UNIVERSITY OF LAUSANNE
FACULTÉ DES HEC

Thesis submitted for the degree

Master of Science in Finance

Modeling and Forecasting cash-flows in Private Investments

Jean-Baptiste GUILLEMIN

August 2023

Supervisors : Yannick Bigeard, Xavier Lambert and Kristoffer Jonsson
Professor : Norman Schürhoff

Acknowledgments

First of all, I would like to thank Pictet for allowing me to carry out a particularly interesting internship on a topical subject. More specifically, I would like to express my gratitude to my supervisors Yannick, Xavier and Kristoffer as their insights, knowledge, and constant support really helped me shape this thesis. I truly appreciate all the advice and encouragement they provided, week after week, throughout this journey. Also, a special thanks to David, who allowed me to access external data, and Christina from PitchBook, who gave her permission to download a large amount of data which helped a lot for this thesis. I would also like to thank Susanne, for her kindness, and Nora, whose many theoretical discussions on modeling helped to bring out new ideas.

I would also like to thank my academic thesis supervisor at the faculty, Professor Norman Schüroff, for guiding me throughout the thesis, and in particular for his help in implementing the models and finding a clustering approach.

Finally, I would like to warmly acknowledge my family, my girlfriend and my friends for their help and support throughout this internship.

Modélisation et prévision des flux de trésorerie dans les investissements privés

Résumé

Sur la base de l'un des ensembles de données les plus complets utilisés pour la recherche universitaire dans ce secteur, nous effectuons une analyse approfondie des flux de trésorerie de contribution et de distribution dans le temps et en fonction de quatre stratégies d'investissement privé différentes. Tout d'abord, nous démontrons que la variabilité des flux de trésorerie n'est pas prise en compte par l'utilisation de moyennes historiques. Nous utilisons ensuite le modèle de Yale pour modéliser les flux de trésorerie, ses paramètres étant estimés sur cet ensemble de données. Nous décomposons ensuite chacun des paramètres par millésime et démontrons l'utilité du modèle de Yale pour établir des analyses de scénarios basées sur l'évolution de ces paramètres, notamment en période de crise. Nous proposons plusieurs améliorations au modèle initial, notamment pour stresser les appels de fonds, et pour surmonter la difficulté d'interprétation du paramètre lié au taux de croissance de la VNI d'un fonds. Enfin, nous montrons que ce modèle ajusté peut être utilisé à des fins de modélisation du risque.

Mots clés— courbe en J, capital-investissement, flux de trésorerie, modélisation, investissements privés, rachat d'entreprise par effet de levier, capital-risque, immobilier, crédit privé, modèle de Yale, analyse de scénario, test de résistance

Modeling and Forecasting cash-flows in Private Investments

Abstract

Based on one of the most comprehensive datasets used for academic research in this sector, we carry out an in-depth analysis of contribution and distribution cash-flows over time and according to four different private investment strategies. First, we demonstrate that the variability of cash-flows is not captured by the use of historical averages. We then use the Yale model to model cash-flows, its parameters being estimated on this data set. By decomposing the evolution of each parameter by vintage, we demonstrate the usefulness of the Yale model for establishing scenario analyses based on the evolution of these parameters, particularly in times of crisis. We propose several improvements to the initial model, notably to stress capital calls, and to overcome the difficulty of interpreting the parameter linked to the growth rate of a fund's net asset value (NAV). Finally, we show that this adjusted model can be used for risk modeling purposes.

Keywords— J-curve, private equity, cash-flows modelling, private investments, leverage buyout, venture capital, real estate, private credit, forecasting, Yale model, scenario analysis, stress testing

Contents

1 Introduction

Private markets investments represent all investment opportunities that are not available on the traditional exchanges markets and public markets. Generally, the time horizon required to invest in them is longer and because of illiquidity and less frequent fair market valuation (usually quarterly) , theses assets classes are less exposed to short-term volatility found in the public markets. Nonetheless, it is still exposed to cyclicality, as we will show later in this paper.

The interest of private investments in the last 10 years has increased tremendously, going from less than \$600bn raised in 2012 to \$1401bn in 2022, with the record breached in 2021 with more than \$1680bn raised capital (Figure 1). The majority of these investments are related to the equity part of the capital structure, mainly private equity strategy (that is often synonymous with leveraged buyout) and venture capital (focused on early investing in promising startups). The third place is shared depending on years by private debt and real assets (especially real estate and infrastructure). The rest is composed of fund-of-funds, co-investments and secondaries markets ; these different ways of investing in PM are discussed the following section. The "other" category include all funds that cannot be assigned to either of these definitions [1].

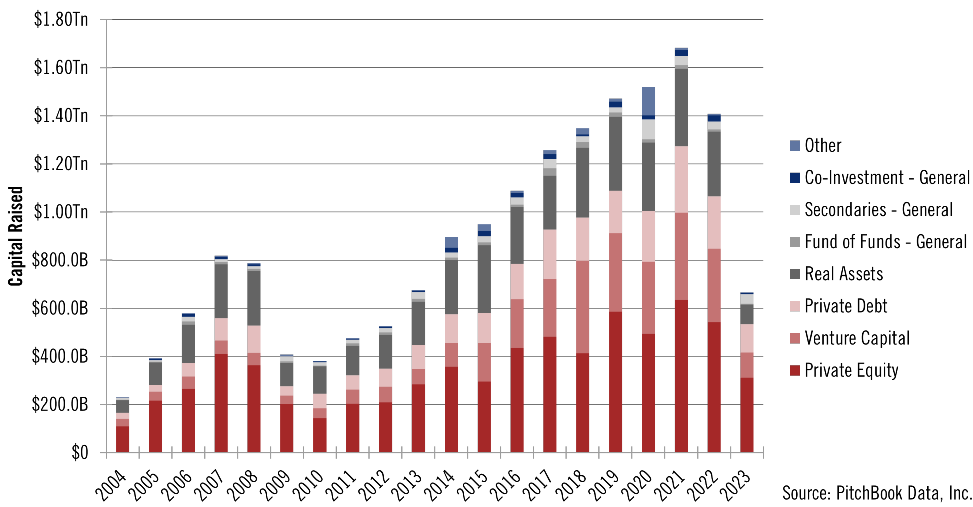


Figure 1 – Fundraising in private markets by strategy

1.1 Typical investments structures

Investing directly in private companies requires in-depth knowledge not only of the company itself, but also of the law applicable to buyouts, and of the structure and resources needed to support the growth of these private companies. Few institutions have the means to do this, and so seek out intermediaries with these capabilities. Investments are made through the limited partnership structure, which according to Bosut (2003)[1], *is the most ideal financial fund management structure, avoiding potential conflicts of interest between fund managers and limited partners, and aligning the incentives of the parties with each other* [2].

1. Such as forestry, energy, farmland, commodities...
2. Quoted in Mathonet and Meyer (2005)

In a company structured with limited partnership agreement, the General Partner (GP) has the management control and is liable for the company's debt ; the Limited Partners (LP) relinquish their right to manage the company and are not liable for the debts of the partnership. The setup of these funds is also influenced by regulatory and tax requirements, with the goal of ensuring transparency (meaning investors are treated as if they're directly investing in the actual portfolio companies) and keeping taxation at a minimum [2].

In a typical private equity fund, the LP will provide the capital needed to invest in the companies, and only the GP will manage it. The terms and conditions are defined in the LP agreement, but usually, the funds will have the following structure (see [3] and [2]) :

— The fund will have a contractual life between 7-10 years, with often a provision for an extension of 2-3 years, after which the partnership (and the fund) is terminated. When the underlying companies are sold, the realization are distributed to investors as soon as possible ; the fund is thus self-liquidating.
— The management fees collected by the GP are high in this industry, the standard being the 2-20 model : a 2% management fee combined with a 20% performance fee if performance exceeds a 'hurdle rate', usually 8%. They can be quite hard to negotiate, given the favorable supply-and-demand situation for the fund managers [4]. The LP sometimes requires the GP to invest its own money in order to align its interests with its own and provide an additional incentive.
— LP's commitments are drawn down only when needed, (i.e. when the GP makes investments) ; the GP does not retain a large pool of non-invested capital. It is therefore important for the LP to manage the timing of these capital calls, so as to be able to respond to them without leaving them in a current account with the consequent opportunity cost of not investing them.

These funds then deploy their capital depending on their strategies, like leveraged buyout of companies (LBO), venture capital (VC), private credit/debt (PC), real estate (RE), infrastructure and other investments in real assets. In this thesis, we focused on 4 main strategies : LBO, VC, PC and RE. To see their definitions, see appendix 7.2.

Another way of investing in private markets is through the fund-of-funds (FoF) structure. Also incorporated as a limited partnership, the FoF pools investors and uses the capital to build up a diversified portfolio of private equity funds. These funds have potentially complementary styles, thus limiting the downturns. This solution is particularly beneficial for small pension funds, endowments, family offices, UHNW individuals, and more broadly, for LPs with no in-house program. However, it comes with a trade-off: a new layer of fees. These fees encompass the cost of manager selection for the FoF, as well as additional portfolio managing activities such as portfolio construction, allocation, and monitoring. On the other hand, if the alternative under consideration is an in-house program, then the FoF structure can, in fact, cheaper due to economies of scale [2].

An LP can also invest in private companies, usually backed by the private equity funds where they participate; in this case, it is called a co-investment. This is in contrast with an investor (like an LP) making direct investments on their own. The difference between co-investing and direct investing is that the co-investment opportunities are already pre-screened and structured by the GP. The financing cycle is then syndicated between the private equity fund and the co-investor, providing the latter with greater upside potential. Due to the major difference of cash-flows with primary funds (only one company to model), we will discuss this point in more detail later in the section 3.5.6.

Finally, the secondary market is a way to fast-pace an allocation to private markets by buying an existing interest or asset from primaries funds LP (primary funds refer to the initial round of capital collection from investors). The seller may want to liquidate or rebalance an existing portfolio, whereas the buyer would be interested in seeing shorter duration and faster return of capital (adding potential discounts on the sale, if the seller is in a hurry). The private secondary market is not part of the thesis's scope, even though it has seen some growth in parallel of the primary private markets.

Figure 2 explains the different ways of investing in private equity. This structure is also relevant for the other main strategies found in private investments, as long as the private fund is structured as a limited partnership.

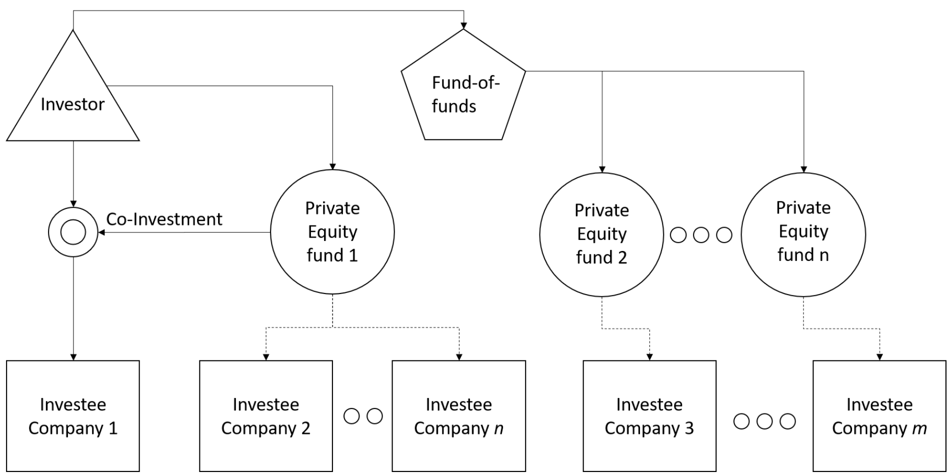


Figure 2 – Alternatives approaches to investing in private companies.

1.2 J-Curve

The private equity J-curve is a pivotal concept within the realm of alternative investments. This phenomenon is notably characterized by an initial period of negative returns followed by an eventual upward trajectory, forming the shape of a "J". The J-curve effect stems from the dynamics inherent to the private equity investment cycle.

The first period in the fund's life cycle is the fundraising period, which lasts 1-2 years and during which LPs commit capital that is then deployed in a variety of businesses during the investment period. This investment period typically lasts 4-5 years, during which new opportunities are identified and invested. It is only when the GP requests cash from its LPs that the binding capital commitment translates into an actual funding obligation. This cash movement from LPs to GP is generally referred to as a "capital call", "drawdowns" or "contributions". The rate of contribution will depend on the fund's strategy, on the GP itself and on macro-economic conditions ; the uncertainty of this pattern (from a LP perspective) is one of the focus of this thesis, especially because these drawdowns are law-binding and any LP that fails to meet its contractual obligations is liable to severe penalties. [3].

The final period of the fund cycle is the harvesting period, where the underlying companies are sold and the proceeds are distributed to the LPs (pro-rata of their commitment). These returns come mostly in the second half of the fund's lifetime, and the fund is terminated when the last realization is distributed. Distributions follow a structure resembling a waterfall : first, LP are returned their investment + fees/expenses, then they attain the preferred return defined in the LP agreement (hurdle rate). This can be followed by a phase known as the "catch-up period". During this phase, the General Partner (GP) receives most, if not all, of the distributions [4]. Finally, when the agreed carried interest split is reached, distributions are shared between the GP and LP according to the terms of the contract (usually 20% and 80%, respectively).

Due to the time required for portfolio companies to mature and generate returns, an apparent lag materializes between capital contributions and tangible profits. Consequently, this temporal misalignment often leads to an early-stage trough in fund performance. As investments mature and exits occur during the realization period, positive returns begin to outpace the initial negative contributions, ultimately resulting in the upward curve of the J-curve (see fig 3). This phenomenon is less pronounced for quicker strategies like PC, and more accentuated for VC strategies that takes longer to distribute.

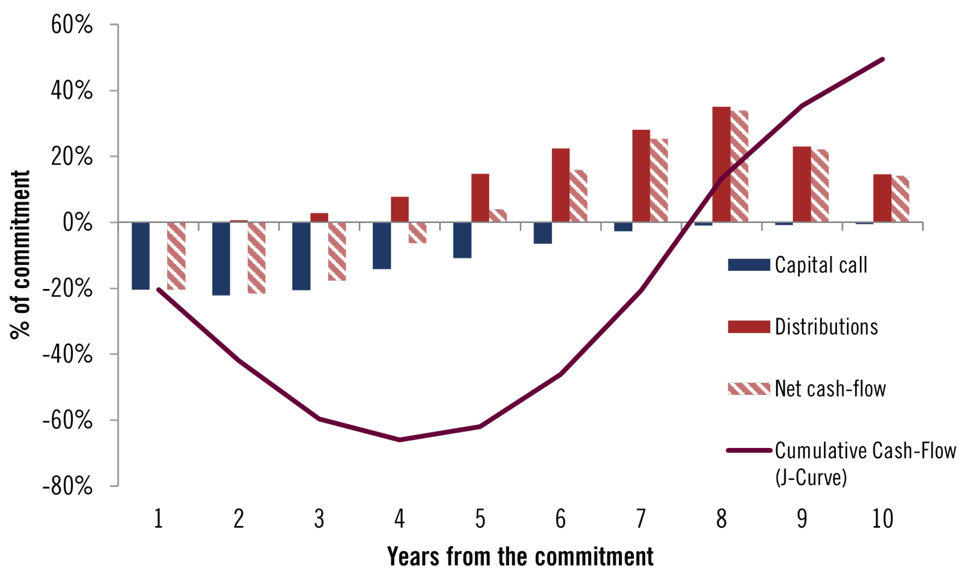


Figure 3 – J-curve example

The comprehension of this J-curve phenomenon assumes paramount significance for LPs, shaping their strategic decision-making, risk assessment, and portfolio management strategies in the realm of private equity investments.

3. Penalties can range from a penal rate of interest, being forced to sell its position on the fund on the secondary markets (potentially at a deep discount), or simply being excluded from committing to future funds raised by the GP. For more information, see [5]

4. This mechanism is not normalized. Sometimes there is no catch-up period, sometimes there is a partial or 10% catch-up ; in this case, the LP receives no money until the GP recovers its entire investment [3].

1.3 The Institution

This thesis was written during a 6-month internship at Pictet, a Swiss financial institution operating in several financial fields, including private markets. More precisely, I worked within Pictet Wealth Management, and the Pictet Investment Office, its unit dedicated to long-term investing with an endowment-style approach [5]; the subject of the thesis was thus of particular interest to them, as LP. Of course, for reasons of confidentiality, all proprietary and historical investment data is anonymous. In the rest of the document, Pictet will be referred as "the Institution".

1.4 Problem description

Building a portfolio of private investments offers distinct challenges compared to traditional public investments. Firstly, liquidity and financing risks are prevalent. Capital pledged to General Partners isn't immediately utilized; instead, it's called upon when managers identify a suitable company for investment. Thus, managing unfunded capital becomes pivotal in an investment strategy : such a strategy aims not only to minimize dormant capital but also to effectively address capital calls without straining liquidity. Additionally, distributions are often unpredictable, both in timing and amount. This unpredictability can lead to significant cash flow uncertainties, particularly during economic downturns when the inherently illiquid nature of these investments might hinder their conversion into cash. Furthermore, private investments typically involve extended holding periods, which can lock in capital, potentially leading to missed opportunities in other arenas.

Modeling and forecasting the cash-flows (and by extension the J-curve) is therefore essential to a good understanding of this sector, particularly for building a targeted allocation in private markets.

1.5 Objectives

The Institution wanted to review its capabilities in private equity cash flow modeling, first in private equity and then in real estate and private credit. At the core of private investment cash flow modelling lies the definition and methodology to estimate future capital calls, distributions and NAVs.

The objectives were stated as follows :

- The first step was to review the current methodology, based mainly on historical averages. We would then analyse the dispersion of cash-flows between vintages and different macro-economical conditions.
- Then, we would to compare and contrast methodology of the existing tools and data to the different academic models proposed during the years.
- Once the best model according to our needs was selected, our aim was to adapt this model to historical data, to analyse the results, limitations, and propose solutions to address them.

5. The endowment style is defined as an investment strategy emphasizing diversification across multiple asset classes, including significant allocations to alternative assets like private equity and hedge funds, pioneered by David Swensen at Yale University's endowment fund. It focuses on long-term investments, active management, and regular rebalancing to achieve optimal returns.

— Afterwards, we would create stress tests scenarios based on historical data in the event of unexpected outlook to stress test cash flows needs. We would also allow the system to create different capital call, distributions, and NAV patterns from buy-out, venture, private debt and real estate strategies if there are significant differences.
— If possible, we would also consider a model to forecast using randomness in distribution, capital call and NAV patterns. Specific patterns for co-investments and exceptions should also be able to be implemented.

At the end of the thesis, a pilot tool would to be proposed.

2 Literature review

In recent decades, the interest in private markets (PE, VC, PC, RE) has gained momentum thanks to the strong returns of this asset class and of its investors (endowments, pensions funds, UHNW individuals, and private banks!), illustrated both in the number of assets invested (Fig 1) and in the number of publications on this sector as shown by Cumming & all [6]. This particular study uses a bibliometric approach to propose a synthesis of PE and VC research areas. They show that particularly for PE, early studies focused on corporate finance mechanisms (buyout/privatization, IPO as illustrated by Kaplan and Strömberg (2009) [7]), but also on performance measures. There are many studies on PE performance, the best known being that of Kaplan and Schoar (2005) [8], but they are not consistent in their conclusion due to the difference in data used (see the study by Korteweg (2019) [9] on this subject).

All of the returns in private equity investments are based on the actual cash-flows, capital calls and distributions. Cash-flow modeling is thus crucial in this sector, and it has been the subject of numerous academical and professional publications.

The first approaches were either heuristic, based on the LP experience or simple rules of thumb. Then, the utilization of historical data to decipher cash-flow patterns soon emerged as a popular tactic. Some practitioners sought to calculate average patterns based on existing cash-flow data, while others leaned towards constructing more refined shape functions, like Weibull distributions [10]. These approaches, although robust, often did not offer insights into the variations around the average. The reliance on historical averages soon transitioned into an industry standard, endorsed and employed now by heavyweights like PitchBook, CEPRES, and Preqin (among others).

In the historical literature, the study of Ljungqvist and Richardson's (2003) is particularly noteworthy in this evolution [11]. Their pioneering work shed light on the nuanced dynamics of capital calls and distributions, providing insights into the different ways of doing things of fund managers, particularly in response to market conditions (for example, when facing greater competition from other private equity funds, fund managers draw down their capital more slowly and hold their investments for longer periods of time).

Robinson and Sensoy further augmented this body of work in 2011 and 2015. Their research not only deepened the understanding of PE cash flow behaviors but also introduced a more nuanced view of how these flows react during economic crises [12, 13]. This sensitivity of PE cash flows to external market conditions is of the utmost importance, as it underscores the necessity for models that can adapt to external shocks and provide accurate forecasts for cash-flow dynamics.

Apart from using historical averages, several models have been proposed during time. The first model allowing to make scenario analysis was proposed by Takahashi & Alexander in 2001 [14]. This model, also called the Yale model because its inventors were part of the Yale University Investment Office, is a deterministic model : one set of inputs will give one output, i.e. capital call, distributions and NAV patterns. This single output is one of the main limitations of the model, along with the joint modeling of contributions and distributions that complicates parameter estimation. Despite being the earliest solution to this modeling challenge, it remains the most esteemed in the sector and often used as benchmark for new approaches, as evidenced by the extensive literature that has developed around it over the years.

For example, researches paper by Jeet and O'Shea (2020) [15] [16] showed their method of estimating the parameters of the Yale model, fitted on the Burgiss data. It is, at the author's knowing, the only study that looks at the evolution of parameters by vintage. However, the exclusivity of one year of historical data might compromise the robustness of these parameters, and especially the 'bow', a crucial component reflecting the convexity of cash flow patterns.

The Yale model wasn't free from scrutiny or the need for enhancement, as several studies exemplified this evolution. Hoek (2007) [17] used the Yale model and proposed improvements, by setting the distribution rate as stochastic and assuming that the return on Private Equity depends on the return on Public Equity, with some delay ; this would change then the growth rate parameter (set on the NAV). As for him, Tolkamp (2009) [18] used it for predicting private equity performance.

Another paper by V. Jeet (2020) [19] proposed that the model's fixed growth parameter is replaced with period-specific growth parameters. This new parameter is estimated on the public market thanks to a lagged regression, a bit like [17]. However, their proposition suffers from large difference between modelled and realized distributions from vintage 2006 onward (in the latter half of lifespan) that could be due to multiple factors including modeling limitations, a change (due to the financial crisis of 2008) in the bow factor, lifespan, or even RC parameters. We will show later in this document how did the bow did evolve before and during the GFC.

More recently, Kieffer et al.'s (2023) foray into the deterministic modeling scene offers another layer of sophistication to the Yale model [20]. Termed the "Yale Plus", this model adjusts cash flows according to frequency and volatility, spanning a range from semi-annual to monthly periodicities. Beyond mere cash-flows predictions, their approach provides actionable insights into recommitment strategies, solidifying the bridge between theory and real-world financial decision-making.

In order to address the issues with the deterministic mode, several studies have been proposed. The first one was de Malherbe (2005) [21], that ventured further into this realm by providing a detailed examination of NAV's role in these forecasts. By highlighting the challenges rooted in the NAV's potential inconsistencies, de Malherbe advocated for a model encompassing diverse parameters, such as fund maturity, vintage year, and investment stage. It relies on the specification of the dynamics of an unobservable fund value and therefore has to account for an inaccurate fund valuation by incorporating an error term. This multi-faceted approach pioneered a more nuanced perspective in cash-flow projections, even though we have not found any practical use of the model in the industry. His work nonetheless foreshadowed the pivot to stochastic approaches, recognizing the limitations of deterministic frameworks, especially in the face of unexpected market anomalies.

Buchner et al.'s (2009) work epitomized this shift [22]. They proposed a continuous-time stochastic model with two independent components that solely relies on observable cash flow data. As the model is time dependant, it can account for the importance of understanding the life-cycle dynamics of private equity investments. Based on the systematic variations over a fund's life cycle, Buchner's approach offers a comprehensive method to capture both the deterministic and random elements of private equity cash flows. This model would be the best suited for modelling risk, as it allows a wide range of possible patterns to be given, in terms of contributions or distributions. A good master's thesis on this model and parameters estimation was written by Ungsgård (2020) [23], but he only had 20 funds to estimate the parameters, so the conclusions are inherently limited.

In 2017, Buchner also extended his model for risk measure in private equity [24], the third point of the study being particularly interesting with a notion called cash-flow-at-risk (CFaR) measure. This measure is defined as the change (loss) in the investor's cash position, which is only exceeded with some given probability, over a given time horizon. In this paper we did not touch risk measures, but this notion can be further explored within the Monte Carlo analysis that we do in section 5.3.

As these models are approaching the problem from different angles, Furenstam and Forsell (2018) [25] studied on the comparison of the performance of the deterministic model and the stochastic model with an empirical analysis on private equity funds. They found that it is hard to make an absolute conclusion on the performance of the two models as they outperform each other on different periods ; their sample is also too small to be exhaustive. In any case, as both model are fitted on historical values, the forecast will always depend on past data ; we'll look in more detail at how these two approaches differ in section 3.2.5.

Another master's thesis on a comparison between the Yale model and stochastic model (this time for private credit) has been done by Virtanen (2021) [26] and he concluded that the stochastic model was better according to backtesting, but as the sample is very limited (only 62 funds) it is difficult to extract meaningful conclusions.

The last few years have been marked by the integration of advanced computing and data analysis capabilities into the industry: cash flow modeling for private equity funds is no exception. Karatas et al. (2021) brought forth an innovative approach, employing supervised neural networks (SNN) to model PE cash flows [27]. Their use of the Yale model as a benchmark demonstrated the adaptability of classic models in the face of modern computational techniques. Due to the complexity of this approach, we did not go in this direction; it is however interesting to note that they have quarterly data in their SNN, and they adjust the Yale model equations accordingly, proving that it is possible to use the deterministic model even with quarterly data.

However, not all modern advancements relied on intricate computational methods. The structural model proposed by Tausch et al. (2022) took a different path [28]. The idea is to predict the final fund/deal performance as precisely as possible and then reverse engineer the intermediate variable paths. This is interesting for GPs and undiversified investors that have the underlying companies information, but not for the LP as he doesn't have this kind of information, only what the GP provides him (from the LP's *point de vue*, the fund is effectively a blind pool).

In this paper, we will start from the beginning with the most simple method : using averages pattern computed on historical data. Then, we will look upon all of the aforementioned alternative to cash-flow modeling, and select the best one according to our needs.

3 Materials and Methods

This section describes first the data used, its cleaning and its limitations. We then begin the analysis of this data by computing historical averages of cash-flow patterns, and discuss the inherent limitations of using averages; by construction, they do not capture the dispersion across vintages that we show using the data. Next, we discuss the existing models presented in section 2 and choose the Yale model as reference. We introduce this model in detail, and present our way of estimating each parameter. We finally propose some improvements to overcome the model's original shortcomings, particularly for co-investment modeling and having a robust method to stress the capital calls.

3.1 Data

Private equity, and by extension private markets, is a sector notorious for its scarcity of data: In the USA, private funds are exempt from the reporting requirements of the Securities Acts of 1933 and 1934 and the Investments Company Act of 1940 [2]. However, several databases exist. They are based on information directly provided by the GP, regulatory disclosures from the LP (like pensions funds), IPO prospectuses (for the companies that go public), filings with the SEC or other public (but often difficult to access) sources. The most well known databases here are Burgiss, Capital IQ, Preqin, Cambridge Associates, CEPRES, and PitchBook (we can also mention the Private Capital Research Institute, for academics). Even though not up to date and with a focus on venture capital, Kaplan and Lerner (2015) [29] have a good comparison of the different databases, especially differences between data recovery. The data in our study comes from PitchBook; you can assume that all data here come from PitchBook or it will be explicitly stated otherwise. In our research of external databases, we found that only PitchBook and Preqin give access to data at the fund level; this was necessary to compute ourselves the relevant patterns and parameters.

3.1.1 Dataset description and general limitation

Compared to previous studies that had access to external data (Buchner, Kaserer & Wagner - 777 funds [22], Furenstam & Foresell - 195 funds [25] or Karatas, Klinkert & Hirsa - 606 funds [27]), we have access to many more funds, as summarized in the table 1 below.

With this kind of dataset, we can push the analysis to many more categories of private investments if needed.

Table 1 – Number of Funds with returns data in the PitchBook Database

Access Point							
Primaries	12828	Funds-of-funds	1644	Co-Investments	592	Secondaries	529
Fund Type within Primaries Funds							
Tier 1		*Tier 2*		*Tier 1*		*Tier 2*	
PE	4995	Buyout	3644	VC	3270	VC general	2349
		Diversified PE	27			Angel Fund	12
		Mezzanine	417			VC Early	770
		Restructuring/Turnaround	30			VC Later	139
Private Debt	1384	Debt general	62	Real Estate	2143	RE general	15
		Direct lending	450			RE Core	175
		Credit special situations	173			RE Core Plus	105
		Distressed debt	357			RE Value Added	870
		RE debt	282			RE Opportunistic	852
		Infrastructure debt	23			RE Distressed	126
		Venture debt	32	Real Assets	491		
		Bridge financing	5	Infrastructure	535		

Note: Data retrieved as of 31.05.2023.

The data is presented as such: for each fund, we have their vintage, the fund size and their location. We also have the cash flows as of period: one limitation here is that the data is yearly. Quarterly data would have been useful to obtain greater granularity in stressful years, or to build precise benchmarks; but for our purposes this was not crucial, since the existing patterns were annual anyway.

PitchBook uses LPs information when the GP disclosed nothing, and then extrapolate the result of the one (or several) LPs that reported to obtain the total cash-flow of the fund, given the pro-rata commitment. But sometimes the GP will report information, and this information will take precedence over the LP and thus will be the data reported for the year in question. However, this data can be quite different and leads to incoherent results. The cash-flow information in the dataset is cumulative, so for example if the total capital reported by the GP in year t was lower than the total capital called reported by the LP in t-1, the capital call in year t will be negative. This can be explained by recallable capital calls, when the GP was not able to find suitable investments opportunities; instead of holding onto this uninvested capital for too long, some funds may have clauses in their agreements that allow for the distribution of this uninvested capital back to the LPs. However, there's no way to tell from the data whether these are indeed uninvested capital calls or simply bad reporting.

Finally, as exposed by Kaplan and Lerner [29], different databases have different ways of defining the vintage, that makes them difficult. PitchBook defines the vintage as: *Vintage Year is the year in which a fund makes its first investment by delivering capital to a project or company. When we cannot confirm the year of first investment, the year of the fund's final close is used as the vintage year*. Some years can show high discrepancies between databases, especially with low amount of funds raised and investments (like 2009-2010 for LBOs). Even in the evolving world of private equity, accessing accurate and comprehensive data remains a formidable challenge, widely recognized as one of the sector's most significant limitations. However, as we progressed in this field, we managed to access data that turned out to be more comprehensive than historical accounts had often presented. This is also good news for research, as it often limited in the past the interpretation of study results in this sector.

3.1.2 Data cleaning

At the beginning we downloaded only funds with returns data, particularly capital calls/distributions. Note that the Institution wasn't interested in LBOs funds with a size below 100m [6] as they invest in larger vehicles, so we have not selected them. Then, we delete the duplicates, remove all information in 2023 (because we only have Q1 2023 information and we use yearly data), and begin the data cleaning.

Initially, we observe that cash-flow information is not reported by LPs of some funds for many quarters, leading to gaps in the cash-flows patterns that may distort the data quality. Therefore, funds are removed from the dataset if at least one of the cash-flows (capital calls and distributions, NAV doesn't matter here because its not relevant to the patterns) have 30% or more missing values. Figure 4 illustrates the distribution of the number of funds in our dataset aligned with their vintage year before and after removing funds based on their missing values. Furthermore, we also trim away all funds with vintage before 1996 because the data quality in funds before this date is poor, rendering it unsuitable for meaningful analysis.

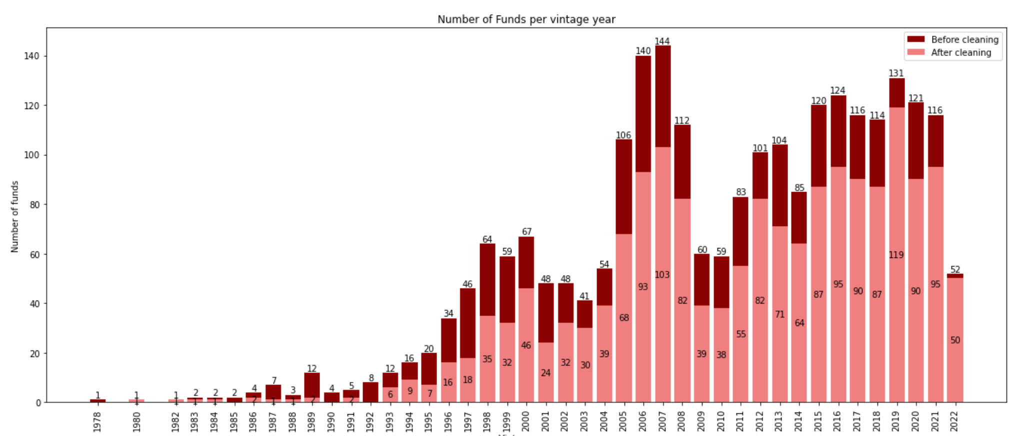


Figure 4 – Number of funds by vintage after cleaning

The total number of the funds in our dataset reduces to 1643 (for buyout funds) after this step. The data cleaning for all others strategies is described in the appendix 7.

In order to fill the missing cash-flows values, we considered different methods of interpolation: linear interpolation, piece-wise polynomial interpolation, or simply filling the missing values with the last value reported.
Like the study [27], we found a cublic spline interpolation to give the best result [7]. Figure 5 presents a fund

6. For other categories, we selected only funds with a size above 50m.
7. See https://medium.com/eatpredlove/natural-cubic-splines-implementation-with-python-edf68feb57aa for detailled application in python.

with missing cash flows and a period where the GP reported something very different from the LP, illustrating both possible problems found in the dataset. The plots shows that the interpolated values for missing cash-flows do not distort the expected patterns of contributions, distributions, and net asset valuations.

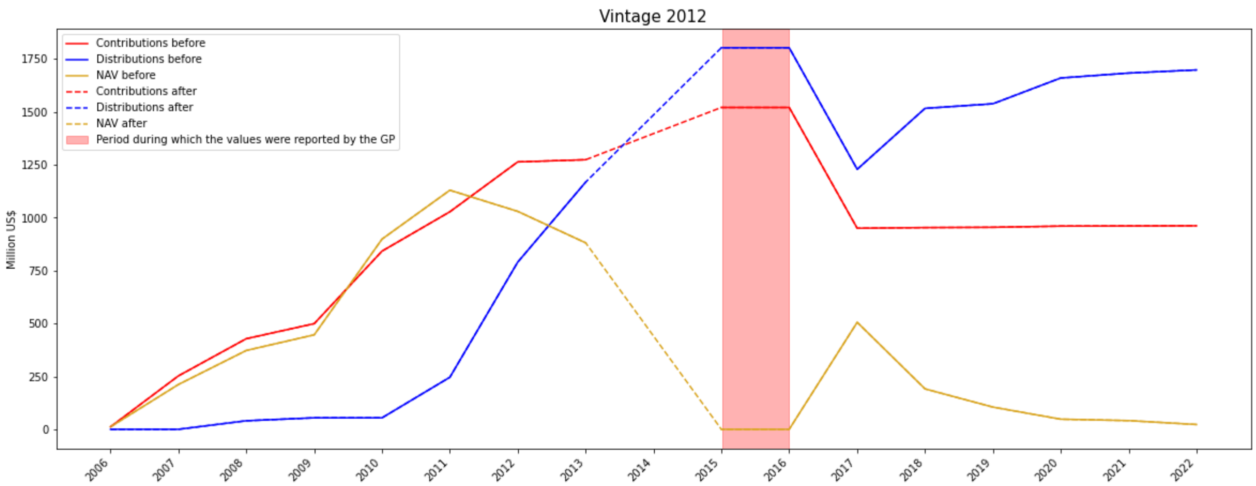


Figure 5 – Example of a fund with interpolation - vintage 2012

3.2 Averages patterns, results and discussion

Once the data had been cleaned, we were able to begin the analysis. The first step of the internship was to determine if the methodology and patterns in use were coherent with the PitchBook database. We selected at first only the dataset of PitchBook buyout funds (N = 1643 after cleaning). The cash-flow data is cumulative, meaning that to compute the cash-flows per year, we need to take the difference between consecutive values: CF_t = Cumulative CF_t - Cumulative CF_{t-1} with $CF_t = CC_t$ or D_t

Then, the process is fairly simple: we make every fund begin at the same time (by resetting their index, effectively normalizing each fund data at a time t > 0), and for each fund$_i$ we compute the coefficient as :

$$\text{CF coefficient} = \frac{\text{CC}_t \text{ or } \text{D}_t}{\text{Fund Size}} \tag{1}$$

In the rest of the thesis, capital call or distributions coefficients is always defined this way, either with respect to the commitment or with the fund size.

We obtain a set of coefficient across time, that we limit to fifteen years for visualization. The scatter plot and boxplot result is shown in figure 6. Whiskers in the box plots correspond to the 1% and 99% percentiles, respectively, and the orange line to the boxplot median.

We raised the issue of recallable capital calls in section 3.1.1, that is here exposed by negative capital calls in the scatter plot. At the same time, distributions also have a lot of visible outliers with negative distributions. This time it can be explained by recallable distributions (also called recycling) that occurs when the GP distributes funds to the LPs and subsequently requests that a portion or the entire distribution be returned to the fund at a later date. For example, the GP identifies a new investment opportunity that requires additional

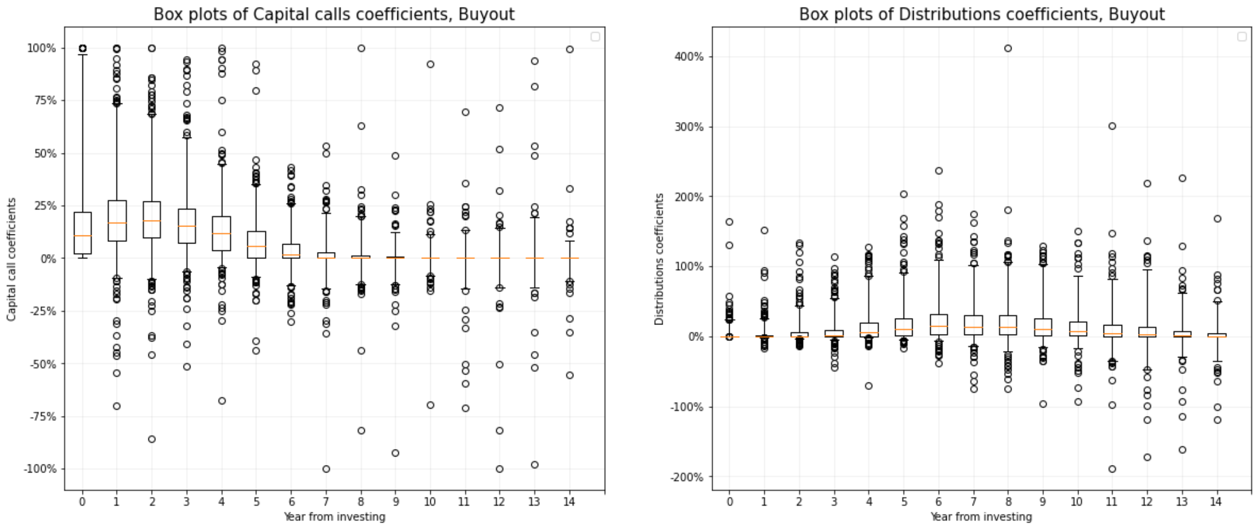


Figure 6 – Capital calls and distributions box plots

capital, but the fund has limited available cash at that moment. In this case, the GP may recall a portion of the earlier distributed capital from the LPs to invest in the new opportunity. The terms and conditions of recallable distributions are typically outlined in the limited partnership agreement. For the LP, they offer on one hand the potential for additional returns if the recalled funds are invested profitably; on the other hand, they may create uncertainty for LPs, as the distributed funds may need to be returned, affecting their cash-flow planning.

If we consider that all theses points have an economic interpretation, then we must keep them in the average calculation. However, given the lack of verification as to whether or not these coefficients are justifiable, we have decided to reduce their effect by applying an outlier "treatment". This allows to compute more robust average patterns.

The most basic method for reducing the effect of outliers would be to use the median instead of the mean, but we have found that using the median does not give satisfactory coefficients. For example, using the median on the pattern of capital calls, we obtain only 85% of capital calls in relation to the fund size over 12 years, whereas we should have 100% (or almost) ; the average gives better results in that regard.

3.2.1 Winsorization

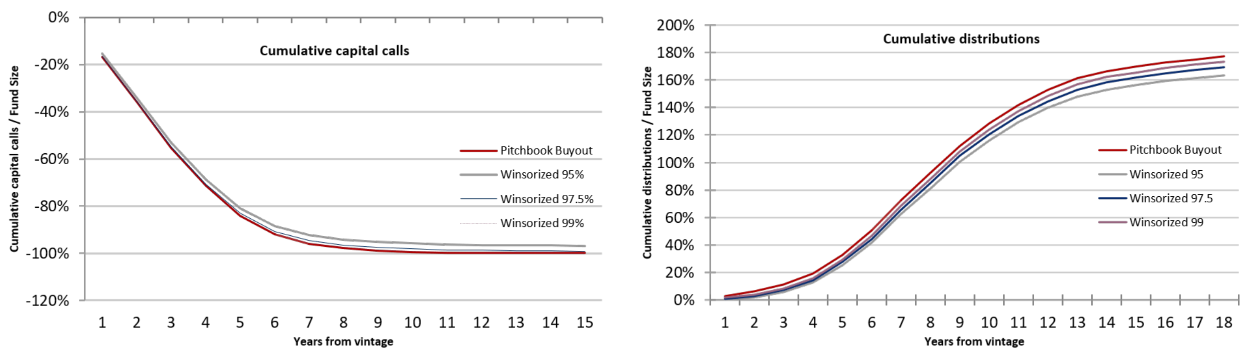
Instead of completely trimming away the outliers, we decided to keep them but in a way that would reduce their effect on the average computation. We selected the winsorization method to this end. In mathematical terms, winsorization is the transformation of outliers above (or below) a pre-specified threshold to the value at the threshold [30]. For example, see table 2.

Winsorization is thus an effective method of lessening the extreme effects of outliers and achieving a less skewed distribution. Question is now, what threshold to use ? Below are the difference between threshold at

Table 2 – Example of winsorized data

Percentile	Original data	Winsorized data (95% percentile)
95th	50	50
...	57	50
...	61	50
97th	120	50

1%, 2.5% and 5%. for the capital calls coefficients [8].



(a) Cumulative capital calls - buyout

(b) Cumulative distributions - buyout

Figure 7 – Winsorized capital calls and distributions

If we compare it, we can see that there is a very high difference between no winsorization and winsorization at 95%, especially with distributions: 15 % less total distributions! This difference is even more pronounced for strategies like VC, where a very small amount of the funds account for the majority of performance. In our case, we selected a threshold at 1% to keep the maximum of funds; for a dataset with 1646 funds, it means effectively winsorizing 17 funds (rounded to the next unit).

3.2.2 Results with the different strategies

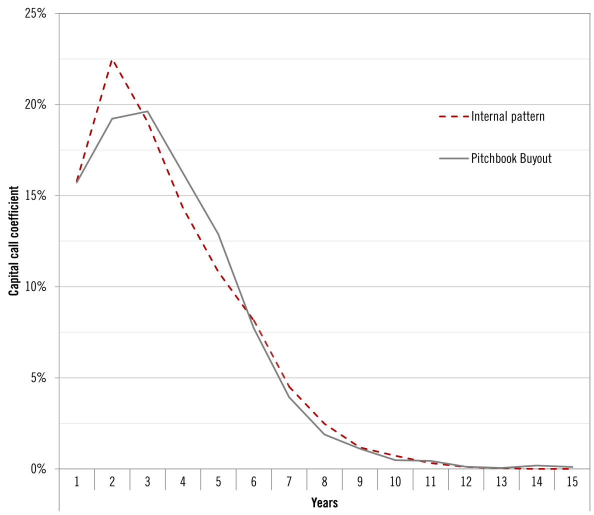
Once we have the cleaned and winsorized (at 99%) data, we could compute historical averages. The graph 8 represents the pattern of capital calls / distributions divided by fund size on the PitchBook data and what was computed by the Institution several years before. The framework used at the Institution until now was to use average capital call/distributions coefficients computed on manually selected buyout funds. The result gives an initial pattern that could be used for forecasting : it is defined here as the internal pattern.

The result is encouraging: there is not a lot of discrepancies between the 2 lines, except that the capital was called a bit sooner in the selection than the PitchBook data, and as expected distributed a bit faster [9]. We

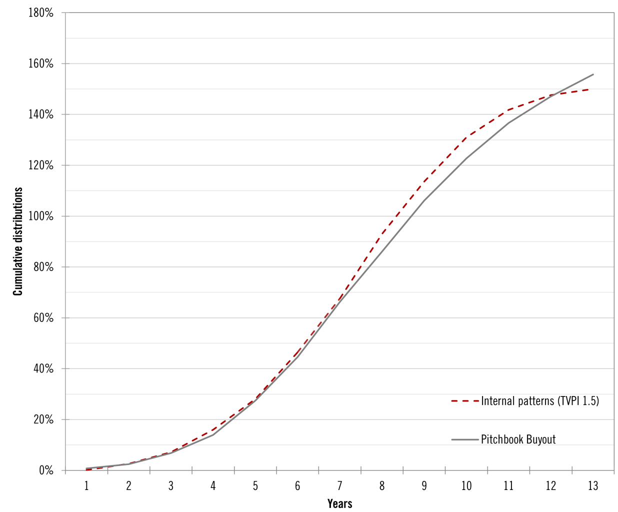
8. Cumulative capital calls coefficients are shown negative with regards to the fund size. However, for visualization purposes they will depicted as positive in the next figures.

9. Because the internal pattern of distributions is normalized at 100%, we need to multiply it by a scaling factor like 1.5 that will represent the TVPI (at the end of the fund, the TPVI is simply equal to the sum of distributions).

could then use this pattern to forecast new funds, if we consider that the past can predict the future to a certain extent.



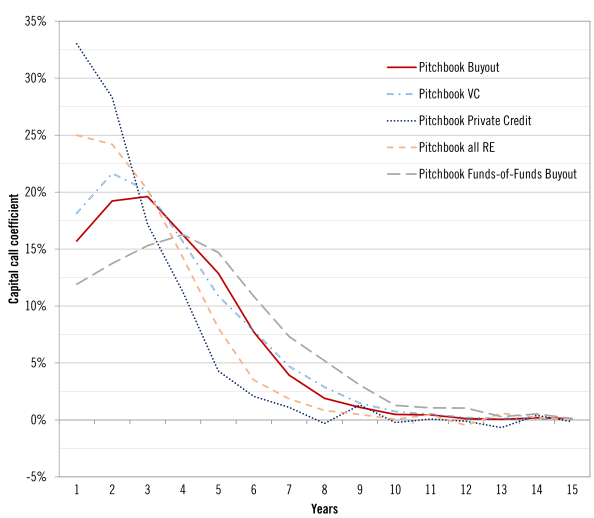
(**a**) Differences in capital call pattern



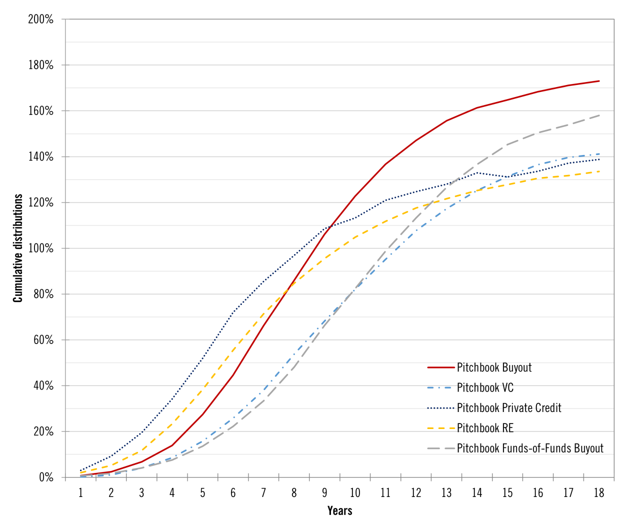
(**b**) Differences in distributions pattern.

Figure 8 – Internal vs PitchBook patterns

The methodology for different strategies is exactly the same: first we select the funds to be downloaded/treated, select only funds with enough data, interpolate if needed, then finally winsorize at 99%. The following graphs illustrates the differences between strategies, first for capital calls and then for distributions. For capital calls, we put the threshold at t = 15 years, and t = 18 for distributions to account for strategies like VC or fund-of-fund buyout that distributed later.



(**a**) Capital calls patterns



(**b**) Cumulative distributions patterns

Figure 9 – Capital calls and distributions patterns by strategy - PitchBook data

As expected, different strategies have different cash-flows patterns that ultimately depends on how quickly the investment are put to work. This is very quick in private credit, where typically the opportunities are selected before the opening of the fund; real estate calls a bit more quickly than the equities strategies that have later drawdowns. Fund-of-fund buyout calls the latest as it should, because of the lag due to the structure of fund-of-fund. VC is a bit faster to call capital than Buyout strategy, and should have the same pattern of distributions; but as showed in the right graph this assumption is not true.

For readability purposes, the distributions are cumulative. Here, the differences between strategies are even clearer. As expected, private credit distributes the fastest because of the short term horizon of the underlying investments. Real Estate follows, with surprisingly less total distribution at the end of the period than private credit, even though the risk profile is different. What is really unexpected here is the huge difference between Buyout and VC, the latter not only distributing the slowest but also the lowest of the 2 equities strategies. It can be explained by the *home-run or bust* model of VC, where the big returns are concentrated in several funds that managed to invest in the few startups that will emerge later as revolutionary (Google, Uber, Facebook...). Venture capital answers to its own particularities, that are outside the scope of the thesis; it is nonetheless important to contrast this asset class to the others before investing in private assets. At last, the FoF buyout distributes the latest and lower than buyout strategies; this difference could be explained by the additional fees and the dilution of capital between too many layers of investment.

3.2.3 Equal-weighted patterns vs asset-weighted patterns

An argument could be made with regards to the computation of the average. In the private equity world, the biggest companies have the biggest funds and thus attract the highest number of LPs. This is particularly true in 2023, as illustrated by figure 10 which shows that the median size of buyout funds has more than doubled from 2022, indicating that the largest and most respected funds capture the majority of capital raised.

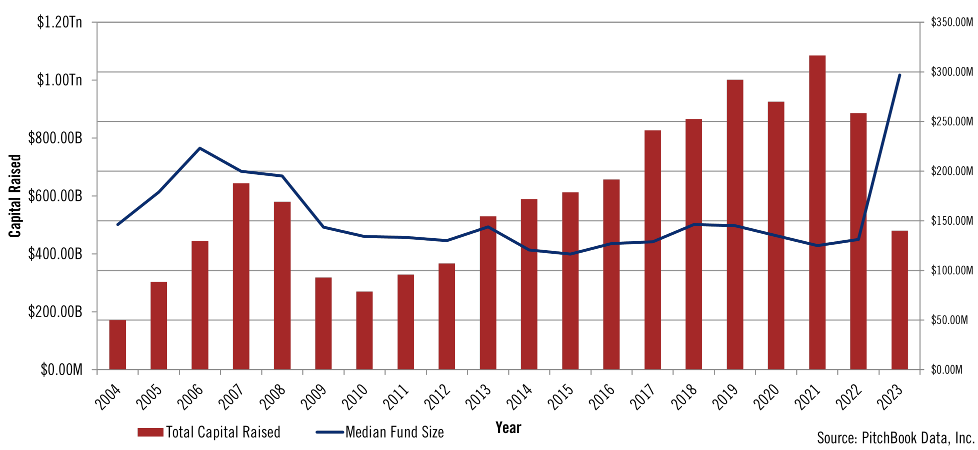


Figure 10 – Total Fundraising for buyout funds, and median size

Thus, computing the pattern using a simple average may not be the best idea. To overcome this problem, we introduce the notion of a fund-sized weighted pattern, very similar to the market capitalization weighted index in public markets. Basically, we have the fund size information for each fund. We compute a fund size coefficient as:

$$\text{Fund size coefficient}_i = \frac{\text{Fund Size}_i}{\sum_0^i \text{Fund Size}_i} \tag{2}$$

Because the number of fund data varies from year to year (see table 8), we must re-compute this coefficient each year. [10]
Then, for each fund, we simply multiply its capital call or distribution coefficient by its fund size coefficient, and sum all the results every year. Note that we use the winsorized pattern of coefficients. The result is presented below (for buyout funds):

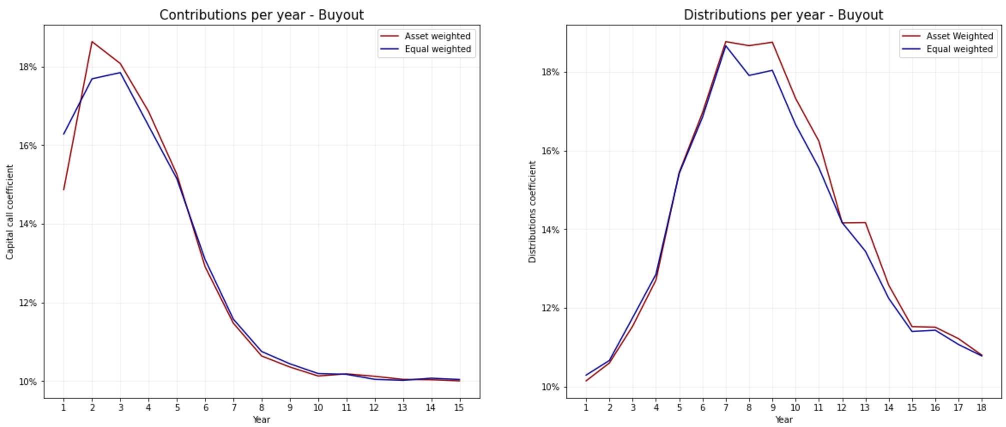


Figure 11 – Asset weighted and equal weighted patterns comparison - Buyout funds, PitchBook data

The difference exists. Surprisingly, the big difference in capital calls comes in the first 2 years, where the asset weighted pattern will have far less capital calls in year one and more in year 2, the rest of the curve is more or less similar. It means that the biggest funds are slower to invest their committed capital than the others, maybe because the size of the targeted companies is higher and thus there is less possible companies in the universe to invest in, or that they first secure credit lines before putting the capital at work.

When it comes to distributions, the beginning until year 7 is very much the same between the 2 patterns; but from year 8 onward the asset weighted pattern distributes more than the equal weighted, indicating a better performance (on average) for the bigger funds. This observation doesn't prove however that the best funds are the biggest; even though much of the research into the persistence performance in private equity tends to prove that the best companies raise funds in the top quartiles consecutively (see Kaplan & Schoar, 2005 [8] [11]), and with success raise increasingly large sums.

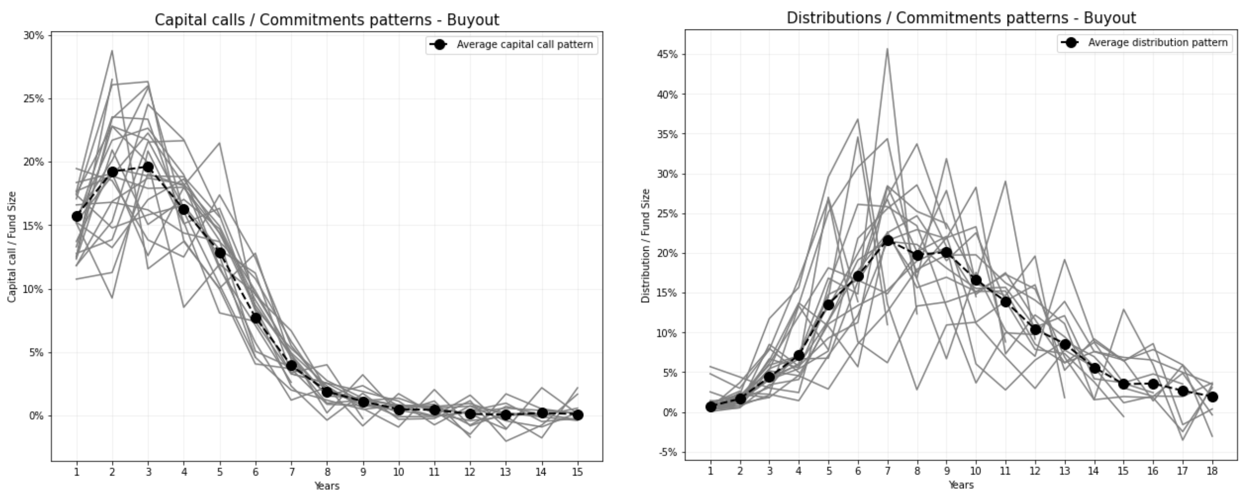
10. Note the number of points is never equal to the total number of funds, because some of them did not report anything in the first years ; and other funds are very recent and have not much data history. The total number of points will decrease simply because the youngest funds are not old enough.

11. Even though not up to date, this study is interesting because it shows the cyclicality in fundraising capacities: *Funds (and partnerships) started in boom times are less likely to raise follow-on funds, suggesting that these funds subsequently perform worse.* Repeating the same type of analysis with more recent data could be the subject of another study.

In our case, we continued to work with simple averages to be more comparable with the existing literature and the methodology used in the Institution. In all cases, averages have their limitations, as the following section will show; however, if one's want to continue working with averages, one needs to take into account this difference in patterns and the impact of fund size.

3.2.4 Averages hide more than they reveal

As finance's most famous saying goes, past performance does not guarantee the future. This is particularly true in private equity, where 2 vintages can have extremely different results (see table 10 in appendix that illustrates the different IRR by vintages according to Burgiss). The cash-flows patterns are no exception to this rule, and the speed of drawdowns/distributions will heavily depend on the state of the economy. To illustrate this problem, we computed the pattern of cash-flows coefficients by vintage, the average shown earlier and we plotted them together in figure 12. Note that each grey line represent a different vintage's pattern.



(a) Capital calls patterns by vintage

(b) Distributions patterns by vintage

Figure 12 – Dispersion of patterns depending on vintage - PitchBook data

These graphs show that there is quite a lot of dispersion over the mean across vintages for capital calls, and this is even more prevalent for distributions. We can show better this dispersion by no longer focusing one the full pattern, but only on one point in time over the full period. In figure 13, the average contributions represents the average capital called (divided by fund size) across buyout funds that were 1,2,3 and 4 years old in a given year, and the period of recessions that are defined as NBER (National Bureau of Economic Research) dated recessions, where at least 3 months in the year have been in recession.

The period of low capital calls seems linked to the period of recessions, which is also logical with the low economical activity in these periods. Uncertainty in those years was high, and visibility on the macro-economic future was low, weighing on investor sentiment and holding back capital calls. This phenomenon can also be linked to the level of interest rates in the economy, which particularly affects LBOs due to the

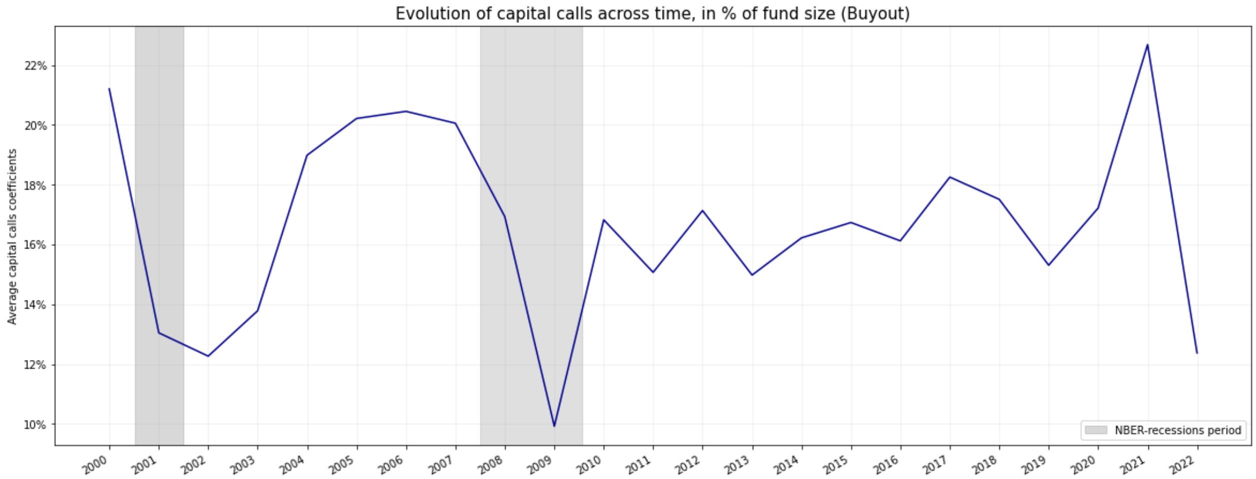


Figure 13 – Average capital call coefficient across buyout funds that were 1,2,3 and 4 years old in a given year - PitchBook data

leverage used to finance company acquisitions (see figure 41). Of course, in one year there is always high dispersion across funds and some of them still manage to continue their investments; as shown by Robinson and Sensoy (2015), the funds with a relatively high propensity to call capital in bad times perform better in both absolute and relative terms [13].

Distributions are also very pro-cyclical, as illustrated by figure 14. In 2009, at the height of the crisis, the landscape dried up completely, only to return to its long-term trend a few years later; the capital calls are also affected but no as much (halved in 2009 compared to 2004-2007, whereas the distributions are divided by 4). This can also be explained by M&A activity, and highlights another aspect in private equity investing: the funds that invested in 2006-2007 at a very high multiple would not be able to sell their companies for several years, until the economy is favorable.

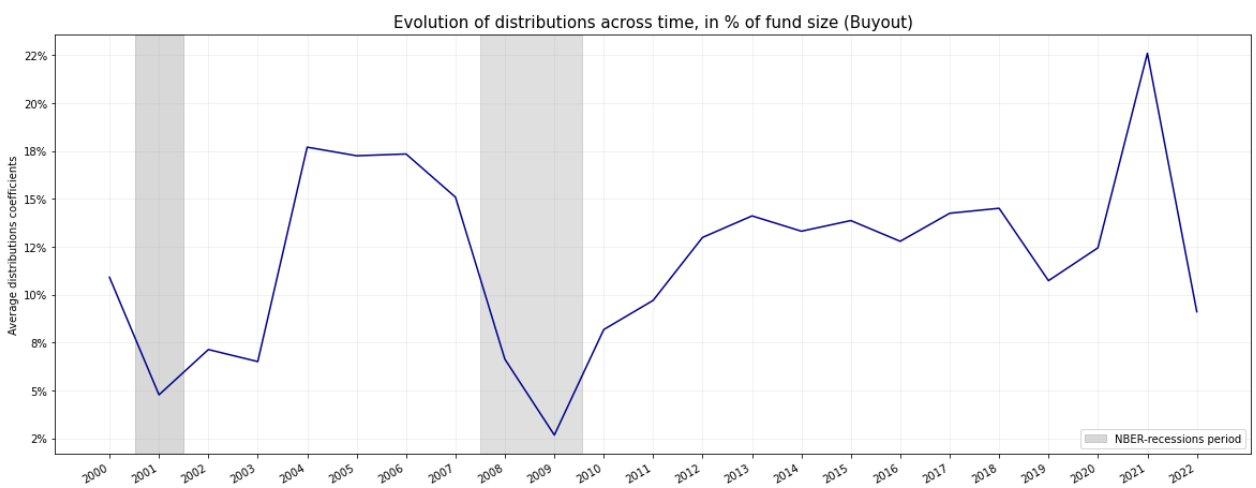


Figure 14 – Average distribution coefficient across buyout funds that were between 5 to 12 years old in a given year - PitchBook data

For other strategies, it can even be more cyclical: for example the number of IPO was down 62% in 2022 compared to 2021, affecting greatly venture capital strategies. Appendix 7.3.2 shows the results for the other strategies.

3.2.5 Alternatives approaches to cash-flow modeling

It is thus necessary to account for this dispersion to forecast the different scenarios possible. Averages are limited to this regard, so a different way of modeling is necessary. For a complete comparison of the different possible approaches, see section 2. Of all these possible approaches, we considered 2 models (apart from using the averages) : the deterministic model [14], in its original form and the stochastic approach [22] of Buchner et al. We did not select Malherbe's approach because we did not find anything more written on this model and because the NAV is considered as the least trustworthy variable in PE (see section 3.5.3). Neither the proposals made by Karatas et al. nor those by Tausch et al. met our need to carry out scenario analyses for a private equity portfolio from the LP point of vue.

Due to its complexity, the lack of interpretation of its parameters, and the lack of information concerning its concrete use in the industry, we have not selected the stochastic model for the implementation and instead focused on the deterministic model. We want the model to be understood both by experts and non-experts, and in this regard, the deterministic model is better suited because the estimated parameters stay intuitive in the private equity context ; the stochastic model is very interesting for risk modelling, but it lacks the interpretability necessary (for the author) to be adopted.

Model	**Strengths**	**Weaknesses**
Historical Averages	Simple, quick insights, fitted on data	Assumes repetition, overlooks variability, lacks adaptability
Deterministic (Yale)	Scalability, structured methodology, fitted on data, allows scenario analysis	Static, joint modeling of cash-flows patterns, one set of input gives one set of outputs
Stochastic (Buchner)	Captures uncertainty, range of outcomes, risk modeling, fitted on data	Complex methodology, dependent on the underlying functions, difficulty of implementation

Table 3 – Strengths and Weaknesses of cash-flow Modeling Approaches

3.3 The Yale model

This section aims to present the Yale model. As the main focus of this thesis, we will begin by presenting the model and all original equations and parameters. The Yale model, also called the deterministic model or the Takahashi & Alexander model (named after its inventors, professors at Yale) is the most well-known model in the industry. [12]. Created more than 20 years ago, it allows to project and model cash-flows and asset

12. Between 2001 and 2005, this paper had been downloaded more than 10,000 times from Yale School of Management International Center for Finance's website. [20]

values with a specific set of parameters; the model is called deterministic because one set of parameters will always returns one output, as opposed to the stochastic model [14].

The Yale model can thus forecast the capital calls, distributions and net asset values, as described in the following sections. We did not consider all the additions made throughout the years, as the simple model can already capture (at least partially) what we want.

3.3.1 Capital Calls modeling

Contrary to historical averages of capital calls coefficients (so divided by commitments), the capital calls (also called contributions or capital drawdowns in the literature) in the Yale model are defined after the unfunded capital. The equation is as follows:

$$C_t = RC_t * U_t \tag{3}$$

Where C_t is the capital call at time t and, U_t is the unfunded capital at time t, the latter being defined as:

$$U_t = Commitment - \sum_{0}^{t-1} C_t \tag{4}$$

And RC_t is the rate of contributions that changes over time. In the original model, only 3 rates of contributions were proposed: 25% in year 1, 33% in year 2 and 33% for all subsequent years: this approximation was sufficient to account for the usual behavior of capital calls, which are concentrated in the early years of the fund. Figure 15 represents the capital call pattern with the original RC parameters.

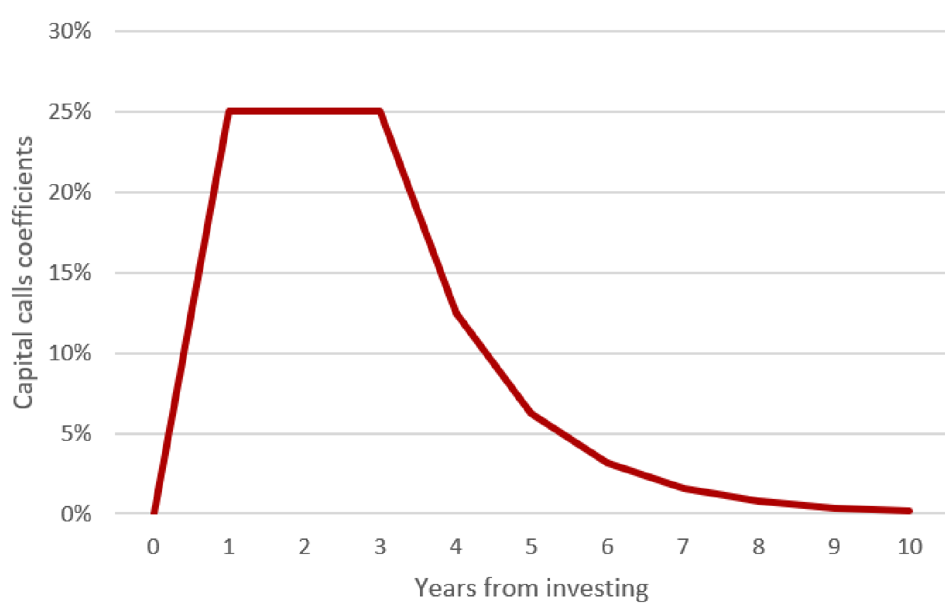


Figure 15 – Example of a capital call pattern, as modeled by the Yale

However, this rate of contribution can be better fitted on the data available or the patterns of capital calls; we will discuss it in the parameter calibration section 3.4. It is interesting to note that the modeling of capital calls in this case is not very different from historical averages, since the RC parameters will be adjusted on the same data as the averages.

3.3.2 Distributions modeling

The big difference between using historical averages and the Yale model is in the definition of the distributions. In the Yale model, the size and timing of the distributions are driven by the performance of the fund; the equation of distributions has thus 2 components, the rate of distributions (RD) and the Net Asset Value (NAV).

$$D_t = RD * [NAV_{t-1} * (1+G)] \tag{5}$$

With D_t the distribution at time t, and G the growth rate of the NAV between t-1 and t.

In the original model, RD is defined as the minimum between a fixed yield and realizations which occur as companies are sold. The yield would account for income generating assets like real estate; but in our case, we deemed it irrelevant and set it always at 0. The RD is thus:

$$RD = (t/L)^B \tag{6}$$

Where L is the Length parameter and B is the so-called Bow, a parameter that controls the speed of distributions. The higher the bow, the slower the initial increase of the distribution rate and the faster the latter acceleration. Figure 16 represents the different rate of distributions with different bows.

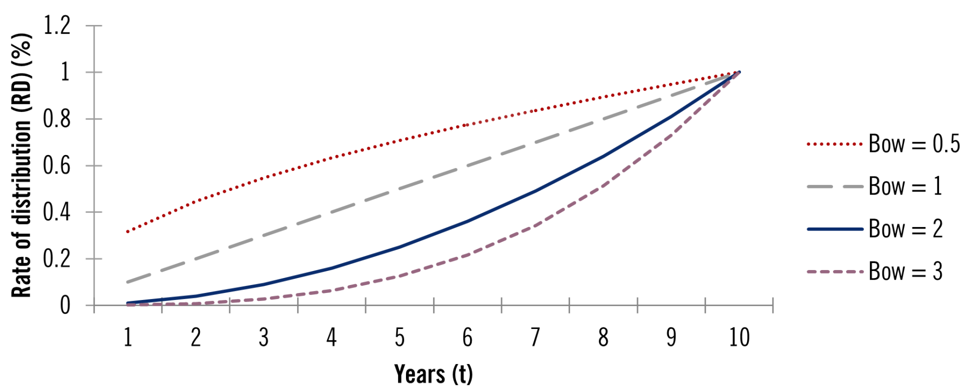


Figure 16 – Effect of different bows on the rate of distribution

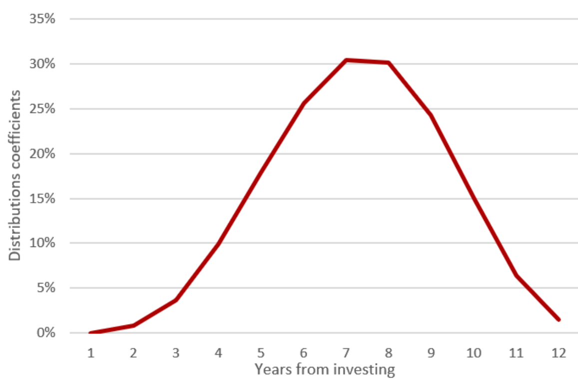
3.3.3 Net Asset Value

Finally, the Net Asset Value (or NAV) of the fund is the last dimension to be calculated. This value is affected by the performance of the fund and its underlying companies, and by capital inflows/outflows (respectively capital calls and distributions). In the Yale model, this is computed as follows:

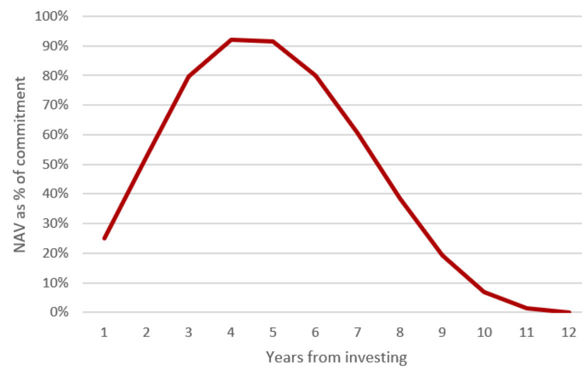
$$NAV_t = NAV_{t-1} * (1+G) + C_t - D_t \tag{7}$$

With G a growth rate that captures the performance of the fund over one year (net of management fees).

Figure 17 illustrates the distribution and NAV pattern with L = 12, Bow = 2 and G = 12% (the capital call pattern is still like figure 15).



(a) Example of the distribution pattern



(b) Example of the NAV pattern

Figure 17 – Example of distributions and NAV patterns, as modeled by the Yale

3.4 Parameters estimation

A model's outputs are only as good as its inputs. The key for a good implementation of the Yale model is thus a good estimation of its parameters. These parameters are the Commitment, the Rate of Contributions (RC), the Bow (B), the Length (L) and the Growth rate (G). They can be estimated each one independently or jointly, depending on the model user. In our case, we have estimated them independently in order to avoid the parameters taking extreme values; for example, in the paper by Furenstam and Forsell (2018) the Length is estimated to be 193 quarters, or 48.25 years (!) ; this result can be mathematically correct but in practice, it is far too high compared to the normal life of a fund.

3.4.1 Commitment and Rate of Contribution

The 2 parameters that modify the capital calls in the original model are straightforward. The commitment is an input of the user, and represents the amount the LP commits to the GP. The rate of contribution can be calculated using two methods :

— The first method would use the available dataset, and computes for each fund i the RC at time t according to the following equation:

$$RC_{t,i} = \frac{C_{t,i}}{U_{t,i}} \tag{8}$$

And then averaging the results for each period in the dataset.

— The second method is based on a given capital calls coefficient pattern, or calculated on the database (see fig 9). The average capital call pattern can be normalized over the desired time interval (e.g.

12 periods), and the unfunded capital is then calculated as $U_t = U_{t-1}$ - C_t; at t = 0, unfunded = commitment = sum of capital calls, effectively normalizing the results so that everything is called at the end of the forecast. Next, we apply equation 8, to find the RC as a function of the unfunded capital. This method is the same as the previous one, except that we calculate the RC directly on the average pattern rather than recalculating it for each fund.

We used the second method because of the good understanding we have of the patterns of capital calls (see section 3.2) and the data cleaning associated; in any case, the RC should be the same as the theoretical definition of this parameter does not change. See part 4.1 for the results computed for each selected sub-strategy.

3.4.2 Length

The length parameter represents the maximum age of the fund, after which it is terminated. Usually, a buyout fund is invested for 10 to 12-13 years, with the prolongation period; this period is smaller for others strategies like private credit, quicker to distribute. It our data, it wasn't possible to see if the fund was terminated or not, so we needed a workaround. Quantitatively, we made the assumption that this parameter can be defined as the year where the NAV is lower than a pre-specified level, usually 10% or 20% of the original fund size. At this stage, the fund could be effectively considered as terminated.

We put the threshold at 15% of the fund size. For every fund we compute the NAV coefficient as NAV_t/Fund size, then aggregate the results as shown in appendix 42. Due to big outliers (that can be explained by the performance of the fund), we preferred here to use the median and to compute the year where the coefficient is below 15%. The results are shown in section 4.1. We noticed in the data that a large number of buyout funds were continuing to distribute after unusually long period of time, a a trend that has also been noticed by the industry:*The average lifespan of funds across the whole private capital industry is increasing beyond the typical 10 years... older funds of vintages 2000-2005 still hold a substantial \$204bn worth of investments, equating to 7.2% of total unrealized assets* (Preqin report, 2016) [13]. The behaviour of funds already older than the Length parameter can be the subject of another research; in this thesis, we didn't find any particular pattern specific to old funds, apart from the fact that they distribute further back in time. This may be expanded by an analysis of top-up or side funds, which are funds created to supported invested companies with limited exit possibilities (like VC after the dot-com bubble). It was difficult in the data to determine what funds were top-up, so we could not dwell deeper on this subject.

Nonetheless, this is a problem when applying the Yale model to old funds, especially when the fund's age is already superior to L; we look at this (among others) limitation in section 5.2.

3.4.3 Growth rate

In the original paper, the growth rate is constant every year. Since it is a return, it can be compared to private equity performance measures such as the internal rate of return (IRR). In fact, the G will be equal by construction to the IRR as it also factors the inflows/outflows.

13. Cited in Gupta and Van Nieuwerburgh, 2021 [31]

The IRR is defined as the rate of return wherein the net present value of the investment equals zero. The equation is the following :

$$0 = NPV = \sum_{i=0}^{I} \frac{CF_i}{(1+IRR)^i} \tag{9}$$

with I the holding period and CF_n the cash-flows.

The IRR is very prone to cyclicality, and will change depending on the vintage. Table 10 represents the median IRR of primaries funds by vintage and strategies, where we can clearly see different performance through time. Changing the growth parameter allows thus to simulate a crisis like the GFC, even though there are limitations to the IRR measure.

For example, time weighted rate of return and IRR are identical only where there have not been any contributions or distributions from a portfolio during the measurement period ; it is thus not the case here. The IRR is very sensitive to small changes in the timing of cash-flows, having the distributions at the end or at the middle of a year can lead to very differents IRR. It also assumes that any positive cash flows are reinvested at the same IRR. Furthermore, the G parameter is assigned to the NAV, and the NAV is considered as the least reliable measure in private markets valuation since it depends (until the sale) on the GP , and we can add to this the lag inherent in private equity valuations. Fixing one particular growth rate rather than another is therefore difficult to justify.

In section 3.5.4, we look at the solutions already proposed and come up with our own solution to this problem, no longer thinking in terms of growth rates but rather in terms of sum of distributions.

3.4.4 Bow

The last parameter to estimate in the Yale model is the bow, that controls the timing of distributions (see 3.3.2 and equation 3.3.2). As the bow is an exponent on the rate of distributions, setting the bow higher would decrease the RD initially and increasing it after, effectively "pushing" the distributions to the right, and vice-versa. For example, see figure 18 where we change the bow from 1 to 3.

Contrary to the others parameters, the bow is not directly observed on a investment and is rather a statistical parameter than an economic one. It can be estimated thanks to the method of ordinary least square (OLS).

OLS definitions : For a random sample x_1, x_2,..., x_n generated from a random variable X and $f_i(\theta)$ be the objective function. The function of least squares is then defined as:

$$Q(\theta) := \sum_{i=1}^{t} (x_t - f_t(\theta))^2 \tag{10}$$

The value θ^* that minimizes $Q(\theta)$ is the least squared estimator of θ.

Here, x is the realized distribution of the fund and f(θ), or f(Bow) in this case the modelled distribution of the TA model. The OLS method would thus minimizes the error between the distribution prediction and the true outcome.

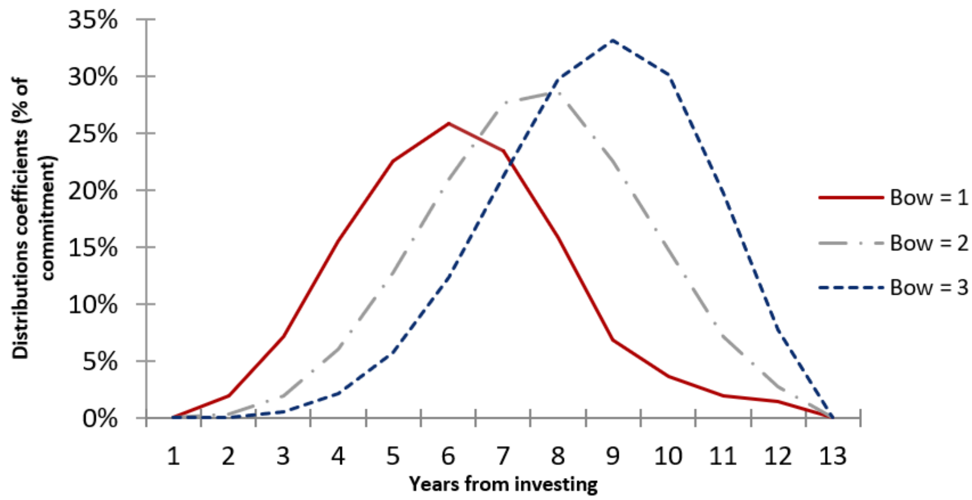


Figure 18 – Distributions coefficients by bow rate, others parameters fixed

In the Yale model, D_t will depend on G and L (see equation 5), since D_t is a function of NAV_{t-1} and RD, we would thus need to estimate several parameters at the same time. To overcome this problem, we fix L as 13 years (see section 3.4.2 and the estimation results for different strategies, 4.1), and G is set at 0 ; if there is no growth on the NAV, the sum of distributions will always be equal to the commitment. And if we set the commitment equal to 1, the distributions will be be expressed as a percentage and we then can compare this pattern of modeled distributions to the realized pattern of distributions coefficient of each fund (normalized with the sum of distributions coefficients at year equal to L, to be comparable). We use the OLS function to fit best the pattern of modeled distributions to the realized one, by changing only the bow.

We obtain a bow by fund, and we then compute this bow for every fund in the cleaned database. If we group them by vintage, we can observe the different bow for different vintages, as observed in figure 19 just below. Whiskers represent the 1 and 99% percentile, with the orange line as median.

This chart computed on buyout funds expresses the evolution of bow over time. The vintages from the beginning of the millennium distributed fairly quickly, thanks to the favorable economic context between the dot-com crisis and the GFC (see chart 1 that shows the rise in raised capital). But from 2003-2004, the bow increased sharply to reach (on average) over 3.25 for vintage 2005. This can be explained by all the funds which, having invested in companies in 2006 and 2007 at high valuations, were unable to sell them for several years; this vintage is also the worst in terms of performance (see section 3.5.5). The bow then decreased from year to year, but remained fairly high, and there was no drastic reduction directly after the crisis. Lastly, the 2016-2017 vintages were the fastest-distributing, thanks to the remarkable performance of private equity in 2021, driven by the number of M&A, IPOs and economic activity that year. It is also interesting to note the dispersion of bows on the vintages 2003 and 2004. This may be due on the fact that some funds managed to sell their underlying quickly before the GFC, and some others did not. Being a top performer these years meant not necessarily the highest multiple possible, but simply to make these multiples a reality quickly.

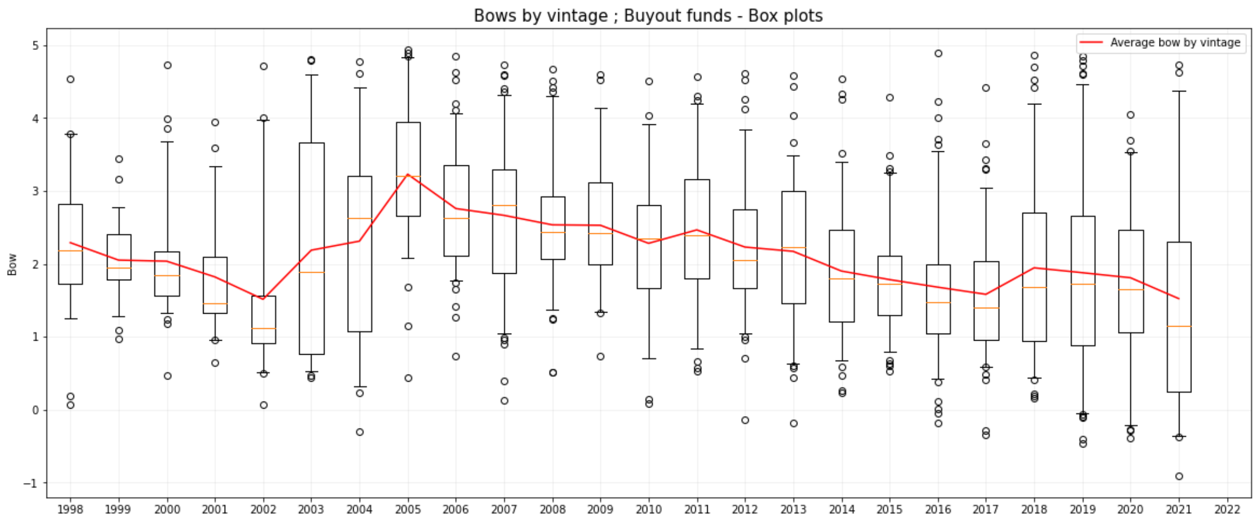


Figure 19 – Evolution of the bow factor depending on vintages - PitchBook data

This radical change illustrates the variation in the timing of private equity fund distributions over different economic cycles, and the need to take this into account when testing the model. This aspect will be further developed in part , where we apply bows that simulate crisis vintage to stress test an existing portfolio. Finally, this analysis can be done also for other strategies. In appendix 7.3.3, you will find the evolution of bow for RE, PC, VC, Growth and Fund of buyout funds, that all react differently to crisis, and the average bow every vintage by strategy.

3.5 Additions to the original model

This section focuses on the model's weaknesses and the proposed solution to remedy them. We first address in detail the problem posed by capital calls, which in the original Yale model are totally fixed, unless the entire pattern is changed. We propose a new parameter, called the Hook, which captures at least part of the dispersion of capital call patterns ; and a multiplier factor on the RC. Then, we look at the problem of the growth rate, and propose a solution to stop reasoning with this parameter while keeping it in the model. Finally, we add a co-investment model, which can be modified according to the LP's assumptions.

3.5.1 Modifying the timing of capital calls - Addition of the "Hook" parameter

In the original Yale model the focus is on distributions, where we can change the convexity/speed of distributions thanks to the bow, and change size of distributions thanks to the growth parameter; the combination of the two parameters will determine the performance of the fund. But the equation for the contribution is not different from a simple average computed on a database, except that its focused on unfunded capital and not committed. This is one of the weakness of the original model because forecasting capital calls and their speed is crucial for the LP :

From an LP's perspective, a capital call is law-binding, meaning that if a LP is not able to provide the money to the GP it can have serious repercussion on the LP's reputation in the sector (see section 1.2) [14]. Stressing your liquidity or funding risk by having more capital calls in the beginning of the fund life can help to avoid this threat.

To account for this problematic, I implemented a new parameter on the capital call equation. This parameter is called the Hook, named after the golf shot that goes to the left; the original idea was to test liquidity risk and thus "bring to the left" the pattern of capital calls.
The new equation becomes:

$$C_t = U_t * RC_t * (\frac{t}{L})^H \tag{11}$$

As you can see, the "hook" parameter has exactly the same shape as the "bow" parameter. It allows us to vary exactly what we want to achieve, i.e. modify the speed of the capital calls. From a practical point of view, this hook can be negative to obtain a tighter pattern at the beginning of the curve, or positive to delay capital calls.

The chart below represents the hook applied to the capital call pattern of buyout funds. A hook between -1 and 0 will put more weight on the first years, and a hook above 0 will delay the capital calls (note that the hook is applied to the RC parameter, and not directly on the CC pattern).

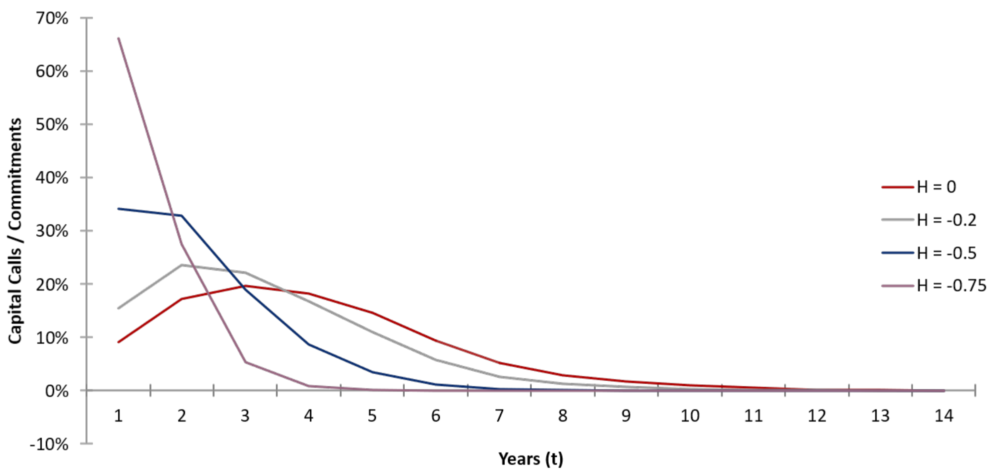


Figure 20 – Change of Capital Calls with different Hook parameters

The hook thus allows to effectively stress test the capital calls, but can it be seen in the data as good as the bow ? First off, we need to fit the parameter to the model. We will use the same methodology as the bow, i.e. use the OLS method. However, we know that the majority of the capital calls is concentrated in the first years of the fund, so we want to put less weight on each residual as it becomes more distant from the beginning. The objective function is the following:

$$\min_{Hook} \sum_{0}^{t} \frac{(\text{Observed CC coefficient}_t - \text{Hooked average CC coefficient}_t)^2}{t} \tag{12}$$

14. The salience of this risk in investors' minds is illustrated by many news accounts of the behavior of many famous institutional limited partners who, during the financial crisis urged general partners not to call capital (Financial Times, 2008) ; quoted in Robinson and Sensoy [13].

The residuals divided by t effectively mean that a later capital call will result in less residuals. We can then vary the hook parameter, changing the RC and thus the capital calls as shown before; effectively giving us different hook for different observed patterns. For example, we can look at averages values by vintage like the bow:

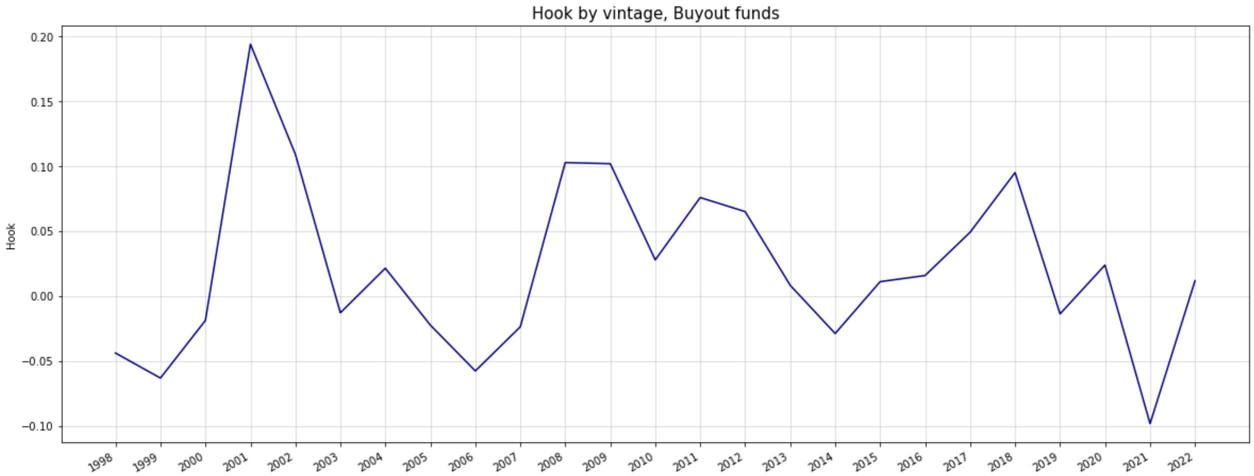


Figure 21 – Evolution of average hook by vintage - PitchBook data

As a reminder, the average pattern will have by construction a hook of 0; lower than zero means that the pattern had quicker capital calls than the average, and higher than zero means that the pattern had lower capital calls than the average. Note that 1996 and 1997 are not shown due to not enough points to compute their patterns (only 2 or 3 funds provided information on contributions in those years).

The hook behaves as expected: it is higher than 0 in period of stress like the dot-com bubble and the GFC where investments where delayed, and lower than 0 in period of high economic activity like before the GFC or in 2021. However, this phenomenon is mitigated by periods such as 2018, when the hook is as high as in 2008, even though there was no stress of the same magnitude (except the rise of interest rate that could have impacted LBOs). This can be explained by the fact that we don't sufficient data to compute the full pattern, (only 5 points for 2018), or that the hook is not sufficient to fully capture changes in the pattern.

If we decompose the patterns by vintages as shown in fig 22, we can see the different behaviours of capital calls. The problem is that the stress period are often only one or two years long, and activity restarts after (like what happened for 1999-2000 vintages, or 2007-2008 vintages). The hook works extremely well when the vintage's is the crisis one (like in 2001 or 2009), but does not capture the negative spikes for vintages just before crisis.

On the other hand, vintages invested during the economic boom between the crisis (2003-2005) have seen their profile shift upwards rather than to the left, with capital calls that are not faster but more important. It's not so much the convexity of the pattern that's changed, but rather its scale.

In its current form, the hook works, at least in part. Changing the convexity of the entire pattern makes sense in the case of distributions, since investments will be prevented from being exited for years; but for entry

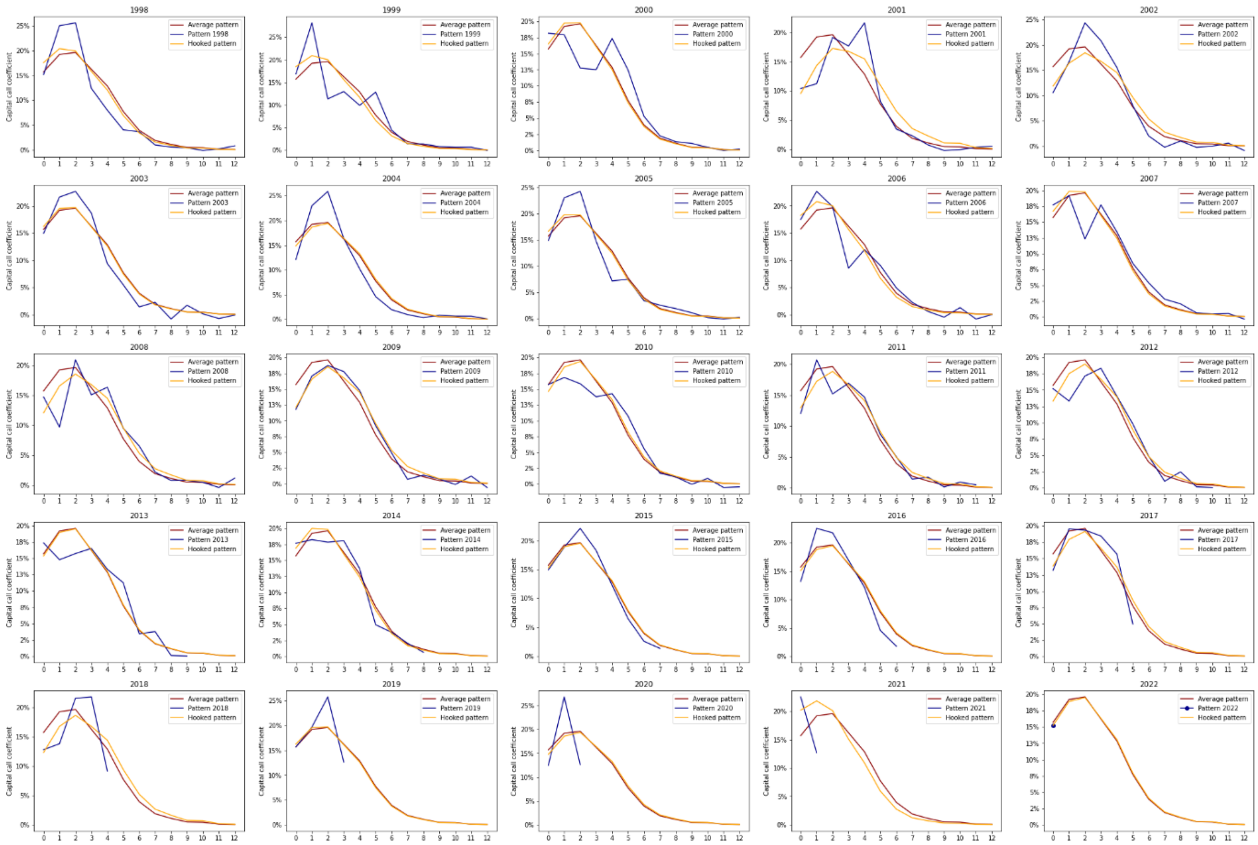


Figure 22 – Average capital call pattern, vintage pattern and hooked patterns

points, once the multiples are adjusted downwards the situation should return to normal (i.e. more capital calls, like 2001-2002 or 2008; this also depends on the economy) ; see figure 40 for the evolution of multiples during time. To achieve a better result, we therefore need another parameter applied only over one or a few years.

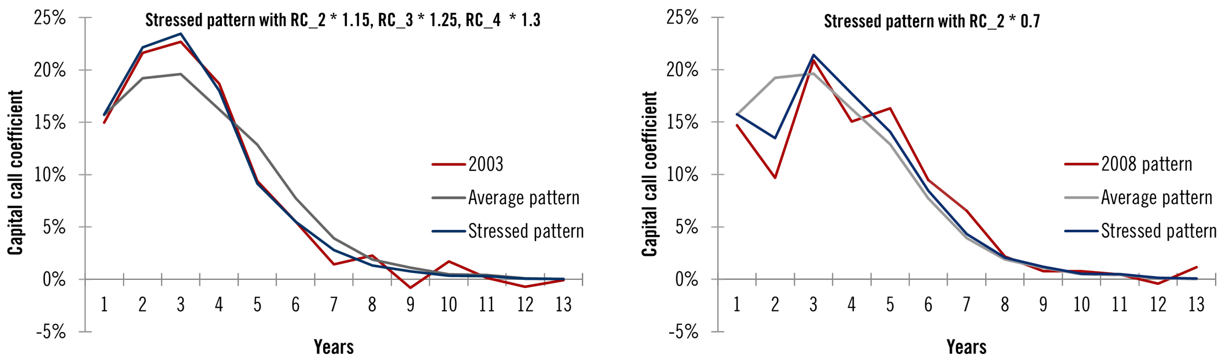
3.5.2 Modifying the scale of capital calls - Addition of a multiplier factor on RC

Implementing another parameter that would multiply the RC by a specific factor (like 1.5) could be a solution. We still begin from the equation 8, and this time it looks like this :

$$CC_t = U_t \cdot RC_t \cdot M_t \tag{13}$$

With M_t a multiplier factor ; putting 1.5 would effectively mean multiplying the RC in s specific year by 1.5.

It is then easy to apply such multiplier on one year or several to simulate a shock lasting a single year, or an entire period of low/high activity.
The following graph shows respectively an curve shifted upwards like the 2003 vintage (1.15 x RC_2, 1.25 x RC_3, 1.3 x RC_4) and a downward spike like the 2008 vintage (RC_2 x 0.7).



(a) Stressed pattern and 2003 vintage pattern

(b) Stressed pattern and 2008 vintage pattern

Figure 23 – Stressed patterns with a multiplier factor and stress vintages patterns

Applying this multiplier to the RC is thus an easy way to stress the capital call pattern. One could argue that directly selecting stress patterns like the 2000 vintage as the RC pattern would also allow such stress, even though using old data might lead to inaccurate results especially in different types of crisis. Changing the RC with different parameters is thus a way to stress test the capital calls. Further work can be done on this subject, but ours shows that the Yale model allows to control for the possible contributions by modifying the RC, be it with an exponent like the Hook factor or a by multiplier factor.

3.5.3 Inconsistencies of the NAV

In private equity, the Net Asset Value (also called remaining value or market value) refers to the current value of a private equity fund's investments minus its liabilities. While the definition of NAV remains consistent across both public and private markets, its application diverges significantly. In public markets, it is possible to break down and sell quickly the positions ; but such an assumption is difficult to hold in private markets, characterized by the illiquid nature of hold positions. The goal of private investments is to earn long-term attractive returns, so a "fire sale" mentality is problematic.

For a GP, determining the NAV can complex due to the illiquid nature of their investments. Since many private equity investments are in private companies or assets that aren't publicly traded, their valuations are often based on financial models, expert assessments, and comparative valuations rather than market prices. This can introduce a degree of subjectivity into NAV calculations, leading to potential inconsistencies or biases in reported values. Not all private markets valuations are born equal. [15]

The NAV is thus commonly described as the least trustworthy variable in the private investment universe. To quote Meyer and Mathonet (2005) : *As a result, the NAV alone is a naïve model of a private equity*

15. see Blaydon & Horvath (2003a), quoted in [3]:*"They (private equity valuations) are an interim report on the performance of the fund and rely on the GP's assessment of unrealized current portfolio company values... The IRRs, publicly reported or not, are only as good as these underlying assessments... But what is missing is sufficient discussion of how the underlying assessment of company value are arrived at, other than to note that some funds may have widely differing assessments of the value of the same company, much to the frustration of LPs to whom these assessments are reported".*

fund's value. Such illiquid and long term oriented assets are essentially marked-to-model and reported NAVs require a review by experienced investors. It is impractical to review the consistency of assumptions made by many fund managers and to ensure them for often thousands of portfolios companies. Finally, we have to deal with the question that committing to a private equity fund is investing into a blind pool.

Considering all these problems, it is then difficult to set a specific growth rate to the NAV like the Yale model does. It is also the main reason why cash-flow modeling approaches like de Malherbe (2005) that first model the NAV and then derive the cash-flows have not been considered in this thesis. It is however important to still keep the NAV as it is in the Yale model, as the NAV will decide *in fine* the level of distributions in the forecast. The next section focuses on finding a workaround inside the Yale model framework.

3.5.4 Change of perspective: set a target TVPI rather than a fixed G

Another weakness of the TA model is the G parameter that defines the growth rate of the NAV. This parameter being constant poses 2 problems:

- The first one is simply that a constant growth is not realist with regards to the cyclicality of private markets, and how an investment could behave during different markets. Some solutions have been proposed, including changing the G depending on the fund age (more mature investments would grow less), or having a market sensitive periodic growth, using a lagged regression like the proposal of V. Jeet (2020) [19]. This brings the Yale model closer to reality, even though a lot of literature has been written on the subject of "smoothed" ans inaccurate private markets returns, like IRR [9]. Moreover, Jeet's model results are not consistent with the realized data after 2006, which could be due according to others factors like the change of Bow due to the GFC crisis.
- The second problem is a modeling limitation, particularly with stress tests. We have shown earlier in this document that the bow and the hook observed on the data move over time, especially during financial crises. Modifying the bow or the hook will have a great impact on performance: the quicker the drawdowns are invested in companies, the quicker the companies will be distributed on average, thus increasing the IRR; the contrary would decrease it. But with our model, if we modify the bow to take account of slower distributions at the start, a constant G means that the longer the holding period, the larger the total distribution will be. If we take again our example shown in figure 18 but this time plot the cumulative distributions (figure 24), we see that a bow of 1 gives rise to 1.260 cumulative distributions, a bow of 2 to 1.453, and a bow of 3 to 1.625. Since the initial desire was to stress the distributions, this is counterproductive.

There are thus two possible solutions: either we change the G according to macro-economic indicators in the period, or stop setting the G directly and instead start from the end and change the final multiple : the sum of distributions, equal to the TVPI at the end of the fund.

TVPI stands for "total value to paid in capital". The formula is :

$$TVPI = \frac{\text{Total distributions + NAV}}{\text{Total capital called}} \tag{14}$$

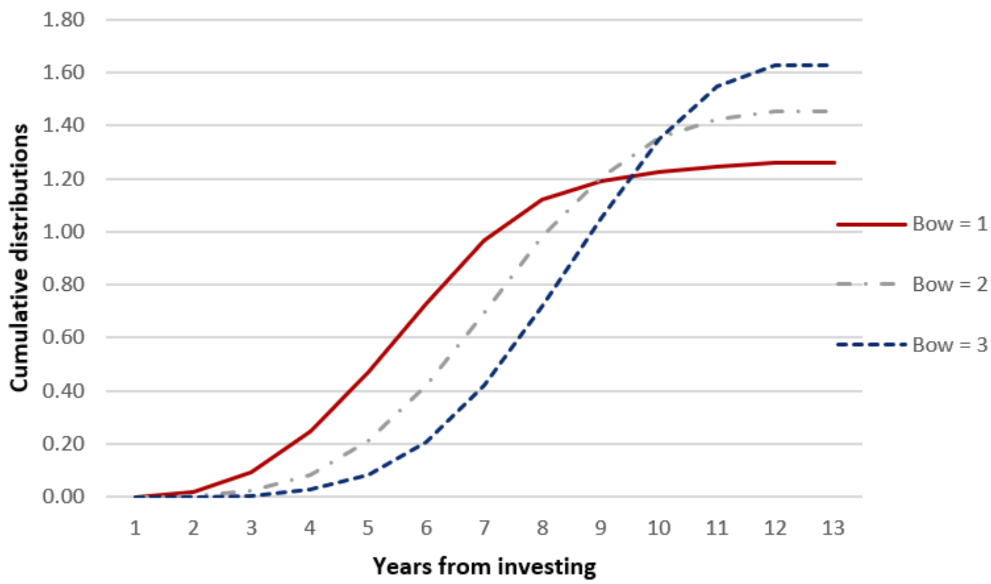


Figure 24 – Cumulative distributions by bow rate, others parameters fixed (1$ commitment)

It is thus simply a multiple of performance on investment : a TVPI higher than 1 indicates that the fund has generated more value than the LPs have invested, suggesting a positive return (and the opposite).

The second solution was preferred because we know the problems associated with private equity returns, and that the TVPI was a more robust measure of performance than the IRR for this framework.

The idea is that at the end of the fund's life all the NAV will be distributed, so the TVPI will be equal to the sum of distributions. In the Yale model, the model's user would select the Length, RC, the Bow and G that would give a pattern of distributions. But if our objective is here the final sum of distributions, we can reason in the opposite way : we would fix all parameters except the G (meaning that only the G can change), and choose a TVPI.

Then, the model would select the G accordingly to arrive at a sum of the distributions equal to the expected TVPI; a larger bow would therefore necessarily decrease the G, and a larger hook would increase it.

This is a non-linear optimization with a constraint (TVPI reached = TVPI selected). The objective function here would be the Yale model, and we would again use the OLS method but this time by varying the G parameter. The optimal G would be found as :

$$\min_{G} \ (\text{TVPI} - \text{TVPI estimated})^2 \tag{15}$$

And we can effectively choose the expected TVPI in our forecast. Of course, the G parameter will still exist (even though not changed direclty by the user), and it will be a constant applied on the NAV ; but considering here that our focus in on the cash-flows, the only verifiable data and not on the NAV, we can consider that we have resolved the problem of the G parameter. This change of perspective, as we call it, also enables another layer of sophistication: analyzing the distribution of TVPIs.

3.5.5 TVPI

As with all other parameters of the model, the expected TVPI will depend on the strategy and the vintage of each fund. A riskier strategy like LBO in a good vintage will have an expected TVPI higher than a PC fund, even though the time to reach the TVPI will also be different (this part can be measured by the Length and Bow parameters). I had access to the TVPI reported by Burgiss depending on strategy and the vintage, which is summarized in the appendix 9. There is high variability between vintages, where very good vintages like the ones just after the GFC reach more than 1.8 TVPI [16] whereas the one invested just before barely reach 1.6 (and by the way confirming that the 2005-2008 were the worst ones as shown by the bow).

It could be interesting to link the TVPI with the bow, with the following reasoning: the GP who invests in companies is primarily concerned with the performance of his fund, a good performance enabling him to charge incentive fees and position himself in the best quartile, making it easier to raise his next fund. If his investments are blocked for a long time, the IRR will be too low to achieve his goals, so he will then focus on the TVPI [17]. As a result, even vintages that take longer to distribute would not see a drastic change in the final TVPI. And this intuition is confirmed at least visually by the data : we plotted in figure 25 the TVPI and bow (mean) reached by vintage, and we can clearly see that that majority of vintage get to at least 1.5 TVPI, with the exception of 2005 funds that were the worst one to invest in all the dataset.

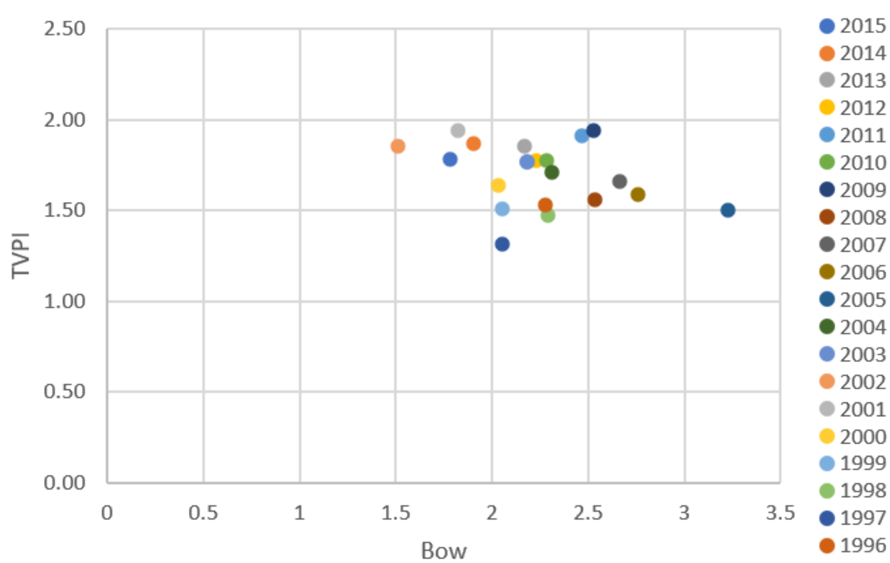
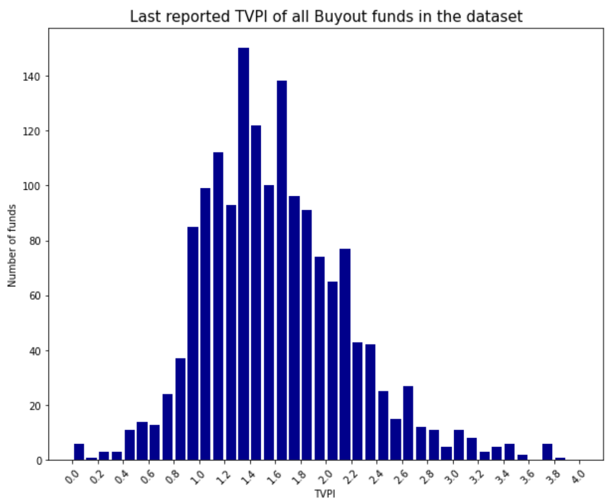


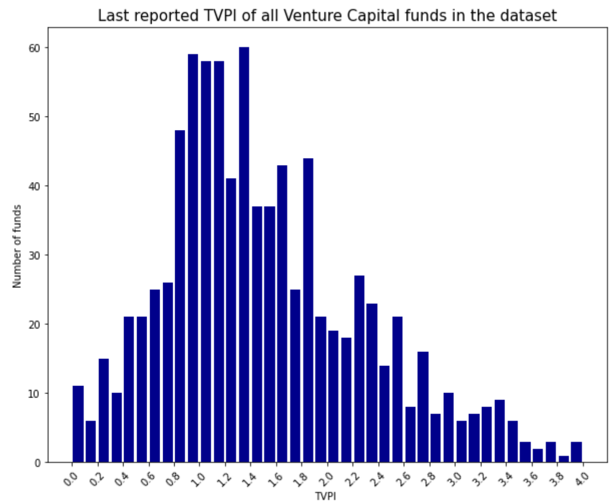
Figure 25 – Scatter plot of TVPI and bow by vintage - PitchBook data

16. Confirmed by [13]: investing in downturns have lead to higher performance
17. The ability to time exits in order to maximize value is one attractive feature of private market investments ; according to Larocque et al (2022) [32] over half of funds' IRR is attributable to cash-flow timing efforts.

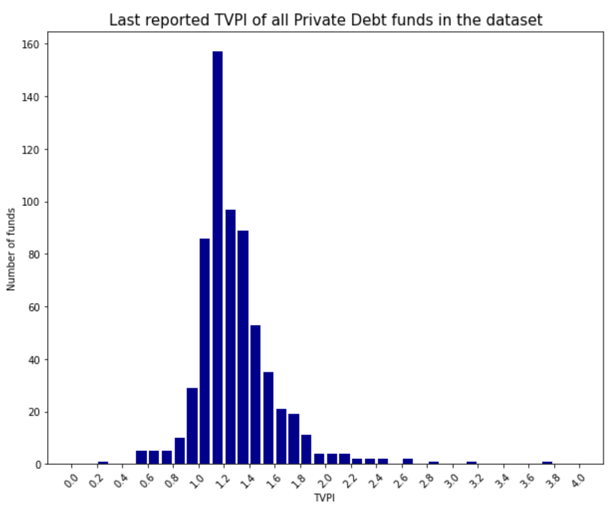
We can also compute the TVPI on the PitchBook database, because every fund in it will have a reported TVPI. The result for LBOs and other strategies are shown below as an histogram for every fund :



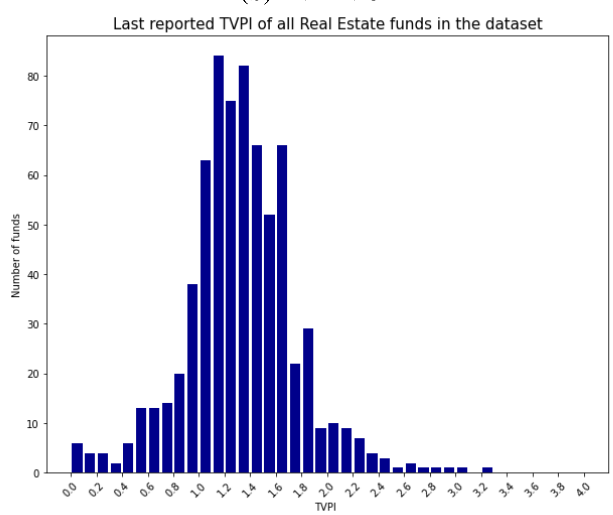
(a) TVPI LBO



(b) TVPI VC



(c) TVPI PC



(d) TVPI RE

Figure 26 – TVPI of all selected strategies - PitchBook data

As expected, LBO and VC will have skewed distributions of TVPI, with a lot of "average" funds and a small amount of very good ones, whereas PC and RE are more concentrated. If we look again at Burgiss data (9), we see the same distinction between strategies but surprisingly the median Buyout's TVPI is higher than VC's TVPI. It can again be explained by the very skewed distribution of VC funds (see figure 26), but it also confirms the chart 9 of distribution patterns.
Seeing these (also skewed) distributions, one could take randomly a value in it to simulate the payoff of a future fund depending on its strategy. For our case, we took the median for each strategy.

3.5.6 How to model co-investments ?

Co-Investment is another possible way of investing in private markets as LP that we have not looked so far, but the Institution was also interested to dig on this subject.

Co-investment allows LPs to invest directly in a company alongside GPs, by syndicating the financing round between both. For the LP, its a way of potentially enhancing their private allocation returns as they are more concentrated, and do not have to pay a management fee or carried interest. For the GP, it seems not interesting as they would relinquish their fees; but in some situations, co-investment is a way of bypassing concentration problems or liquidity needs. For example, if the terms and conditions of the LP agreement states that the fund can not put more than 10% of the fund size into a single investment, then co-investment can be a solution because the GP can invest in bigger and more targets.

It may be even more relevant for smaller funds investing in venture capital, as they may find interesting deals but lack the capital to pursue these opportunities through to exit without being diluted ; co-investment can help providing the necessary amount to take a sufficient stake, diminishing the downside risk. It ts also less problematic than syndicating with other PE/VC funds, as there is less potential conflicts with regards to the management.

Furthermore, as financing with debt becomes extremely costly (see fig 41) financing though equity becomes more relevant than before, putting pressure on GPs to find the capital needed.

It is even more difficult to find data on this type of investment, be it on the timing of cash-flows or simply performance due to transparency issues. Co-investing offers more upside, but also more risks as LPs here will participate partially or totally in the companies management and are responsible for their liabilities. Some investors are very enthusiastic about it and report that "on average, we have had signicantly better returns on co-investments than fund investments, but that is because we have been involved in very good ones" [18], whereas others investors experienced very bad results [19].

Implementing co-investments in the Yale model is difficult, and we would lose the economic intuition of the parameters. As a first approach, we consider that a co-investment is simply one investment, regardless of the number of financing rounds that can happen in reality. Here, the full commitment must be called up at the outset and then repaid when the investment is sold, multiplied by a factor called MOIC in the sector. Here, we can only play with the time to exit and the exit multiple, which can range from 0 when the underlying company is in default to 20 (for example) if it's a very good investment, potentially with an IPO. Incorporating co-investment cash-flows would resemble a Dirac function, with a single capital call and a single distribution. The only thing to compute here would be the NAV, that will depend on the exit multiple and the time to exit.

18. Hans van Swaay, Pictet & Cie (quoted in [2])

19. One such example is the case of Brazilian data center company Aceco T1. Private equity firm KKR Co. acquired the company in 2014 along with its co-investors, the Singaporean investment firm GIC and the Teacher Retirement System of Texas. The company was found to have cooked its books since 2012 and KKR wrote down its investment in the company to zero in 2017. Quoted in https://www.investopedia.com/terms/e/equity-coinvestment.asp

We define the NAV of co-investments as:

$$NAV_{\text{coinv}_t} = (NAV_{\text{coinv}_{t-1}} \cdot \text{Growth})^{\frac{1}{\text{Time to exit}}}$$
$$\text{With Growth} = \frac{\text{Commitment} \cdot \text{Exit multiple}}{\text{Commitment}} - 1 \tag{16}$$

Figure 27 presents an example of a co-investment that took 6 years to exit from the first capital call, with a MOIC of 1.8. In this example, the NAV will grow constant to reach the expected amount to be distributed in

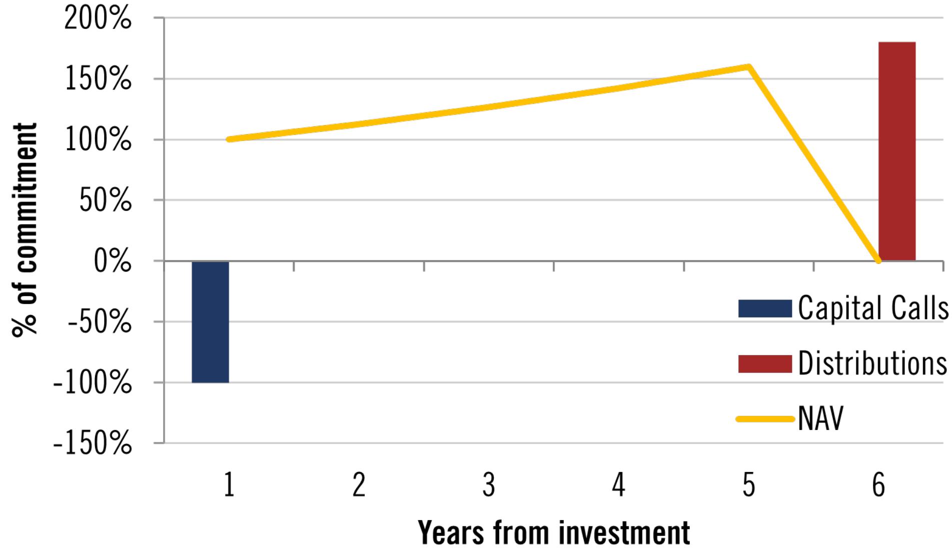


Figure 27 – Forecast of a co-investment, exit after 6 years, TVPI 1.8

the final year. A constant growth rate on the NAV is not realistic, but (like using the TVPI in the Yale model) the focus here is on the cash-flows and not the NAV, an unreliable measure of performance. The parameters of the co-investment forecast will change accordingly to the LBO model on the outset of the co-investment. For example, in a buyout co-investment, the parameters are closely linked to the assumption made by the GP (like expected IRR).

It is then possible to merge the co-investment forecast with the Yale model forecast in a single portfolio forecast. We show the result of this implementation in section 4.2.

4 Results

This section presents the estimation results of the Yale model parameters with the calibration method explained in the previous section. Then, we implement the Yale model for an existing portfolio, and extend the forecast with a future commitment strategy based on 4 different sub-strategies that proves the relevance of the Yale model for cash-flow forecasting purposes, especially for stress-testing needs. We also present and compare the result of this implementation by applying back-tests on the already existing data, comparing a pattern or directly a portfolio.

4.1 Parameters estimation results

Table 4 presents the estimated parameters for each strategy and vintage. Note that the RC does not change depending on vintage, only the hook would change the pattern according to the user ; by definition, the hook is equal to 0 for the average pattern of RC computed here. For every parameter, we have used the theory and method described in section 3.4.

Table 4 – Parameters estimation results, fitted on PitchBook data

Parameters/Strategy	LBO	PC	RE	VC
Bow	2.1356	1.5141	1.7906	2.4483
L	13	10	10	15
TVPI (median)	1.54	1.23	1.31	1.43
RC	0.158	0.337	0.254	0.173
	0.230	0.436	0.329	0.249
	0.304	0.468	0.410	0.308
	0.362	0.576	0.491	0.346
	0.449	0.520	0.546	0.369
	0.490	0.521	0.522	0.421
	0.490	0.585	0.576	0.438
	0.461	-0.390	0.601	0.475
	0.501	1.204	0.898	0.461
	0.442	1	1	0.434
	0.711			0.512
	0.677			0.435
	1			0.457
				0.793
				1

It is interesting to note the difference in bow between LBO, private credit and RE, which correspond well to the idea of these investment classes with a different risk/return. On the other hand, VC has both a higher bow and a lower TVPI, due to all the funds that have taken years to exit their investments. It is also the strategy the most sensitive to outliers; changing the median parameter to the mean would give a TVPI of 1.83. However, we have kept these values unchanged in the following sections, in order to be conservative in our assumptions. All forecasts are therefore based on parameters taken from this table of results.

4.2 Application for a sample portfolio

All coding was done in python.

The first step to forecast the cash-flows of an sample portfolio is to retrieve the historical cash-flows and

NAV of the portfolio's funds. With it, we can then use the Yale model as defined before to forecast the next cash-flows until the end of the fund. Several rules are put in place depending on the fund age and its already existing performance (see section 5.2 for more details).

Once we have the forecast for one fund, we can forecast every other fund in the sample portfolio. The parameters will change if it is a co-investment, real estate fund, leveraged buyout fund, venture capital fund or private credit fund. Here. we take all parameters shown in table 4, with the exception of the TPVI that was put at 1.5 for LBOs, and 1.8 for co-investments to be comparable with previous forecasts. Then we just have to merge all funds information to obtain the final forecast as shown below :

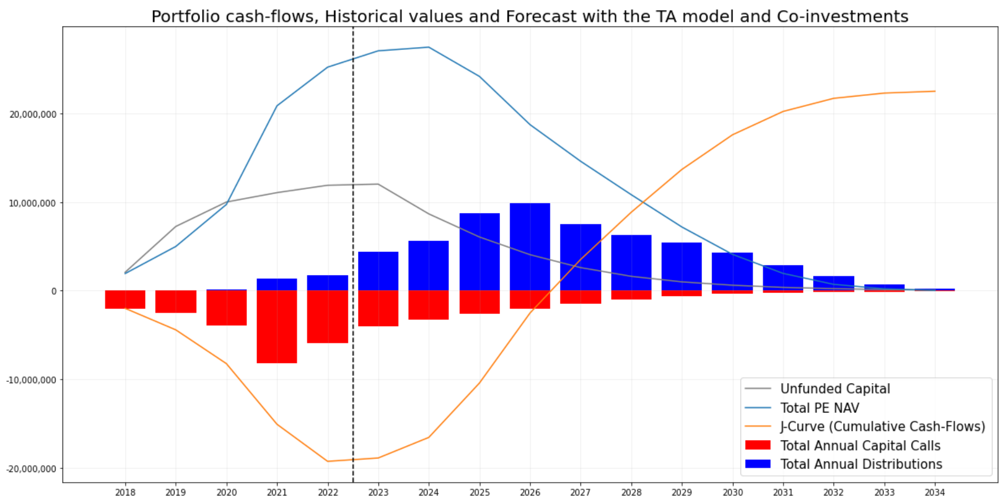


Figure 28 – Forecast of a sample portfolio, considered as standard

And voila ! We recognize here the J-curve, with a peak of distributions in 2026 due to a high number of co-investments made in 2021 (exit after 5 years here). A portfolio of private funds is simply a linear combination of the private funds ; it is thus expected that the portfolio will exhibit the J-curve pattern but more extended in time (but not much deeper, as it should be as deep as the average of the funds in the portfolio).

4.2.1 Commitment planning - constant for illustration

At this point, we can add a simple commitment planning to the model, constant in this case, to simulate a true PE program. Each year, we will allocate money to one strategy (5m each year in Buyout, 2m in Private Credit and 2m in Co-Investments for example), and we simulate with this commitment the expected cash-flows according to parameters set in part 4.1. We then merge the result every year, resulting in a full private markets allocation forecast. The parameters are static here so every commitment (respectively of the strategy) is planned in the same way, but we can easily change the parameters according to their dispersion as shown in part 3.4. The next part focuses more on this aspect.

We see that the NAV is relatively constant over time, as it should because the parameters are fixed (even the commitment). Then, we could add this forecast to a multi-asset portfolio, and discuss the construction of this type of portfolio (commitment based, NAV based, target return...) ; but this is not the subject of this thesis.

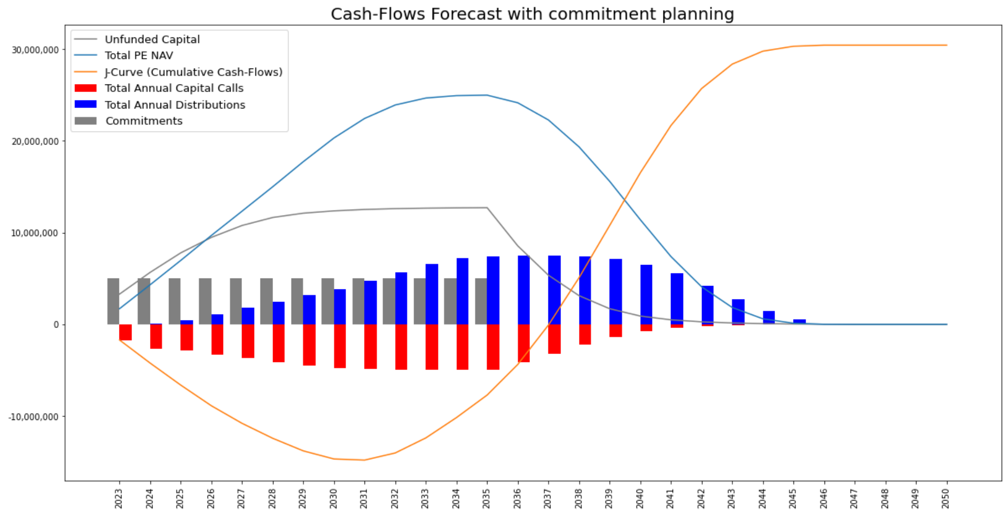
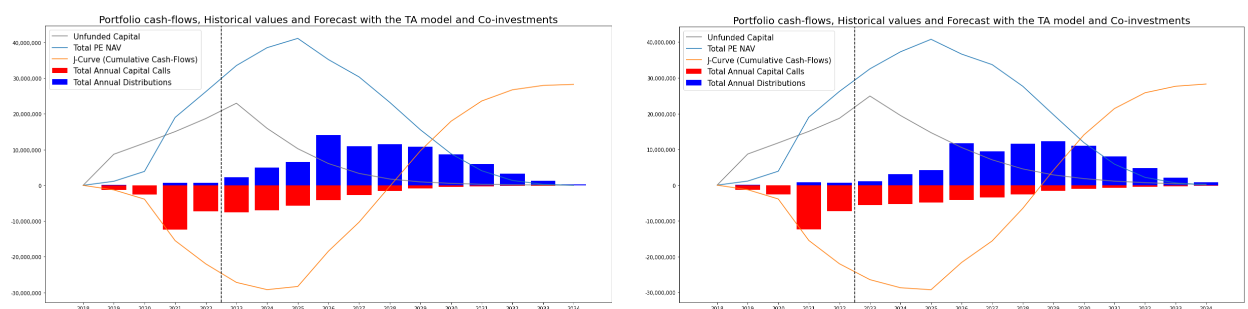


Figure 29 – Forecast of a constant commitment strategy over 15 years

4.2.2 Stress testing

One of the main advantages of the Yale model is its ability to do scenarios analysis, by varying its parameters. We have demonstrated the high variability between vintages in the section 3.4. If we want to simulate an adverse scenario, we can for example take the bow estimations of the vintages 2004-2008 and apply them to the vintages 2018-2022 to forecast a big delay in the distributions. We can also delay the capital calls at the same time by varying the hook but only for the 2022 vintages, implying that the lack of M&A dynamism in the first half of 2023 will negatively impact the capital calls. Finally, we didn't touch the TVPI here, by making the assumption that the underlying companies will be sold eventually at the same price but later in time ; we are more interested in stressing the J-curve convexity than its final destination.

For visualization purposes, we forecasted and stressed the sample portfolio without adding future commitment, with the final objective of comparing only what has been invested already.



(a) Sample portfolio forecast with no stress testing

(b) Sample portfolio forecast with stress testing

Figure 30 – Difference between normal and stressed sample portfolios

In the stress test, the J-curve is effectively delayed and becomes positive in 2029, one year after the standard forecast, due to later capital calls and later distributions, with a NAV that stays high for longer. Note that we didn't change the assumption of co-investment here, even though it represents a significant part of the portfolio ; putting the exit year at 8 instead of 5 after the investment would delay even more the J-curve.

Bad macro-economic conditions will affect the aggregated J-curve for several years, but this stress test shows that vintage diversification is essential to mitigate the effect of very bad years. Moreover, we only changed the timing of cash-flow here and not the size : as seen in section 3.5.5, the worse vintages in 2005-2008 have been succeeded by very good vintages (helped by the longest bull run of public markets in history).

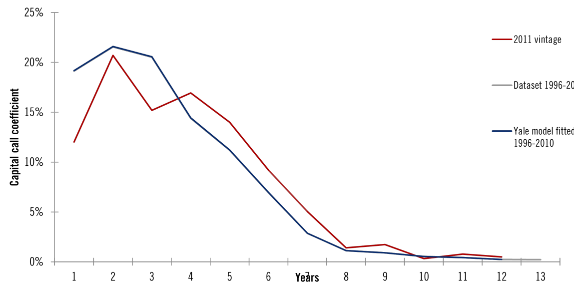
This effective framework allows us to prove the utility of the Yale model for stress testing and scenario analysis ;it can also be adapted to future commitment strategies, using the same methodology.

4.3 Back-testing

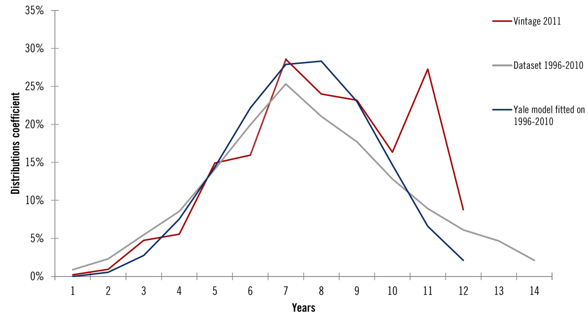
4.3.1 Back-testing on patterns

To see if the model behaved well on historical data ex-post, we used back-testing. The idea is to compare the patterns of a specific vintage to the pattern based on historical averages that happened before this vintage, and the Yale model fitted on the same historical data. In this case we have the in-sample dataset used to fit the model, and the out-of-sample analysis to evaluate the performance of the model. We selected the vintage 2011 for testing purposes ; it has the advantage of being a good vintage (not a crisis vintage) and with sufficient points in the pattern.

The process for computing a backtest is the following : first, we compute the average patterns on the vintages 1996-2009 of capital calls and distributions divided by the commitment. We normalize these patterns (both capital calls and distributions), and multiply the distribution pattern by 1.5 to be comparable with the Yale model, and because 1.5 TVPI is the conservative (and standard) assumption of the Institution for LBOs. Then, we fit the parameters of the Yale model to the same dataset. The pattern for the capital calls would be exactly the same because the RC is fitted on the CC pattern. However, the pattern of distributions is quite different : the bulk of the Yale model's distributions arrive in year 8 instead of 7 for the average pattern, because the latter is flatter and with still high distributions in year 12-13-14. The Yale model being terminated in year 12, the distributions will be more concentrated than the average but still show the positive skew of the average pattern. Finally, we compare the resulting patterns to what really happened in the vintage 2011 in the figure 31. Other strategies are compared in appendix 7.3.5.



(a) Capital calls - Backtest 2011



(b) Distributions - Backtest 2011

Figure 31 – Average LBO pattern compared to Yale and vintage 2011

The vintage 2011 saw late capital calls, and thus later distributions than expected ; the difference between the average, fitted model and the historical cash-flows is thus quite high. Below are the MSE computed on both model for every strategy (still vintage 2011). [20]

Table 5 – MSE of Averages vs Vintage 2011 & Yale Output vs Vintage 2011

Capital calls	LBO		PC		RE		VC	
	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale
MSE	0.0105	0.0106	0.0079	0.0111	0.0060	0.0078	0.0119	0.0111
Distributions	LBO		PC		RE		VC	
	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale	Avg vs 2011	Avg vs Yale
MSE	0.0434	0.0541	0.0188	0.0178	0.0514	0.0303	0.0967	0.1786

On the capital calls, average patterns and the Yale model perform very similarly for LBOs because it is computed with the same data, and the length of the pattern is 12 in both cases. The small differences for private credit, real estate and venture capital is because the length is less for the Yale model (respectively 11, 10 and 11 - fitted on 1996 - 2010 data), thus compressing the pattern more than the average which is always fixed at 12 ; see appendix 7.3.5 for the graphs of capital calls for the other strategies. The 2011 pattern called less than the average, explaining the differences in MSE for capital calls ; if the year called more than expected, then having a more compressed pattern would have been better. Putting the L as the length of the average pattern could be a solution.

There are more differences on distributions. The Yale model performs better for PC and RE, and worse for VC and PE ; for the latter two, it is due to the unusual pattern of the 2011 vintage with huge distributions in year 11 (the 2021 effect) leading to high residuals : 79% of the Yale's MSE is in this year for buyouts. For

20. As a reminder, MSE stands for Mean Squared Error and is one of the preferred measure to compute the precision of a model. The equation is :

$$MSE = \sum_{i=1}^{n} (\bar{x} - x_i)^2 \quad (17)$$

With $\bar{x}$ the observed coefficient and x_i the modelled coefficient.

PC and RE, the Yale model was able to predict better than the averages because the patterns looked more like their usual distributions. Again, see appendix 7.3.5 for the graphs by strategy.

This analysis is not enough by itself to say if one model performs better than the other, especially because both are fitted to the same historical data ; replicating the analysis on another vintage could lead to very different results. We could also compute other measures of fit than the MSE, but it would not change the conclusions of this analysis :

Where the model truly shines is in its ability to change the parameters according to different scenarios ; for example, increasing the bow to account for high distributions in one of the last years of the fund is one the way to approximate the 2011 vintage scenario. By doing a grey sky and blue sky scenario, we'll actually predict the range of possibilities in which the true pattern should be found.

4.3.2 Backtesting on a sample portfolio

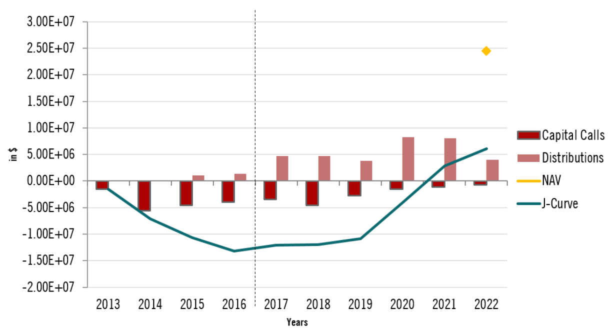
Where the model also shines is its direct application to a portfolio, as shown in section 4.2 where we used it on a sample portfolio. Forecasting is easy ex-ante, but how does this forecast holds up ex-post ? We had access to another sample portfolio, that was for the sake of the exercise created in 2013, with both :

— the data as of 2016

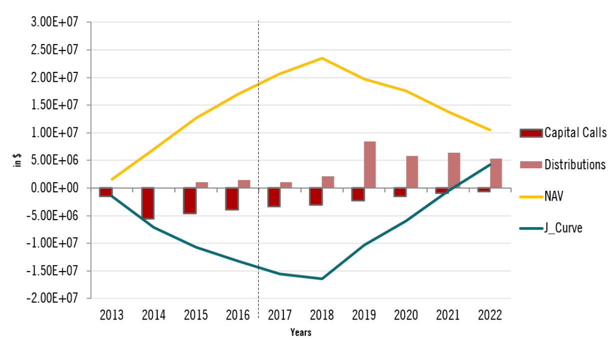
— the data as of 2022.

The sample portfolio is composed of 10 leveraged buyouts funds, and 7 co-investments ; this number is enough to test our model.

We thus forecast the 2016 sample portfolio using the Yale model (usual parameters) and compare the (realized) results of the sample portfolio up until 2022.



(a) Portfolio benchmark



(b) Portfolio forecasted

Figure 32 – Forecasted vs Benchmark. Application to a sample portfolio

We can see that the model has succeeded in modeling the elements of the portfolio and is pretty close to reality, except for the NAV. A good point is that the capital calls pattern follow closely the reality ; the differences have more to do with the distributions. The model forecasted low distributions in 2017-2018, much higher in 2019 (due to co-investments exiting) and then stable around 5m for 3 years. In reality, the big jump in distributions is not in 2019 but in 2020, followed by another in 2021, while 2022 has seen a big drop in realizations. We can also note that the true portfolio distributed more than forecasted in 2017-2017-2018.

The true J-curve becomes positive in 2021, while the forecast curve is just below 0 in 2021, but catches up in 2022, when both J-curves are at 2m apart (the forecast J-curve in 2022 reaches 4.22m, compared with 6.08m for the actual data). In fact, the big difference is not really the J-curve but more the NAV, because in 2022 the portfolio has a much higher NAV than what is forecasted, indicating further distributions down the line. This is due to some co-investments that have not yet distributed, and potentially underlying companies that are now public. These are very good news for this example portfolio, and it means that the total distributions will be higher than the average PE program.

We demonstrated here that the Yale model exhibits robustness when compared to historical data. The model accurately approximates the actual distribution of cash flows. When applied to an extant portfolio, its performance stands up well to scrutiny, even in portfolios with a significant proportion of co-investments, which inherently possess a binary nature.

5 Discussion

This section discusses the relevance of the Yale model, its limitations and several points and subjects of research linked to the cash-flow modeling in private investments. In particular, we look at a potential stochastic implementation for risk modeling and a clustering approach to discuss the optimal numbers of strategies.

5.1 Relevance of the Yale model

Despite its age, the Takahashi-Alexander model still serves as a foundational framework upon which many newer models are built or compared. It captures the behavior of private investment patterns well enough, while remaining simple enough to be understood by all even in an era where stochastic models (which account for a wider array of possible outcomes and risks) have gained prominence.

As the model only need at one point in time the NAV and the unfunded capital to forecast the rest of the fund's life, the main advantage of the Yale model is in fact its versatile adaptation to existing funds/cash-flows. Furthermore, The dispersion analysis of its parameters allows stress testing : for example, Changing the parameters of the model to stress existing assumptions, especially for scenario analysis like the GFC. Here, one could take the different bows/TVPI (or even hooks) between 2005 and 2008, and putting them as the bows for vintages between 2018-2022 (or even in the future of a commitment strategy), to see like we did in section 4.3.2 how a sample portfolio would react, and adjust the commitment strategy if needed.

It is also possible to use the Yale model for different strategies, like we did in our study. The parameters change according to the behaviour of each strategy, and by extension the J-curve will also change. In section 5.3, we extend this analysis of different j-curves by trying Monte Carlo analysis.

However, the model is not perfect, and they are some limitations that one must consider before adopting this model. We dwell further on this subject in the next section. Nonetheless, it is interesting to note that nearly 23 years after its creation, the Yale model is still relevant for cash-flow modeling needs.

5.2 Limitations

There are some limitations to the model, especially with longer funds than expected, a TVPI already reached, or limitations inherent to the Yale model.

One of the limitation is when an existing fund is already above the selected Length, the rate of distributions will be higher than 1 (remember that RD = t/L), causing $D_T > NAV_t$ and NAV_{t+1} to be negative. If we go on like this, the next distribution will be negative, the next one positive, then negative, etc ; and the model would simply break. To counter this problem, we have several solutions when t > L :

— Either we distribute all remaining NAV in time t+1 without a RD, effectively deleting the problem but causing a big cliff in distributions (especially for funds with still a very high portion of their NAV remaining)

— Or, we smooth the distributions on several years, by also deleting the RD but this time distributing the remaining NAV over several years (like 1/3 in t+1, 1/3 in t+2, 1/3 in t+3) - with or without growth rate, depending on a conservative assumption). This would smooth the cliff.

Another limitation is when the TVPI is already higher than the set parameter, as the model would expect a negative G value to reach the entered TVPI. Here, we simply check the state of the fund at first, and if it's already above the expected TVPI, we set the G to 0, effectively distributing the NAV as of today without further growth.

In the industry, forecasts are often quarterly, as this is also in line with the frequency of the investment reporting. [21]. It is easy to adjust the Yale equations to be quarterly, as shown by [27] ; but as the model is continuous in essence, it will always predict some cash-flows during each period. As there can be long periods during the years (even several quarters) of non-activity, (capital calls or distributions), implementing some periods with 0 cash-flows can be an extension of this model as de Malherbe said (for his model) : *A natural extension would be to include some jumps that could be helpful to tackle some of the extreme cash flow movements that are sometime observed. The model would lose in parsimony but not in tractability*.

Finally, one of the biggest limitation of the model is the fact that contributions and distributions are modeled jointly, unlike the stochastic model for example. This is not a problem for the contributions as they are only defined by the RC ; but as the distributions are in function of the NAV and the NAV depends on capital calls, the distribution pattern will always be in function of the first pattern, and by extension the RC, even though the contributions and distributions may not always move in tandem or be influenced by the same factors. This means that when we estimate the bow, we have two choices :

— Either, for each year and each fund we take the corresponding RC of the fund, and estimate the bow next.

— Or, we take always the same RC for every fund of the dataset, regardless of the state of the economy.

We chose the second approach so that every bow is comparable ; a bow calculated with the second method

21. At least for short-term projections, LPs may also need monthly forecasts.

would be difficult to be interpreted considering that every fund has a different RC. A more robust method to address this problem can be the subject of further research.

5.3 Stochastic implementation

The stochastic model presented by Buchner et al. [22] has the advantages of giving several outputs for one input. This is especially useful for risk modeling, where we are interested in the worst possible scenario, similar to the VaR used in the banking sector. However, its difficult implementation and parameter calibration has prevented us from putting it to concrete use. The basic idea behind was still relevant : can we obtain a theoretical distribution of all the possible J-curves making up a portfolio?

We tried to answer this question by setting up an easy to understand and implement framework, still using the Yale model. We implemented a Monte Carlo simulation, with a portfolio as follows : we have over a 10 year period one commitment per year, gradually increasing to take account of a (theoretical) increase in the liquid portion of the portfolio (the growth rate of the commitment is not important here). For each vintage, we can thus make a forecast by selecting a set of parameters (B, TVPI, H, L) ; we merge every forecast to obtain the J-curve of the total portfolio.

By selecting a random set of parameters for every vintage in the portfolio, we get a random J-curve. Here, we select the parameters in an uniform distribution (with large enough number it becomes a normal distribution - see central limit theorem) : the TVPI taking random value between 1.25 and 1.75, the bow between 1.5 and 3.5, the Length between 11 and 14 (taking only integer values) and the hook between -0.2 and 0.2.

We repeat this process 10'000 times to get 10'000 different j-curves, and this is the result :

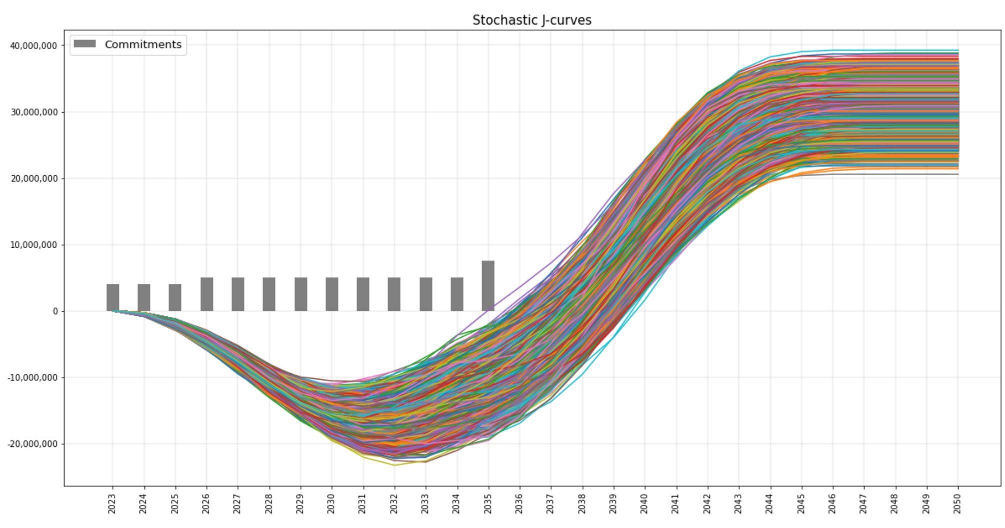


Figure 33 – Monte carlo simulation of J-curves using the Yale model

All of the j-curves represent a different path, some with quicker and higher distributions, some with higher capital calls but later distributions, etc. This simulation shows that the J-curve becomes positive between 12 and 17 years of investing, and the value of final assets is between 20 and 40 million, i.e. almost double.

Of course, this is a naïve implementation of the parameters : in real life, there will be high correlation between the bow and the hook for example, when in times of crisis no acquisitions are made in the leveraged buyout world. A more robust method of stochastic modeling by computing the correlation matrix between coefficients is outside the scope of this thesis, even though we demonstrated here that it was possible using the Yale model.

5.4 Optimal numbers of patterns

5.4.1 Regression approach

The Institution was also interested in resolving a simple question : How many patterns and different set of parameters should we take into account ? Is separating strategy sufficient, or should we separate depending on fund size, geography, performance resilience, subs-subs strategies (VC early/later, RE opportunistic/value-add), etc.

This is not an easy question to answer. One possible way of answering it would be cluster analysis, which involves dividing unlabeled data or data points into different groups, so that similar data points fall into the same group as those that differ from others. Several studies were done with clustering, on individual stock [33], mutual funds ([34]) or hedge funds ([35]) but also with private assets ([36]). This particular study focuses on private equity factors return, similar to the famous Fama and French 3 factors model : "Estimating private factors requires reducing the dimensionality of the private fund universe to a small set of portfolios that summarize well the dispersion in returns and risk exposures ".

They found 8 private equity factors minimizes the auto-correlation between clusters. However, this particular study is interested in private fund returns, whereas this thesis is more focused on the cash-flows timing and modeling. To reduce the dimensionality of our data, we introduce a notion called average year of cash-flow. The idea is to pass from 2 dimensional data (coefficient, year) to a single point. For example, we have the following datapoints (random numbers) :

Table 6 – Example of the sum product result

Année	1	2	3	4	5	Sum product
Coef	10%	21%	28%	19%	12%	2.72

The sum product of years and coefficient yields a single point, which can be defined as the average year around which the coefficients (here, the capital calls) are located.

We then compute it for every fund in the database. However, to be comparable across funds of different size of capital calls/distributions we divide the sum product by the sum of the cash-flows. Note that t = 12 for capital calls and t = 15 for distributions. If the result is negative due to bad reporting, we don't take into account for the average. The full equation in the following :

$$YWACF = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ t \end{bmatrix} \cdot \begin{bmatrix} CF_1 \\ CF_2 \\ CF_3 \\ \vdots \\ CF_t \end{bmatrix}}{\sum_{i=t}^{12 \text{ or } 15} CF_t} \tag{18}$$

YWACF stands for Year-weighted average cash flow. We can also make the distinction between AYCC (Year of Average Capital Call) and AYD (Year of Average Distribution).

Contrary to the coefficients patterns computed on the fully cleaned dataset, we apply here another layer of cleaning by selecting only mature funds (defined as having at least 80% of the fund size called) and only terminated funds (defined as a fund with a NAV in 2022 < 10% of the initial fund size). This step is necessary to avoid recent funds that can skew the data and yield a not interpretive result. The result of this cleaning is presented in appendix 11.

As a result, we obtain one point for capital calls and one for distributions. The following graph presents the average across strategies to illustrate the first differences between strategies for the timing of capital calls :

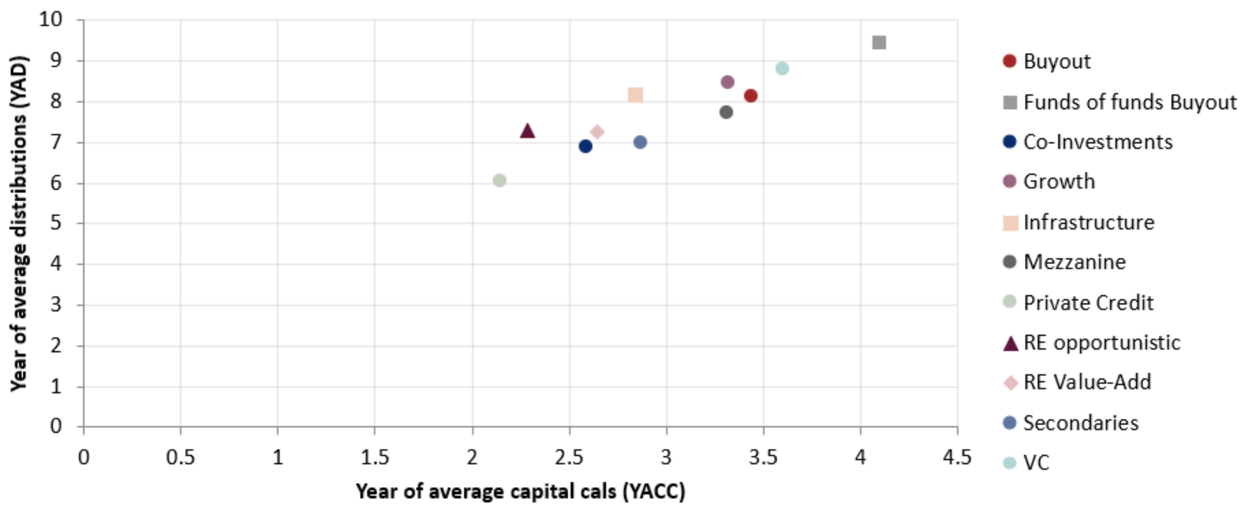


Figure 34 – Scatter plot of average year of cash-flows

There seems to be a clear pattern across fund types in when they call and when they distribute, and the two are positively correlated. Then, we can extend this analysis to not only capture the differences between strategies but also across fund sizes, subs-subs strategies (VC early/Late), their performance (noted by the IRR) and geography.

The equation is thus :

$$YWACF = \alpha + \beta * \text{fund types} + c * X + e \tag{19}$$

With α a constant, β the factor of fund type and X the factor of other determinants.

As the strategies are expressed as text, we need to transform them as categorical data : if the fund is an LBO, the LBO column will have 1 as value and the others strategies 0. We will thus have a vector of dependent values and a corresponding vector of independent variables, upon which we will conduct a regression analysis.

Given the high number of variable, we first do the regression only with fund size and fund strategy/substrategy. The results of the regression is shown in the appendix, for AYCC and AYDD (respectively figures 43 and 44).

The result is mitigated. For the capital call regression, the R^2 is not very good at 0.154, but several parameters are significant statistically at 5%. First, the constant is equal to 2.93, a value coherent with the usual behaviour of capital calls. The fund size variable is not significant, but that can be explained by the high existing variability between strategies ; do again the regression but this time only on one strategy could give better results.

Strategies like buyout and VC early, mezzanine, growth/expansion, VC general and VC later are all statistically significant and with a positive coefficient, meaning that they are strategies that call later. Of them, mezzanine, growth and VC later have lower coefficient (0.56, 0.59 and 0.57) than buyout at 0.8, VC general is close to buyout at 0.73 and VC later is the highest with a coefficient equal to 1.4. On the opposite side, strategies like distressed lending/debt and credit special situations are significant with a negative coefficient, meaning that they will call faster (confirming what we saw in the data).

The distribution regression follows the same path, with strategies like buyout, mezzanine, growth, VC early/general/later that have positive coefficients, but also infrastructure opportunistic/core that are here significant (VC early has the highest coefficient, equal to 2.86). And the quicker strategies are still direct lending, credit special situation, and here RE debt.

These regressions already yield good results, because they captures the difference in existing strategies. They however have their limits, as shown by all the variables statistically insignificant even if there is a high number of funds, like real estate ; more importantly, the R^2 is relatively low. Adding more explicating variables (like geography) would not give better results in this regard, because then the number of funds for example would be too small for the majority of strategies.

5.4.2 Clustering approach

Maybe we should then take another direction. Here, we will aim to apply clustering, a data analysis methodology designed to identify inherent groupings, termed clusters, within a dataset. Notably, cluster analysis does not require categorization into predetermined groups, signifying its nature as an unsupervised learning technique.

First, we begin simply by plotting in figure 35 the coefficients of every fund in the dataset, with their strategies as colors and fund size as point size.

As the chart is not very clear, we get rid of the strategies with less than 10 funds in it. Then, they are several

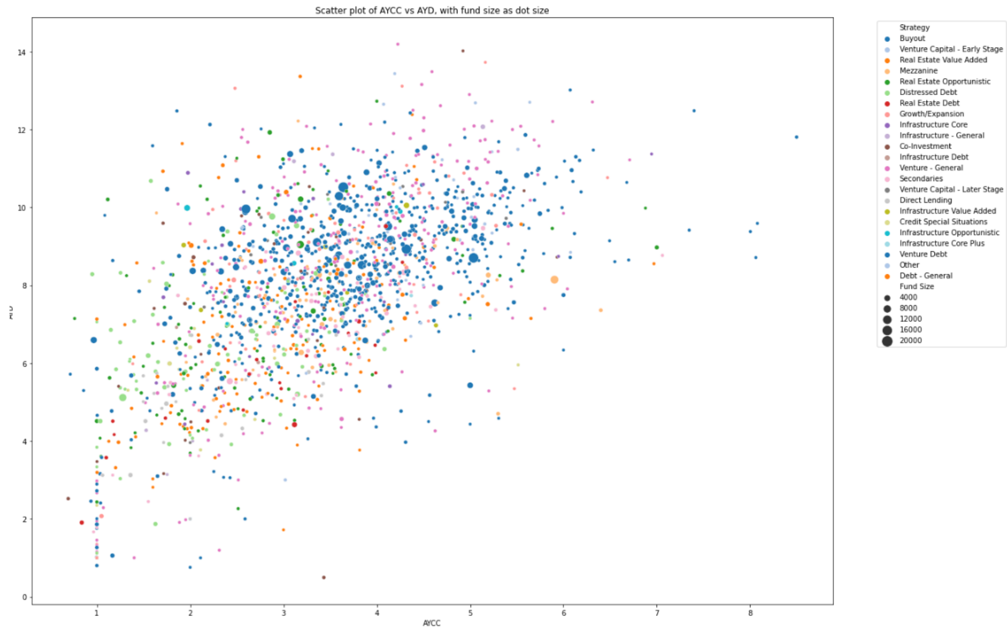


Figure 35 – Scatter plot of YACFs

possible clustering algorithms, like K-means or hierarchical clustering ; we tried both and they would not give very different clusters. We thus focused on K-means clustering, a partitioning method that divides a dataset into k distinct, non-overlapping subsets (or clusters) based on their distances to the mean of the cluster, aiming to minimize the variance within each cluster.

As Kmeans is sensitive to scale, we first we need to standardize the data [22]. Then, we compute the optimal number of clusters by calculating the sum of squared distances for different number of clusters solutions, denominated by inertia [23]. The result is shown in figure 36. To determine the optimal number of clusters, we have to select the value of k at the “elbow” ie the point after which the distortion/inertia starts decreasing in a linear fashion. Thus for the given data, we conclude that the optimal number of clusters for the data is 5. In any case, selecting higher number of cluster did not give insightful results.

We then use the K-means clustering algorithm with 5 clusters. Each point is assigned to a cluster, and we can plot it again like figure 37 :

The result is not very clear, but he 5 clusters are recognizable, and we can analyze them as follows:

— In green, this could be the cluster of statistical outliers, as illustrated by the number of points with a YACC equal to 0.
— In red, funds that call quickly/distribute quickly, and which should therefore be strategies linked to private debt or even real estate.
— The purple and orange clusters are more difficult to interpret, since they have approximately the same abscissa; the storm funds have nevertheless called up capital perhaps a little later but distributed it

22. The library ’sklearn’ in python was used for all the clustering analysis.
23. More precisely, inertia is the sum of the squared distances of samples to their closest cluster center.

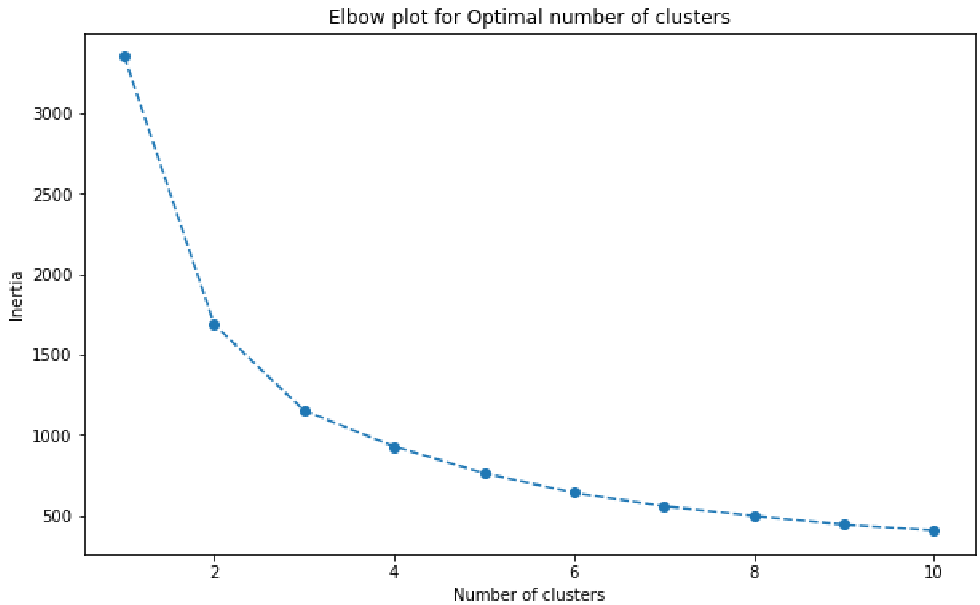


Figure 36 – Elbow plot

faster, and the purple funds are the exact opposite.

— Finally, the blue cluster represents all the funds that called very late and distributed just as late; logically, we should find equities strategies in it, like VC.

By visualizing solely the strategies, while retaining the identified clusters, we aim to validate our preliminary insights. Figure 38 illustrates this perspective.

The distinction emerges here with more clarity. Initially, it's evident that Buyout dominates, representing a significant majority of the data points, and spans across all clusters. This heavy representation of the Buyout strategy may introduce a bias in our interpretations in contrast with other strategies; however, several observations still stand out. it becomes evident that it predominantly aligns with equity strategies—namely LBO, VC, and Growth - with a slight presence in Real Estate Opportunistic. ; In contrast, debt strategies such as Distressed, Real Estate Debt, Direct Lending, and Credit Special Situations are noticeably absent from this cluster. The Mezzanine strategy is intriguing; it exhibits a scarcity of violet data points, suggesting faster capital calls and distributions (eventh though still slower than private debt). Real Estate Opportunistic and value add seems also to call relatively quickly, but it is more difficult to draw conclusions given that they have several clusters. Several strategies lack the volume of data points necessary for discerning clear patterns.

Considering the pronounced cyclicality highlighted previously, it would be insightful to investigate cluster evolution in relation to vintage years. But again, we don't have simply enough points to extract anything meaningful for several strategies. From this analysis, one could infer the existence of at least three distinct patterns: those for the private debt sector, those for equity strategies like LBO/VC, and those for intermediary strategies such as Real Estate and Mezzanine.

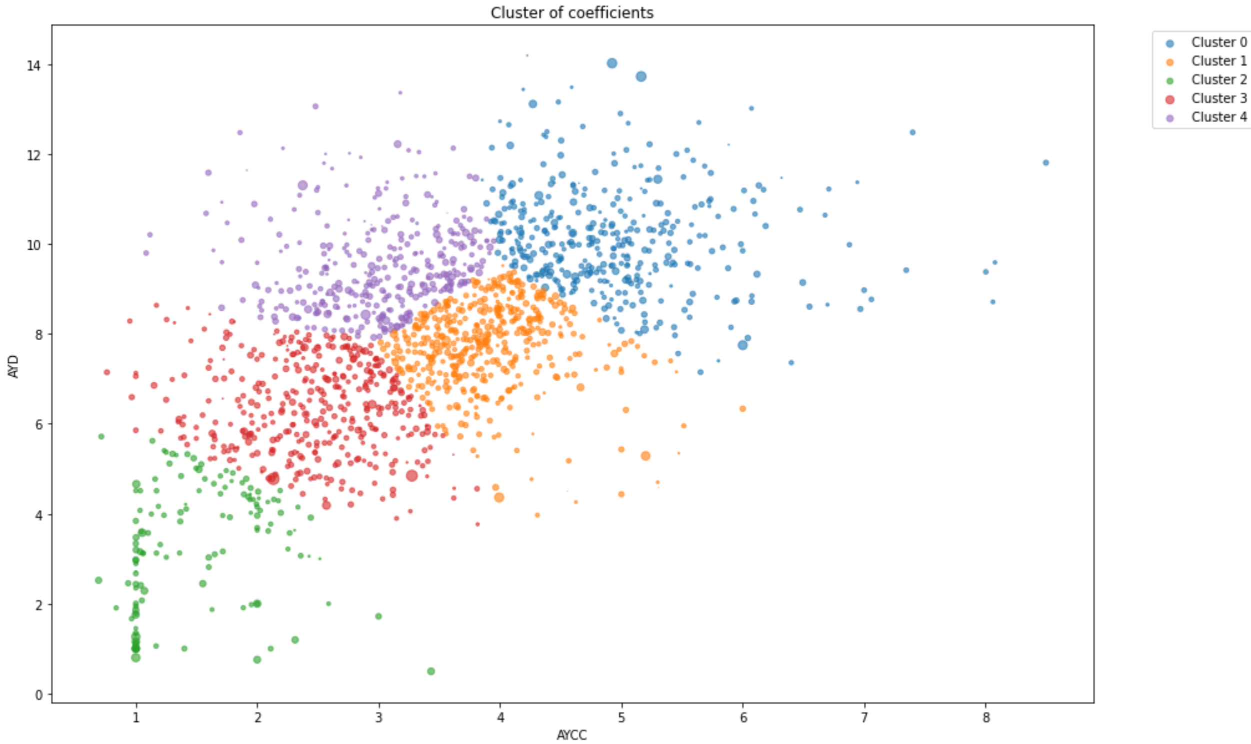


Figure 37 – Results of clustering - full dataset

Finally, we can also add a TVPI axis to the scatter plot, to see if different clusters would give different performance as shown by figure 39.

There seems to be no any titled plan for any strategies. A tilted plan would that distributions happening later in time would be of higher size, if the underlying investments had had more time to grow. But as we shown earlier with the bow and TVPI analysis, there is no positive correlation between the two ; the clustering approach allows us to see it again from another perspective.

Any further analysis was deemed impossible due to the lack of data. It is nonetheless very encouraging that we managed to perform such an analysis on this sector, and shows how much the data access has evolved. Doing the same analysis in 5 years' time will give much more clear-cut results to answer the question of the number of patterns; for the moment, we see at least 3 (PC/RE/LBO), with perhaps another distinction between buyout and VC.

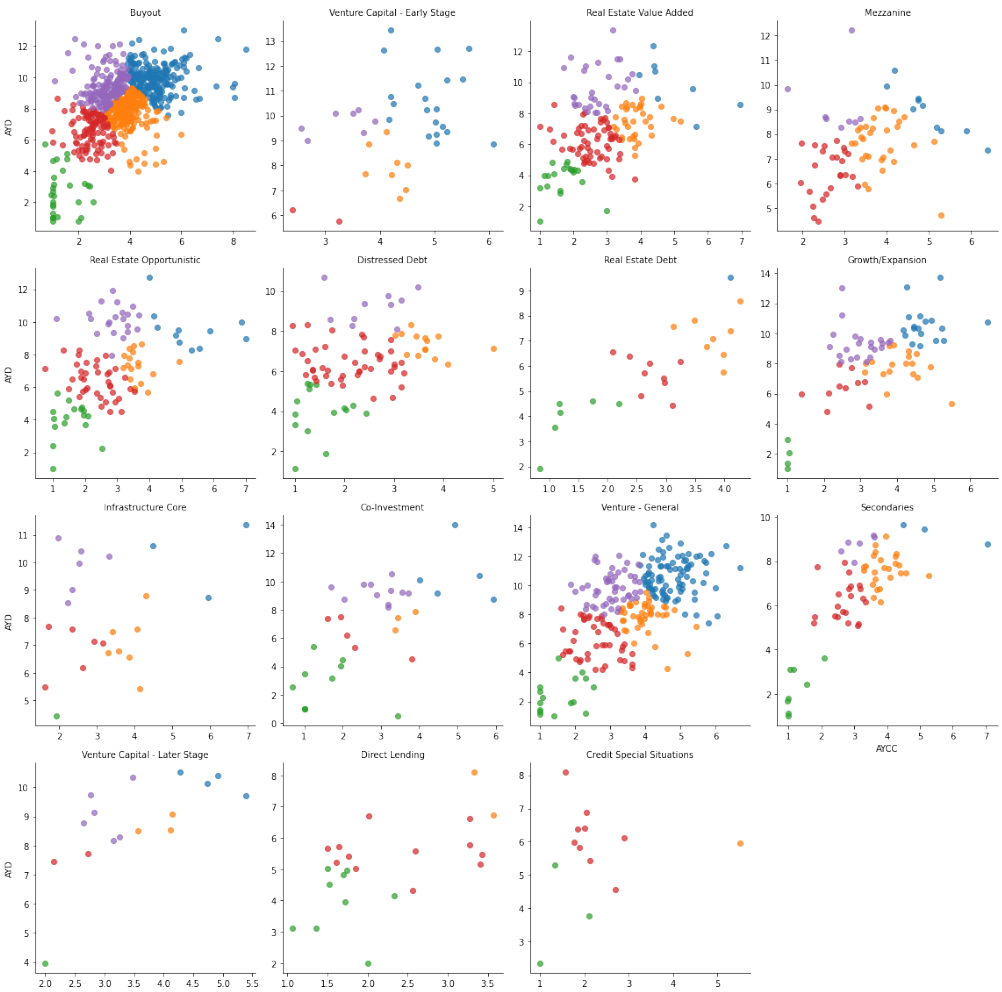


Figure 38 – Results of clustering - by strategy

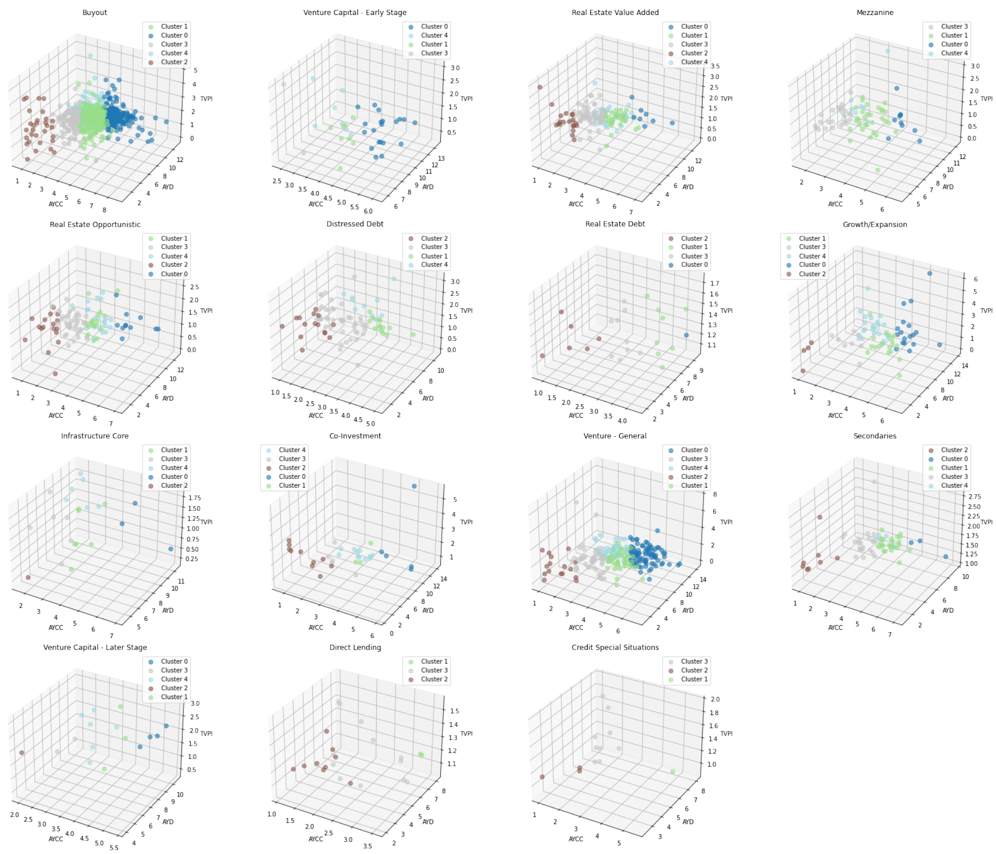


Figure 39 – 3D clustering with TVPI as Z axis

6 Conclusion

The purpose of this thesis was to review the Institution's capabilities in cash-flow modeling applied to private investments. Using PitchBook's database and after a comprehensive data cleaning, we began the analysis of contributions and distributions patterns, and showed the differences between strategies. This analysis is done on 4 different strategies : LBO, VC, RE and PC. However, we found that the historical averages were not sufficient to capture all of the dispersion found during the selected time period (1996-2022).

We then compared the various existing approaches and chose the Yale model as the basic framework for continuing our analysis. We estimate its parameters on the PitchBook database, and demonstrate that the parameters evolve over time and across different macroeconomic conditions, thus laying the foundations for a robust stress-testing methodology. Like the historical averages, we also compute the parameters for all selected strategies.

However, the initial model had a number of limitations, some of which had already been addressed in the literature, and some of which were new. As the Institution wished to emphasize its liquidity needs, the original form of the model was not sufficient as it lacked a parameter for modifying the contribution rate. In this thesis, we propose a parameter called Hook designed specifically to modify the timing of cash calls. We then compared historical cash flows to the patterns calculated with the Hook, and saw that this parameter only partially captured dispersion. To this end, we are also proposing another parameter (here a multiplier) on the contribution rate, making it possible to simulate peaks (positive or negative) or higher economic activity.

Another limitation concerns the growth rate of NAV, a topic already partially addressed in the existing literature, but we didn't find the existing solutions convincing as the NAV is considered in the industry to be the least trustworthy and reliable measure. We therefore propose a way of selecting an expected sum of distributions instead of a specific growth rate, an assumption consistent with historical data. We also introduce a way of implementing and forecasting co-investments in the overall model.

After presenting the result of our parameters estimation, we implement a framework based on the (improved) Yale model for the institution. This model can use fund-level data to forecast an existing portfolio, and the parameters can easily be modified to account for different assumptions. As the final objective of the internship was to propose a pilot tool to be implemented for cash flow modeling purposes, we consider this thesis accomplished.

Finally, we have several suggestions for further analysis in this area. Using the Yale model, we have tried to implement Monte Carlo analysis to encompass all the different possible scenarios. As this analysis was carried out without taking into account the correlation between the coefficients, it would be interesting to examine this correlation (and auto-correlation) where appropriate. Further research could be carried out on the regression of parameters on different macroeconomic data, in order to test the model's predictive power.

For the author, a significant discovery was the extensive size of the database used, which was far superior to all previous studies on this subject. With the increasing availability of PE dataset of higher quantity and quality, and with recent developments in methodology, this looks to be a promising area for future research.

References

[1] Bosut. *Private Equity and Venture Capital*. Tech. rep. AltAssets, 2003.

[2] Pierre-Yves Mathonet and Thomas Meyer. *J-Curve exposure : Managing a portfolio of venture capital and private equity funds*. 2nd ed. Wiley Finance, 2007.

[3] Thomas Meyer and Pierre-Yves Mathonet. *Beyond the J-Curve : Managing a portfolio of venture capital and private equity funds*. 2nd ed. Wiley Finance, 2005.

[4] Tjibbe Hoekstra. « 'Painful' private equity fees are hard to avoid ». In: *Investment and Pensions Europe* (2022).

[5] Chris Meads, Nik Morandi, and Andrea Carnelli. *Cash Management Strategies for Private Equity Investors*. Tech. rep. Pantheon, 2016.

[6] Cumming D. J. et al. *Venture capital and private equity research: A bibliometric review and future research agenda.* 2022. DOI: 10.2139/ssrn.4034812.

[7] S.N. Kaplan and Per Strömberg. « Leveraged buyouts and private equity ». In: *Journal of Economic Perspectives* (2009).

[8] Steven Kaplan and Antoinette Schoar. « Private equity performance: Returns, persistence, and capital flows ». In: *Journal of Finance* (2005).

[9] Arthur G. Korteweg. *Risk Adjustment in Private Equity Returns*. 2018. DOI: https://ssrn.com/abstract=3293984.

[10] Christian Diller, Ivan Hergera, and Marco Wulff. *The private equity J-Curve: cash flow considerations from primary and secondary points of view*. Tech. rep. Capital Dynamics, 2009.

[11] Alexander Ljungqvist and Matthew Richardson. « The investment behavior of private equity fund managers ». In: (2003).

[12] David T. Robinson and Berk A. Sensoy. « Private equity in the 21st century: Liquidity, cash flows, and performance from 1984-2010 ». In: *NBER Working paper* 17428 (2011).

[13] Robinson and Sensoy. « Cyclicality, Performance Measurement, and Cash Flow Liquidity in Private Equity ». In: *Journal of Financial Economics* (2015).

[14] Dean Takahashi and Seth Alexander. « Illiquid Alternative Asset Fund Modeling ». In: *The Journal of Portfolio Management 28* (2002).

[15] Vishv Jeet and Luis O'Shea. *Calibrating the Parameters of Takahashi and Alexander's Model*. 2020.

[16] Vishv Jeet and Luis O'Shea. *Calibrating the Parameters of the TA Model to Recessionnary periods*. 2020.

[17] Hank Hoek. *An ALM analysis of private equity*. 2007.

[18] Coen Tolkamp. *Predicting Private Equity Performance*. 2009.

[19] Vishv Jeet. *Modeling private investment cash flows with market-sensitive periodic growth.* PGIM IAS. 2020. DOI: https://dx.doi.org/10.2139/ssrn.3775008.

[20] Kieffer E. et al. *Learning private equity recommitment strategies for institutional investors*. 2023. DOI: doi:10.3389/frai.2023.1014317.

[21] Etienne de Malherbe. « Modeling private equity funds and private equity collateralised fund obligations ». In: *International Journal of Theoretical and Applied Finance* (2003).

[22] Axel Buchner, Christoph Kaserer, and Niklas Wagner. « Modeling the Cash Flow Dynamics of Private Equity Funds: Theory and Empirical Evidence ». In: *Journal of Alternative Investments* (2010).

[23] Oscar Ungsgård. *Stochastic Modelling of Cash Flows in Private Equity*. 2020.

[24] Axel Buchner. « Risk management for private equity funds ». In: *Journal of risk* (2017).

[25] Elias Furenstam and Johannes Forsell. *Cash Flow Simulation in Private Equity*. 2018.

[26] Jussi Virtanen. *Cash flow simulation and model comparison in private credit*. 2021.

[27] Tugce Karatas, Federico Klinkert, and Ali Hirsa. *Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting*. 2021. DOI: https://doi.org/10.48550/arXiv.2108.02853.

[28] Tausch C., Buchner A., and Schlüchtermann G. « Modeling the exit cashflows of private equity fund investments ». In: *Journal of Risk* (2022).

[29] Steven Neil Kaplan and Josh Lerner. *Venture Capital Data: Opportunities and Challenges*. 2016. DOI: https://ssrn.com/abstract=282330.

[30] John W. Tukey. *The future of data analysis*. The annals of mathematical statistics, 1962.

[31] Arpit Gupta and Stijn Van Nieuwerburgh. « Valuing Private Equity Investments Strip by Strip ». In: *Journal of Finance* (2021).

[32] Stephannie Larocque, Sophie Shive, and Jennifer Sustersic Stevens. « Private Equity Performance and the Effects of Cash-Flow Timing ». In: *Journal of Portfolio Management* (2022).

[33] Brown S. J., W. N. Goetzmann, and M. Grinblatt. *Positive Portfolio Factors*. 1997.

[34] S. Brown and W. Goetzmann. « Mutual fund styles ». In: *Journal of financial Economics* (1997).

[35] Nandita Dass. *Hedge Funds Classifications using K-means Clustering method*. 9th internation conference on computing in economics and Finance. 2003.

[36] William N. Goetzmann, Elise Gourier, and Ludovic Phalippou. *How Alternative Are Private Markets?* 2019.

7 Appendix

7.1 Tables

Table 7 – Number of funds before and after cleaning

	Number of Downloaded funds	Number of Funds that answers the cleaning criteria
Buyout	2455	1643
RE	1222	678
PC	965	640
VC	1830	885

As a reminder, the exclusion criteria are: fund size > 100 mio, no vintage 2023, no vintage before 1996, at least 70% of existing cash-flows values.

Table 8 – Number of datapoints per year

Year from investing	Buyout	Private Credit	Real Estate	VC
1	1490	545	556	780
2	1564	590	634	804
3	1508	546	596	770
4	1435	507	579	729
5	1322	452	536	680
6	1237	390	478	640
7	1148	333	432	604
8	1053	289	390	554
9	965	234	337	513
10	901	194	295	473
11	829	163	262	452
12	745	139	229	431
13	689	120	200	409
14	651	100	184	386
15	612	89	166	367
16	530	69	139	325
17	425	50	92	277
18	333	38	60	237

Table 9 – Median TVPI by vintage and strategy, Burgiss data

Vintage	50th Buyout	50th VC	50th Private Debt	50th Real Estate
(All)	1.55	1.34	1.24	1.28
2023	0.80	0.75	1.03	0.87
2022	0.92	0.89	1.03	0.93
2021	1.11	1.00	1.08	1.13
2020	1.27	1.24	1.13	1.18
2019	1.41	1.43	1.19	1.20
2018	1.56	1.78	1.24	1.29
2017	1.80	2.02	1.23	1.37
2016	1.85	2.06	1.28	1.37
2015	1.80	2.15	1.32	1.39
2014	1.89	2.49	1.28	1.35
2013	1.89	2.36	1.29	1.41
2012	1.76	2.57	1.29	1.50
2011	1.89	2.37	1.31	1.40
2010	1.77	2.36	1.38	1.44
2009	1.92	2.25	1.48	1.44
2008	1.57	1.84	1.43	1.29
2007	1.60	1.72	1.32	1.15
2006	1.56	1.40	1.32	0.84
2005	1.49	1.35	1.34	1.02
2004	1.62	1.03	1.22	1.16
2003	1.76	1.25	1.38	1.44
2002	1.79	1.00	1.25	1.52
2001	1.93	1.22	1.51	1.57
2000	1.64	0.86	1.34	1.66
1999	1.33	0.69	1.43	1.50
1998	1.43	1.09	1.28	1.48
1997	1.32	2.11	1.60	1.55
1996	1.46	2.84	1.33	1.59

Table 10 – Median IRR by vintage and strategy, Burgiss data

Vintage	50th Buyout	50th VC	50th Private Debt	50th Real Estate
(All)	13.77	8.38	8.84	8.61
2023	-18.35	-24.67	5.76	-19.45
2022	-12.29	-15.63	4.57	-9.95
2021	11.55	-0.04	9.25	12.01
2020	19.69	14.37	9.86	13.73
2019	21.85	17.82	10.70	11.35
2018	21.14	23.26	10.08	11.83
2017	21.98	24.14	8.59	11.55
2016	19.53	20.46	9.30	9.83
2015	17.72	17.92	8.39	10.03
2014	17.51	19.86	8.30	7.45
2013	16.47	15.36	7.35	10.79
2012	15.13	17.91	8.07	12.35
2011	16.04	15.72	8.87	11.10
2010	12.92	13.78	10.48	11.38
2009	13.22	12.95	11.58	9.97
2008	10.64	9.29	10.31	6.95
2007	9.59	9.45	6.53	2.94
2006	8.32	5.24	5.45	-1.81
2005	8.16	3.97	7.02	0.88
2004	11.80	0.32	3.19	3.03
2003	12.90	2.50	9.88	12.17
2002	16.25	-0.06	10.77	13.40
2001	20.89	3.49	14.51	19.32
2000	12.10	-1.92	13.33	18.35
1999	6.44	-5.33	11.34	10.77
1998	7.90	1.54	5.38	10.27
1997	5.78	28.26	10.15	13.10
1996	9.34	45.03	5.62	10.49

Table 11 – YACF and number of funds after cleaning

Strategies	WACPY	WADPY	Number of funds
Buyout	3.44	8.13	585
Funds of funds Buyout	4.10	9.46	94
Co-Investments	2.59	6.90	51
Growth	3.32	8.47	75
Infrastructure	2.84	8.17	41
Mezzanine	3.31	7.73	75
Private Credit	2.14	6.07	148
RE opportunistic	2.28	7.28	117
RE Value-Add	2.64	7.25	153
Secondaries	2.87	6.99	79
VC	3.60	8.79	305

7.2 Definitions

Buyout/LBO Purchase of at least a controlling percentage of a company's capital stock [24] by a PE firm to take over its assets and operations. A leveraged buyout (LBO) involves borrowing money to finance a portion of the purchase price.

VC Private equity financing that is provided by venture capital firms to startups and early-stage companies that have been deemed to have high growth potential. Unlike traditional financing, VC usually involves offering capital in exchange for an equity stake in the company. The primary goal of venture capitalists is to achieve a high return on investment through the growth and eventual exit, either through a company sale or an IPO, of their portfolio companies.

Private debt/credit Private debt refers to loans or credit extended to companies or individuals by entities other than traditional banks. These loans are typically structured, negotiated, and held privately, often by specialized private debt funds, rather than being publicly syndicated or traded on a secondary market.

Real estate Real estate investing pertains to the direct acquisition, ownership, and management of properties, often bypassing public exchanges. Investment strategies in this sector can vary, including core strategies focusing on stable, income-generating assets, or value-add strategies that seek to enhance properties through improvements and repositioning to achieve higher returns.

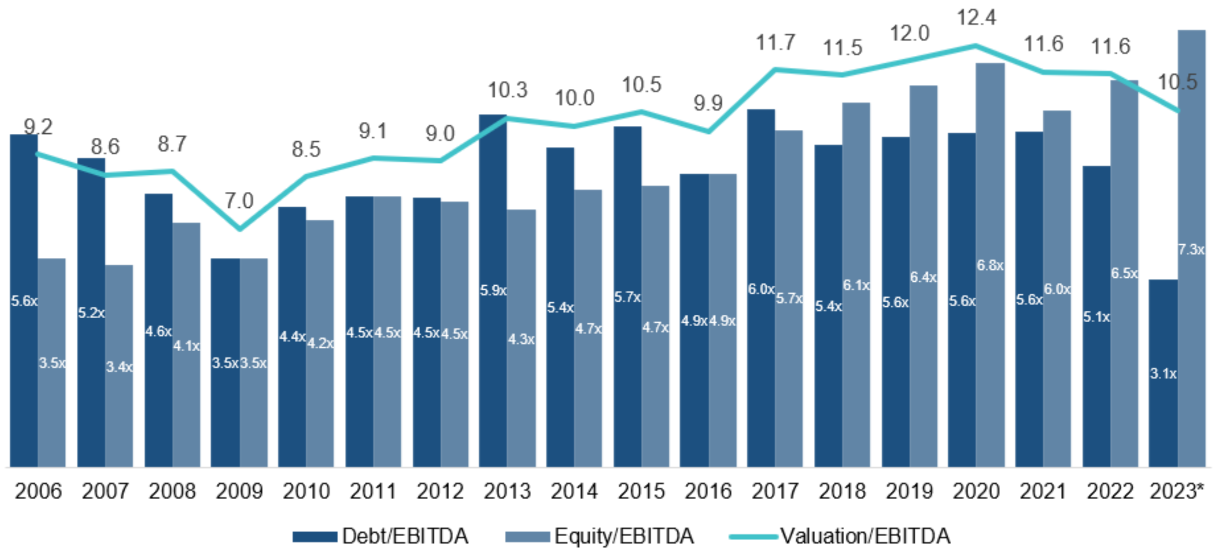


Figure 40 – North America & Europe multiples for financial M&As

7.3 Figures

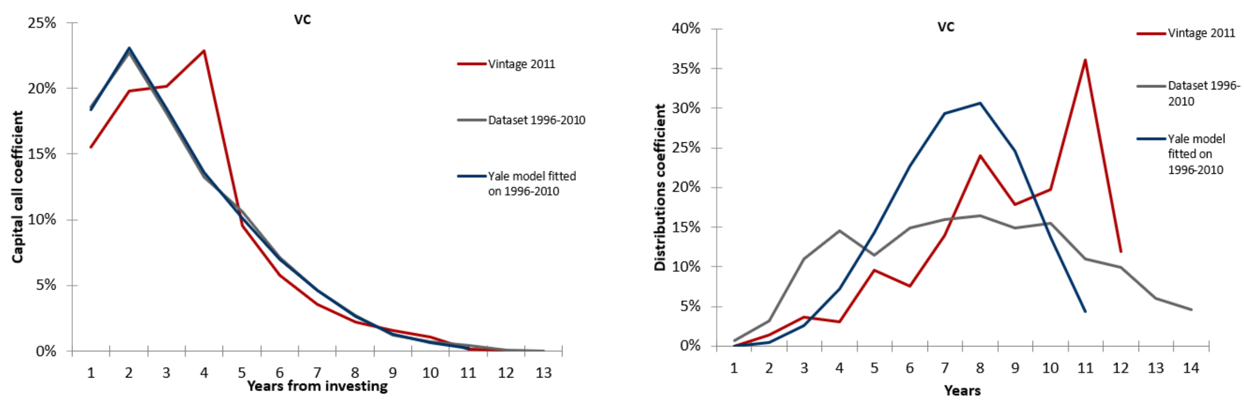
7.3.1 Asset vs equal weighted by strategy

7.3.2 Cash-flows coefficients during time by strategy

7.3.3 Bow by vintage and by strategies

7.3.4 Hook by vintage and by strategy

7.3.5 2011 vintage comparison by strategy



(a) Capital calls - Backtest 2011 - VC

(b) Distributions - Backtest 2011 - VC

Figure 57 – Average VC pattern compared to Yale and 2011 vintage

24. On a fully diluted basis

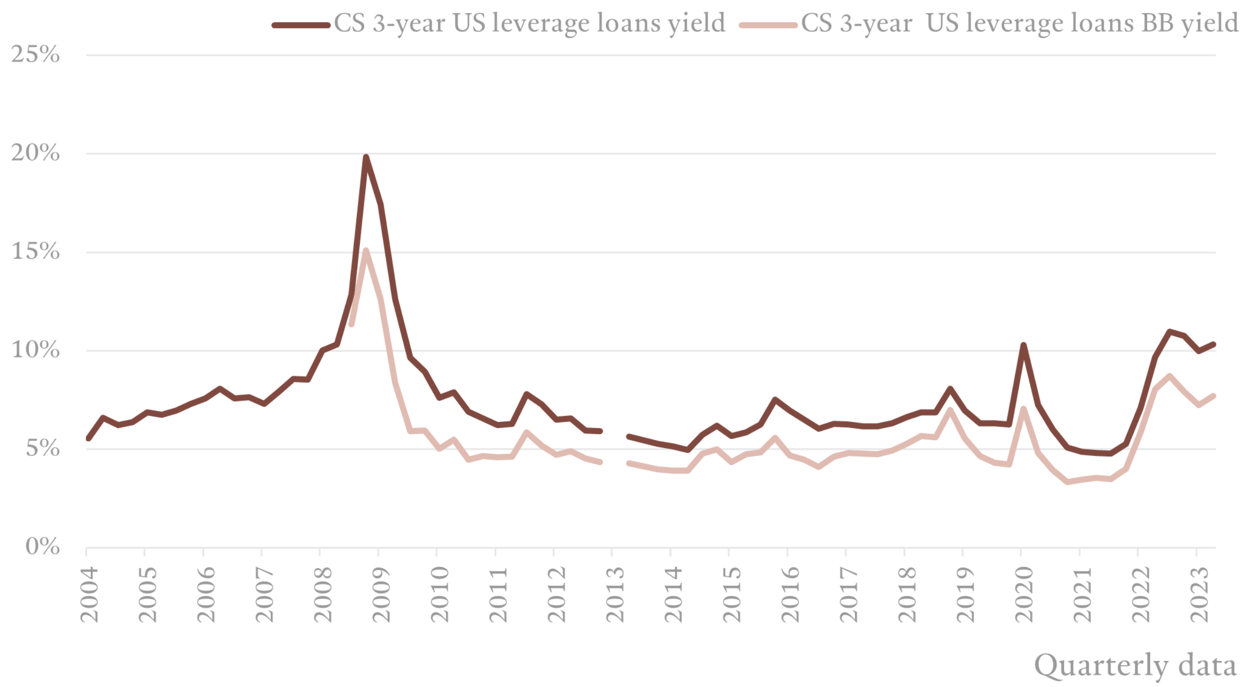
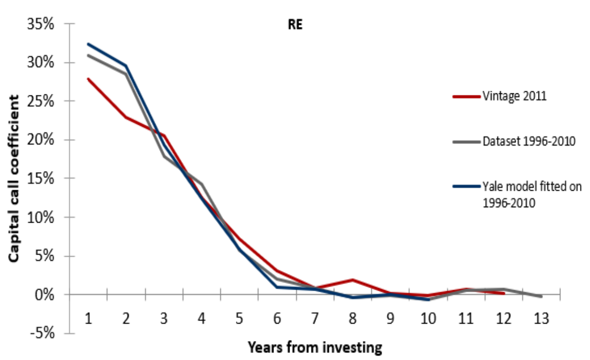
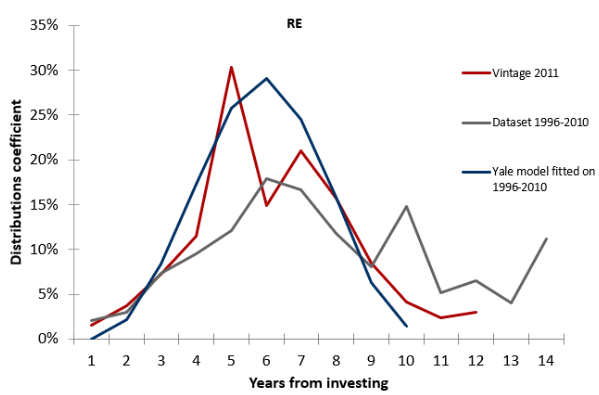


Figure 41 – Proxy of the refinancing's cost for LBOs



(a) Capital calls - Backtest 2011 - RE



(b) Distributions - Backtest 2011 - RE

Figure 58 – Average RE pattern compared to Yale and 2011 vintage

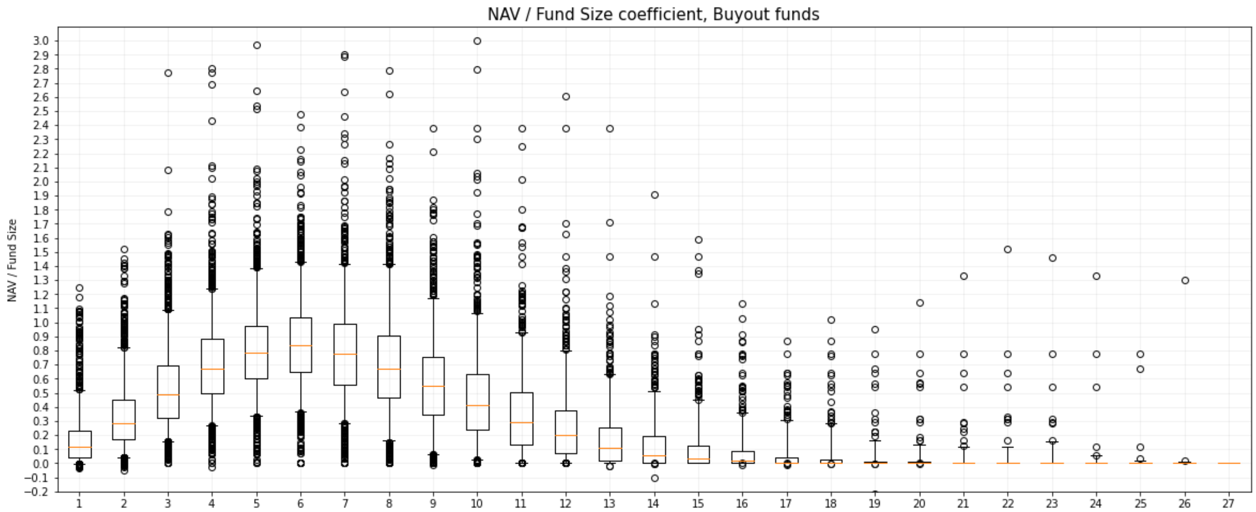
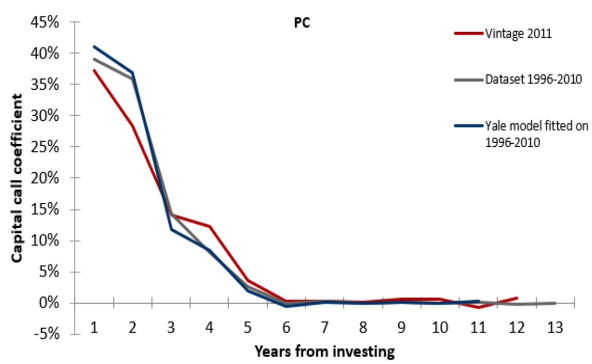
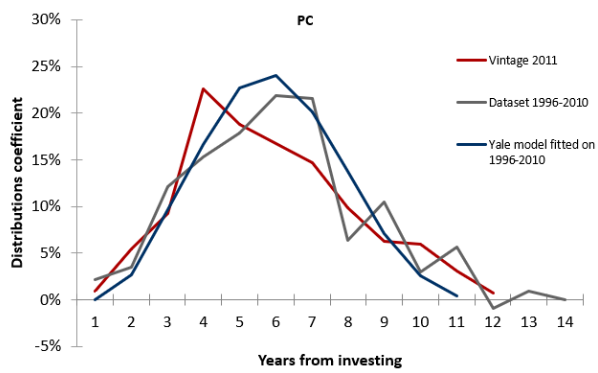


Figure 42 – Box plot of NAV coefficient, PitchBook data



(a) Capital calls - Backtest 2011 - PC



(b) Distributions - Backtest 2011 - PC

Figure 59 – Average PC pattern compared to Yale and 2011 vintage

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   AYCC   R-squared:                       0.154
Model:                            OLS   Adj. R-squared:                  0.142
Method:                 Least Squares   F-statistic:                     13.30
Date:                Thu, 17 Aug 2023   Prob (F-statistic):           1.32e-46
Time:                        08:54:38   Log-Likelihood:                -2614.3
No. Observations:                1706   AIC:                             5277.
Df Residuals:                    1682   BIC:                             5407.
Df Model:                          23
Covariance Type:            nonrobust
==================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                              2.9317      0.097     30.152      0.000       2.741       3.122
Fund Size                      -2.252e-06   1.52e-05     -0.148      0.882   -3.21e-05    2.76e-05
Buyout                             0.8098      0.105      7.746      0.000       0.605       1.015
Venture Capital - Early Stage      1.4061      0.202      6.956      0.000       1.010       1.803
Real Estate Value Added           -0.0369      0.130     -0.284      0.777      -0.292       0.218
Mezzanine                          0.5606      0.158      3.547      0.000       0.251       0.871
Real Estate Opportunistic         -0.0814      0.143     -0.571      0.568      -0.361       0.198
Distressed Debt                   -0.6153      0.154     -4.008      0.000      -0.916      -0.314
Real Estate Debt                  -0.1177      0.240     -0.489      0.625      -0.589       0.354
Growth/Expansion                   0.5980      0.160      3.748      0.000       0.285       0.911
Infrastructure Core                0.3506      0.245      1.432      0.152      -0.130       0.831
Infrastructure - General          -0.1387      0.394     -0.352      0.725      -0.911       0.634
Co-Investment                     -0.1491      0.209     -0.714      0.475      -0.559       0.260
Infrastructure Debt               -0.6381      0.770     -0.829      0.407      -2.148       0.872
Venture - General                  0.7360      0.118      6.213      0.000       0.504       0.968
Secondaries                        0.2448      0.167      1.470      0.142      -0.082       0.572
Venture Capital - Later Stage      0.5773      0.287      2.013      0.044       0.015       1.140
Direct Lending                    -0.7237      0.245     -2.956      0.003      -1.204      -0.244
Infrastructure Value Added         0.3775      0.421      0.897      0.370      -0.448       1.203
Credit Special Situations         -0.7143      0.315     -2.270      0.023      -1.331      -0.097
Infrastructure Opportunistic       0.2181      0.452      0.483      0.629      -0.668       1.104
Infrastructure Core Plus           1.1534      0.770      1.498      0.134      -0.356       2.663
Venture Debt                      -0.8547      0.631     -1.355      0.176      -2.092       0.383
Other                              0.0890      1.084      0.082      0.935      -2.038       2.216
Debt - General                    -0.1199      1.084     -0.111      0.912      -2.247       2.007
==============================================================================
Omnibus:                       36.685   Durbin-Watson:                   1.665
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               49.567
Skew:                           0.249   Prob(JB):                     1.72e-11
Kurtosis:                       3.671   Cond. No.                     3.00e+19
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.21e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Figure 43 – Results regression AYCC

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    AYD   R-squared:                       0.154
Model:                            OLS   Adj. R-squared:                  0.143
Method:                 Least Squares   F-statistic:                     13.34
Date:                Thu, 17 Aug 2023   Prob (F-statistic):           9.57e-47
Time:                        08:54:39   Log-Likelihood:                -3689.7
No. Observations:                1706   AIC:                             7427.
Df Residuals:                    1682   BIC:                             7558.
Df Model:                          23
Covariance Type:            nonrobust
=================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
const                            6.7256      0.183     36.825      0.000       6.367       7.084
Fund Size                        0.0001   2.86e-05      3.858      0.000    5.42e-05       0.000
Buyout                           1.4766      0.196      7.519      0.000       1.091       1.862
Venture Capital - Early Stage    2.8650      0.380      7.545      0.000       2.120       3.610
Real Estate Value Added          0.2152      0.244      0.881      0.379      -0.264       0.694
Mezzanine                        0.7459      0.297      2.513      0.012       0.164       1.328
Real Estate Opportunistic        0.3877      0.268      1.448      0.148      -0.137       0.913
Distressed Debt                 -0.4705      0.288     -1.632      0.103      -1.036       0.095
Real Estate Debt                -0.9753      0.452     -2.159      0.031      -1.861      -0.089
Growth/Expansion                 1.6518      0.300      5.511      0.000       1.064       2.240
Infrastructure Core              1.2065      0.460      2.624      0.009       0.305       2.108
Infrastructure - General         0.6141      0.740      0.830      0.406      -0.837       2.065
Co-Investment                    0.0976      0.392      0.249      0.803      -0.672       0.867
Infrastructure Debt             -0.3449      1.446     -0.239      0.811      -3.181       2.491
Venture - General                1.8657      0.223      8.384      0.000       1.429       2.302
Secondaries                     -0.2162      0.313     -0.691      0.490      -0.830       0.398
Venture Capital - Later Stage    2.0095      0.539      3.731      0.000       0.953       3.066
Direct Lending                  -1.7413      0.460     -3.787      0.000      -2.643      -0.839
Infrastructure Value Added       1.2176      0.790      1.541      0.124      -0.332       2.768
Credit Special Situations       -1.1846      0.591     -2.005      0.045      -2.344      -0.026
Infrastructure Opportunistic     1.7643      0.849      2.079      0.038       0.099       3.429
Infrastructure Core Plus         0.2712      1.446      0.188      0.851      -2.565       3.107
Venture Debt                     0.0967      1.185      0.082      0.935      -2.228       2.421
Other                           -3.7917      2.037     -1.862      0.063      -7.787       0.203
Debt - General                  -1.0355      2.037     -0.508      0.611      -5.030       2.959
==============================================================================
Omnibus:                      163.364   Durbin-Watson:                   1.684
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              259.191
Skew:                          -0.692   Prob(JB):                     5.22e-57
Kurtosis:                       4.315   Cond. No.                     3.00e+19
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.21e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Figure 44 – Results regression AYD

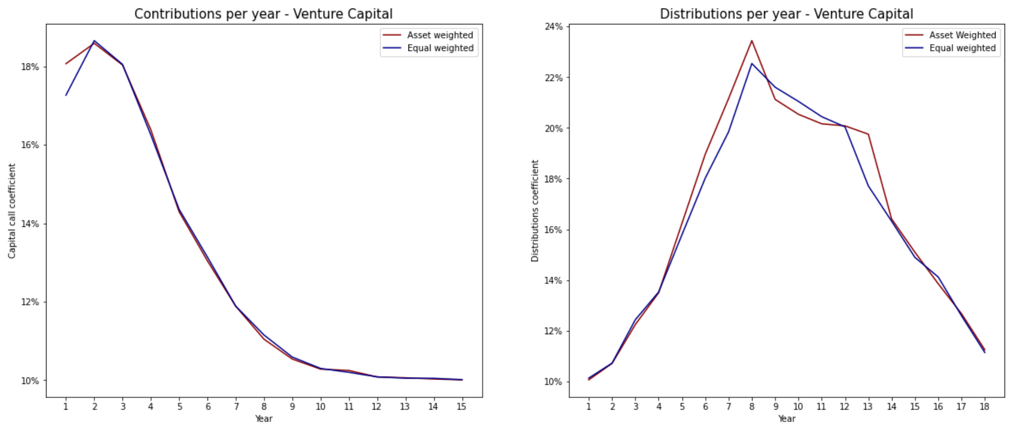


Figure 45 – Asset weighted vs Equal weighted - VC

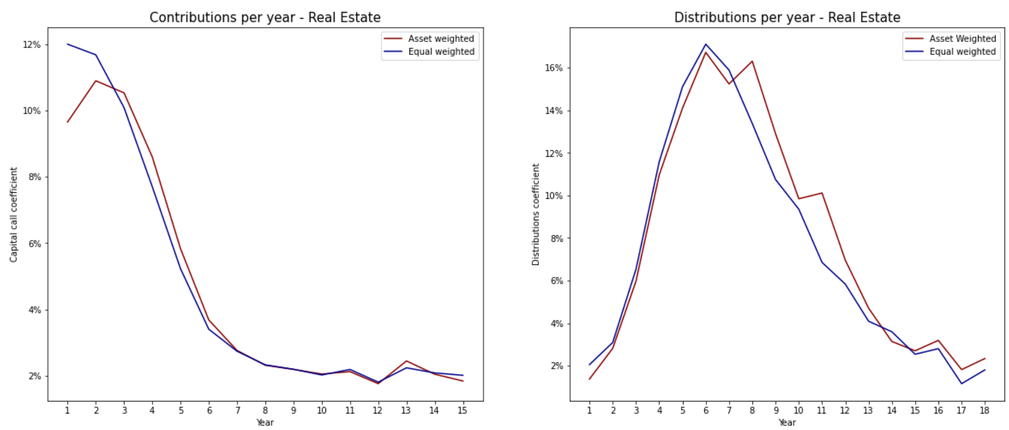


Figure 46 – Asset weighted vs Equal weighted - RE

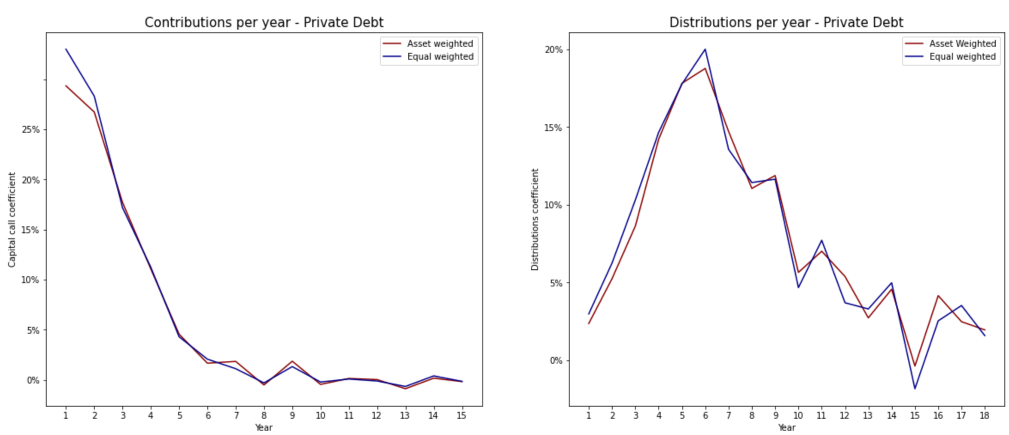


Figure 47 – Asset weighted vs Equal weighted - PC

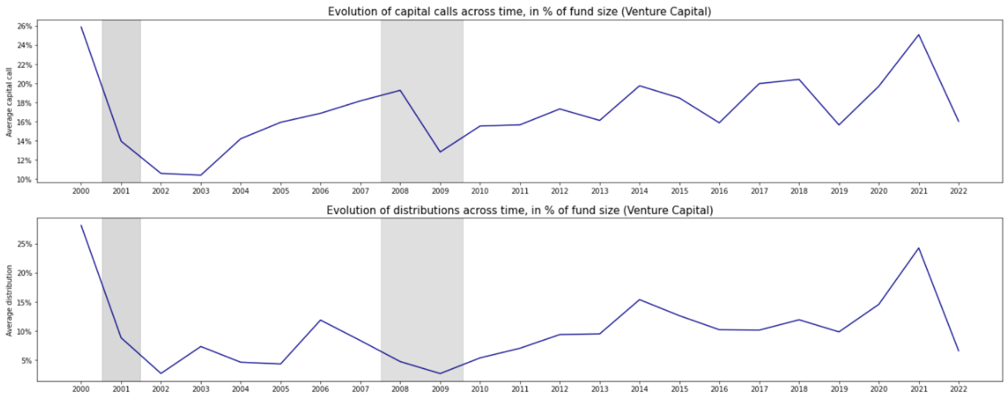


Figure 48 – Cash-flows coefficients across time, VC

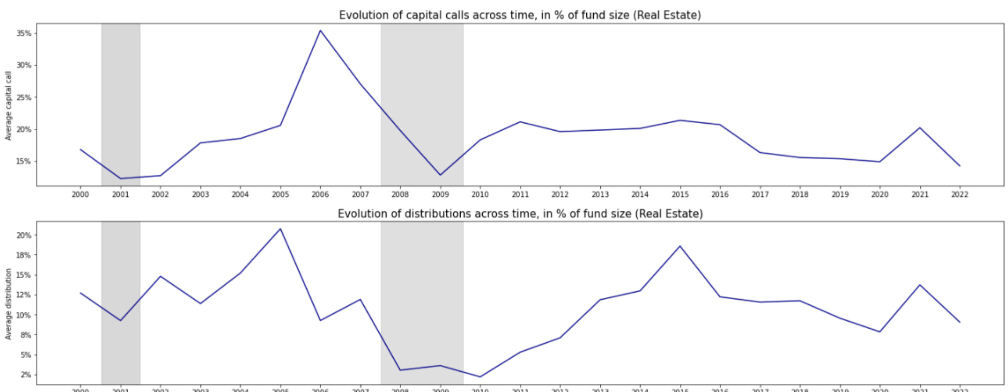


Figure 49 – Cash-flows coefficients across time, RE

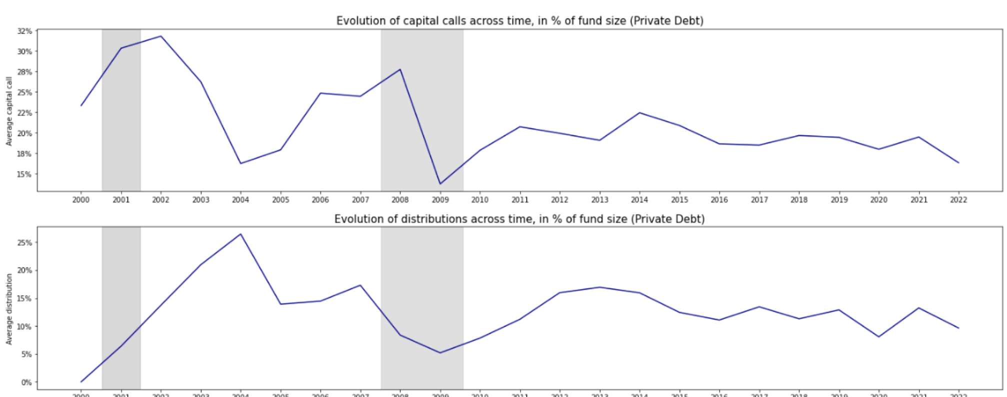


Figure 50 – Cash-flows coefficients across time, PC

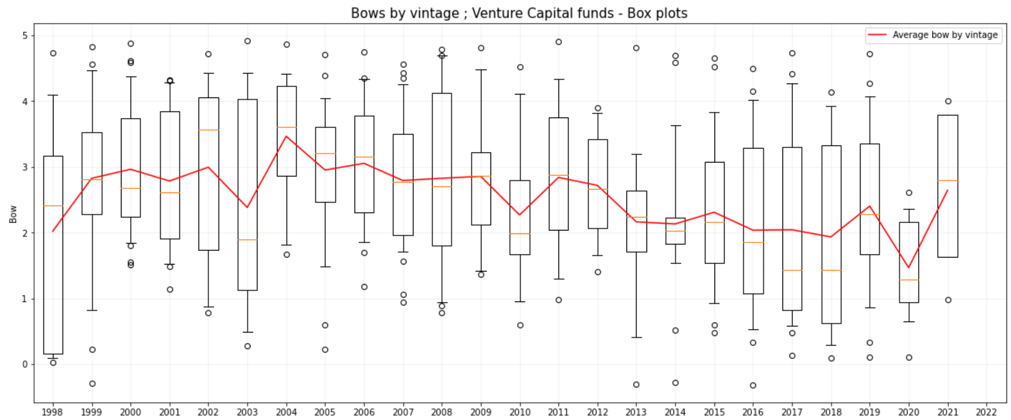


Figure 51 – Bow by vintage - VC

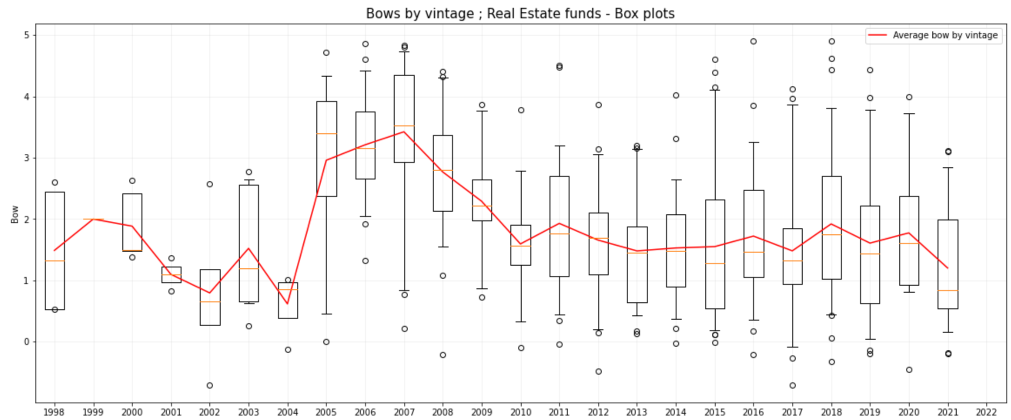


Figure 52 – Bow by vintage - RE

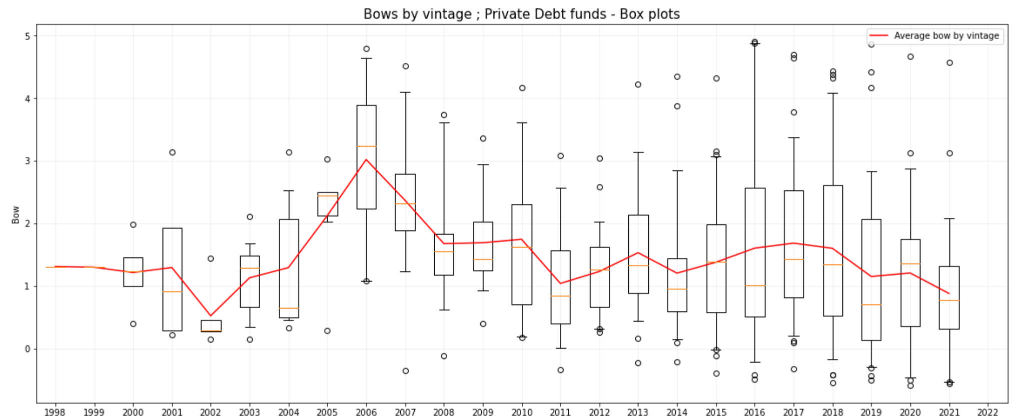


Figure 53 – Bow by vintage - PC

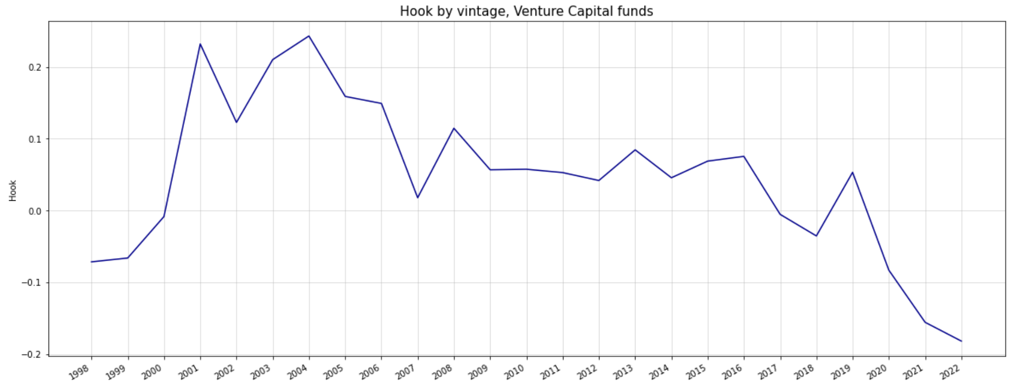


Figure 54 – Hook by vintage - VC

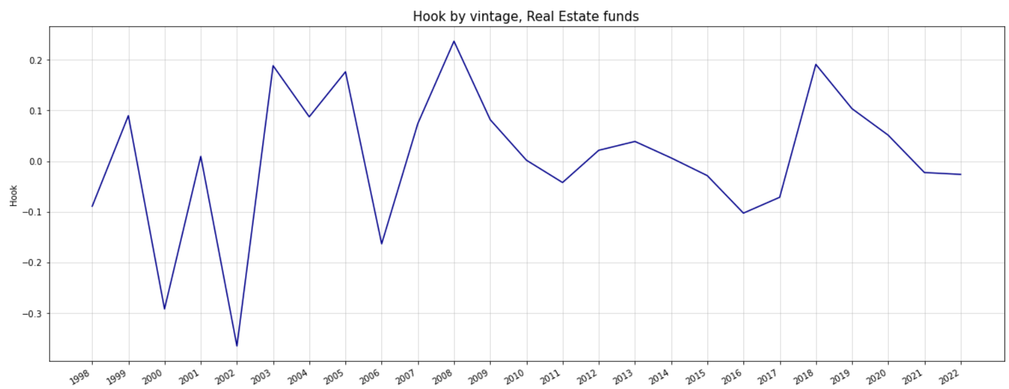


Figure 55 – Hook by vintage - RE

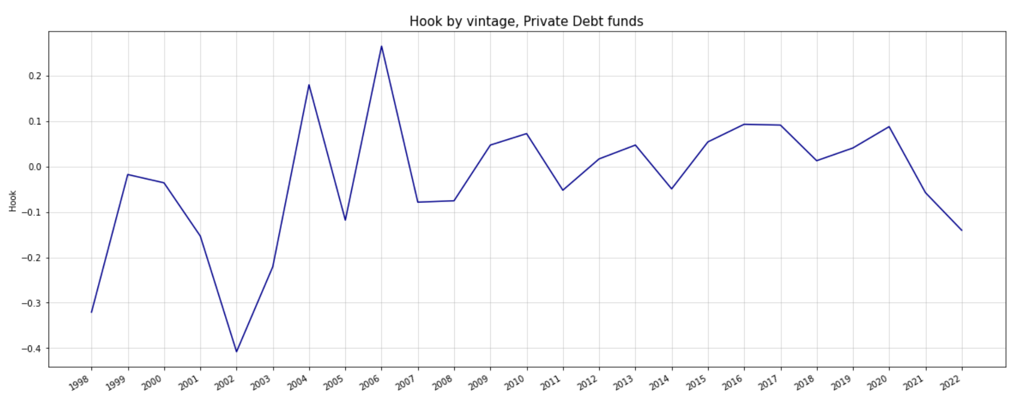


Figure 56 – Hook by vintage - PC