030
35	Primo Ventures	Primo Space Fund	85	85	09-2019	09-2030
36	Synergo Capital	Sinergia Venture Fund	150	30	03-2021	03-2032
37	United Ventures	United Ventures One	70	70	10-2014	10-2025
38	United Ventures	United Ventures 2	120	120	12-2019	12-2030
39	United Ventures	UV T-Growth	100	100	07-2021	07-2032
40	Vertis	Vertis Venture	25	25	03-2009	03-2021
41	Vertis	Vertis Venture 2 Scaleup	36	36	07-2017	07-2027
42	Vertis	Vertis Venture 3 Technology Transfer	40	40	08-2017	08-2027
43	Vertis	Vertis Venture 4 Scaleup Lazio	8	8	03-2019	07-2027
44	Xyence (formerly Principia)	Principia Fund	25	25	06-2005	06-2016
45	Xyence (Principia)	Principia II	64	64	06-2009	06-2020
46	Xyence (Principia)	Principia III - Health	206	206	12-2014	12-2025

Notes

1. Data on target amount to raise and commitment UtD were obtained from PitchBook, press releases and AMCs' websites.

2. Kick-off date was assumed to coincide with first closing date. Data were obtained from PitchBook, press releases and AMCs' websites.

3. Unless fund length was explicitly found (e.g. on AMC's website), expected closing date was computed assuming a total fund life of 11yrs. This results from: 5-year investment period, 5-year portfolio management & divestment period and 1-year grace period.

Table II. List of CVCs (and relating funds) covered in the analysis

	CVC	Fund	Kick-off date[1]	Fund Status[2]
1	Barilla	Blu1877	02-2018	ACTIVE
2	Healthware	Healthware Ventures	06-2005	ACTIVE
3	Mediaset	Ad4Ventures	03-2013	ACTIVE
4	Reale Group	Reale Group Corporate Venturing	01-2018	ACTIVE
5	Tim	Tim Ventures	07-2014	ACTIVE
6	Zanichelli	Zanichelli Venture	05-2019	ACTIVE
7	ZernikeMeta Ventures	Ingenium Catania	01-2010	CLOSED
8	ZernikeMeta Ventures	Ingenium Emilia Romagna I	01-2004	CLOSED
9	ZernikeMeta Ventures	Ingenium Emilia Romagna II	01-2010	CLOSED
10	ZernikeMeta Ventures	Ingenium Sardegna	01-2009	CLOSED
11	ZernikeMeta Ventures	Ingenium Umbria	01-2004	CLOSED

Notes

1. Kick-off date was retrieved from PitchBook, press releases and CVCs' websites.

2. Only ZernikeMeta Ventures has funds with an explicit closing date (retrievable from the company's website), while the other CVCs are still seeking investment opportunities.

Table III. Classification of the business verticals into the 9 sectors

Sectors	Verticals			
Digital	Drug Delivery[1]	E-Commerce	Marketplace	Mobile
	Printing Services	Second Hand		
Education & HR	Dental Education	EdTech	HR Tech	
FinTech	Accelerator	Banking	Crowdfunding	Cryptocurrency/ Blockchain
	FinTech	InsurTech	LegalTech	Payments
Food & Agriculture	AgTech	E-Grocery	Food and Beverage	Food Delivery
	FoodTech	Restaurant Technology		
Healthcare & Biotech	Diabetes	Digital Health	Health Services	HealthTech
	Life Sciences	Medical Device	Nanotechnology	Oncology
Media	AdTech	AudioTech	Marketing Tech	Phototech
	Price Comparison	Publishing	TMT	
SaaS & Software	Application Performance Management	CloudTech & DevOps	Customer Service	Cybersecurity
	eSports	Event Management	Gaming	Mobile Apps
	SaaS	Social Impact[2]	Sport Management App	
Smart City	Autonomous cars	CleanTech	Cycling	Delivery
	Green Energy	Home Rental	Mobility Tech	Real Estate Technology
	Smart Cities	Storage	Supply Chain Tech	Travel
Tech	3D Printing	Advanced Manufacturing	Artificial Intelligence & Machine Learning	Augmented Reality
	Big Data	Engineering	Industrials	Internet of Things
	Manufacturing	Materials	Oil & Gas	RFID
	Robotics and Drones	Security	Space Technology	Virtual Reality
	Wearables & Quantified Self			

Notes

1. Pharma Prime S.r.l. is the only start-up belonging to this business vertical. The attribution of the vertical to Digital instead of Healthcare & Biotech offers a better representation of the start-up's business model.

2. Mygrants S.r.l. S.B. is the only start-up belonging to this business vertical. The attribution of the vertical to SaaS & Software offers an appropriate representation of the start-up's business model.

Table IV. Attribution of deal stage based on round series

	Investment Stage	Investment Type
1	**Acceleration/Incubation**	Acceleration
2	**Early Stage VC**	Pre-seed, Seed, Series A
3	**Later Stage VC**	Series B, Series C, Series D, Series E

Table V. Attribution of field of study based on subject of study

Field of study	Subject of study			
Technology & Science	Actuarial & Financial Science	Aerospace engineering	Architecture	Artificial Intelligence
	Astronomy	Astrophysics	Biochemistry	Biology
	Biomedical engineering	Biotechnology	Chemistry	Computer Science
	Data Science	Dental Technician Institute	Economics	Electronics & Computer Science
	Electronics	Engineering	Finance	Genetics
	Imaging Science	Mathematics	Mathematics & Computer Science	Medicinal Chemistry
	Medicine	Natural Science	Neural Systems	Neuroscience
	Optics	Pharmacy	Physics	Psychiatry
	Robotics	Science	Scientific High School	Software
	Sport Science	Statistics	Technical & Commercial Institute	Technical Institute
	Telecommunications	Telecommunications engineering	Tourism	
Human & Social Sciences	Cinematic Arts	Classical High School	Classical Literature	Communication Studies
	Design	Diplomatic Studies	Graphics	International Relations
	Journalism	Languages & Modern Literature	Law	Linguistic High School
	Linguistics	Literature	Marketing & Communication	Media
	Media & Telecommunications	Music	Philosophy	Political Science
	Psychology	Social Science	Sociology	

ANNEX 2 – Founders' and partners' summary stats

Table I. Founders' summary stats, by AMC

AMC	Sample Size	Avg Age Founders	Avg Age at incorporation date	Avg Age at deal's closing date	# Male Founders	Avg Age Male Founders	# Female Founders	Avg Age Female Founders
Indaco Venture Partners	122	51	38	44	111	51	11	53
United Ventures	94	42	33	37	89	42	5	40
Vertis	92	46	36	40	81	46	11	45
Primo Ventures	101	40	33	37	93	40	8	43
P101	101	40	32	35	88	40	13	40
Innogest Capital	87	48	37	41	80	48	7	49
360 Capital Partners	279	42	34	37	259	42	20	42
Xyence (former Principia)	103	55	42	46	95	55	8	55
Liftt	38	47	42	46	33	46	5	54
Panakès Partners	27	53	43	49	24	54	3	46
Claris Ventures	7	50	44	48	6	50	1	48
Eureka! Ventures	19	37	33	36	19	37	0	n.a.
CDP Venture Capital	127	43	36	40	108	43	19	41
Synergo Capital	4	38	31	38	3	40	1	32
Neva	31	47	41	45	30	47	1	39
Azimut Libera Impresa	61	40	34	39	56	41	5	38
AVM Gestioni	2	44	38	43	2	44	0	n.a.
Lumen Ventures	5	42	36	39	5	42	0	n.a.
Oltre Impact	28	52	43	47	22	52	6	50

Note: *the table divides founders based on the AMC that invested in their startup. The number of unique founders (979) does not coincide with the sum of the founders in each AMC (1344) because some startups were financed by more than one AMC.*

Table II. Founders' summary stats, by CVC

CORPORATE	Sample Size	Avg Age Founders	Avg Age at incorporation date	Avg Age at deal's closing date	# Male Founders	Avg Age Male Founders	# Female Founders	Avg Age Female Founders
ZernikeMeta Ventures	67	49	37	38	57	50	10	43
Mediaset	48	40	30	35	43	40	5	40
Tim	43	40	31	33	41	39	2	46
Zanichelli	20	39	33	37	15	39	5	38
Healthware	17	41	35	37	16	41	1	45
Barilla	14	39	32	36	10	39	4	40
Reale Group	11	37	31	33	11	37	0	n.a.

Note: *the table divides founders based on the CVC that invested in their startup. The number of unique founders (214) does not coincide with the sum of the founders in each CVC (220) because some startups were financed by more than one CVC.*

Table III. Partners' summary stats, by AMC

AMC	Sample Size	Avg Age Partners	Avg Age at deal's closing date	# Male Partners	Avg Age Male Partners	# Female Partners	Avg Age Female Partners
Indaco Venture Partners	9	54	47	6	54	3	54
United Ventures	6	47	50	6	47	0	n.a.
Vertis	6	57	53	6	57	0	n.a.
Primo Ventures	6	52	52	6	52	0	n.a.
P101	5	44	44	4	44	1	41
Innogest Capital	11	51	44	11	51	0	n.a.
360 Capital Partners	16	49	45	14	50	2	47
Xyence (former Principia)	9	52	41	9	52	0	n.a.
Liftt	6	55	55	5	53	1	65
Panakès Partners	3	56	55	2	59	1	51
Claris Ventures	4	47	43	3	38	1	71
Eureka! Ventures	4	55	54	3	54	1	57
CDP Venture Capital	5	47	50	2	43	3	49
Synergo Capital	2	47	47	2	47	0	n.a.
Neva	3	48	46	3	48	0	n.a.
Azimut Libera Impresa	3	39	42	3	39	0	n.a.
AVM Gestioni	4	61	60	3	60	1	62
Lumen Ventures	6	38	35	5	35	1	53
Oltre Impact	3	59	56	2	67	1	43

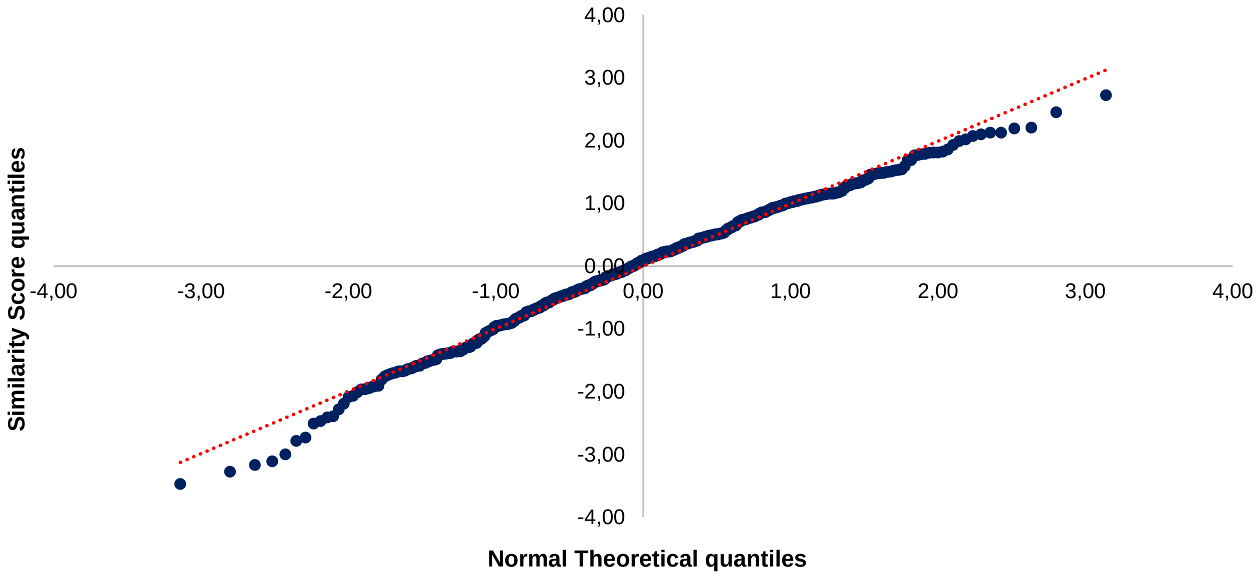
Note: *the number of unique partners (106) does not coincide with the sum of the partners in each AMC (111) because 5 partners overlaps among 2 AMCs*

Table IV. Partners’ summary stats, by CVC

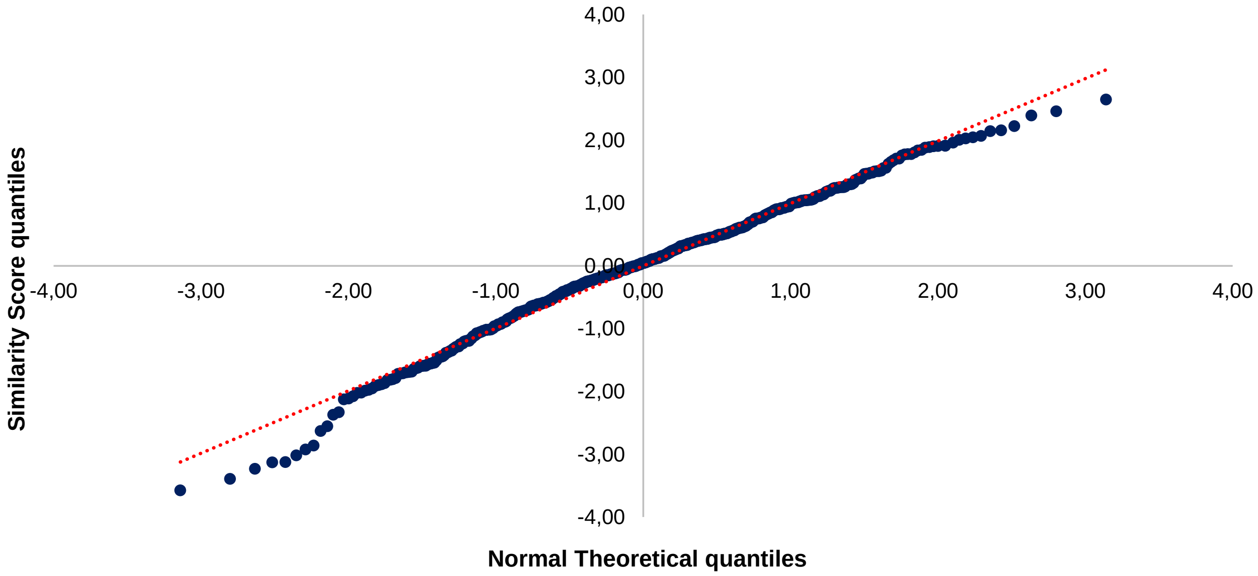
Corporate	Sample Size	Avg Age Partners	Avg Age at deal's closing date	# Male Partners	Avg Age Male Partners	# Female Partners	Avg Age Female Partners
ZernikeMeta Ventures	2	57	46	1	60	1	53
Mediaset	3	51	46	2	53	1	47
Tim	3	56	50	3	56	0	n.a.
Zanichelli	1	47	45	1	47	0	n.a.
Healthware	1	50	46	1	50	0	n.a.
Barilla	2	61	58	1	66	1	57
Reale Group	2	47	43	2	47	0	n.a.

ANNEX 3 – Tables and graphs on Sample 1 similarity score distribution

Graph I. Similarity Score Q-Q plot - base specification



Graph II. Similarity Score Q-Q plot - background-based specification



Graph III. Similarity Score Q-Q plot - ethnicity-based specification

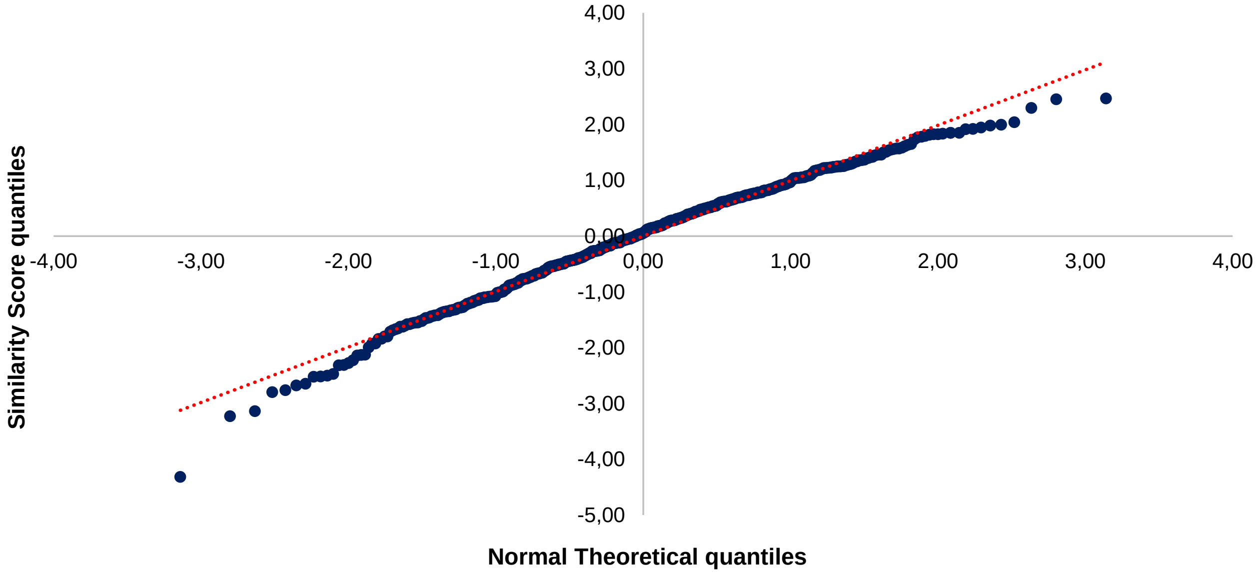


Table I. Sample 1 similarity score summary stats, by AMC (base specification)

AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,71	0,71	0,86	0,49	0,08	-0,36	0,39
United Ventures	34	0,69	0,67	0,91	0,53	0,09	0,54	-0,20
Vertis	37	0,73	0,71	0,85	0,54	0,08	-0,35	0,01
Primo Ventures	41	0,66	0,65	0,92	0,37	0,11	-0,11	0,45
P101	43	0,64	0,64	0,83	0,33	0,09	-0,71	2,24
Innogest Capital	43	0,75	0,77	0,91	0,52	0,10	-0,51	-0,32
360 Capital Partners	108	0,69	0,70	0,89	0,38	0,09	-0,59	0,59
Xyence (former Principia)	52	0,74	0,75	0,99	0,44	0,12	-0,22	-0,41
Liftt	18	0,75	0,74	0,93	0,61	0,08	0,37	0,87
Panakès Partners	12	0,76	0,77	0,84	0,63	0,06	-0,67	-0,10
Claris Ventures	3	0,71	0,77	0,82	0,56	0,14	-1,52	n.a.
Eureka! Ventures	8	0,62	0,62	0,70	0,54	0,05	0,02	-0,43
CDP Venture Capital	62	0,54	0,53	0,84	0,32	0,10	0,58	0,64
Synergo Capital	2	0,82	0,82	0,88	0,76	0,09	n.a.	n.a.
Neva	13	0,81	0,79	0,96	0,71	0,08	0,53	-0,61
Azimut Libera Impresa	30	0,68	0,70	0,89	0,33	0,11	-1,17	3,55
AVM Gestioni	1	0,92	0,92	0,92	0,92	n.a.	n.a.	n.a.
Lumen Ventures	4	0,74	0,76	0,82	0,62	0,08	-1,28	2,15
Oltre Impact	20	0,64	0,66	0,83	0,29	0,13	-0,88	1,46

Table II. Sample 1 similarity score summary stats, by funding series (base specification)

Funding Series	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Pre-seed	29	0,71	0,71	0,94	0,53	0,09	0,31	1,10
Seed	247	0,68	0,70	0,99	0,29	0,12	-0,46	0,29
Series A	172	0,69	0,68	0,92	0,37	0,10	-0,13	-0,13
Series B	46	0,66	0,66	0,91	0,32	0,13	-0,39	0,51

Note:

In order to avoid the risk of reaching conclusions based on too small samples, only funding series with at least 25 investments were considered.

Table III. Sample 1 similarity score summary stats, by sector (base specification)

Sector	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Education & HR	26	0,63	0,60	0,86	0,41	0,10	0,43	0,41
Food & Agriculture	27	0,64	0,64	0,89	0,38	0,13	-0,03	-0,21
FinTech	58	0,72	0,74	0,99	0,33	0,13	-0,59	0,17
Media	60	0,67	0,66	0,96	0,45	0,12	0,47	-0,17
Digital	64	0,67	0,70	0,85	0,29	0,12	-1,20	1,71
Smart City	67	0,67	0,68	0,94	0,32	0,11	-0,75	2,15
SaaS & Software	69	0,71	0,72	0,92	0,46	0,10	-0,36	-0,21
Healthcare & Biotech	109	0,70	0,72	0,91	0,42	0,11	-0,47	-0,51
Tech	113	0,70	0,70	0,92	0,46	0,10	-0,22	-0,25

Table IV. Sample 1 similarity score summary stats, by AMC (background-based specification)

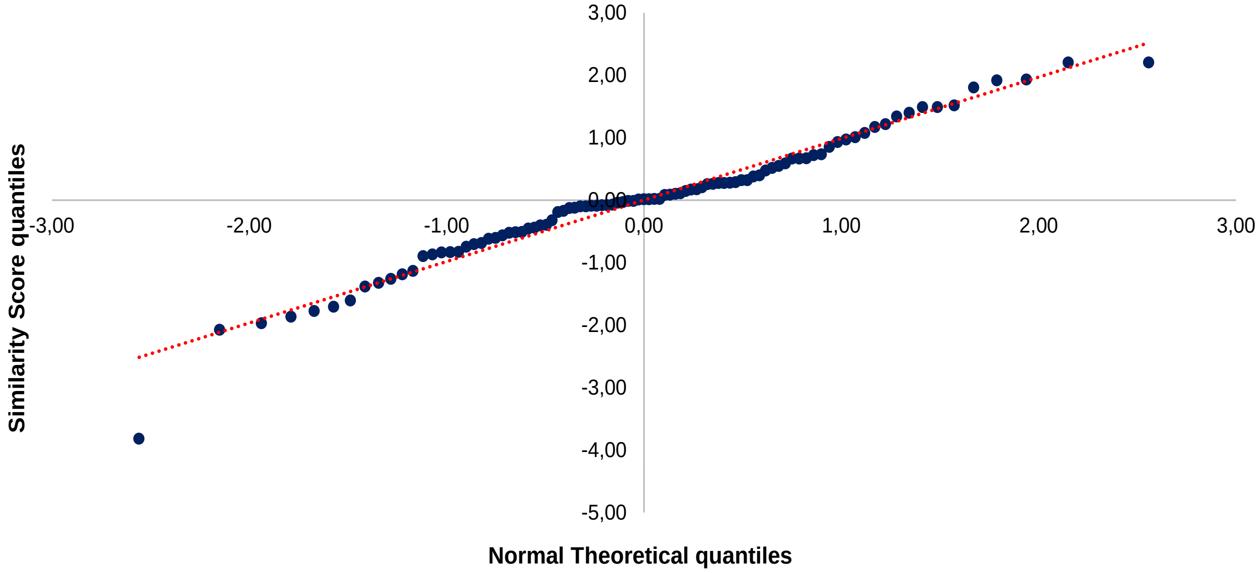
AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,69	0,70	0,84	0,47	0,07	-1,28	2,27
United Ventures	34	0,69	0,69	0,88	0,48	0,08	-0,35	0,63
Vertis	37	0,73	0,74	0,90	0,49	0,09	-0,60	0,08
Primo Ventures	41	0,67	0,67	0,91	0,36	0,11	-0,20	1,54
P101	43	0,67	0,67	0,93	0,32	0,11	-0,68	2,00
Innogest Capital	43	0,76	0,80	0,93	0,53	0,11	-0,63	-0,81
360 Capital Partners	108	0,70	0,70	0,91	0,43	0,09	-0,38	0,70
Xyence (former Principia)	52	0,73	0,73	0,99	0,33	0,13	-0,29	0,88
Liftt	18	0,77	0,77	0,97	0,62	0,08	0,50	2,18
Panakès Partners	12	0,71	0,72	0,78	0,59	0,06	-0,70	-0,27
Claris Ventures	3	0,69	0,72	0,76	0,60	0,08	-1,20	n.a.
Eureka! Ventures	8	0,63	0,64	0,67	0,58	0,03	-0,91	0,53
CDP Venture Capital	62	0,55	0,55	0,81	0,30	0,09	0,04	1,18
Synergo Capital	2	0,79	0,79	0,85	0,73	0,08	n.a.	n.a.
Neva	13	0,78	0,82	0,97	0,62	0,11	-0,06	-1,10
Azimut Libera Impresa	30	0,73	0,76	0,90	0,35	0,12	-1,28	2,16
AVM Gestioni	1	0,92	0,92	0,92	0,92	n.a.	n.a.	n.a.
Lumen Ventures	4	0,75	0,79	0,84	0,59	0,11	-1,58	2,44
Oltre Impact	20	0,67	0,70	0,92	0,28	0,15	-0,55	0,69

Table V. Sample 1 similarity score summary stats, by AMC (ethnicity-based specification)

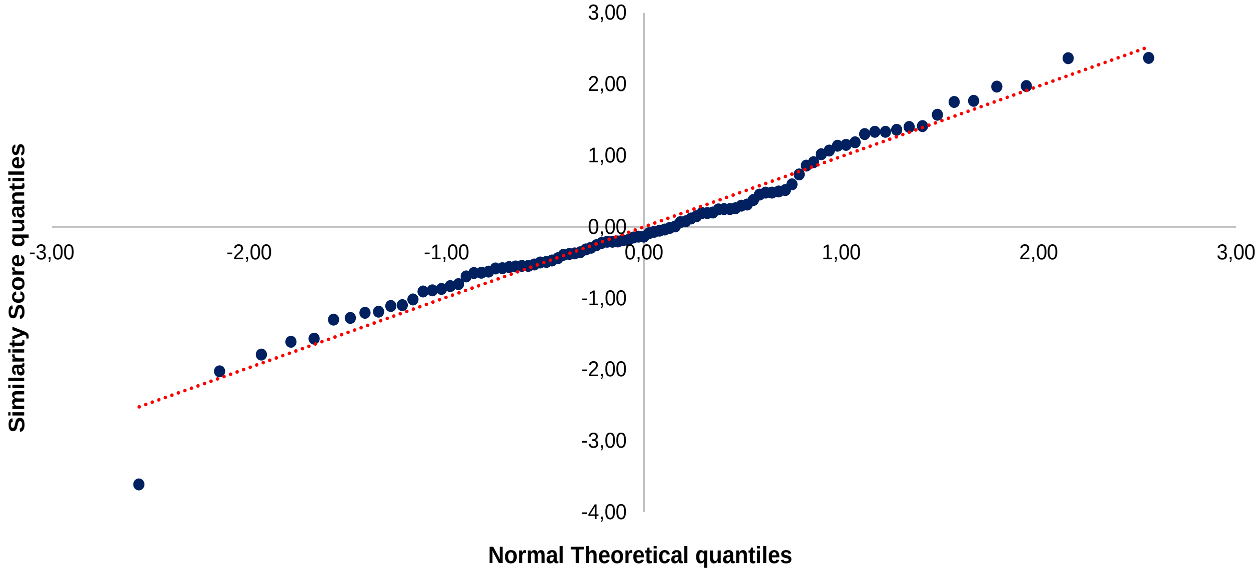
AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,76	0,78	0,90	0,49	0,09	-0,80	0,55
United Ventures	34	0,72	0,76	0,90	0,31	0,11	-1,65	4,20
Vertis	37	0,75	0,76	0,88	0,56	0,08	-0,54	-0,37
Primo Ventures	41	0,71	0,69	0,93	0,53	0,09	0,65	0,25
P101	43	0,75	0,77	0,90	0,58	0,08	-0,39	-0,65
Innogest Capital	43	0,81	0,82	0,94	0,62	0,08	-0,74	-0,18
360 Capital Partners	108	0,72	0,72	0,92	0,47	0,08	-0,19	0,70
Xyence (former Principia)	52	0,78	0,80	0,99	0,42	0,11	-1,05	1,89
Liftt	18	0,79	0,79	0,97	0,62	0,09	-0,20	0,27
Panakès Partners	12	0,76	0,77	0,87	0,55	0,09	-1,18	1,45
Claris Ventures	3	0,78	0,74	0,88	0,72	0,09	1,53	n.a.
Eureka! Ventures	8	0,69	0,70	0,76	0,60	0,06	-0,86	-0,68
CDP Venture Capital	62	0,65	0,65	0,82	0,47	0,09	-0,10	-0,73
Synergo Capital	2	0,83	0,83	0,91	0,75	0,11	n.a.	n.a.
Neva	13	0,81	0,81	0,99	0,65	0,10	0,11	-0,85
Azimut Libera Impresa	30	0,78	0,79	0,93	0,46	0,11	-1,04	1,44
AVM Gestioni	1	0,83	0,83	0,83	0,83	n.a.	n.a.	n.a.
Lumen Ventures	4	0,80	0,82	0,87	0,69	0,08	-1,38	2,60
Oltre Impact	20	0,73	0,73	0,89	0,43	0,12	-0,80	0,65

ANNEX 4 – Tables and graphs on Sample 2 similarity score distribution

Graph I. Similarity Score Q-Q plot - base specification



Graph II. Similarity Score Q-Q plot - background-based specification



Graph III. Similarity Score Q-Q plot - ethnicity-based specification

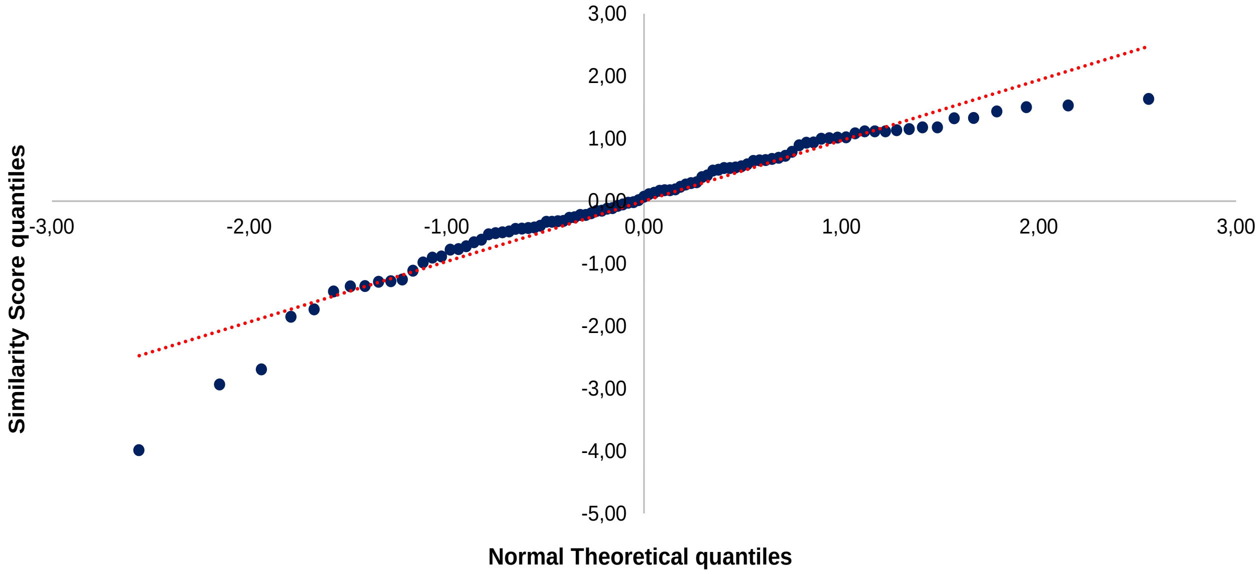


Table I. Sample 2 similarity score summary stats, by CVC (base specification)

Corporate	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Kurtosis
ZernikeMeta Ventures	29	0,73	0,72	0,85	0,64	0,05	0,77	0,58
Mediaset	22	0,64	0,64	0,80	0,51	0,08	0,11	-0,99
Tim	16	0,75	0,72	0,89	0,65	0,07	0,87	0,09
Zanichelli	10	0,67	0,69	0,85	0,34	0,14	-1,31	2,82
Healthware	7	0,80	0,74	0,92	0,71	0,10	0,40	-2,54
Barilla	6	0,70	0,68	0,84	0,52	0,12	-0,27	-0,17
Reale Group	5	0,71	0,65	0,85	0,59	0,12	0,45	-3,02

Table II. Sample 2 similarity score summary stats, by funding series (base specification)

Funding Series	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Kurtosis
SEED	53	0,74	0,73	0,92	0,52	0,09	-0,03	0,32
SERIES A	15	0,68	0,70	0,79	0,51	0,08	-0,86	0,71

Note:

In order to cushion the risk of reaching conclusions based on too small samples, only Seed and Series A were considered, as the other funding series presented a negligible number of data points.

Table III. Sample 2 similarity score summary stats, by sector (base specification)

Sector	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Kurtosis
Education & HR	12	0,66	0,67	0,85	0,34	0,13	-1,33	3,45
Food & Agriculture	8	0,66	0,66	0,81	0,52	0,09	0,03	-0,27
FinTech	8	0,76	0,77	0,89	0,54	0,12	-0,77	0,19
Media	7	0,70	0,71	0,85	0,53	0,09	-0,52	3,08
Digital	22	0,69	0,71	0,92	0,51	0,11	0,26	-0,23
Smart City	4	0,71	0,72	0,82	0,59	0,11	-0,20	-3,11
SaaS & Software	11	0,74	0,72	0,89	0,66	0,07	1,35	2,12
Healthcare & Biotech	12	0,76	0,73	0,92	0,70	0,07	1,45	1,65
Tech	11	0,71	0,70	0,77	0,65	0,03	-0,03	0,44

Table IV. Sample 2 similarity score summary stats, by CVC (background-based specification)

Corporate	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Kurtosis
ZernikeMeta Ventures	29	0,67	0,67	0,83	0,53	0,06	0,22	1,50
Mediaset	22	0,66	0,69	0,83	0,51	0,08	-0,06	-0,53
Tim	16	0,71	0,69	0,87	0,61	0,07	1,10	0,54
Zanichelli	10	0,72	0,74	0,89	0,36	0,15	-1,73	3,78
Healthware	7	0,85	0,83	0,92	0,79	0,06	0,32	-2,11
Barilla	6	0,71	0,67	0,85	0,62	0,09	1,03	-0,47
Reale Group	5	0,73	0,72	0,83	0,60	0,09	-0,52	-0,65

Table V. Sample 2 similarity score summary stats, by CVC (ethnicity-based specification)

Corporate	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Kurtosis
ZernikeMeta Ventures	29	0,75	0,77	0,89	0,64	0,07	0,00	-1,16
Mediaset	22	0,71	0,72	0,84	0,52	0,08	-0,33	0,11
Tim	16	0,72	0,71	0,88	0,60	0,08	0,60	0,17
Zanichelli	10	0,63	0,61	0,81	0,29	0,15	-1,00	1,91
Healthware	7	0,80	0,82	0,86	0,70	0,05	-1,36	1,17
Barilla	6	0,55	0,55	0,69	0,40	0,12	0,02	-1,80
Reale Group	5	0,76	0,75	0,86	0,69	0,08	0,35	-2,46

ANNEX 5 – Outputs of Sample 1 similarity score regressions

Table I. Sample 1 base similarity score regressed on partners' age at deal date (P_AGE_DEAL)

Regression Statistics	
Multiple R	0,1672
R Square	0,0279
Adjusted R Square	0,0263
Standard Error	0,1119
Observations	589

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,2112	0,2112	16,8768	0,0000
Residual	587	7,3457	0,0125		
Total	588	7,5569			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,8441	0,0386	21,8877	0,0000	0,7683	0,9198
P_AGE_DEAL	-0,0034	0,0008	-4,1081	0,0000	-0,0050	-0,0018

Note

The sample includes 589 observations because the deal closing date of 4 investments was not disclosed, so that it was impossible to compute the regressor.

Table II. Average Sample 1 base similarity score regressed on number of investments (N_INV)

Regression Statistics	
Multiple R	0,4287
R Square	0,1838
Adjusted R Square	0,1358
Standard Error	0,0768
Observations	19

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0226	0,0226	3,8284	0,0670
Residual	17	0,1003	0,0059		
Total	18	0,1229			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,7563	0,0271	27,8630	0,0000	0,6990	0,8136
N_INV	-0,0013	0,0007	-1,9566	0,0670	-0,0027	0,0001

Note

Average similarity score computed at AMC level.

Table III. Average Sample 1 base similarity score regressed on fund partners' female ratio (F_RATIO)

Regression Statistics	
Multiple R	0,5179
R Square	0,2683
Adjusted R Square	0,2516
Standard Error	0,0595
Observations	46

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0571	0,0571	16,1306	0,0002
Residual	44	0,1558	0,0035		
Total	45	0,2129			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,7283	0,0102	71,5619	0,0000	0,7078	0,7488
F_RATIO	-0,1619	0,0403	-4,0163	0,0002	-0,2432	-0,0807

Note

Average similarity score computed at fund level.

Table IV. Average Sample 1 base similarity score regressed on fund target size (SIZE)

Regression Statistics	
Multiple R	0,1014
R Square	0,0103
Adjusted R Square	-0,0122
Standard Error	0,0692
Observations	46

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0022	0,0022	0,4573	0,5024
Residual	44	0,2107	0,0048		
Total	45	0,2129			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,6989	0,0165	42,3648	0,0000	0,6656	0,7321
SIZE	0,0001	0,0002	0,6762	0,5024	-0,0002	0,0005

Note

Average similarity score computed at fund level.

Table V. Average Sample 1 valuation step-up regressed on average base similarity score (AVG_SS)

Regression Statistics	
Multiple R	0,2376
R Square	0,0565
Adjusted R Square	-0,0161
Standard Error	3,9580
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	12,1876	12,1876	0,7780	0,3938
Residual	13	203,6521	15,6655		
Total	14	215,8397			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	11,8768	10,2667	1,1568	0,2682	-10,3031	34,0566
AVG_SS	-12,6852	14,3817	-0,8820	0,3938	-43,7550	18,3846

Note

The majority of the investments mapped in the sample was relatively recent, so that for some funds the average valuation step-up was not significantly different from 1. In these cases, this index cannot be considered as a reliable metrics of fund performance, as it simply reflects the status quo *at the investment date. Thus, only the funds with average valuation step-up different from 1 were considered in the regression.*

ANNEX 6 – Outputs of Sample 2 similarity score regressions

Table I. Sample 2 base similarity score regressed on partners' age at deal date (P_AGE_DEAL)

Regression Statistics	
Multiple R	0,0178
R Square	0,0003
Adjusted R Square	-0,0104
Standard Error	0,0967
Observations	95

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0003	0,0003	0,0296	0,8637
Residual	93	0,8701	0,0094		
Total	94	0,8704			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,7286	0,1179	6,1804	0,0000	0,4945	0,9627
P_AGE_DEAL	-0,0004	0,0025	-0,1721	0,8637	-0,0054	0,0045

Table II. Average Sample 2 base similarity score regressed on number of investments (N_INV)

Regression Statistics	
Multiple R	0,1876
R Square	0,0352
Adjusted R Square	-0,1578
Standard Error	0,0571
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0006	0,0006	0,1823	0,6872
Residual	5	0,0163	0,0033		
Total	6	0,0169			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,7289	0,0408	17,8738	0,0000	0,6241	0,8337
N_INV	-0,0011	0,0025	-0,4270	0,6872	-0,0076	0,0055

Note

Average similarity score computed at CVC level.

“Felix qui potuit rerum cognoscere causas”

Vergilius, Georgica II, 490

Ringraziamenti

Ai miei genitori, grazie ai quali ho imparato che il sacrificio, presto o tardi, paga sempre. Se oggi sono Marco, lo devo soprattutto a voi.

A De, che nell’ultima estate della mia infanzia mi ha insegnato a leggere, temperando i capricci di un bambino indisponente. Tu sei già stata ciò che io sono, e io farò tutto il possibile per diventare, un giorno, quello che tu sei.

A Loris, che da quasi venti anni mi mostra come due personalità opposte possano completarsi e rafforzarsi a vicenda. Il contenuto della mia tesi è, in fondo, anche merito tuo.

A Martina, che alla mia vita ha dato un senso diverso, più dolce. Spero, tra qualche anno, di potermi voltare indietro e sorridere pensando ai lati di te che sono ormai divenuti parte di me.

A tutti voi, spero di essere riuscito, in questo primo capitolo della mia storia, a rendervi fieri di me.

Contents

Glossary

Word	Definition
Asset Management Company (AMC)	Each of the 19 professional firms managing Venture Capital funds mapped in the analysis.
Capitalization Table (CapTable)	A chart typically used by startup' founders to show how ownership is distributed among the company's shareholders.
Failure rate	The proportion of write-offs on the total number of initial investments made by an AMC.
Followers	In the context of syndicate financing, the followers are all the investors different from the lead (as defined in this glossary).
Follow-on investment	Any subsequent investment made by an AMC after the initial investment in a portfolio company.
Initial investment	First investment made by an AMC in a startup.
Investment status	The current condition of an investment made by an AMC, can be: - *active*, if the VC fund still owns the shareholding; - *transferred to new fund*, if the AMC has moved its shareholding from one of its funds to another; - *write-off*, in case the invested startup has gone bankrupt; - *exited*, if the fund has already sold its shareholding.
Lead Investor	In the context of syndicate financing, the lead investor conducts the due diligence on the startup and is responsible for direct negotiation of the round's terms and conditions.
Liquidity event	Transaction allowing the VC fund to exit from the investment. For the purpose of this paper, three liquidity events were considered: • *M&A*, if the investee was acquired by a financial or strategic buyer, or merged with another company; • *IPO*, in case the investee was listed on a public stock exchange; • *Secondary Purchase*, in case the AMC sold its shares on the secondary market to either existing or new shareholders.
Post-money valuation	Value of a company after receiving a capital injection by a VC fund. It is computed by adding the round size to the company's pre-money valuation.
Pre-money valuation	Value of a company before receiving capital injection by a VC fund. other investments such as external funding or financing.
Sector	Broad aggregation of business verticals sharing similar traits. This paper defines a total of 9 sectors based on the investees' business verticals.

Syndicate	Group of investors participating in a deal. Normally, a syndicated is made up of a lead investor and a series of followers.
Valuation step-up	A measure of a shareholding’s appreciation. It is computed as the ratio between the last available pre-money valuation of the company and its post-money valuation on the occasion of the fund’s first investment.
VC Fund	The vehicle through which an AMC makes professional Venture Capital investments.
Vertical	A business vertical describes a group of companies that focus on a shared niche or specialized market spanning multiple industries. The business vertical of each startup was retrieved from PitchBook.

1. Introduction

A substantial research literature has established that individual and collective decisions often diverge from the paradigm of rationality. This observation applies to a variety of contexts: for instance, people assess probabilities incorrectly (Kahneman & Tversky, 1973), they violate the axioms of utility theory (Kahneman & Tversky, 1978) and they interpret information in a way that confirms their prior beliefs or values (Klayman & Ha, 1987; Nickerson, 1998).

Some of the most impactful departures from normative decision-making are caused by the affinity bias, i.e. the unconscious tendency to gravitate towards people to whom we feel to be close for interests, background, ethnicity and other personal traits. The affinity bias induces preference for what is similar and distrust for what is diverse, which can ultimately lead to suboptimal outcomes in many areas, including hirings (Ross, 2008; Gompers & Wank, 2017), access to credit (Hunter & Walker, 1995) and even quality of medical treatment (Marcelin *et al.*, 2019).

This paper analyses the role of affinity bias in the Italian Venture Capital (VC) ecosystem. The background idea is that VC funds' partners may be subject to psychological biases when deciding the startups to invest in. Specifically, they could unintentionally prioritise founders who share with them cultural and genetic features. Clearly, this approach is not guaranteed to produce the best investment decisions, as it is not coherent with rational economic theory and utility maximization.

To check for the affinity bias influence on VC partners, a sample of 593 initial investments made by 19 Italian Asset Management Companies (AMC) over the last 21 years was collected. For each transaction, a percentage similarity score was built to capture the degree of proximity between the startup's founding team and the partners of the fund that participated in the deal. The similarity score, computed as a weighted average of seven variables, is presented in three different specifications, which depend on the system of weights applied.

The results obtained are quite interesting. Each specification of the similarity score presents realistic distributional features and intuitive links with selected sample variables. In particular, the affinity bias seems to affect all professional investors covered in the analysis, with average similarity scores well above 50%. Notably, the differences at AMC level can be at least partially linked to return performance, which suggests that those AMCs that suffer the most from irrationality-induced decisions reach poorer financial results. Furthermore, the impact of the affinity bias, as measured by the score, is negatively related to partners' experience.

The remainder of the paper is organized as follows. Chapter 2 offers a brief overview of existing literature contributions on the affinity bias and its implication for rational decision-making. Chapter 3 describes the data gathering process. Chapter 4 provides a general description of sample features. Chapter 5 describes the construction of the similarity score. Chapter 6 analyses the distribution of the similarity score across its three specifications, with a focus on how results change when segmenting data according to several criteria (AMC, time, financing round and sector). Chapter 7 links the similarity score to certain VC partners' features (average age at deal date and number of investments closed) and VC funds (size and overall performance). Chapter 8 provides conclusions.

2. Literature review

Unconscious biases are an unavoidable component of human life: according to Wilson (2002), we are faced with approximately 11 million bits of information at any given moment, while our brain is able to process only 40 at a time. This makes it impossible to always analyse the reality through a rational paradigm, creating the need to use non-fully rational shortcuts.

In this regard, Stanovich & West (2000) make a useful distinction between System 1 and System 2 cognitive functioning: the former resorts to intuition and is typically fast, automatic, effortless and emotional; the latter uses rationality and, as a consequence, is slower, conscious and effortful. As a matter of fact, most decisions in life are made using System 1 thinking, and while this can be helpful in many cases[1], it can lead to make serious mistakes in others[2].

In fact, cognitive biases are much more likely to happen under System 1 thinking than under System 2. One of the most discussed in literature is affinity bias, i.e. the tendency to prefer people, things and situations with which we feel a certain degree of familiarity. Affinity bias influences many of the most important decisions we make and has a profound effect on others' lives.

For example, Hunter & Walker (1995) notice that, *ceteris paribus*, US white loan agents penalize the access to credit of minorities, and they argue that this discrimination could result from the lack of cultural affinity between the two ethnic groups. Indeed, since loan agents feel to know little about minorities, they prefer to rely more on objective loan application information in appraising their creditworthiness. As a consequence, hard metrics (e.g. credit history and the ratio of total monthly obligations-to-total monthly income) have a substantially greater impact on the probability to receive a loan for minorities than for whites.

---

[1] For instance, it would be impractical (and potentially confusing) to rationally ponder every choice we make when shopping groceries.

[2] For example, people usually lose much money when gambling because they badly assess probabilities, or they overstate their level of control on the events.

Also, affinity bias can induce significant distortions in hiring practices (Louis, 2019). When evaluating candidates, recruiters tend to favour those who are more similar to them. This can create a vicious cycle whereby the newly selected members of an organization will, in turn, choose people who are affine to them, and so on. It is easy to see, then, that the affinity bias can lead to suboptimal hirings and harm diversity, especially in small firms.

The impact of the affinity bias has been evaluated also in relation to the healthcare industry. For example, Marcelin *et al.* (2019) study the US medical system and find that minority groups suffer from cognitive-bias-induced discriminations when seeking treatments. This happens because the increasing diversity in the US population is reflected in patients, but it is often missing in healthcare professionals. Therefore, under-represented categories risk experiencing health inequities caused by cultural stereotypes.

Finally, Chhaochharia & Laeven (2009) collect data on approximately 30,000 equity investments by sovereign wealth funds and find that they concentrate most of their allocations in countries displaying common cultural traits. This suggests that sovereign wealth funds prefer to "invest in the familiar", which may depend on the exploitation of informational advantages, but also on the influence of irrational affinity considerations.

Within all this framework, a growing attention has been given in the last years to the role of affinity bias in the dynamics of the VC industry. Gompers *et al.* (2016) investigate how personal traits affect VC partners' desire to collaborate and whether this attraction influences VC funds' performance. Specifically, they consider four characteristics: two (educational and professional background) are related to abilities and, as such, shall have a key role in venture capitalists' success; the other two (ethnicity and gender) are affinity-related features which do not depend on ability and, thus, shall not influence investment performance. Interestingly, the authors find that ethnicity and gender have a non-negligible impact on VC partners' desire to collaborate with other venture capitalists through syndicated investments. Notably, the authors show that this behaviour dramatically reduces returns: for instance, if two partners belonging to the same ethnic minority group invest together, performance[3] can drop by as much as 20%.

Another relevant contribution comes from Gompers & Wang (2017), to whom I partly owe the inspiration for the title of this paper. The authors analyse the impact of the affinity bias on new VC partners' hirings from an innovative perspective: specifically, they gather data on the gender of VC partners' children and find that, when existing partners have more daughters, they are more likely to hire a female investor partner, naturally increasing diversity within the organization. As a further step, they assess the consequences for the fund returns and show that greater gender diversity increases performance by a meaningful amount: on average, if existing partners have a daughter rather than a son, deal success[4] rises by almost 3% and net excess IRR[5] increases by 3.20%.

As it can be seen, the studies cited mainly focus on the internal dynamics of VC funds. Conversely, the role of affinity bias in the interplay between VC partners and invested startups is still a relatively unexplored area. This paper tries to fill this gap by quantifying the degree of cultural, ethnic, educational and professional similarity between teams of founders and partners involved in VC transactions. Having confirmed the strong presence of the affinity bias in the sample through the computation of a similarity score, the analysis assesses whether this latter can be related to specific VC funds' and partners' characteristics.

[3] Performance is measured by the probability of realizing a successful exit through IPO.
[4] Deal success is defined as a dummy variable taking value of 1 in case the VC fund realized an exit through IPO or acquisition with acquisition value higher than the invested capital.
[5] Net excess IRR is defined as the difference between a fund's net IRR and the median fund return in the same region and year.

3. Data collection

The sample includes 593 VC initial investments made by 19 Italian AMCs from January 2000 to December 2021. It is important to underline that the sample does not coincide with the population of deals closed by the AMCs over the period under analysis because of two main reasons. Firstly, follow-on transactions (as defined in the glossary) were not considered: this happens because the sample of investments is tuned for the similarity score computation, which must depend only on initial investments, as considering subsequent financings would have implied double counting. Secondly, information in the VC industry is traditionally opaque: *inter alia,* this implies that several transactions remain undisclosed and cannot be mapped in the sample.

The AMC considered are all headquartered in Italy and represent the most active professional investors in the Italian VC ecosystem.

When an AMC operates both VC and Private Equity investments (as it happens, for instance, for Vertis), then only the VC funds it manages have been considered. An exhaustive list of the AMCs (and relating funds) covered in this paper is provided in Table I of Annex 1.

For each investment, four main categories of information have been collected: i) company-specific data (e.g. date of incorporation and headquarters location); ii) deal-specific data (e.g. deal date and size); iii) data on investee's founders (e.g. gender and age); iv) data on AMC's partners who participated in the deal (same information as investee's founders). A more detailed explanation is furnished below.

3.1 Company-specific data

3.1.1 Date of incorporation

In almost all cases, the company's date of incorporation was obtained from Orbis. When not available, the foundation year was taken from PitchBook and the date of incorporation was assumed to coincide with the 1st of January. For instance, if a company's foundation year were 2010, then the date of incorporation would be set to January 1, 2010.

3.1.2 Headquarters city and country

To find geographic information on each investee, the proprietary websites and Orbis were used as primary sources. If data were not found in this way, then PitchBook and Crunchbase were checked, with priority given to the former because of its greater reliability.

3.1.3 Primary Business Vertical and Sector

Each company was assigned a sector based on its main business vertical, as provided by PitchBook. When information on primary vertical was not found, sector attribution was done by looking both at the company's business description (as provided by PitchBook) and at its website – when available.

Specifically, the following 9 sectors were defined:

- Digital;
- Education & HR;
- FinTech;
- Food & Agriculture;
- Healthcare & Biotech;
- Media;
- SaaS & Software;
- Smart City;
- Tech.

An exhaustive list of the verticals covered by each sector is provided in Table II of Annex 1.

3.2 Deal-specific data

3.2.1 Deal date

Deal date was retrieved from either PitchBook or the investment's press release. In case only the investment year was found, the deal date was assumed to coincide with the 1st of January, applying the same rationale which was followed for the startups' date of incorporation.

3.2.2 Transaction type and stage

The deals analyzed were segmented by type (i.e. by round series) and by stage (acceleration, early stage VC and later stage VC).

Information on round series was obtained from either PitchBook or press releases. When not available, the series was assigned case by case by looking at the startup's funding history: the first round was always regarded as "Seed", the second, if bigger, "Series A", otherwise "Seed" and so on. When it was not possible to unambiguously assign the series because of lack of precise information on the company's equity story, the round type was labelled "Undisclosed".

Round stage attribution directly descends from round series (see Table III of Annex 1 for more details).

3.2.3 Round size and pre-money valuation

Information on round size and pre-money valuation was obtained from either PitchBook or press releases. While in the vast majority of cases it was possible to find the round size, pre-money valuation was disclosed for fewer deals.

As Chapter 7 will show, round size and pre-money valuation enter the computation of the company's valuation step-up, which can be used as an approximate measure of the investment performance.

3.2.4 Investment Status

The investment status was labelled as:

- *active,* if the VC fund is still on the company's capitalization table (CapTable);
- *transferred to new fund,* if the AMC has moved its shareholding from one of its funds to another[6];
- *write-off,* in case the company went bankrupt;

[6] This typically happens if the fund that has originally invested in the company enters the divestment phase, but the AMC still wants to keep the shareholding.

- *exited,* if the fund sold its shareholding on the occasion of a liquidity event. For the purpose of this paper, three liquidity events were taken into account, which leads to three potential exit clusters:
    - *M&A*, if the investee was acquired by a financial or strategic buyer, or merged with another company;
    - *IPO,* in case the investee was listed on a public Stock Exchange;
    - *Secondary Purchase,* in case the AMC sold its shares on the secondary market to either existing or new shareholders.

Information on M&A and IPO activity was obtained from both Zephyr and PitchBook, while secondary market transactions were inferred by looking at changes of companies' CapTables on Orbis.

3.3 Data on founders

The founders of each startup were identified by looking at the company's profile on PitchBook and, when available, at its website and LinkedIn page.

In case either the founders were unidentifiable, or they had already left the company when the deal took place, they were replaced with C-level members.

By using this approach, a total of 979 different profiles was found.

Having identified founders (or C-level members), the following set of information was retrieved.

3.3.1 Gender

Data on gender was derived from founders' names and pictures found on the company's website and LinkedIn page. There were no cases in which identification was not possible.

3.3.2 Birth date

The birth date was either directly obtained by looking at the founder's profile on Orbis or indirectly inferred from information on graduation/high school completion date.

In case only the birth year was found, the birth date was assumed to coincide with the 1st of January of that year.

There were cases in which it was not possible to find the founders' birth date, but this was not detrimental to the similarity score computation shown in Chapter 5.

3.3.3 Nationality

Founders' nationality was retrieved by looking at their personal profiles on Orbis. As with the birth date, there were cases in which it was not possible to collect the data but, as it happened for the birth date, this lack of information did not severely impact the similarity score calculation.

3.3.4 Role start date and role end date

Logically, founders were assumed to begin their role on their startup's date of incorporation. Conversely, C-level members' start date was found on either their personal LinkedIn pages or the company's website.

For active startups, role end date was obtained by looking at founders' (C-level members') LinkedIn pages, while for bankrupt companies it was assumed to coincide with the company's dissolution date, as given by Orbis.

3.3.5 Previous professional experience

Founders were categorized based on the prevalent professional experience they had before the deal date. In this respect, the following alternatives were identified:

- *academic,* in case a founder had at least one relevant academic experience (e.g. professorship);
- *financial,* in case a founder had at least one relevant professional experience in a financial institution (e.g. bank or asset management company);
- *entrepreneurial,* in case a founder had at least one relevant experience in a startup or a corporate;
- *mixed – entrepreneurial/financial,* in case a founder had at least one relevant entrepreneurial experience and one financial experience, as previously defined;
- *mixed – entrepreneurial/academic,* in case a founder had at least one relevant entrepreneurial experience and one academic experience, as previously defined.

Information on professional experience was found by examining the founders' LinkedIn pages and *curricula* – when available.

There were cases in which it was not possible to retrieve founders' professional experience, but this had a limited impact on the similarity score computation.

3.3.6 Education level

The education level was defined by the qualification held by a founder. In this respect, the following qualifications were identified: High School Diploma, BSc, MSc, PhD, MBA, Post-Doctoral research.

Information on education level was collected by examining the founders' LinkedIn pages and *curricula* – when available.

There were cases in which it was not possible to find the data, but this was not detrimental to the similarity score computation.

3.3.7 Subject of study

Information on the subject of study was obtained by examining the founders' LinkedIn pages and *curricula* – when available.

To avoid excessive sample fragmentation, granular distinctions among subjects of the same type (e.g. mechanical engineering and electronic engineering) were not considered.

The cases in which it was not possible to find the data had a minor impact on the similarity score calculation.

3.3.8 Field of study

Field of study attribution directly descends from the subject of study (see Table IV of Annex 1 for more details). With respect to the traditional classification proposed by the Italian Ministry of Education[7], it is worth mentioning that the following subjects have been reassigned to the Scientific field: i) Economics, ii) Finance and iii) Actuarial & Financial Science.

---

[7] Ministero Italiano dell'Istruzione, dell'Università e della Ricerca. "Raggruppamenti dei corsi di studio per Area disciplinare"

This choice avoids the creation of a large (and unrealistic) gap between those subjects and others belonging to the Scientific field (especially Engineering) when building the similarity score.

3.4 Data on VC partners

In order to compute the similarity score, the same information collected for startup founders was retrieved for AMCs' partners. A total of 106 profiles was mapped.

The only point of attention concerns the previous professional experience: indeed, some VC partners do not fall into the classification detailed in paragraph 3.3.5, as they have performed a mix of academic and financial roles. Those individuals were attributed the "Mixed – academic/financial" professional background.

4. Sample description

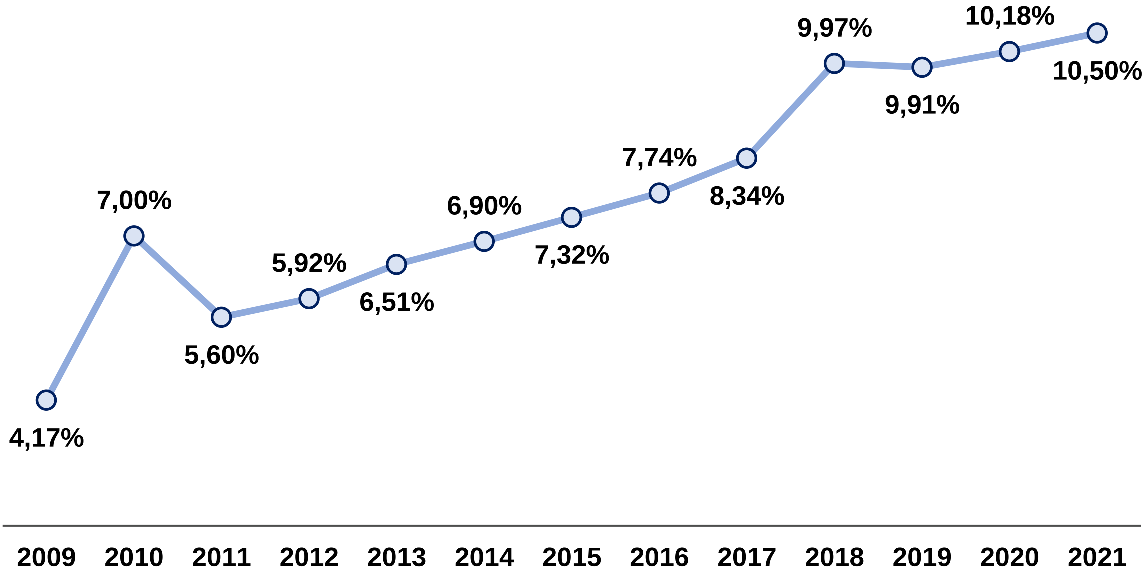
Before moving to the construction of the similarity score, it is worth providing more details on certain sample features. In this respect, paragraphs 4.1 and 4.2 focus on the descriptive analysis of founders and partners, while paragraph 4.3 gives an overview on the investments.

4.1 Founders

The sample of founders is made up of 979 individual profiles.

As far as the gender is concerned, male founders outnumber female founders by a factor of 9.5 (886 vs. 93). This suggests that the Italian VC ecosystem is heavily dominated by men, while women still struggle to emerge. It should be noticed, however, that the incidence of female founders has steadily increased over the last years. In this respect, *figure 1* shows the evolution of the percentage of women in the sample, which has more than doubled since late 2000s.

Figure 1. Cumulative incidence of female founders (%)



Note: *investments are cumulated over time. Years from 2000 to 2008 are not displayed in the graph due to the smallness of the sample.*

It is worth looking at founders' age when they launched their companies. The result is slightly higher than 36 years old, but this is likely to be an overestimation of the real datum, since many founders in the sample were not at their first entrepreneurial experience. This evidence is partially coherent with Azoulay *et al.* (2020), who used confidential administrative data sets from the U.S. Census Bureau covering the 2007-2014 period and found that, on average, US entrepreneurs start new ventures at the age of 42.

A complete set of founders' summary stats, divided by AMC, is provided in Table I of Annex 2.

4.2 Partners

The sample of partners is made up of 106 individual profiles. 360 Capital Partners and Innogest Capital are the AMCs with most partners involved in deal execution over the period analysed (16 and 11 respectively). For 360 Capital Partners, this result can be motivated by the significant investment activity (see paragraph 4.3), while for Innogest Capital the sectorial specialization may have required a higher number of partners with strong technical expertise.

As for the gender, the Italian funds are heavily dominated by men. Nearly half of the AMCs never had a female partner over the sample period, and even when women are present, they are significantly outnumbered (CDP Venture Capital is the only AMC with more female than male partners). Moreover, female partners are generally older than male partners (53 vs. 50 years across the overall sample), albeit this shall not be interpreted as a proof that the time needed to reach the apical roles in Italian AMCs depends on the gender. In fact, when considering the age at which the 106 individuals in the sample became partners, there is no clear evidence that women are penalized.

It is also interesting to look at the average age that the partners were at the time they participated in the different deals. The result for the whole sample is approximately 47 years, but this cannot be taken as a good generalization of the dynamics affecting each investment firm analysed. Indeed, when looking at data per AMC, a great variability arises, with the highest value (60 years for AVM Gestioni) and lowest value (35 years for Lumen Ventures) differing by 25 years.

A complete set of partners' summary stats is provided in Table II of Annex 2.

4.3 Investments

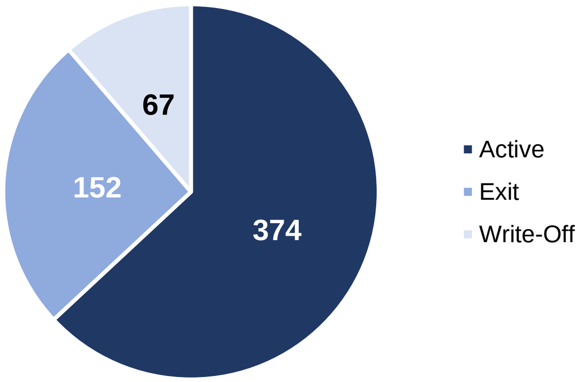
The sample includes 593 VC initial investments made from January 2000 to December 2021. As already explained in Chapter 3, follow on transactions were not taken into account.

4.3.1 Investment status breakdown

360 Capital Partners is the most active AMC (108 deals). This comes as no surprise: it is the oldest AMC in the sample (its first fund was started in the early 2000s), the largest (8 funds, with more than € 650M raised) and the most international (more than three quarters of investments made outside Italy).

Figure 2 shows the distribution of the investments by status.

Figure 2. Investments' breakdown by status (#)



As it can be noticed, most of the investments are still active (374), which depends on the combination of two main factors.

One the one hand, only 10 out of the 46 funds mapped are completely divested, while the others are either in the investment or in the portfolio management phase.

On the other hand, public coverage on Italian funds' activity has considerably increased with time, so that it has been easier to retrieve information on the most recent investments. Not by chance, albeit the sample spreads over 21 years, 55% of deals analysed happened in the last 5 years (70% in the last 7).

The remaining 219 shareholdings were liquidated because of either the investee's failure (67) or a liquidity event as defined in paragraph 3.2.4 (152).

In detail, Xyence (former Principia) is the AMC with the highest number of write-offs (17, almost one third of the investments made), but this datum is likely to be influenced by the typical VC's lack of transparency on the least successful transactions. In other words, the number of write-offs in the sample and the implied failure rate (11%) are likely to be an underestimation of the real figures. This intuition is confirmed when looking at VC-backed companies write-off rates reported by the literature, which, according to the investment stage and the definition of failure, vary from 30% to 75% (Gage, 2012).

When it comes to the 152 exits, two thirds stem from M&A (100), while only a small part is due to public market listing (10). This is fully coherent with the overall maturity of the Italian VC market, which is far behind the top European countries (e.g. UK and Germany), where IPOs are much more frequent[8].

Interestingly, although more than 70% of investments involved Italian startups (see paragraph 4.3.3 for more details), the incidence of foreign companies on liquidity events reaches 43%, peaking at ca. 50% for M&A and IPOs.

Moving to the analysis of the data at AMC level, 360 Capital Partners is responsible for almost one third of the exits (47). This depends, at least partially, on the better performance of its investments when compared to the other AMCs in the sample.

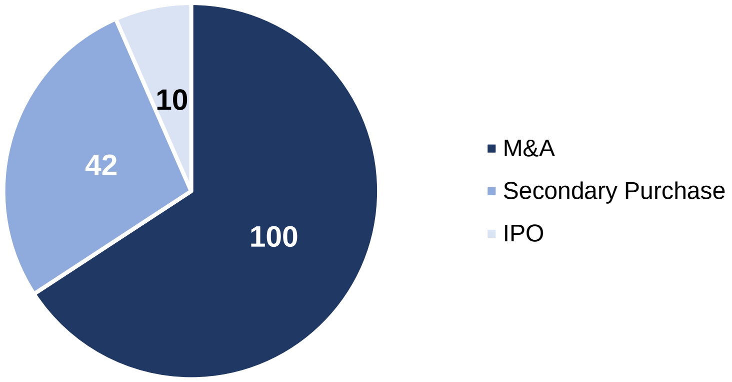
The median time to exit across the overall sample is 44 months, which is coherent with the typical investment horizon of VC funds. Moreover, among the AMCs with at least 10 exits realized[9], 360 Capital Partners has the quickest median time to exit (44 months), while Vertis displays the worst result (70 months). Among the factors explaining this marked discrepancy, the difference in geography of investments (liquid foreign markets for 360 Capital Partners, mainly south of Italy for Vertis) may have played a key role.

---

[8] Since January 2000, there have been 181 IPOs of UK-based companies backed by UK-based VC funds, and 79 IPOs of German startups backed by German VCs. Data have been extracted from PitchBook.

[9] The choice of imposing a cut-off of 10 investments when comparing averages and medians aims at reducing the bias in data. When samples are bigger (as it happens in subsequent chapters of the paper) the threshold is raised to 30.

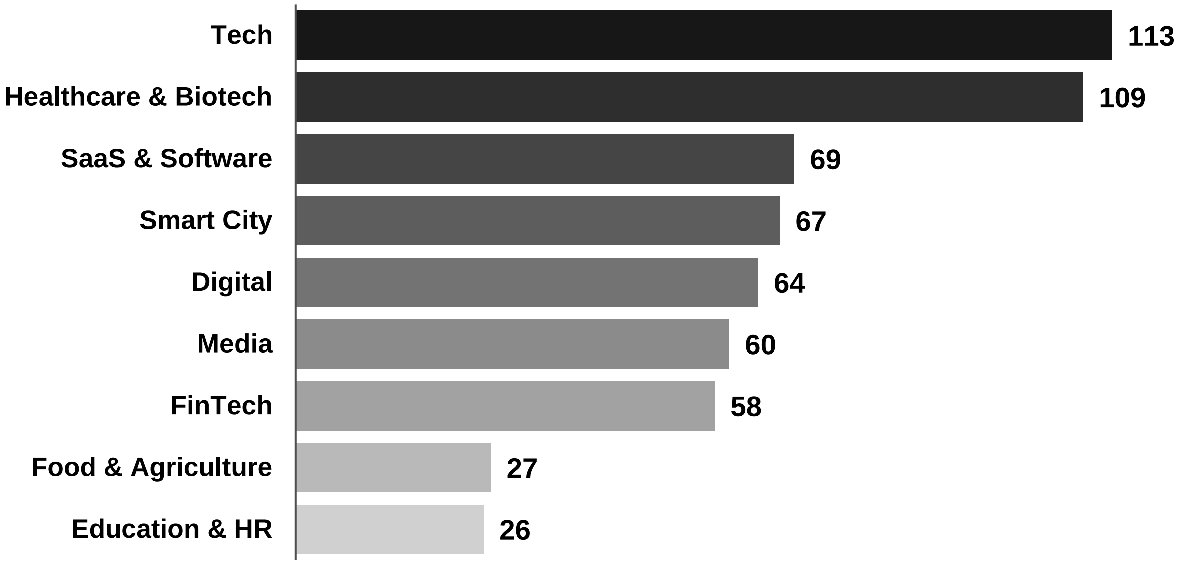
Figure 3. Exits' breakdown by cluster (#)



4.3.2 Sector and primary vertical breakdown

Figure 4 offers a sectorial breakdown of the investments covered by the analysis.

Figure 4. Investments' breakdown by sector (#)



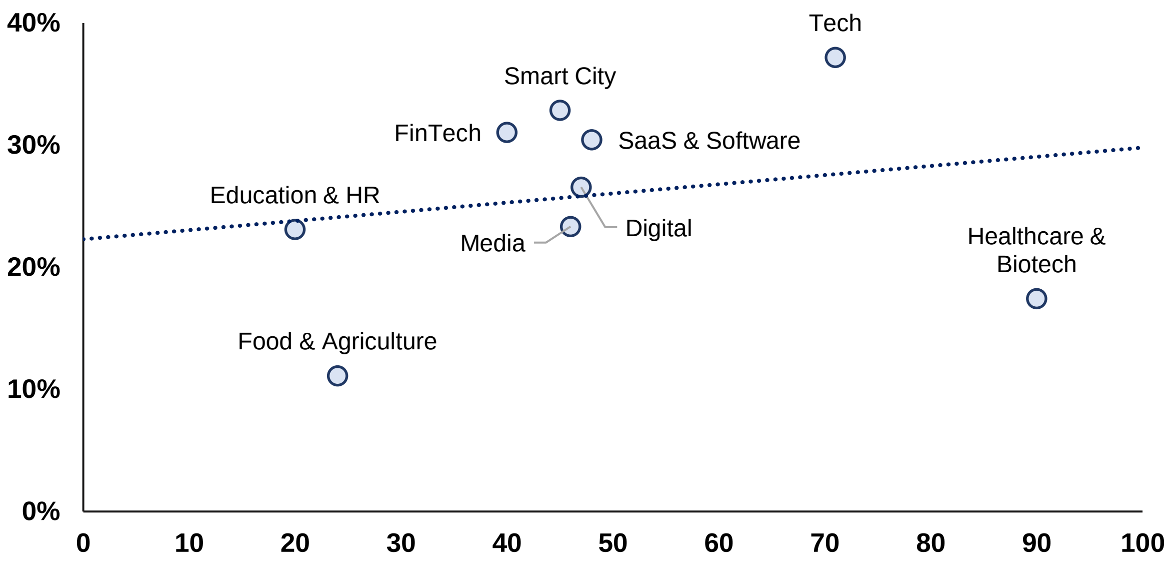
The sample appears rather concentrated, with the top 3 sectors combined representing almost 50% of the investments. Specifically, Tech is the most popular sector (113), followed by Healthcare & Biotech (109) and SaaS & Software (69).

The large importance of Tech and SaaS & Software is quite expected, as they are the two sectors normally accommodating the most innovative ventures. On the other side, the eminent role of Healthcare & Biotech is explained by the presence of three AMCs strongly focused on this sector, i.e. Innogest Capital, Panakès Partners and Xyence.

Not by chance, the three aforementioned AMCs account for almost half of the investments in Healthcare & Biotech.

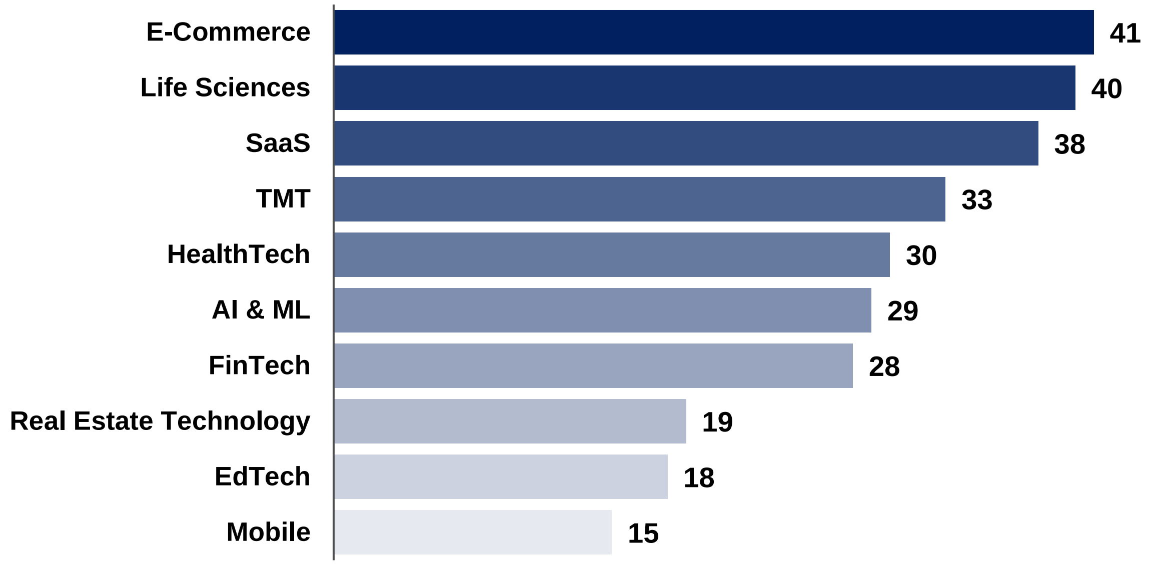
Another relevant observation is obtained by crossing the sectorial breakdown with the investees' headquarters country. Apparently, the incidence of non-Italian startups markedly varies based on the sector considered, ranging from 11% (for Food and Agriculture) to 37% (for Tech). However, aside from Healthcare & Biotech (which, as said before, is characterized by peculiar dynamics), the presence of foreign companies is higher in the most invested sectors (see *figure 5*). This result is perfectly understandable: as the AMCs' appetite for certain sectors increases, their search for investment opportunities may encourage them to explore potential targets outside national boundaries, which ultimately leads to a higher incidence of foreign financings.

Figure 5. Incidence of non-Italian startups (%) as a function of deals closed (#), by sector



When it comes to business verticals, the analysis distinguishes among 78 unique items. However, data presents a certain homogeneity, with the 10 most invested verticals covering almost half of the deals. Specifically, E-Commerce is the most popular choice (41), followed by Life Science (40) and SaaS (38). *Figure 6* offers an overview of the top 10 verticals by number of deals associated (AI & ML stands for Artificial Intelligence & Machine Learning).

Figure 6. Investments' breakdown - top 10 verticals (#)

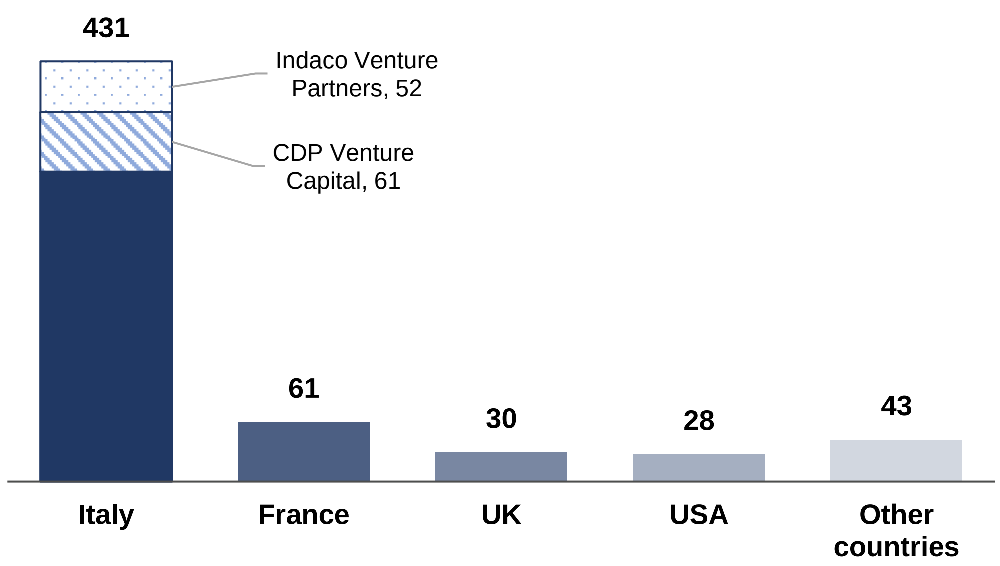


4.3.3 Geography, stage & series breakdown

The large part of the investments involved Italian startups (431): this observation alone already suggests that investment decisions are somewhat biased by affinity considerations.

Figure 7 shows the distribution of investments according to the investee's headquarter country.

Figure 7. Investments' breakdown by geography (#)



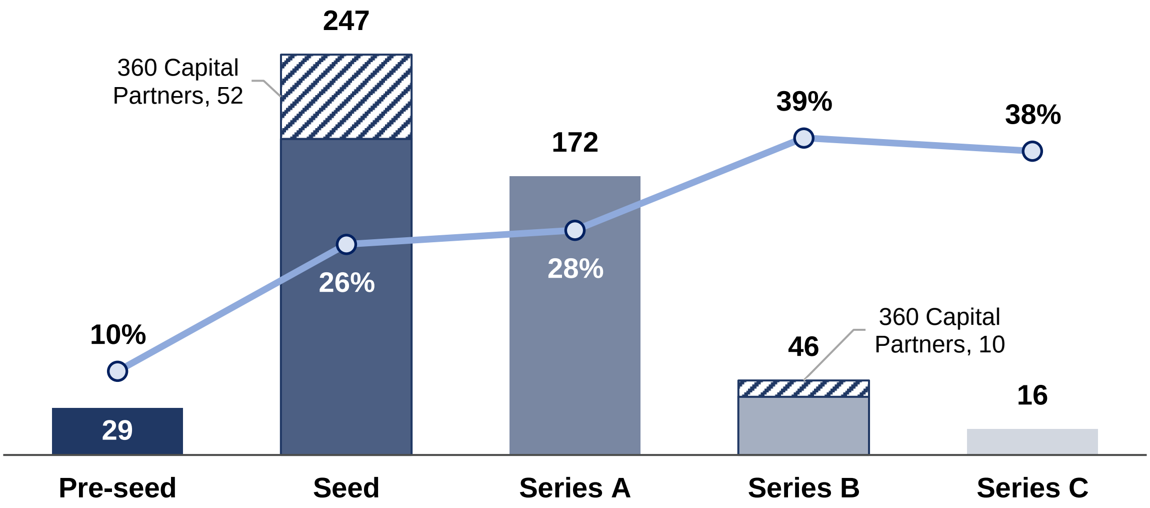
Note: *other countries include Spain (13), Germany (7), Israel (6), Switzerland (6), Netherlands (5), Austria (1), Finland (1), Ireland (1), Singapore (1), Sweden (1) and UAE (1).*

Predictably, 360 Capital Partners is the most internationally oriented AMC, with 82 deals made abroad (56 in France). On the other hand, CDP Venture Capital and Vertis are the two AMCs with the lowest activity in foreign markets among those with at least 30 investments made. This is explained by government-imposed regulatory constraints (in the former case) and by the explicit choice to allocate considerable financial resources to the south of Italy (in the latter case).

As far as the investment stage is concerned, more than 80% of transactions involved early stage financing, with an overall median deal size of € 2.5M. These data furnish a further proof of the relative underdevelopment of the Italian VC market: indeed, more mature realities (e.g. UK and Germany) typically have higher median values because of the higher incidence of later stage financing.

With respect to the funding series, Seed (243) and Series A (173) were the sweet spot of Italian AMCs' investments. Notably, the incidence of 360 Capital Partners is significant across all deal types, reaching its peak in Seed (11%) and Series B (15%). Additionally, the data suggest a strong correlation between funding series and investee's geography: the relevance of non-Italian startups increases as later stage transactions are considered (see *figure 8*).

Figure 8. Investments' breakdown (#) and incidence of deals in non-Italian startups (%), by funding series



Note: *there are two reasons why the graph does not consider all 593 investments. First, it was not always possible to identify the funding series. Secondly, it appeared more sensible to stop the analysis at Series C, as for superior series there were not enough investments to reach statistically meaningful conclusions.*

5. Model Construction

For each transaction mapped in the sample, a similarity score was built in order to capture the degree of affinity between the startup's founding team and the partners of the fund that participated in the deal. The score was computed by taking the weighted average of seven variables measuring cultural, ethnic and personal differences between investors and founders:

- difference in gender (Δgen)
- difference in age (Δage)
- difference in nationality (Δnat)
- difference in previous professional experience (Δexp)
- difference in education level (Δed)
- difference in field of study (Δfst)
- difference in subject of study (Δsst)

Thus, the full expression of the similarity score is given by:

$$\begin{aligned} &Similarity\ score_i \\ &\qquad = \alpha_1 \cdot \Delta gen_i + \alpha_2 \cdot \Delta age_i + \alpha_3 \cdot \Delta nat_i + \alpha_4 \cdot \Delta exp_i + \alpha_5 \cdot \Delta ed_i \\ &\qquad + \alpha_6 \Delta fst_i + \alpha_7 \cdot \Delta sst_i \end{aligned}$$

where each difference is computed for the i-th investment.

Each variable was built to oscillate between 0 and 1, so that the similarity score takes minimum value of 0 (signalling absence of the affinity bias) and maximum value of 1 (signalling strong presence of the affinity bias).

Paragraphs 5.1 and 5.2 describe the two-step procedure which was followed to calculate each differential in the formula, while paragraph 5.3 focuses on the identification of the system of weights.

5.1 Step 1: Computation of the mode

First of all, individual data on founders and partners were aggregated at team level by computing either their statistical mode (for non-numerical variables) or their average (for numerical variables). Notably, age was the only term for which it was possible to apply the average, while the others required mode calculations.
The criteria applied to compute the average (mode) of the (non-)numerical terms vary.

5.1.1 Mode of Gender

When it was possible to compute the mode, gender was treated as a dummy variable taking value of 1 if most founders (partners) were male and 0 if they were female.
When the mode could not be calculated – consider, for instance, teams made up of 1 male founder (partner) and 1 female founder (partner) – gender was given the intermediate value of 0.5.

5.1.2 Average Age

As said before, age is the only numerical variable included in the similarity score computation. Therefore, the average age was computed for both the founding team and the group of VC partners participating in the investment.

5.1.3 Mode of Nationality

For each transaction, the most frequent nationality among the founders (partners) was taken. For instance, a team of 2 Italian founders (partners) was considered Italian.
The cases in which it was not possible to compute the mode were treated differently depending on whether the problem concerned founders or partners.

For founders, there were situations in which the mode was not retrievable, but nonetheless there was 1 Italian member in the team. In this scenario, the founding team was attributed the Italian nationality in order to capture the known tendency of Italian VC funds to invest in startups somewhat linked with Italy.
The other cases of mode unavailability were those in which two foreign nationalities had the same frequency: in this scenario, the two nationalities were given the same importance in the similarity score computation. For example, a team made up of 1 French founder and 1 Dutch founder was considered both Dutch and French. Clearly,

this choice impacts the value of the difference in nationality (Δnat): for instance, consider the case of two founding teams, one made up of 2 French individuals and the other comprising 1 French and 1 American individual. When calculating the difference in nationality (Δnat) between each of these two teams and a group of, say, Italian partners, the former was given a lower value than the latter because discrepancies between the French and the Italian nationalities are less evident than those between the American and the Italian ones.

There were no situations with 3 or more equally frequent nationalities.

As for partners, there was no need to check whether at least 1 member of the team was Italian, as all VC funds were predominantly composed of Italians. Thus, mode calculation was straightforward.

5.1.4 Mode of Previous professional experience

For each team of founders (partners), the most frequent previous professional experience was taken – when available.

In case the mode was not computable, the partial overlapping between different professional experiences was enhanced. For instance, a team consisting of 1 founder (partner) with “Mixed – entrepreneurial/financial” experience and 1 founder (partner) with pure *financial* experience was assumed to have a *financial* background.

5.1.5 Mode of Education level

For each transaction, the most frequent education level among the founders (partners) was taken.

In case of lack of a statical mode, the member with the highest education level determined the value applied to the whole team. To this purpose, the following study path was considered:

- level 1: High School Diploma
- level 2: Bachelor of Science (BSc)
- level 3: Master of Science (MSc)
- level 4: MBA and PhD
- level 5: Postdoctoral

For example, a team made up of 1 founder (partner) with a BSc and 1 founder (partner) with a MSc was ascribed the latter's education level.

5.2.6 Mode of Field of study and Subject of study

For each team of founders (partners), the most frequent field and subject of study were considered – when available.

The cases where it was impossible to compute the mode were not solved at this stage but rather in the second step of the similarity score calculation procedure (see paragraphs 5.3.6 and 5.3.7 for more details).

5.2 Step 2: Computation of the differences

The second step of the procedure consists in measuring how much the mode (average) of each variable in the similarity score differs between founders and partners. The approach followed varies based on the variable.

5.2.1. Difference in Gender

The difference in gender between founders and partners was computed simply by taking the absolute difference of the modes:

$$\Delta gen_i = \left| gen_{f,i} - gen_{p,i} \right|$$

where $gen_{f,i}$ is the mode of founders' gender and $gen_{p,i}$ is the mode of partners' gender for the i-th investment.

5.2.2 Difference in Age

For each investment in the sample, the difference in age was computed by taking the absolute standardized distance between the average founders' and partners' age.

Standardization was made with respect to the maximum absolute difference in age found in the sample and reflects the need to make the variable oscillate between 0 and 1.

The expression of the difference in age in the i-th investment can then be written as:

$$\Delta age_i = \frac{\left| age_{f,i} - age_{p,i} \right|}{\max \left| age_f - age_p \right|}$$

where $age_{f,i}$ is the average founders' age and $age_{f,i}$ is the average partners' age. Both measures relate to the i-th investment.

5.2.3 Difference in Nationality

Each pair of founders' and partners' nationality was assigned a distance taking into account geographical, linguistic and cultural differences. For obvious reasons, the distance attributed vary between 0 and 1, where the 0 (1) signals identical (very different) nationalities.

5.2.4 Difference in Previous professional experience

Similar to what was done for nationality, each pair of professional experiences received a distance based on their degree of affinity. For instance, the pair "Financial" & "Mixed-entrepreneurial/academic" was attributed the highest distance (1), as these two backgrounds have weak commonalities. Conversely, the pair "Financial" & "Mixed-entrepreneurial/financial" was assigned a smaller distance (0.5) because the two backgrounds share the financial component. A null distance was assigned in case of identical previous professional experiences.

5.2.5 Difference in Education level

Each pair of education levels was attributed a distance based on the study path illustrated in paragraph 5.1.5.
For instance, the pair "High School Diploma – Postdoctoral" was given the maximum distance (1), as these two education levels are at opposite ends of the study path. Conversely, the pair "BSc – MSc" was assigned a smaller distance (0.25), since these two education levels are contiguous in the study path. A null distance was attributed in case of identical education levels.

5.2.6 Difference in Field of study

When it was possible to compute the mode of the field of study both for founders and partners, the distance was set to 1 if the modes were equal, 0 if they were different.
In case of absence of one of the two modes, a distance of 0.5 was attributed if there was a match in the field of study between at least 1 founder and 1 partner, otherwise

the variable was excluded from the computation of the similarity score because of the lack of unambiguous data.

5.2.7 Difference in Subject of study

As for the subject of study, a similar approach to that applied for the area of study was followed. Therefore, when it was possible to compute the mode both for founders and partners, the distance was set to 1 if the modes were equal, 0 if they were different.
When the mode was not available, but pairs of founders and partners involved in a certain investment shared at least one common value, then the distance was linked to the ratio between the number of shared subjects and the maximum number of shared subjects in the sample:

$$distance_i = 1 - \frac{(number\ of\ shared\ subjects)_i}{(number\ of\ shared\ subjects)}$$

where the subscript is referred to the i-th investment in the sample.
In all the other cases, the distance was set to the maximum value (1) because of lack of commonalities between founders and partners on this particular aspect.

5.3 Weights identification

Three specifications of the similarity score were identified by varying the weights applied to each variable in the formula.

In the *base specification*, each differential was given the same weight, so that the similarity score was computed as a simple average.

In the *background-based specification,* a higher weight was given to the variables relating to founders' and partners' academic and professional background.

Finally, in the *ethnicity-based specification*, a higher weight was attributed to gender, age and nationality. Moreover, Δed, $\Delta fstu$ and $\Delta sstu$ received a lower weight so as to cushion the partial overlapping of information among them.
Notably, the third specification gives greater importance to variables which, according to the rational economic theory, shall not be taken into account when making an investment decision. Nevertheless, as Chapter 6 shows, this version of the similarity score produces interesting results in terms of distributional features. Moreover, as

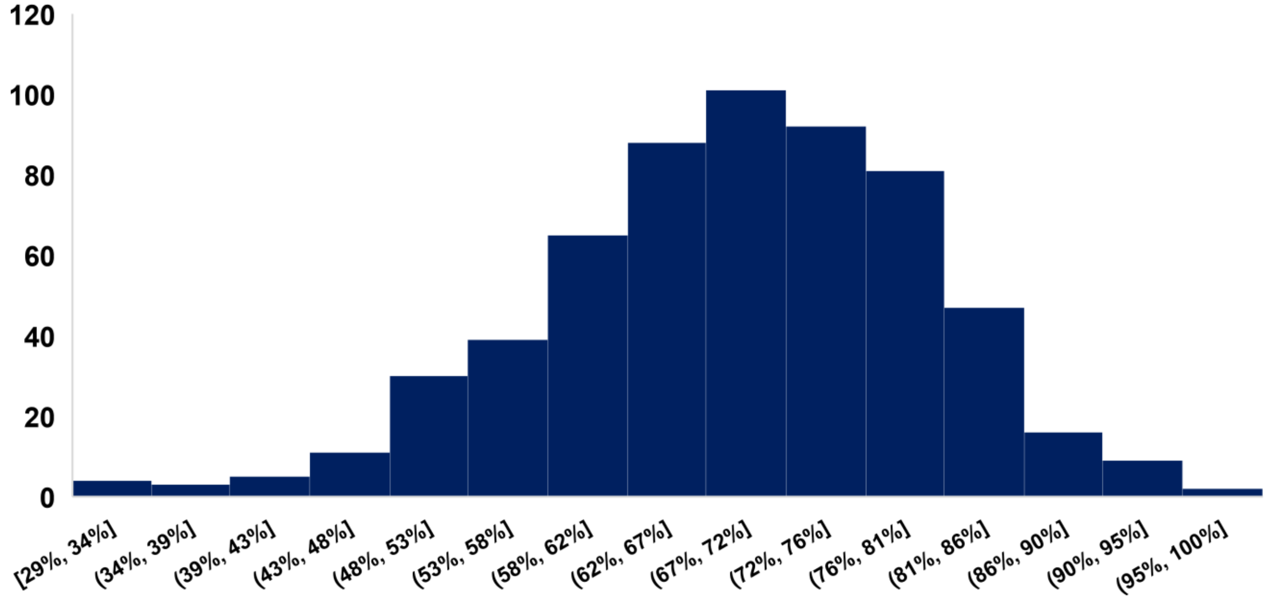
highlighted in Chapter 7, it presents a non-negligible correlation with AMCs' performance (as measured by average valuation step-up) and partners' experience (as measured by their age at deal date and by the number of investments they made).

6. Empirical results

6.1 Base specification

Figure 9 shows the probability distribution function (pdf) of the similarity score across the whole sample under the base specification.

Figure 9. Similarity Score distribution - base specification



The similarity score is bounded between 0 and 1, so that its pdf cannot be properly defined as Normal. However, the distribution depicted in the graph resembles a Gaussian, and this intuition is confirmed by the Normal Q-Q plot presented in Graph I of Annex 3.

The similarity score generally takes quite high values and is characterized by a limited variability: data range from 0.29 to almost 1.00, with an average of 0.69 and a standard deviation of 0.11. Moreover, almost 95% of observations is above the 0.50 threshold. The distribution presents a light negative asymmetry (-0.39) and is slightly concentrated around the mean, as proven by its small positive excess kurtosis (0.24).

Useful insights on the behaviour of the similarity score can be obtained by looking at the distribution for individual AMCs. A complete set of summary stats for each AMC under the base specification is provided in Table I of Annex 3.

As a first observation, the distribution of all AMCs is rather concentrated around the mean, with standard deviations ranging from 0.07 to 0.10.

More heterogeneous results are found for the difference between the maximum and the minimum: considering the AMCs with at least 30 investments[10], the value oscillates between 0.31 (for Vertis) and 0.56 (for Xyence). Additionally, the difference increases with the number of investments, which is a plausible outcome (as the sample size rises, so does the probability of finding outliers in the distribution).

Among the AMCs with at least 30 deals, Innogest Capital is the one with the highest average similarity score (0.75), while CDP Venture Capital is the one with the lowest (0.54, i.e. 0.15 less than the average across the whole sample). To check whether this difference is statistically significant, the classical t-test for mean difference was applied[11]: the extremely small p-value (1.4E-18) signals that the impact of affinity bias on the two AMCs can be considered statistically different.

Interestingly, for the three funds with a strong specialization in the medical sector (i.e. Innogest Capital, Panakès Partners and Xyence) the probability mass is shifted to the right when compared to the distribution of the overall sample[12]. As this result persists also in the other two specifications of the similarity score, this could suggest that the affinity bias is stronger when investments are concentrated on specific sectors.

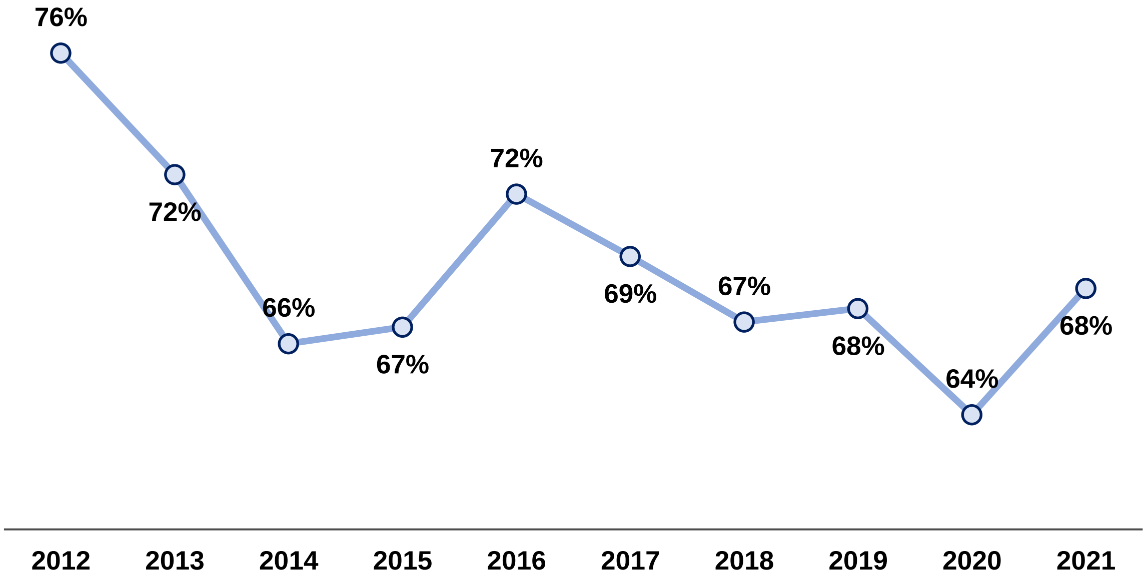
Furthermore, it is useful to focus on the historical evolution of the similarity score, which is shown in *figure 10*.

[10] As previously explained, the choice to impose a cut-off to the investments is made to avoid the risk of formulating statistically meaningless conclusions.

[11] As said before, the distribution of the similarity score under the base case is quite close to a Gaussian. Under these conditions, the use of the standard t-test to check for statistically relevant differences in means is a safe approach.

[12] For Panakès Partners, however, there are only 12 observations, so that this result is not particularly solid from a statistical standpoint and can well depend on the small sample size.

Figure 10. Similarity Score historical evolution - base specification



Note: *years from 2000 to 2012 are not displayed in the graph due to the smallness of the sample.*

As said before, the Italian market is still far from being a mature environment. Nonetheless, the last years have seen considerable growth in both the number of deals executed and the amounts financed. Within this framework, the VC funds' activity has intensified, which may have resulted in partners being more experienced and competent in investment decision-making. Therefore, a sensible measure of the similarity score shall capture this effect and show a decreasing tendency over time. *De facto*, this is exactly what happens: the average value switches from 0.76 in 2012 to 0.68 in 2021, and this change is statistically significant at any confidence (p-value 6E-4)[13].

The similarity score shall also decrease when comparing earlier rounds with later financings. Indeed, at pre-seed and seed stage, the founding team plays a key role in investment decisions because there is still a relative lack of hard metrics to analyse (e.g. revenue evolution, number of customers). As a company grows and its business expands, however, partners focus more on quantitative aspects before choosing to

[13] The reason why the analysis starts from 2012 is that this is the first year of the sample with at least 25 investments made. In this way, the risk of making conclusions based on too few observations is cushioned.

invest, so that they shall be less exposed to the affinity bias. The decreasing dynamics of the similarity score across financing series is observed in the data: the average is 0.71 for pre-seed and 0.66 for Series B, and this difference is statistically significant at 5% and 10% confidence levels (p-value 0.011).

Table II of Annex 3 provides an exhaustive set of summary stats for each financing series under the base specification.

A final analysis concerns the distribution of the similarity score across the 9 sectors identified in paragraph 3.1.3, whose summary stats are presented in Table III of Annex 3. As a preliminary note, since all sectors in the sample have a satisfactory number of observations, there is no need to exclude some of them from the general conclusions. That said, the similarity score displays quite little variability when segmented according to this criterion: the biggest average (0.72 for FinTech) and the smallest one (0.63 for Education & HR) are separated by less than 10 percentage points, albeit this difference is statistically significant at any confidence level (p-value 0.002). Furthermore, the standard deviation displays quite low values for all sectors, oscillating between 0.10 (for 3 sectors) and 0.13 (for 2 sectors).

Moreover, in 7 out of the 9 cases the similarity score distribution is negatively asymmetric, with the index taking the smallest value for Digital (-1.20). Less unanimous evidence is found for the excess kurtosis: the pdf is platykurtic for 5 sectors (with a minimum of -0.51 for Healthcare & Biotech) and leptokurtic for the remaining 4 (with a maximum of more than 2.00 for Smart City).

In summary, the study of the similarity score pdf results more informative when segmenting data by AMC than when using sectors. This outcome is quite intuitive: discrepancies in similarity scores should arise from heterogeneity of personalities and approaches of different partners' groups, something which can be captured only dividing the sample by AMC.

Furthermore, interesting suggestions come from the analysis by investment date and funding series.

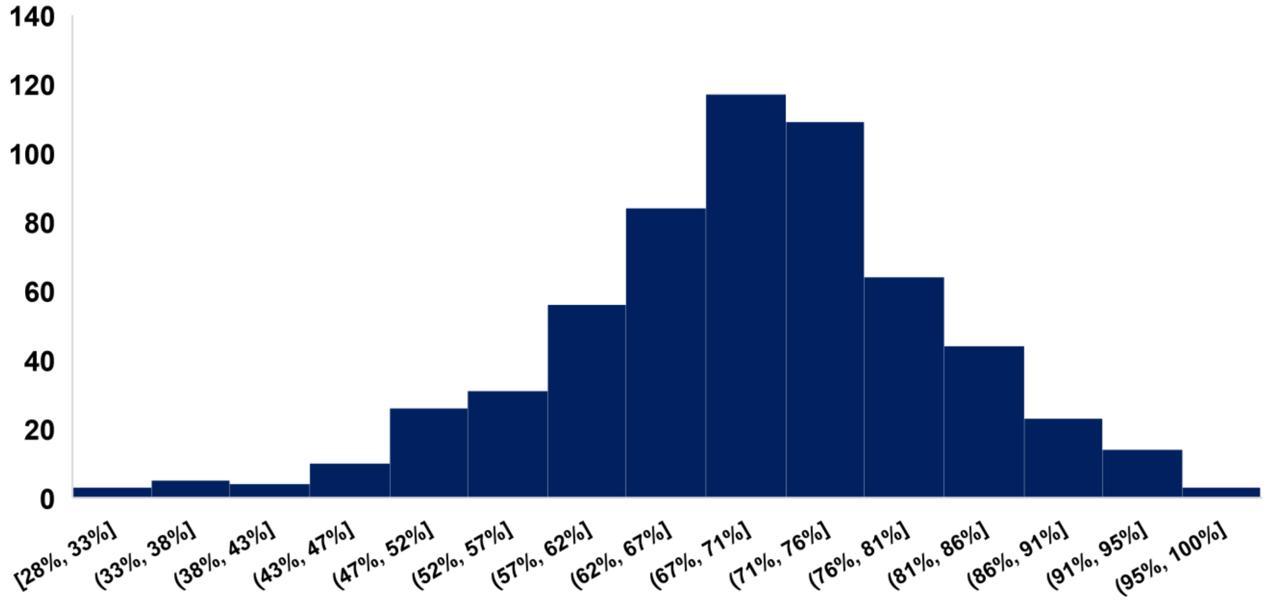
The next two paragraphs describe the other two specifications of the similarity score. Rather than replicating the same analysis made for the base version, it appears convenient to focus on the differences with this latter.

6.2 Background-based specification

Among the variables entering the computation of the similarity score, founders' education and professional background are the only two which shall be explicitly considered by partners when evaluating a potential investment. Therefore, an interesting extension of the analysis consists in studying how the similarity score distribution changes when a higher weight is given to differences in previous professional experience (Δexp) and field of study ($\Delta fstu$).

This gives rise to the background-based specification of the similarity score, whose pdf is shown in *figure 11.*

Figure 11. Similarity Score distribution - background-based specification



Moments up to the third are almost unaltered when compared to the base specification[14]. A more perceivable change concerns the excess kurtosis, which more

[14] For the first moment, the classical t-test for mean difference was performed. The related p-value (0.42) led not to reject the null of equal means between the two specifications at any confidence level.

than doubles (from 0.24 to 0.59). This suggests that the distribution departs from the Gaussian more than before, even if it still looks quite similar to it. These intuitions are confirmed by the Normal Q-Q plot shown in Graph II of Annex 3: the empirical and theoretical quantiles are almost identical in the central part of the pdf, while they diverge in the tails (especially in the left one).

Table IV of Annex 3 provides a complete set of summary stats at AMC's level. The average similarity score remains extremely similar to the base case, while asymmetry and excess kurtosis display a more variable behaviour. In some cases (e.g. Eureka! Ventures), this can be linked to the restricted sample size, but in others it signals the impact of background-related variables on the similarity score. In this regard, an interesting example is provided by United Ventures: the similarity score distribution switches from being positively asymmetric (index 0.54) and slightly platykurtic (-0.20) to be left-skewed (-0.35) and leptokurtic (0.63).

On a general basis[15], the study at AMC level leads to conclude that, even if previous professional experience and field of study do not shift the distribution mean and standard deviation, they do alter the way in which the probability mass concentrates around the center and the extreme values.

As for the data breakdown by funding series, sector and year, the background-based specification produces the same logical outcomes as the base version, which supports the plausibility of its construction.

6.3 Ethnicity-based specification

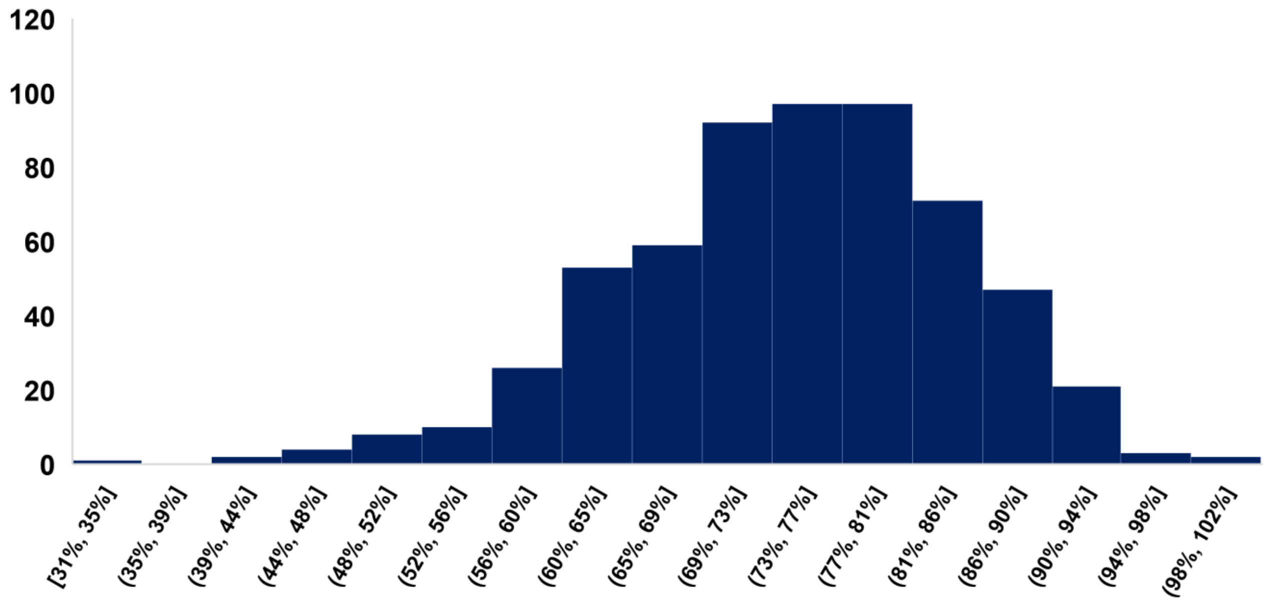
The third and last specification of the similarity score was named *ethnicity-based* because a higher weight was attributed to the terms capturing gender and nationality differences among founders and partners. This responds to the need of verifying the degree to which the variables connected to the most irrational component of decision-making influence AMCs' investments.

For the second moment, the classical F-test was run. The resulting p-value (0.40) led not to reject the null hypothesis of equal variances at any confidence level.

[15] For obvious reasons, this conclusion applies only to the AMCs with statistically significant sample sizes.

However, before moving to the analysis at AMC level, it is useful to briefly describe the general distribution of data. To this purpose, *figure 12* displays the pdf of the similarity score under this specification across the whole sample.

Figure 12. Similarity Score distribution - ethnicity-based specification



As a first observation, apart from a slight negative asymmetry, the distribution appears quite close to a Normal. This intuition is confirmed by the Q-Q plot shown in [Graph III of Annex 3](#).

Moreover, the general distribution is more concentrated around the mean and slightly shifted to the right than in the base case.

More in detail, the average similarity score increases by 4 percentage points when compared to the base specification, and this change is statistically significant, as confirmed by the t-test for mean difference (p-value 8.07E-19). Conversely, the standard deviation slightly decreases (from 0.13 to 0.10), and this negative shift is statistically significant[16].

[16] In order to check for changes of the second moment, the classical F-test was performed. The resulting p-value (0.001) led to reject the null of equal variances between the two specifications at any confidence level.

Similar to what happened in the second specification, asymmetry modestly reduces, while excess kurtosis rises considerably (even if less than in the background-based case).

A more interesting discussion concerns how the similarity score varies at AMC level with respect to the base specification. An exhaustive set of summary stats is provided in Table V of Annex 3.

Among the AMCs with at least 30 observations, CDP Venture Capital presents the highest change in the average similarity score (+0.11), but nevertheless it remains the AMC with the lowest value in the sample (0.65). However, data at AMC level result now more concentrated: the distance between the highest and the lowest average shrinks to 0.16, i.e. 5 percentage points less than in the other versions of the similarity score.

As for the other moments, standard deviations do not vary, while asymmetry and kurtosis change in different ways based on the AMC. Notably, United Ventures confirms to be the AMC with the most unstable third and fourth moments: with respect to the base case, the pdf becomes markedly left-skewed (-1.65) and extremely leptokurtic (4.20). This outcome may be in part influenced by the limited number of observations available for this AMC (34).

To summarize, apart from small-sample-induced effects, giving more importance to the irrational components of the similarity score leads to a general increase in similarity score levels, but the various distributions remain minimally dispersed.

As for the data breakdown by funding series, sector and year, the ethnicity-based specification produces the same intuitive results as the base version, which supports the plausibility of its construction.

7. Relations with partners' and funds' features

The last part of the analysis verifies whether the similarity score can be related to specific partners' and funds' features[17].

7.1 Similarity score and partners' features

As already noticed, sample data are quite heterogeneous when it comes to partners' age. Thus, an interesting question concerns the potential link between the similarity score of a each investment and the average age of partners participating in it. Intuitively, older partners could be less influenced by the affinity bias because of their longer professional experience, which shall make them less likely to fall prey to irrational decisions. Therefore, a negative correlation between the two variables could be observed.

To check that, the similarity score was regressed on partners' average age at deal closing date. Table I of Annex 4 shows the summary output for the base similarity score[18]. As expected, there is a slightly negative correlation between the two variables, with the regression coefficient equal to -0.003 and statistically significant at any confidence level. However, the R square is quite low (less than 3 percent), which implies that the overall fit is poor.

Similar conclusions are reached when using the other two specifications of the similarity score. The regression coefficient remains statistically significant at any confidence level, while the R square increases up to a more interesting 0.056 for the ethnicity-based similarity score.

Another measure of partners' experience is given by the number of investments they took part in. Intuitively, there should be a negative relationship between the average similarity score of the AMCs and the number of transactions that each of them made. In fact, this is exactly what is observed: as Table II of Annex 4 displays, the

---

[17] As suggested by the sentence, the analysis in paragraph 7.2 was made at fund rather than at AMC level. This choice is motivated by the fact that information of return performance, treated in paragraph 7.2.2, was available only at fund level (and for a restricted number of funds).

[18] A comment on the difference of results when the other specifications of the score are used is provided in the subsequent section of the paragraph.

regression coefficient for the base case is negative (-0.001) and statistically significant at 10% confidence level. Furthermore, the overall regression fit is quite high (R square 0.184).

The result slightly worsens as the background-based specification is considered, even if the regression coefficient continues to be statistically significant at 10% confidence level and the overall fit remains above 0.15.

Conversely, the situation improves with the ethnicity-based specification: the regression coefficient becomes statistically significant at 5% confidence level (but not at 1%) and the R square overcomes 20 percent.

Therefore, all in all data seem to confirm that more experienced partners tend to be less influenced by the affinity bias when making investment decisions.

7.2 Similarity score and funds' features

The following step of the analysis checks whether the similarity score depends on some relevant funds' features.

7.2.1 Fund size

The first relationship being tested is with funds' size, as measured by their target size. Ideally, bigger funds are more likely to participate in larger rounds, where more attention is devoted to hard metrics (e.g. revenue growth, client base expansion) rather than to "soft elements" (including the founding team). Moreover, they have a higher probability to act as lead investors, since they have the financial and human resources to oversee the round's progression: clearly, this means that they perform deeper analyses on potential investments. In light of these considerations, a negative relationship could be observed between fund size and average similarity score.

To corroborate this idea, the fund average similarity score was regressed on its target dimension. Table III of Annex 4 displays the full summary output for the base specification. There does not seem to exist a meaningful linear relation between the two variables: the regression coefficient (which is actually positive) is not statistically significant (p-value 0.502) and the overall fit is very low (R square 0.01).

The situation does not improve with the other versions of the similarity score: in both cases, the regression coefficient is far from being of any relevance and the R square remains extremely small.

In conclusion, the theoretical intuition about the relationship between affinity bias and fund dimension is not confirmed by the evidence found in the data.

7.2.2 Fund performance

Chapter 6 has highlighted that Italian AMCs' investment decisions are somewhat influenced by affinity considerations. Indeed, the similarity score reaches high values (typically more than 0.60) in any sample breakdown considered, peaking at almost 1 in certain investments.

The question that naturally follows from this observation is whether and how the affinity bias impacts fund performance. Theoretically, it should induce suboptimal asset allocations, eventually leading to poorer returns. Therefore, the average similarity score of a fund should be negatively correlated with its IRR.

Unfortunately, data on funds' IRR were publicly available only in 5 cases, so that it was impossible to reach statistically solid conclusions. Anyway, it is worth mentioning that the worst performing funds in terms of IRR were those with the highest average similarity score, while the best performing ones showed the lowest sensitivity to the affinity bias.

An alternative (albeit less precise) approach to measure the performance of an investment consists in computing the investee's valuation step-up. This is obtained by taking the ratio between the latest available pre-money valuation of the company and its post-money valuation on the occasion of the fund's first investment. Clearly, if a startup grows and its business expands, its value is likely to increase, which will cause the valuation step-up to rise; on the other side, poor market performance will lower the valuation, driving down the ratio.

It is important to recall that the valuation step-up is an approximate measure of the investment return, since it presents a number of deficiencies when compared to the IRR. Firstly, being a cash-on-cash multiple, it does not take into account the time value of money. Secondly, it just compares the latest valuation of the company with the post-money valuation when the fund invested for the first time. Therefore, it does not consider potential follow-on investments. Thirdly, it neglects the operating costs, which normally erode the fund performance. Finally, it is invariant to the company's percentage stake acquired by the fund.

That said, data on companies' pre- and post-money valuations were obtained from PitchBook and Zephyr – when publicly disclosed. Clearly, in case the fund entry coincided with the latest round made by the company, the valuation step-up was given the value of 1.

Then, the fund average valuation step-up was regressed on fund average similarity score. The results of the regression are displayed in Table IV of Annex 4. The correlation between the two variables seems quite weak: as expected, the regression coefficient is negative (-12.69), but it is not statistically significant at any confidence level; moreover, the R square is slightly higher than 5 percent, which indicates an overall bad fit.

When the background-based similarity score is used, the results are even worse, with the regression coefficient remaining statistically non-significant (p-value 0.583) and the R square decreasing to a modest 0.024.

Interestingly, the scenario improves with the use of the ethnicity-based version of the score. Indeed, even if the regression coefficient is still not significantly different from 0 (p-value 0.166), the overall fit increases to 0.142.

Another (and even more approximative) way to check whether the similarity score is connected to fund performance consists in comparing the average scores of the best and the worst performing AMC[19]. If affinity bias negatively impacts performance, the average similarity score of the best performing AMC should be lower than the one of the worst performing AMC, and the difference between the two should be statistically significant.

[19] This is the only analysis of paragraph 7.2 that is done at AMC rather than at fund level. This choice is motivated by the willingness to avoid excessive sample fragmentation, which would have made the results of the test difficult to interpret.

This paper uses a “quantitative” and a “qualitative” criterion to identify the best and the worst AMC in the sample: the former defines the best (worst) performing AMC as the one with the highest (lowest) average valuation step-up, while the latter concentrates on the number of exits (for the best AMC) and write-offs (for the worst one).

Evidently, both methods present some limitations: the quantitative criterion is impacted by the lack of data on companies’ pre-money valuations, which makes it difficult to compute valuation step-ups; conversely, applying the qualitative criterion exposes to mistakes when identifying the best AMC (since not all exits are successful) and the worst one (since data on write-offs are often opaque).

That said, 360 Capital Partners results the best performing AMC according to both criteria. A divergent result is obtained for the worst performing AMC, with Indaco Venture Partners (Xyence) being selected by the quantitative (qualitative) criterion.

Firstly, let us consider the quantitative criterion. As expected, the average similarity score of 360 Capital Partners (best performing AMC) is lower than that of Indaco Venture Partners (worst performing AMC). When the base and the background-based specifications are used, the difference between the two means is not statistically significant. Conversely, the situation changes with the ethnicity-based specification: the mean difference becomes statistically significant at any confidence level (p-value 0.003). This result confirms that the ethnicity-based specification is the most sensitive to performance metrics.

Secondly, let us apply the qualitative criterion. As for the previous case, the difference between the best performing AMC (360 Capital Partners) and the worst performing one (Xyence) is negative. This time, though, it is statistically significant regardless of the specification used.

To summarize, fund performance appears to be somewhat dependent on the affinity bias, even if the relationship is not particularly strong. Obviously, the results obtained are strongly influenced by the lack of detailed information on investments, which makes it necessary to adopt rough and approximative measures of fund and AMC performances.

8. Conclusions and implications for future work

This paper studies the impact of the affinity bias on the investment decisions of Italian AMCs operating in the VC industry.

The results obtained confirm that unconscious mental processes are a non-negligible component of human actions, even for highly professional individuals.

Intuitively, heterogeneity in the bias strength among AMCs is partially explained by the different degree of partners' experience: those with more transactions performed tend to favour more diversity in investments.

Moreover, the affinity bias seems to have a certain influence on funds' performance: the best performing funds are those with the lowest similarity scores, and this result is robust to the three specifications used in the analysis. Interestingly, the highest correlation is obtained when giving a higher weight to the least rational variables of the score (gender and nationality).

The conclusions reached by this paper are subject to improvements brought by future research on the topic.

First of all, the analysis is limited to Italian AMCs, which, as mentioned several times, still operate in a relatively underdeveloped VC market. In this regard, it would be interesting to extend the study to more mature European environments (e.g. the UK, Germany and France) to check whether more structured dynamics can cushion the impact of the affinity bias.

Finally, the similarity score built in this paper directly compares a number of characteristics of partners and founders. Elaborating on Gompers & Wang (2017), the score could be integrated with additional variables measuring the ethnic, educational and professional differences between partners' children and startups' founders. This would allow to verify the relevance of another channel through which the affinity bias could operate, namely venture capitalists' tendency to invest in founders who remind them of their children.

References

Azoulay, P. *et al*. (2020). "Age and high-growth entrepreneurship." *American Economic Review: Insights*, 2(1): 65-82

Bazerman, M. H. & Moore, D. A. (2012). "Introduction to Managerial Decision-making." In: M. H. Bazerman & D. A. Moore *Judgement in Managerial Decision-making, 8th ed.* John Wiley & Sons, pp. 1-13

Chhaochharia, V. & Laeven, L. (2009). "The investment allocation of sovereign wealth funds." *Available at SSRN 1262383*

Gage, D. (2012). "The Venture Capital Secret: 3 Out of 4 Start-Ups Fail". *Wall Street Journal.* <https://www.wsj.com/articles/SB10000872396390443720204578004980476429190> [last access: 28/02/2022]

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

Gompers, P. A. & Wang, S.A. (2017). *And the children shall lead: gender diversity and performance in Venture Capital*. Working Paper 23454. National Bureau of Economic Research

Hunter, W. C. & Walker, M.B. (1995). *The Cultural Affinity Hypothesis and Mortgage Lending Decisions*. Working Paper, n.95-8. Federal Reserve Bank of Chicago

Kahneman, D. & Tversky, A. (1973). "On the psychology of prediction". *Psychological Review*, 80: 237–251.

Kahneman, D. & Tversky, A. (1979). "Prospect theory: An analysis of decision under risk." *Econometrica*, 47(2): 263–291.

Klayman, J., Ha & Y.-w. (1987). "Confirmation, disconfirmation, and information in hypothesis testing". *Psychological Review*, 94(2): 211–228.

Louis, C. (2019). "Unconscious bias: The affinity bias". <https://podojo.com/unconscious-bias-the-affinity-bias/> [last access: 01/03/2022]

Marcelin, J. R. *et al.* (2019). "The Impact of Unconscious Bias in Healthcare: How to Recognize and Mitigate It." *The Journal of Infectious Diseases*, 220(2): S62–S73,

Ministero Italiano dell'Istruzione, dell'Università e della Ricerca. "Raggruppamenti dei corsi di studio per Area disciplinare". <http://attiministeriali.miur.it/media/275438/tabella1.pdf> [last access: 08/02/2022]

Nickerson, R. S. (1998). “Confirmation bias: A ubiquitous phenomenon in many guises.” *Review of General Psychology*, 2: 175–220.

Ross, H. (2008). “Exploring Unconscious Bias”. *Best Diversity Practices*, 2(5)

Stanovich, K. E. & West, R. F. (2008). “On the relative independence of thinking biases and cognitive ability”. *Journal of Personality and Social Psychology*, 94(4): 672.

Wilson, T. D. (2002). “Strangers to ourselves.” *Harvard University Press*.

ANNEX 1 – Additional information on collected data

Table I. List of AMCs (and relating funds) covered in the analysis

	AMC	Fund	Target (€M)[1]	Total raised UtD (€M)[1]	Kick-off date[2]	Expected closing date[3]
1	360 Capital Partners	Nestor 2000	130	130	01-2000	01-2011
2	360 Capital Partners	360 Capital One	100	100	02-2008	02-2019
3	360 Capital Partners	360 Capital 2011	75	75	10-2012	10-2023
4	360 Capital Partners	Robolution	80	80	01-2014	01-2025
5	360 Capital Partners	360 Square	35	35	12-2015	12-2026
6	360 Capital Partners	Poli360	54	54	01-2018	01-2029
7	360 Capital Partners	A+360 Fund	30	30	01-2020	01-2031
8	360 Capital Partners	360 Fund V	150	90	05-2020	05-2031
9	AVM Gestioni	Cysero	100	15	03-2021	03-2032
10	Azimut libera impresa	Italia 500	40	40	01-2020	01-2031
11	Azimut libera impresa	Azimut Digitech Fund	65	65	01-2021	01-2028
12	CDP Venture Capital	Fondo Italia Venture I	80	80	10-2015	10-2026
13	CDP Venture Capital	Fondo Italia Venture II	150	150	04-2018	04-2030
14	CDP Venture Capital	Fondo Evoluzione	200	100	03-2021	03-2032
15	Claris Ventures	Claris Biotech I	30	30	09-2020	09-2031
16	Eureka! Ventures	Eureka! Fund I	40	40	01-2020	01-2031
17	Indaco Venture Partners	Atlante Ventures	25	25	12-2007	12-2018
18	Indaco Venture Partners	TT Venture	65	65	01-2008	01-2023
19	Indaco Venture Partners	Atlante Ventures Mezzogiorno	25	25	04-2009	04-2020
20	Indaco Venture Partners	Atlante Seed	10	10	07-2011	07-2022
21	Indaco Venture Partners	Indaco Ventures Fund I	250	130	06-2018	06-2029
22	Innogest Capital	Innogest Capital I	80	80	05-2006	05-2017
23	Innogest Capital	Innogest Capital II	85	85	09-2015	09-2026
24	Liftt	Liftt	21	21	12-2019	12-2030
25	Lumen Ventures	Lumen Ventures Fund	25	25	07-2020	07-2031
26	Neva	Neva First Fund	250	180	08-2020	08-2032
27	Oltre Impact	Oltre I	8	8	01-2006	01-2017
28	Oltre Impact	Oltre II	43	43	03-2016	03-2027
29	P101	Programma101	67	67	11-2014	11-2025

30	P101	Programma102	103	103	05-2018	05-2029
31	Panakès Partners	Panakes Fund I	100	100	03-2016	03-2027
32	Primo Ventures	Digital Investments	6	6	10-2010	10-2021
33	Primo Ventures	Barcamper Venture	44	44	09-2016	09-2027
34	Primo Ventures	Barcamper Venture Lazio	8	8	08-2019	08-2030
35	Primo Ventures	Primo Space Fund	85	85	09-2019	09-2030
36	Synergo Capital	Sinergia Venture Fund	150	30	03-2021	03-2032
37	United Ventures	United Ventures One	70	70	10-2014	10-2025
38	United Ventures	United Ventures 2	120	120	12-2019	12-2030
39	United Ventures	UV T-Growth	100	100	07-2021	07-2032
40	Vertis	Vertis Venture	25	25	03-2009	03-2021
41	Vertis	Vertis Venture 2 Scaleup	36	36	07-2017	07-2027
42	Vertis	Vertis Venture 3 Technology Transfer	40	40	08-2017	08-2027
43	Vertis	Vertis Venture 4 Scaleup Lazio	8	8	03-2019	07-2027
44	Xyence (former Principia)	Principia Fund	25	25	06-2005	06-2016
45	Xyence (Principia)	Principia II	64	64	06-2009	06-2020
46	Xyence (Principia)	Principia III - Health	206	206	12-2014	12-2025

Notes

1. Data on target amount to raise and commitment UtD have been obtained from PitchBook, press releases and AMCs' websites.

2. Kick-off date is assumed to coincide with first closing date. Data have been obtained from PitchBook, press releases and AMCs' websites.

3. Unless fund length was explicitly found (e.g. on AMC's website), expected closing date has been computed assuming a total fund life of 11yrs. This results from investment period of 5yrs, portfolio management & divestment period of 5yrs and grace period of 1yr.

Table II. Attribution of sectors based on main business vertical

Sectors	Verticals			
Digital	Drug Delivery[1]	E-Commerce	Marketplace	Mobile
	Printing Services	Second Hand		
Education & HR	Dental Education	EdTech	HR Tech	
FinTech	Accelerator	Banking	Crowdfunding	Cryptocurrency/ Blockchain
	FinTech	InsurTech	LegalTech	Payments
Food & Agriculture	AgTech	E-Grocery	Food and Beverage	Food Delivery
	FoodTech	Restaurant Technology		
Healthcare & Biotech	Diabetes	Digital Health	Health Services	HealthTech
	Life Sciences	Medical Device	Nanotechnology	Oncology
Media	AdTech	AudioTech	Marketing Tech	Phototech
	Price Comparison	Publishing	TMT	
SaaS & Software	Application Performance Management	CloudTech & DevOps	Customer Service	Cybersecurity
	eSports	Event Management	Gaming	Mobile Apps
	SaaS	Social Impact[2]	Sport Management App	
Smart City	Autonomous cars	CleanTech	Cycling	Delivery
	Green Energy	Home Rental	Mobility Tech	Real Estate Technology
	Smart Cities	Storage	Supply Chain Tech	Travel
Tech	3D Printing	Advanced Manufacturing	Artificial Intelligence & Machine Learning	Augmented Reality
	Big Data	Engineering	Industrials	Internet of Things
	Manufacturing	Materials	Oil & Gas	RFID
	Robotics and Drones	Security	Space Technology	Virtual Reality
	Wearables & Quantified Self			

Notes

1. Pharma Prime S.r.l. is the only startup belonging to this business vertical. The attribution of the vertical to Digital instead of Healthcare & Biotech offers a better representation of the startup's business model.

2. Mygrants S.r.l. S.B. is the only startup belonging to this business vertical. The attribution of the vertical to SaaS & Software offers an appropriate representation of the startup's business model.

Table III. Attribution of deal stage based on round series

	Investment Stage	Investment Type
1	**Acceleration/Incubation**	Acceleration
2	**Early Stage VC**	Pre-seed, Seed, Series A
3	**Later Stage VC**	Series B, Series C, Series D, Series E

Table IV. Attribution of field of study based on subject of study

Field of study	Subject of study			
Technology & Science	Actuarial & Financial Science	Aerospace engineering	Architecture	Artificial Intelligence
	Astronomy	Astrophysics	Biochemistry	Biology
	Biomedical engineering	Biotechnology	Chemistry	Computer Science
	Data Science	Dental Technician Institute	Economics	Electronics & Computer Science
	Electronics	Engineering	Finance	Genetics
	Imaging Science	Mathematics	Mathematics & Computer Science	Medicinal Chemistry
	Medicine	Natural Science	Neural Systems	Neuroscience
	Optics	Pharmacy	Physics	Psychiatry
	Robotics	Science	Scientific High School	Software
	Sport Science	Statistics	Technical & Commercial Institute	Technical Institute
	Telecommunications	Telecommunications engineering	Tourism	
Human & Social Sciences	Cinematic Arts	Classical High School	Classical Literature	Communication Studies
	Design	Diplomatic Studies	Graphics	International Relations
	Journalism	Languages & Modern Literature	Law	Linguistic High School
	Linguistics	Literature	Marketing & Communication	Media
	Media & Telecommunications	Music	Philosophy	Political Science
	Psychology	Social Science	Sociology	

ANNEX 2 – Founders' and partners' summary stats

Table I. Founders' summary stats, by AMC

AMC	Sample Size	Avg Age Founders	Avg Age at incorporation date	Avg Age at deal's closing date	# Male Founders	Avg Age Male Founders	# Female Founders	Avg Age Female Founders
Indaco Venture Partners	122	51	38	44	111	51	11	53
United Ventures	94	42	33	37	89	42	5	40
Vertis	92	46	36	40	81	46	11	45
Primo Ventures	101	40	33	37	93	40	8	43
P101	101	40	32	35	88	40	13	40
Innogest Capital	87	48	37	41	80	48	7	49
360 Capital Partners	279	42	34	37	259	42	20	42
Xyence (former Principia)	103	55	42	46	95	55	8	55
Liftt	38	47	42	46	33	46	5	54
Panakès Partners	27	53	43	49	24	54	3	46
Claris Ventures	7	50	44	48	6	50	1	48
Eureka! Ventures	19	37	33	36	19	37	0	n.a.
CDP Venture Capital	127	43	36	40	108	43	19	41
Synergo Capital	4	38	31	38	3	40	1	32
Neva	31	47	41	45	30	47	1	39
Azimut Libera Impresa	61	40	34	39	56	41	5	38
AVM Gestioni	2	44	38	43	2	44	0	n.a.
Lumen Ventures	5	42	36	39	5	42	0	n.a.
Oltre Impact	28	52	43	47	22	52	6	50

Note: *the table divides founders based on the AMC that invested in their startup. The number of unique founders (979) does not coincide with the sum of the founders in each AMC (1344) because some startups were financed by more than one AMC.*

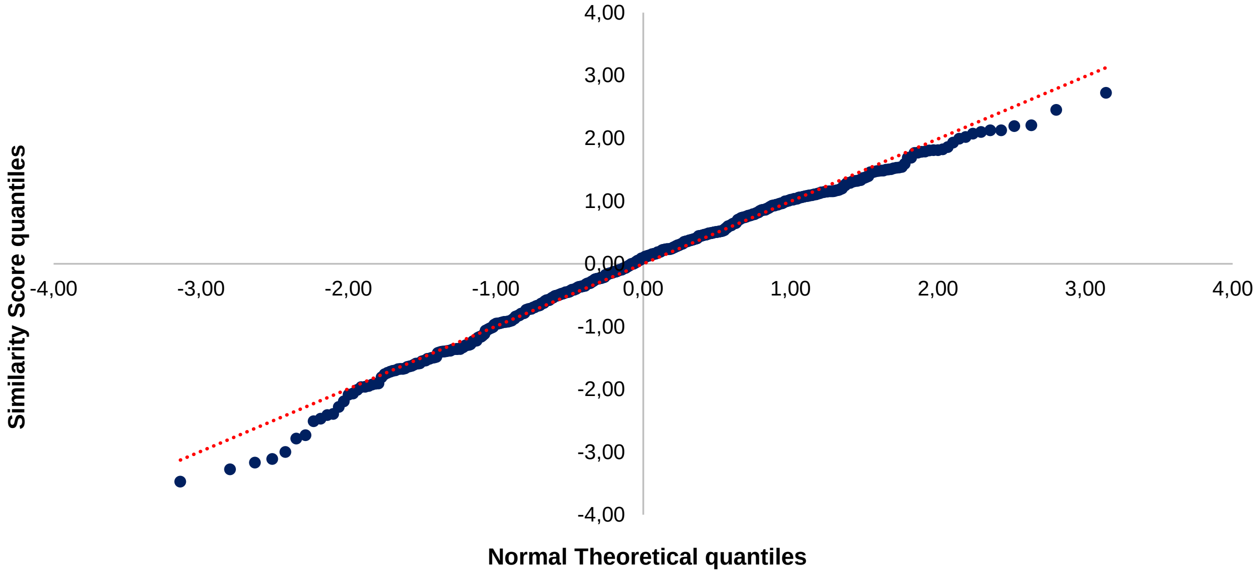
Table II. Partners' summary stats, by AMC

AMC	Sample Size	Avg Age Partners	Avg Age at deal's closing date	# Male Partners	Avg Age Male Partners	# Female Partners	Avg Age Female Partners
Indaco Venture Partners	9	54	47	6	54	3	54
United Ventures	6	47	50	6	47	0	n.a.
Vertis	6	57	53	6	57	0	n.a.
Primo Ventures	6	52	52	6	52	0	n.a.
P101	5	44	44	4	44	1	41
Innogest Capital	11	51	44	11	51	0	n.a.
360 Capital Partners	16	49	45	14	50	2	47
Xyence (former Principia)	9	52	41	9	52	0	n.a.
Liftt	6	55	55	5	53	1	65
Panakès Partners	3	56	55	2	59	1	51
Claris Ventures	4	47	43	3	38	1	71
Eureka! Ventures	4	55	54	3	54	1	57
CDP Venture Capital	5	47	50	2	43	3	49
Synergo Capital	2	47	47	2	47	0	n.a.
Neva	3	48	46	3	48	0	n.a.
Azimut Libera Impresa	3	39	42	3	39	0	n.a.
AVM Gestioni	4	61	60	3	60	1	62
Lumen Ventures	6	38	35	5	35	1	53
Oltre Impact	3	59	56	2	67	1	43

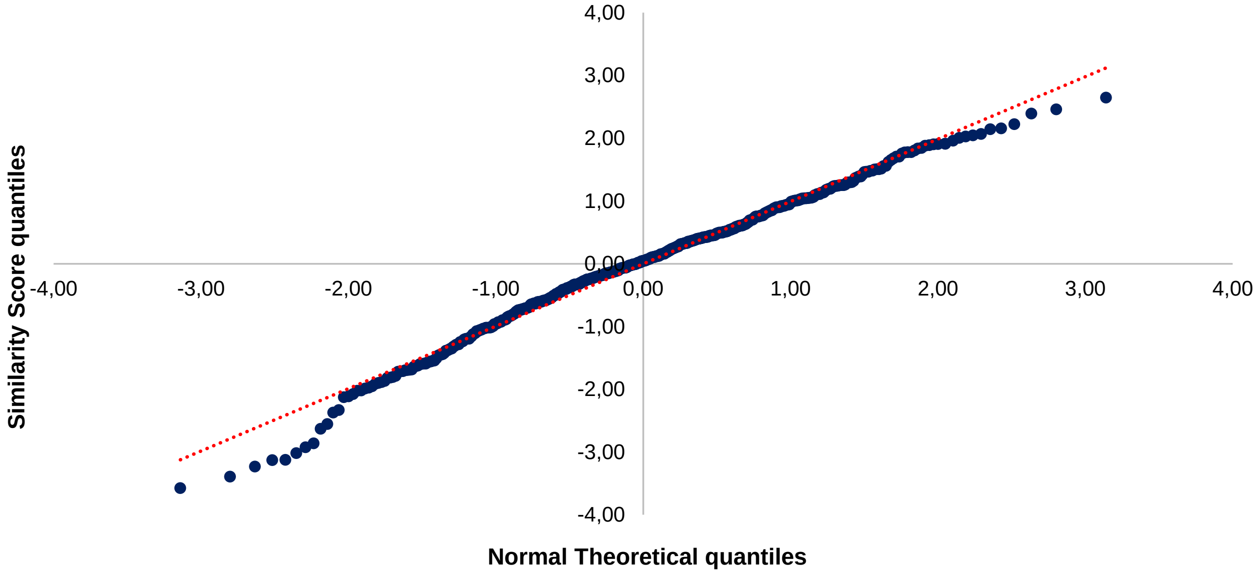
Note: *the number of unique partners (106) does not coincide with the sum of the partners in each AMC (111) because 5 partners overlaps among 2 AMCs*

ANNEX 3 – Tables and graphs on similarity score distribution

Graph I. Similarity Score Q-Q plot - base specification



Graph II. Similarity Score Q-Q plot - background-based specification



Graph III. Similarity Score Q-Q plot - ethnicity-based specification

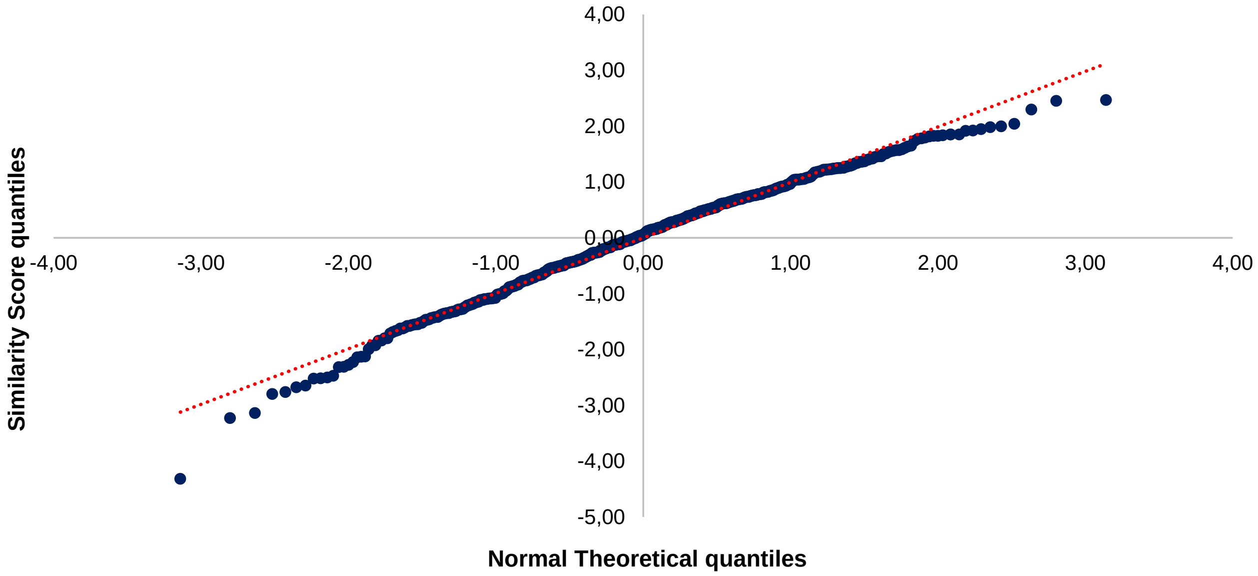


Table I. Similarity score summary stats, by AMC (base specification)

AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,71	0,71	0,86	0,49	0,08	-0,36	0,39
United Ventures	34	0,69	0,67	0,91	0,53	0,09	0,54	-0,20
Vertis	37	0,73	0,71	0,85	0,54	0,08	-0,35	0,01
Primo Ventures	41	0,66	0,65	0,92	0,37	0,11	-0,11	0,45
P101	43	0,64	0,64	0,83	0,33	0,09	-0,71	2,24
Innogest Capital	43	0,75	0,77	0,91	0,52	0,10	-0,51	-0,32
360 Capital Partners	108	0,69	0,70	0,89	0,38	0,09	-0,59	0,59
Xyence (former Principia)	52	0,74	0,75	0,99	0,44	0,12	-0,22	-0,41
Liftt	18	0,75	0,74	0,93	0,61	0,08	0,37	0,87
Panakès Partners	12	0,76	0,77	0,84	0,63	0,06	-0,67	-0,10
Claris Ventures	3	0,71	0,77	0,82	0,56	0,14	-1,52	n.a.
Eureka! Ventures	8	0,62	0,62	0,70	0,54	0,05	0,02	-0,43
CDP Venture Capital	62	0,54	0,53	0,84	0,32	0,10	0,58	0,64
Synergo Capital	2	0,82	0,82	0,88	0,76	0,09	n.a.	n.a.
Neva	13	0,81	0,79	0,96	0,71	0,08	0,53	-0,61
Azimut Libera Impresa	30	0,68	0,70	0,89	0,33	0,11	-1,17	3,55
AVM Gestioni	1	0,92	0,92	0,92	0,92	n.a.	n.a.	n.a.
Lumen Ventures	4	0,74	0,76	0,82	0,62	0,08	-1,28	2,15
Oltre Impact	20	0,64	0,66	0,83	0,29	0,13	-0,88	1,46

Table II. Similarity score summary stats, by financing series (base specification)

Funding Series	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Pre-seed	29	0,71	0,71	0,94	0,53	0,09	0,31	1,10
Seed	247	0,68	0,70	0,99	0,29	0,12	-0,46	0,29
Series A	172	0,69	0,68	0,92	0,37	0,10	-0,13	-0,13
Series B	46	0,66	0,66	0,91	0,32	0,13	-0,39	0,51

Note:

In order to avoid the risk of reaching conclusions based on too small samples, only funding series with at least 25 investments were considered.

Table III. Similarity score summary stats, by sector (base specification)

Sector	Sample Size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Education & HR	26	0,63	0,60	0,86	0,41	0,10	0,43	0,41
Food & Agriculture	27	0,64	0,64	0,89	0,38	0,13	-0,03	-0,21
FinTech	58	0,72	0,74	0,99	0,33	0,13	-0,59	0,17
Media	60	0,67	0,66	0,96	0,45	0,12	0,47	-0,17
Digital	64	0,67	0,70	0,85	0,29	0,12	-1,20	1,71
Smart City	67	0,67	0,68	0,94	0,32	0,11	-0,75	2,15
SaaS & Software	69	0,71	0,72	0,92	0,46	0,10	-0,36	-0,21
Healthcare & Biotech	109	0,70	0,72	0,91	0,42	0,11	-0,47	-0,51
Tech	113	0,70	0,70	0,92	0,46	0,10	-0,22	-0,25

Table IV. Similarity score summary stats, by AMC (background-based specification)

AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,69	0,70	0,84	0,47	0,07	-1,28	2,27
United Ventures	34	0,69	0,69	0,88	0,48	0,08	-0,35	0,63
Vertis	37	0,73	0,74	0,90	0,49	0,09	-0,60	0,08
Primo Ventures	41	0,67	0,67	0,91	0,36	0,11	-0,20	1,54
P101	43	0,67	0,67	0,93	0,32	0,11	-0,68	2,00
Innogest Capital	43	0,76	0,80	0,93	0,53	0,11	-0,63	-0,81
360 Capital Partners	108	0,70	0,70	0,91	0,43	0,09	-0,38	0,70
Xyence (former Principia)	52	0,73	0,73	0,99	0,33	0,13	-0,29	0,88
Liftt	18	0,77	0,77	0,97	0,62	0,08	0,50	2,18
Panakès Partners	12	0,71	0,72	0,78	0,59	0,06	-0,70	-0,27
Claris Ventures	3	0,69	0,72	0,76	0,60	0,08	-1,20	n.a.
Eureka! Ventures	8	0,63	0,64	0,67	0,58	0,03	-0,91	0,53
CDP Venture Capital	62	0,55	0,55	0,81	0,30	0,09	0,04	1,18
Synergo Capital	2	0,79	0,79	0,85	0,73	0,08	n.a.	n.a.
Neva	13	0,78	0,82	0,97	0,62	0,11	-0,06	-1,10
Azimut Libera Impresa	30	0,73	0,76	0,90	0,35	0,12	-1,28	2,16
AVM Gestioni	1	0,92	0,92	0,92	0,92	n.a.	n.a.	n.a.
Lumen Ventures	4	0,75	0,79	0,84	0,59	0,11	-1,58	2,44
Oltre Impact	20	0,67	0,70	0,92	0,28	0,15	-0,55	0,69

Table V. Similarity score summary stats, by AMC (ethnicity-based specification)

AMC	Sample size	Mean	Median	Max	Min	St. Deviation	Asymmetry	Excess Kurtosis
Indaco Venture Partners	62	0,76	0,78	0,90	0,49	0,09	-0,80	0,55
United Ventures	34	0,72	0,76	0,90	0,31	0,11	-1,65	4,20
Vertis	37	0,75	0,76	0,88	0,56	0,08	-0,54	-0,37
Primo Ventures	41	0,71	0,69	0,93	0,53	0,09	0,65	0,25
P101	43	0,75	0,77	0,90	0,58	0,08	-0,39	-0,65
Innogest Capital	43	0,81	0,82	0,94	0,62	0,08	-0,74	-0,18
360 Capital Partners	108	0,72	0,72	0,92	0,47	0,08	-0,19	0,70
Xyence (former Principia)	52	0,78	0,80	0,99	0,42	0,11	-1,05	1,89
Liftt	18	0,79	0,79	0,97	0,62	0,09	-0,20	0,27
Panakès Partners	12	0,76	0,77	0,87	0,55	0,09	-1,18	1,45
Claris Ventures	3	0,78	0,74	0,88	0,72	0,09	1,53	n.a.
Eureka! Ventures	8	0,69	0,70	0,76	0,60	0,06	-0,86	-0,68
CDP Venture Capital	62	0,65	0,65	0,82	0,47	0,09	-0,10	-0,73
Synergo Capital	2	0,83	0,83	0,91	0,75	0,11	n.a.	n.a.
Neva	13	0,81	0,81	0,99	0,65	0,10	0,11	-0,85
Azimut Libera Impresa	30	0,78	0,79	0,93	0,46	0,11	-1,04	1,44
AVM Gestioni	1	0,83	0,83	0,83	0,83	n.a.	n.a.	n.a.
Lumen Ventures	4	0,80	0,82	0,87	0,69	0,08	-1,38	2,60
Oltre Impact	20	0,73	0,73	0,89	0,43	0,12	-0,80	0,65

ANNEX 4 – Outputs of similarity score regressions

Table I. Base similarity score regressed on partners' age at deal date (P_AGE_DEAL)

Regression Statistics	
Multiple R	0,1672
R Square	0,0279
Adjusted R Square	0,0263
Standard Error	0,1119
Observations	589

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,2112	0,2112	16,8768	0,0000
Residual	587	7,3457	0,0125		
Total	588	7,5569			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,8441	0,0386	21,8877	0,0000	0,7683	0,9198
P_AGE_DEAL	-0,0034	0,0008	-4,1081	0,0000	-0,0050	-0,0018

Note

The sample includes 589 observations because the deal closing date of 4 investments was not disclosed, so that it was impossible to compute the regressor.

Table II. Average base similarity score regressed on number of investments (N_INV)

Regression Statistics	
Multiple R	0,4287
R Square	0,1838
Adjusted R Square	0,1358
Standard Error	0,0768
Observations	19

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0226	0,0226	3,8284	0,0670
Residual	17	0,1003	0,0059		
Total	18	0,1229			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,7563	0,0271	27,8630	0,0000	0,6990	0,8136
N_INV	-0,0013	0,0007	-1,9566	0,0670	-0,0027	0,0001

Note

The average similarity score is computed at AMC level.

Table III. Average base similarity score regressed on fund target size (SIZE)

Regression Statistics	
Multiple R	0,1014
R Square	0,0103
Adjusted R Square	-0,0122
Standard Error	0,0692
Observations	46

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,0022	0,0022	0,4573	0,5024
Residual	44	0,2107	0,0048		
Total	45	0,2129			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0,6989	0,0165	42,3648	0,0000	0,6656	0,7321
SIZE	0,0001	0,0002	0,6762	0,5024	-0,0002	0,0005

Note

The average similarity score is computed at fund level.

Table IV. Average valuation step-up regressed on average base similarity score (AVG_SS)

Regression Statistics	
Multiple R	0,2376
R Square	0,0565
Adjusted R Square	-0,0161
Standard Error	3,9580
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	12,1876	12,1876	0,7780	0,3938
Residual	13	203,6521	15,6655		
Total	14	215,8397			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	11,8768	10,2667	1,1568	0,2682	-10,3031	34,0566
AVG_SS	-12,6852	14,3817	-0,8820	0,3938	-43,7550	18,3846

Note

The majority of the investments mapped in the sample was relatively recent, so that for some funds the average valuation step-up was not significantly different from 1. In these cases, this index cannot be considered as a reliable metrics of fund performance, as it simply reflects the status quo *at the investment date. Thus, only the funds with average valuation step-up significantly different from 1 were considered in the regression.*