Venture Capital Data: Opportunities and Challenges

by

Steven N. Kaplan* and Josh Lerner**

June 2015

Abstract

This paper describes the available data and research on venture capital investments and performance. We comment on the challenges inherent in those data and research as well as possible opportunities to do better.

*University of Chicago Booth School of Business and NBER, and **Harvard Business School and NBER. This paper was prepared for the NBER-CRIW Conference on Measuring Entrepreneurial Businesses: Current Knowledge and Challenges. We thank Rick Townsend for very helpful comments. Kaplan and Lerner have consulted to venture capital general partners and limited partners and have invested in venture capital funds. Lerner thanks Harvard Business School's Division of Research for support. Kaplan thanks the Fama-Miller Center at Chicago Booth. Address correspondence to Steve Kaplan, University of Chicago Booth School of Business, skaplan@uchicago.edu and Josh Lerner, Harvard Business School, josh@hbs.edu. All errors and omissions are our own.

I. Introduction

Venture capital is a relatively small financial institution. In the five years from 2009 to 2013, the NVCA – National Venture Capital Association (2014) –reports that an average of fewer than 1,200 firms received venture capital for the first time annually in the U.S. This is a very small fraction – roughly one in 500 or 0.2% – of the 600,000 firms (with employees) that are started each year (U.S. SBA (2012)). Over the same five-year period, U.S. venture capital partnerships received an average of less than $18 billion in new capital commitments from investors each year. And these figures are for the U.S., by far the largest market for venture capital in the world.

So why then does venture capital receive a large amount of theoretical, empirical, policy and media interest? From a theoretical perspective, venture capital is particularly interesting because it encompasses the extremes of many corporate finance challenges: uncertainty, information asymmetry and asset intangibility. At the same time, from an empirical and policy perspective, venture capital has had a disproportionate impact. Kortum and Lerner (2000) find that venture capital is three to four times more powerful than corporate R&D as a spur to innovation. Kaplan and Lerner (2009) find that roughly 50% of the "entrepreneurial" IPOs in recent years are venture-backed despite the fact that only 0.2% of all firms receive venture funding.

But despite the extent of interest in venture capital, substantial misunderstandings about this intermediary persist. This is particularly true in policy circles, which have seen the launch of ill-considered efforts to promote venture activity in many geographies (see Lerner (2009)), and media discussions. This reflects the facts that venture capital is a form of private equity, and that aspect of private equity is that it is indeed private. Unlike mutual funds, venture capitalists are

typically exempt from the Investment Company Act of Act of 1940, and typically do not disclose much information to the United States Securities and Exchange Commission or other regulators. This has led to a shortage of 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.

This lack of a comprehensive dataset has also posed challenges to academic research. One of the most important ways that academic research in the social sciences proceeds is by researchers replicating and exploring the limitations of earlier studies. Instead, in venture capital, because the studies often rely on proprietary datasets that are not shared more generally, studies are difficult to replicate or refute. Another unappealing consequence is that dubious or misleading studies can linger for many years without rebuttal.

Sadly, this problem may be getting worse, rather than better. The past decade has seen of the rise of "individualized entrepreneurial finance": angels, groups of angels, crowdfunding platforms, and the like. While venture capital remains concentrated in a few metropolitan areas, mostly in the United States, the amount of angel investments appear to be increasing in many nations (Wilson and Silva, 2013). Active involvement in the investment and close social ties between angels and entrepreneurs may help to overcome the lack of minority shareholder and legal protections that are important for the development of more institutionalized capital markets. These investors are typically very reluctant to share information about their activities, both for strategic reasons as well as due to a reluctance of personal exposure.

In this paper, we describe the available data and research on venture capital investments and performance. As we do so, we comment on the challenges inherent in those data and research as well as possible opportunities to do better. We begin by describing the data and

research on investments by venture capital funds in portfolio companies. We follow that by describing the data and research on investments (by institutional investors and wealthy individuals) in the venture capital funds.

II. Investment Data and Research

A. Longstanding databases

Much of the early research into venture capital relied on information available in IPO prospectuses and S-1 registration statements. For the subset of venture-backed firms that eventually go public, voluminous information is available. Investments in firms that do not go public are more difficult to uncover, since these investments are usually not publicized. Unfortunately, because only a relatively modest fraction of venture-backed companies go public, researchers must dig deeper.

There are two longstanding databases that characterize the investments of venture capital funds into portfolio companies, regardless of the investment outcome. VentureXpert (VX), a unit of Thomson Reuters, began collecting data in 1961. Venture Source (VS), a unit of Dow Jones, began collecting data in 1994.

The basic story here is that there are large inconsistencies in both databases and a general problem of incompleteness. Furthermore, qualitatively, both show deterioration in data quality over the past decade. That said, VX has more complete coverage of investments while VS measures outcomes more accurately.

Maats et al. (2011) focus on investments by 40 VC funds with vintage years 1993 to 2003. They obtain data about the investments and exits from outside sources and, for two VC funds, from a major limited partner. They then compare the actual data to the data in VS and

VX. This follows and expands on an earlier iteration of this research design by Kaplan, Sensoy and Stromberg (2002).

First, they find that VX has more complete coverage of the investments in the funds. Second, they find that both VX and VS understate the fraction of companies that are defunct, with VX having more incorrect. In fact, VX reports less than 10% of investments as defunct when, in fact, more than 20% are defunct. Third, VX exit / status coverage has dropped dramatically in recent years suggesting a lack of investment in collecting new data.

Maats et al. (2011) then do a firm level comparison for 449 venture-financed firms that are in both VX and VS. Figure 1 shows that VX appears to have somewhat better coverage. VX has 40% more financing rounds. While VX and VS have post-money valuations for roughly the same number of firms, VX has roughly 10% more post-money valuations for financing rounds. Figure 2 provides a round-level comparison for 173 firms that are in both VS and VX. Again, VX has roughly 40% more rounds and roughly 10% more post-money valuations.

Maats et al. (2011) also compare the accuracy of the two databases for two specific funds where they obtain data from a limited partner investor in the two funds. VX does a much better job of including firms in the database that the funds actually invested in. The funds that VS excludes tend to be predominantly funds that failed, leading to a likely upward performance bias in VS.

The earlier comparison by Kaplan et al. (2002) had suggested some valuation advantages for VS. They compared the actual valuations in 143 financings to their reported values in VS and VX (prior to 2000). They found that VS included almost twice as many valuations as VX and the average absolute error of those valuations was only 60% of those in VX.

There is an important additional caveat in measuring valuations. They do not reflect the impact of transaction terms, instead simply reporting the "pre" or "most-money" valuation, which are defined as the product of the nominal price per share paid in transaction times the number of shares outstanding (typically, assuming all shares are converted into common stock) before and after the transaction. In other words, these calculations ignore the implicit call and put options associated with these securities. See Kaplan and Stromberg (2003) for a catalog of these features.

Liquidation preferences, in particular, can have a large impact on values. Metrick and Yasuda (2010) provide examples where valuations change by 75% when deal terms are properly analyzed. To correctly analyze valuations across different investments, it is necessary to have access to the actual deal terms. This requires access to the underlying deal documents which are not easy to obtain.

There is one other difficulty in both databases – firm name changes. Both databases only index on the current (or latest) portfolio company name. The recording of former names is desultory at best. Of course, this makes matching to historical records challenging.

Finally, the results in Maats et al. (2011)—as well as anecdotal accounts—suggest that there has been substantial subsequent deterioration in the quality of both databases. In particular, the initial focus by VS on valuations seems to have been largely abandoned. In part, this may reflect the challenges associated with the reliance on commercial data providers, who may decide on an investment in ensuring data quality that while profit-maximizing, is less than an academic financial economist would prefer.

B. More recent alternatives

There are a number of recent alternatives to VX and VS. Several databases that focus on tracking private equity (buyout) funds and transactions also include some VC funds and deals. These databases are typically based on disclosures from limited partners, filings with the SEC, and other public (but often difficult to access) sources. Examples include Capital IQ, Pitchbook and Preqin. VCExperts is a newer database that specializes in VC deals and is sourced from state and federal regulatory filings by private companies.

The SEC maintains Form D filings of private financings, but these provide only the amount of funding and not the names of investors.

There are some websites that track venture capital financings. Tech Crunch's, Crunchbase, is the best known. While many of these newer databases are promising, they have not gotten the kind of scrutiny that VS and VX have. Thus, their ability to support academic research is still to be fully determined.

C. The bottom line on portfolio company data

As mentioned above, the basic story on portfolio company data is not a great one. There are large inconsistencies in the two major existing databases, VX and VS, and a general problem of incompleteness. Furthermore, qualitatively, both show deterioration in data quality over the past decade. As we will discuss in the conclusion, there is an opportunity for a new provider—whether for-profit or non-profit – to significantly improve on these data.

It also seems possible that the fund performance data providers described in the next section, particularly Burgiss, Cambridge Associates and Preqin will be able to augment their fund data with data on individual portfolio companies.

III. Performance data

There are currently three major providers of data on VC (and private equity) performance – Burgiss Private I, Cambridge Associates (CA) and Preqin. Pitchbook is a fourth newer entrant with more of a focus on private equity performance. Until recently, there was a fifth, Thomson Venture Economics (TVE). For reasons likely related to poor quality data that we describe below, TVE decided to discontinue its database and, instead, make CA available on TVE's platform.

As with the data on VC firm investments in portfolio companies, VC fund performance data are also potentially subject to biases:

- First, the data from any one provider may be incomplete. For instance, a number of leading venture capital funds have pressured pension funds not to post on-line or to report their performance to data providers such as Preqin. Some have gone as far as to drop institutions that cannot make such commitments as limited partners (Lerner, et al. (2011)). Given the highly skewed nature of performance in venture capital, even a handful of omissions can have a substantial impact on reported performance figures.
- Second, it is possible there is a backfill bias in that the databases report positive past returns for funds that are newly added to the database. Many first-time funds do not have any institutional investors, and may not be captured by commercial data providers unless they successfully raise a second fund.
- Third, to the extent that the databases rely on data directly reported by the GPs, it is possible that poorly performing funds stop reporting or never report at all.
- Fourth, to the extent that database providers rely on information from GPs—or the LPs report data from GPs without adjustment—the quality of the information can suffer from

deliberate distortions of the valuations. One example is the valuation of still privately companies in the venture capitalists' portfolios. Particularly with early stage companies, valuations assigned by venture firms to their own portfolio of investments are often based not on quantitative metrics (such as price-to-earnings or discounted cash flow) because the company may not have any prior earnings or reliable projections. Instead, the partners rely on complex, frequently subjective assessments of a venture's technology, expected market opportunity, and its management team's prowess. Less established groups, or those seeking to raise new funds in the near future, may be tempted to shade these valuations upwards. Similar concerns have been raised by stock distributions to LPs, a technique often employed by venture funds to unwind large positions in recently public (and often thinly traded) firms. While venture groups may value these distributions at the price prior to the distribution, the sales that ensue after the distribution often mean that the realized price is substantially lower. Again, because many LPs do not adjust the GPs' data, these inflated valuations may find their way into databases.

- Finally, the commercial platforms use different data definitions that complicate cross-platform comparisons. For example, funds are generally grouped by vintage year – the year they began. However the different platforms define beginning differently. Burgiss groups funds by the year in which the year the fund first takes down money from investors. CA groups funds by the year the fund is legally formed. Preqin groups funds by the year the fund makes its first investment in a company. While these three definitions will often coincide, they do not always do so.

In addition, some funds not only make investments in venture capital / early stage companies, but also in growth stage companies and in buyouts. Indeed, it is frequently difficult

to define where early-stage investing ends and later-stage transactions begin. While traditional buyout groups such as TPG have increasingly taken part in the later rounds of social media companies, many venture funds have undertaken growth investments in traditional manufacturing firms in markets such as India and China. In some cases, one commercial platform will classify a multi-asset class investor as a VC fund while a different platform will classify the same investor as a buyout fund.

In the rest of this section, we describe the coverage of the major platforms and their advantages and disadvantages.

A. Coverage

Figure 3 presents data on fund coverage by four of the commercial platforms as of the first quarter of 2011 using data from Harris et al. (2014). TVE had the highest number of funds in the 1980s and the 1990s. In the 2000s, however, TVE declined to the lowest coverage with Preqin and CA moving to the highest number of funds represented. Though not illustrated in the graphs above, Burgiss had increased its representation for the most recent vintage years, 2006 to 2008, to roughly the same coverage as Preqin and CA. In its most recent release, the second quarter of 2014, Burgiss' coverage had increased markedly to 538 VC funds with vintage years from 2000 to 2008, up from the 423 funds in 2011.

Figure 4 presents total capital commitments represented in the commercial platforms as a proportion of total capital committed to VC, using the data in Harris et al. (2014) as of 2011 Q1. Total committed capital is taken from the annual totals provided by the Private Equity Analyst. Burgiss and Preqin have a higher proportion of total commitments from 2000 to 2008. Capital commitments for CA funds were not available for the study.

As with the number of funds, TVE had strong coverage in the 1980s and 1990s with over 100% of committed capital in the 1980s and almost 80% in the 1990s. In the 2000s, TVE dropped off. Preqin had performance data on funds with roughly 70% of committed capital; Burgiss performance data on funds with 60% of committed capital.

In its most recent release, 2014 Q2, Burgiss has coverage of 72% of committed capital for 2000 to 2008 vintages. Its coverage reaches 89% of committed capital for vintages from 2006 to 2013.

B. Commercial platforms

1. Burgiss Overview

The data are derived from LPs for whom Burgiss' systems provide record-keeping and performance monitoring services. The Burgiss data are sourced exclusively from a diverse array of LPs for whom Burgiss provides record-keeping and performance monitoring services. This includes a complete transactional and valuation history between the LPs and fund investments. As a result, Burgiss is able to record exact cash outflows / investments made by LPs to GPs and distributions from GPs back to LPs. Burgiss also cross-checks across investors in the same fund. This feature results in investment histories that are free from any reporting bias. For instance, Burgiss has the complete investment history of LPs who allow Burgiss to aggregate their data. In addition, the Burgiss data are current because Burgiss' LPs receive their data currently from GPs and Burgiss uses the quarterly reporting used by most investors.

Harris et al. (2014) report that their data come from over 200 institutional investors representing over $1 trillion in committed capital. Two-third of the LPs have PE commitments

of over $100 million. Of these 60% are public/private pensions and 20% are endowments or foundations.

Over time, the number of funds in the Burgiss database has increased as Burgiss has gained permission to access the investment performance of an increasing number of LPs. The one potential bias in the Burgiss data is that the LPs who allow access are selected. In particular, it is possible that the LPs who allow access, as a group, have tended to invest in above average funds and, therefore, exclude some below average funds. For this bias to be in the data, however: (1) there would have to be a group of institutional investors who invested in the worst VC funds, had poor performance, and do not use Burgiss to measure their fund performance; (2) no other institutional investors who do use Burgiss invested in those same VC funds, so the poorly performing PE funds do not show up in the data set. Given the size of the Burgiss data set, this seems unlikely. Furthermore, the fact that Burgiss covers almost 90% of the total capital committed to venture capital in post-2005 vintages suggests that this bias, even if it were to exist, is likely to be small for those vintages.

2. Preqin

Preqin's performance data are sourced primarily from public filings by pension funds, from FOIA requests to public pension funds, and voluntarily from GPs (about 60% of performance data) and LPs.

Preqin (and Pitchbook) are the only major data sources that identify GPs by fund name This means that the Preqin data are transparent and can be verified / corrected. The authors know GPs who have voluntarily contacted Preqin to correct erroneous data for their funds.

At the same time, Preqin has at least three potential biases. First, Preqin may miss some high performing funds that do not have public pension fund investors or have reporting restrictions. Notably, Preqin does not have performance data for a number of funds raised by very high performing VC's like Sequoia and Accel.

Second, because Preqin relies on voluntary reporting, Preqin often has somewhat stale data because of tardy responses.

Third, Preqin reports performance for a number of funds for which it does not have the granular cash flow data. In other words, some LPs simply report IRRs and multiples without reporting the cash flows that generated them.

3. Cambridge Associates (CA)

CA sources its data from voluntary disclosures by LPs and by GPs who have raised or are trying to raise capital. Because GPs typically do not try to raise a new fund if their performance is poor, CA may have a bias towards successful GPs. Also favoring this bias is CA's traditional orientation to providing services to endowments, who appear to have (historically at least) selected the most successfully venture capital LPs with which to invest (Lerner, et al. (2007)).

Whatever its other strengths and weaknesses, CA also is the least transparent of the commercial platforms.

4. Thomson Venture Economics (TVE)

TVE has traditionally sourced its data from both LPs and GPs in a manner similar to that used by CA. The major issue with TVE was that it appeared to stop updating performance on roughly 40% of the venture capital and private equity funds in the VE sample. Stucke (2011)

finds that of 488 buyout funds with 1980-2005 vintage years, 43% have constant NAVs and no cash flow activity for at least two years prior to December 2009. Phalippou and Gottschalg (2009) find that 300 of 852 sample funds are inactive for over 3 years, with most for 6+ years.

Stucke (2011) compares the performance of individual buyout funds in TVE to the actual performance of those funds provided by a large LP in those funds. He finds a substantial downward bias in the TVE data. While he does not study VC funds, it seems likely that the VC performance data had a similar downward bias.

Consistent with such a downward bias, Harris et al. (2014) find that VC fund performance in the TVE data is lower than that in Burgiss, CA and Preqin. Also strongly consistent with data problems, in March 2014, TVE decided to discontinue its benchmark data and, instead, contracted with CA to provide CA's private equity benchmarking data to TVE subscribers.[1]

C. Performance results

Harris et al. (2014) present VC (and PE) performance data from the major commercial databases as of the first quarter of 2011. Harris et al. (2015) present performance data updated to the second quarter of 2015. They find that venture capital (VC) funds outperformed public markets (as measured by the S&P 500) substantially until the vintages of the late 1990s. Coinciding to some extent with the tech bust, vintages from 1999 to 2003 underperformed public markets. Vintages from 2004 to 2010 have rebounded, performing better than or equal to public markets. That performance has likely further improved since then. This performance contrasts

[1] See, "Thomson Reuters partners with Cambridge Associates on benchmark data," March 2014, https://www.pehub.com/2014/03/thomson-reuters-partners-with-cambridge-associates-on-benchmark-data/.

with the view held by some that VC has been a poorly performing asset class as a whole in this century.

Harris et al. (2014) also find that Burgiss, Cambridge Associates and Preqin yield qualitatively and quantitatively similar performance results. Tables 1a and b reproduce these results from Harris et al. (2014). There is little reason to believe that the Burgiss and Preqin data sets, in particular, suffer from performance selection biases in the same direction. At the same time, consistent with Stucke (working paper 2011), they find that performance is lower in the Venture Economics data (particularly for buyout funds).

Kaplan and Sensoy (forthcoming) provide a broader summary of the performance of PE and VC funds. Other research is broadly consistent with the results in Harris et al. (2014 and 2015).

D. The bottom line on performance providers

Based on the research done to date, Burgiss is likely the best of the commercial data providers. The data it has are current and do not appear to be selected. Given the similar results in Preqin and CA, it is unlikely there is any appreciable bias across these databases. The fact that Burgiss now covers performance for almost 90% of the total capital committed to venture capital in post-2005 vintages suggests that the ability to do research on venture capital funds will continue to improve over time. This is particularly encouraging given that Burgiss makes its data available to researchers through proposals to the PERC (Private Equity Research Consortium). Kaplan serves on PERC's academic advisory board.

While Preqin (and Pitchbook) have potential selection biases, they are also powerful and valuable because they identify the performance of individual funds. This allows a better fix on the potential selection biases at work.

Thomson Venture Economics (TVE) should not be used. Its database has been discontinued. Results in past work using TVE should be with viewed with caution.

It is also worth noting that this is a dynamic field, with a number of new entrants. Examples include eFront and State Street Bank, which have gathered data as part of their work with general and limited partners, and analytics solutions providers such as Bison. While it is still early to evaluate many of these efforts, the promise of more and higher quality data augurs well for future research opportunities.

A "horse of a different color" is the Private Capital Research Institute, in which both of the authors are involved (Kaplan as an academic advisory board member and Lerner as director). This foundation-supported non-profit is in the process of developing an database exclusively for academic research, modelled after the architecture for compiling confidential information employed by the U.S. government. By restricting the data use to these applications, it is hoped that a broader swath of the industry will consent to the utilization of their data.

The heart of the PCRI effort is high quality data about private capital investments. While commercial data vendors typically piece this together from a variety of sources, including security filings and disclosure statements by institutional investors, frequently the information is incomplete and inconsistent.

The vision of the PCRI is to focus very much on obtaining data from the private equity firms themselves. To date, over 40 of the 100 largest private equity firms world-wide have provided data to the PCRI, or are in the process of doing so. It might be plausibly wondered 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. The answers are several:

1) The constraints the PCRI places on the use of the data. In particular, the PCRI is designed to be a project run by academics and for academics. The information is used exclusively for academic research, rather than for any commercial purpose.
2) The research protocol simultaneously allows academics to undertake high-quality research while protecting the confidentiality of the data being provided by the private equity firms. In particular, following the model employed by the United States Bureau of the Census when making available information that it and the United States Internal Revenue Service collect, academics can undertake detailed cross-tabulated analyses but not download or view individual data entries. Essentially the academics would be able to upload queries and download results without "touching" the individual data entries.
3) A third reason for the success of the PCRI in generating participation in the private equity community has to do with the fact that the industry itself is under much greater 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 from the private equity firms has limitations. Even if every active group chose to participate, there would still be a number of groups that have gone out of business. As a result, the PCRI is complementing the data gathered from the private equity firms

with data from commercial sources. In addition, the PCRI is working with a commercial group that has developed an extremely efficient and cost-effective manner to collect the cash flow data of private equity funds from public regulatory and disclosure filings. As regulations push private equity groups to undertake more and more security filings, this will likely be an increasingly fruitful methodology. This relationship will allow us to gain more experience with the harvesting of such data. Thus, the use of commercial data sources allow the PCRFI to get a more holistic picture of the activity in the private capital industry, as well as to quantify any potential biases that may affect rigorous scientific analysis.

In addition to our own efforts to acquire data for the PCRI, the support of the institutional investor community has proved valuable. Because there are ambiguities about whether institutional investors can share data on existing funds, the PCRI initially did not ask them for data directly. Nonetheless, a number of institutional investors –including some formally on our practitioner board—have been very helpful in encouraging the private capital firms in which they have invested to share data with the PCRI.

IV. Conclusions

Venture capital is an increasingly important intermediary, able to transform capital into new firms and innovations in an apparently highly productive manner. This intermediary is attracting increasing interest by policymakers and investors, but the availability of data as well as the consistency of the academic findings using these data are still lacking.

This paper attempted to take a careful look at the availability of information about this intermediary. Several conclusions emerge from our review of the major data sources for venture capital investments and funds:

- Reflecting the relative lack of disclosure and the substantial information asymmetries surrounding venture capital, it is difficult to paint in definitive terms the level of investment activity and fund performance.
- Existing databases differ in methodologies, and analyses frequently produce discrepancies and varying conclusions. These problems are particularly prevalent when it comes to transaction-level data.
- That being said, the venture data space has seen substantial entry, particularly in regard to performance measurement. As a result, the quality of information available has increased in recent years and can be expected to continue to do so going forward.

References

Gottschalg, O., and L. Phalippou, 2009, "The Performance of Private Equity Funds," Review of Financial Studies 22, 1747-1776.

Harris, R., T. Jenkinson and S. Kaplan, 2014, "Private Equity Performance: What Do We Know?" Journal of Finance, 1851-1882.

Harris, R., T. Jenkinson and S. Kaplan, 2015, "How Do Private Equity Investments Perform Compared to Public Equity?" Journal of Investment Management, forthcoming.

Kaplan, S. and J. Lerner, 2010, "It Ain't Broke: The Past, Present, and Future of Venture Capital," Journal of Applied Corporate Finance 22 (2), 36–46.

Kaplan, S. and B. Sensoy, forthcoming, "Private Equity Performance: A Survey," Annual Review of Financial Economics.

Kaplan, S., B. Sensoy and P. Strömberg, 2002, "How Well Do Venture Capital Databases Reflect Actual Investments?" Unpublished working paper, University of Chicago.

Kaplan, S. and P. Strömberg, 2003, "Financial Contracting Theory Meets the Real World: Evidence from Venture Capital Contracts," Review of Economic Studies 70, 281-315.

Kortum, S., and J. Lerner, 2000, "Assessing the Contribution of Venture Capital to Innovation," Rand Journal of Economics 31, 674-92.

Lerner, J., 2009, The Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed—and What to Do About It. Princeton: Princeton University Press.

Lerner, J., A. Leamon, and F. Hardymon, 2011, Private Equity, Venture Capital, and the Financing of Entrepreneurship: The Power of Active Investing, New York: John Wiley & Sons.

Maats, F., A. Metrick, A. Yasuda, B. Hinkes and S. Vershovski, 2011, "On the Consistency and Reliability of Venture Capital Databases," Unpublished working paper, University of California at Davis.

Metrick, A. and Yasuda, A., 2010, Venture Capital and the Finance of Innovation, New York: John Wiley & Sons.

National Venture Capital Association, 2014, Venture Capital Yearbook, Washington: National Venture Capital Association.

Stucke, R., 2011, “Updating History,” Unpublished working paper, Oxford University (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1967636).

U.S. Small Business Administration, 2012, The Small Business Economy, Washington: U.S. Government Printing Office.

Karen Wilson and Filipe Silva, 2014, “Policies for Seed and Early Stage Finance,” Science, Technology and Innovation Directorate Policy Paper #9, OECD.

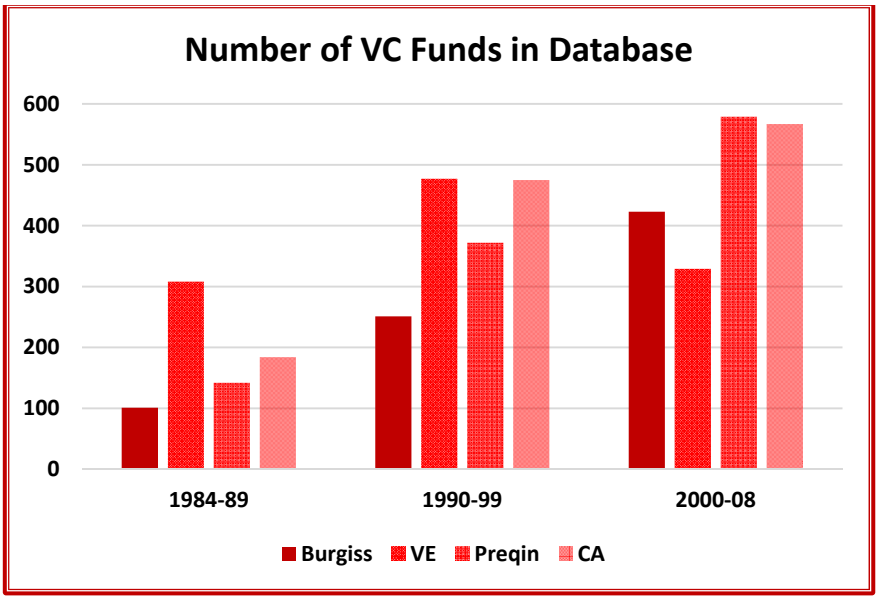
Figure 1: Firm-level comparison. Source: Maats et al. (2011)

	VS	VX
Total firms compared	449	449
Number with post-money valuations	285	286
Rounds with post-money valuations	693	764
Number with same number of valued rounds in VS and VX	85	85
Count of investors	3485	3405
Number with same number of investors in VS and VX	107	107
Average investment per company ($M)	43.9	52.0
Number with same* investment amount in VS and VX	113	113
Count of rounds	1507	2145
Number with same round count	134	134
* Within $1 million.		

Figure 2: Round-level comparison. Source: Maats et al. (2011)

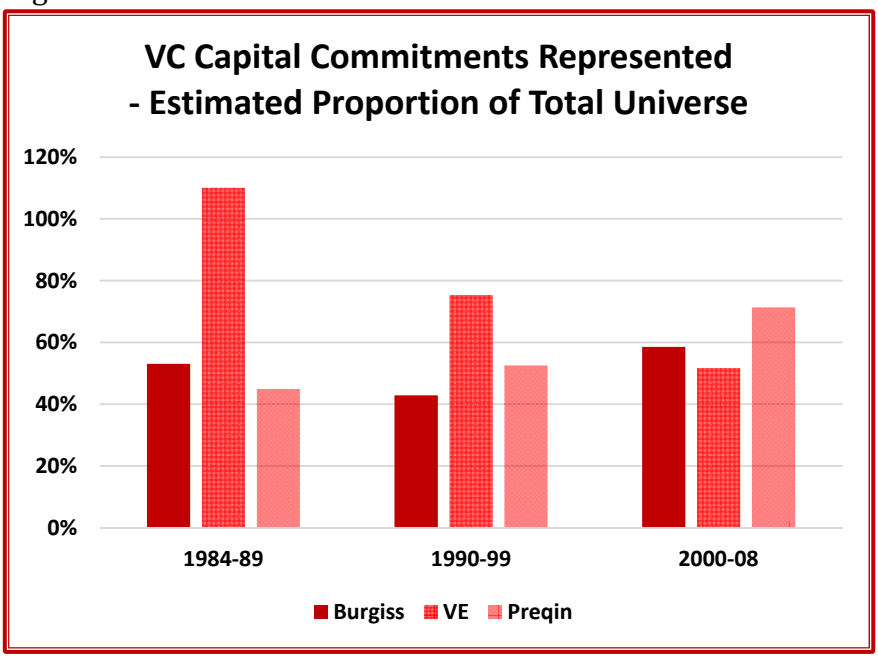
	VS	VX
Total firms compared	173	173
Count of rounds	592	857
Number with investment date	591	857
Rounds with same date in VS and VX	190	190
Number with post-money valuation	262	288
Number with same* post-money valuation in VS and VX	95	95
Number with same number of investors in VS and VX	203	207
Number of rounds with investment amount	570	857
Average investment per company ($M)	$7.8	$8.6
Number with same* investment amount in VS and VX	248	258
Number with same* date, amount, investor count, and valuation	13	13
* Within $1 million.		

Figure 3



Source: Harris, Jenkinson, Kaplan, (2014)

Figure 4



Source: Harris, Jenkinson, Kaplan, (2014)

Table 1a: Venture Capital Multiples from Commercial Databases

Panel B : Venture Capital Funds

Vintage	Weighted Average Multiples			Average Multiples				Median Multiples		
	Burgiss	Venture Economics	Preqin	Burgiss	Venture Economics	Preqin	Cambridge Associates	Burgiss	Venture Economics	Preqin
1984	1.73	1.59	2.74	1.78	1.57	2.35	1.70	1.71	1.37	1.99
1985	1.93	2.08	2.24	1.96	2.02	2.59	2.61	1.81	1.84	2.15
1986	1.82	2.79	2.19	1.83	1.72	1.85	1.83	1.93	1.54	1.67
1987	2.77	2.39	2.54	2.70	2.02	2.51	2.56	2.35	1.65	2.22
1988	2.88	2.59	2.98	2.45	2.04	2.58	2.15	2.55	1.86	2.57
1989	3.09	2.42	2.80	2.92	2.13	2.55	2.44	2.41	1.89	2.27
1990	3.30	2.82	2.81	2.96	2.32	2.50	2.74	2.48	1.81	2.22
1991	2.92	3.26	2.95	3.11	2.47	4.41	3.07	2.70	2.23	2.75
1992	2.72	3.45	2.84	2.69	3.46	3.26	3.54	2.07	2.18	2.07
1993	6.34	3.31	3.38	6.65	2.92	3.86	3.49	3.28	1.67	2.58
1994	6.58	4.65	7.19	5.27	3.32	3.98	3.83	3.05	2.23	2.08
1995	3.55	4.16	6.33	3.64	3.83	5.16	5.20	2.50	2.43	1.89
1996	6.33	4.78	2.81	5.92	4.68	3.49	4.51	2.06	2.14	1.87
1997	3.28	2.38	2.83	3.03	2.49	2.69	2.52	1.87	1.63	1.52
1998	1.60	1.74	1.77	1.55	1.68	1.67	1.51	0.93	1.10	1.08
1999	0.94	0.74	0.87	0.81	0.84	0.93	0.93	0.73	0.72	0.81
2000	0.97	1.06	1.04	0.91	0.94	1.14	0.90	0.88	0.88	0.99
2001	1.01	1.21	1.19	0.97	1.11	1.08	1.09	0.87	1.00	1.04
2002	1.07	0.96	1.12	1.01	0.98	0.98	1.06	0.99	0.93	0.96
2003	1.11	1.16	1.07	0.99	1.13	1.01	1.08	1.00	1.06	1.02
2004	1.07	1.08	1.19	1.01	1.04	1.22	1.36	0.97	0.99	0.98
2005	1.31	1.14	1.11	1.37	1.13	1.04	1.06	1.02	0.98	1.04
2006	1.04	1.01	1.05	1.01	0.93	1.02	1.03	0.95	0.96	0.98
2007	1.09	1.01	1.06	1.06	1.05	1.12	1.22	1.06	0.96	1.01
2008	0.97	0.94	1.00	0.99	0.93	1	1.06	0.98	0.92	0.95
Average	**2.46**	**2.19**	**2.36**	**2.34**	**1.95**	**2.24**	**2.18**	**1.73**	**1.48**	**1.63**
Average 2000s	1.07	1.06	1.09	1.03	1.03	1.07	1.10	0.97	0.96	1.00
Average 1990s	3.76	3.13	3.38	3.56	2.80	3.20	3.13	2.17	1.81	1.89
Average 1980s	2.37	2.31	2.58	2.27	1.92	2.41	2.22	2.13	1.69	2.15

Source: Harris, Jenkinson, Kaplan, (2014)

Table 1b: Venture Capital IRRs from Commercial Databases

Panel B : Venture Capital Funds

Vintage	Weighted Average IRR				Average IRR				Median IRR			
	Burgiss	Venture Economics	Preqin	Cambridge Associates	Burgiss	Venture Economics	Preqin	Cambridge Associates	Burgiss	Venture Economics	Preqin	Cambridge Associates
1984	7.9	6.0	16.7	8.1	8.2	5.0	14.7	7.7	6.9	3.6	12.8	6.3
1985	7.1	9.2	14.6	12.4	5.5	8.2	10.4	11.6	8.7	8.6	13.5	12.7
1986	9.4	11.9	14.3	13.2	9.0	7.3	13.6	8.8	9.3	6.3	8.9	9.4
1987	20.2	12.9	12.4	17.9	15.8	7.6	23.7	14.5	16.7	7.2	14.9	15.7
1988	24.4	19.8	28.6	18.6	17.9	12.3	21.5	14.3	21.6	9.5	23.1	11.9
1989	25.7	16.4	25.8	18.1	20.5	13.0	16.7	17.1	15.3	10.9	14.7	13.3
1990	29.5	24.8	20.3	31.4	25.3	18.5	55.7	24.3	21.7	14.3	19.3	21.9
1991	28.5	28.9	36.5	23.5	28.1	18.7	26.7	23.0	24.4	17.8	28.7	18.6
1992	24.8	31.5	25.1	24.9	21.0	27.8	33.3	28.7	14.2	13.4	21.5	21.0
1993	51.9	28.5	44.2	37.6	47.1	21.7	29.9	29.5	40.9	12.4	36.5	18.8
1994	41.4	42.9	47.0	47.2	41.7	26.9	55.9	34.6	31.8	23.7	27.1	26.5
1995	46.4	56.6	59.6	65.0	49.2	44.4	35.8	56.8	28.9	19.3	20.0	42.9
1996	76.7	61.2	22.7	71.4	64.5	67.0	48.6	61.1	25.2	28.2	14.8	37.1
1997	76.1	44.1	58.1	69.3	65.9	48.5	26.1	52.8	26.3	19.9	22.8	8.9
1998	15.5	24.4	19.8	13.2	16.3	25.8	-4.2	18.3	-1.2	2.0	4.9	0.4
1999	-4.5	-6.5	-4.4	-3.6	-7.4	-4.6	-0.2	-3.6	-5.6	-5.9	-4.9	-4.4
2000	-1.3	0.4	-0.4	-1.5	-2.7	-1.5	-1.0	-3.7	-2.1	-2.0	-0.5	-3.3
2001	-0.7	4.6	4.0	1.0	-1.7	2.0	-2.8	-1.5	-2.4	0.0	1.2	-0.7
2002	0.6	-0.8	2.1	-0.1	-1.1	-0.7	-1.5	0.7	-0.2	-1.2	-0.6	0.4
2003	0.9	3.6	1.7	2.8	-2.1	3.1	1.5	0.6	0.1	1.3	0.1	0.5
2004	0.3	1.9	2.3	2.6	-1.5	0.7	-0.5	2.0	-1.0	-0.3	-0.2	0.9
2005	3.3	2.9	2.0	1.0	2.2	2.5	-1.6	-1.1	0.5	-0.4	0.8	1.0
2006	0.6	0.0	1.1	2.9	-1.3	-3.7	3.8	-0.2	-2.4	-1.9	-1.4	0.3
2007	3.2	-0.9	3.9	4.1	1.7	1.3	-1.3	7.1	2.6	-1.7	1.5	3.9
2008	-4.5	-5.2	-2.7	3.7	-2.8	-7.1		1.6	-1.6	-6.5	-3.9	1.1
Average	**19.3**	**16.8**	**18.2**	**19.4**	**16.8**	**13.8**	**16.9**	**16.2**	**11.1**	**7.1**	**11.0**	**10.6**
Average 2000s	0.3	0.7	1.6	1.8	-1.0	-0.4	-0.4	0.6	-0.7	-1.4	-0.3	0.5
Average 1990s	38.6	33.6	32.9	38.0	35.2	29.5	30.8	32.5	20.7	14.5	19.1	19.1
Average 1980s	15.8	12.7	18.7	14.7	12.8	8.9	16.8	12.3	13.1	7.7	14.7	11.5

Source: Harris, Jenkinson, Kaplan, (2014)