

## CHAPTER 10

# The Private Capital Research Institute: Making Private Data Accessible in an Opaque Industry

Josh Lerner (Harvard Business School)

Leslie Jeng (Private Capital Research Institute)

Therese Juneau (Private Capital Research Institute)

---

### 10.1 Summary

The Private Capital Research Institute (PCRI) is a non-profit corporation that seeks to understand the fundamental economics of private capital.<sup>1</sup> A wide variety of forms of private capital are examined, including angel investors, venture capital and private equity organizations, and public providers of private capital (e.g., sovereign wealth funds).

---

Copyright © Josh Lerner, Leslie Jeng, and Therese Juneau.

Cite as: Lerner, Josh, Leslie Jeng, and Therese Juneau. “The Private Capital Research Institute: Making Private Data Accessible in an Opaque Industry.” In: Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. 2020.

<sup>1</sup>The PCRI is a non-profit corporation devoted exclusively for charitable purposes within the meaning of Section 501 (c)(3) of the Internal Revenue Code of 1986 as amended.

The PCRI grew out of a multi-year research initiative sponsored by the World Economic Forum that studied the economic impact of private equity.<sup>2</sup> The PCRI received initial support from the Ewing Marion Kauffman Foundation and continues to be funded through grants and strategic relationships.

The principal activities of the PCRI are (a) build data sets related to private capital that can be made available to researchers for analysis,<sup>3</sup> (b) build up a community of scholars and sponsor independent academic research on the nature and effects of private capital, and (c) disseminate the findings of this research to policymakers and the public at large to foster deeper understanding of the role that private capital plays in the economy and society.

The PCRI collects data from commercial data vendors as well as the private equity firms themselves. In addition, the PCRI collects data from primary sources, such as publicly available filings. One of the more recent projects is the gathering of public filings called Certificates of Incorporation (CoIs) from states of incorporation.

The PCRI databases are available to all academic researchers with a credible research agenda. Additionally, safeguarding PCRI's independence from outside influence is critically important. Thus, the PCRI only accepts funding from entities or individuals who recognize that the value of the PCRI's research and analysis depends on an analytically rigorous and unbiased process.

## 10.2 Introduction

### 10.2.1 Motivation and Background

The level of interest in alternative investments, and private capital in particular (which encompasses both venture capital (VC) and private equity), has been intense over the past decade. This interest has

---

<sup>2</sup>This work was collected in Anuradha Gurung and Josh Lerner, editors (2008).

<sup>3</sup>See Appendix A for summary information on the PCRI database.

stemmed from both investors' desires for attractive returns and the policy questions around this rapidly growing asset class.

Returns from the United States publicly traded equities, the mainstays of investment portfolios for individuals and institutions, are projected by many analysts to be substantially weaker going forward, while exceedingly low interest rates suggest limited future returns for bonds (Perianan, 2020). Many other classes of alternative investments, such as hedge funds and real estate, have struggled in recent years to match market benchmarks. Concurrently, many public pension funds are facing severe shortfalls, and other institutional investors—from university endowments to sovereign wealth funds—are seeking additional funds to fulfill ambitious agendas. As a result, institutions are increasingly looking to private capital investments such as venture capital, buyout, and growth funds. The global private capital pool reached US\$714 billion in 2018, up from US\$324 billion a decade ago (Bain & Company, 2019).

This growth has, in turn, raised questions about the consequences of these investments for companies, workers, and the economy more generally (Private Capital Research Institute, 2017). In particular, policymakers have enacted and proposed several initiatives in the past decade to address the perceived harms of private equity. For example, the European Union implemented an Alternative Investment Fund Managers Directive to prevent asset stripping from private firms after acquisition by private equity or other financial sponsors (see Chapters IV and V, especially Chapter V, Section 2, Articles 26–30).<sup>4</sup> As another example, the European Central Bank (ECB) guidance on leveraged transactions<sup>5</sup> requires stringent internal review of “all types of loan or credit exposures where the borrower is owned by one or more financial sponsors.”<sup>6</sup> Additionally, in 2019 United States Senator Eliz-

---

<sup>4</sup>See the directive: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011L0061&from=EN> (accessed 2020-12-11).

<sup>5</sup>See the guidance: [https://www.banksupervision.europa.eu/ecb/pub/pdf/ssm.leveraged\\_transactions\\_guidance\\_201705.en.pdf](https://www.banksupervision.europa.eu/ecb/pub/pdf/ssm.leveraged_transactions_guidance_201705.en.pdf) (accessed 2020-12-11).

<sup>6</sup>See Section 3 of the ECB Guidance, which states that “Syndicating transactions presenting high levels of leverage … should remain exceptional … and form part of the credit delegation and risk management escalation framework of the credit institution.”

abeth Warren introduced the Stop Wall Street Looting Act (see Section 3(13) of the Act) to broadly regulate private equity in the United States.<sup>7</sup> Fears about the high indebtedness of buyouts and their potential risk to the stability of the financial system animated United States regulatory guidance of leveraged lending to facilitate buyouts and post-buyout activities of target firms.

Although the global economy and individual investors are increasingly dependent on private capital, much remains poorly understood about these investments. A salient aspect of private capital is that it is indeed private. Traditionally, the general partners (GPs) who manage these funds have not disclosed much information to the United States Securities and Exchange Commission, other regulators, or even to their own investors (limited partners, or LPs). A shortage of reliable industry data leads to an unappealing setting where industry advocates make sweeping claims about the benefits and critics make broad charges on very shaky empirical foundations (Kaplan and Lerner, 2017).

This lack of transparency has led to two important barriers to private capital research. First, there have been barriers to entry: it has been difficult for academic researchers, graduate students and junior faculty, to get access to these records. Second, much of the research has been undertaken using commercial databases (most notably, Thomson Reuters, which has a licensing program, and Burgiss, which has made its data available to the Private Equity Research Consortium<sup>8</sup>) based at the University of North Carolina at Chapel Hill or else using data provided to researchers directly by limited and general partners on a one-off basis (e.g., Gompers and Lerner, 1997).

It is typically difficult to compare any but the basic facts about the various commercial databases, as these databases draw from different sources, some of which may be proprietary.<sup>9</sup> As a result, there are contradictory findings on a number of important topics, such as the

---

<sup>7</sup>See the act: <https://www.warren.senate.gov/imo/media/doc/2019.7.17%20Stop%20Wall%20Street%20Looting%20Act%20Text.pdf> (accessed 2020-12-11).

<sup>8</sup><http://uncipc.org/index.php/initiativecat/private-equity/> (accessed 2020-12-11).

<sup>9</sup>For recent efforts along these lines, see Brown et al. (2015), Maats et al. (2011), and Kaplan and Lerner (2017).

risk-adjusted performance of private equity and the extent of persistence of performance of funds: the differences in results appear to be at least in part a function of the differences between databases. On the former topic, see Korteweg (2019); on the latter, see Braun, Jenkinson and Stoff (2017), Harris et al. (2014), and Korteweg and Sorensen (2015). These issues are akin to the more general issues of access to private data raised by the American Economic Association's Committee on Economic Statistics.<sup>10</sup>

The Private Capital Research Institute (PCRI) was initially founded to address these data issues and to provide a greater fact-based understanding of private capital's global impact. Thus, the PCRI's goal is to create a standardized database on the private capital industry. Since 2010, an important part of the PCRI effort has been the building of a series of comprehensive private capital databases to serve as the foundation for independent analysis of the economic impact of private capital and the performance of funds and individual transactions.

The PCRI uses two strategies to gather data. One approach is to collect data directly from primary and secondary sources: the private capital firms themselves and commercial data vendors, respectively. Thus, a large part of this process has been formulating licensing agreements with the two types of data providers. An alternative strategy to further address these data issues is to gather data on private firms from public filings.

An example of this second strategy is a recent initiative of the PCRI: the creation of a library of CoIs and related documents to allow more researchers to explore the important topics in this area. A few academic papers have utilized the information found in CoIs to explore questions around capital structure and contractual terms of private capital investments and corporate governance issues. These studies, however, have used extremely limited proprietary data sets, making it difficult to replicate or refute the studies.

---

<sup>10</sup>The American Economic Association's Committee on Economic Statistics issued a report in March 2020 illustrating some of these points (AEA Committee on Economic Statistics, 2020).

For example, one of the pioneering academic papers to explore topics in this area is Kaplan and Stromberg (2003), which examines 213 venture capital investments in 119 portfolio companies by fourteen VC firms and provides an empirical analysis of the contracts used. To obtain the data for this study, the researchers created a proprietary data set by asking fourteen VC firms to provide detailed information on their portfolio companies, which included financing terms, the firms' equity ownership, and contingencies to future financing. Two concerns—acknowledged by the authors—were that the firms that were willing to share their data were not necessarily representative of the universe of venture firms and that these firms may select non-representative transactions to share. A similar critique can be offered of Lerner and Schoar (2005), which employed a very related methodology.

Furthermore, an alternative approach has been to use a selected sample of CoIs for firms collected by VC Experts, a commercial data vendor that collects data on a contractual basis. For instance, Bengtsson (2011) studies the restrictive covenants in 182 venture capital contracts. Chernenko, Lerner and Zeng (2019) study the implications of mutual funds making private investments in firms, an activity that has historically been done by venture capital firms. Again, they use VC Experts data, focusing on approximate unicorn (or near-unicorn) firms (privately held companies with valuations greater than US\$1 billion). While these authors have been able to negotiate for access from VC Experts, the process was protracted, expensive, and highly limited in scope. Other academics have attempted to get access to these data and been unable to obtain it. Moreover, the representativeness of the sample of CoIs collected by VC Experts seems unclear. Given the substantial access problems associated with this data source, the PCRI believes there is a huge opportunity to create a resource that is a broadly available resource to academics.

The PCRI's CoIs collection process mitigates these concerns. First, the PCRI does not rely on proprietary data from specific VC firms who are willing to share their data. Instead, the PCRI creates a random sample of venture-backed portfolio companies and manually collects the CoI documents from the states in which the firms were originally incor-

porated. However, as a result, researchers only interested in studying a specific set of companies would unlikely find the companies' documents available in the PCRI CoI library. Second, the PCRI makes its CoI database available without charge to all academic researchers with a credible research agenda.

It is virtually impossible to pinpoint the exact size of the private equity industry or to verify the completeness of any data set. However, the PCRI universe is one of the most comprehensive and complete databases on private capital funds and transactions.<sup>11</sup> The unique feature of the PCRI database is that it draws from multiple data sources, including the private capital firms themselves, several commercial data vendors, private capital associations, limited partners, and the PCRI's own research.

### **10.2.2 Data Use Examples**

The PCRI databases are available for use by academic researchers for academic research purposes only. As of May 2020, over 25 academic researchers were using PCRI databases.

As mentioned, one of the primary objectives of the PCRI is to promote a better understanding of the private capital industry. The PCRI hopes to encourage research in this area through access to its databases. Highlighted below are research projects that have been submitted for publication or are near completion:

- Researchers Steven J. Davis (University of Chicago), John Haltiwanger (University of Maryland), Kyle Handley (University of

---

<sup>11</sup>The Private Equity Growth Capital Council (2013) reported that 2,800 private equity firms were headquartered in the US investing in buyout, growth equity, infrastructure, and energy funds. Over the same time period, the PCRI database has recorded 1,600 US private capital firms that solely invest in buyouts. In addition, the National Venture Capital Association reported 874 US venture capital firms were in existence in 2013 with 1,331 VC funds and US\$192.9 billion under management. By comparison, for 2013, the PCRI database has 2,082 US venture capital firms seeking investments. Some of the differences between the PCRI database and the reports can be explained by different firm-type classifications (for example, it is challenging to distinguish growth equity firms, which are often classified as venture capital), as well as the fact that the PCRI is missing firm type for about 30 percent of the data (Jeng and Lerner, 2015).

Michigan), Josh Lerner (Harvard Business School), Ben Lipsius (University of Michigan), and Javier Miranda (United States Census Bureau) completed a major project titled “The Social Impact of Private Equity Over the Economic Cycle” (Davis et al., 2019). They explored the broader economic effects of private equity buyouts over business cycles, a topic in which there has been very little investigation but is critical to understand for the regulation of private equity buyouts as well as leveraged bank lending. Using the PCRI’s private equity buyout transaction data matched to Census Bureau microdata, their research answers some important questions about how private equity buyouts affect employment growth and the pace of job reallocation and wages.

- Professor Andrea Rossi at the University of Arizona’s Eller College of Management submitted for publication his paper titled “Decreasing Returns or Reversion to the Mean? The Case of Private Equity Fund Growth” (2019). In April 2019, Rossi presented this paper at the European Investment Forum held at the University of Cambridge and sponsored by FTSE Russell and was awarded a runner-up award (best five papers, technically). This paper explores the phenomenon that when a private equity firm raises a larger fund, performance tends to decline relative to the previous funds it managed. Rossi uses the PCRI’s data on portfolio investments to study the relationship between the amount of investments a fund makes and the fund’s size.
- Jun Chen from California Institute of Technology completed his study on the scale, scope, and dynamics of non-VC early-stage financing markets. In particular, the paper “What role does angel finance play in the early-stage capital market” (2017) used PCRI data to examine the interaction between the ways in which angel investors complement and substitute for venture capital financing and the broader economic implications on how to promote economic growth and entrepreneurship. In this paper, Chen assembles the first comprehensive data set on angel financing and characterizes its size, scope, and role in the early-stage capital market.
- PCRI data was used in 2017 in the latest research on Smart So-

cieties and the accompanying *Harvard Business Review* article of the Digital Planet initiative at the Fletcher School at Tufts University (Chakrovorti and Chaturvedi, 2017).<sup>12</sup> PCRI data helped form part of a benchmark that would assist policymakers to better identify their country's investment environment. The PCRI data are combined with several measures to provide a more complete view of what is actually happening in terms of investments, especially investments in technology.

## **10.3 Making Data Useable for Research**

### **10.3.1 Collecting Data on the Private Capital Industry**

The PCRI collects information on private capital firms, funds, portfolio companies, transaction data, and investment exits. In particular, the Institute focuses on buyouts, growth equity, and venture capital investing. One of the main strengths of the data collection strategy is that it relies on gathering data from multiple sources to mitigate sample selection biases. In the future, the PCRI would like to include more information on angel investments and sovereign wealth funds.

PCRI's first goal is to gather data directly from the private capital firms. In the outreach to these firms, the Institute has relied primarily on the relationships of its team members with the individual private capital firms. Thus, the PCRI has had to approach each private capital firm one at a time—a time-consuming endeavor. To date, the Institute has approximately fifty groups that have provided data or are in the process of contracting to do so. It might be questioned why private equity firms would be willing to share data with the PCRI when the commercial databases have often struggled to get data from these institutions. There are several answers:

- There are constraints that the PCRI places on the use of the data. In particular, the PCRI is designed to be a project run by academics

---

<sup>12</sup>For research reports, see <https://sites.tufts.edu/digitalplanet/> (accessed 2020-06-21).

and for academics. The information is used exclusively for academic research rather than for any commercial purpose.

- The research protocol simultaneously allows academics to undertake high-quality research while protecting the confidentiality of the data provided by the private equity firms. The PCRI follows the model employed by the Census Bureau for making information available: academics can undertake detailed cross-tabulated analyses but not download or view individual data entries. Essentially the academics are able to upload queries and download results without “touching” the individual data entries.
- The private equity industry has been under much scrutiny. In particular, in the aftermath of the financial crisis there has been much greater attention to institutions such as hedge funds and private capital groups that traditionally were exempt from most regulatory oversight in the United States and Europe. As a result of these pressures, industry leaders have increasingly appreciated the need for high-quality independent research.

Gathering information directly from private capital firms has its own limitations. Even if every active group chose to participate, there would still be some groups that have gone out of business and no longer keep their records or would be difficult to contact. In addition, as the PCRI began collecting data from individual private capital firms, one of the major concerns raised was that it would take too long for the PCRI to get a database large enough to disguise the data to preserve anonymity. The PCRI thus realized the importance of quickly building a large foundation for the database. As a result, the Institute is complementing the data that is gathered from the private capital firms with data from commercial sources, even if it is not always of the same quality as that provided directly by the general partners.

The commercial sources include the Emerging Markets Private Equity Association (EMPEA), Alternatives Data Cell (“Alternatives”), Refinitiv (formerly Thomson Reuters Financial & Risk ), Unquote (a UK-based data collection company acquired by Mergermarket), Start-Up Nation Central (a company that focuses on collecting data on Israeli private equity transactions and funds), and Venture Intelligence (a leading

**Table 10.1:** Number of distinct private capital firms provided by source of information

Vendor	Distinct Private Capital Firms
EMPEA	2,964
NYPPEX	6,100
Thomson Reuters	11,491
Unquote	5,291
<b>PCRI Unique</b>	<b>17,633</b>

Note: As of 2015. Source: Jeng and Lerner (2015).

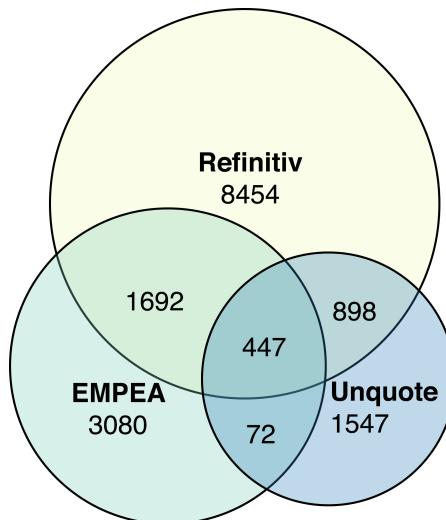
source of information on private company financials, private capital transactions, and their valuations in India). Table 10.1 provides the coverage of information for the PCRI's original top four sources.<sup>13</sup> Refinitiv has the largest coverage of private capital firms with 11,491 firms. By combining the sources and eliminating duplicates, the PCRI finds that the overlap of private capital firms in the databases is roughly 32 percent.<sup>14</sup> After eliminating duplicates, the PCRI combined data set contains 17,633 *unique* private capital firms. Figure 10.1 shows a diagram of the overlap between the sources. The key features of the PCRI database are summarized in Jeng and Lerner (2015).

### 10.3.2 Processing of Data Received From Data Vendors or Primary Sources (LPs or GPs)

The process of combining and cleaning the various data sources is an arduous task. At the PCRI, the research staff consists of one full-time director of research, two full-time research associates, and approximately six part-time undergraduate research assistants. Research

<sup>13</sup>PCRI focuses on the four original sources: EMPEA, Alternatives, Refinitiv, and Unquote. Start-up Nation Central and Venture Intelligence were added later and represent a small fraction of the total database.

<sup>14</sup>As of 2015. The majority of the overlap in the data is between the Alternatives and the Refinitiv data sets. By not including the Alternatives database in this analysis, the number of unique capital firms is 16,190 (Jeng and Lerner, 2015).



**Figure 10.1:** Overlap of private capital firms in PCRI database by source of information

associates work to understand, clean, and research the various data sources as well as to make those sources consistent and to develop a file-matching protocol. Part-time undergraduate research assistants help with name matching, researching missing data items, and manually collecting data.

The PCRI databases include data for over 17,000 private capital firms, 33,000 private capital funds, and 110,000 portfolio companies covering a time span from the early 1970s to 2018. The portfolio companies are geographically diverse with over 50 percent outside the United States, including 32 percent in Europe and 10 percent in Asia (Jeng and Lerner, 2015). The PCRI database contains eight different data tables: company, deal, exit, fund, fund performance, fund quarterly cash flow, general partner (GP), and investment. See Appendix A for more information on the database. Figure 10.2 provides details on the tables, including the variables in each data table and the relationship between the tables. For more information on the PCRI databases, please see the data user manual available on the PCRI website.

The PCRI has developed an internal data processing system that is used as a training guide for new research associates. The PCRI database

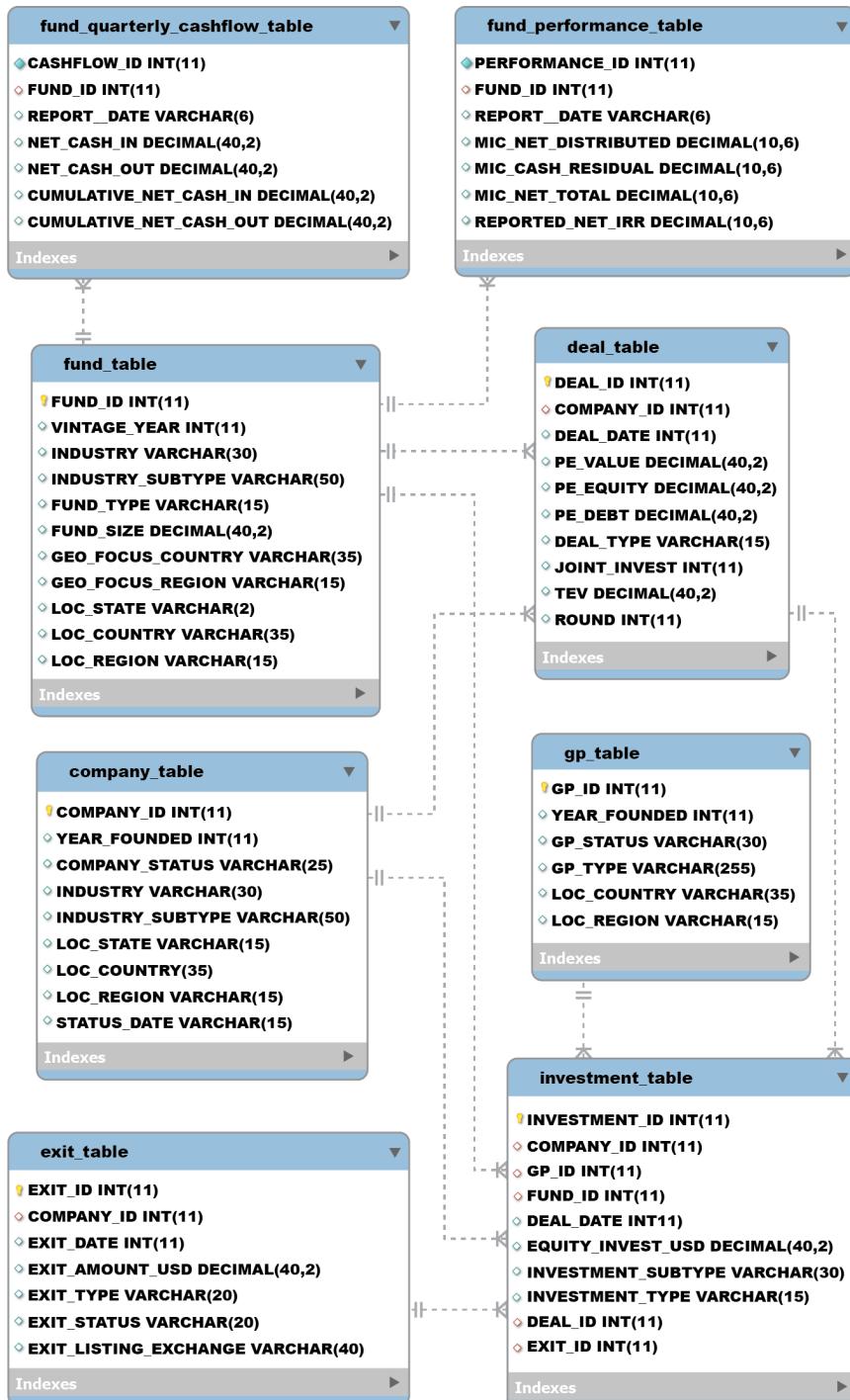


Figure 10.2: Table relationship diagram of PCRI database

is a relational database. When new data are received from either a commercial vendor or a private capital firm, the new data are separated into variables and put in a consistent format to be read in by Stata. Various rudimentary tables are created that correspond to tables in the PCRI database (e.g., investment, exit, GP, fund). For each data item in a table, the PCRI then identifies a key that is a unique identifier specific to a data provider. This key is then mapped to a unique identifier for each unique observation within that table (the source ID). This process, whereby the key of the data provider is mapped to a source ID, is referred to as local aliasing. These newly created source IDs are then mapped to a unique identifier within the larger PCRI database (the ID). This process is called global aliasing. The global aliasing file contains the unique identifier-name-source link.

Once this step is complete, all variables are processed in Stata to match existing codes in the PCRI database. For instance, country names are standardized to match the PCRI standard names in the supplemental tables (e.g., *Fr* would be converted to “France”).

The resulting tables are stored as Stata data files in a file location specific to that data provider. Next, in the append file, all the data files are aggregated into a set of long tables containing data collected from all data providers and independently researched information. The append file saves these files to the pre-stacked file location as .csv files. This is where the PCRI eliminates formatting inconsistencies (e.g., date formatting). Once these operations have been completed, the files are saved to a file location called Stacked within each data vendor’s directory, once again in .csv format.

The final consolidation stage is handled using a Python MySQL script. Wherever there are multiple sources for a variable, the PCRI creates a ranking system to determine the information to keep. An internal pecking order determines this ranking: PCRI internal research data ranks the highest, then GP-provided data, and finally commercial vendor data. The Python program also keeps track of discrepancies between data providers that meet certain thresholds (e.g., if a dollar amount invested differs by more than 10 percent, this data item is flagged). Discrepancies are added to a file for further research.

The Python script then runs a SQL query that builds the Merged Files. The data are anonymized (i.e., names and other identifiers are removed). Lastly, these .csv files are converted to Stata data files in the CSV to DTA do-file and saved to the Uploaded Files folder where they are accessible to researchers.

### **10.3.3 Collaboration for Certificate of Incorporation Data Acquisition**

In October 2018, we signed a memorandum of understanding with the Stanford Graduate School of Business (GSB) to collaborate with GSB Professor Ilya Strebulaev. The memorandum supports the development of a library of venture capital-backed firms' CoIs and related documents and a database of the information contained within these documents. Additionally, the memorandum encourages the diffusion of the library and database to academic researchers.

Using an agreed-upon methodology, the PCRI and Stanford GSB created an initial random sample of 622 venture-backed companies. In the summer of 2019, the PCRI completed collecting CoIs for this random sample and has made these documents available to researchers through an electronic library hosted on a new platform at the Harvard Business School (HBS) called SmartRoom. At the same time, Professor Strebulaev and his team are completing the coding of the documents for this random sample. This data set will also be made available for approved researchers via the National Opinion Research Center (NORC) at the University of Chicago. The NORC houses all of the PCRI databases behind a secure firewall.

The PCRI is currently working with the GSB to collect and code CoIs for a second random sample of 250 venture-backed companies. The Institute has created the list of companies for this sample and has received and processed half of the CoIs. In addition to these two random samples created in collaboration with the GSB, the PCRI independently has collected CoIs for another 750 companies, including many unicorns. The final sample will contain almost 2,500 companies.

### **10.3.4 Certificate of Incorporation Document Acquisition Process**

This section provides the details on the creation of the PCRI CoIs library. A CoI is a public legal document filed with the state in which the company is incorporated. It is essentially a license issued by a state government for a company to form a corporation. In addition, whenever a corporation receives private capital funding, an amended CoI is filed with the state in which it is registered. The availability of CoIs varies by state. Some states make this document available online. Other states require a written request and payment of a fee to obtain a copy (Masters, n.d.).

Based on consultation with academics interested in CoIs and colleagues at the GSB, the PCRI has opted to get a complete set of CoIs for each funding date/portfolio company pair because it gives a comprehensive overview of the firms' financing histories. To begin the process of obtaining CoIs, a PCRI research associate prepares a list of corporation names, addresses, and investment dates in an Excel spreadsheet. A research associate then determines where each corporation is incorporated/registered by going to the state website for any state in which a corporation maintains a business location. Once it is determined in which state a corporation is registered, a research assistant orders the appropriate documents from that state's website for business incorporations.

As seen in Table 10.2 below, the vast majority (82 percent) of companies in the database are registered in Delaware, which has been the focus of most of our efforts. To obtain documents for corporations registered in Delaware, order forms must be filled out and submitted online. Delaware charges US\$10 for the first page of a document and US\$2 for each additional page. On average, one document costs around US\$30 dollars, but sometimes a single document could cost up to hundreds of dollars. On average, the total cost to obtain all the documents for one company is approximately US\$125. In the case of Delaware, a hard copy of a CoIs arrives within several weeks of submitting a request.

Other states, such as New York, accept orders by mail. In the case

**Table 10.2:** Breakdown of state of incorporation of the 622 venture-backed portfolio companies

State	Percentage
Delaware	82.0%
California	10.0%
Georgia	1.0%
Washington	1.0%
Texas	0.7%
Colorado	0.7%
Other States	4.6%

of New York, a form must be filled out for each request. The cost is US\$5 for each plain copy of a document. Processing takes ten to twelve business days, not including shipping, and the copies are sent thereafter. For an unusually large order, arrangements can be made in advance, but the work is still performed on a first-come, first-serve basis. For other states, the CoIs are usually available online for no cost. Thus, the PCRI can access the business entity database for those states (e.g., California and Massachusetts) and download copies of the documents.

In most states, the Corporations Division of the Secretary of State's office handles business incorporations and related filings. In a handful of states, business registrations are handled by a different state agency. The United States Small Business Administration maintains a list of state business registrars to help find the appropriate state agency. When the PCRI receives a hard copy of a CoIs, the file is scanned and loaded to be stored electronically. At this time, Delaware and some other states only provide hard copies of documents, making the CoI acquisition process more laborious.

### 10.3.5 Metadata

Information about metadata is available in a PDF version of the PCRI Data User Manual on its website for use by academic researchers.<sup>15</sup> The PCRI has a staff dedicated to maintaining the PCRI databases. The Institute revises the databases annually to account for data updates from the data provider and then places updated researcher-accessible files in a designated folder for users to access. The PCRI data are stored as Stata data files. The manual provides a detailed description of the eight different data tables, specifically company, exit, fund, fund performance, fund quarterly cash flow, GP, investment, and deal tables. This reference also cites the location of the files and the size of each file (i.e., number of rows). Additionally, the manual catalogs the names of the different variables and the definitions contained within each table. Furthermore, the manual indicates whether a data item is a Primary Key or Foreign Key. This distinction provides the link between two tables, facilitating the merging of different data tables. Primary Keys are unique within a table, whereas Foreign Keys are not unique but link to another table. Since the PCRI does not permit users to look at the data, the Institute provides a small, artificial sample of each data table in the manual.

## 10.4 Legal and Institutional Framework

### 10.4.1 Institutional Setup

The PCRI is an independent non-profit organization, which seeks to provide a greater fact-based understanding of private capital's global impact. The PCRI is governed by a Practitioner Advisory Committee, which is comprised of experts in the private capital industry, and an Academic Advisory Committee, consisting of leading researchers in the field. Representing diverse affiliations, these two committees provide guidance on the data collection process and periodically review

---

<sup>15</sup>[http://privatecapitalresearchinstitute.org/images/news/pcri\\_manual\\_2-4.pdf](http://privatecapitalresearchinstitute.org/images/news/pcri_manual_2-4.pdf) (accessed 2020-12-11).

and provide advice on the administration, finances, and research program of the PCRI. A subcommittee of the Academic Advisory Committee meets to approve research proposals requesting access to the PCRI databases. Projects are approved as long as they have a credible research agenda, are not for commercial purposes, and do not jeopardize the security of the data. The Practitioner Advisory Committee is not able to veto approved projects. A list of members of the Practitioner Advisory Committee<sup>16</sup> and Academic Advisory Committee<sup>17</sup> is available on the PCRI website.

#### **10.4.2 Legal Context for Data Use**

One of the biggest challenges faced in the data collection effort was creating a standardized licensing agreement for all data sources, in particular for the private capital firms. By working closely with the PCRI's lawyers at Debevoise & Plimpton and a few prominent private capital firms, a standardized licensing agreement was developed. The agreement not only allowed the PCRI to obtain, use, and administer highly confidential data but also alleviated the major concerns (i.e., confidentiality and data security) of the private capital firms. The authors highlight some of the chief features of the Private Equity Sponsor Data Agreement.<sup>18</sup>

First, the data licensing agreement grants the PCRI a royalty-free, non-transferable license. Second, the PCRI is permitted to receive, store, reproduce, and combine the data. Third, since the PCRI is a project run by academics and for academics, the PCRI database is to be used exclusively for academic research rather than for any commercial purpose. In accordance with the licensing agreement, the PCRI carefully reviews research proposals and monitors output files to ensure that data are being used appropriately. Additionally, as a primary objective of the PCRI is to promote unbiased, academic research, the licensing agreement

---

<sup>16</sup>List available at <http://privatecapitalresearchinstitute.org/advisory-committee.php> (accessed 2020-12-11).

<sup>17</sup>List available at <http://www.privatecapitalresearchinstitute.org/academic-advisory-board.php> (accessed 2020-12-11).

<sup>18</sup>A sample agreement can be found in the Online Appendix.

mandates that the data sponsors would not be able to limit the areas of academic research. However, under the licensing agreement, data sponsors can obtain a preview of working papers and are also given the option to be acknowledged for their contribution to the PCRI research effort. Lastly, the licensing agreement allows either party to terminate the agreement.

In cases where data disclosure harms data sponsors, liability issues were a major source of discussion in the creation of this licensing agreement. Given the limited resources of the PCRI, the licensing agreement puts a cap on PCRI's liability at US\$1,000, which only applies in cases not resulting from gross negligence, strict liability, fraud, misconduct, or misrepresentation on the part of PCRI. In such cases, there would be no liability cap. Also, the PCRI agrees not to bring any claims against any of the data sponsors.

While the PCRI has been successful with private equity groups, venture capital organizations have been much more resistant to sharing information. Thus, with a grant from the Alfred P. Sloan Foundation, the PCRI is taking an alternative approach to obtain information for researchers to understand venture capital activity. As mentioned previously, the Institute has been developing a compilation (and an associated taxonomy) of CoIs over the past two years. Extremely detailed information on venture capital transactions is available in CoIs, which are typically compiled by regulators in the state of the firms' incorporation. These corporate filings include important details on deal structure (i.e., the capital structure and key terms) as well as important valuation information. While this information is publicly available, the costs to obtain these documents are prohibitively expensive. For example, in some locations, a request for these documents must be made in person, and US\$1.00 to US\$2.00 is charged per page. As a parallel process, the PCRI is working to create a data set containing the twenty to thirty most critical variables contained in these documents.

### **10.4.3 Collaboration with the Census Bureau**

Over the past several years, the PCRI has been working with the Census Bureau team regarding ways in which the PCRI data could be integrated with the Longitudinal Business Database (LBD). In the spring of 2016, the PCRI signed a confidentiality agreement that allowed it to tie certain portfolio-company level data to the LBD in a way that would respect the various confidentiality restrictions on the government and PCRI data but greatly increase the usefulness to researchers. The agreement established between the Census Bureau and the PCRI allows for the sharing of specific PCRI data items, which include portfolio company name, location information, and a PCRI source code (PCRI Data). The purpose was to link the data to the Census Bureau Business Register. As of May 2020, one academic is using the Census-PCRI linked data to explore the effect of acquisitions of startups on their newly acquired employees.

### **10.4.4 Legal Framework for Granting Data Access**

Approved data users interested in using either PCRI data or the coded CoIs data are required to sign a data use agreement (DUA) between both the PCRI and its data host, the NORC. The contract is standard and has been vetted by a few users, but there is some flexibility for negotiation. If changes are made, they are reviewed extensively by the NORC legal staff and the PCRI. However, very few users have requested any changes. Data users are granted access for a term of one year; at the end of the term they are able to request an extension. Access to the data can be revoked if the NORC deems that any aspect of the DUA is violated. The NORC and PCRI reserve the right to ensure that the PCRI data are used in compliance with the agreement. The output is reviewed to ensure that it preserves the anonymity of the individual observations and that the analyses match the original intent of the research proposal. Additionally, according to the DUA, both the user and the NORC have the right to terminate the DUA without cause at any time. Currently, there is no cost to accessing the PCRI databases, as

the fixed fee paid by the PCRI to the NORC accommodates a generous number of users.

The PCRI maintains the intellectual property (IP) rights on the architecture, computing systems, and computing environment of the data enclave, including the data set and all other data, information, documents, programs, trade secrets, and confidential information. However, the PCRI does not assert IP rights on products created by the data user, such as research papers and independent data analyses.

## **10.5 Protection of Sensitive and Personal Data: The Five Safes Framework**

### **10.5.1 Safe Projects**

Before obtaining access to the data, interested academic researchers are required to submit a two-to-three-page written research proposal. This proposal must clearly state the objective of the project, the PCRI data that would be used in the study, and the research methodology. A subcommittee of the PCRI Academic Advisory Committee evaluates research proposals submitted to the PCRI. The subcommittee safeguards the data by ensuring that the proposals use the data for only academic research purposes. Users may not use the PCRI databases for commercial purposes. The review process typically takes less than a week and is free of charge. The PCRI is not required by the DUA to obtain consent from the data providers before approval of a research project.

For researchers interested in gaining access to linked Census-PCRI data, the same protocol is used—researchers are required to submit a research proposal for approval by a subcommittee of the PCRI Research Advisory Committee. Again, the subcommittee reviews projects to ensure the safety of the data and that the data use is for only academic research purposes. For the combined data stored at the Census Bureau, approved researchers also need to follow the access protocol at the Census Bureau.

## **10.5.2 Safe People**

Only approved users are granted access to the PCRI databases. The PCRI primarily accepts research proposals from academic researchers from accredited universities, not-for-profit research organizations, and research groups of government organizations. Academic researchers applying for access do not need to obtain special training to use the data. Additionally, to protect the confidentiality of the data, approved academic researchers must sign a DUA with the PCRI and the NORC. Currently, the PCRI does not place a limit on the number of users that it is able to host.

## **10.5.3 Safe Settings**

The security of the PCRI data is paramount. The Institute's ability to obtain data from the various data vendors and private equity firms rests primarily on the ability to maintain the security and confidentiality of their data. The PCRI has designed a protocol that simultaneously allows academics to undertake high-quality research while protecting the confidentiality of the data provided by the data sponsors. To this end, PCRI's first step was to host the PCRI databases at the NORC, which has experience hosting highly sensitive federal (e.g., Medicare) and private sector data (e.g., hedge fund data used by federal investigators as part of the Financial Crisis Inquiry Commission). The PCRI initially explored creating its own platform on its own servers, but it was too costly to construct and maintain. As mentioned in section 10.4.4, the PCRI and NORC reserve the right to ensure that the PCRI data are used in compliance with the DUA agreement. Thus, privileges to access the data may be revoked if the NORC deems that any aspect of the DUA is violated.

The data are accessed using a secure remote access protocol. Users login through the NORC portal, which requires a two-factor authentication: a password and a security token password. The PCRI provides Stata software to the researcher to access the data files. To ensure safety, the PCRI employs a methodology whereby academics

can undertake detailed cross-tabulated and regression analyses but not download or view individual data entries. Also, for added security, the Institute disables certain features of Stata (e.g., outsheets, list, and browse commands) that could allow the user to identify companies in the database. The cost of hosting the data at the NORC is around US\$40,000 per year.

The Census data typically can only be accessed at Federal Statistical Research Data Centers (FSRDCs). Researchers require approval from both the PCRI and the Census Bureau.

For the CoIs library, the PCRI has created a searchable database using SmartRoom, a content management platform, for researchers to view, but not download, the documents. The Institute does not charge for the use of this CoIs library or the ultimate coded database that the Stanford GSB is creating. The cost to use the SmartRoom is about US\$10,000 per annum and varies depending on the number of users. The Institute chooses to use this platform not only because of its security features but also its user-friendly interface.

#### **10.5.4 Safe Data**

To make the data safe, the PCRI databases are anonymized (i.e., data are de-identified), and only PCRI research staff have access to identified data. Furthermore, researchers cannot see the data as they are unable to print, browse, or outsheets the data. Researchers are only able to run queries and view the results of analyses.

External data sets can be linked to the PCRI databases. Data sets can be uploaded to the NORC platform, requiring a PCRI staffer to do the matching. Only in exceptional circumstances can data be downloaded from the servers and matched to an external database. For instance, in collaboration with the Census Bureau, certain PCRI variables were linked to Census data and then stored with the Census Bureau. The Census Bureau assumes an obligation to use the PCRI data for only statistical purposes and requires maintaining data confidentiality.

### **10.5.5 Safe Outputs**

The PCRI-NORC DUA outlines how the PCRI data can be used. Essentially, the researchers can upload queries without “touching” the individual data entries. Even though the data are anonymized, the PCRI further protects the data by prohibiting the viewing of individual observations. The NORC helped us limit STATA so that only summarized outputs can be viewed provided there are at least 100 observations used in each analysis. Outputs can be downloaded off the NORC platform but must first be approved by a PCRI staff member. Furthermore, software program log files provide a paper trail of activity, which is monitored periodically by NORC staff and PCRI to ensure that the data are being used for research purposes.

## **10.6 Data Life Cycle and Replicability**

### **10.6.1 Preservation and Reproducibility of Researcher-Accessible Files**

The PCRI periodically releases new versions of the databases. As of May 2020, the database is on Version 2.5. The versioning is based on content and, if necessary, structure updates. Each version is preserved and archived and can be made available upon request for replication purposes. When a new version is released, it is copied and uploaded to a separate folder for sharing on the NORC data platform.

Researcher-accessible files can be regenerated. The data are processed using a series of Python and Stata scripts to unpack raw data files from the various sources and to put them in a useable format. The PCRI receives new data feeds periodically and thus regenerates the data files annually.

### **10.6.2 Preservation and Reproducibility of Researcher-Generated Files**

Researcher-generated files are preserved within each researcher’s directory/folder and are backed-up regularly. Files are not shared

amongst the researchers unless a researcher gives permission subject to PCRI approval.

### **10.6.3 Disposal of Data**

The PCRI is not required to delete data from data providers. Upon termination of any agreements, the PCRI would be required to purge the sponsor's data from future versions of the databases within a reasonable time frame. However, any previously approved researcher would continue to have access to the previously consolidated databases. Also, the data sponsor agreements could be reassigned to another non-profit such as another academic institution, provided the PCRI gives the data sponsors prior written notice.

## **10.7 Sustainability and Continued Success**

### **10.7.1 Outreach**

As mentioned in section 10.3.1, outreach to private equity sponsors is a slow process because each potential data provider is contacted individually. Moreover, these private equity firms are very wary of sharing data and need to have their legal departments review the data provider agreement, necessitating frequent communication between the parties. One benefit of participation that is highlighted to the data sponsors is that they would be able to obtain a preview of working papers and would be invited to attend PCRI conferences featuring the Institute's research.

Going forward, PCRI hopes to make this process more efficient by working with some organizations that are already actively collecting information from general partners. Such organizations include a large custodian bank with whom the PCRI has signed a data sharing agreement and national venture capital associations.

To create more awareness of the PCRI databases and to get more researchers using the databases, a call for research proposals has been

posted on the SSRN (Social Science Research Network). In addition, the PCRI has participated in numerous conferences (American Economic Association, the American Library Association, and the National Association for Business Economics) to present the PCRI's mission and talk more about the PCRI databases. Lastly, the PCRI hosts conferences twice a year to bring together industry leaders, academics, and policymakers to discuss relevant topics in the private capital industry. For outreach purposes, summaries of the conferences are released on the PCRI's website.<sup>19</sup>

A recap of some of the more recent dissemination activities is highlighted here:

- On October 11, 2019, the PCRI, along with the Private Capital Project at Harvard Business School, sponsored a small workshop entitled "The Rise of the Asset Owner-Investor in Private Markets" on the HBS campus. The past decade has seen an extraordinary surge of interest on the part of asset owners in direct investing in private markets. Not only are these institutions investing in traditional funds, but they are eager to build up their own capabilities to invest. The motivations for these initiatives include a desire to avoid the fees charged by traditional partnerships, the belief that their long-run time horizons will facilitate the identification of attractive investment opportunities, and the quest to better manage the assets in their portfolios.
- On June 21, 2019, the PCRI partnered with the PBC School of Finance, Tsinghua University to bring together a group of industry thought leaders in Beijing to share perspectives on the changing landscape of private capital in China. Over the past three decades, the private capital industry in China has grown and evolved: it is now a US\$1.6 trillion industry with over US\$94 billion in private equity investment value last year alone. The maturing of the industry has challenged both Chinese GPs and LPs to rethink their value creation strategies and their relationships with each other.
- On September 11, 2018, the PCRI partnered with the Private Capital Project and the Impact CoLaboratory (Impact CoLab), both at

---

<sup>19</sup><http://www.privatecapitalresearchinstitute.org/> (accessed 2020-12-11).

the Harvard Business School, to bring together a group of industry thought leaders—comprised of prominent limited partners, general partners, and academics—to discuss the rise of impact investing. While some investors understand the basic concept of impact investing, there is still widespread confusion about the practice, its various approaches, and the difference it can make.

- On June 15, 2018, the PCRI and the Institute for Business Innovation (of the Haas School of Business at the University of California, Berkeley) co-sponsored a roundtable discussion on “New Models of Entrepreneurial Finance.” Industry leaders (GPs and LPs) and academics were brought together to discuss new developments in early-stage (e.g., syndicated angel investments) and later-stage (e.g., direct investments by sovereign funds) financing, how venture groups are responding to increased competition, and what the implications are for entrepreneurs and society more generally.

### **10.7.2 Revenue**

The PCRI is a non-profit that relies entirely on grants and strategic partnerships to fund its endeavor. It currently receives funds from the Alfred P. Sloan Foundation and the HBS Private Capital Project.

The PCRI does not charge a fee for the use of its databases as the fixed fee agreement with the NORC allows for a certain number of users. If the maximum number of users is exceeded, there would be a charge to cover additional costs.

### **10.7.3 Metrics of Success**

The PCRI gauges success on three fronts: (1) building a comprehensive database of private capital information; (2) sponsoring independent academic research on questions of relevant policy interest; and (3) arranging thought leadership forums that bring together academics, policymakers, regulators, investors, and industry practitioners to examine private capital’s role in the economy.

To date, the PCRI database contains global information on approximately 23,000 general partners, 44,000 funds, 141,000 portfolio companies, 432,000 investments, 229,500 private capital deals, and 36,000 private equity exits that are tied to deals. Given the strategy of combining data from multiple sources, this is one of the most comprehensive private capital databases currently available. As of May 2020, the PCRI has more than 25 active academic researchers utilizing the PCRI databases, resulting in several successful research papers submitted to academic journals.

#### **10.7.4 Concluding Remarks**

An increasing share of economic activity today is taking place in settings that elude traditional federal data collection mechanisms or fail to capture the richest of the activity at work. Against this backdrop, economists are increasingly turning to private data. This chapter underscores the experience of the PCRI, specifically the process of creating a database to facilitate access to private equity information for academics to address the myriad major concerns regarding private data. While this effort is certainly a work in progress, hopefully the experience can guide researchers who want to address similar issues in other fields.

## About the Authors

Josh Lerner is the PCRI Director and the Head of the Entrepreneurial Management Unit and the Jacob H. Schiff Professor of Investment Banking at Harvard Business School. He graduated from Yale College with a special divisional major that combined physics with the history of technology. He worked for several years on issues concerning technological innovation and public policy in various settings, including the Brookings Institution, a public-private task force in Chicago, and Capitol Hill. He then earned a PhD from Harvard's Economics Department. Much of his research focuses on venture capital and private equity organizations. (This research is collected in three books: *The Venture Capital Cycle*, *The Money of Invention*, and *Boulevard of Broken Dreams*.) He also examines policies on innovation and how they impact firm strategies. (That research is discussed in the books *Innovation and Its Discontents*, *The Comingled Code*, and *The Architecture of Innovation*.) He co-directs the National Bureau of Economic Research's Productivity, Innovation, and Entrepreneurship Program and serves as co-editor of their publication *Innovation Policy and the Economy*.

Leslie Jeng is the PCRI Director of Research and has been with the PCRI since its inception. She graduated from the University of Pennsylvania with degrees in finance and mathematics and has a PhD in Business Economics from Harvard University. In addition, she has work experience in academia as well as the private sector in investment banking and quantitative investment research. Her areas of interest are corporate finance, financial institutions and regulation, and public policy. Her publications include "The Determinants of Venture Capital Funding: Evidence Across Countries," *The Journal of Corporate Finance*, 2000 and "Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective," *The Review of Economics and Statistics*, 2003.

Therese Juneau is a Research Assistant at the PCRI. She graduated from the University of Texas at Austin in 2019 with a bachelor of science and arts degree in mathematics.

## Acknowledgements

We would like to thank the Kauffman, Sloan, and Smith-Richardson Foundations, as well as numerous data sources, for their support of PCRI.

## References in Chapter 10

- AEA Committee on Economic Statistics.** 2020. “Successful Private Firm-Academic Researcher Arrangements for Access to and Replicable Use of Private Data.” American Economic Association report. <https://www.aeaweb.org/content/file?id=11794>.
- Bain & Company.** 2019. “2019 Global Private Equity Report.” [https://www.bain.com/contentassets/875a49e26e9c4775942ec5b86084df0a/bain\\_report\\_private\\_equity\\_report\\_2019.pdf](https://www.bain.com/contentassets/875a49e26e9c4775942ec5b86084df0a/bain_report_private_equity_report_2019.pdf) (accessed 2020-10-05).
- Bengtsson, Ola.** 2011. “Covenants in Venture Capital Contracts.” *Management Science*, 57: 1926–1943. <https://doi.org/10.1287/mnsc.1110.1409>.
- Braun, Reiner, Tim Jenkinson, and Ingo Stoff.** 2017. “How persistent is private equity performance? Evidence from deal-level data.” *Journal of Financial Economics*, 123: 273–291. <https://doi.org/10.1016/j.jfineco.2016.01.033>.
- Brown, Gregory W., Robert S. Harris, Tim Jenkinson, Steven N. Kaplan, and David Robinson.** 2015. “What do different commercial data sets tell us about private equity performance?” SSRN 2701317. <https://ssrn.com/abstract=2701317> (accessed 2020-10-05).
- Chakrovorti, Bhaskar, and Ravi Shankar Chaturvedi.** 2017. “The “Smart society” of the future doesn’t look like science fiction.” *Harvard Business Review*. <https://hbr.org/2017/10/the-smart-society-of-the-future-doesnt-look-like-science-fiction> (accessed 2020-10-05).
- Chen, Jun.** 2017. “What role does angel finance play in the early-stage capital market?” California Institute of Technology Ph.D. thesis Chapter 1. <https://doi.org/10.7907/3mg0-2622>.
- Chernenko, Sergey, Josh Lerner, and Yao Zeng.** 2019. “Mutual funds as venture capitalists? Evidence from unicorns?” European Corporate Governance Institute Finance Working Paper 675/2020. <http://doi.org/10.2139/ssrn.2897254>.
- Davis, Steven J., John Haltiwanger, Kyle Handley, Ben Lipsius, Josh Lerner, and Javier Miranda.** 2019. “The social impact of private equity by buyout type and stage of the economic cycle.” <https://www.aeaweb.org/conference/2019/preliminary/paper/5nsZRYTz#:~:text=Abstract%3A%20We%20study%20the%20impact,per%20worker%2C%20and%20labor%20productivity.&text=Slower%20GDP%20growth%20after%20the,a%20widening%20of%20credit%20spreads>.
- Gompers, Paul, and Josh Lerner.** 1997. “Risk and reward in private equity investments: The challenge of performance assessment.” *The Journal of Private Equity*, 1(2): 5–12. <https://www.jstor.org/stable/43503183>.
- Gurung, Anuradha, and Josh Lerner.** 2008. “Globalization of alternative investments.” World Economic Forum Working papers.
- Harris, Robert S., Tim Jenkinson, Steven N. Kaplan, and Ruediger Stucke.** 2014. “Has persistence persisted in private equity: Evidence from buyout and venture capital funds.” Darden Business School Working Paper. <https://ssrn.com/abstract=2304808>.
- Jeng, Leslie, and Josh Lerner.** 2015. “Insights into the private capital industry using

the private capital research institute database. Part 1: Data sources and overview of private capital firms, funds, and portfolio companies.” [http://www.privatecapitalresearchinstitute.org/images/news/data\\_description%20article%201.pdf](http://www.privatecapitalresearchinstitute.org/images/news/data_description%20article%201.pdf) (accessed 2020-10-05).

**Kaplan, Steven N., and Josh Lerner.** 2017. “Venture capital data: Opportunities and challenges,” chapter in measuring entrepreneurial businesses: Current knowledge and challenges.” National Bureau of Economic Research Working Paper w22500. <https://www.nber.org/papers/w22500>.

**Kaplan, Steven N., and Per Stromberg.** 2003. “Financial contracting theory meets the real world: An empirical analysis of venture capital contract.” *The Review of Economic Studies*, 70(2): 281–315. <https://doi.org/10.1111/1467-937X.00245>.

**Korteweg, Arthur.** 2019. “Risk Adjustment in Private Equity Returns.” *Annual Review of Financial Economics*, 11(1): 131–152,. <https://doi.org/10.1146/annurev-financial-110118-123057>.

**Korteweg, Arthur, and Morten Sorensen.** 2015. “Skill and luck in private equity performance.” *Journal of Financial Economics*, 124(3): 535–562. <https://doi.org/10.1016/j.jfineco.2017.03.006>.

**Lerner, Josh, and Antoinette Schoar.** 2005. “Does legal enforcement affect financial transactions? The contractual channel in private equity.” *Quarterly Journal of Economics*, 120(1): 223–246. <https://doi.org/10.1162/0033553053327443>.

**Maats, Frederike H.E., Andrew Metrick, Ayako Yasuda, Brian Hinkes, and Sofia Vershovski.** 2011. “On the Consistency and Reliability of Venture Capital Databases.”

**Masters, Terry.** n.d.. “How to obtain a copy of a certificate of incorporation.” <https://legalbeagle.com/12717162-how-to-obtain-a-copy-of-a-certificate-of-incorporation.html> (accessed 2020-10-05).

**Perianan, Veerapan.** 2020. “Why market returns may be lower and global diversification more important in the future.” <https://www.schwab.com/resource-center/insights/content/why-market-returns-may-be-lower-in-the-future#:~:text=The%20main%20factors%20behind%20the,investing%20in%20a%20diversified%20portfolio>. (accessed 2020-10-05).

**Private Capital Research Institute.** 2017. “The consequences of private equity on employment and management.” <https://www.hbs.edu/private-capital/Documents/ILPA-PCRI-Roundtable-2017.pdf> (accessed 2020-10-05).

**Private Equity Growth Capital Council.** 2013. “2013 Annual Report.” <https://www.slideshare.net/pegccouncil/pegcc-annual-report825x825012314a2> (accessed 2020-10-05).

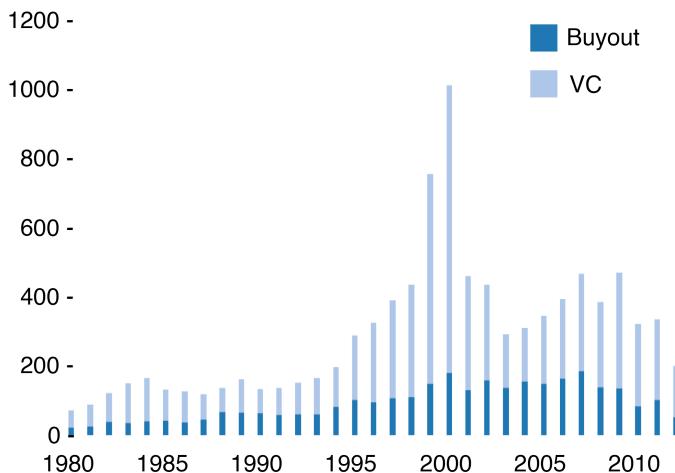
**Rossi, Andrea.** 2019. “Decreasing returns or reversion to the mean? The case of private equity fund growth.” SSRN. <https://ssrn.com/abstract=3511348> (accessed 2020-10-05).

## Appendix

### Appendix A

#### Summary Information of the PCRI Private Capital Database

The following tables and figures, from Jeng and Lerner (2015), provide a summary overview of the data collected on private capital firms, funds, and portfolio companies. In particular, the PCRI focuses on buyouts, growth equity, and venture capital investing.



**Figure 10.3:** Number of private capital firms by year founded

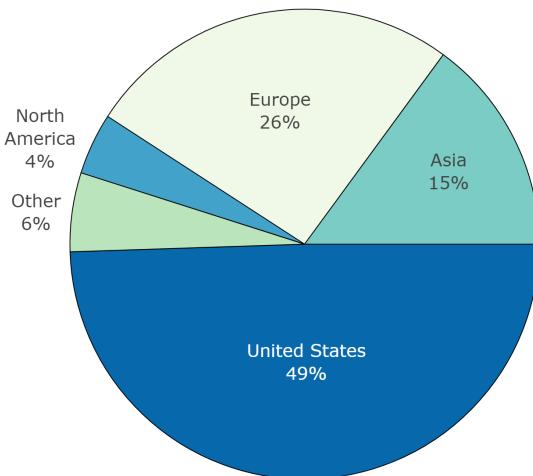
**Table 10.3:** Private capital firms by location of company headquarters and year founded

Regions	1980-1989	1990-1999	2000-2009	2010-2015	Total
Africa	0.3%	1.3%	1.7%	1.0%	1.3%
Asia	9.0%	10.4%	18.2%	27.6%	15.3%
Eurasia	0.0%	0.6%	1.0%	2.3%	0.9%
Europe	22.0%	24.5%	27.1%	22.0%	25.2%
Middle East	1.0%	2.4%	2.7%	1.9%	2.3%
Multi Geography	0.1%	0.0%	0.2%	0.2%	0.1%
North America	4.1%	5.6%	5.5%	4.6%	5.2%
Oceania	1.2%	2.2%	1.9%	0.9%	1.8%
South America	0.5%	1.1%	1.7%	1.7%	1.4%
United States	62.9%	51.8%	40.1%	37.9%	46.5%

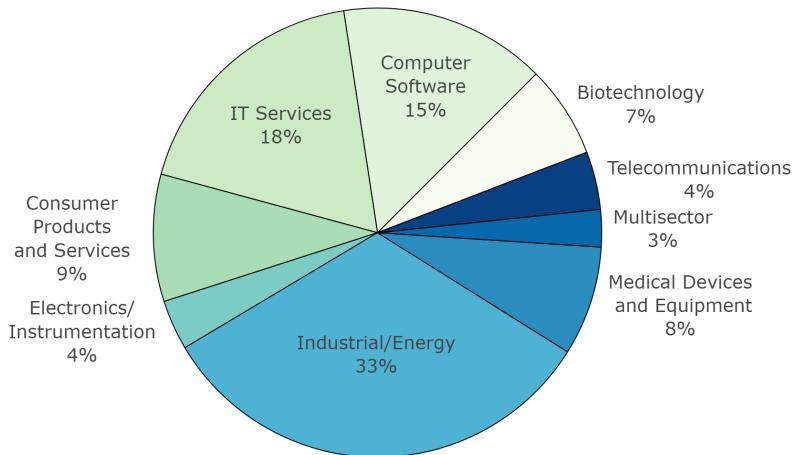
Notes: As of 2015. Source: Jeng and Lerner (2015).

**Table 10.4:** Private capital firms by location of company headquarters split by year founded

Fund Type	1980-1989	1990-1999	2000-2009	2010-2015	Total
Buyout	19.0%	26.0%	27.3%	26.4%	26.1%
Growth Equity	0.9%	0.7%	1.9%	8.5%	2.3%
Other	21.5%	9.4%	14.2%	11.9%	13.4%
Second	0.1%	0.4%	0.3%	0.2%	0.3%
VC	58.6%	63.6%	56.3%	53.0%	57.9%



**Figure 10.4:** Funds by region (N = 25,238)



**Figure 10.5:** Funds by industry (N = 12,333)

**Table 10.5:** Portfolio companies by region and year founded

Regions	1980-1989	1990-1999	2000-2009	2010-2015
Africa	0.6%	0.6%	0.5%	0.1%
Asia	11.5%	13.5%	14.9%	10.3%
Eurasia	0.1%	0.5%	0.5%	2.2%
Europe	31.2%	25.8%	30.3%	32.5%
Middle East	0.9%	1.4%	1.8%	1.9%
Multi Geography	0.0%	0.0%	0.0%	0.0%
North America	7.3%	4.9%	3.7%	2.8%
Oceania	2.5%	1.5%	1.2%	0.6%
South America	0.7%	0.8%	0.7%	0.6%
United States	45.2%	51.1%	46.4%	49.0%

*Notes:* As of 2015. Source: Jeng and Lerner (2015).

## **Appendix B and Appendix C**

*A sample data sponsor agreement and sample list of certificate of incorporation variables can be found in the Online Appendix at [admindatahandbook.mit.edu/book/v1.0/pcri.html#pcri-appendix](http://admindatahandbook.mit.edu/book/v1.0/pcri.html#pcri-appendix)*

