

Please cite this paper as:

Breschi, S., J. Lassébie and C. Menon (2018), "A portrait of innovative start-ups across countries", *OECD Science, Technology and Industry Working Papers*, 2018/02, OECD Publishing, Paris.
<http://dx.doi.org/10.1787/f9ff02f4-en>



OECD Science, Technology and Industry
Working Papers 2018/02

A portrait of innovative start-ups across countries

Stefano Breschi, Julie Lassébie,
Carlo Menon

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS

The release of this working paper has been authorised by Andrew Wyckoff, OECD Director for Science, Technology and Innovation and by Luiz de Mello, OECD Deputy Director and Special Advisor to the Chief Economist.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s). Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to Directorate for Science, Technology, and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Nick Johnstone provided ideas, guidance and constructive feedback along the whole process of preparation of this report. The authors are also grateful to Dominique Guellec, Caroline Paunov, and Andrew Wyckoff for useful discussions and suggestions; to Steinar Liverud and Invest Europe for support with the identification of government venture capital investors; and to Guido de Blasio and Emanuele Ciani for useful discussions on machine learning algorithms.

© OECD 2018

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

ABSTRACT

The report presents new cross-country descriptive evidence on innovative start-ups and related venture capital investments drawing upon Crunchbase, a new dataset that is unprecedented in terms of scope and comprehensiveness. The analysis employs a mix of different statistical techniques (descriptive graphics, econometric analysis, and machine learning) to highlight a number of findings. First, there are significant cross-country differences in the professional and educational background of start-ups' founders, notably the share of founders with previous academic experience and in the share of "serial entrepreneurs". Conversely, the founders' average age is rather constant across countries, but shows a fair degree of variability across sectors. Second, IP assets, and in particular the presence of an inventor in the team of founders, are strongly associated with start-ups' success. Finally, female founders are less likely to receive funding, receive lower amounts when they do receive financing, and have a lower probability of successful exit, when other factors are controlled for.

TABLE OF CONTENTS

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS.....	2
ABSTRACT	3
A PORTRAIT OF INNOVATIVE START-UPS ACROSS COUNTRIES.....	6
Introduction.....	6
Literature review.....	9
Business dynamics and the role of young innovative firms	9
VC-backed start-ups.....	10
Data and descriptive statistics.....	12
Data sources	12
Descriptive evidence	14
Econometric analysis	28
Start-up characteristics associated with VC funding.....	30
Start-up characteristics associated with successful exit and patenting.....	34
Policy discussion and next steps.....	38
REFERENCES	41
ANNEX 1: COVERAGE AND REPRESENTATIVENESS OF CRUNCHBASE.....	48
ANNEX 2: ADDITIONAL TABLES AND FIGURES.....	52
NOTES	60

Tables

Table 1. Age of founder at the time when the start-up was created	24
Table 2. Descriptive statistics of main variables	29
Table 3. Determinants of VC funding	32
Table 4. Successful exit and post-VC patenting	36

Figures

Figure 1. Number of start-ups by country and sector.....	19
Figure 2. Number of IPOs and acquisitions by country.....	20
Figure 3. Patent intensity	21
Figure 4. Difference between the start-up's founding year and the patent priority date.....	22
Figure 5. Closeness to science	23
Figure 6. Age of founders at the time when the start-up was created.....	24
Figure 7. Student entrepreneurship	25
Figure 8. Previous academic or entrepreneurial experience	26
Figure 9. Share of founders with a Ph.D.....	27
Figure 10. Predictors of receiving one or more rounds of VC funding	33
Figure 11. Machine learning: predictors of the amount of VC funding.....	34

Figure 12. Machine learning: predictors of acquisition	37
Figure A.1.1. Start-ups in digital-related sectors having attracted equity funding in OECD and BRIICS, 2011-16	49
Figure A.1.2. Top digital-related sectors having attracted equity funding, 2011-16	49
Figure A.1.3. Benchmarking Crunchbase against PwC	50
Figure A.1.4. Benchmarking Crunchbase against the OECD Entrepreneurship Financing Database	51
Figure A.2.1. Age of the company at the time in which the patent is filed	53
Figure A.2.2. Closeness to science	54
Figure A.2.3. Age of the founder at the time when the start-up was created.....	55
Figure A.2.4. Student entrepreneurship	56
Figure A.2.5. Share of founders with previous academic experience.....	57
Figure A.2.6. Share of founders with previous entrepreneurial experience.....	58
Figure A.2.7. Share of founders with a Ph.D.....	59

Boxes

Box 1. Visualising “robust” country and sector averages.....	18
--	----

A PORTRAIT OF INNOVATIVE START-UPS ACROSS COUNTRIES

Introduction

Creating the right conditions for innovative start-ups to experiment and thrive is a policy priority across all OECD countries. Recent OECD work has shown that new and young firms contribute disproportionately to job creation across OECD countries.¹ Start-ups may also be more effective in exploiting new technologies and introducing radical innovations, which can help address some of the major policy challenges of our times (e.g., climate change, aging society).² Innovative start-ups can also be instrumental in achieving more inclusive societies by promoting social mobility. Nevertheless, countries, regions, and sectors differ significantly in the degree to which innovative businesses are created, attract resources, and prosper (Calvino, Criscuolo, & Menon, 2016). The enabling factors of innovative entrepreneurship are still being explored, and it is unclear for policy makers which levers can be activated in this domain. While the analysis of business registers data has provided with invaluable insights on young firm dynamics (e.g., Criscuolo, Gal, & Menon, 2014; Haltiwanger, Jarmin, & Miranda, 2013), these studies typically cannot account for initial differences across firms in the ambitions of their founders or their inherent growth potential (Guzman & Stern, 2016).

Therefore, the analysis of venture capital (VC) and related risk-finance instruments data gives an important complementary view, as these instruments play a critical role in fostering the scaling-up of the tiny proportion of start-ups that prove to be successful (Kortum & Lerner, 2001; Samila & Sorenson, 2011). For instance, some estimates show that, while VC funds invest in only around 0.2% of new U.S. businesses, 43% of U.S. public companies founded between 1979 and 2013 were VC-backed, and they accounted for 82% of the total research and development (R&D) expenditure of public companies founded in the same period (Strebulaev & Gornall, 2015). Similarly, Puri & Zarutskie (2012) calculate that the amount of employment generated by VC-backed firms accounts for nearly 10% of employment in the US in the late 1990s and early 2000s, steadily rising from about 5% in the 1980s. The emergence of industries such as semiconductors, biotechnology, and the Internet, as well as the introduction of several innovations across a spectrum of sectors, has been driven in large part by VC investments (Kerr, Nanda, & Rhodes-Kropf, 2014). However, across OECD economies there are striking differences in the relative size of the venture capital (VC) market, with economies like Israel and the United States having a VC-to-GDP ratio ten times as large as the average European country (OECD, 2015).

There is another important reason why the analysis of venture capital for high-growth start-ups is important for policy: governments are among the main VC investors in several OECD countries. For instance, up to 50% of VC-backed start-ups receive some forms of government VC funding in e.g. Korea and Canada, according to some estimates (Brander, Du, & Hellman, 2015). Investments in government VC may be motivated by a number of different objectives, including supporting start-ups with high social returns, or reducing information asymmetries in the VC market (Lerner, 2002). Government VC might also act as a substitute for private VC in the presence of frictions or policy failures in the financial market, or in the case in which the private VC market is not “mature” yet. Irrespective of the ultimate policy objectives of public VC investments, the descriptive analysis presented in this paper may better inform their allocation.³

The analysis of cross-country differences in innovative start-up activity and VC investments has however been hampered by the lack of suitable and comparable data. This paper tries to fill this gap by exploiting innovative micro-data sources to characterize high-growth innovative start-ups across OECD countries. The ultimate aim is providing some evidence on the sources and the enabling factors of high-

growth and innovative entrepreneurship, with a particular focus on the two main pillars of an innovative venture: knowledge and people.

Technological and scientific knowledge is a crucial asset for an innovative start-up. In a knowledge-based society, knowledge accounts for a larger share of the value of increasingly “weightless” goods and services. This is an area where governments play a pivotal role: across OECD countries, universities and public research institutions are major producers of knowledge as the results of substantial public investments. Measuring knowledge is an extremely challenging task; however, in some circumstances data on IP assets have proven to be a valuable – albeit partial – source, given the detailed information available for each individual patent (citations, technological classifications, etc.).

Another crucial asset for start-ups is the people who create them. Knowledge and ideas flow through the economy via individuals; start-up success critically depends on its founders’ experience and incentives. Little is known, however, on who the “start-uppers” are, and whether their profile and characteristics vary across countries, technological fields, and type of start-ups. Of particular interest for policy making is the characterization of specific groups of start-uppers who are highly successful in the market, and/or who promote innovations that fulfil wider social objectives (e.g., environmental technologies or breakthrough innovations).⁴

Furthermore, the analysis of the individual characteristics of start-up founders is relevant for policy because innovative entrepreneurship can promote inclusiveness, which is high in the policy agenda given growing concerns that economic inequality would undermine social cohesion. For instance, there is evidence that innovative entrepreneurship fosters social mobility in the United States (Aghion et al., 2016), while minority communities, particularly those of South/East Asian origin, have played increasingly important roles in US science and technology sectors.⁵ At the same time, the gender gap in entrepreneurship is striking and persistent, with men being on average seven times more likely to be start-up founders than women.

The report first describes in detail the cross-economy and cross-technology differences in a large sample of innovative start-ups, by looking in particular at the start-ups’ knowledge assets and at the profile of their founders. This analysis provides with new, detailed descriptive evidence on VC-backed start-ups across countries. The descriptive analysis shows that there is considerable cross-economy and cross-sector variation in the educational and professional background of start-ups’ founders. For instance, the share of founders with previous experience in academia and PhD education is significantly higher in countries such as Switzerland and Finland, and in the biotechnology sector. While the median age of founders is similar across countries, a few countries and sectors stand out in the share of individuals founding start-ups during their undergraduate or graduate education. There is also significant heterogeneity in start-ups’ patenting activity.

Subsequently, the factors associated with VC funding and successful exit (initial public offerings and acquisitions) are identified through econometric analysis and supervised machine-learning algorithms, which are used to unbundle the distinct effects of correlated factors, and their interactions. This analysis is motivated by the evidence that VC funding has been shown to constitute a critical turning point for start-up growth patterns. Beyond financial backing to entrepreneurs, VC investors help recruit talented managers, formulate new strategies, and use their networks to garner resources for the company (Gompers & Lerner, 2004). Public offers and acquisitions are even clearer indicators of start-up fast growth and success, as typically the company’s evaluation at exit is a multiple of the amount of capital invested during the early stage, and it therefore represents a very strong incentive for founders. The findings of the econometric analysis show that the educational and professional background of founders, their number, as well as the start-ups’ IP assets, are important predictors of the probability of receiving VC funding. The size of the founders’ team, their age, and their personal history as patent inventors are also good predictors of the

amount of funding received, as well as of the probability that the start-up is eventually acquired. On the other hand, everything else being held constant, female founders appear to be significantly less likely to receive funding and, when they do, the amount they receive is substantially lower; the probability of acquisition is also lower for start-ups having at least one woman among the founders.

By focusing on VC-backed entrepreneurship, this work is complementary to other ongoing OECD analyses that either look at aggregate entrepreneurship dynamics, or consider other important dimensions of start-up heterogeneity, like e.g. the size, the sector, or the employment or productivity growth rates.⁶ The project is also complementary to a number of other projects of the OECD Directorate for Science, Technology, and Innovation (DSTI), namely the work of the OECD Committee for Scientific and Technological Policy (CSTP) and the related Working Party of National Experts on Science and Technology Indicators (NESTI) and Working Group on Innovation and Technology Policy (TIP).

Literature review

The work discussed in this paper is related to several different strands of economic, policy, and managerial literature. To streamline the review of these contributions, the Section is organized around two main topics: business dynamics and the role of young innovative firms, and VC-backed start-ups. It is also important to highlight a difference in the use of the term “start-up” across the two main topics: while the first stream of literature typically defines as start-ups all new firms, including the majority of small businesses with limited growth ambitions, the venture-capital literature generally adopts a more targeted and narrow definition of start-ups, which corresponds to a venture-capital backed company with fast-growth ambitions and often with a heavy focus on sector.

Business dynamics and the role of young innovative firms

Recent OECD work has also shown that new and young firms contribute disproportionately to job creation across OECD countries (Criscuolo, Gal, & Menon, 2014; Calvino, Criscuolo, & Menon, 2016). Previous research has also shown that a limited number of successful or breakthrough innovations commercialized by both established incumbents and high-growth start-ups are responsible for a disproportionate amount of job creation and productivity growth (Andrews, Criscuolo, & Menon, 2014; Henderson, 1993; Tushman & Anderson, 1986). Findings from the DynEmp project have also shown that only a tiny proportion of start-ups grow very fast. For instance, on average only 3% of micro start-ups (i.e., new companies created with less than 10 employees) have more than 10 employees five years later. This also consistent with earlier findings by Wong, Ho, & Autio (2005) based on the Global Entrepreneurship Monitor (GEM) 2002.

Combined, the two findings imply that a tiny proportion of start-ups grow very fast and create a disproportionate number of jobs. Nevertheless, countries, regions, and sectors differ significantly in the degree to which innovative businesses are created, attract resources, and prosper (Andrews, Criscuolo, & Menon, 2014; Calvino, Criscuolo, & Menon, 2016). Beyond employment creation, it has also been shown that innovative start-ups contribute substantially to productivity growth (Henderson, 1993; Tushman & Anderson, 1986) and promote radical innovations (Andrews, Criscuolo, & Menon, 2014; Schneider & Veugelers, 2010). Venture capital investors have been shown to play a key role in funnelling resources to high growth start-ups operating in young and dynamic industries (Kortum & Lerner, 2000; Gompers & Lerner, 2002; Samila & Sorenson, 2011). Furthermore, as mentioned in the introduction, there is evidence for the United States that, while only a very tiny proportion of newly created businesses (around 0.2%) are VC-backed in their first years of activity, those account for a disproportionate share (more than 40%) of companies that become public (Strebulaev & Gornall, 2015).

Recently, there has been growing attention on the economic risks associated with a persistent downward trend of entrepreneurship. On the one hand, a number of recent empirical contributions based on administrative data have shown that entrepreneurship as a whole has been declining before and in the aftermath of the 2007-9 crisis. This is particularly evident in the United States, where a “secular” decline in business dynamism and new firm creation since the 1970s has been observed (Decker et al., 2016); the decline in entrepreneurship and business dynamism has also been observed – over a shorter period of time – in several other OECD countries (Blanchenay et al., 2017a). On the other hand, more recent data on venture capital (VC) funding suggest that VC-funded entrepreneurship has been booming over the last couple of years, with the total amount of VC funding granted across OECD countries in year 2015 being 50% higher than in 2007 (OECD Entrepreneurship Financing Database). In the United States, there is also evidence of an upward trend since year 2014 in the number of so-called “unicorns,” i.e., private start-ups valued at \$1 USD billion or higher.⁷ This is also consistent with the analysis of (Guzman & Stern, 2016), showing that the number of high-quality start-ups has not declined in 15 US States in the period 1988 to 2015.

Beyond measurement issues, this mixed evidence also reflects the inherent complexity of the entrepreneurship concept and the definitory differences around the term “start-up” discussed above. While a small number of entrepreneurs play a “transformational” role in society, the majority of individuals engaged in new business creation typically belong to the category of the so-called “subsistence” entrepreneurs (Schoar, 2010). This is also consistent with a recent working paper by (Ng & Stuart, 2016). The authors analyse the career histories of around two million individuals in the U.S. showing that the group of individuals possibly classifiable as “entrepreneurs” actually comprises two distinct clusters of individuals (which they name “hobos” and “highflyers”) that have diametrically opposed characteristics, namely in terms of social positions and career pathways.

VC-backed start-ups

The analysis of innovative, fast-growing young firms is therefore strongly intertwined with the economic and managerial literature on VC-backed start-ups. Indeed, there is clear evidence that in the United States, where the VC market is operating at large scale since at least two decades and therefore the middle- and long-term repercussions on the economy can be assessed, VC-backed firms have represented the backbone of the disproportionate contribution of start-ups to job creation, innovation, and productivity growth. For instance, (Puri & Zarutskie, 2012) calculate that the amount of employment generated by VC-backed firms accounts for nearly 10% of employment in the US in the late 1990s and early 2000s, steadily rising from about 5% in the 1980s. Beyond financial backing to entrepreneurs, there is also evidence that VC investors help recruit talented managers, formulate new strategies, and use their networks to garner resources for the company (Gompers & Lerner, 2004). Furthermore, the backing of venture capitalist firms act as a signal of the quality of a new venture (Brav & Gompers, 1997; Brav & Gompers, 2003; Carter & Manaster, 1990; Stuart, Hoang, & Hybels, 1999). After closing a VC deal, start-ups typically experience sales growth, initial public offerings, and acquisitions (e.g., Arikian & Capron, 2010; Gulati & Higgins, 2003; Lee, Lee, & Pennings, 2001; Ozmel, Reuer, & Gulati, 2013; Pollock & Gulati, 2007; Ragozzino & Reuer, 2011). Hellmann & Puri (2000) and (2002) show that VC backed companies aim at more radical innovations, are significantly faster in introducing their products to the market, and pursue more aggressive market strategies than other start-ups.

Several papers in the economic and entrepreneurship literature explore the factors affecting the likelihood of receiving VC funding or of achieving a successful exit. Among the factors that appear to be associated with VC funding, the most recurrent ones are the professional experience of the founders, and in particular prior firm-founding experience (Zhang, 2011; Burton, Sørensen, & Beckman, 2002; Hsu, 2007); doctoral education (Hsu, 2007); and personal financial resources (Chandler & Hanks, 1998). The policy environment has also been shown to play a role: e.g., Criscuolo & Menon (2015) analyse micro-data on venture capital investments in the green sector across 29 countries over the period 2005–2010 to identify the role that environmental policies play in explaining observed cross-country differences.

A few studies also analyse the factors correlated with the probability of exit (going public or getting acquired). The prominence of the VCs affiliated with the new venture, the number of VCs investing in the company, as well as the timing, duration, and magnitude of their investments in the new ventures all positively affect the exit probability (Ragozzino & Blevins, 2016; Shane & Stuart, 2002). There is evidence that VC funding is associated with higher takeover premiums when companies become acquisition targets (Ivanov & Xie, 2010). Patenting activity also appears to be positively correlated with successful exit (Mann & Sager, 2007). Studies that look at the effect of entrepreneurs' characteristics find mixed results: Williams (2013) find that entrepreneurs' human capital is associated with successful exit, while Lee & Lee (2014) argue that amount of effort exerted by the founder is more important. Finally, Brau, Francis, & Kohers (2003) argue that IPO and acquisitions are very different exit outcomes. In particular, they show that the choice between these two types of restructuring is driven by industry characteristics (concentration, average market to book ratio, sector), financial markets specificities (cost of

debt, state of the IPO market), firms characteristics (firm size, insider ownership percentage) as well as deal-specific factors.

The empirical analysis presented in this report also builds on previous contributions looking at the relationship between venture capital funding and patents. The theoretical and empirical literature has generally recognized the existence of a positive correlation between patenting activity and the amount of VC funding (e.g., Hall & Ziedonis, 1979; Mann & Sager, 2007; Häussler, Harhoff, & Müller, 2012; Cockburn & MacGarvie, 2009). A growing stream of the literature highlights the role of patents as signals of quality and commercial viability (e.g., Baum & Silverman, 2004; Engel & Keilbach, 2007; Hsu & Ziedonis, 2013; Conti, Thursby, & Thursby, 2013). Furthermore, patents can act as a form of collateral of firms' intangible assets (Amable, Chatelain, & Ralf, 2010) increasing the "salvage value" of funded companies, in case the investment proves to be unsuccessful.

A few contributions look instead at aggregate VC data across countries and sectors to identify the policy and framework condition variables that are associated with higher investments (Gomes Santana Félix, Pacheco Pires, & Gulamhussen, 2013; Gompers & Lerner, , 1998; Jeng & Wells, 2000; Schertler, 2003; Romain & Van Pottelsberghe, 2004; Da Rin, Nicodano, & Sembenelli, 2006). Corporate and capital gain tax regimes, stock markets regulations, volatility of interest rates, and employment protection legislation appear to play a role in attracting VC investments.

However, most of the empirical, micro-data studies mentioned above typically rely on a relatively small database, limited to a specific technological sector or country, and usually focus on one dimension of the relationship between VC funding, patenting activity, founders characteristics, and exit events. The main novelty of the stream of work initiated with this report steams from the analysis of a large micro-level database encompassing many countries and sectors, and containing detailed information on start-up financing, their founders, and their IP portfolio. This new database makes it possible to study the different determinants of successful start-ups together and identify the most important factors, as well as drawing nuanced conclusions for different countries and sectors.

Data and descriptive statistics

Data sources

The empirical evidence presented in this paper builds on substantial work to collect, classify, and integrate different micro-data sources. For instance, job and education title lists originally contain tens of thousands different entries and have been grouped into meaningful clusters using supervised machine-learning techniques. Similarly, the matching of the start-up and founder repositories with patent data also required developing an ad-hoc procedure. While this technical work is only briefly mentioned in this Section, the interested reader can find more information in the background papers (Tarasconi & Menon, 2017; Dalle, den Besten, & Menon, 2017) or by contacting the authors of this report. The rest of this subsection gives a general overview of the main data sources that have been used for the empirical analysis.

Data on start-ups, start-uppers, investors, and deals

The main source of data for start-ups and VC activity is Crunchbase.⁸ Crunchbase is a commercial database on innovative companies maintained by Crunchbase Inc., an innovative start-up in itself, located in California, US. The database was created in 2007 but its scope and coverage has increased significantly over the past few years. As reported by Kaufmann Foundation, the database is increasingly used by the venture capital industry as a “the premier data asset on the tech/startup world”.⁹ Dalle, den Besten, & Menon (2017) present a detailed discussion of the database and its potential for economic, managerial, and policy-oriented research. Compared to commercial databases covering similar information and frequently used for economic research (see e.g. Da Rin, Hellman, & Puri, 2011, for an overview of available data sources), Crunchbase has major advantages. It is free of access for academic research (conditional on applying for a license and on complying with the terms of use); is partially crowd-sourced, i.e., users can add and revise contents, which add to the comprehensiveness and timeliness of the database; is updated on a daily basis; contains cross-linked information on companies, their funders, and their staff; and it is structured in an accessible way. Furthermore, it lists both companies that have received VC and start-ups that have not been funded yet but that are presumably actively looking for funding, and thus permits a meaningful comparison between both types of firms.

Consequently, academic interest in Crunchbase has recently grown and research using this database has been published in major journals. Examples include (but are not restricted to) Alexy, Block, Sandner, & Ter Wal (2012), Bertoni & Tykvová (2015), and Block, Fisch, Hahn, & Sandner (2015). For a more detailed literature review, see Dalle, den Besten, & Menon (2017), who discuss more than 80 academic studies in the field of economic, managerial, and entrepreneurship research based on the Crunchbase data.

In the version used for this report, downloaded in January 2017, the database contains information on more than 490 000 distinct entities located in 199 different countries; 447 000 are companies, while the remaining ones are VC investors, schools, or business groups. Of those 447 000 companies, 137 000 report a founding year later than 2010, and 210 000 later than 2005. The database started being populated in May 2007, and for every company the date in which the related record was created is reported. The pace of new record creation was however rather limited until the beginning of 2013, when it stabilized at around 260 records per day, on average, with the exceptions of two dates in 2013 and 2014, respectively, when around 6 000 and 8 000 records were added, probably as the consequence of the acquisitions of additional sources. Therefore, the historical dimension of the database is mainly limited to the snapshot of companies that have been active until recently. The raw data are obtained through two main channels: a large investor network and community contributors. More than 3 000 global investment firms submit monthly portfolio updates to Crunchbase, in exchange of free data access. In

addition, around 500 000 executives, entrepreneurs, and investors contribute to update and revise Crunchbase company profile pages. This wealth of data is processed by the Crunchbase analyst team with the support of artificial intelligence (AI) and machine learning algorithms, in order to ensure accuracy and scan for anomalies. Additionally, algorithms continuously search the web and thousands of news publications for information to enrich profiles.

A brief overview of the database is presented here, while Annex 1 discusses the coverage and representativeness of the database, compared to some benchmark data sources that are more commonly used in the literature. The general message of the benchmarking exercise is that Crunchbase has a better coverage of VC deals and start-ups than comparable data sources. The country-year comparison with aggregated sources on VC investments also suggests that the coverage of Crunchbase is sufficiently exhaustive across OECD member countries and four large emerging economies (Brazil, People's Republic of China (hereafter 'China'), India, Russia), with few exceptions. It is recognized that micro-level databases on VC all suffer from some sample selection issues (Da Rin, Hellman, & Puri, 2011) and CrunchBase is not an exception. However, Crunchbase is quickly becoming a reference for professionals seeking to invest in start-ups, and young firms have strong incentives to appear in the database. Therefore, the results presented in this report can be generalized to start-ups that are actively looking for funding opportunities.

Companies are classified into 45 different economic activity groups, which henceforth are referred to as "sectors". This classification does not appear to follow any major industry classification system, but rather to be especially customized to the start-up world and to emerging sectors. It is however possible to benchmark the Crunchbase classification against the NACE rev. 2 industry classification in the sample of 27 000 companies for which it is possible to find a unique correspondent in the ORBIS database with the same name and country code. The comparison shows that Crunchbase categories are overall meaningful, as the distribution of 2-digit NACE codes within categories show a fair degree of concentration. At the same time, the benchmarking also shows that in 37 out of 45 categories the relative majority of companies is classified either in the "Computer programming, consultancy and related activities" NACE sector (code 62) or in the "Manufacture of computer, electronic and optical products" one (code 26). This may suggest that the NACE classification may not fully capture the technological diversification of new start-ups, at least at 2-digit level.

Crunchbase also contains around 580 000 records on people who are connected to at least one company listed in the database. The following variables are reported: the full name, location (city and region), gender, job title, and the dates on which the record was created and updated, respectively. Most people are classified as founder, co-founder, or CEO. In addition, the database also contains two linked tables reporting the education and the employment history, respectively, of the listed individuals. Whenever feasible, this information has been complemented and cross-validated with data taken from Breschi et al. (2017). While education and employment history is not available for the full sample of listed individuals, the data allow to analyse the "curriculum vitae" of approximately 130 000 people listed as founders or managers of more than 25 000 start-ups.

Furthermore, the database covers 230 000 VC deals, 11 000 IPOs, and 34 000 acquisitions. The table on investors cover 50 000 entities, which are classified by their country code and their "type" (e.g., investment bank, business angel, incubators, etc.). This latter classification enables to identify government venture capital (GVC) funds, incubators, and other public investors. The list of GVC investors has been further refined using several additional sources: the 2012 OECD Financing Questionnaire (Wilson & Silva, 2013), the membership list of InvestEurope (the European association of VC investors), and a list of government VC funds compiled manually by the authors. These three lists have been combined and matched to the main database with a fuzzy matching procedure on investors' names in each country.¹⁰

Data on patents

The “EPO worldwide PATent STATistical Database” (PATSTAT) database developed by the European patent Office (EPO) in cooperation with WIPO, OECD and Eurostat covers 100 million patents from 90 patent authorities. PATSTAT is distributed twice a year in a set of 30 tables that create a relational database, allowing advanced, large scale statistical analysis of patent data.

Tarasconi & Menon, 2017 describe in detail the procedure used to match Crunchbase with information on intellectual property (IP) contained in PATSTAT. While other scholars have matched Crunchbase with IP data for specific subsamples of the two databases, this is the only matching exercise that covers the entirety of the Crunchbase. The match covers both companies and inventors. Given that neither administrative nor other unique identifiers are available in either of the two databases, the matching is based on a “fuzzy” procedure that exploits the available overlapping information across the two databases: the company names, their location, and the names of the people linked to them. The matching procedure needs to be carefully designed in order to maximize the number of correct matches, while at the same time minimizing both “false positive” and “false negative” errors. This is not straightforward as the spelling of companies’ and people’s name is not always consistent across the two databases. Furthermore, in the PATSTAT database there is neither a unique internal identifier for either applicants or inventors, which have therefore to be adequately disambiguated before engaging in the matching exercise. This is particularly critical for inventors, where homonymy is very frequent (the so-called ‘John Smith’ problem). The matching exercise is limited to patents applied for at the five main patent offices (IP5): the the European Patent Office (EPO); the Japan Patent Office (JPO); the Korean Intellectual Property Office (KIPO); the State Intellectual Property Office of the People’s Republic of China (SIPO); and the United States Patent and Trademark Office (USPTO).

Almost 50 000 companies, out of the 447 000 listed in Crunchbase in January 2017 (excluding venture capital companies), are found to own one or more patents, for a total of around 12 million patents. Around 220 000 of those have been applied for by companies created after 2005.¹¹ The share of patentees for US companies is 15%, but the share doubles for companies reporting at least one funding round. Regarding individuals, out of the 578 000 professionals listed in Crunchbase who could be potential patent inventors, around 25 000 are found to have a corresponding matched individual in PATSTAT. These inventors account for 2,2 million patent applications. More evidence on start-up patenting activity is discussed in the following of the report.

Descriptive evidence

This Section presents and discusses some general descriptive evidence of the database of start-ups under investigation, with a particular attention to the main subjects of analysis of this report: the start-ups intellectual assets, and the people who created the start-ups (founders). The graphs generally report simple averages across companies, patents, or individual founders, aggregated by country or sector. In order to report statistics that are based on a sufficiently large number of observations, the graphs on founders are limited to countries and sectors for which at least 200 observations are available. Depending on the specific variable under scrutiny, the sample of countries and sectors can therefore slightly change accordingly. In the graphs reporting company or patent level information, for which the number of observations is larger, the sample is limited to the top 20 countries or sectors in number of observations (individual companies or patents) to ease visualization.

Furthermore, the graphs reporting average values across countries or sectors can partially reflect composition effects. For instance, the share of VC-backed start-ups varies across countries, and this may affect the average value of other variables. In order to partial-out the potentially confounding effect of sample composition, Annex 2 reproduces the main graphs reported here using a simple econometric

technique that calculates “robust” country and category averages via a linear regression model. This procedure is explained in Box 1.

The interested reader can therefore cross-check the robustness of the core messages emerging from the descriptive analysis by comparing the simple graphs reported here with those derived from regression analysis in Annex 2.

The first group of graphs illustrates the cross-country and cross-sector distribution of the sample of start-ups under investigation. In particular, Figure 1 reports the number of start-ups founded after year 2000 (i.e., that are maximum 15 year old at January 2017, when the database has been built) included in the database, by country and sector group, distinguishing between companies with and without VC backing. Not surprisingly, as it is also confirmed by aggregate VC investments data, it appears clearly that the United States account for the largest share in the sample (around 35% of companies). United Kingdom, India, Canada are the largest contributors after the United States. The distribution across sectors is more even, with retail, internet services, and advertising being the largest group. The upper graph also shows share of companies with VC funding tends to vary across countries, which suggests that country-specific factors may influence the likelihood for a start-up looking for VC to appear in the database. For this reason, all main findings highlighted in this report are cross-validated in the sample of VC recipients only. The share of VC-backed start-ups varies significantly also across sectors, with biotechnology and energy showing a much higher value than the average; this pattern, however, is likely mainly to reflect actual technological differences, rather than measurement “noise”.

Figure 2 reports the number of acquisitions and IPOs in the period 2007 to 2016, by country (upper panel) and by sector (lower panel). The graphs show that the number of acquisitions is one order of magnitude higher than the number of IPOs, although there are sizeable differences across sectors. In “digital” sectors like software, data analytics, or digital services, acquisitions account for almost the totality of successful exit events, while in sectors like biotechnology and energy acquisitions are only twice as frequent as IPOs.

The second group of graphs focuses on the start-ups’ intellectual property (IP) assets, as measurable from patent data. Figure 3 shows a simple measure of patent intensity, calculated as the ratio of the number of patents owned by VC-backed start-ups over their total amount of VC funding (in nominal USD billion). The patent number is reported in both unweighted terms and as weighted by the inverse of each patent simple family size.¹² Japan, France, and Netherlands are the economies characterized by the highest number of patents per USD billion of VC. Across sectors, start-ups in the manufacturing sector (which comprises a relatively small number of start-ups) patent with the highest intensity, followed by software and hardware.

Figure 4 reports the frequency histogram of the distribution of the difference between the start-up’s founding year and the patent priority date. In order to have an adequate observation window before and after the founding year, the statistic is calculated on start-ups founded between 2000 and 2005, and still active in 2016. The graph shows that the majority of patents are filed for in the first six years of start-up activity, decreasing smoothly afterwards. A small but non-negligible share of patents is filed for before the company is created. Figure 15 in Annex 2 shows that there are not significant differences across countries and sectors in the average age of the start-up at the time the patent is filed for, once other confounding factors are controlled for.

Figure 5 reports the average “closeness to science” of start-up owned patents. The measure is taken from Squicciarini, Dernis, & Criscuolo (2013) and is calculated as the share of non-patent literature citation over total backward citations; as it is available only for USTPO and EPO patents, the sample is restricted accordingly. The graph shows that patenting start-ups in Belgium, Denmark, and France are

characterized by the greatest closeness to science. The corresponding “robust” graph (Figure 16 of Annex 2) confirms this pattern, but it also shows that a bigger group of economies – which also includes United Kingdom, Switzerland, and Australia – show a statistically higher average value than the US. Across sectors, the biotech sector is standing out as the closest to science, while all other sectors show a remarkably similar value. The case of Switzerland appears particularly interesting, as the country is also characterized by higher shares of founders with a PhD and with a previous academic experience, as it is discussed in the following of the Section. All these metrics therefore point to describing a start-up ecosystem that appears particularly effective in leveraging academic knowledge for innovative entrepreneurship, or in attracting highly qualified and internationally mobile founders.

A number of interesting findings emerge from the descriptive analysis regarding intellectual assets. First, patents are filed at a relatively constant rate in the first years of start-up activity, while the share of patents filed for before the start-up is officially founded is rather small. This suggests that patents are not exclusively linked to the main idea that triggers the birth of the business venture, but rather are the outcome of the start-up activity in the first five-ten years of activity. Second, patent propensity and patent closeness to science show a remarkable degree of heterogeneity across countries and – in the first case – sectors. IT-related sectors, namely software and hardware, are characterized by a higher average patent propensity, second only to manufacturing.

The third group of graphs summarizes the information available on the founders, namely their age and their professional and educational background. Figure 6 reports the box-plots summarizing the distribution of founders’ age across countries, measured as the difference in years between the beginning of the undergraduate education (B.A. or B.Sc.) and the start-up birth. For founders for which information on B.Sc. is not available, the age has been inferred from the start year of other education titles, assuming that the difference with the B.Sc. is equal to the median difference in the sample of founders for which information is available for both titles. The small number of outliers correspond individuals who completed higher education after company founding. The chart shows that there is little variation in the age of the majority of founders across economies. This is also confirmed by Table 1: the median age ranges between 10 and 15 (therefore, the median founder is approximately 28 to 33 year old, depending on the country). However, in some countries, like the United States and United Kingdom, there is larger variability in the age variables, meaning that there are a larger number of younger or older founders.

Figure 7 looks at student entrepreneurship. The phenomenon recently gained interest, spurred mostly by the notorious cases of world-leading IT companies founded by undergraduate students (e.g., Facebook or SnapChat). In this report, student entrepreneurs are defined as individuals who found their start-up within four years since the beginning of the B.Sc.. The upper panel of Figure 7 shows that Canada, Brazil, and India are characterized by the highest shares of student entrepreneurship, while across sectors (lower panel) the gaming and education groups are – not surprisingly – those with the highest incidence of the phenomenon.

Figure 8 reports the share of founders with previous academic and entrepreneurial experiences. Building on findings of Ng & Stuart, 2016, the self-employed group is distinguished from the entrepreneur group. The former group includes job titles that contain the words “self”, “independent”, or “free-lance”, while the latter includes job titles that contain the words “entrepreneur”, “founder”, or “owner”. The academic group contains job titles like “professor”, “post-doc”, “research associate”, etc. Consistently with the finding of (Ng & Stuart, 2016), the share of founders with previous self-employment experience is very marginal across all countries and sectors, and it is not reported in the graph to ease visualization. The previous work experience is considered only if it ends before the company is founded.

Figure 8 shows that between 20 to 30% of founders have one or more previous entrepreneurial experiences, i.e., are “serial entrepreneurs”, with Sweden, Netherlands, Israel, and United States being the

economies with the highest shares. The corresponding “robust” graph in Annex 2 (Figure 20) confirms that for these four countries the positive difference in the share of serial entrepreneurs with the other countries is statistically significant, even when other basic factors are controlled for. The share of academic founders, instead, is much more variable across economies, being lower than 5% in several countries, while reaching 20% in Finland. Again, the “robust” graph of Annex 2 (Figure 19) confirms that cross-country differences are significant, with a group of three countries characterized by a higher value than United States (Finland, Ireland, and Switzerland), and several countries showing a lower value. Across sectors (Figure 8, lower panel), biotech is a clear outlier for the share of academic entrepreneurs, which is close to 30%; it is also characterized by the lowest share of serial entrepreneurs. On the other side of the ranking, less sector-intensive sectors – like financial services and sales and marketing – appear to have higher shares of serial entrepreneurs.

Finally, Figure 9 reports the share of founders with a PhD. Switzerland appears to be the clear outlier in the country group, with around 25% of founders listing a doctorate among their educational achievements. Panel A of Figure 21 in Annex 2 shows that Germany is the only other country for which the share of founders with a PhD is higher than in the United States in a statistically significant way. Across sectors, the differences are even larger, with more than 50% of founders in biotech having earned a PhD.

Summing up, the graphs illustrating the founders’ characteristics uncover some interesting new facts. First, the age of the majority of founders, expressed as number of years since completion of undergraduate education, is similar across countries at 14 years. However, in some countries there is a fair degree of dispersion, with a non-trivial share of founders starting their venture during, or immediately after, undergraduate education. This is reflected in the large variation observed across sectors. This finding contrasts with evidence that the average age at first patent invention is strikingly similar across fields (Jones, 2009). This suggests that in a few specific fields innovative entrepreneurship may not require knowledge accumulation in specific fields, but rather be the domain of talented generalists. This might be in line with the theories of a few sociologists who argue that extensive training in a particular field can impede cognitive insight, with marginal experts in multiple intellectual domains being more likely to introduce creative breakthroughs than well-established specialists in a single domain (see Baumol, Schilling, & Wolff, 2009, for an in-depth discussion).

Another important finding relates to the prominence of “serial” entrepreneurs. Across countries and sectors, the share of founders with previous entrepreneurial experience ranges from 20 to 30%, approximately. The share is consistent with previous findings (Plehn-Dujowich, 2010). The share of founders with previous “self-employment” experience is instead much lower. This is in line with the study of Ng & Stuart (2016), showing that ambitious entrepreneurs and the self-employed belong to two radically different groups, with very few transitions across them. Conversely, the share of founders with previous academic experience or with doctoral education shows a much larger degree of heterogeneity across countries. This is also true for student entrepreneurship, which accounts for a non-trivial share of founders in a few countries (e.g., for Canada, Brazil, Australia, and India the share of “student founders” is higher than 10%).¹³ This suggests that different national policies and framework conditions could have an important impact on the level and quality of academic knowledge that is made accessible or flows to innovative new businesses.

Box 1. Visualising “robust” country and sector averages

The graphs reporting average values across different groups (e.g., countries or sectors) can reflect possible composition effects. For instance, Figure 8 shows that Finland is characterized by the highest share of start-up founders who have a previous academic job spell; however, this might e.g. reflect a possible marked technological specialization of Finnish start-ups in sectors like biotechnology, which is also characterized by a very high share of founders with previous academic experience, as shown in the same Figure.

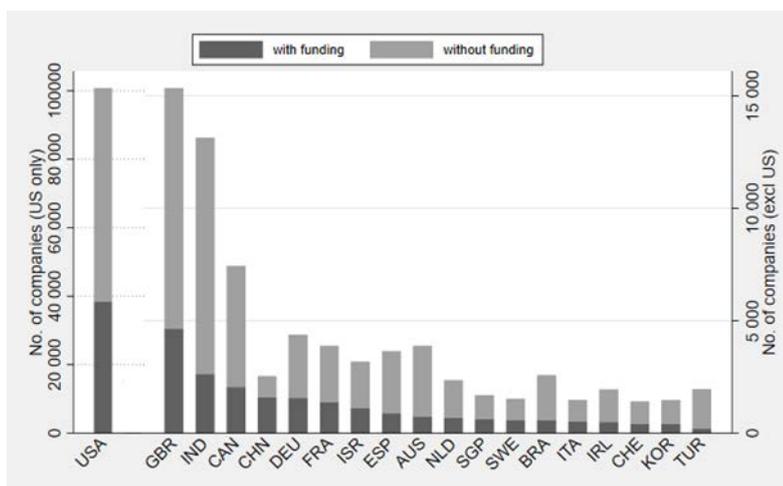
There is an easy way to partial-out composition effects while at the same time preserving most of the immediateness of the graph content using basic econometric techniques. Specifically, it consists in running a simple ordinary least squares regression of the variable under scrutiny with the microdata (e.g., a binary variable indicating whether founders have previous academic experience) on country fixed effects and all the other variables that are expected to play a “confounding” role, like e.g. sector fixed effects, year in which the start-up was founded, the gender, etc. The values of the fixed effects are then visualised in the graph. As the estimation of the fixed effects requires dropping a “baseline” group to avoid the so-called “dummy variable trap” (i.e., perfect collinearity with the constant term), the fixed effect values are expressed in difference from the excluded group, which is therefore mentioned in the graph as the baseline, and which take a zero value by construction. The graphs also report the 90% confidence interval around the point estimates of the coefficients on the fixed effects, which allow to assess whether cross-country or cross-sector differences are statistically significant. The resulting graphs are reported in Annex 2 and mirror those reported in this Section.

These values therefore indicate whether, keeping all other variables constant, each group is statistically different from the baseline group along the specific dimension under scrutiny. For instance, going back to the example discussed earlier, Figure 19 shows that Finland is still the country with the highest share of founders with previous academic experience; however its value is not statistically different from those of Switzerland, Denmark, and Ireland. For all the four economies, the measure is however higher than the value for United States, and the difference is statistically significant.

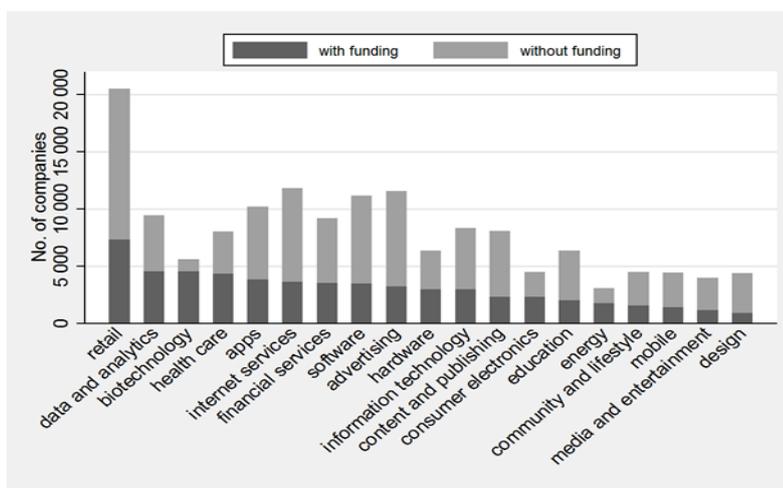
Figure 1. Number of start-ups

With and without VC funding; start-ups founded during the period 2001-16 and still operating in 2016

Panel A: by economy



Panel B: by sector



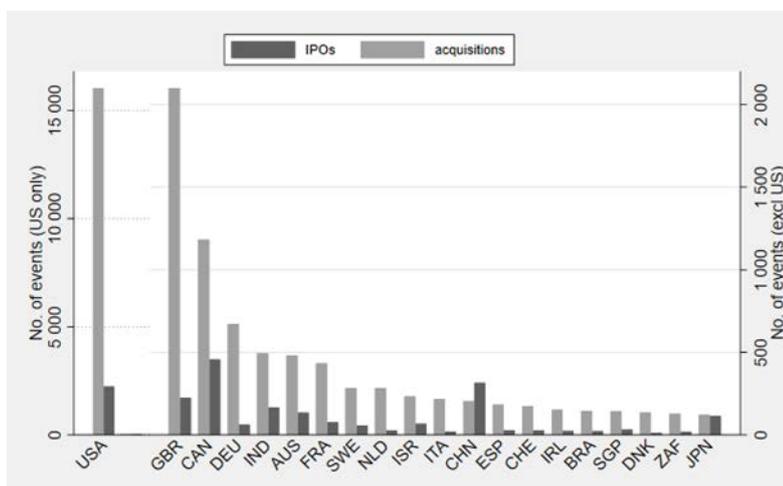
Note: Start-ups with funding are companies for which at least one VC deal is reported. The graphs are limited to the top 20 economies or sectors for number of listed companies. Companies for which the founding date, the country code, or the primary sector group classification is not available are excluded. For panel A, numbers for the US are reported on the left axis while numbers for other economies are reported on the right axis.

Source: Authors' elaboration on www.crunchbase.com

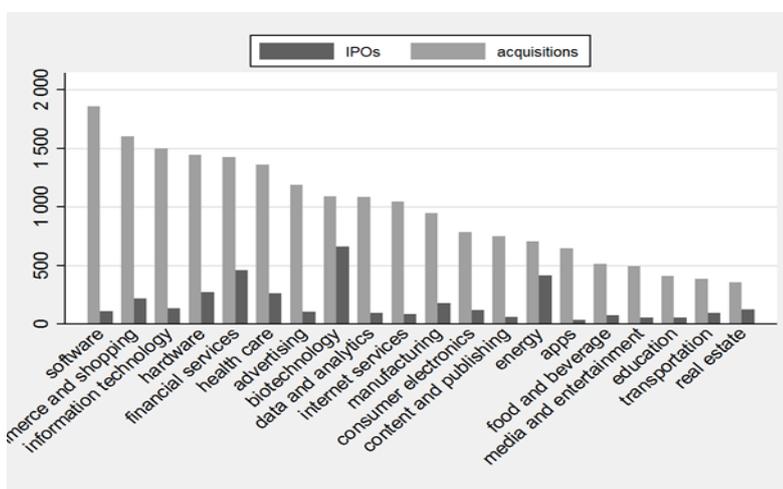
Figure 2. Number of IPOs and acquisitions

During the period 2007-16

Panel A: by economy



Panel B: by sector



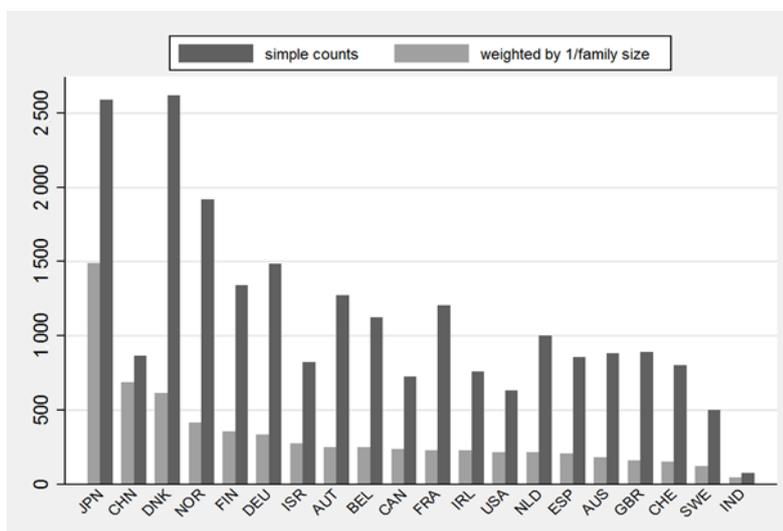
Note: The graphs are limited to the top 20 economies or sectors for number of listed exit events. Companies for which the founding date, the country code, or the primary sector group classification is not available are excluded. For panel A, numbers for the US are reported on the left axis while numbers for other economies are reported on the right axis.

Source: Authors' elaboration on www.crunchbase.com

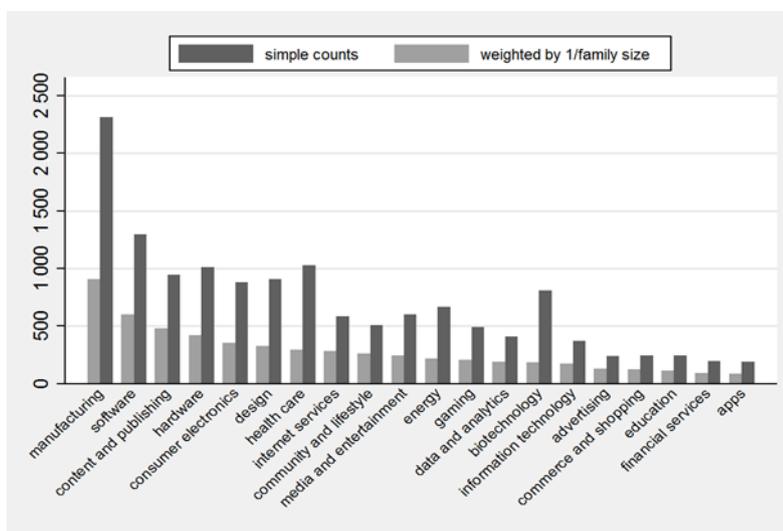
Figure 3. Patent intensity

Number of patents over VC investment (USD billion)

Panel A: by economy



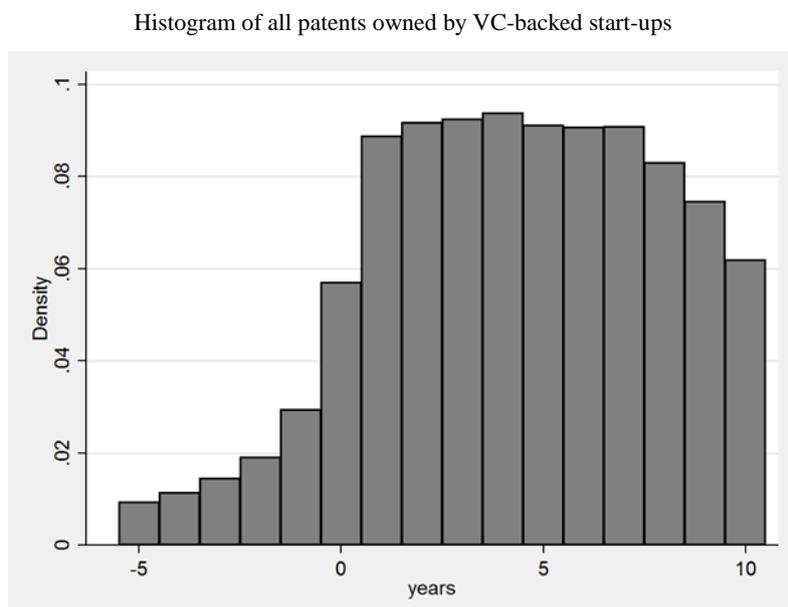
Panel B: by sector



Note: The sample is limited to companies created after 2001 and having received at least one VC investment. The graphs are limited to the top 20 economies or sectors for total amount of VC investments. Companies for which the founding date, the country code, or the primary sector group classification is not available are excluded. A patent family is “a set of patents filed in several countries which are related to each other by one or several common priority filings” (OECD, 2009). A patent family generally comprises all patents protecting the same invention.

Source: OECD elaborations on www.crunchbase.com and PATSTAT

Figure 4. Difference between the start-up's founding year and the patent priority date



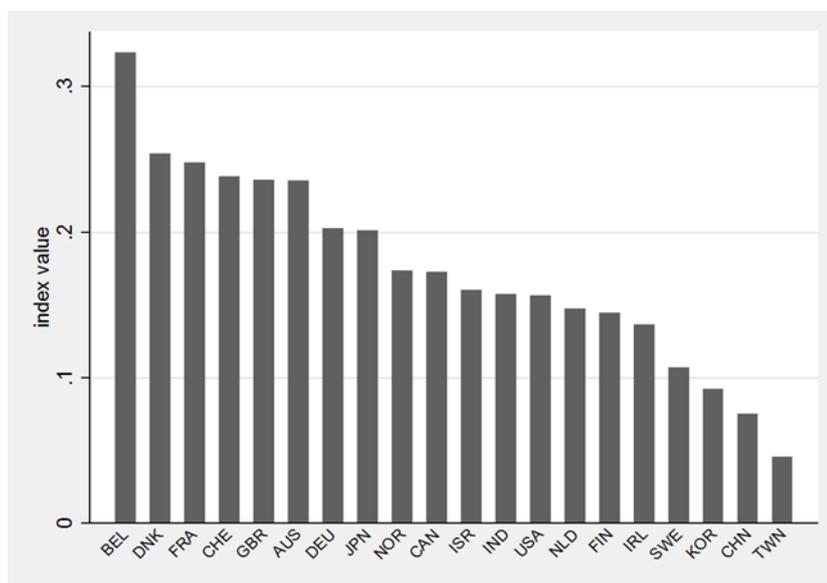
Note: The sample is composed by patents owned by VC-backed companies founded between 2000 and 2005 and that filed a patent application within 10 years of company founding. Companies for which the founding date, the country code, or the primary sector group classification is not available are excluded.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

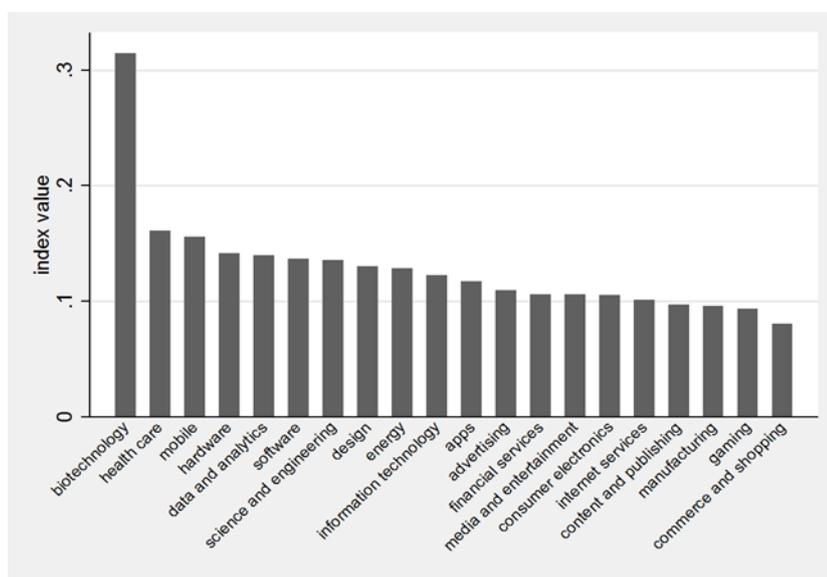
Figure 5. Closeness to science

Average value across all USPTO and EPO patents owned by start-ups

Panel A: by economy

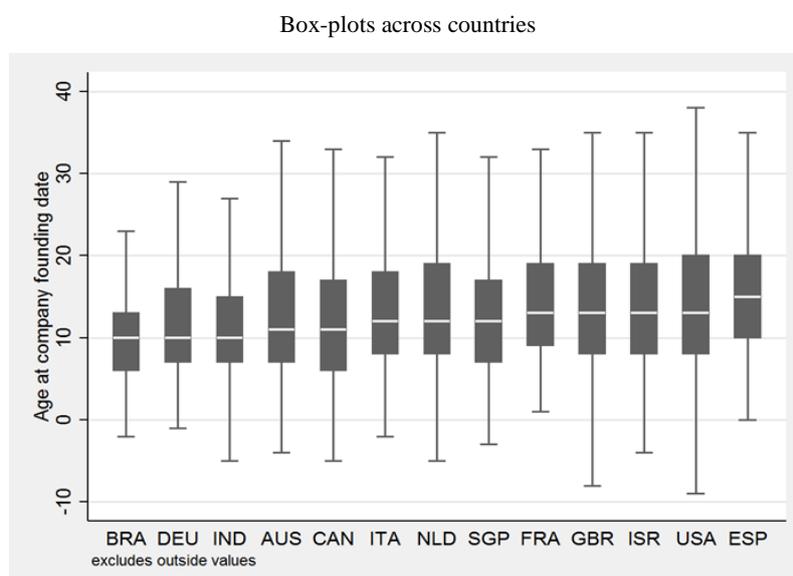


Panel B: by sector



Note: The sample is limited to companies created after 2001. The graphs are limited to the top 20 economies or sectors for total number of patents. Companies for which the founding date, the country code, or the primary sector group classification is not available are excluded. Closeness to science is computed as the average share of citations to non-patent literature in total backward citations (OECD, 2009). A higher index value indicates that the patented sector builds on academic advances in science and is closer to the research frontier.

Source: www.crunchbase.com; PATSTAT; Squicciarini, Dernis, & Criscuolo (2013)

Figure 6. Years since the beginning of undergraduate education at the time of start-up founding

Note: The sample is limited to companies created after 2001 with at least one founder listed in the database. The graphs are limited to economies with more than 200 founders. Companies for which the founding date or the country code is missing are excluded. The box identifies the lower adjacent value (low bar/ whisker below the box), the 25th percentile (lower end of the box), the median (bar inside the box) the 75th percentile (upper end of the box), and the upper adjacent value (bar/ whisker above the box) of countries' average-size-at-entry distribution. For further information and definitions of adjacent and outside values see Cox (2009). Age variable is normalized to 0 at date of entry in B.A. or B.Sc. Therefore this variable indicates the number of years between date of entry in B.A. or B.Sc. (usually around 18, but this varies across individuals and countries) and date of company creation. For founders for which information on B.Sc. is not available, the age has been inferred from the start year of other education titles, assuming that the difference with the B.Sc. is equal to the median difference in the sample of founders for which information is available for both titles. The small number of outliers correspond individuals who completed higher education after company founding.

Source: Authors' elaboration on www.crunchbase.com

Table 1. Years since the beginning of undergraduate education at the time of start-up founding

Descriptive statistics						
	Number of observations	Mean	Standard deviation	Median	1st quartile	3rd quartile
BRA	290	11	6	10	6	13
IND	1758	12	7	10	7	15
DEU	446	12	8	10	7	16
CAN	676	13	8	11	6	17
SGP	222	13	8	12	7	17
AUS	316	13	8	11	7	18
ITA	223	13	7	12	8	18
NLD	317	13	7	12	8	19
ISR	675	14	8	13	8	19
GBR	1388	14	8	13	8	19
USA	12386	15	9	13	8	20
FRA	432	15	8	13	9	19
ESP	495	16	7	15	10	20
Total	19624	14	9	12	8	19

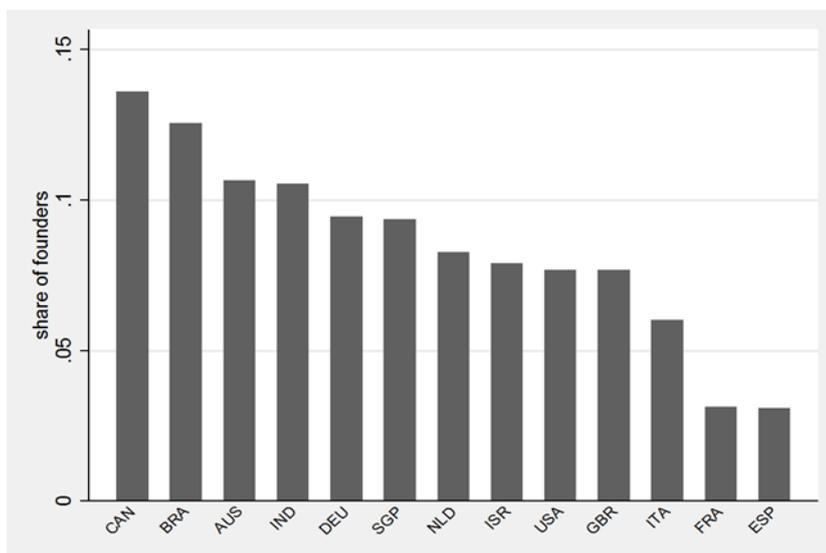
Note: The sample is limited to companies created after 2001 with at least one founder listed in the database. Companies for which the founding date or the country code is missing are excluded. The age variable is normalized to 0 at date of entry in B.A. or B.Sc. Therefore this variable indicates the number of years between date of entry in B.A. or B.Sc. (usually around 18, but this varies across individuals and countries) and date of company founding. The variable is winsorized at 1st and 99th percentiles.

Source: Authors' elaboration on www.crunchbase.com

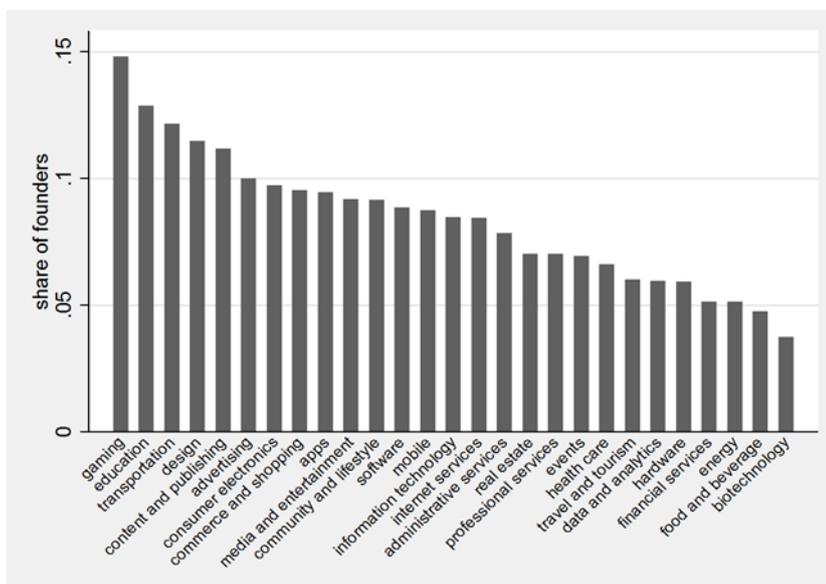
Figure 7. Student entrepreneurship

Share of founders who founded the company within four years since the start of the B.Sc.

Panel A: by economy



Panel B: by sector



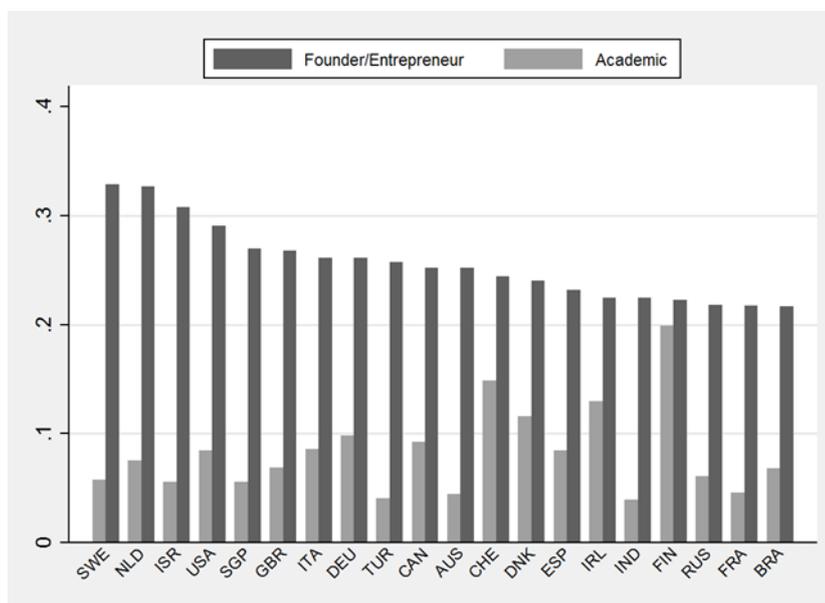
Note: The sample is limited to companies created after 2001 with at least one founder listed in the database. The graphs are limited to economies with more than 200 founders. Companies for which the founding date or the country code is missing are excluded. Student entrepreneurs are individuals founding a company within four years after the start of B.Sc.

Source: Authors' elaboration on www.crunchbase.com

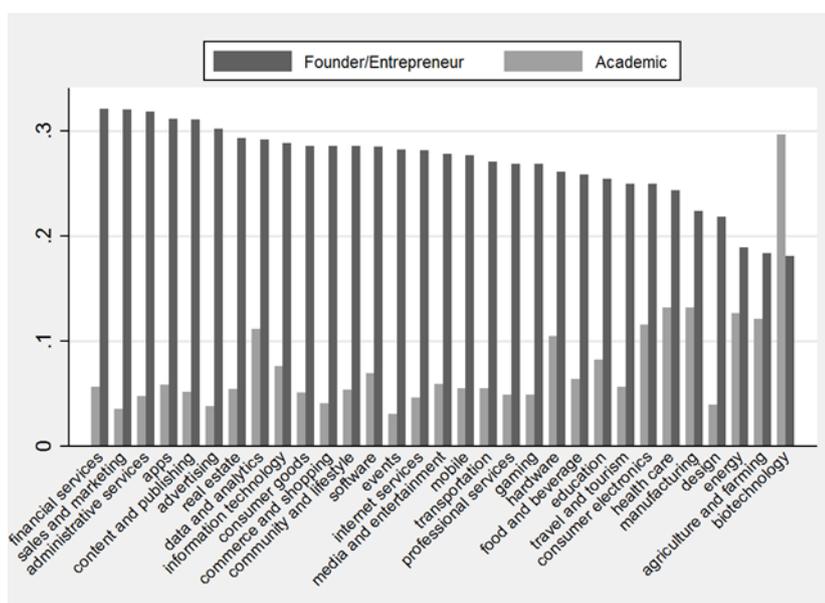
Figure 8. Previous academic or entrepreneurial experience

Share of founders with previous academic or entrepreneurial experience

Panel A: by economy



Panel B: by sector

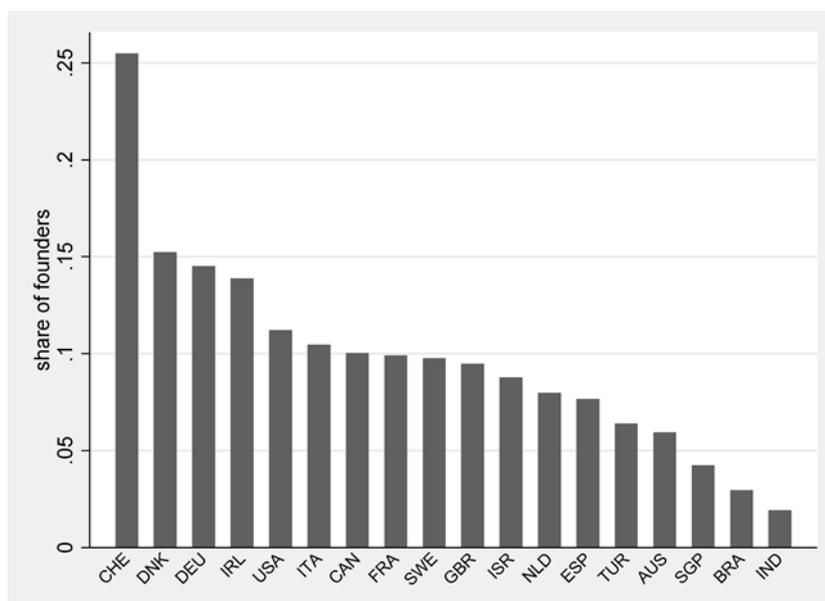


Note: The sample is limited to companies created after 2001 with at least one founder listed in the database. The graphs are limited to economies with more than 200 founders. Companies for which the founding date is missing are excluded.

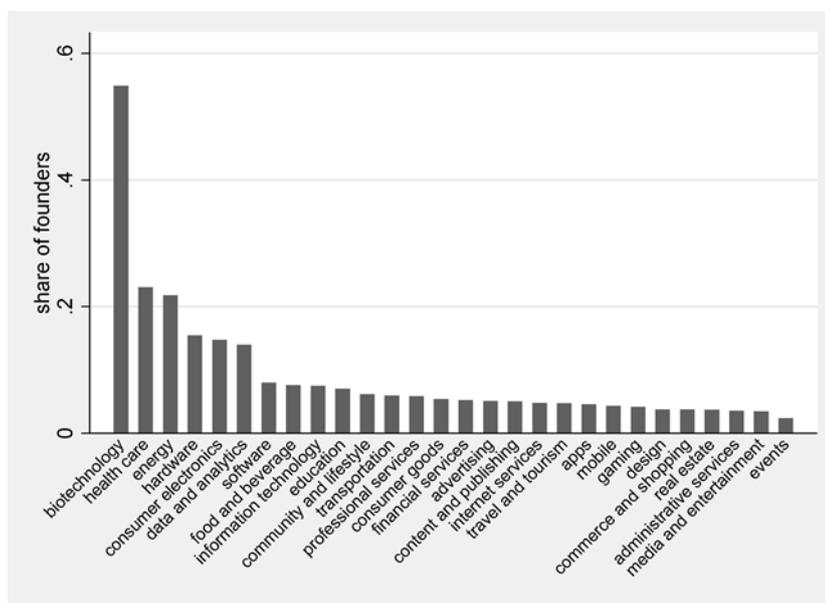
Source: Authors' elaboration on www.crunchbase.com

Figure 9. Share of founders with a Ph.D.

Panel A: by economy



Panel B: by sector



Note: The sample is limited to companies created after 2001 with at least one founder listed in the database. The graphs are limited to economies with more than 200 founders. Companies for which the founding date is missing are excluded.

Source: Authors' elaboration on www.crunchbase.com

Econometric analysis

From the data sources described above, a database containing firms and founders characteristics is created. The sample is restricted to firms founded since year 2001, located in OECD and four large emerging economies (Brazil, China, India, Russia), and for which at least one founder is listed in the database. This first selection leaves 56,311 companies in the sample. Among those, detailed founders' employment history is available for 25,237 companies and education information is available for 25,240 companies¹⁴. The econometric estimates across those three different samples confirms that the main findings on founders' characteristics – such as previous employment and education history – are very similar, thus they do not appear to be driven by sample selection.

Table 2 below reports descriptive statistics of the variables used in regression analyses and machine learning algorithms. All the variables are computed at company level, but the possible heterogeneity within the team of founders is taken into account. The empirical analysis discussed in this Section is based on both traditional econometrics and on more experimental machine-learning techniques. The latter are becoming increasingly popular in economic analysis (see (Mullainathan & Spiess, 2017) for a comprehensive introduction). In this context, the two alternative statistical approaches give complementary sets of results. Traditional regression analysis is used to precisely estimate a number of coefficients on a limited number of variables selected *a priori* based on some theoretical hypotheses. The estimated coefficients allow the calculation of marginal effects and semi-elasticities, e.g., it is possible to state that start-ups with at least one female founder receive on average 70% less funding, everything else equal. However, this approach is not informative on whether the linear combination of selected variables is doing a good job in explaining the phenomenon under scrutiny, compared to possible alternative variable combinations and selections.

Supervised machine learning techniques, instead, are in general designed to solve a prediction problem. Given all the available information on start-ups, the objective is to identify the variables that are the best predictors of the outcome under scrutiny (e.g., probability of getting funding, amount of funding received, probability of being acquired). The algorithm is fed with the largest available set of variables, and the result is a list of variables ranked by importance, i.e., by their explanatory power in the prediction exercise. While the algorithm also estimates variable coefficients, their values are not necessarily consistent, and they should not be used to estimate marginal effects and elasticities.

The major advantage of machine learning techniques lies in their flexibility. Because no functional form is imposed to the data *a priori*, these techniques are capable of finding appropriate models for data with varied structure and complexity. However, this comes at a cost: the risk of over-fitting the data at hand, which would lead to poor out-of-sample performance. Because machine learning algorithms are typically used for prediction problems, their focus is thus to minimize out-of-sample prediction error. To do so, the dataset at hand is randomly partitioned into equally sized subsamples. The estimation process then involves successively holding out one of the subsample for evaluation while tuning the algorithm on all remaining subsamples to achieve best out of sample prediction accuracy on the evaluation sample (a technique called cross-validation) (Mullainathan & Spiess, 2017).

A large number of machine learning algorithms are available. Among them, the LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method similar to OLS but that involves penalizing the absolute size of the regression coefficients. Its theoretical properties have been studied in the econometrics literature, and it has been shown that the LASSO is consistent under certain conditions (Hastie, Tibshirani, & Wainwright, 2014). The LASSO is used for variable selection: because of the penalization parameter, it forces some regression coefficients to be exactly zero. This is useful in this case to detect potentially important variables interactions: the LASSO is fed with all OLS regressors reciprocally interacted (thus with more than 200 independent variables) and the algorithm selects the most

important variables for predicting the outcome. It allows to answer questions such as: is previous experience in entrepreneurship more important for female founders? These properties therefore seem appropriate to fruitfully complement the econometric analysis of this report.

While the machine-learning exercise presented here is experimental and limited in scope, its potential applicability is much wider. There are indeed numerous additional explanatory variables that could feed the same kind of algorithm, like e.g. the detailed job and education titles of all founders, the start-ups' description texts, the detailed characteristics of their IP portfolio, etc. While this would require a substantial amount of preparatory data work that is beyond the scope of this report, it could possibly lead to an effective tool to predict start-ups' VC funding and acquisition. In principle, such a tool could be used in policy analysis to attempt to identify start-ups with high-growth potential, for instance in order to inform the eligibility criteria of entrepreneurship policies.

Table 2. Descriptive statistics of main variables

	Number of observations	Mean	Standard deviation	Min	Max
Founders characteristics					
Number of founders	62343	1.59	1	1	15
At least one founder has experience in entrepreneurship (%)	25237	40.06	49	0	100
At least one founder has research experience (%)	25237	11.43	32	0	100
At least one founder has been self-employed (%)	25237	3.18	17.55	0	100
Founders are male and female dummy (%)	62048	7.99	27.11	0	100
Founders are all female dummy (%)	62048	7.78	26.78	0	100
At least one founder holds a PhD (%)	25240	11.52	31.93	0	100
At least one founder holds an MBA (%)	25240	19.16	39.35	0	100
At least one founder holds an M.Sc (%)	25240	34.89	47.66	0	100
Age of youngest founder at date of company founding	20515	14.11	8.54	-8	48
Age of oldest founder at date of company founding	20515	15.14	8.76	-8	48
Patents					
Company has filled patent application before VC (%)	178793	3.31	17.88	0	100
Company has been granted patent before VC (%)	178793	1.01	10.02	0	100
Venture Capital funding					
Company has received VC funding (%)	178793	23.05	42.11	0	100
Number of VC fundings	178793	0.40	0.95	0	24
Total VC amount received (millions USD)	178793	3.68	53.47	0	11457
Company has received public VC (%)	178793	1.13	10.55	0	100
Exit information					
Company has been acquired (%)	178793	6.71	25.02	0	100
Company went public through IPO (%)	178793	1.32	11.41	0	100
Patent application after last VC (%)	178793	3.78	19.08	0	100

Note: Age variable is normalized to 0 at date of entry in B.Sc. Therefore this variable indicates the number of years between date of entry in B.Sc. (usually around 18, but this varies across individuals and countries) and date of company founding, for the founder closest to his B.Sc. studies. Variable winsorized at 1st and 99th percentiles.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data.

Start-up characteristics associated with VC funding

This subsection presents the results of econometric analyses aimed at understanding what type of start-ups receive VC funding, and what determines the number of funding rounds and total amount of VC funding that a company receives. It also explores whether firms receiving government VC support differ from the ones backed with private VC only.

Results from OLS regressions

Results from OLS regressions of four dependent variables (a dummy variable indicating whether the company has ever received VC, and if so, number of VC rounds, total amount received, and a dummy variable indicating whether a firm has ever received government VC funding) on founders and firms characteristics are reported in Table 3. Regressions all include country and sector fixed effects that control for time-invariant country and sector heterogeneity. Standard errors are robust to heteroscedasticity and clustered at country-sector level to take into account possible correlation of shocks. The results are robust to estimating probit models for binary outcomes, and Poisson models for count data instead of linear models.

Table 3 shows that VC backed start-ups typically differ from the ones that are not, both in terms of founders and firms characteristics. Firms founded by a team of individuals are more likely to receive VC funding. If they do, they benefit from a higher number of funding rounds and a greater amount of VC funding than firms with only one founder. Previous experience in entrepreneurship and in academia increases the probability of getting VC funding by 2 and 5% respectively, but is not associated with the number and amount of VC investments. On the contrary, previously self-employed individuals receive less VC funding, if anything. These results are consistent with literature showing that the vast group of entrepreneurs is heterogeneous, with self-employed individuals and high potential start-up creators belonging to two separate sociodemographic groups (Ng & Stuart, 2016). The results also show that companies with founders that are inventors of patented technologies (that are not assigned to the company they founded) are more likely to receive VC funding by 8% and receive more in total.

Start-ups with young founders only (defined as founders being younger than 25, thus including student entrepreneurs) receive lower amounts of VC investments, but this effect is reversed when they are paired with older founders. Start-ups with a heterogeneous team of founders in terms of gender receive less VC investments than those with a team of founders composed of males only: they are 3% less likely to receive VC, and if they do they receive approximately 70% less. Companies with a team of founders composed of women only are as likely as all-male companies to receive VC funding, but if they do, the amount is significantly lower (also about 70% less). This result holds when controlling for observable firm and founders characteristics such as sector, patenting activity, education level, professional experience, age.

Start-ups' patent applications are positively associated with the probability of getting VC, and negatively associated with the number of funding rounds. A possible explanation for these results is the following: companies with patent applications before the first VC funding are more advanced in the process of product development and as such represent a less risky investment (hence the higher probability to get funding). However, they may need only few rounds of investments, for instance to market the product, thus they receive a lower number of VC rounds.

The last column of Table 3 shows whether start-ups receiving government VC funding differ from the ones receiving exclusively private VC. It indicates that government VC focuses on firms with a smaller number of founders, and with less experience in entrepreneurship. Applying for a patent before the

first VC round also appears to be less important for government VC than for private VC. The effect of other founders' characteristics, such as gender and diploma, is the same for both types of funding.

Results from machine learning algorithms

Figures 10 and 11 summarize LASSO results with the following outcome variables: the probability of receiving VC funding, and the amount received. More specifically, LASSO algorithm is used to confirm results obtained with simple OLS regressions, and to detect possible meaningful interactions of the explanatory variables. In order to partial-out country, sector and time unobserved heterogeneity, the outcome variables that feed the LASSO are residuals of the regression of each outcome variable on country, sector and time fixed effects.

Figure 10 displays the main predictors of VC funding. The first graph reinforces the OLS results and allows determining what variables are most important predictors of the probability to receive VC investments. Patents seem crucial to get VC funding as two variables relating to patenting behaviour of company and its founders are among the three best predictors. Results from OLS regressions concerning the number of founders are confirmed: teams of three founders are more likely to receive VC while start-ups with a unique founder are less likely. The second graph of Figure 10 permits to draw conclusions concerning the interaction of explanatory variables considered in OLS regressions. The most important variable in this regression appears to be the interaction between a dummy variable indicating a unique founder and a dummy variable indicating that she independently holds a patent. The strong negative effect of unicity of founder is reversed when the founder is an inventor of a patented technology. It seems that in this case having developed a patent sends a compelling signal to the market. Surprisingly, patenting does not play a role for other under-invested groups (e.g. mixed gender start-ups).

Figure 11 illustrates the results when the outcome variable is the amount of VC funding received by start-ups. Again, patent inventors are clearly favoured by VC funds, and this variable reinforces the effect of other variable that have a positive effect on VC funding in OLS regressions, i.e., patent behaviour of founders does not send a signal to offset the effect of negative variables, but rather strengthens the positive impact of some variables (e.g. teams of three founders).

Table 3. Determinants of VC funding

OLS regressions of VC funding variables on firms and founders' characteristics

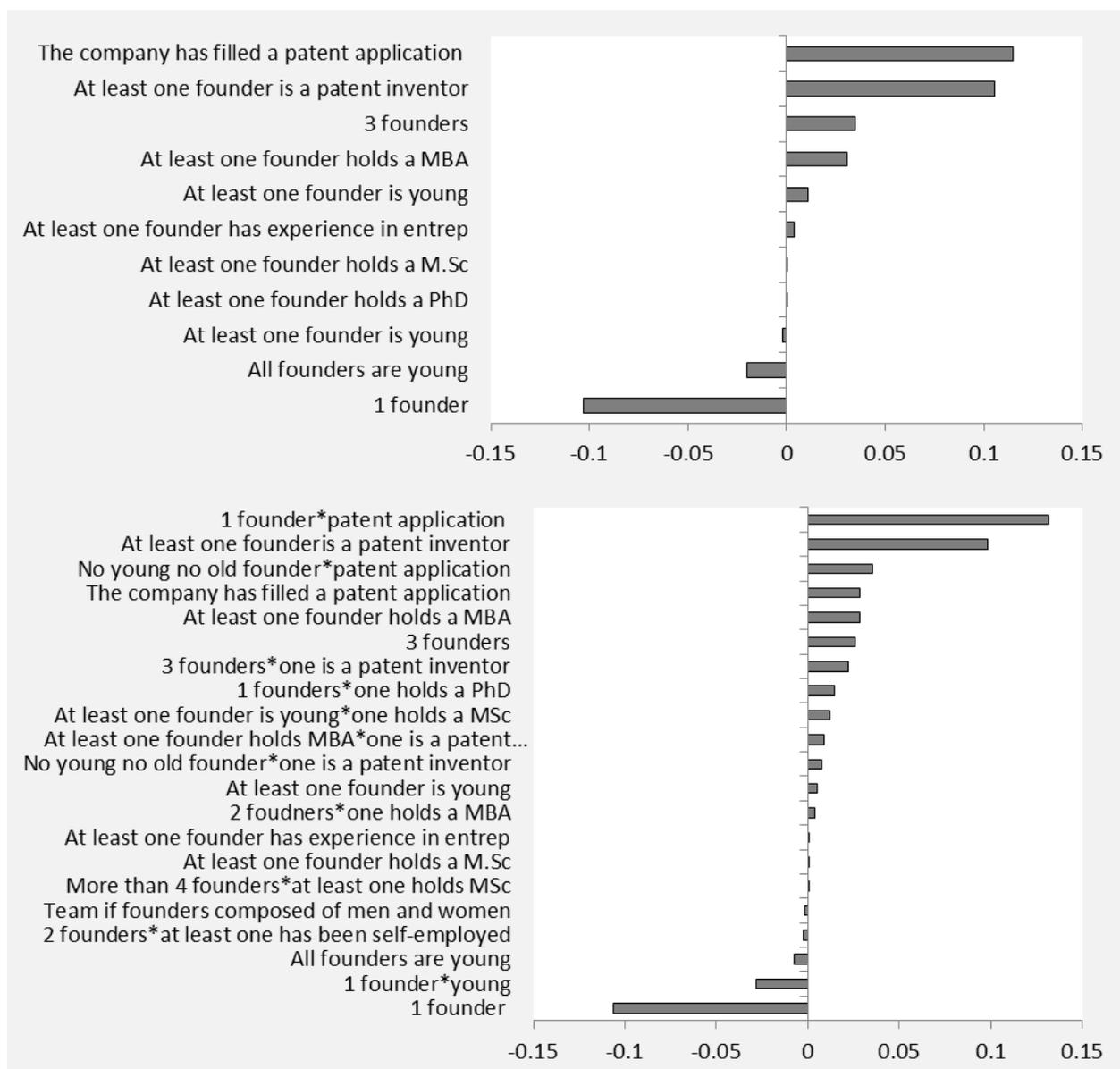
Dependent variables:	VC	Number VC	Amount (USD, in log)	Government VC
2 founders dummy	0.100*** (0.00906)	0.150*** (0.0338)	0.334** (0.161)	-0.00857 (0.00578)
3 founders dummy	0.149*** (0.0111)	0.220*** (0.0465)	0.693*** (0.169)	-0.0171*** (0.00586)
4 founders or more dummy	0.115*** (0.0162)	0.225*** (0.0531)	1.147*** (0.245)	-0.0205** (0.00947)
At least one founder has experience in entrepreneurship dummy	0.0188** (0.00838)	0.0588* (0.0346)	0.0534 (0.102)	-0.0144*** (0.00524)
At least one founder has research experience dummy	0.0457*** (0.0127)	0.00298 (0.0345)	-0.0369 (0.177)	0.00520 (0.00674)
At least one founder has been self-employed dummy	-0.00962 (0.0227)	-0.0503 (0.0512)	-0.00440 (0.306)	-0.00519 (0.0112)
At least one founder is a patent inventor dummy	0.0805*** (0.0135)	0.349*** (0.0456)	1.102*** (0.161)	0.00233 (0.00620)
Founders are male and female dummy	-0.0331*** (0.0126)	-0.0788* (0.0408)	-0.716*** (0.191)	-0.0116* (0.00692)
Founders are all female dummy	0.00743 (0.0187)	-0.0485 (0.0557)	-0.656** (0.277)	-0.0146 (0.00914)
At least one founder holds a PhD dummy	0.0229* (0.0124)	-0.0873 (0.0539)	0.00829 (0.198)	0.0169* (0.00985)
At least one founder holds an MBA dummy	0.0453*** (0.00944)	0.0119 (0.0320)	0.137 (0.128)	0.00725 (0.00529)
At least one founder holds an M.Sc dummy	0.0129* (0.00714)	-0.0184 (0.0312)	-0.371*** (0.143)	0.0144*** (0.00475)
Founders are all younger than 25	-0.0106 (0.0128)	0.0361 (0.0415)	-0.387** (0.161)	0.00797 (0.00715)
Founders are all older than 50	-0.00436 (0.0231)	-0.267*** (0.0607)	0.675*** (0.256)	0.0286** (0.0142)
Founders are young and old	-0.235** (0.115)	-0.568 (0.645)	2.155*** (0.563)	-0.0545*** (0.0141)
At least one founder is young	0.0584*** (0.0182)	0.0531 (0.0634)	-0.350 (0.256)	-0.00310 (0.0101)
At least one founder is old	-0.0390 (0.0279)	-0.330** (0.151)	-0.529 (0.476)	0.0342 (0.0250)
At least one founder is young and one is old	0.116*** (0.0277)	-0.311 (0.276)	-6.514 (6.405)	-0.0639** (0.0252)
Company has filled patent application before VC dummy	0.153*** (0.0278)	-0.236*** (0.0557)	0.0720 (0.291)	-0.0276** (0.0109)
Company has been granted patent before VC dummy	0.00793 (0.0325)	-0.470*** (0.0861)	0.251 (0.475)	0.0217 (0.0217)
Observations	14,829	8,556	8,556	8,556
R-squared	0.157	0.241	0.165	0.053
F-test of joint significance of country fixed effects	6.657	32.47	5.632	7.473
F-test of joint significance of sector fixed effects	30.58	11.89	25.04	6.038

Note: * p<0.1, ** p<0.05; *** p<0.01. Fixed effects for country, sector, and year in which the start-up was founded included in all regressions. Robust clustered standard errors in parentheses. The sample of column 1 is composed of all firms for which the data are available; the sample of column 2 to 4 is restricted to VC recipients only. The reported statistics on the fixed effects are F-tests of joint significance of all fixed effects relative to that group (countries or sectors).

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data.

Figure 10. Machine learning: predictors of receiving one or more rounds of VC funding

LASSO regressions of VC funding dummy on set of firms and founders characteristics (top) and their interactions (bottom)

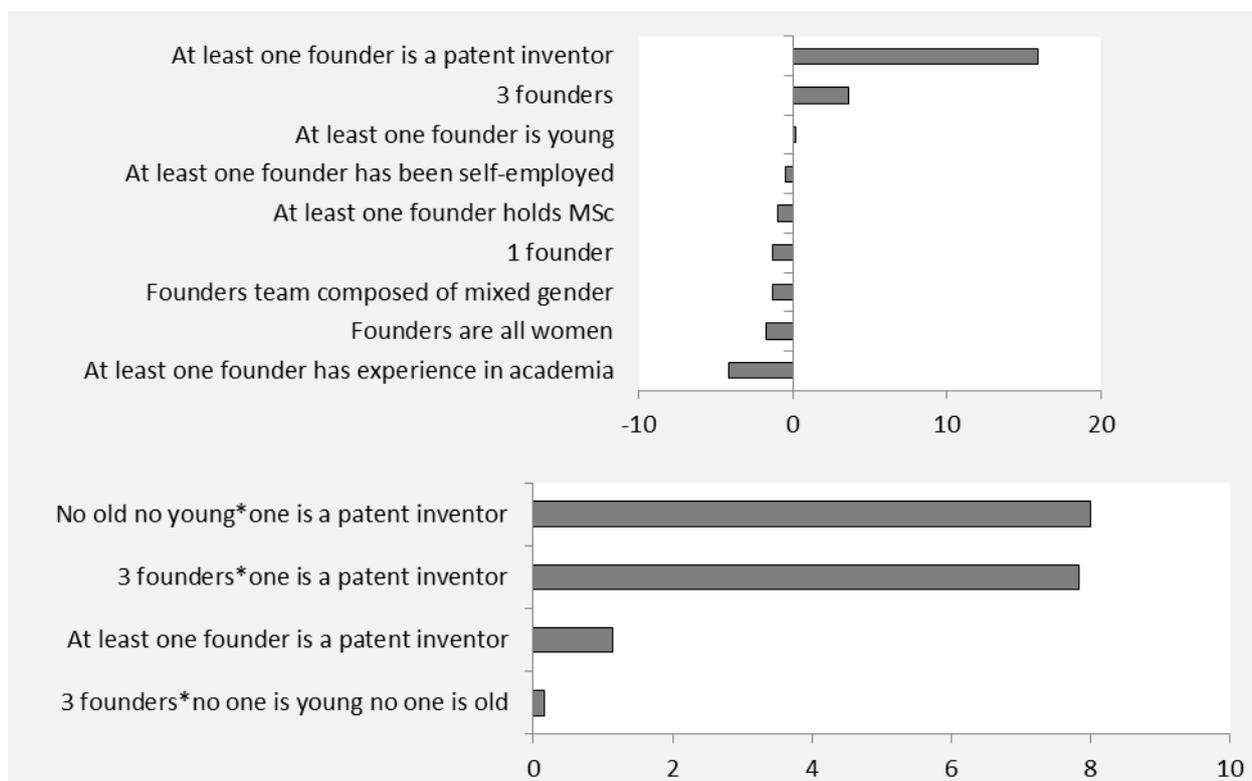


Note: Coefficients from LASSO regressions of residuals of VC funding dummy (obtained by regressing the dummy variable on country and sector fixed effects). Sample is composed of all firms.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data..

Figure 11. Machine learning: predictors of the amount of VC funding

LASSO regressions of amount of funding (for VC recipients only) on set of firms and founders characteristics (top) and their interactions (bottom)



Note: Coefficients from LASSO regressions of residuals of amount of VC funding (obtained by regressing the variable on country and sector fixed effects). Sample is restricted to VC recipients only.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data.

Start-up characteristics associated with successful exit and patenting

This subsection presents the results of econometric analyses aimed at understanding which start-ups are “successful”. Successful start-ups are identified based on three different outcomes. The first two variables relate to successful exit events: acquisition and IPO. In Crunchbase, as well as in other databases on VC, there is little information on failures, and a few start-ups that are officially still operating might actually be “living dead”. The literature has therefore focused on these two exit events to identify successful start-ups (Da Rin, Hellman, & Puri, 2011). A third outcome is filing one or more patents after the last round of VC funding. While patent applications are definitely not the end goal of a start-up, they usually represent a stepping stone in the firm’s development and can be seen as an indicator of success for firms not mature yet for being acquired or going public. It is indeed recognized in the literature that start-ups that hold patents experience higher growth in terms of employment and sales (Farre-Mensa, Hegde, & Ljungqvist, 2017).

Results from OLS regressions

Table 4 presents results from OLS regressions of three dependent variables (a dummy variable indicating whether the company has been acquired, has had an IPO, and has filed a patent application) on founders and firms characteristic, and VC funding variables. Regressions include country and sector fixed effects.

Standard errors are robust to heteroscedasticity and clustered at country-sector levels. Results are robust to estimating probit instead of linear models.

In general, VC funding (both receiving funding and amount raised) is positively associated with successful exit, even if the magnitude of the effect varies. For instance, receiving VC funding increases the probability of being acquired by 3%, the probability of IPO by 0.2%, and the probability of a later patent application by 2%. Notably, the fact that a start-up receives government VC support instead of only private VC funding does not impact the probability of acquisition, IPO, or patent application.

Experience in academia is positively correlated with the probability to be acquired, while holding a PhD matters for patent application. Companies where at least one founder is a patent inventor are more likely to experience a successful exit or file a patent application. Companies with at least one female founder are less likely to be acquired, but as likely as male companies to initiate an IPO or to file a patent application. Start-ups founded by young individuals only (including students only) are less likely to be acquired, but the effect is mainly reversed when the team of founders contain older individuals. Finally, patent applications and publications are strongly negatively correlated with the probability of IPO, and to a lesser extent with probability of acquisition.

Table 4. Successful exit and post-VC patenting

	Acquisition	Acquisition	IPO	IPO	Patent application	Patent application
2 founders dummy	0.00977** (0.00403)	0.00879** (0.00411)	-0.00357 (0.00242)	-0.00324 (0.00232)	-0.00582 (0.00446)	-0.00695 (0.00455)
3 founders dummy	0.00724 (0.00564)	0.00609 (0.00575)	-0.00442 (0.00301)	-0.00384 (0.00286)	-0.00266 (0.00402)	-0.00417 (0.00409)
4 founders or more dummy	0.0131* (0.00789)	0.0130 (0.00791)	0.00278 (0.00471)	0.00340 (0.00487)	-0.000754 (0.00434)	-0.00157 (0.00437)
At least one founder has experience in entrepreneurship dummy	0.00139 (0.00350)	0.00109 (0.00353)	-0.000165 (0.00103)	-8.66e-05 (0.00102)	-0.000736 (0.00289)	-0.000935 (0.00288)
At least one founder has research experience dummy	0.0109* (0.00560)	0.0103* (0.00563)	0.000184 (0.00242)	0.000259 (0.00246)	-0.00361 (0.00516)	-0.00425 (0.00516)
At least one founder has been self-employed dummy	-0.00930 (0.00739)	-0.00914 (0.00740)	-0.000234 (0.00089)	-0.000219 (0.00088)	0.00381 (0.00485)	0.00397 (0.00493)
At least one founder is a patent inventor dummy	0.0534*** (0.0122)	0.0539*** (0.0121)	0.0357*** (0.0135)	0.0361*** (0.0137)	0.0596*** (0.0134)	0.0595*** (0.0135)
Founders are male and female dummy	-0.0213*** (0.00432)	-0.0218*** (0.00434)	0.00237 (0.00198)	0.00211 (0.00188)	0.00423 (0.00404)	0.00430 (0.00397)
Founders are all female dummy	-0.0159* (0.00828)	-0.0169** (0.00829)	-2.69e-06 (0.00200)	-0.000132 (0.00203)	0.000720 (0.00488)	0.000330 (0.00494)
At least one founder holds a PhD dummy	-0.00742 (0.00696)	-0.00769 (0.00699)	0.00294 (0.00252)	0.00299 (0.00253)	0.0161*** (0.00536)	0.0158*** (0.00537)
At least one founder holds an MBA dummy	0.00449 (0.00562)	0.00406 (0.00567)	-0.00164 (0.00175)	-0.00151 (0.00170)	0.000205 (0.00284)	-0.000336 (0.00287)
At least one founder holds an M.Sc dummy	-0.00160 (0.00395)	-0.00212 (0.00397)	-0.00223 (0.00174)	-0.00234 (0.00172)	-0.00237 (0.00277)	-0.00275 (0.00276)
Founders are all younger than 25 dummy	-0.00811* (0.00488)	-0.00830* (0.00489)	-0.00118 (0.00120)	-0.00130 (0.00118)	-0.00225 (0.00254)	-0.00228 (0.00251)
Founders are all older than 50 dummy	-0.0308*** (0.00890)	-0.0300*** (0.00888)	0.00785 (0.00555)	0.00790 (0.00563)	0.0170 (0.0119)	0.0173 (0.0119)
Founders are young and old dummy	-0.0566*** (0.0185)	-0.0522*** (0.0191)	-0.00838 (0.00935)	-0.00848 (0.00891)	-0.0586 (0.0377)	-0.0552 (0.0381)
At least one founder is young dummy	0.0106 (0.00839)	0.00928 (0.00831)	-0.00220 (0.00253)	-0.00219 (0.00252)	0.00246 (0.00446)	0.00147 (0.00447)
At least one founder is old dummy	-0.0132 (0.0170)	-0.0132 (0.0168)	0.000590 (0.00595)	0.000291 (0.00597)	-0.0108 (0.0154)	-0.0107 (0.0154)
At least one founder is young and one is old dummy	-0.0655** (0.0279)	-0.0791** (0.0383)	-0.0335 (0.0218)	-0.0356* (0.0206)	-0.0399 (0.0304)	-0.0462* (0.0257)
Company has filled patent application	-0.00170 (0.00961)	-0.00286 (0.00952)	-0.0337** (0.0138)	-0.0334** (0.0138)	0.235*** (0.0183)	0.234*** (0.0183)
Company has been granted patent	-0.0221* (0.0116)	-0.0202* (0.0119)	0.0194*** (0.00558)	0.0189*** (0.00529)	0.176*** (0.0254)	0.176*** (0.0255)
Raised amount (USD, in log)	0.00194*** (0.000428)		0.00040*** (0.00015)		0.000827** (0.00025)	
Company has received VC funding dummy		0.0349*** (0.00556)		0.00196* (0.00100)		0.0219*** (0.00327)
Company has received public VC dummy		-0.00958 (0.0135)		0.00282 (0.00743)		0.00319 (0.0113)
Observations	14,746	14,746	13,935	13,935	14,829	14,829
R-squared	0.070	0.072	0.101	0.100	0.400	0.401
F-test for country effects	12.87	12.57	2.072	2.148	1.222	1.213
F-test for sector effects	7.092	6.795	5.055	4.962	4.990	5.407

Note: * p<0.1, ** p<0.05; *** p<0.01. Fixed effects for country, sector, and year in which the start-up was founded included in all regressions. Robust clustered standard errors in parentheses. The sample of columns 1 and 2 excludes firms that did an IPO; the sample of column 3 and 4 excludes firms that got acquired. The sample of columns 5 and 6 includes all firms. The reported statistics on the fixed effects are F-tests of joint significance of all fixed effects relative to that group (countries or sectors). All dependent variables are binary variables.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data.

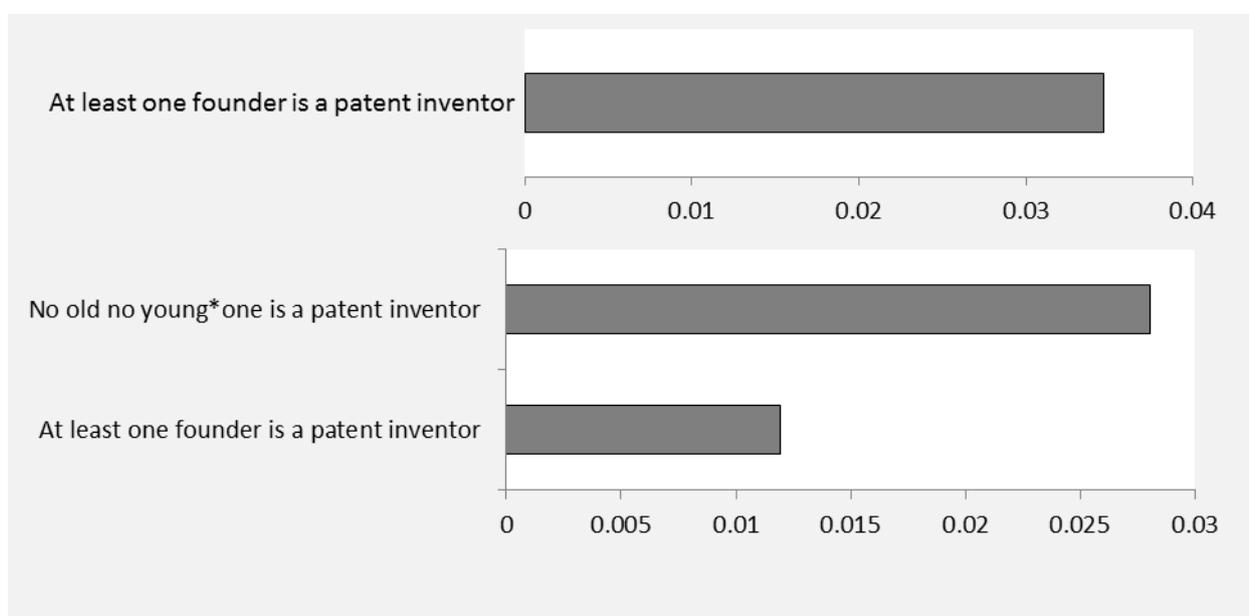
Results from machine learning algorithms

Figure 12 shows results from LASSO for the probability of acquisition of start-ups. In order to get rid of country, sector and time unobserved heterogeneity, LASSO is used on residuals of the regression of each outcome variable on country, sector and time fixed effects.

These results confirm conclusions drawn from OLS regressions: whether one founder is a patent inventor or not is the single most important predictor of successful exit (top of Figure 12). Firms with founders of intermediate age holding an independent patent are the ones that are highly likely to be acquired (bottom of Figure 12).

Figure 12. Machine learning: predictors of acquisition

LASSO regressions of acquisition dummy on set of firms and founders characteristics (top) and their interactions (bottom)



Note: Coefficients from LASSO regressions of residuals of acquisition dummy (obtained by regressing the dummy variable on country and sector fixed effects). The sample is composed of all firms.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT data.

Policy discussion and next steps

The main objective of this interim report is to illustrate a newly assembled database on innovative start-ups and VC investments. The descriptive part of this report points to substantial differences across countries in the educational and professional background of start-ups founders, as well as in the importance and composition of start-ups' IP assets. Similarly, in the regression analysis country fixed effects appear to be highly significant in explaining the probability of receiving VC, the amount of funding, and the probability of achieving a successful exit (acquisition or IPO) – even once founders' characteristics, IP assets, and sector are controlled for. These results suggest that national policies and framework conditions are likely to play an important role in explaining start-ups' success.

The descriptive analysis presented in this paper can be expanded along a number of different dimensions. For example, it is possible to describe with much more detail the fields of study of the founders, as well as their previous employment career. Furthermore, for most start-ups the location (at municipality level) is also available, which would allow an analysis of the geography of innovative entrepreneurship. The information available from the patent documents can also be exploited much further, e.g. by a detailed analysis of the backward citations of the start-up patents. The dynamics of some specific group of start-ups – e.g. operating in climate change mitigation sectors or in digital sectors, or developing disruptive innovations – can be specifically explored.

More importantly, this report sets the basis for several possible avenues of future policy research. As stressed in the introduction, there are several reasons why a better understanding of the innovative start-ups and VC funding ecosystem in a cross-country perspective is useful for policy making. Notably, these motivations rest on the evidence that in economies like the United States, where the venture capital market is operating at large scale since at least three decades, its primary role in fostering the growth of the most successful and innovative businesses is undisputed. In countries where the venture capital market is still operating at a smaller scale, the policy strategies to develop it are often explicit, through sizeable public investments in government venture capital and related policy initiatives. Beyond venture capital, this report presents a detailed portrait of an important and large sample of innovative start-ups. A better understanding of these companies and their founders is useful since such firms can potentially provide a significant contribution not only to economic growth, but also to address some of the grand challenges of our times, like e.g. climate change and growing economic divides.

The findings also point to a few policy areas that appear to be particularly important for innovative entrepreneurship. The first one is the role of gender. The regression analysis provides striking preliminary evidence that start-ups with at least one female founder are significantly less likely to receive VC; and, if they are funded, the amount invested is substantially lower.¹⁵ This evidence suggests that there could be important market failures and other barriers to be addressed in order to unleash the full potential of female innovative entrepreneurs. Further analysis could assess whether: the gender gap is at least partially explained by observable differences between male and female founders; there are significant differences across countries, sectors, investment stages, and type of investors; and some characteristics of the founders, of the investors, or of the investment network may attenuate it.

On the other hand, innovative entrepreneurship can also promote ethnic and social inclusiveness. For instance, there is evidence that innovative entrepreneurship fosters social mobility in the United States (Aghion et al., 2016), while minority communities, particularly those of South/East Asian origin, have played increasingly important roles in US science and sector sectors (Stephan & Levin, 2000; Chellaraj, Maskus, & Mattoo, 2008; Stuen, Mobarak, & Maskus, 2012). The links between government venture capital and the demographic characteristics of entrepreneurship could be an interesting extension of this work.

More generally, the data can be used to analyse the role played by government VC across OECD member countries and the major emerging economies. As mentioned in the introduction, up to 50% of VC-backed start-ups receive some sort of public VC funding in some countries. These public investments (which increasingly take the form of indirect investments or “funds of funds”) can be motivated by a number of different objectives, including supporting start-ups that fulfil wider societal needs, counteracting information asymmetries in the private venture capital market by providing quality signals, and promoting knowledge spillovers. However, government VC may also crowd-out private investments, especially if their investment portfolio mirrors those of private investors. In this case, government VC would imply lower returns for private investors, and would ultimately be of little support to innovative entrepreneurship. The richness of the micro-data assembled for this project allows testing a number of alternative hypotheses, advancing our knowledge on the role played by government VC across countries.

The role played by public research in fostering innovative entrepreneurship is another policy area that can be explored using the database presented in this report, possibly in conjunction with the two topics described above. More specifically, the project would address a number of detailed research questions: how important is public research for innovative entrepreneurship? Which kind of public research (e.g. distinguishing by sector fields, or basic vs. applied research) is more likely to trigger innovative entrepreneurship? Does spatial, social, and technological distance affect the intensity of the links? What role do policies and public institutions play in fostering knowledge transfer and public-private collaborations? Is there a link between inclusiveness and closeness to public research in innovative entrepreneurship?

Another policy-relevant finding of this report relates to the importance of IP assets. Both traditional econometric analysis and the machine-learning algorithms show that start-up IP assets are an important predictor of receiving VC. Surprisingly, patents invented by founders but not owned by its start-up play an even more important role – indicating that patents are first and foremost a signalling device. This suggests that faster, cheaper, and more efficient IP systems can further foster innovative entrepreneurship, notably by reducing information asymmetry in the VC market. In future research, the role of intellectual property assets in helping secure VC funding, or in achieving successful exit, can be explored. In line with the existing economic and managerial literature, this stream of work could in particular assess whether patents are more important as collateral or as a signal.

Further work could be done in a number of additional areas, including student entrepreneurship. The phenomenon is rapidly gaining attention in the policy debate, also spurred by anecdotal evidence on fast-growing “superstar” companies founded by undergraduate students. This report shows that the incidence of student entrepreneurship is non-trivial in a few countries and sectors. However, the topic is under-investigated in the economic literature, and it is not clear yet whether some policy levers could be possibly activated in order to fully unleash its potential for economic growth.

The evidence on serial entrepreneurship is also relevant for policy, in particular with reference to bankruptcy and insolvency regimes. As stressed by a number of other OECD studies (e.g., Calvino, Criscuolo, & Menon, 2016; Bravo-Biosca, Criscuolo, & Menon, 2016; Lee et al., 2011) tighter bankruptcy regulation can hamper entrepreneurship for both serial and new entrepreneurs, as it prospectively poses a greater burden on entrepreneurs in event of failure. This, however, is at least partially counterbalanced by the fact that tighter bankruptcy laws also represent a strong guarantee for investors, ultimately streamlining access to credit. For serial entrepreneurs, however, insolvency regimes and bankruptcy regulations bear additional importance, as several specific provisions directly affect the capability of starting a new business after insolvency over a period varying length. Furthermore, economic and civic disabilities imposed on the debtor are linked to different consequences in terms of stigma of failure. Landier (2005) provide a theory of how a stigma of failure may become a self-fulfilling prophecy: if the environment is not forgiving of failure, only low ability individuals are willing to start risky companies, and the pool of failed

entrepreneurs consists of low ability entrepreneurs that cannot be financed profitably. Across OECD countries and four large emerging economies (Brazil, China, India, Russia), there is a fair degree of heterogeneity in the stringency, scope, and effectiveness of bankruptcy regulations (McGowan, Andrews, & Millot, 2017), which could potentially explain the observed variability in the incidence of serial entrepreneurship.

Finally, after having stressed the main findings, it is also important not to overlook the limitations of this work. The sample under scrutiny provide with unprecedented coverage of start-ups' founders and VC deals across countries. However, the coverage and scope of the database is not as clearly defined as it is the case, for instance, of business register data, which typically cover the universe of firms in a given country. In the case of innovative start-ups and VC deals, the economic concepts itself have intrinsically fuzzy definitory boundaries. While the concept of "firm" is quite uncontroversial in economics, terms like "innovative start-up" or "VC investment" present several different nuances. As Da Rin, Hellman, & Puri (2011) clearly spell out, in the area of VC, all micro-level databases have some sample selection issues.

Furthermore, many variables analysed in this report are sourced from repositories in which this information is self-reported by the start-up founders. Although the system of openly editable content, reputation incentives, and other verifiability mechanisms increase the reliability of the reported information, this kind of data sources are still experimental in economic analysis. Although the robustness of the main results is carefully assessed by using different statistical approaches and by referring – whenever possible – to supporting existing evidence, these caveats should be kept in mind while assessing the findings of this report

REFERENCES

- Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, D. Hémous (2015), “Innovation and top income inequality”. Working paper n. 21247, National Bureau of Economic Research.
- Alexy, O. et al. (2012), “Social capital of venture capitalists and start-up funding”. *Small Business Economics*, 39(4).
- Almeida, P. and B. Kogut (1997), “The Exploration of Technological Diversity and the Geographic Localization of Innovation”. *Small Business Economics*, 9(1).
- Amable, B., J. Chatelain and K. Ralf (2010), “Patents as collateral”. *Journal of Economic Dynamics and Control*, 34(6).
- Andrews, D., C. Criscuolo and C. Menon (2014), “Do Resources Flow to Patenting Firms?: Cross-Country Evidence from Firm Level Data”, *OECD Economics Department Working Papers*, No. 1127, OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/5jz2lpmk0gs6-en>
- Arikan, A. and L. Capron (2010), “Do newly public acquirers benefit or suffer from their pre-IPO affiliations with underwriters and VCS?” *Strategic Management Journal*, 31(12), pp. 1257-1289.
- Baum, J. A. and B. Silverman (2004), “Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups”. *Journal of Business Venturing*, 19(3), pp. 411-436.
- Baumol, W. J. (2004), “Entrepreneurial enterprises, large established firms and other components of the free-market growth machine”. *Small Business Economics*(23), 9–21.
- Baumol, W. J., M. A. Schilling and E. N. Wolff (2009), “The Superstar Inventors and Entrepreneurs: How Were They Educated?” *Journal of Economics Management and Strategy*, 18(3), pp.711–728.
- Bertoni, F. and T. Tykvová (2015), “Does governmental venture capital spur invention and innovation?” Evidence from young European biotech companies. *Research Policy*, 44(4), pp. 925-935.
- Bertoni, F. and T. Tykvová (2012), “Which Form of Venture Capital is Most Supportive of Innovation?” *ZEW Discussion Paper*.
- Blanchenay, P., F. Calvino, C. Criscuolo and C. Menon (2017a), “Cross-country Evidence on Business Dynamics over the Last Decade: From Boom to Gloom?” OECD Science, Technology and Industry Working Paper, (forthcoming).
- Blanchenay, P., G. Berlingieri, S. Calligaris, C. Criscuolo (2017b), “The Multiprod project: a comprehensive review”. *OECD Science, Technology and Industry Working Papers*. No. 2017/04, OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/2069b6a3-en>
- Block, J. et al. (2015), “Why do SMEs file trademarks? Insights from firms in innovative industries”. *Research Policy*, 44(10).
- Brander, J. A., Q. Du and T. Hellman (2015), “The Effects of Government-Sponsored Venture Capital: International Evidence”. *Review of Finance*, pp. 571-618.

- Brau, J. C., B. Francis and N. Kohers (2003), “The Choice of IPO versus Takeover: Empirical Evidence”. *The Journal of Business*, 76(4), pp.583-612
- Brav, A. and P. Gompers (2003), “The Role of Lockups in Initial Public Offerings”. *Review of Financial Studies*, 16(1).
- Bravo-Biosca, A., C. Criscuolo and C. Menon (2016), “What drives the dynamics of business growth?” *Economic Policy*, 31(88).
- Breschi, S., D. Tzabbar, B. Cirillo, H.D. Park (2017), “Knowledge and Social Defences: The Effect on Knowledge and Social Interdependencies on Inventor Outward Mobility”. *Mimeo*.
- Burton, M., J. Sørensen and C. Beckman (2002), “Coming From Good Stock: Career Histories and New Venture Formation”. In Lounsbury, M., & Ventresca, M. J. (eds), *Social structure and organizations revisited* (pp. 229-262). Emerald Group Publishing Limited.
- Calvino, F., C. Criscuolo and C. Menon (2016), “No Country for Young Firms?: Start-up Dynamics and National Policies”. *OECD Science, Technology and Industry Policy Papers* n. 2016/29. OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/5jm22p40c8mw-en>
- Carter, R. and S. Manaster (1990), “Initial Public Offerings and Underwriter Reputation”. *The Journal of Finance*. 45(4), pp.1045-1067.
- Chandler, G. and S. Hanks (1998), “An Examination of The Substituability of Founders Human and Financial Capital in Emerging Business Ventures”. *Journal of Business Venturing*, 13(5), pp. 353-369
- Chellaraj, G., K. Maskus and A. Mattoo (2008), “The contribution of international graduate students to US innovation”. *Review of International Economics*, 16(3), pp. 444-462.
- Cockburn, I. and M. MacGarvie (2009), “Patents, thickets and the financing of early-stage firms: Evidence from the software industry”. *Journal of Economics and Management Strategy*, 18(3).
- Conti, A., J. Thursby and M. Thursby (2013), “Patents as signals for startup financing”. *Journal of Industrial Economics*, 61(3).
- Cox, N.J. (2009). “Speaking Stata: Creating and varying box plots”. *Stata Journal*, 9, pp. 478–496.
- Criscuolo, C. and C. Menon (2015), “Environmental policies and risk finance in the green sector: Cross-country evidence”. *Energy Policy* 83, pp. 38-56.
- Criscuolo, C. and C. Menon (2014), “Environmental policies and risk finance in the green sector: cross-country evidence”. *OECD Science, Technology and Industry Working Papers*. Doi: <http://dx.doi.org/10.1787/5jz6wn918j37-en>
- Criscuolo, C., P. N. Gal and C. Menon (2014), “The Dynamics of Employment Growth: New Evidence from 18 Countries”. *OECD Science, Technology and Industry Policy Papers*, no. 14. OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/5jz417hj6hg6-en>
- Da Rin, M., T. F. Hellman and M. Puri (2011), “A Survey of Venture Capital Research”. *NBER Working papers*.

- Da Rin, M., G. Nicodano and A. Sembenelli (2006), "Public policy and the creation of active venture capital markets". *Journal of Public Economics*, 90(8-9).
- Dalle, J., M. den Besten and C. Menon (2017), "Using Crunchbase for economic and managerial research", *OECD Science, Technology and Industry Working Papers*, No. 2017/08, OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/6c418d60-en>
- Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2016). "Where has all the skewness gone? The decline in high-growth (young) firms in the US". *European Economic Review*, 86, pp. 4-23.
- Engel, D. and M. Keilbach (2007), "Firm-level implications of early stage venture capital investment - An empirical investigation". *Journal of Empirical Finance*, 14(2).
- Farre-Mensa, J., D. Hegde and A. Ljungqvist (2017), "What is a Patent Worth? Evidence from the U.S. Patent 'Lottery'". *NBER Working Papers*.
- Gomes Santana Félix, E., C. Pacheco Pires and M.A. Gulamhussen (2013), "The Determinants of Venture Capital in Europe—Evidence Across Countries". *Journal of Financial Services Research*, 44, pp. 259–279.
- Gompers, P. A. and J. Lerner (2004), *The Venture Capital Cycle*. The MIT Press.
- Gompers, P. A. and J. Lerner (2002), "The money of invention". *Ubiquity*, 1(43).
- Gompers, P. A. and J. Lerner (1998), "Venture capital distributions: Short-run and long-run reactions". *The Journal of Finance*, 53(6), pp. 2161-2183.
- Gompers, P. et al. (2010), "Performance persistence in entrepreneurship". *Journal of Financial Economics*, 96(1).
- Grilli, L. and S. Murtinu (2014), "Government, venture capital and the growth of European high-techentrepreneurial firms". *Research Policy*, pp. 1523-1543.
- Grilli, L. and S. Murtinu (2014), "New technology-based firms in Europe: market penetration, public venture capital, and timing of investment". *Industrial and Corporate Change*, pp. 1109-1148.
- Gulati, R., & Higgins, M. C. (2003). "Which ties matter when? The contingent effects of interorganizational partnerships on IPO success". *Strategic Management Journal*, 24(2), pp. 127-144.
- Guzman, J. and S. Stern (2016), "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014". *National Bureau of Economic Research working paper*, No. 22095.
- Hall, B. and R. Ziedonis (1979), The Patent Paradox Revisited: An Empirical Study of Patenting in the US". *The RAND Journal of Economics Journal of Economics*, 32(1), pp. 101-128.
- Haltiwanger, J., R. Jarmin and J. Miranda (2013), "Who Creates Jobs? Small Versus Large Versus Young". *The Review of Economics and Statistics*, 95(2), pp. 347-361.

- Hastie, T., R. Tibshirani and M. Wainwright (2015). “Theoretical results for the LASSO”. In Hastie, T., Tibshirani, R., & Wainwright, M. (eds.). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.
- Häussler, C., D. Harhoff and E. Müller (2012), “To Be Financed or Not...-The Role of Patents for Venture Capital-Financing”. *ZEW - Centre for European Economic Research Discussion Paper, No. 09-003*.
- Hellmann, T. and M. Puri (2002), “Venture capital and the professionalization of start-up firms: Empirical evidence”. *Journal of Finance*, 57(1).
- Hellmann, T. and M. Puri (2000), “The interaction between product market and financing strategy: The role of venture capital”. *Review of Financial Studies*, 13(4).
- Henderson, R. (1993), “Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry”. *The RAND Journal of Economics*, 24(2), pp. 248-270.
- Hsu, D. (2007), “Experienced entrepreneurial founders, organizational capital, and venture capital funding”. *Research Policy*, 36(5).
- Hsu, D. and R. Ziedonis (2013), “Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents”. *Strategic Management Journal*, 34(7).
- Ivanov, V. I. and F. Xie (2010), “Do Corporate Venture Capitalists Add Value to Start-Up Firms? Evidence from IPOs and Acquisitions of VC-Backed Companies”. *Financial Management*. 39(1), pp. 129-152.
- Jeng, L. and P. Wells (2000), “The determinants of venture capital funding: Evidence across countries”. *Journal of Corporate Finance*, 6(3).
- Jones, B. F. (2009), The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? ” *The Review of Economic Studies*, 76(1), pp. 283-317.
- Jorgenson, D. (2001), “Information technology and the U.S. Economy”. *American Economic Review*, 91(1).
- Kerr, W., R. Nanda and M. Rhodes-Kropf (2014), “Entrepreneurship as Experimentation”. *Journal of Economic Perspectives*, 28(3), pp. 25-48
- Kortum, S. and J. Lerner (2000), “Assessing the contribution of venture capital to innovation”. *RAND Journal of Economics*, 31(4).
- Kortum, S. and J. Lerner (2001), “Does venture capital spur innovation?” In G. D. Libecap, *Entrepreneurial inputs and outcomes: New studies of entrepreneurship in the United States* (pp. 1-44). Emerald Group Publishing Limited.
- Kortum, S. and J. Lerner (2000), “Assessing the contribution of Venture Capital to innovation”. *RAND Journal of Economics*. 3(4), pp. 674-692.
- Landier, A. (2005), “Entrepreneurship and the Stigma of Failure”. *Working Paper*.

- Lee, C., K. Lee and J. Pennings (2001), "Internal Capabilities, External Networks, and Performance: A Study on Technology-Based Ventures". *Strategic Management Journal* 22, pp. 615-640.
- Lee, S. M. and B. Lee (2014), "Entrepreneur characteristics and the success of venture exit: an analysis of single-founder start-ups in the U.S." *International Entrepreneurship and Management Journal*. 11(4), pp. 891-905.
- Lee, S.-H. et al. (2011), "How do bankruptcy laws affect entrepreneurship development around the world?" *Journal of Business Venturing*, 26(5), pp. 505-520
- Lerner, J. (2002), "When Bureaucrats Meet Entrepreneurs: the Design of Effective Public Venture Capital Programmes". *The Economic Journal*, pp. 73-84.
- Mann, R. J. and T. W. Sager (2007), Patents, venture capital, and software start-ups". *Research Policy*. 36(2), pp. 193-208.
- McGowan, A., M.D. Andrews and V. Millot (2017), "Insolvency regimes, zombie firms and capital reallocation". *OECD Economics Department Working Papers* no. 1399. Doi: <http://dx.doi.org/10.1787/5a16beda-en>
- Mullainathan, S. and J. Spiess (2017), "Machine Learning: An Applied Econometric Approach". *Journal of Economic Perspectives* 31(2), pp. 87-106.
- Ng, W. and T. E. Stuart (2016), "Of Hobos and Highfliers: Disentangling the Classes and Careers of Technology-Based Entrepreneurs". *Working paper*.
- OECD (2015), *Entrepreneurship at a Glance 2015*. Paris: OECD Publishing. Doi: http://dx.doi.org/10.1787/entrepreneur_aag-2015-en
- OECD (2009), *Patent Statistics Manual 2009*. OECD Publishing. Doi: <http://dx.doi.org/10.1787/9789264056442-en>
- Ozmel, U., J. Reuer and R. Gulati (2013), "Signals across Multiple Networks: How Venture Capital and Alliance Networks Affect Interorganizational Collaboration". *Academy of Management Journal*, 56(3), pp. 852-866.
- Plehn-Dujowich, J. (2010), "A theory of serial entrepreneurship". *Small Business Economics*, 35(4), pp. 377-398.
- Pollock, T. and R. Gulati (2007), "Standing out from the crowd: the visibility-enhancing effects of IPO-related signals on alliance formation by entrepreneurial firms". *Strategic Organization*, 5(4), pp. 339-372.
- Puri, M. and R. Zarutskie (2012), "On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms". *The Journal of Finance*, 67(6), pp. 2247-2293.
- Ragozzino, R. and D. P. Blevins (2016), "Venture-Backed Firms: How Does Venture Capital Involvement Affect Their Likelihood of Going Public or Being Acquired?" *Entrepreneurship Theory and Practice*, 40(5), pp. 1540-6520.
- Ragozzino, R. and J. Reuer (2011), "Geographic distance and corporate acquisitions: Signals from IPO firms". *Strategic Management Journal*, 32(8).

- Romain, A. and B. Van Pottelsberghe (2004), “The economic impact of venture capital”. *Working Paper*.
- Samila, S. and O. Sorenson (2011), “Venture capital, entrepreneurship, and economic growth”. *The Review of Economics and Statistics*, 93(1), pp. 338-349.
- Schertler, A. (2003), “Driving forces of Venture Capital investments in Europe: a dynamic panel data analysis”. *Working Paper*.
- Schneider, C. and R. Veugelers (2010), “On young highly innovative companies: Why they matter and how (not) to policy support them”. *Industrial and Corporate Change*, 19(4), pp. 969-1007.
- Schoar, A. (2010), “The divide between subsistence and transformational entrepreneurship”. *Innovation policy and the economy*, 10(1), pp. 57-81.
- Shane, S. and T. Stuart (2002), “Organizational Endowments and the Performance of University Start-ups”. *Management Science*. 48(1), pp. 154-170.
- Squicciarini, M., H. Dernis and C. Criscuolo (2013), “Measuring Patent Quality : Indicators of Technological and Economic Value”. *OECD Science, Technology and Industry Working Papers* n. No. 2013/03, OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/5k4522wkw1r8-en>
- Stephan, P. and S. Levin (2000), “Exceptional contributions to US science by the foreign-born and foreign-educated”. *Population Research and Policy Review*. 20(1), pp. 59-79.
- Strebulaev, I. A. and W. Gornall (2015, October 21), *Insights by Stanford Business*. (Accessed August 22, 2017) from <https://www.gsb.stanford.edu/insights/how-much-does-venture-capital-drive-us-economy>
- Stuart, T., H. Hoang and R. Hybels (1999), “Interorganizational Endorsements and the Performance of Entrepreneurial Ventures”. *Administrative Science Quarterly*. 44(2), pp. 315-349.
- Stuen, E., A. Mobarak and K. Maskus (2012), “Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US Doctoral Programmes”. *Economic Journal*. 122(565), pp. 1143-1176
- Tarasconi G. and C. Menon (2017). "Matching Crunchbase with patent data", *OECD Science, Technology and Industry Working Papers*, No. 2017/07, OECD Publishing, Paris. Doi: <http://dx.doi.org/10.1787/15f967fa-en>
- Tushman, M. L. and P. Anderson (1986), “Technological Discontinuities and Organizational Environments”. *Administrative Science Quarterly*, pp. 439-465.
- Williams, D. R. (2013), “Human and financial capital as determinants of biopharmaceutical IPO de-listings”. *Journal of Business Research*. 66(12), pp. 2612-2618.
- Wilson, K. E. and F. Silva (2013), “Policies for Seed and Early Stage Finance: Findings from the 2012 OECD Financing”. *OECD Science, Technology and Industry Policy Papers*, No. 9. Doi: <http://dx.doi.org/10.1787/5k3xqsf00j33-en>
- Wong, P., Y. Ho, and E. Autio (2005), “Entrepreneurship, innovation and economic growth: Evidence from GEM data”. *Small Business Economics*. 24(3), pp. 335-350

- Zhang, J. (2011), “The advantage of experienced start-up founders in venture capital acquisition: evidence from serial entrepreneurs”. *Small Business Economics*, 36(2), pp. 187-208.
- Zucker, L., M. Darby and M. Brewer (1998), “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises”. *American Economic Review*, 88(1).

ANNEX 1: COVERAGE AND REPRESENTATIVENESS OF CRUNCHBASE

Similarly to other commercial databases like e.g. ORBIS maintained by Bureau Van Dijk, which are not created with the particular needs of statisticians and economists in mind, the coverage of Crunchbase is not clearly defined and may vary significantly across countries and sectors. This is undoubtedly the most important issue requiring careful examination by researchers.

However, given that Crunchbase is quickly becoming a primary data source for investors, it is plausible to assume that dynamic and ambitious entrepreneurs have a strong incentive to register on the website and to keep their information updated. The crowd-sourcing process works in the same spirit of e.g. Wikipedia, where registered users can complement and revise information added by other users. This represents an important innovation compared to other commercial databases and public data sources commonly used in economic research, which may provide with unprecedented opportunities to analyse phenomena that have been under investigated so far because of lack of suitable data.

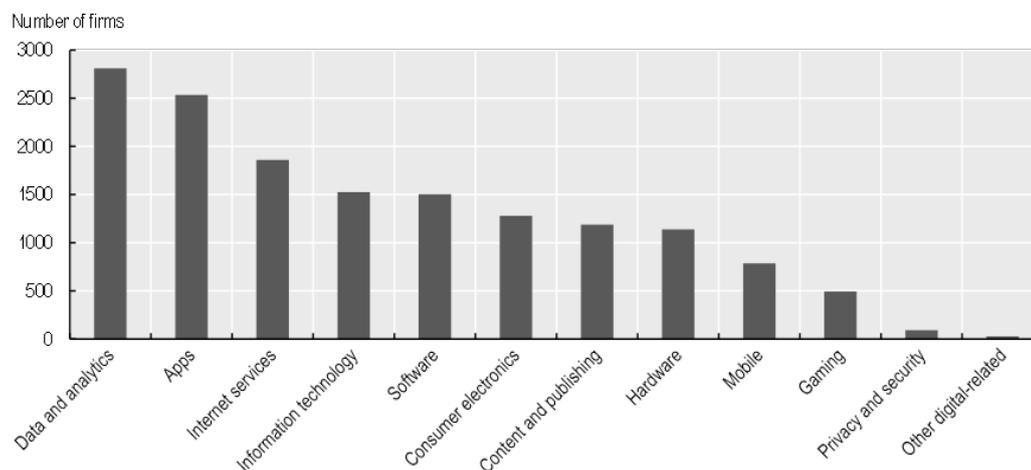
The coverage appears particularly large for start-ups operating in digital-related sectors, the top two sectors in terms of number of firms receiving VC being Data and analytics and Apps (Figure 13). From 2011 to 2016, funding in these two sectors reached more than 50% of total funding in digital-related sectors in a number of countries (Figure 14).

Furthermore, comparisons with other sources suggest that the coverage is quite comprehensive for start-ups located in the United States. Figure 15 shows that the number and amount of US investments calculated using Crunchbase and a trusted source of data on start-ups, PwC, coincide over time; if anything, Crunchbase seems more comprehensive. An unreported comparison with VentureXpert in 2013 and with VC deals listed in Eikon database from 2013 to 2016 (available from the authors upon request) also shows that numbers are consistent.

In addition, while other data sources tend to concentrate on US companies, one key advantage of Crunchbase is its international coverage. Again, aggregate statistics on VC funding by economy and year tend to be reassuringly similar to the same figures produced with an alternative and more established source, the OECD Entrepreneurship Financing Database (see Figure 16).¹⁶

Figure A.1.1. Start-ups in digital-related sectors having attracted equity funding in OECD and BRIICS, 2011-16

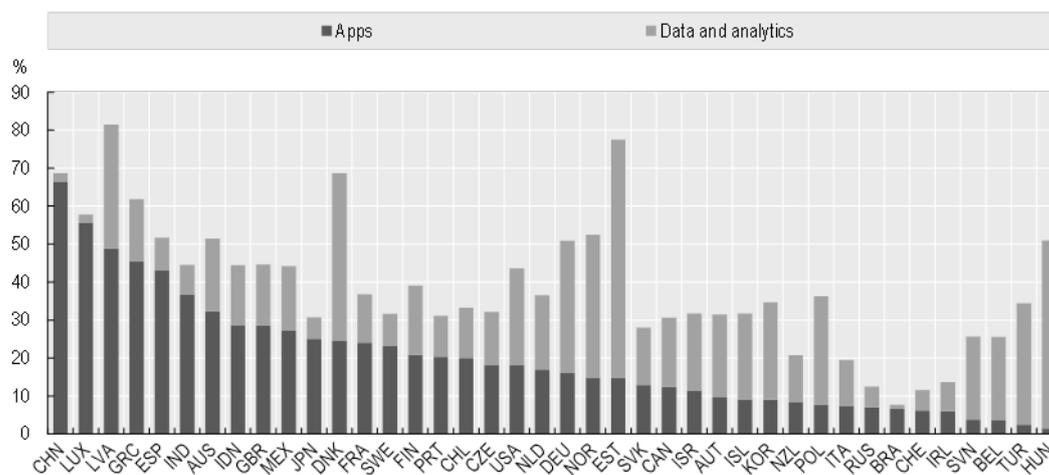
Firms aged 5 years old or less



Source: Authors' elaborations on www.crunchbase.com

Figure A.1.2. Top digital-related sectors having attracted equity funding, 2011-16

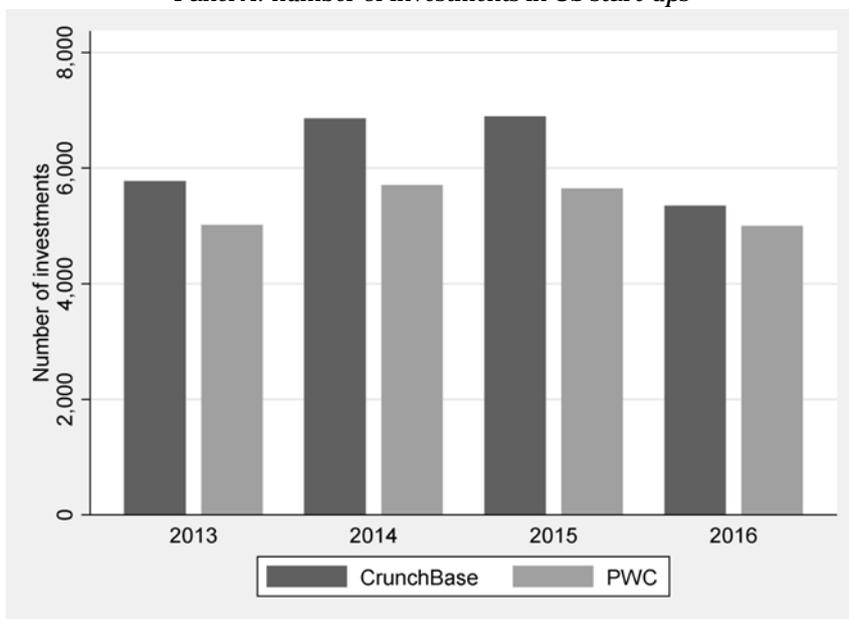
As a percentage of total equity funding in the digital-related sectors



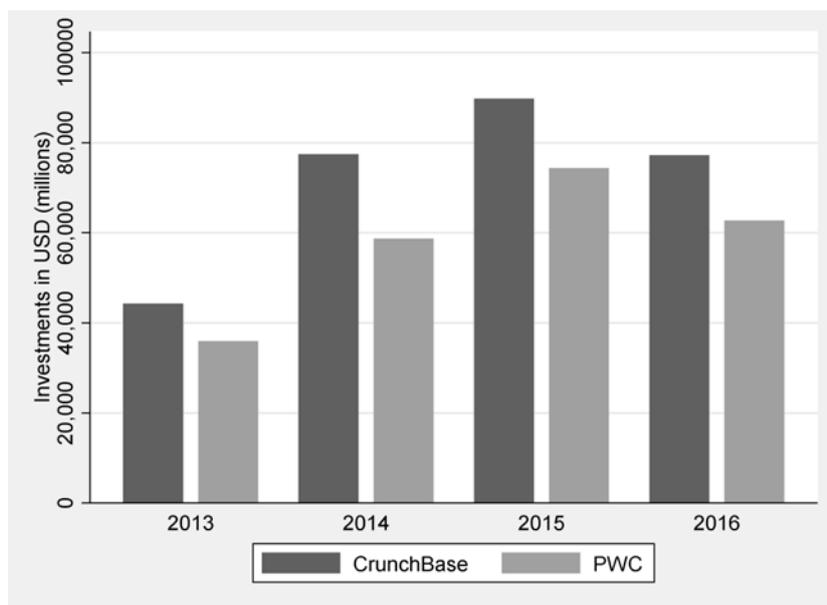
Source: Authors' elaborations on www.crunchbase.com

Figure A.1.3. Benchmarking Crunchbase against PwC

Panel A: number of investments in US start-ups



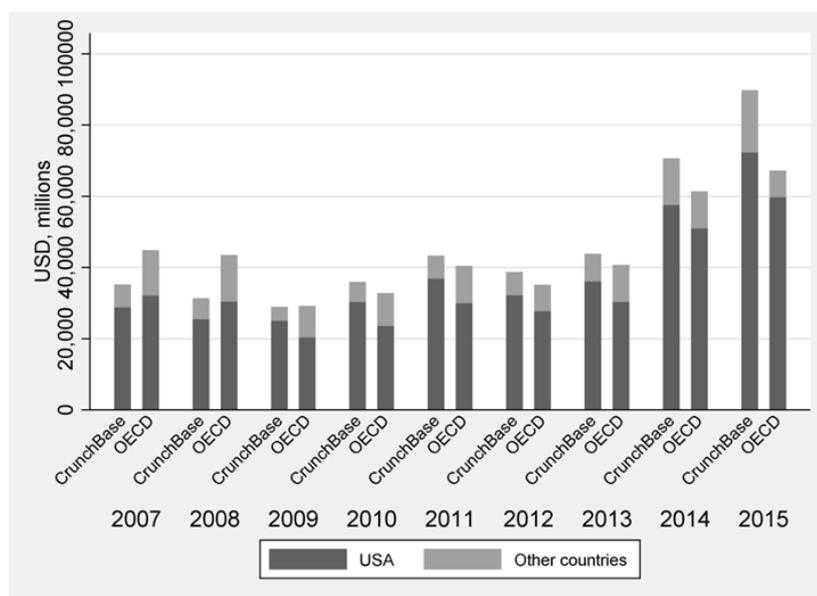
Panel B: total investment amount (millions USD) in US start-ups



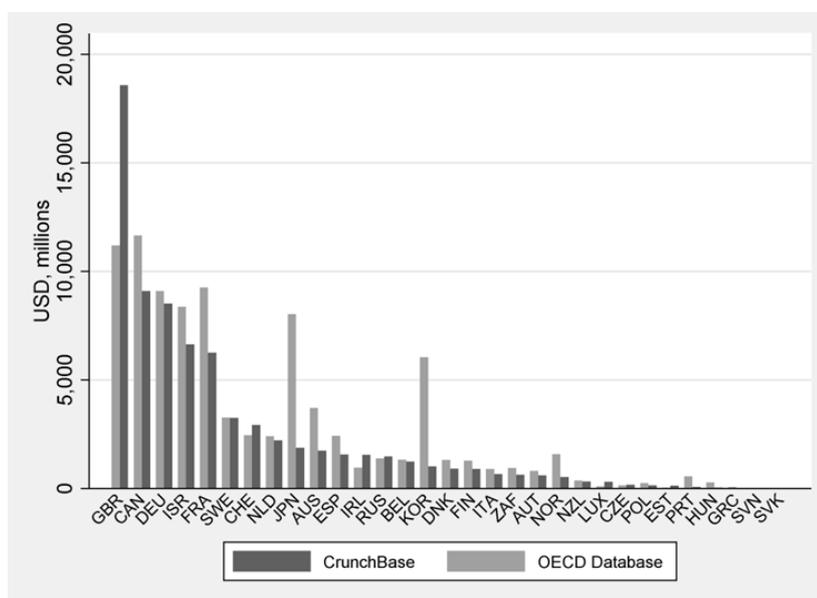
Source: Authors' elaborations on www.crunchbase.com and PwC/CB Insights MoneyTree™ Report

Figure A.1.4. Benchmarking Crunchbase against the OECD Entrepreneurship Financing Database

Panel A: over time, United States and other countries



Panel B: across countries, excluding United States



Source: Authors' elaborations on www.crunchbase.com and OECD Entrepreneurship Financing Database

ANNEX 2: ADDITIONAL TABLES AND FIGURES

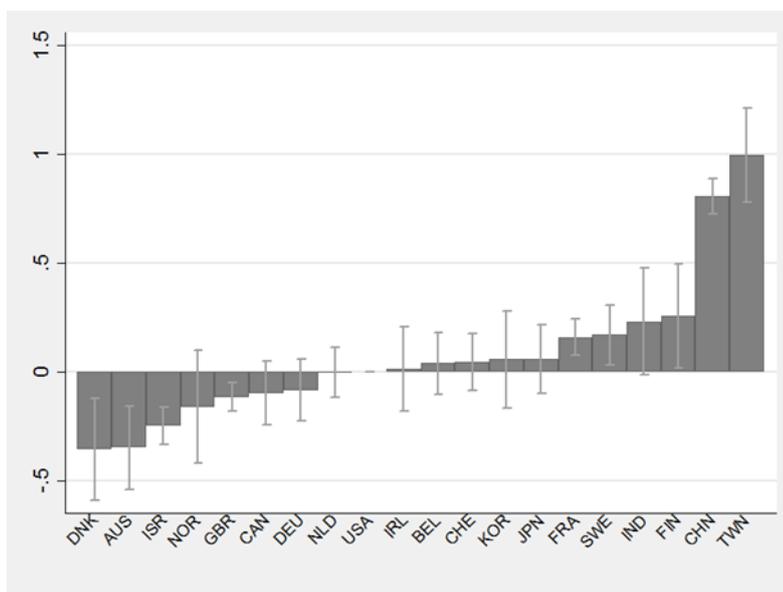
This Annex (Figure 17 to Figure 23) contains the graphs reporting the economy and sector values of the different measures illustrated in the descriptive part of the paper. The scope and interpretation of these graphs is explained in Box 1. A few additional methodological details are reported in the following.

The economy and sector fixed effects for the graphs related to patent variables (age of the company at the time when the patent was filed, Figure 17; and closeness to science, Figure 18) are estimated (via OLS) with a regression model at patent-level (one observation for each patent) that also includes fixed effects for: the year in which the start-up was founded; the patent office at which the patent was filled; the patent technological classification (first letter of the CPC classification). The model also include a dummy equal to one if the patent has been granted, and another dummy equal to one if the company received VC funding. The economy and sector fixed effects for the graphs related to founder variables (previous entrepreneurial experience; student entrepreneurship; age; education) reported in Figure 19 to Figure 23 are estimated via OLS with a regression model at founder-level (one observation for each founder) that also includes fixed effects for: the year in which the start-up was founded, plus a dummy equal to one if the company received VC funding, and if another dummy equal to one if the founder is male. In both cases, the standard errors are clustered at economy-sector pair level.

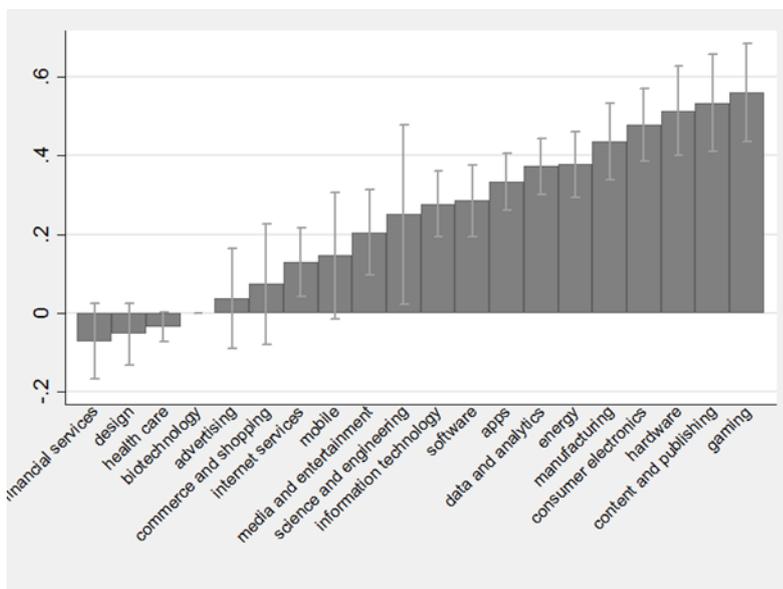
Figure A.2.1. Age of the company at the time in which the patent is filed

Average across patents; the filing year is based on the priority date; controlling for basic start-up and patent characteristics

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



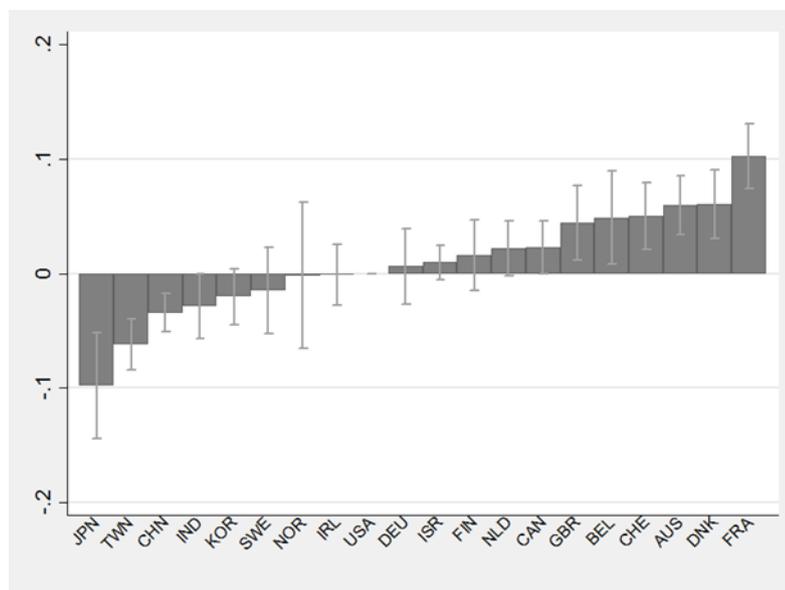
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Wisker lines report 10% confidence intervals. For more details see Box 1.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

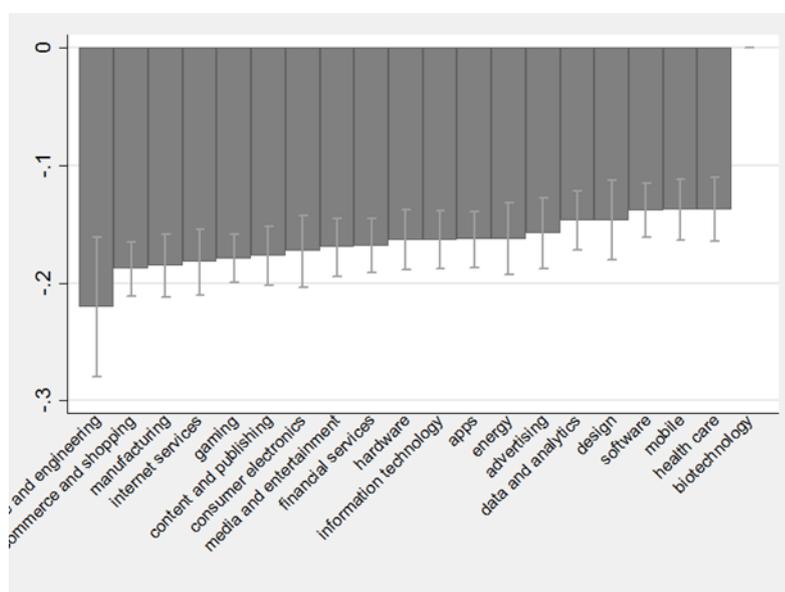
Figure A.2.2. Closeness to science

Controlling for basic start-up and patent characteristics

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



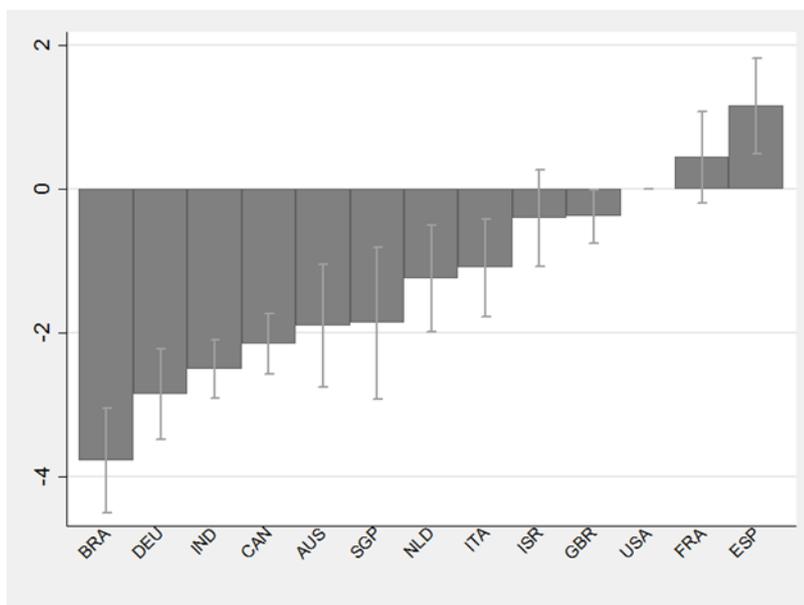
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Wisker lines report 10% confidence intervals. For more details see Box 1. Closeness to science is computed as the average share of citations to non-patent literature in total backward citations (OECD, 2009). A higher index value indicates that the patented sector builds on academic advances in science and is closer to the research frontier.

Source: www.crunchbase.com; PATSTAT; (Squicciarini, Dernis, & Criscuolo, 2013)

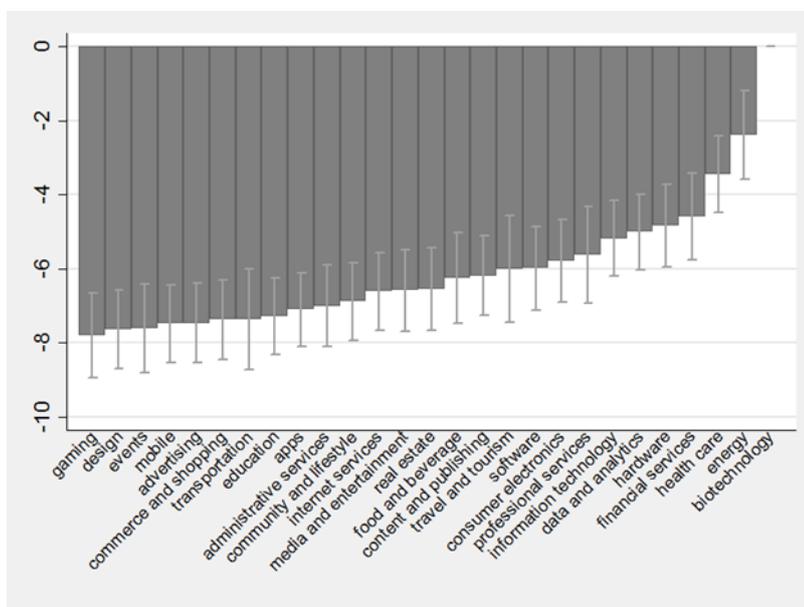
Figure A.2.3. Age of the founder at the time when the start-up was created

Controlling for sector basic start-up characteristics and founder gender

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



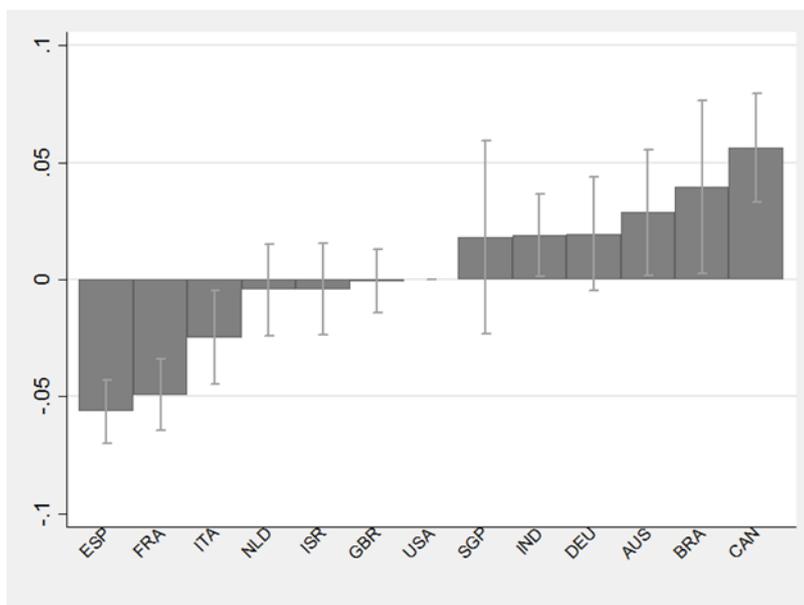
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Whisker lines report 10% confidence intervals. For more details see Box 1. Age variable standardized at 0 at date of entry in B.Sc.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

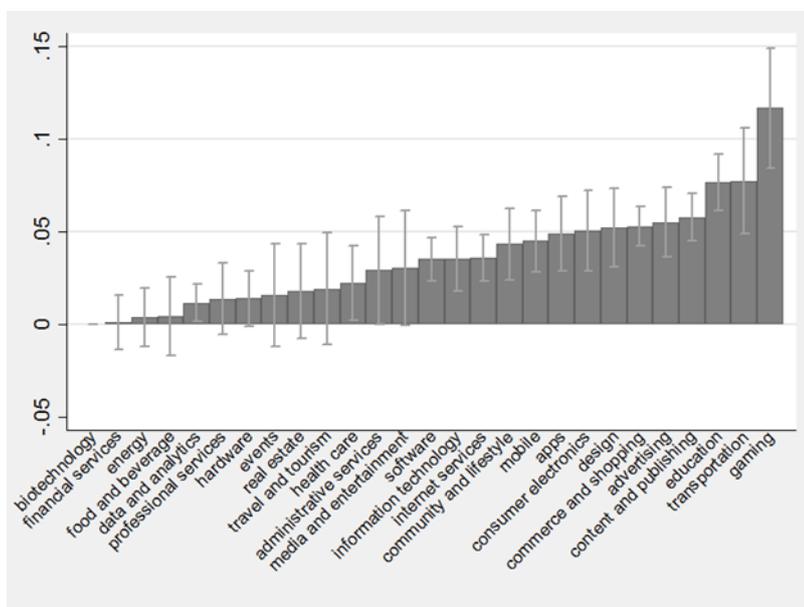
Figure A.2.4. Student entrepreneurship

Controlling sector basic start-up characteristics and founder gender

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



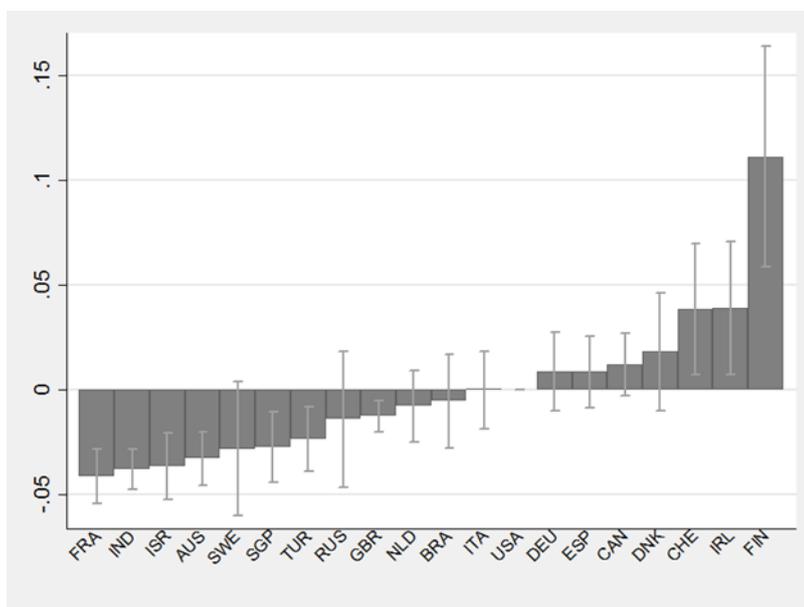
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Whisker lines report 10% confidence intervals. For more details see Box 1. Student entrepreneurs are individuals founding a company within four years since the start of the B.Sc.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

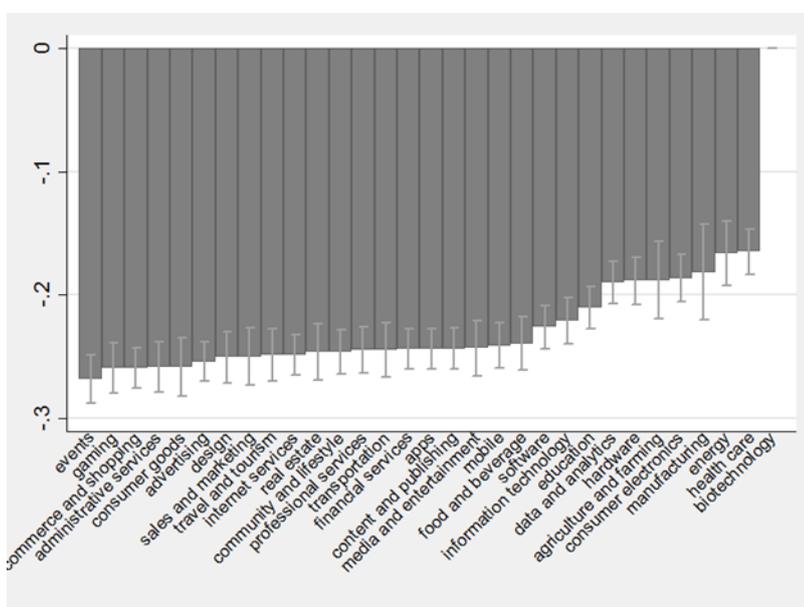
Figure A.2.5. Share of founders with previous academic experience

Controlling for sector basic start-up characteristics and founder gender

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



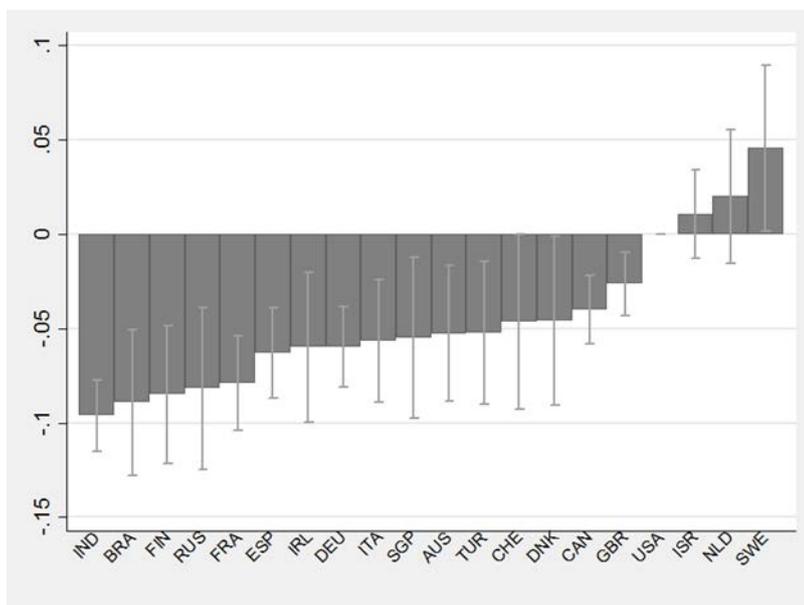
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Wisker lines report 10% confidence intervals. For more details see Box 1.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

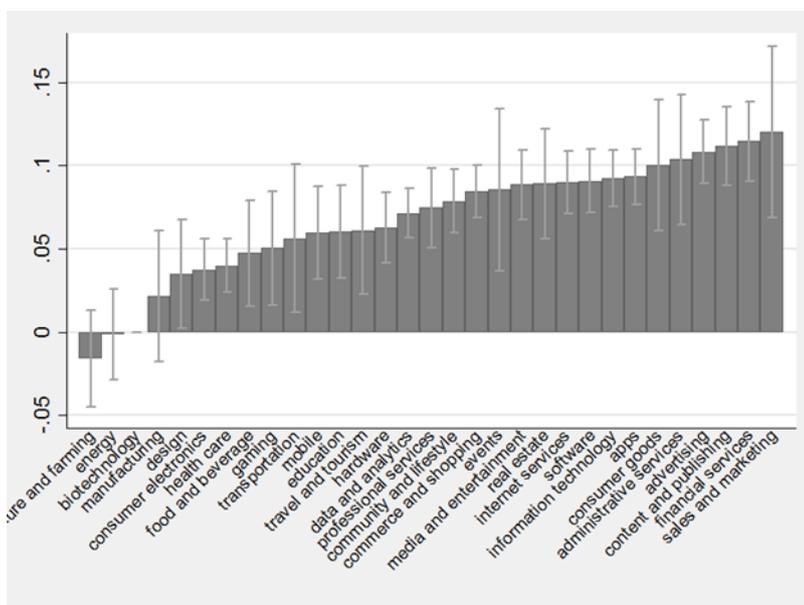
Figure A.2.6. Share of founders with previous entrepreneurial experience

Controlling for sector basic start-up characteristics and founder gender

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



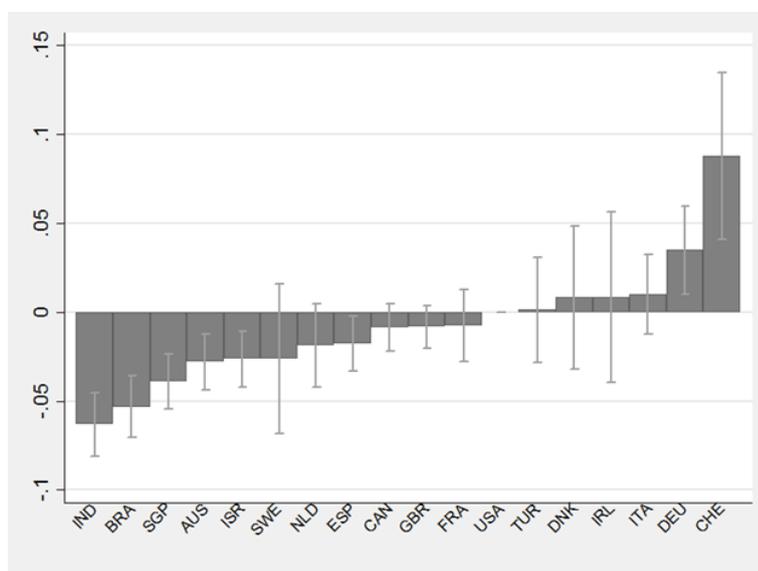
Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Whisker lines report 10% confidence intervals. For more details see Box 1.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

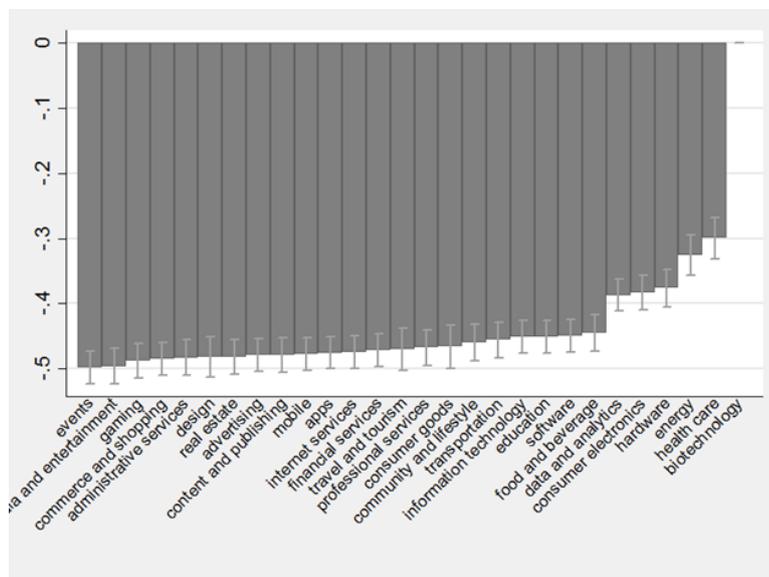
Figure A.2.7. Share of founders with a Ph.D.

Controlling for sector basic start-up characteristics and founder gender

Panel A: by economy, difference from the United States



Panel B: by sector, difference from biotechnology



Note: The graphs report the value of economy (upper panel) and sector (lower panel) fixed effects calculated from the same OLS regression model controlling also for basic start-up and patent characteristics. Whisker lines report 10% confidence intervals. For more details see Box 1.

Source: Authors' elaborations on www.crunchbase.com and PATSTAT

NOTES

1 See e.g. Calvino, Criscuolo, & Menon (2016).

2 See e.g. Almeida & Kogut, (1997); Baumol (2004); Zucker, Darby, & Brewer (1998); Jorgenson (2001).

3 On-going work within this research stream is focussing specifically on government VC.

4 For instance, a recent working paper by Ng & Stuart (2016) applies machine learning technique to classify entrepreneurs in large sample of individual careers. They find that two very distinct groups of workers can be identified under the entrepreneur umbrella: self-employed entrepreneurs, who often depart relatively low-wage jobs and may further sacrifice income for the autonomy of self-employment, compose the first group; conversely, high flyers exit high-wage, high-advancement careers to launch high potential companies.

5 See e.g. Stephan & Levin (2000); Chellaraj, Maskus, & Mattoo (2008); Stuen, Mobarak, & Maskus, (2012).

6 See e.g. Calvino, Criscuolo, & Menon (2016); Criscuolo, Gal, & Menon (2014); Blanchenay et al. (2017b); OECD (2015).

7 See e.g. “Tech ‘unicorn’ IPOs are booming” by Joe Ciolli, published at <http://www.businessinsider.fr/us/tech-unicorn-ipos-are-booming-2017-8/> (visited on August 22nd, 2017)

8 The database can be accessed on line at www.crunchbase.com

9 <http://www.kauffman.org/microsites/state-of-the-field/topics/finance/equity/venture-capital> accessed on September 11th, 2017.

10 This procedure compares for each country investors names in CrunchBase and names in the list of government VC described above, computes a similarity score for each pair of names and matches for each investor in CrunchBase its closest counterpart in the list of government VC, given that they are located in the same country. An investor is then classified as public if the similarity score exceeds a certain threshold. From the list of investors in CrunchBase, the different procedures described above allow for the identification of approximately 1,000 public investors (2% of the sample, while 35% remain unclassified).

11 Crunchbase includes companies such e.g. Microsoft that, while being innovative startups with no or few patents at the time of their founding, are now large established players with very large patent portfolios.

12 A simple patent family is a collection of patent documents filed at different patent offices that are considered to cover a single invention. The technical content covered by the applications is considered to be identical. More information is available at <https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/docdb.html> (accessed on Sept. 5th, 2017). It should be noted that patent families may extend beyond the IP5 patent offices considered in this study, therefore the sum of patents weighting by the inverse of the family size does not necessarily correspond to the number of distinct patent families.

13 The share of student founders is not dissimilar when the sample is restricted to founders of VC-backed start-ups.

- ¹⁴ Age is available only for a smaller subset of start-ups. While age is computed thanks to education variables (see Note of Table 2), it is available for a smaller number of start-up founders as some of them do not report relevant diplomas or date of entry in these diplomas.
- ¹⁵ Further preliminary regression analysis at deal-level shows that when woman-led investors are involved, the mixed-gender penalty disappears, and start-ups with female founders receive more funding. However, it must be noted that the share of female investors in the database is relatively low.
- ¹⁶ The OECD Entrepreneurship Financing Database contains aggregate statistics, by country and year, of venture capital investments. These data are typically compiled by national or regional Private Equity and Venture Capital Associations, often with the support of commercial data providers. The creation of such database is complicated because of the lack of a standard international definition of venture capital, and of the diverse methodologies employed by data compilers. For a detailed description of the underlying sources see OECD (2015).