Making Private Data Accessible in an Opaque Industry:
The Experience of the Private Capital Research Institute

By Leslie Jeng[1] and Josh Lerner[2]

Introduction: The Rationale

The level of interest in alternative investments, and private capital in particular (by which we mean both venture capital and private equity), has been intense over the past decade. This interest has stemmed from both investors' needs 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, have been weak for much of this period, while low (and rising) interest rates suggest limited future returns for bonds. Many other classes of alternative investments, such as hedge funds, 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 equity sector grew from $870 billion a decade ago to $3.8 trillion in mid-2014.[3] This growth has in turn raised questions about the consequences of these investments for companies, workers, and economic growth more generally.

Despite the fact that the global economy and individual investors are increasingly dependent on private capital, much remains poorly understood about these investments. One key aspect of private capital is that it is indeed private. Traditionally, the general partners of these funds have not disclosed much information to the U.S. Securities and Exchange Commission or other regulators, or even to their own investors (limited partners, or LPs). A shortage of highly reliable industry data and to an unappealing setting where industry advocates make sweeping claims about the benefits and critics make broad charges on very shaky empirical foundations.

[1] *Director of Research*, Private Capital Research Institute, 114 Western Avenue, Boston, MA 02134 (email: ljeng@pcrinstitute.org)
[2] *Director*, Private Capital Research Institute and Harvard Business School, Arthur Rock Center for Entrepreneurship, Room 314 Boston, Massachusetts 02163 (email: jlerner@hbs.edu)
[3] 2015 Preqin Global Private Equity Report; https://www.preqin.com/docs/reports/.

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 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 been licensed to many schools, and Burgiss, which has made its data available to the Private Equity Research Consortium based in the University of North Carolina) or else data provided to researchers directly by limited and general partners on a "one-off" basis. It is typically difficult to compare any but the basic facts about the various commercial databases [for recent examples, see Brown et al. (2015), Maats et al. (2011), Kaplan and Lerner (2015)]. As a result, there are contradictory papers on a number of important topics, such as the risk-adjusted performance of private equity and the extent of persistence of performance of funds, [on the latter topic, see Braun et al (2015), Harris et al. (2013), Korteweg and Sorensen (2015)] where the results appear to be a function of the differences between databases, but it is difficult to replicate or refute the key findings. These issues are akin to the more general issues with private data raised by the American Economic Association's Committee on Private Data.

The PCRI

The Private Capital Research Institute ("PCRI") was founded with the objective to solve these data issues and to provide a greater fact-based understanding of private capital's global impact. Since 2010, an important part of the PCRI effort has been the building of a comprehensive and authoritative private capital database to serve as the foundation for independent analysis of the economic impact of private capital and the performance of funds and individual transactions. A large part of the process has been working out licensing agreements with commercial data vendors and the private capital firms themselves.

It is virtually impossible to pinpoint the exact size of the private equity industry or to verify completeness of any dataset. However, we believe that the PCRI universe is one of the most comprehensive and complete databases on private capital funds and transactions. 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 PCRI's own research.

The PCRI collects information on private capital firms, funds, portfolio companies, transaction data, and investment exits. In particular, we focus on buyouts, growth equity, and venture capital investing. In the future, we would like to include information on angel investments and sovereign wealth funds.

In our outreach to private capital firms, we have relied primarily on the relationships of our team members with the individual private capital firms. Thus, we have had to approach each private capital firm one at a time, an extremely time-consuming endeavor. At this point, we have approximately fifty groups that have provided data or are in the process of contracting to do so. Going forward, we hope to make this process more efficient by working with some organizations which are already actively collecting information from general partners. These include a large custodian bank with whom we have signed a data-sharing agreement and which has collected similar data from many private capital organizations, major pension funds (one has already provided such information), and national venture capital associations.

Gathering information from the private capital firms has its own limitations. Even if every active group chose to participate, there would still be a number of groups that have gone out of business. In addition, as we began collecting data from the 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" their data. Thus, we realized the importance of quickly building a large foundation for our database. As a result, we are complementing the data that we gather from the private capital firms with data from commercial sources, even if it is not always of the same quality as that of that provided directly by the general partners. The commercial sources include the Emerging Markets Private Equity Association (EMPEA), NYPPEX FUNDSIQ ("NYPPEX"), Thomson Reuters, and unquote. The features of the PCRI database are summarized in the PCRI's most recent data description article [Jeng and Lerner (2015)], which is available on the PCRI's website.

Data Sponsor Licensing Agreements

The largest challenge we faced in our data collection effort was creating a standardized licensing agreement for all our data sources, in particular the private capital firms. Working closely with our lawyers at Debevoise & Plimpton and a few key private capital firms, we were able to develop a standardized licensing agreement that not only allowed the PCRI to obtain, use, and make available to academics highly confidential data, but also alleviated the major concerns, i.e., confidentiality and security, of the private capital firms. Below, we highlight some of the key features of our Private Equity Sponsor Data Agreement.

The data licensing agreement grants the PCRI a royalty-free, non-transferable license. The PCRI is permitted to receive, store, reproduce, and combine the data. In addition, the PCRI is free to distribute the data to academic researchers from reputable institutions (accredited colleges and universities, not-for-profit research organizations, and research groups of government organizations). As part of the process, prior to obtaining access to the data, academic researchers are required to submit their research plans for approval. Furthermore, since the PCRI is designed to be a project run by academics and for academics, the PCRI database would be used exclusively for academic research, rather than for any commercial purpose.

Next, to protect the confidentiality of the data, academic researchers are required to sign a Data Access Agreement with the PCRI. As such, the PCRI has designed a protocol that simultaneously allows academics to undertake high quality research, while protecting the confidentially of the data being provided by the data sponsors. The PCRI data is hosted at the National Opinion Research Center (NORC) at the University of Chicago, which currently hosts highly sensitive federal and private sector data. The PCRI databases are anonymized; only PCRI research staff have access to identified data. PCRI researchers are able to access PCRI databases only through special computers with disabled USB ports and no screen capture. Furthermore, the PCRI employs a methodology whereby academics can undertake detailed cross-tabulated and regression analyses but not download or view individual data entries. Essentially the academics can upload queries and download results without "touching" the individual data entries. Finally, software program log files would provide a paper trail of activity and would be monitored.

One of the primary objectives of the PCRI is to promote *unbiased*, academic research. Thus, another key feature of the licensing agreement is that the data sponsors would not be able to limit the areas of academic research. However, the data sponsors would be able to obtain a preview of working papers and would be invited to attend PCRI conferences featuring our research. Lastly, the data sponsors would be given the option to be acknowledged for their contribution to the PCRI research effort.

Not surprisingly, liability issues were a major source of discussion. Given the limited resources of the PCRI, the licensing agreement puts a cap on PCRI's liability at $1000, which only applies in cases not resulting from fraud, misconduct, or misrepresentation on the part of PCRI, in which case there would be no liability cap. Also, the PCRI agrees not to bring any claims against any of the data sponsors.

A final feature is that the licensing agreement allows either party to terminate the agreement. The PCRI would be required to purge the sponsor's data from future versions of its databases within a reasonable time frame. However, any previously approved researcher would still continue to have access to the consolidated databases. Also, this agreement could be reassigned to another non-profit provided the PCRI gives the data sponsors prior written notice.

Final Thoughts

Arguably, an increasing share of economic activity today is taking place in settings which elude traditional Federal data collection mechanisms, or else where these data sources cannot capture the richest of the activity at work. Against this backdrop, economists are increasingly turning to private data. This brief essay has highlighted the experience of the PCRI, which been in the process of creating a database to facilitate access to private equity information by academics in a way that addresses the many of the major concerns regarding private data. While this effort is definitely a work in progress, hopefully the experience can provide guidance to researchers who want to address similar issues in other fields.

Citations

Braun, Reiner, Tim Jenkinson, and Ingo Stoff, "How Persistent is Private Equity Performance? Evidence from Deal-Level Data," *Working Paper*, December 2015.

Brown, Gregory W., Robert S. Harris, Tim Jenkinson, Steven N. Kaplan, and David Robinson. "What do Different Commercial Data Sets tell us about Private Equity Performance?." *Working Paper*, December 2015.

Harris, Robert S., Tim Jenkinson, Steven N. Kaplan, and Ruediger Stucke, "Has Persistence Persisted in Private Equity: Evidence from Buyout and Venture Capital Funds," *Working Paper,* February 2014.

Jeng, Leslie, and Josh Lerner, "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," *Private Capital Research Institute Working Paper*, September 2015

Kaplan Steven N. and Josh Lerner, "Venture Capital Data: Opportunities and Challenges," *Working Paper*, June 2015.

Korteweg, Arthur, and Morten Sorensen, "Skill and Luck in Private Equity Performance," *Rock Center for Corporate Governance, Stanford University Working Paper No. 179,* October 2015

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