Towards a Centralized Venture Capital Data Source - The Key to Increased Minority Access to Venture Capital?

Maya X. Trujillo

Follow this and additional works at: https://repository.upenn.edu/wharton_research_scholars

Part of the Business Commons

Trujillo, Maya X., "Towards a Centralized Venture Capital Data Source - The Key to Increased Minority Access to Venture Capital?" (2022). Wharton Research Scholars. 230. https://repository.upenn.edu/wharton_research_scholars/230

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/wharton_research_scholars/230
For more information, please contact repository@pobox.upenn.edu.
Towards a Centralized Venture Capital Data Source - The Key to Increased Minority Access to Venture Capital?

Abstract
In 2016, Steven Kaplan and Josh Lerner wrote about how the shortage of reliable and comprehensive industry data has created challenges for academics who wish to replicate or build on previous venture capital (VC) studies (414). The author of this paper experienced a similar lack of data first-hand while attempting to validate findings from a 2008 study which found that minority focused VC funds tend to earn equivalent or higher returns on their investments than the returns of the VC industry (Bates and Bradford 2008). The data required was virtually nonexistent, which prompted a search for databases which collect VC data. This paper is a call to action for greater VC fund financial transparency. This paper’s original goal of validating a prior study should be replicable by any researcher. The findings, which result from first-hand experiences and secondary research sources, conclude that while there are various companies and resources that offer VC data, none can provide the necessary compilation of VC fund IRR data efficiently or inexpensively. This suggests that such a collection is needed.

Disciplines
Business
TOWARDS A CENTRALIZED VENTURE CAPITAL DATA SOURCE – THE KEY TO INCREASED MINORITY ACCESS TO VENTURE CAPITAL?

By

Maya X. Trujillo

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

WHARTON RESEARCH SCHOLARS

Faculty Advisor:

Harvey Rubin, PhD, MD

Professor of Medicine at the University of Pennsylvania;
Professor of Computer and Information Sciences at the University of Pennsylvania School of
Engineering and Applied Sciences

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2022
ABSTRACT

In 2016, Steven Kaplan and Josh Lerner wrote about how the shortage of reliable and comprehensive industry data has created challenges for academics who wish to replicate or build on previous venture capital\(^1\) (VC) studies (414). The author of this paper experienced a similar lack of data first-hand while attempting to validate findings from a 2008 study which found that minority focused VC funds tend to earn equivalent or higher returns on their investments than the returns of the VC industry (Bates and Bradford 2008). The data required was virtually nonexistent, which prompted a search for databases which collect VC data. This paper is a call to action for greater VC fund financial transparency. This paper’s original goal of validating a prior study should be replicable by any researcher. The findings, which result from first-hand experiences and secondary research sources, conclude that while there are various companies and resources that offer VC data, none can provide the necessary compilation of VC fund IRR data efficiently or inexpensively. This suggests that such a collection is needed.

\(^1\) “Venture capital (VC) is a high-touch form of financing used primarily by high-growth, innovative, and risky companies. VC funds invest in these companies on behalf of limited partners, who are mostly large institutional investors. Venture capitalists provide not only financing, but also non-financial support such as mentorship, strategic guidance, and network access” (Gornall and Strebulaev 2021, 2).
INTRODUCTION

Motivation

In 2008, a pair of researchers and professors from Wayne State University and the University of Washington published a remarkable conclusion: VC firms which invest in minority owned companies tend to earn equivalent or higher returns on their investments compared to the returns of the “VC industry mainstream” (Bates and Bradford 2008). This finding was at odds with a wide-standing belief that exists to this day surrounding minority (and broader social impact) focused funds (Morgan Stanley 2018; Morgan Stanley 2020).

For example, a 2018 report by Morgan Stanley found that investors think women and minority-owned businesses (WMBEs) “are twice as likely to perform below market average, as compared to non-minority and male-owned businesses. This assumption prevents WMBEs from receiving funding, and does not actually align with reality” (Germano 2018). Further research made it impossible to substantiate or disprove Bates and Bradford’s findings (2008) because many databases did not offer fund IRR data or it was unduly expensive to obtain. In addition, some data was anonymized rendering it unusable. Only one expensive data provider had minority focused VC fund data broken out.

I set out to explore the validity of the claims with more recent data in large part because of the great change in VC since the early 2000s when their research was conducted. Over the past twenty years, there has been dramatic growth in VC funding (TrueBridge Capital Partners 2021), a steep rise in minority group populations (particularly the Latinx population) (Krogstad 2019).

Impact investments are investments made with the intention to generate positive, measurable social and environmental impact alongside a financial return. Impact investments can be made in both emerging and developed markets, and target a range of returns from below market to market rate, depending on investors' strategic goals. The growing impact investment market provides capital to address the world’s most pressing challenges in sectors such as sustainable agriculture, renewable energy, conservation, microfinance, and affordable and accessible basic services including housing, healthcare, and education” (Global Impact Investing Network. n.d.).
and Noe-Bustamante 2021), and an increased number of entrants to the minority VC space (van Romburgh and Teare 2021). Minority focused VC funds have recently raised unprecedented funds and focus on investing in minority founders (Konrad 2021; "VamosVentures" 2021). Additionally, the focus on the racial wealth gap (Hernández Kent and Ricketts 2020; Bhutta, Chang, Dettling, and Hsu 2020) and racial inequity is more pronounced than ever. This disparity was highlighted with the emergence of the Black Lives Matter (BLM) movement in 2020 (Adams 2020) and exacerbated by the COVID-19 pandemic (Wilson 2020; Lopez III, Hart III, and Katz 2021). The BLM movement also ignited heightened investor interest in the minority focused VC space and VC fund interest in investing in minority founders (Franck 2020; Konrad 2021; Jan, McGregor, Merle, and Tiku 2020; Morgan Stanley 2020).

Early research into the data needed to address Bates and Bradford’s findings (2008) yielded limited results. Considering how many third-party sources (such as academic institutions, non-profits, private companies etc.) appeared to collect and offer VC data, it seemed reasonable to find and analyze an existing dataset, rather than collecting the data individually from a select number of VC funds. Bates and Bradford utilized the latter approach, and individually collected year-by-year cash flow data from 24 VC funds (out of 36 that were eligible from a larger subset of funds who had active National Association of Investment Companies membership in March 2001) who responded to their “detailed questionnaire regarding fund characteristics and monetary returns on their individual small-business investments” (Bates and Bradford 2008, 491).

To address Bates and Bradford’s conclusions (2008), my approach was to follow their methodology: calculate a sample of minority focused VC funds’ yearly aggregated internal rate
of return (IRR) and compare it to the VC industry’s “typical IRR.” The rationale for using IRR is as follows:

The primary measure of return in this industry is the internal rate of return (IRR). It is well-known (Ross, Westerfield, and Jordan 2000, chap. 9) that the IRR is an imperfect measure of returns. For the minority-oriented funds specifically, the problem of aggregating individual investments’ IRRs to the IRRs of individual funds was noteworthy. The IRR does not quantify negative and positive returns symmetrically; thus, a weighted average of the individual investments’ IRRs in the portfolio may not accurately measure the portfolio’s return. (Bates and Bradford 2008, 494)

The main difference was the timeframe the data would cover: 2010-2020. In their 2008 study, Bates and Bradford analyzed the VC funds’ IRR data they collected over a six-year timeframe, from January 1, 1989, to December 31, 1995. The researchers accounted for the downsides of an aggregated IRR by analyzing the performance of the minority-focused funds using four measures of investment returns: the overall IRR of each fund’s investments, the public market equivalent-vintage, the public market equivalent (PME)-NASDAQ performance measures for each fund, and the fund’s equally weighted average IRR for its individual investments (Bates and Bradford 2008, 496)

For the IRR of the fund, the PME-vintage funds, and the PME-NASDAQ, Bates and Bradford combined the cash flows of all funds being analyzed and their number of investments, treating the cash flows as one investment. The PME-vintage measure “provides a basis for comparing the returns generated by minority-oriented VC funds to those produced by the mainstream VC industry” (Bates and Bradford 2008, 496). It is calculated by dividing the present value of the minority fund’s returns (cash inflows) by the present value of its investments (cash outflows) (Bates, 496). Bates and Bradford did note that some of their results from the PME-vintage analysis were not statistically significant, so they could not draw conclusions about the relative overall performance of minority and mainstream VC funds from those numbers (497-
The third measure they utilized is the PME-NASDAQ, which compares investing in the NASDAQ index to investing in minority-focused funds (Bates, 498). The fourth measure used by Bates and Bradford is the IRRs for each of the individual investments made by the minority-focused funds they analyzed in the study. They utilized cash flow data for all investments initiated by the funds during 1989–95. Their findings, after conducting an equally weighted average calculation, showed that larger individual investments tended to have higher IRRs than smaller ones (Bates, 498).

Bates and Bradford’s conclusions were:

Regarding returns, our analyses of the PMEs and results of the VC returns of previous studies support a more conservative position that the returns to the minority-focused funds are broadly consistent with those of mainstream funds. Conclusions about the relative returns generated by minority- and mainstream-focused VC funds must be tempered by the reality of underlying databases that are not perfectly comparable, a small number of minority VC funds, and mature investment sample sizes that are small. Our ongoing research seeks to clarify the nature of the comparative returns earned by minority-focused funds and their counterparts in the mainstream VC industry. (Bates and Bradford 2008, 503)

The results of Bates and Bradford (2008) sparked the research which eventually led to this investigation of the current VC data landscape.

What led to this topic

Initial research revealed that websites such as Wharton Research Data Services (WRDS), Pitchbook, Preqin, Refinitiv, CB Insights, and others failed to possess the one data point necessary to replicate more recent findings – a VC fund’s year-by-year aggregated IRR. A similar datapoint, the per-year return on investment (ROI) of a VC fund’s investment in a specific company, was relatively easy to locate through WRDS and PitchBook. But why not IRR? How can this be?
The VC industry and VC funds have always been somewhat of a “black box” (Dickey 2022) to outsiders, but this level of obscurity was surprising. Why can’t an investor or researcher easily replicate the findings of reports from Morgan Stanley, TrueBridge Capital Partners, PitchBook, or even Bates and Bradford’s 2008 study?

**VC funds are largely exempt from disclosure requirements**

Continued investigation led to the conclusion that good and reliable collections of data on VC are difficult to find (Hellmann and Puri 2002; Kaplan and Lerner 2016). One of the most apparent reasons for the lack of IRR data arises from the fact that the VC’s have virtually no public reporting requirement. VCs are exempt from the Dodd-Frank Wall Street Reform and Consumer Protection Act, section 207 of which states: “Exempts an investment adviser who advises solely venture capital funds from registration requirements with respect to the provision of investment advice relating to a venture capital fund” (H.R.4173 2022). Not long after the act’s passing in 2010, the Securities and Exchange Commission, or SEC, adopted the Dodd-Frank Act’s amendments to the Investment Advisers Act in 2011 (SEC 2011). Simply put, since VC funds are privately held and not publicly traded, they do not have to publicly disclose their financial information and they only divulge a limited amount of information to the SEC (Kaplan and Lerner 2016, 414).

**Databases collect VC data through voluntary collection**

In 2002, Steven Kaplan, Per Strömberg, and Berk A. Sensoy published the paper, “How Well Do Venture Capital Databases Reflect Actual Investments?,” which discusses a grand total of two databases where researchers can access venture capital data: VentureSource (from VentureOne) and VentureXpert (from Venture Economics). As would be expected, in 2022, more websites collect VC data, and they will be discussed later in the paper. Databases collect
data from public research, mining, and filtering (Prequin 2022, "Data Collection") along with third-party providers (Refinitiv, “Third-party”). But for private company data, VCs and portfolio companies typically report their individual data voluntarily (Kaplan et al. 2002, 1; Hellmann and Puri 2002). This finding, the research, and the data begged the following question: what would motivate a VC fund to voluntarily share its financial data with the world?

**Importance of institutional investors in VC funds**

Large institutions are the most common type of LP, examples of which include university endowments, pension funds, insurance companies, and financial firms (Zider 1998). Institutional investors tend to be the largest investors in VC funds (Napoletano and Schmidt 2022). Another significant category of VC investors is accredited investors\(^3\). Bates and Bradford reported that the funds they surveyed raised more than $2.1 billion from institutional investors (2008, 492). The largest of the investors in the Bates and Bradford study were public pension funds (492).

Public pension funds did not always have the legal ability to invest in VC. A policy change in 1979, motivated by a positively performing economy, granted pension funds the ability to invest in VC. In just eight years, “the share of funds committed to VC by pension funds” skyrocketed to 50% in 1986 from just 15% in 1978 (Lerner and Tag 2013, 164). The amount of dollars invested has continued to increase to the present day.

---

\(^3\) “Under the federal securities laws, only persons who are **accredited investors** may participate in certain securities offerings. One reason these offerings are limited to accredited investors is to ensure that all participating investors are financially sophisticated and able to fend for themselves or sustain the risk of loss, thus rendering unnecessary the protections that come from a registered offering. Unlike offerings **registered** with the SEC in which certain information is required to be disclosed, companies and private funds, such as a **hedge fund** or venture capital fund, engaging in these exempt offerings do not have to make prescribed disclosures to accredited investors…

An **accredited investor**, in the context of a **natural person**, includes anyone who:

- earned income that exceeded $200,000 (or $300,000 together with a spouse or spousal equivalent) in each of the prior two years, and reasonably expects the same for the current year, OR
- has a net worth over $1 million, either alone or together with a spouse or spousal equivalent (excluding the value of the person’s primary residence), OR
- holds in good standing a Series 7, 65 or 82 license.” (SEC Office of Investor Education and Advocacy 2021).
Contribution & Significance of paper

**LPs benefit from greater access to VC fund financial data**

The lack of aggregated VC data is important to LPs, VC funds themselves, and government and media entities. Typically, before deciding to invest in a VC fund, investors want to see as much financial performance data as possible from a fund (for example, data on how portfolio companies have performed and the overall performance of past funds year-by-year). LPs can access data from an individual fund if they are considering investment, and from a fund in which they have already invested. However, a collection of meaningful VC fund performance data would allow LPs to evaluate the performance of their current investments and potential investments compared to other VC funds and the industry. LPs would greatly benefit from stricter regulation which would require VC funds to disclose all their data. This creates an intriguing mismatch of goals where LP interests are aligned with government interests for transparency and LP interests are aligned with VC funds because both parties want high returns.

If Bates and Bradford’s findings (2008) prove to be accurate in today’s VC landscape, both minority focused VC funds and companies that are funded by minority focused VC funds would benefit. More investors would likely become interested in the space as it presents a validated economic opportunity in addition to the social investment. As a result, investment in minority focused VC funds would likely increase. Additionally, VC fund managers may want to avoid any negative regulation changes by showing they are being transparent, even if not to the SEC. Furthermore, VC funds should care about the lack of data because funds could measure their own performance against competitors. To do this, VCs also require a comprehensive, trustworthy, and accessible database.
Government regulators (particularly the SEC) and the media would also benefit from compiled VC data

Stakeholders, including government regulators and the media, may benefit from greater VC industry transparency because “substantial misunderstandings about this intermediary persist” (Kaplan and Lerner 2016, 414). The lack of reliable VC data has resulted in an unproductive and potentially harmful situation where people in powerful positions of government or media attention take bold positions on the VC industry, either positive or negative, based more on observation than hard facts (Kaplan, 414). Moreover, the SEC specifically should want to know about these gaps in data to evaluate whether further action or transparency is necessary. Also, the dearth of publicity suggests that readers of this paper, primarily LPs, fund managers, government and media are unaware of the publicly available data sparsity within VC because it hasn’t been noticeably discussed in the news/outside of scholarly essays.

The field of VC research would also benefit from more accessible VC data

VC researchers and future VC studies (some of which are discussed in this paper) would certainly benefit from greater data transparency as it would make it easier to conduct studies, have meaningful results, and make valid conclusions.

Summary of paper’s purpose and main results

This paper updates findings by Steven Kaplan and Josh Lerner, who conducted a thorough overview of the pre-existing and, in some cases, obsolete VC data options (2016). This paper summarizes the author’s personal experience trying to find data on VC fund returns for the originally intended thesis. Additionally, this paper is a call to action for greater VC fund financial transparency and proposes potential solutions for the problem. The findings, which
result from first-hand experiences and secondary research sources, conclude that while there are various companies and resources that offer VC data, none can provide the necessary compilation of VC fund IRR data efficiently or inexpensively.

**The findings of Bates and Bradford should be able to be replicated**

This all comes back to this paper’s original research question about minority-owned business performance and whether VC funds which invest in minority-owned businesses perform on par or better than the broader industry. Investigating Bates and Bradford’s findings should be able to be accomplished by any researcher. This author’s experience suggests their findings have not been updated because there are no reasonably publicly available VC data collections. Given the growth and impact of VC, validation of Bates and Bradford’s findings has the potential to be impactful.

**Implications of validating Bates and Bradford’s findings**

Regardless of whether Bates and Bradford’s conclusions (2008) are validated, updated findings about minority focused VC fund performance will help VC industry stakeholders. On one hand, validation may help change the perception of minority focused VC fund performance. Currently, minority focused VC is viewed as a subset of social impact investing. There is a widely held perception that investing in social impact funds sacrifices returns (Knowledge at Wharton Staff 2014). If minority focused VC fund returns are found to be equal to or greater than the industry average, minority focused VC funds have a compelling selling point that is substantiated by data and academic research. Even further than investor perception, VC funds themselves may realize there could be a competitive advantage in investing in minority entrepreneurs. If minority owned business investment increases, the VC racial capital access gap may start to close, impacting the racial wealth gap. On the other hand, if minority focused VC
funds are proven to underperform, then investors will know with clarity they are investing in social impact funds for social impact, rather than pure returns. Either way, the results of validating Bates and Bradford’s 2008 study are helpful for investors

**LITERATURE REVIEW**

*The growth in VC has been swift and impactful*

The VC industry has grown exponentially in the past couple of decades and shows no signs of slowing. VC investment value in U.S. based companies doubled from 2020 to 2021, from $166.6 billion to $329.6 billion. Even before the coronavirus pandemic, VC investment value in U.S. based companies more than tripled from $45.4 billion in 2011 to $144.5 billion in 2019 (PitchBook and NVCA 2022, 5). This investment value growth suggests an increasing reliance on VC financing.

*VC has a large impact on society*

Even though VC is not a primary vehicle for entrepreneurs to access capital (Stangler, Tareque, and Morelix 2016, 1), it is critical for financing some of the world’s largest, most innovative, and influential companies (i.e., Alphabet, Amazon, Tesla, and Apple) (Gornall and Strebulaev 2021, 2). About half of all public U.S. companies founded after 1968 (and that went public after 1978) have received VC funding (Gornall, 3). It is estimated over one-sixth of the largest 300 U.S. public companies would not have achieved their level of success or even existed without VC funding (30). There are many benefits for companies that received VC funding, the most significant of which include the possibility of expansion that comes with venture capitalist mentoring and network connections and the fact that venture capital money is not debt (Kuhlor 2022).
Along with VC funds playing a major role in the success of today’s most influential companies, a study published last year found that companies with VC funding account for 41% of the total U.S. market capitalization, 62% of research and development (R&D) spending, and 48% of patent value (Gornall and Strebulaev 2021, 3). VC is also:

…A major economic engine that generates job growth, spurs innovation and creates new business models that change the world. The funding VCs provide give nascent businesses and — industries — the chance to flourish. They help to bring ideas to life and fill the void that capital markets and traditional bank debt leave due to the high risk associated with limited operating history, lack of collateral and unproven business models. VC funds play a particularly important role when a company begins to commercialize its innovation” (Kuhlor 2022).

What academic resources on VC are available?

Various extensively cited sources delve into the details of VC (Da Rin, Hellmann, and Puri 2013; Gompers and Lerner 2001): what it is, how it works, who are the main stakeholders, what the return landscape looks like, etc. One of the most comprehensive overviews is “A Survey of Venture Capital Research,” published in 2013. “This survey reviews the growing body of academic research” on VC but also includes information about the VC investment lifecycle, investment strategies, estimating ROI, and VC’s impact on the economy (Da Rin et al. 2013). To summarize the scope of VC fund performance literature, there have been numerous analyses conducted which have, using their own methods, attempted to index or generalize VC fund rates of return (Bygrave, Fast, Khoylian, Vincent, and William 1989; Peng 2001; Hwang, Quigley, and Woodward 2005; Korteweg and Nagel 2016). None of these studies specifically attempted to validate the work of Bates and Bradford (2008); they were more generalized.

Databases: WRDS

Companies that possess VC data collections include WRDS, PitchBook, Preqin, Refinitiv, CB Insights, Burgiss, Dow Jones Factiva, VentureSource, and VentureXpert. WRDS is a data platform that grants researchers access to 600+ vetted and accurate datasets from over 50
vendors across various disciplines ("Wharton Research Data Services" 2022). The only two vendors which have select data on VC are PitchBook and Preqin. The PitchBook data available included investor information (i.e. headquarters information, total investments in a certain time period, total exits, preferred deal size), deal information (i.e. deal size, raised to date, financing status, CEO information, EBITDA, valuation revenue, price 30 days after offering), VC fund information (i.e. fund size, preferred verticals, fund investments), and company information (i.e. total raised, company financing status, headquarters information, employee count, first financing size, deal type, date, growth rate change, size multiple). One important note is that much of PitchBook’s deal information appeared to be about deals/companies that had become publicly traded, which suggests this data was easier to collect.

There are fewer Preqin data offerings on WRDS compared to PitchBook. Preqin’s available data included variables such as stage of the deal, deal date, deal financing size, portfolio company information in general, total known funding USD, primary industry of the deal, and the year established.

**PitchBook**

PitchBook is a financial data and software company which offers data on more than “3.3 million companies, 1.6 million deals, 390,000 investors and 79,000 funds” in the VC, Private Equity, and Mergers and Acquisitions spaces (PitchBook 2022). PitchBook’s website contains market analysis available on verticals (i.e., AI, AgTech and FinTech), emerging spaces, market size estimates, and market maps. PitchBook also has a research center with private market research, emerging tech and private company research, public market thematic research, coronavirus research, public center research, and even more categories of research. Additionally, PitchBook has a rich array of data available on companies and deals (i.e. recent PE/VC exits,
IPOs, unicorn companies, PE/VC-backed companies/ exits, M&As), public companies in various industries like healthcare, technology etc., investors and funds (i.e. currently fundraising, top seed investors, invests in women and minority-owned, PE/VC investors, closed funds), limited partners (i.e. current mandates, public pension funds, fund of funds, family offices), people (i.e. women founders, board members, executives at VC/PE-backed companies), debt and lenders (i.e. top lenders, buyout deals, refinancing opportunities), and service providers (i.e. law firms, buyside/sellside investment banks, accounting, lenders/commercial banks).

Within each category, there are options to filter the data such as deal criteria, deal types, industry, location, company status, exit/liquidity, investors, financial data, debt characteristics, and others. Inside each of these categories are even more specific filters. For example, in financial data, a viewer can filter by revenue growth percent (within minimum and maximum range), time period of growth, and by ranges of revenue, net income, enterprise value and EBITDA. In the companies and deals category, after searching for VC fund financial data, the results show very specific information on thousands of investors with information such as number of investments, active portfolio, investments in the last 12 months, AUM, HG location, and investor status.

The site also contains graphs which detail overall statistics (for example, the deal count of investments over time in PE, VC, M&A, and IPO/Liquidity). Selecting an investor (or VC fund) allows the viewer to access general information, contact information, team information, recent notes, affiliates, board seats, lead partners on deals, investments & M&A, investments by industry, investments by year, exits, co-investors, funds, funds by type, fund performance,
limited partners, service providers, associated deal lenders, investment preferences, and news (see Appendix I and II) (PitchBook 2022).4

**Preqin**

Preqin is the self-proclaimed “…Home of Alternatives™, the foremost provider of data, analytics, and insights to the alternative assets community” (Preqin 2022, "Who We Are") (alternative assets refer to investments which are not classified as “traditional asset classes” like bonds, stocks, or cash investments. Examples include VC, Private Equity, Hedge Funds, and Real Estate (Preqin 2022, "What Are Alternative Assets?")). Similar to PitchBook, Preqin also appears5 to collect abundant data with categories such as investors (i.e. large public institutions, individual family offices; future investment plans, full investment history, contact details), fund managers (i.e. profiles, history, strategy, dry powder, AUM), investment consultants, and service providers), funds (i.e. complete profiles, strategic information, contacts, investors, deals, companies), performance (i.e. Net-to-LP performance data for VC funds including IRR, multiples, called capital, cash flow), companies and deals (VC deals & exits globally, portfolio company profiles including deal history and investors), service providers, investment consultants, and secondary market. Additionally, Preqin has aggregated charts and league tables (on private capital fundraising, dry powder & AUM, companies and deals, and private capital performance), private capital and hedge fund benchmarks, and insights (including research, conferences and events, and premium publications) (Preqin 2022, "Venture Capital Database").

---

4 I used my account/log in to view.
5 Through the free account I created, I was only able to view a limited amount of information without upgrading to Preqin Pro, especially compared to how much was accessible through PitchBook via the University of Pennsylvania Libraries level of access.
Refinitiv

Refinitiv is one of the leading providers of financial markets data, insights, and technology across the world (Refinitiv, “About Us”) and it indeed contains a robust collection of data. Upon first opening the workspace, a viewer sees a sample layout called the Global Markets Monitor which lists indices across the world and their price changes along with a graph of the S&P 500 index, and a suggested news feed that is constantly updating (see Appendix III).

Refinitiv offers data on markets with the options: global markets, equities, equity indices, equity market movers, industries, fixed income, FX, money markets, commodities, economic monitor, countries & regions, macro explorer, deals markets, and deals screener. The workspace also has up-to-date news including categories like top news, news monitor, IFR, Reuters Breakingviews, Refinitiv Newscasts, United States Policies, Ukraine Crisis, and COVID-19.

A user also can create their own chart of an index, for example, with time length options, chart visual options, the ability to add analysis directly onto the chart, add event markers and trendlines, annotate, add shapes, export the chart, create a data table, view the news at a certain date, and add the historical price change. I did not have access to Refinitiv Datastream, their chartbook. Refinitiv also has deep dives available on individual companies, and the categories for this segment include overview, estimates, financials, valuation, relative valuation, ESG, segments, peer analysis, price history, significant developments, debt structure, CDS, ownership summary, company tree, value chains, research, company events, and company filings. The other research options are advanced research and search and discover. The additional filings and events information include advanced research, search and discover, dividends, and corporate actions.
Refinitiv also allows a user to monitor companies, view a watchlist pulse of a company, and create portfolios of indices. This level of customizability makes Refinitiv stand out from the other databases. Lastly, there is a detailed search option which search options including currency, investment universe, asset class, deal type, investment date, company nation, and more filters the user can add. The search outputs can either be reports, league tables, or volume analysis.  

**CB Insights**

CB Insights offers financial markets data and research to customers along with a free trial, but they specialize in technology markets information (CB Insights 2022, “About Us”). More specifically, CB Insights offers extensive research into industries (automotive & mobility, banking & payments, consumer products, energy, financial services, healthcare & lifesciences, industrials, insurance, retail & services, TMT, and wealth management), geographies (Asia, Europe, Latin America, Middle East & Africa, North America), technologies & topics (artificial intelligence, cybersecurity, downround tracker, enterprise IT, internet of things, robotics, and unicorn tracker), infographics (business social graphs, geo-graphics, market maps, NExTT framework, periodic tables, timelines), research reports and webinars.

There is also an advanced search feature which allows the user to search CB Insights by research, companies, deals, investors, patents, earnings transcripts, and news. Within companies, the user can search by keywords, organization name, and industry, along with company attributes (i.e. location, company status), financing & exits (i.e. latest funding round, investment stage, deal size range, VC backed yes or no), and performance metrics/mosaic (i.e. percentile, market cap

---

6 I accessed the Refinitiv platform via the Refinitiv Workspace Application with an account through the University of Pennsylvania Libraries.
range, stock price range). Additionally, CB Insights allows the user to create heat maps, exits, and geographics using their data (CB Insights 2022).7

Burgiss

Burgiss provides data and analytics solutions globally for investors specializing in private markets (Burgiss 2022, “About Burgiss”). The company offers a variety of platform options for customers, but there is no free trial or free level of access. One platform is the Private i® Platform which is

…a secure, modern, web-based private capital portfolio management platform, expressly designed for private capital investors, to support the work of institutional investors, service providers, and asset managers, such as funds of funds. Private i centralizes the work of investment, risk, and operations teams to help make better investment decisions (Burgiss 2022).

The other platform Burgiss offers is the Caissa Platform which is

…the industry-leading multi-asset class investment analytics platform, developed for institutional allocators, including endowments, foundations, OCIOs, pension funds, and family offices. This web-based platform is designed to perform exposure, risk, liquidity, attribution, and private equity modeling analysis across multi-asset class portfolios (Burgiss 2022).

Additionally, Burgiss offers a license to their Burgiss Transparency Data, which contains “information on the holdings of private capital funds” (Burgiss 2022). Burgiss also offers access to their Burgiss Manager Universe, “the largest, best-in-class, private capital dataset of its kind: nearly $9 trillion from over 11,600 private capital funds and funds of funds, including private equity, private debt, and real assets” (Burgiss). Burgiss also has a Carbon Footprinting of Private Equity and Debt Funds. This tool “combines private company data from Burgiss…with modeling developed by MSCI ESG Research, for estimating corporate carbon emissions”

7 I used a 7-day free trial to view the information.
Lastly, Burgiss provides their Investment Book of Record (IBOR) Service, an “administrative and record-keeping service to investors in private capital funds” (Burgiss).

**Dow Jones Factiva**

Like Burgiss, Dow Jones Factiva requires a corporate ID and password/paid account to access their platform. They offer three main solutions for clients, including Dow Jones Risk & Compliance, Dow Jones Factiva, and Dow Jones Newswires. Dow Jones Risk & Compliance allows customers to “Conduct comprehensive due diligence on your customers and connected parties as part of a robust Know Your Customer (KYC) program,” “verify the integrity of your supply chain and business partners by assessing risks, such as bribery, corruption and dealings with sanctioned parties,” keep pace with sanctions developments and requirements, even as international and national regimes shift rapidly,” and conduct trade compliance to identify red flags for illicit activity (Dow Jones 2022, “Dow Jones Risk & Compliance”). Dow Jones Factiva enables customers to access “an unrivaled selection of global news and data accessible via a powerful research platform, on mobile devices or integrated via advanced feeds and APIs” (Dow Jones 2022, “Factiva”).

More specifically, the platform offers research, advanced analytics & data mining, media monitoring & corporate communications, competitive intelligence, sales funnel & business development, application development, content & workflow integration, and customer engagement & marketing. Dow Jones Newswires allows financial firms, professionals, and investors to “uncover trading, investing and deal opportunities and deepen client relationships with market-moving news, exclusive insights and rich data sets from trusted Dow Jones sources” (Dow Jones 2022, “Newswires”).
Confirming/updating past findings is important

Even in 2022, a shortage of quality data persists, particularly fund performance data. Even though the findings of Bates and Bradford’s study (2008) have not been validated or updated, the study has been referenced/cited in papers published as recently as 2021. One of these recent papers finds that female and minority owned companies are outperformed by male and nonminority owned companies when looking at company profitability and sales. The authors state that their findings are congruent with one study from 2016 but also contradict previous research which “made no conclusive statements in favor of either diversity or non diversity” (citing Bates and Bradford’s 2008 study as an example) and research which “found diversity as factors for performing even better” (Johansson and Li 2021, 20). This divergence of findings suggests that further and more timely research needs to be conducted to reach a clear consensus. However, the research can only be as thorough, consistent, and valid as the data itself.

For instance, a recent working paper disproves a long-held belief that very few VC funds outperform the S&P 500 (Harris, Jenkinson, Kaplan, and Stucke 2022). This study utilized what appears to be more reliable and recent data to show that “VC funds with previous performance in both the top and second quartiles [and even those above the median] outperform the S&P 500” (Harris et al. 2022, 23). Academic literature is ever evolving and constantly being updated. The implications for potentially positive and reaffirming findings like those from Bates and Bradford (2008) can materially change perception in the VC ecosystem and among its stakeholders.

The specific data and methodological challenges faced by VC researchers

Steven Kaplan and Josh Lerner (2016) wrote about how the shortage of reliable industry data has created challenges for academics who wish to replicate or build on previous VC studies. VC researchers face certain difficulties when deciding to conduct research such as what this
Computing financial returns requires good data and a solid methodology widely shared by researchers. In the case of VC, both these ingredients are scarce. Only recently has there been progress in obtaining reliable estimates. Most of these research challenges are not unique to VC, but apply to research on alternative asset (most notably buyouts) more broadly. Yet VC returns also pose some additional challenges due to the way VC firms obtain capital from their LPs, invest it, and return it back to LPs (Da Rin et al. 2013, 620).

One main challenge with conducting VC research is consistency of methods between studies

Analyzing VC data comes with a few inherent challenges. For example, calculating VC fund returns is far from simple. Not only is the industry volatile (Peng 2001, 3), but the performance of individual VC investments varies greatly (Bates and Bradford 2008, 494-95). Also, the risky nature of venture capital allows for potential scenarios where one or two investments become wildly successful and more than make up for the thirty other companies that either failed or stagnated.

METHODS AND RESULTS

While seeking data to explore the initial objective for this paper, I researched and contacted multiple different companies which publicized they had VC data collections. The companies were: WRDS, PitchBook, Preqin, Refinitiv, CB Insights, Burgiss, Dow Jones Factiva, VentureSource, and VentureXpert.

WRDS Takeaways

The author spent many hours downloading datasets, looking through the variables, and contacting WRDS support. Ultimately, WRDS did not have access to a dataset from PitchBook or Preqin with individual VC funds’ IRRs for a given year. After no success with WRDS, I reached out to PitchBook and Preqin separately.
**PitchBook Takeaways**

PitchBook's offerings, while initially promising, proved to be inadequate with my level of access. After reaching out to the company, a representative said “We do cover fund returns and we do label those that have a stated preference for Women and Minority Owned businesses. We do not sell our data individually, but many universities have access to some of it so I would recommend speaking to your school's library to see what level of access you may have.”

PitchBook account creation through Penn Libraries’ subscription followed and the wealth of information available was impressive. Penn does not currently have a subscription which allows a viewer to see a VC fund’s performance (see Appendix I and II), specifically the IRR of a fund within the larger VC firm, so this resource’s data could not be leveraged.

**Dow Jones Factiva Takeaways**

Similar to Pitchbook, Dow Jones Factiva appeared to offer the data I needed. After reaching out through the company website, a sales executive reached out quickly and said “…we do have the data that you are looking for as well as research on minority funds. Typically we provide this data via an API extraction & researchers have access for 12 months or longer. The data & extraction is around $22,000 and gives access to up to 2 million articles.” The price was shocking. After subsequent communications, I learned that this quote was, in fact, the student discounted price. The sales executive explained that “It essentially comes down to 2 components: the extraction process our team is required to complete & the data itself. The majority of the sources we own, but many we also license from the publications themselves because they have paywalls & require royalties from us.” By the time this company was discovered, it was too late to even attempt to find funding to buy this data. For full disclosure, this data was never viewed,
so whether Dow Jones Factiva actually possessed the exact data required for the original thesis topic is unconfirmed.

**Prequin Takeaways**

Conversely, Prequin was dismissed almost immediately. A live chat with a company representative on their website revealed that they do not collect the data I needed.

**Refinitiv, CB Insights, and Burgiss Takeaways**

Refinitiv, CB Insights, and Burgiss were also unhelpful. Refinitiv was an expansive, quantitative-heavy database that was difficult to use and maneuver, especially compared to PitchBook. The data required on their platform was inaccessible. While it may be in the database, it was unattainable. Like Refinitiv, CB Insights offers a free trial which allows access to some illuminating reports on VC, but no data that could be leveraged for this paper. Another similar company, Burgiss, unfortunately proved fruitless as well. After reaching out via the company website, a client coverage associate responded promptly directing the academic research request to the Private Equity Research Consortium (PERC). He also said “Given the nature in which our data is sourced, it is completely anonymized, so no individual fund names are listed. We also don’t categorize any funds with ESG attributes (likely something we would look to add in the future).” PERC never responded to my request.

Two other sources that had to be vetted were VentureSource and VentureXpert. These two databases were described as the primary resources for researchers available in 2002 (Kaplan, Strömberg, and Sensoy). Ironically, VentureSource was acquired by CB Insights in 2020 ("Our First Acquisition" 2020). and VentureXpert still exists, but only within the Refinitiv platform (Refinitiv Eikon Private Equity Content 2022). So, the once only sources of data were absorbed by larger companies.
DISCUSSION OF RESULTS AND CONCLUSIONS

After all the research conducted on companies who collect VC data, the initial conclusion is that while scattered VC data exists, none of the databases are fully consolidated to offer companies, investors, and academics reasonable access to useful information. In addition, no single database exists which has all the components necessary to replicate Bates and Bradford’s 2008 study – easy accessibility, simple navigation, non-anonymized financials, and free (or, at the very least, affordable) data for the researcher.

Some of the databases were accessed through the University of Pennsylvania’s extensive library relationships and resources. This institutional accessibility allows an authorized Penn user to avoid having to sign up for a free trial and run into a financial barrier. Refinitiv was a particularly complex database which appeared to require training and experience to use efficiently and effectively (see Appendix III). Dow Jones Factiva had a steep price tag that would have required extensive time and resources from my university to sponsor. I conclude that in 2022, updating the findings of Bates and Bradford (2008) is an unattainable goal for a university student and most researchers. For well-financed investigators who have years to dedicate, there may be a path to further vet Dow Jones Factiva’s data, perhaps by partnering with a foundation like others appear to have done. With more minority-focused VC funds than ever, and an even greater spotlight on racial inequity than ever, there is much work left to be done.

PROPOSED SOLUTION

There is a two-fold solution. First, at least one of the databases discussed in this paper can and should seek to become the VC industry database that collects and distributes data. Nearly all (except for WRDS) claim to be the best for private equity or alternative assets, but this more inclusive category is likely too broad. For instance, this thesis originally required two primary
variables: a VC fund’s IRR and whether the fund is investing in a minority-focused company. Most databases lack at least one of the two. This is not to imply that collecting this data is an easy feat or that it can be done without expense but having at least one database as the go-to, reliable first choice like VentureSource or VentureXpert used to would be valuable for researchers. Such a resource could provide the media one or two definitive sources for information, interviews, quotes, and most topics related to VC.

This all-inclusive database (or databases) should collect and offer accurate, anonymized data such as the year-by-year aggregated IRR of a VC fund, the yearly ROI of a single investment by a VC fund, the amount of money invested in each portfolio company, and the length of time that money has been invested/deployed. These numbers would be meaningful to LPs who are evaluating a fund’s performance and to researchers in the VC space.

Another option is for institutional investors to pressure VC funds to disclose their financial data to the proposed centralized database. Since databases rely on VC funds to provide their financials, existing databases cannot fill the holes in the currently missing data without data from the funds themselves. Why would private VC funds make their data (at least partially) publicly available? If institutional investors drive greater VC data transparency, VC funds will likely need to comply. VC funds rely on their LPs for reinvestment, network connections, and positive word of mouth reputation from their current LPs. Also, disclosure does not have to be the complete publication of financial statements or nothing. It could entail a variety of options. The least intrusive option may be to send the fund’s financials to the new and improved database(s) discussed above. The database(s) would then compile all VC fund data in one place that is easily accessible. Why might institutional LPs care enough to make this sweeping
requirement? First, for their own self-interests as already discussed. But LPs may also care to prevent any further government or regulatory body intervention, which is already taking place.

On February 9, 2022, the SEC “voted to propose new rules and amendments under the Investment Advisers Act of 1940 (Advisers Act)” to enhance the regulation of private fund advisers and to protect private fund investors by increasing transparency, competition, and efficiency in the $18-trillion marketplace” (U.S. Securities and Exchange Commission 2022). The rules would require registered private fund advisors to create quarterly reports for investors containing information such as expenses, fund fees, and performance, among other requirements and prohibitions, some of which apply to non-registered private fund advisors as well.

Gary Gensler, SEC Chair, strongly supports the proposal because the provisions are advantageous not only for investors, but also for companies raising capital from funds. He stated, “Private fund advisers, through the funds they manage, touch so much of our economy. Thus, it’s worth asking whether we can promote more efficiency, competition, and transparency in this field” (U.S. Securities and Exchange Commission 2022).

Furthermore, the SEC fact sheet explicitly states, “there is a need to enhance the regulation of private fund advisers” (Fact Sheet: Private Fund Proposed Reforms 2022). If the SEC is already prioritizing greater fund transparency for investors and greater regulation is on their minds, what could be next? LPs are likely pleased that the SEC is looking after their

---

8 “The Investment Advisers Act defines a venture capital fund as a private fund that, among other things outlined in Rule 203(l)-1 under the Investment Advisers Act: (1) represents to investors that it pursues a venture capital strategy; (2) generally limits redemption rights; (3) holds no more than 20% of the amount of the fund’s aggregate capital contributions and uncalled capital commitments in non-qualifying investments (often referred to as the 20% non-qualifying basket); and (4) limits the use of leverage” (“Venture Capital Fund” 2022).

9 “An investment adviser is an individual or firm that is engaged in the business of providing investment advice to others or issuing reports or analyses about securities for compensation. Investment advisers may include money managers, investment consultants, financial planners, general partners of private funds, and others who are compensated for providing advice about securities. Investment advisers are required to register with the SEC or applicable state securities regulators as a registered investment adviser unless they are exempt from applicable registration requirements (for example, as an exempt reporting adviser)” (“Investment Advisor” 2022).
interests, but also wary about the day when sentiments could shift in favor of even greater transparency. Perhaps this is the push LPs and VC funds need to demonstrate to the SEC that the VC industry can be financially transparent without compulsory governmental data collection.

**CONCLUSION**

VC continues to grow, and it is not poised to stop any time soon. VC helps fund young companies and drive their growth, but this capital is difficult to access. Minority focused venture capital funds have worked to offer capital to minority founders who may be overlooked. Investors remain wary about the profitability of minority (and social impact) focused VC funds because of their dual mandates of social impact along with providing returns. Validating or disproving Bates and Bradford’s findings (2008) would provide clarity to VC industry stakeholders. However, this study is not possible without accurate, accessible, and comprehensive data. The findings of this paper demonstrate that the current VC database offerings do not offer the compiled data necessary to replicate Bates and Bradford’s 2008 study. The VC industry requires an easily accessible, simple-to-navigate, non-anonymized, and free (or, at least, affordable) database. This study recommends that at least one database should become the go-to resource for researchers and take charge of collecting all VC data in one place. Additionally, this paper suggests that institutional investors use their leverage to require VC funds to disclose their financials to this database, as it would benefit LPs. The findings of Bates and Bradford (2008) should be replicable, particularly at this time where more minority-focused VC funds exist, and racial inequity has gained increasing attention. Much work remains to close the gap of minority access to venture capital, but at least one centralized VC data source can play a large role in changing the status quo.
REFERENCES


*Journal Of Money, Credit And Banking* 40, no. 2-3 (March): 489-504.


https://www.dowjones.com/professional/risk/.


https://doi.org/10.2139/ssrn.281804.

PitchBook and NVCA. 2022. "Venture Monitor Q4 2021."


Note: A user must upgrade their subscription to view the fund performance data (including IRR).
Source: PitchBook
Appendix II: Fund Terms & Fees Data

### Fund Terms & Fees

<table>
<thead>
<tr>
<th>Term</th>
<th>10 yrs</th>
<th>Management Fee (0)</th>
<th>10 yrs</th>
<th>Additional Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment Type</td>
<td>10 yrs</td>
<td>N/A</td>
<td>10 yrs</td>
<td>N/A</td>
</tr>
<tr>
<td>Exit Commitment</td>
<td>3 yrs</td>
<td>2%</td>
<td>10 yrs</td>
<td>N/A</td>
</tr>
<tr>
<td>Management Fee</td>
<td>1.5%</td>
<td>N/A</td>
<td>10 yrs</td>
<td>N/A</td>
</tr>
<tr>
<td>Final Stage UPF</td>
<td>N/A</td>
<td>N/A</td>
<td>10 yrs</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note: Another example of how the user must upgrade their subscription to view the Fund Terms & Fees data, which includes data on IRR and fund returns.

Source: PitchBook
Appendix III: The Refinitiv Application’s Sample Layout (Global Markets Monitor)

Note: This is what a user sees upon first opening the Refinitiv application.
Source: The Refinitiv application