Thematic Investing With Big Data: The Case of Private Equity

Ludovic Phalippou*

University of Oxford Said Business School

Using natural language processing, companies globally can be scored based on the frequency with which news articles contain both their names and private equity (PE) related vocabulary. An index can then be created, with the weight of each component set as a function of both their liquidity and their PE exposure scores. This procedure generates a large set of firms whose underlying business is PE-related. Even though the algorithm does not optimize on either return or correlation, we find that the listed PE index is highly correlated to, and has similar performance to, the PE fund market index. This low-cost and scalable process can be generalized to any theme an index seeks to capture.

Keywords: Big Data, Natural Language Processing, Private Equity, Thematic Investing

* Ludovic Phalippou is Professor of Financial Economics at the University of Oxford Said Business School.
Part of the research behind this publication was commissioned by J.P. Morgan Securities plc'

1. Introduction

Private equity (PE) funds are illiquid investment vehicles that have grown exponentially in size and numbers over the last two decades. In 2021, their assets under management (AUM) passed the $10 trillion mark.[1] Within PE, we commonly distinguish between real assets (and within this category: real estate, infrastructure and natural resources) and corporate PE (and within corporate PE, the main categories are leveraged buyout (LBO) and venture capital (VC)). In this paper, I focus on the largest segment in terms of AUM: the LBO segment, which is simply referred to as PE. The objective is to design a flexible, low-cost, liquid and innovative index that has high correlation with PE fund market returns.

To build any public index around a theme, the first step consists of assembling a relevant set of publicly listed companies. For the case at hand, we need companies whose profits are related to the PE fund market portfolio returns.

A traditional approach would consist of reviewing the accounting data of thousands of listed companies worldwide to evaluate their respective exposure to the PE industry. Such an approach seems impractical, costly, and necessarily subjective. It is also worth noting that it is not possible to use SIC codes because there is no such code for PE. In addition, SIC assignments are quite arbitrary (Hoberg and Phillips (2010)). In practice, in the case of existing listed PE indices, we observe that different providers use different sets of companies, hence apparently disagreeing with the definition of 'PE-related companies.'

In contrast, this paper uses a big data technique – natural language processing – and applies it to the huge volume of press articles to rank publicly traded companies according to how frequently each company is mentioned in the context of private equity. Once all the publicly traded companies in the world are scanned for PE exposure, which only takes a few minutes, the second step consists of assigning a weight to each company.

The traditional approach would be to set weights as a function of market capitalization. This choice is not only arbitrary, but in the case of the PE industry, results in unnecessarily volatile indices because a few large firms dominate the industry in terms of size. Moreover, consider the largest PE-related company: Blackstone. Blackstone has a large hedge fund division, is the largest private landlord in the US, has the largest global real estate portfolio in the world and is the largest shadow bank in the world. Of Blackstone's $571bn AUM, only $183bn (32%) is directly related to PE. Using the whole market capitalization of Blackstone to compute its weight in a PE index would therefore not be appropriate.

[1] Source: https://pro.preqin.com/analysis/dryPowderAUM

Instead, our weighting scheme is a direct function of the frequency with which a company is mentioned in news articles alongside terms such as 'private equity' and 'leveraged buyouts.' However, we need to impose strong liquidity constraints because many companies are thinly traded. We first restrict ourselves to stocks with a minimal dollar trading volume and those with a mention of PE/LBO in more than 25% of their news articles. Next, we set up an optimization algorithm that consists of minimizing the distance to the unconstrained weights (those derived from textual analysis) subject to a maximum dollar holding per stock that is equal to a week's worth of trading volume (trailing average).

The resulting weights vary over time, but are relatively stable. We also observe that the weights we obtain are well distributed with a minimum of about 0.5%, and a maximum of about 5% even though we have not imposed any upper cap. Also, the index contains 50 stocks on average and is therefore reasonably well diversified across different companies (but obviously not across sectors).

Specifically, as of October 2021, the top 10 holdings in the index are, from the highest to the lowest: HarbourVest (4.8%), KKR, 3i Group, Carlyle, Apollo, Eurazeo, Golub Capital, Ares, Partners Group, and EQT (3.3%). Blackstone is 12th at 3.2% and Hamilton Lane is 15th at 2.7%. The weights are therefore well distributed across a large set of companies even though these weights are *endogenously* determined by the algorithm.

Our index is most highly correlated with the S&P Listed PE index, which is that used by the listed private equity ETF of ishares. The main difference between this index and ours is that our index is less volatile, as expected given the above discussion. Importantly, our index is a priori more reactive (stocks are included or excluded quickly), and less costly to both trade and assemble. Our index also happens to have higher returns, but this is not a feature that results directly from our approach.

We also compare our index to that of the PE fund market. Several specialized firms calculate the PE fund market index return each quarter by using the net asset value reported by all funds at the beginning and at the end of each quarter, and accounting for the net cash flows occurring during the quarter. The most often cited index is the Cambridge Associates Global PE Index, but this index is no longer publicly available. Thus, we use the Burgiss global pooled LBO index instead.[2]

[2] We have the quarterly return of both indices for the period 2008-2019. Burgiss averages 9.96% and Cambridge Associates averages 10.01% (annualized); and the correlation between the two is 99.6%.

The Burgiss global pooled LBO index has an average arithmetic return of 12% (annualized) between (January) 2008 and (September) 2021. The S&P 500 Index average return over the same time-period is about the same: 11.7%. Our PE index, however, averages 14.8% and the S&P Listed PE Index averages 12.9%. Again, there is no fundamental reason for our index to outperform other listed PE indices or stock indices. However, given the boom in the PE industry over the last decade, we would expect that an index that captures best the industry outperforms indices that capture it less well.

Given the smoothness of the illiquid index, correlation at quarterly frequency is biased downwards. Yet, we find that the correlation between our index and the illiquid PE market fund index is as high as 80%. Obviously, however, the volatility of our liquid index is much higher than that of the illiquid index, and is also higher than that of the S&P 500 Index. To reduce the volatility of our index, it is possible, and trivial, to use derivative instruments such as put and call options.

Our index can be used in different applications. First, a liquid listed PE index can be used as a hedging instrument. Investing in PE requires significant liquidity risk management due to the 'commit and drawdown' model. It is necessary to plan for capital calls and have the cash ready when such calls are issued. There can be a mismatch between the speed of capital calls and distributions and thus a listed PE index should offer a hedge against this risk. In addition, managing this liquidity risk is particularly challenging in defined contribution pension plans, like the 401(k), therefore having a close liquid substitute might be attractive.

Second, another consequence of the 'commit and drawdown' model is that from a strategic asset allocation (SAA) perspective, investing in PE creates frequent deviations between target and effective capital allocations. A so-called denominator effect occurs when public markets go down but PE returns react with a delay. The PE allocation, therefore, becomes higher than the target allocation. Conversely, in up markets, investors often overcommit to PE to meet any given allocation to PE because only part of the capital committed is deployed at any point in time. Being able to trade a liquid index that mimics as closely as possible PE fund returns can offer a low-cost and effective solution to SAA deviations for institutional investors.

Third, for retail investors a liquid solution might be a more cost-effective way to obtain an exposure to PE fund returns. Investing in any PE fund comes with significant fixed costs (e.g., due diligence, legal review; see DaRin and Phalippou (2017)). Such fixed costs, as well as sometimes a lack of access, put smaller investors at a disadvantage in obtaining exposure to the growing PE industry. A cost-effective liquid solution might be particularly appealing to them. Nonetheless, large asset owners and hedge funds, too, might use a liquid substitute to place timely tactical bets on this growing industry.

Fourth, such an index can be used as a benchmark for this fast-growing asset class, which, at present, does not have any commonly accepted one.

The rest of the paper is as follows. Section 2 reviews the composition of other PE listed indices, and discusses issues with these approaches. Second 3 describes our novel and systematic approach to collect companies that are in the PE industry using natural language processing (NLP) and big data. Section 4 details the methodology to attribute weights to selected stocks in the index. Section 5 goes over the top 25 holdings in the index, and contrasts the top holdings of other listed indices with companies included in our index. Section 6 compares our index return to that of other listed PE indices and to that of the Burgiss global pooled LBO index. Section 7 concludes.

2. Existing Listed Private Equity Indices

2.1. Background on the PE sector landscape

As shown in Phalippou (2017), defined benefits pension plans and sovereign wealth funds are important sources of capital for PE funds. However, most of the capital is channeled by a few capital aggregators, sometimes called 'solution providers.' These service providers allocate PE capital either directly via a spectrum of vehicles they control (e.g. separate managed accounts, feeder funds, fund of funds, etc.) or indirectly via their advice.[3] These solution providers include banks with large private wealth departments (e.g. UBS) and specialized firms (e.g. Hamilton Lane, Stepstone). A firm like Hamilton Lane influences the allocation of as much as $500 billion, i.e. it effectively channels more than 10% of all the investments in PE worldwide (assets under management plus asset under advisory; see Phalippou (2017), chapter 15 for details).[4]

In addition, Phalippou (2017) highlights how the boundary between investors and fund managers has become blurred over the years. For instance, firms like Hamilton Lane, and Partners Group, were initially channeling money from institutional investors to PE funds sponsored by firms such as Blackstone, KKR, Carlyle, and Apollo. Nowadays, both Hamilton Lane and Partners Group act as PE fund sponsors. Thus, a large part of their business overlaps with that of traditional fund sponsors such as Blackstone and KKR. Conversely, fund sponsors like Blackstone and Carlyle have developed an advisory business very much like the historic business of a firm like Hamilton Lane. As a result, there is not much difference between a company like Blackstone and Hamilton Lane, as both have profits that primarily depend on how good the returns on the $10 trillion PE fund market are.

[3] See Lerner et al. (2020) for a comprehensive discussion and empirical analysis of side vehicles.
[4] https://www.hamiltonlane.com/Who-We-Are/Our-Firm/

Interestingly, both Blackstone and Hamilton Lane are publicly traded; as are Partners Group, KKR, Carlyle and Apollo. In this paper, we show that there are many other publicly traded companies whose profits are arguably related the PE fund market returns. These companies include placement agents, specialized investment banks and specialized law firms. In addition, there are several listed PE funds, fund of funds and holding companies that use similar investment techniques as PE funds. Moreover, there are firms that lend to companies that are subject to LBOs. The debt tranches they hold should have returns correlated to the equity tranche in the same deals, and these are held by traditional PE funds. Importantly, the fees they generate from providing debt to LBO transactions depend on LBO volume, which is itself primarily a function of the PE fund market returns.

To sum up, there are many publicly traded companies that are operating in the PE sector and whose stock market returns should have a reasonably high correlation with the return of the PE fund market.

2.2. Existing Listed PE indices

BlackRock and Standard & Poors Listed PE index

iShares Listed Private Equity (IPRV) is the largest private equity ETF, with assets totaling about $400mn as of June 2020. It tracks the S&P Listed Private Equity Index, which is weighted by market capitalization (with various upper limits). The index aims to capture "the performance of large, liquid, and listed private equity companies from developed markets." These companies meet certain size, liquidity, exposure and business activity requirements. Table 1 shows the top 10 holdings, which together represent 54% of the total asset value.

Table 1: Top 10 holdings of the iShares Listed Private Equity ETF

	Name	Weight (%)
1	PARTNERS GROUP	8.39%
2	BLACKSTONE	8.18%
3	BROOKFIELD ASSET MANAGEMENT	6.99%
4	KKR	5.98%
5	APOLLO GLOBAL MANAGEMENT	5.46%
6	3I GROUP	5.10%
7	CANNAE HOLDINGS	3.80%
8	ARES CAPITAL	3.61%
9	CARLYLE GROUP	3.61%
10	ONEX	3.54%

Invesco & RedRocks

The Invesco Global Listed Private Equity Portfolio (NYSE Arca: PSP) has assets totaling about $150mn as of June 2020. This ETF tracks the Red Rocks Global Listed Private Equity Index, and adjusts market capitalization with a purity score to emphasize a representation of firms that "pursue a PE model - buying private companies, enhancing them over 7-10 years, and then selling them." For example, alternative asset managers, holding companies, and investment banks, which have limited direct ownership in private companies, receive a lower purity rating.

The ETF includes business development companies (BDCs), special purpose acquisition companies (SPAC), master limited partnerships (MLPs), and other securities whose principal business is to invest in, lend capital to or provide services to privately held companies. Purity rating assignments are recalculated quarterly. Constituents have an adjusted market capitalization in excess of $200mn, and a minimum average daily trading value in excess of $500k.[5] Table 2 shows the top 10 holdings of the ETF, which together represent 42% of the total asset value.

Table 2: Top 10 stock holdings of the Invesco Global Listed Private Equity Portfolio ETF[6]

	Name	Weight (%)
1	WESFARMERS	5.93%
2	PARTNERS GROUP	5.66%
3	3I GROUP	4.93%
4	MATCH GROUP	3.84%
5	KINNEVIK	3.59%
6	KKR	3.51%
7	MELROSE INDUSTRIES	2.92%
8	BLACKSTONE	2.76%
9	FOSUN INTERNATIONAL	2.48%
10	IAC/INTERACTIVECORP	2.40%

Table 3: Top 10 holdings of the DWS ETF

	Name	Weight (%)
1	BLACKSTONE	12.47%
2	KKR	12.39%
3	PARTNERS GROUP	10.67%
4	APOLLO GLOBAL MANAGEMENT	9.47%
5	3I GROUP	9.17%
6	ARES CAPITAL	5.49%
7	WENDEL	4.17%
8	ONEX	4.13%
9	EURAZEO	4.09%
10	INTERMEDIATE CAPITAL	3.52%

[5] RedRocks claims they identified 402 securities, but the index includes only about 50 stocks.

[6] One of the top holdings is a money market fund and we have taken this out of the table.

DWS & LPX

DWS has an ETF written on the LPX Major Market Index, which aims to reflect the performance of the 25 most liquid listed private equity companies worldwide. The index is weighted by annual net dividend yield. Total assets are about $130mn. Table 3 shows the top 10 holdings, the top five of which alone make up 54% of the total weight.

Other Listed PE indices

ProShares Global Listed Private Equity Portfolio (BATS Trading: PEX) tracks another LPX index (LPX Direct Listed Private Equity Index). Its asset value is about $14mn. LPX Direct Listed Private Equity Index has up to 30 listed private equity companies worldwide whose primary business is to direct investment into private enterprises, and excludes companies whose primary business is private equity fund management (i.e. fund of funds).

The VanEck Vectors BDC Income ETF (BIZD) is not a pure PE index but rather includes many companies that are included in other listed PE indices. Its goal is to replicate the MVIS US Business Development Companies Index (MVBIZDTG), which tracks the overall performance of publicly traded business development companies. Its assets total about $200mn as of June 2020. The largest holding is Ares with a 19% weight and there are only 26 holdings in total. The top 10 holdings make up 63% of the index. Business development companies (BDCs) invest in the debt and equity of mid-sized private firms. Most BDCs provide debt financing.

Refinitiv offers a different approach. Its Thomson Reuters Private Equity Buyout Index tracks the fraction of LBOs across seven different sectors: 1. Consumer Non-Cyclical, 2. Consumer Cyclical, 3. Technology, 4. Industrials/Materials, 5. Healthcare, 6. Financials, 7. Energy/Utilities. In a nutshell, it tracks the fraction of LBOs that occur in any one of these industry over a given period of time and weights corresponding industry ETFs accordingly. The index is therefore a portfolio of seven ETFs (one for each of these sectors) whose weight is mostly based on the fraction of LBOs in that sector.[7] Its AUM is $20mn. The construction of the Refinitiv index is, therefore, very different to that of the other listed PE indices. As of June 2020, the composition is Technology 34%, Industrials and Materials 17%, Healthcare 10%, Financials 9%, Consumer Non-Cyclical 10%, Consumer Cyclical 14%, Utilities and Energy 6%.

[7] https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/private-equity-buyout-index-methodology.pdf

2.3. Limitations of current PE replication

The high concentration of these indices (except for Refinitiv) generates a high volatility, which contrasts with what investors experience when allocating capital to PE funds. In addition, many stocks included in these indices have low liquidity which can generate high transaction costs, especially as and if the corresponding ETF grows in value. For example, Onex, which is in the top 10 for both ishares and DWS has a daily trading volume of about $8mn.

There is overlap in the companies included in the different indices however each use different definitions of PE (e.g. some include real estate, others may exclude it, etc.). Weights are often based on market capitalization but as a handful of firms in this industry are much larger than others, the resulting indices are highly concentrated.

As mentioned in the introduction, and will be further highlighted below, there are many companies whose profits, hence stock returns, should be related to PE fund returns. Conversely, there are many PE firms that have multiple business segments, including some that are not PE. For example, the largest holding in most listed private equity indices is Blackstone (BX). However, BX has a large hedge fund division, and is the largest private landlord in the US, and the largest shadow bank in the world. Hence, BX is not purely a PE firm. In contrast, KKR is mostly PE focused.

Any index construction, at present, includes significant manual research and judgement. It is costly to monitor all the stocks trading in the world, keeping track of IPOs and delisting, mergers and acquisitions, companies changing their business segments etc. For example, Blackrock is growing a few divisions that are involved in PE investing and advisory. At some point, these divisions combined might be comparable in size to, say, Hamilton Lane. Should Blackrock not be added to a listed PE index given that Hamilton Lane is? This decision would require extensive research and ultimately would contain a degree of judgement and subjectivity.

Along the same lines, assume there is a publicly listed company in South Korea that manages PE funds in the Asia region. Research teams are often based in London or New York, and thus they may miss this company.

3. Index construction using Natural Language Processing and Big Data technologies

Researchers in the field of statistics have developed tools to transform unstructured and big datasets into structured and readily analyzable datasets. A private company called RavenPack has used these techniques to develop a natural language processing algorithm, which led to the creation of a structured dataset of news articles. RavenPack builds its dataset from 19,000 different sources (including premium news providers, regulatory and press wires, and publications). We use only press releases and newspaper articles; and require these articles to be in English.

Company names are identified in each article, tagged, and given a *relevance score*, between 0 and 100. Variables that are used to establish the score include where the company was mentioned (headline, first paragraph, second paragraph), the number of times it is referenced, and the number of companies mentioned in the article. Higher values indicate greater relevance: a score of 0 means the firm is passively mentioned. RavenPack states that scores above 75 corresponds to situations where the firm is significantly involved with the main event described in the news article.[8] The number of unique articles increased from 520,419 in 2004, which is when the digitalization of news articles effectively started to 2,214,891 in 2019, covering 43,000 unique companies.

News articles are transformed into structured data that can be searched for using specific tags. As we aim to capture the return of private equity leveraged buyout funds, we use the following unambiguous tags (plural and singular): Leveraged buyout, Private Equity, Private Market.

Such an approach to index construction is systematic, flexible, and data-driven. We quickly capture information over a large set of companies at minimal cost and avoid any potential cognitive biases. As an extension to this approach, we can over-weight certain strategies, industries, or countries. For example, we could add words like 'venture capital.' In addition, we can quickly estimate the exposure of a company to PE. To capture the size of Blackrock's PE business, we may study its accounting data and with some further judgment, decide on the fraction of its earnings that are PE-related. Our method, in contrast, simply and quickly measures the frequency with which Blackrock appears in news articles in the context of PE. Using our dataset, we observe that the frequency has indeed grown over the years, but remains low: as soon as it is otherwise, the index will reflect this new situation.

[8] RavenPack also enables to detect different stories on similar events using a Novelty score, thereby filtering for stories within the day to avoid duplicates (referred as novel articles in this paper).

4. Index Construction

We start with the 43,000 companies in RavenPack and select those that are publicly listed and have an average daily trading volume (ADTV) of at least $100,000 over the past month and at least $250,000 over the past year (and require them to have been trading for at least one year).

For each calendar month *t*, we count the number of novel articles that mention company *i* with a relevance score above 75, and any word from our list above and denote this PE score for company *i* $SPE_{i,t}$. We also count all the novel articles that mention company *i* with a relevance score above 75 and denote this variable $STOT_{i,t}$. Small companies might have no news coverage in some months. A company is included in our sample only if the sum of $STOT_i$ over the last 36 months is greater than 50. We define the PE Exposure Score for company *i* at the end of month *T* as:

$$PEES_{i,t} = \frac{1}{N} \sum_{t=T-35}^{T} \left(\frac{SPE_{i,t}}{STOT_{i,t}} | STOT_{i,t} > 0\right)$$

Where the time unit *t* is a calendar month, and *N* is the number of months with a strictly positive value for $STOT_{i,t}$.

Taking an average over three years (36 months) as shown in the formula above is preferable to, say, adding up the numerator and denominator separately. For example, when companies are in talks with PE firms to sell to or acquire a division, or are involved in a legal dispute with a PE firm, they are cited alongside PE-related words for a few months. By taking an average, these temporary spikes in PE-related content have a low influence on the PE Exposure Score. If a company's business is private equity, it will be mentioned in that context each month.

While we could simply weight all the companies in the index according to their PEES this would result in a few practical issues. First, there would be many companies with negligible weights, which makes index tracking more costly. In addition, companies that are not directly involved in the business of private equity can have a non-zero PEES because they dealt with PE on occasions; for example, they sold a division to a PE fund. We set the threshold for inclusion at PEES>25%. Empirically, this coincides with a minimum weight in the index of about 0.4%, and the results are not significantly sensitive to this choice.

Second, many stocks are thinly traded. As a result, we split the index tracking fund into two funds. The first fund trades over the first half of the month (about 10 days) and the second fund trades over the second half of the month (about 10 days). In addition, we impose two liquidity constraints.

The first liquidity constraint is based on the index Size in comparison to the Average Daily Trading Value (ADTV) over the previous 365 days (12 months).[9] No position can be larger than one week's (five days) worth of trading volume (or more than 5% of the index for diversification purposes).

$$Cap_{i,T} = Min\left(\frac{5 * ADTV_{i,T,12}}{Size_T}, 5\%\right)$$

The second liquidity constraint is to prevent large turnover. No position can change by more than twice the (previous month) ADTV. As exposures are rebalanced over 10 consecutive days within one fund, it means that no more than 20% of ADTV will be traded on any given day:

$$Tcap_{i,T} = \frac{10 * 20\% * ADTV_{i,t,1}}{Size_T/2}$$

To illustrate, if the total index size is $200mn, each sub-fund has $100mn. A stock with a $1mn of average daily volume (say over both the previous year and month) cannot have a weight of more than $\left(\frac{5\times 1mn}{200mn}\right) = 2.5\%$ and the weight cannot change within each fund from one month to the next by more than 2% $\left(\frac{2\times 1mn}{100mn}\right)$. These constraints are, therefore, relatively loose, but they do bind for many stocks given their low trading volume.

The weight of the index invested in company *i* in month *T*, $Pos_{i,T}$, is the solution to the following constrained optimization problem

$$Min\left(Pos_{i,T} - \frac{PEES_{i,T}}{\sum_{s=1}^{Np} PEES_{s,T}}\right)^2$$

Subject to (1) $Cap_{i,T} > Pos_{i,T} > 0$

(2) $|Pos_{i,T} - Pos_{i,T-1}| < Tcap_{i,T}$

[9] ADTV is converted into US dollars using the relevant currency conversion rates. We exclude the highest three and lowest three values within the lookback period to avoid extreme values.

5. Index Composition

We start our PE index in January 2008, three years after the effective start of news digitalization. Table 4 shows the 25 highest weights in our index as of May 2020 out of the 55 companies that satisfied all the criteria (PE relevance and liquidity). The weights obtained from the optimization setup are well distributed across a range that spans from 0.5% to 3.7%. With a $125 million sub-fund size, the first liquidity constraint in the problem above is binding for 22 companies.

We search for the business description of the selected companies and group them into four types. Importantly, the line between these categories is blurred, and it is difficult to isolate the PE-related business part of each company from their other businesses. Business descriptions and accounting data do not offer a clear way to assess PE exposures, which is precisely what makes an NLP-based approach attractive. The top 10 holdings together represent only 29% of the total asset value.

Table 4: Top 25 Index Components and Weights as of May 2020

	Company Name	Weight	Main Stock Exchange	Type	Number of listed PE indices stock included in	PE AUM[10]
1	HARBOURVEST GLOBAL	3.4%	USA	Fund of funds	2	68
2	3I GROUP	3.4%	UK	Fund sponsor	3	16
3	EURAZEO	3.2%	France	Fund sponsor	3	22
4	KKR	3.0%	USA	Fund sponsor	3	75
5	PANTHEON INTERNATIONAL	3.0%	UK	Fund of funds	2	49
6	CARLYLE GROUP	2.8%	USA	Fund sponsor	3	81
7	PARTNERS GROUP	2.8%	Switzerland	Fund sponsor	3	45
8	OAKLEY CAPITAL	2.7%	UK	Fund sponsor	0	4
9	GOLUB CAPITAL	2.4%	USA	BDC	3	30
10	HG CAPITAL TRUST	2.2%	UK	Investment trust	1	11
11	BLACKSTONE	2.2%	USA	Fund sponsor	3	183
12	INTERMEDIATE CAPITAL	2.2%	UK	Fund sponsor	3	45
13	SANNE GROUP	2.2%	UK	Serv. provider	0	n.a.
14	APPOLO GLOBAL MANAGEMENT	2.1%	USA	Fund sponsor	2	68
15	HAMILTON LANE	2.0%	USA	Fund of funds	2	68
16	HOULIHAN LOKEY	2.0%	USA	Serv. provider	0	n.a.
17	ARES MANAGEMENT	1.9%	USA	Fund sponsor	0	22
18	BARINGS	1.9%	USA	BDC	1	n.a.
19	NB GLOBAL	1.9%	USA	Investment trust	0	n.a.
20	ARES CAPITAL	1.9%	USA	BDC	3	16
21	PJT PARTNERS	1.8%	USA	Serv. provider	0	n.a.
22	NOAH HOLDINGS	1.8%	USA	Fund sponsor	1	25
23	CAPITAL SOUTHWEST	1.8%	USA	BDC	1	0.6
24	F&C	1.8%	UK	Investment trust	0	0.5
25	ICG ENTERPRISE TRUST	1.8%	UK	Fund of funds	1	1

[10] As of December 2019. Source: Company's website.

5.1. Fund sponsors

PE funds are run by independent, specialized legal entities called PE firms, a.k.a. management (or advisory) companies or *fund sponsors*. For example, Permira Advisors is a firm located in London with several employees, and manages a set of PE funds. Permira Advisors is a fund sponsor. These PE funds invest third party capital into various companies. Holding the stock of a PE fund sponsor gives an exposure to the fee stream generated from managing third party (clients) that is earned-marked for PE investing. More specifically, revenues are mainly derived from i) management fees, which are mainly a function of PE activity level, which in turn depends primarily on PE fund market returns, and ii) carried interest, which directly depends on PE fund returns. Hence, PE sponsor stock returns are expected to be correlated to the PE fund market returns and should therefore offer a close substitute to an investment into the underlying PE funds.

The four largest PE fund sponsors in the world are publicly listed: KKR, Carlyle, Blackstone (BX), and Apollo. Our methodology picks up these four PE firms but attributes slightly different weights to each one of them; they are respectively positioned 4th, 6th, 11th and 14th in Table 4. BX, although the largest PE firms in the world, does not have the highest weight. We note that BX has a large hedge fund division worth $81bn, is the largest private landlord in the US and has a global real estate portfolio of $163bn and is the largest shadow bank in the world with a $144bn portfolio of mostly corporate loans (although most of them linked to companies subject to LBOs). Of BX's total $571bn AUM, only $183bn (32%) is directly related to private equity. Similarly, Apollo's $316bn AUM is made up primarily of credit ($210bn), which is mostly corporate and structured credit (although, again, most of these are linked to companies subject to LBOs). PE makes up 22% of Apollo's total AUM. In contrast, KKR has $75bn of its total $218bn AUM as PE (34%) and a sizeable portion of its other funds are related to LBOs via leveraged credit channels ($73bn) – putting the PE-related funds closer to 68%. Carlyle has $81bn of its total $216bn as pure PE, with the rest of its AUM split almost equally between credit, real assets and investment solutions (which include some fund of funds, and separate managed accounts).

Fund sponsors with the highest weights are 3i Group and Eurazeo positioned at 2nd and 3rd place respectively in Table 4. These organizations started essentially as publicly traded closed ended funds, and increasingly managed third-party capital. As a result, an equity stake in these stocks is increasingly an exposure to the fee income generated from sponsoring PE funds, just like the big four we just covered. For example, Eurazeo has an AUM of $21bn, $14bn of which is the management of third-party capital. Meanwhile, 59% of its AUM is in private equity, while the rest is tightly related to PE as it sponsors PE fund of funds, and private debt funds (whose capital is

mainly invested into the debt of PE-held companies). Finally, 63% of 3i's total AUM is in PE ($10bn of a total $16bn), with the rest in private equity infrastructure.[11]

Partners Group was founded in 1996 and was initially a publicly listed closed-ended fund of funds, like 3i Group but investing in PE funds rather than directly in deals. Over time, Partners Group launched a wide range of private market funds which it sponsors. Partners Group currently has an AUM of $94bn, 48% of which is in PE, with the rest split across private equity infrastructure and real estate, plus private debt.

Ares Management is the next fund sponsor on the list (17th). Ares has an AUM of $149bn but only $22bn PE, while the rest is in private debt: direct lending to middle market companies, mostly in connection with leveraged buyouts. Ares Management is also the parent company of Ares Capital (a BDC, mentioned below; 20th largest index holding) which currently has a portfolio of 355 companies backed by multiple PE sponsors. Oddly, no other index picked up Ares Management.

Noah Holdings has three business segments, which include an asset management division, which sponsors a range of private equity and real estate funds, as well as some fund of funds. Noah Holdings (22nd) is present in only one of the existing PE indices. In addition, although trading on the NYSE, it operated mainly in China with an AUM of $25bn.

Intermediate Capital Group (ICG) was founded in 1989 as a provider of mezzanine finance for LBO transactions. In recent years, it has grown its fund sponsoring activities, mainly by acquisitions. In 2015, it acquired Graphite Enterprise Trust, which is one of the oldest fund sponsors in the world, having been founded in 1981. This division was renamed ICG Enterprise Trust, and sponsors a fund of funds, and is publicly listed. This was picked up by our algorithm (25th highest weight). ICG recently announced ambitious plans to build a private equity business in North America. Only one other listed PE index picked it up.

Oakley Capital's (8th) sole business is to sponsor PE funds and provides growth capital to relatively mature middle-market companies. It completed several high-profile exits (education institution Inspired, to Warburg Pincus in May 2019; real estate platform Home, to Mayfair in February 2020). In 2019, it raised $2bn for Fund IV. Meanwhile, its AUM is about $4bn. Surprisingly, no other listed PE index includes this stock.

[11] As of 2019, 10% of the infrastructure division were invested in Wireless Infrastructure Group, a UK-based wireless infrastructure company, which it sold that December to Brookfield.

5.2. Fund of funds (FoF)

Table 4 is topped by the world's largest fund of funds almost in order of their founding date and AUM. Modern FoFs provide a range of services, that are, in fact, similar to the range of services that large private equity firms such as Carlyle or Blackstone offer. Like large PE firms, their revenue also mostly comes from management fees (and carried interest). FoFs manage separate accounts (customized to each individual client and structured as single client vehicles) and act as fund sponsors for many private market funds on the primary market and on the secondary market. They also do advisory work (due diligence, strategic portfolio planning, monitoring and reporting services), in addition to providing reporting and analytics solutions. Hence, these companies could have very well been included with those in the previous section, or included with the service providers (fourth category).

HarbourVest Global was one of the first closed ended FoF globally, having been established in 1982, and has an AUM of $68bn. Most of its funds are in HarbourVest's private equity FoF. Pantheon is also primarily a FoF and is listed on the London Stock Exchange with an AUM of $49bn. Of its new investments in 2019, for example, most were in primary funds (IK Investment Partners, European Mid-market Buyout Fund, and LYFE Capital, an Asian growth equity fund) while one third of its portfolio was through co-investments in companies themselves (e.g. a commitment with HG Capital to invest in Visma, a financial software provider in Northern Europe).

Hamilton Lane is the largest investor in PE in the world with assets under management plus assets under advisory at a whooping half a trillion. Launched in 1991 as a private equity advisory firm, Hamilton Lane has grown to sponsor several private market funds (funds of funds, private debt funds etc.). Hamilton Lane provides a range of services: separate accounts (customized to each individual client and structured as single client vehicles); specialized strategies (fund of funds, secondaries, co-investments); advisory (including due diligence, strategic portfolio planning, monitoring and reporting services); and reporting and analytics solutions. Hamilton Lane had an IPO in March 2017 and was added to our index in May 2019.

5.3. Investment Trust & Business Development Companies

Investment trusts are closed ended publicly listed funds which invest in private companies. They invest their balance sheets, and raise capital via initial public offerings and seasoned equity offerings. They are a legal category recognized in the United Kingdom and Japan. Investment trusts in our index include HG Capital Trust, which is a midsize firm with all of its portfolio invested in PE (mainly investments in mid-market European companies in services or technology-related

sectors), and has an AUM of $11bn. F&C is part of the list with a lower weight, as only part of its portfolio is invested in PE (7.7% of its total $5.6bn AUM, with a current upper limit on long-term exposure to PE investments of 20% of total AUM). NB Global Floating Rate Income Fund Limited invests in a global portfolio of below investment grade senior secured corporate loans, often from PE-backed companies.

A business development company (BDC) is a type of investment trust. The U.S. Congress created the BDC category in the 1980s. To qualify as a BDC, a company must invest at least 70% of its assets in private or public U.S. firms with market values of less than $250 million and must provide managerial assistance to its portfolio companies. BDCs are, therefore, the closest equivalent to PE funds; they are effectively publicly listed PE funds. However, most BDCs only provide loans to private companies and would therefore not be directly involved in the PE industry except when those loans are for PE-backed businesses. In addition, most BDCs are small and thinly traded. There are 87 BDCs; however, only 28 have more than $1bn AUM.

Golub (9^{th}) has $30bn AUM (but a $2bn market cap). Most of its investments are in middle market companies backed by PE funds. For example, in February 2020 Golub participated in a $1.6bn debt facility that financed the acquisition of MRI Software by Harvest Partners, a PE firm.

Barings BDC is managed by Barings LLC, a $327bn AUM asset manager, itself a subsidiary of MassMutual and is the closest competitor to Golub. Ares Capital is the BDC with the highest market cap ($7bn), and is a subsidiary of Ares Management which, as mentioned above, works primarily with its credit team to invest in private companies in transactions sponsored by PE funds. Ares Capital's AUM is $16bn. Meanwhile, Capital Southwest was formed in 1961 and elected to be regulated as a BDC in 1988, and changed strategy to be more PE-focused in 2014, which is what our algorithm seems to have picked up. Its total balance sheet assets are worth $600mn. Investments include private debt to PE-backed and equity co-investments.

5.4 Service providers

PE was pioneered by a few investment banks; these include HSBC in China, and AIG in Latin America and Asia. Nearly all large UK PE firms were spun out of investment banks. Nowadays, investment banks often act as placement agents for PE firms or advisors for M&A deals; as such, their revenues are dependent on the volume of LBO activity. Holding a share of investment banks that are highly involved in PE is therefore not dissimilar to holding a share of, say, KKR. In both cases, corporate revenue is a function of the fees it collects, which is a direct function of LBO volumes.

Investment banks collect at least the M&A fees coming from LBO activity, hence at least the typical 2% of the LBO volume each year. They also arrange debt packages for LBOs, making this source of income a function of LBO volumes as well. For most banks, PE activities is still a relatively small fraction of their revenues, this is why most large generalist investment banks do not appear on our list. However, those that do are tightly linked to LBO activity volume.

The investment bank with the highest weight is Houlihan Lokey. It is a placement agent but also an advisor in PE restructuring and a sell-side advisor for mid-cap companies. Its private funds group has raised 221 PE funds aggregating more than $334bn, plus offering a wide range of investor-relations activities.

PJT is an interesting case because it is a spin-off of Blackstone. Hence, all the listed PE indices used to hold PJT via their holdings in BX. Once PJT spun out, none of the other PE indices picked it up. There is no valid reason for this, rather it is purely arbitrary. Our algorithm picked up the fact that PJT is involved in PE activities and therefore includes it as part of the index. PJT works in restructuring, and private equity and alternative assets advisory. In addition, PJT acts as a PE placement agent, and these activities make up 19% of their 2019 revenues.

Sanne (13th) conducted its IPO in 2015, and specializes in PE fund administration, i.e. the services offered include fund establishment, regulatory applications, transaction management, financial reporting, consulting on fee arrangements and verification. Sanne's revenues therefore primarily depends on LBO volume, just like all the other companies included in the index. Again, the main driver of its LBO volume is PE fund returns, hence the expected correlation between this stock return and that of PE funds.

Surprisingly, no other listed PE index include these service providers whose revenues and returns are tightly related to PE fund returns.

5.5 Companies present in other listed PE indices that we missed

As just discussed, our index includes many companies that are missing in other listed PE indices. But it is also the case that some companies are included in other indices and are not in ours. For instance, eight of the top 10 in the S&P Listed Private Equity Index are included in our index.

Brookfield represents as much as 7% of the S&P LPE Index. It was not included in our index in May 2020 because it is cited in the context of PE in 20% of news articles and we set our threshold at 25%. Brookfield does some PE – as shown by its 20% score – but is primarily a real asset

management company (real estate, infrastructure, renewable power). However, this firm has continued to expand its PE activities and entered the index later in the year. It is part of the index as of October 2021 with a weight of 2.3%.

Cannae Holdings, which is the 7th largest holding of the S&P LPE Index is missing. This company has not been cited in news article in a PE context. Cannae is a relatively small and diversified holding company (restaurants, healthcare, finance). From the company's website, it is difficult to tell whether it is more akin to a hedge fund, or a growth capital fund. It does not seem to use leverage and that might be why terms such as 'LBOs' and 'PE' are never mentioned in related news articles. It also seems to be mainly holding the Restaurant Group which is a publicly traded company that holds several restaurant franchises.

The largest holding in the RedRocks Index is Wesfarmers, which has 6% of the index, and is missing in our index. Wesfarmers is the largest Australian company by revenue. Wesfarmers was founded in 1914 as a co-operative to provide services and merchandise to Western Australian farmers. It was listed on the Australian Securities Exchange in 1984. Although Wesfarmers grew into a major retail conglomerate, the expected correlation between its stock return and that of illiquid PE indices is not obvious.

The fourth largest holding in RedRocks Index is also missing from our index: Match Group – 3.9% of the index. Match Group is an internet company that owns and operates several online dating services including Tinder, Match.com, Meetic, and OkCupid. It was briefly a subsidiary of IAC (until July 2020). Match Group does not appear in the RedRocks Index though, but IAC does. IAC is an American holding company that buys and sells companies in the media and internet industry (e.g. Ticketmaster Group, LendingTree, TripAdvisor, Newsweek). IAC can be seen as a venture capital or growth capital fund, but probably not as a PE/LBO-related company.

Three other holding companies are among the top 10 holdings but are not present in our index. The first is Kinnevik AB, which is similar to IAC, but for Sweden. Kinnevik was founded as an investment company in 1936. Currently, its two most significant holdings are a 26% stake in Zalando – a European e-commerce company with about $7bn in revenue – and a stake in a publicly traded company called Tele2 which operates in the telephony, cable TV, and internet industries.

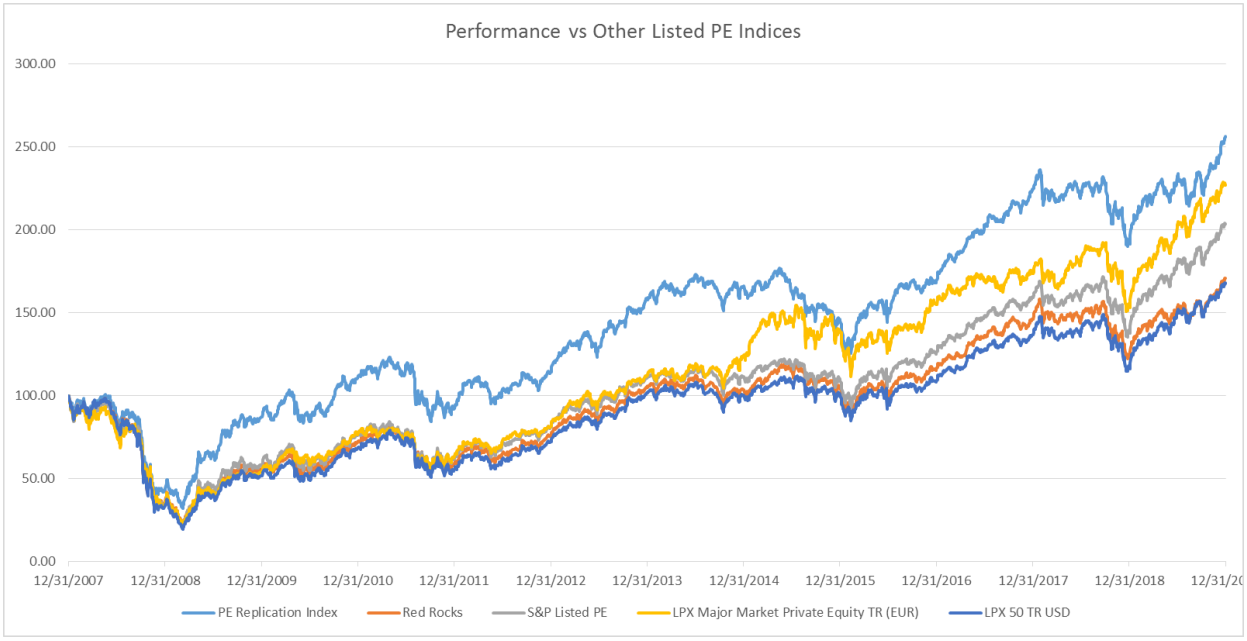
Second is Fosun International, one of the largest Chinese conglomerates, whose interests span across healthcare, finance and entertainment. In 2010, it partnered with Carlyle to create a Chinese PE fund, but this seems to represent only a small part of its activities. Fosun received a 6% PE-relevance score.

The third is Melrose Industries, a UK-based holding company. The company aims to buy and turn around underperforming businesses. Melrose has bought and sold numerous engineering and manufacturing companies.

Finally, the top 10 holdings of the LPX Index are all included in our index, except for Wendel, a holding company, which received a PE-relevance score of 10% and is therefore not included.

6. Performance statistics

The figure below shows the cumulated performance of our index and that of the other listed PE indices discussed above. The index starts in January 2008, which is the first date on which there is enough news coverage to compute the weights.



Outperformance is not what is sought after, but in a period of strong expansion of the PE industry we would expect a well-constructed PE index to perform well, and this is indeed the case for our index. Table 5 shows statistics on the performance of several indices. We note that our index, despite being focused on a single industry is about as volatile as the S&P 500.

The Burgiss global pooled LBO index has an average arithmetic return of 11.9% (annualized) between 2008 and (September) 2021. The S&P 500 Index's average return over the same time period turns out to be nearly the same: 11.7%. Our PE index, however, averages 14.8% and the S&P Listed PE Index averages 13%. Given the smoothness of the illiquid index, the correlation at quarterly frequency are biased downwards. Yet, we find a correlation between our index and the illiquid PE one to be as high as 80% (non-tabulated). Obviously, the volatility of our liquid index is much higher than that of the illiquid index, and is also higher than that of the S&P 500 Index.

Table 5: Return Characteristics – Q1 2008 to Q3 2021 [12]

	Our PE Index	S&P Listed PE	S&P 500	S&P IG Bond	Burgiss Global LBO
Arithmetic Average	14.8%	13.0%	11.7%	5.4%	12.0%
Volatility	27%	32%	17%	6%	11%
Geometric Average	11.7%	7.9%	10.4%	5.2%	12.0%

7. Conclusion

This paper shows that it is possible to use big data and natural language processing techniques to construct indices in a cheap and systematic way. We also demonstrate that the resulting index is meaningful and requires much less ad hoc choices than the commonly used index construction approaches. Several limitations, such as language being restricted to English, can easily be solved. Other limitations such as time periods are unfortunately binding: our approach is only technologically feasible over the last 12 years. Another limitation we have dealt with but is common to other indices is the lack of liquidity of many stocks. To have realistic return statistics, we have used strong liquidity restrictions at the cost of limiting ourselves to less than 100 stocks.

The volatility of our index is lower than that of comparable indices, probably due to better diversification, but it remains much higher than that of non-traded private equity indices. Applying derivative over-layers and/or non-linear fee structures to our listed index can reduce volatility dramatically.

8. References

DaRin, Marco, and Ludovic Phalippou, 2017, The Importance of Size in Private Equity: Evidence from a Survey of Limited Partners, *Journal of Financial Intermediation* 31, 64–76.

Hoberg, Gerard, and Gordon Phillips, 2010, Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis, *Review of Financial Studies* 23, 3773–3811.

Lerner, Josh, Jason Mao, Antoinette Schoar, and Nan R Zhang, 2020, Investing Outside the Box: Evidence from Alternative Vehicles in Private Equity. NBER working papers.

Phalippou, Ludovic, 2017, *Private Equity Laid Bare* (Self Published).

[12] Arithmetic Quarterly Average Return are annualized by multiplying by four. Volatility is also computed from quarterly returns (times two to annualize).