DEGREE PROJECT IN MATHEMATICS,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2020

Stochastic Modelling of Cash Flows in Private Equity

OSCAR UNGSGÅRD

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING SCIENCES

Stochastic Modelling of Cash Flows in Private Equity

OSCAR UNGSGÅRD

Degree Projects in Financial Mathematics (30 ECTS credits)
Master's Programme in Applied and Computational Mathematics
KTH Royal Institute of Technology year 2020
Supervisor at KTH: Sigrid Källblad Nordin
Examiner at KTH: Sigrid Källblad Nordin

TRITA-SCI-GRU 2020:062
MAT-E 2020:025

Royal Institute of Technology
School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci

Abstract

An investment in a private equity is any investment made in a financial asset that is not publicly traded. As such these assets are very difficult to value and also give rise to great difficulty when it comes to quantifying risk. In a typical private equity investment the investor commits a prespecified amount of capital to a fund, this capital will be called upon as needed by the fund and eventually capital will be returned to the investor by the fund as it starts to turn a profit. In this way a private equity investment can be boiled down to consist of two cash flows, the contributions to the fund and distributions from the fund to the investor. These cash flows are usually made within a prespecified time frame but at unspecified intervals and amounts. As an investor in a fund, carrying too little liquid assets when contributions are called upon will cause trouble, but carrying significantly more than needed is also not desirable as it represents a loss in potential revenue from having less capital in more profitable investments. The goal of this thesis was to attempt to find a way to reliably model these cash flows and to find a way to represent the results in a meaningful way for the benefit of the investor by constructing value at risk like risk measures for the necessary liquid capital to carry at a given time in case contributions are called upon.

It was found that the distributions could be modelled very well with the chosen stochastic processes, both as it related to predicting the average path of the cash flows and as it relates to modelling the variability of them. Contrary to this it was found that the contributions could not be modelled very well. The reason for this was found to be an observed lag in the speed of contributions at the start of the funds lifetime, this lag was not taken into account when constructing the stochastic model and hence it produced simulated cash flows not in line with those used in the calibration.

Stokastisk Modellering av Kassaflöden i Private Equity - Abstrakt

En investering i private equity är en investering i en tillgång som inte är börsnoterade. På grund av detta är sådana tillgångar väldigt svåra att värdera och medför även store svårigheter när det kommer till att kvantifiera risk. I en typisk private equity investering so ingår en investerare i ett löfte att under en viss förbestämd tidsperiod bidra med en fixt mängd kapital till en private equity fond. Detta kapital kommer att gradvis kallas på av fonden vid behov för att sedan mot slutet av fondens livstid ge utdelning när private equity fonden börjar göra en vinst. På detta viset kan en private equity investering brytas ner i två kassaflöden, kontributioner in i fonden, och distributioner ut ur fonden. Dessa kassaflöden sker under en förbestämd tidsperiod men ej förbestämda belopp. Som en investerare i denna typen av fond är därför en risk att bära för lite likvid kapital när kontributioner blir kallade på men även oattraktivt att bäre på för mycket de detta representerar förlorar potentiell avkastning.
Målet i denna uppsatts är att hitta ett sätt att på att tillförlitligt vis modellera dessa kassaflöden och representera resultaten på ett meningsfullt sätt från perspektivet av en investerare. För att uppnå detta skapades value-at-risk liknande mått för mängden likvid kapital som krävs under en tidsperiod för att säkra sig mot påkallade kontributioner.
Slutsatsen blev att distributioner kunde modelleras väl, både när det kom till att efterlikna den genomsnittliga vägen av kassaflöden och även för att modellera risken. I kontrast till detta så kunde inte kontributioner modelleras mot tillräckligt hög säkerhet för att användes i det ämnade syftena. Anledningen till detta var en eftersläpning i hastigheten som kontributioner kallades med som inte tågs i beaktande av den tillämpade matematiska modellen.

Preface

This thesis was written at the behest of one of the largest banks in Sweden using transaction data from private equity investments dating back to 2009. Because of potential confidentiality issues the institution within the bank that provided the data and deal with the private equity investments analysed in this thesis will henceforth be referred to simply as The Institution when mentioned.

Acknowledgements

For their help in making this thesis a reality I would like to extend thanks to the following people:

Sigrid Källblad, my supervisor at KTH, for providing valuable and in-depth feedback at key times during the work on the project.

A special thanks to Alexander Bea at The Institution without whose continued support and guidance this thesis would not have been possible.

Finally a word of gratitude to all the people at The Institution who helped out in acquiring the data and made themselves available to respond to questions on the subject of private equity.

Contents

1 Introduction

In recent times investments in private equity have increased in popularity among investment firms. One reason for this is lower interest rates causing investors to seek out different ventures in order to maximize profit, private equity investments being among those. Naturally this has caused an increased demand for models analysing these investments and highlights the importance of research such as this. The illiquid character of these investments, however, make much of the well established methods in traditional financial analysis ill suited for the purpose and ordinary calculations like valuation, risk analysis and cash flow forecasting prove much more difficult. This thesis will place its focus on the latter and attempts to find a way to model these.

1.1 The investment explained

A private equity investment is typically made through what is known as a private equity fund. A private equity fund is a pooled investment into an unlisted company, it is an agreement between the general partner (GP), who manages the fund, and the limited partners (LP) that are the primary investors in the fund. There may be several LP invested in a single fund and these investors usually fall into the categories of pension funds, foundations and other institutional investors [10]. In this thesis we take the perspective of a LP investing in a fund, and will commonly refer to them simply as "the investor" and the private equity fund being invested in simply as "the fund". The inner workings of the fund allocations by the GP is not of interest in this thesis as we seek to model the cash flows quantitatively using advanced mathematical statistics by looking at large amounts of data from previous investments.
The investment is built upon an agreement between the investor and the GP of the fund specifying the terms of the investment. Two key figures are specified in the contract that will be of necessity, on top of the cash flow transactions, to perform the analysis in this these. These are:

- The initial committed capital (C). This is the max amount of capital the fund may demand from the investor during the investment period

- The investment time period. This is the amount of time the fund has in which it may make calls from the investors, this time period is usually around 10 years long [11].

In this way the life cycle of the investment is divided into two time periods, the investment period and the harvesting period. The investment period being characterized by the capital calls from the fund and the following contributions made to the fund by the investor, and conversely the harvesting period is characterized by distributions made back to the investor by the fund. Although the distributions are usually centered in the harvesting period they may also occur during the investment period.

1.2 The cash flows explained

The cash flows for a private equity investment is composed of two parts, the contributions from the investor to the fund, and the distributions from the fund back to the investor when the investment starts to turn a profit. These two parts are modeled separately and together make up the cash flow forecasting for the investment. The net cash position, that is the cumulated distributions minus the contributions, often give rise to a typical curve initially showing a sharp decline in the net cash positions as funds are called upon that then slowly starts to turn and exhibit a sharp increase as the investment turns a profit and is eventually liquidated. This curve is aptly named the "j-curve", an example of which is shown in figure 3. The typical behaviour of the contributions can be seen in figure 1, the analysis of which may be of great interest when it comes to managing an investors liquid positions. Figure 2 shows the behaviour of the distribution. Notice how the contributions are centered close to the start period of the fund while the distributions happen primarily towards the end, this along with the difference in magnitude between the sum of the distributions and the contributions is what gives rise to the well known J-curve shape of the net cash position.

1.3 Goals

The goal of this thesis is to model the contributions and distributions associated with private equity in order to reliably predict the J-curve. To achieve this we will model the cash flows with stochastic processes, calibrated with transaction data from previous investments, in order to be able to simulate the possible paths the contributions and distributions of an ongoing investment may take. The analysis in this thesis is based on- and builds upon the work done by Buchner, Axel[1]; Kaserer, Christoph[2] and Wagner, Niklas [3] in their extensive quantitative analysis of private equity. In order to suit the model better for the purposes of this thesis we will also make some augmentations to the model, see section 2.2.1. The data used in this thesis consist of raw transaction data dating back to the start of 2009, the data was provided by The Institution but because of the data being spread out over many different databases with metadata for the investments missing many of the key figures needed in the analysis will have to be inferred from the transaction data itself. This is in contrast to [1,2,3] where much more information about the funds was available as well as information on the precis nature of the investment, e.g. if it was a venture capital investment or a buyout [23]. Because of this we are forced to make no distinction between the investments in our analysis.
The primary benefit of modelling the cash flows of a private equity investment, apart from getting an overview of the performance of the fund, is to better help manage the liquidity of the investor. Two common difficulties in handling liquidity that may arise are:

- Carrying too little liquid capital when contributions are called upon, causing risk of defaulting on the agreement or having to sell assets at below market price in order to quickly raise the needed capital.

- Carrying too much liquid capital as a result of not being able to optimally reinvest distributions in new investments. In light of the short notice often given before distributions are made by the fund, and the hard to predict nature of these distributions, the investor may find itself suddenly being in a position where it is not optimally investing its capital. Often the investor will be seeking to continually reinvest distributions in new funds so this poses a real issue. [12]

With this in mind we seek to answer the follow questions in particular:

- To what certainty can future contributions and distributions for private equity be predicted, and can useful enough metrics be derived from those predictions so that you are able to advice an investment firm on how much liquid capital they ought to have in order to answer called upon contributions?

2 Mathematical model

There are two key components to the mathematical model used to model the cash flows of a private equity investment: the contributions to the fund and the distributions gotten from the fund. These two components are modelled separately and then aggregated together to form the results. This section will go over the key properties of the mathematical models used for each component of the analysis and the arguments for using them. The theory of the model is based on the works of Buchner, Kaserer and Wagner published in a series of articles on the analysis of private equity[1,2,3].

2.1 Modelling contributions

It is known empirically that contributions to a fund are usually concentrated in the early and middle stages of a funds lifetime only to stagnate as only very little undrawn capital remains from the initial committed capital. It is therefore suitable to model the speed at which funds are drawn as a function of the remaining capital, from which we arrive at the first equation governing the contributions:

$$dD_t = \delta_t U_t dt. \tag{1}$$

Equation (1) is an ordinary differential equation where D_t denotes the cumulative distributions up to the point t, $U_t = C - D_t$ with C being the total commited capital. We will call the parameter δ_t the "drawdown rate" which is a stochastic process affecting the speed at which contributions are drawn, hence its name. Equation (1) can be solved straight forwardly in closed form for the cumulated distributions to obtain:

$$D_t = C - Cexp(-\int_0^t \delta_u du) \tag{2}$$

The economic explanation of the drawdown rate is that it plays the role of the current state of the market, a high appetite from investors to invest in private equity funds would lead to a high long term mean of the drawdown rate and increased speed at which contributions are drawn.

We may also derive an expression for the instantaneous capital drawdowns, the instantaneous change in the cumulated contributions, given by $\frac{dD_t}{dt}$ as:

$$\frac{dD_t}{dt} = \delta_t C \exp(-\int_0^t \delta_u du) \tag{3}$$

Mathematically the drawdown rate is modelled by a mean reverting stochastic process, analogous to the well known CIR[4] model proposed by Cox, Ingersoll and Ross for modelling the short rate term structure and provides us with many of the same desirable attributes, such as remaining non negative. The stochastic differential equation for the drawdown rate is:

$$d\delta_t = k(\theta - \delta_t)dt + \sigma_\delta \sqrt{\delta_t} dB_{\delta,t} \tag{4}$$

where $B_{\delta,t}$ is a standard brownian motion. Analogous to [4], the parameters of the model that need to be estimated are: θ, the long run mean of the drawdown rate indicating the average level of the drawdown rate in current market conditions; k, the coefficient of mean reversion, which deicides the speed at which normality is reached after period of volatility in the drawdown rate, and finally σ_δ which is the volaility of the drawdown rate itself.
By calibrating and simulating the drawdown rate (see section 3.1 and 3.5) we may solve equation (2) numerically for each simulated value of the drawdown in order to model the contributions.

2.2 Modelling distributions

Contrary to the contributions, the distributions from a fund are typically centered towards the middle and end of the lifetime, increasing quickly as the fund starts to turn a profit and then decreasing once enough distributions have been made and the fund starts to get liquidated.
To model the distributions We make the assumption that the speed of distributions p_t, where $p_t = \frac{dP_t}{dt}$, P being the cumulated distributions, follows a geometric Brownian motion such that:

$$d \log p_t = \mu_t dt + \sigma_P dB_{P,t}, \tag{5}$$

where σ_P is a constant volatility, μ_t is a time dependant drift and $B_{P,t}$ is a second Brownian motion. We will later see how by using a suitable definition of μ_t this allows us to create a model with the expected behavior. By modelling $\log p_t$ instead of p_t directly we restrict p_t to be non-negative at all times, which is important since we are modelling the distributions and contributions separately.
We now need to find a suitable representation for the time dependant drift that gives the distributions the desired properties, in order to do which we introduce the concept of the fund multiple M_t defined as

$$M_t = \frac{P_t}{C} \tag{6}$$

where P_t is the cumulated distributions up to time t and C is the commited capital. We see that the fund multiple is the normalized return of the fund at time t, e.g. a fund where the commited capital is C that when liquidated has distributed 1.5C would have a fund multiple of $M_T = 1.5$ at the end of the lifetime. We now make the assumption that the expected value of the fund multiple M_t, seen from the time s, i.e $E_s[M_t]$ are governed by the following dynamics:

$$d(E_s[M_t]) = \alpha t(m - E_s[M_t])dt \tag{7}$$

where $t \geq s$ a.s.
We have introduced two new parameters here: m, the long run mean of fund mulitiple and the constant α which together with the time parameter governs the speed that the expectation of the fund multiple $E_s[M_t]$ is pulled towards the long run mean m. In this way shifting α will affect how early the distributions

happen and shifting m will affect how big they are, this way we retain a very flexible model. Equation (6) can further be solved for $E_s[M_t]$ to produce

$$E_s[M_t] = m + c_1 \exp(\alpha(t^2 - s^2)\frac{1}{2}), \tag{8}$$

where the initial condition $E_s[M_s] = M_s$ gives $c_1 = M_s - m$, yielding:

$$E_s[M_t] = m - (m - M_s) \exp(\alpha(t^2 - s^2)\frac{1}{2}). \tag{9}$$

From the definition of the fund multiple we have that $M_t = \frac{P_t}{C} = \frac{1}{C}\int_0^t p_s ds$ leading to $p_t = C\frac{dM_t}{dt}$ and $E_s[p_t] = C\frac{dE_s[M_t]}{dt}$. Inserting (7) into this equation we find, noting that $M_s = \frac{P_s}{C}$:

$$E_s[p_t] = \alpha t(mC - P_s) \exp(-\frac{1}{2}\alpha(t^2 - s^2)). \tag{10}$$

We now turn back to equation (5). From stochastic calculus we know that the solution of (4), with the initial value p_s is

$$p_t = p_s \exp(\int_s^t \mu_u du + \sigma_P(B_{P,t} - B_{P,s})) \tag{11}$$

and since $B_{P,t} - B_{P,s} = \sqrt{t-s}\epsilon$ is normal, ϵ being from a standard normal distribution, we find the well known expectation of this to be:

$$E_s[p_t] = p_s \exp(\int_s^t \mu_u du + \frac{1}{2}\sigma_p^2(t - s)). \tag{12}$$

Stopping to access for a moment we find that we have 3 equations of note, (10), (11) and (12). Equation (11) is our desired process for the distributions that we wish to simulate from, but it contains the unknown drift expression $\int_s^t \mu_u du$. Equations (10) and (12) however are both expressions of $E_s[p_t]$ with (12) containing the unknown drift term. By setting the two equal to each other allows us to solve for the unknown drift term yielding:

$$\int_s^t \mu_u du = \log(\frac{K}{p_s}) - \frac{1}{2}\sigma_p^2(t - s)) \tag{13}$$

where K is the expression on the right hand side of equation (10). Inserting equation 13 into 11 now give us an expression for p_t as:

$$p_t = \alpha t(mC - P_s) \exp -\frac{1}{2}(\alpha(t^2 - s^2) + \sigma_p^2(t - s) + \sigma_p(\epsilon\sqrt{t-s}) \tag{14}$$

having used again that $B_{P,t} - B_{P,s} = \sqrt{t-s}\epsilon$ where ϵ is from a standard normal distribution. With equation (14) we may then simulate values of p_t and P_t.

2.2.1 Additional augmentations to the model

The model as described above is for the most part derived from the work done by Buchner, Axel[1]; Kaserer, Christoph[2] and Wagner, Niklas [3]. It offers a model that with suitable on average produces the desirable J-Curve behaviour and allows for a very flexible model to calibrate. It does however lack some, for the type of research questions posed in this thesis, desirable features. One of these is that the distributions will towards the end of the lifetime tend very strictly to the long run mean of the fund, m, as described above. This being due to the $\alpha t(mC - P_s)$ term in equation (14) causing P_s to always converge to m. This is the expected outcome of the investment, but if we are interested in the uncertainty of the distribution this convergence deprives any such information past a certain point in the lifetime of the fund. We also run into difficulties if we are to analyse current investments where the cumulated distributions have exceeded, or have come close to exceeding, the expected return of the investment m. In [1,2,3] it is mentioned that the long run mean could also be made dependant on the information available at time s which is what we will try to model in order to circumvent the aforementioned issues.

In this thesis the following additional augmentation to the model is proposed in the form of a dynamic fund multiple m_t that is dependant on the information available at time t. The dynamics of the time dependant fund multiple m_t is:

$$dm_t = m_t(p_t - E_s[p_t])dt \tag{15}$$

where p_t is the observed instantaneous distributions at time t, inferred from the up to t observed distributions, and $E_s[p_t]$ is the expectation of m_t calculated at time $s < t$, which is available to us from (10). In this way m_t increases (decreases) with investments that that give rise to higher (lower) than expected distributions. Note that as m_t increases (decreases) as does the expected instantaneous returns, causing m_t to fall down again if the investment does not live up to its (new) expected distributions.

Note also that this augmentation only impacts the spread of the simulations, not the expected outcomes of them. The long run mean of the fund is still m.

3 Calibration and Simulation

This section will outline how the calibration is done and how the analysis has been carried out in this thesis.

3.1 Conditional least squares

A powerful tool in parameter estimation of stochastic differential equations is conditional least squares (CLS) which we will partly make use of when estimating the in-going parameters of the model. As the name suggests CLS is based on minimizing the sum of squares of the difference between the realized value at time t_i, X_{t_i}, and the expected value of X_t at time t_{i-1}, i.e. $E[X_t|t_{i-1}$ where t_{i-1} and t_i are discrete points in time where the value of the stochastic process X_t has been observed. In mathematical notation we have that the CLS estimator $\hat{\theta}$ of θ is given by the θ that minimizes: $\sum_{k=1}^{n}(X_k - E_\theta[X_k|F_{k-1}])^2$ where F_{k-1} is the information available at the point $k-1$ [5].

3.2 Calibrating the contributions

This section will outline how to calibrate the parameters going into the process modelling the contributions.

3.2.1 Estimating the long run mean and the coefficient of mean-reversion of the drawdown rate

Much work have been done on the CIR process used to model the contributions and there is therefore a lot of previous work done which enables us to easily cast the model into the context of CLS estimation; see e.g. [6]. Specifically, the conditional expectation of the (3) may analytically be derived as was done in [7]. To do this first note that the solution of (3) can be found analytically simply by multiplying both sides by $\exp k * \Delta t$ and integrating to produce:

$$\delta_t = \exp(kt)\delta_0 + \theta(1 - \exp(-kt)) + \sigma_\delta \exp(-kt) \int_0^t (\exp(ks)\sqrt{\delta_s} dW_s. \tag{16}$$

Taking the expectation of this we get

$$E[\delta_t] = \exp(-kt)\delta_0 + \theta(1 - \exp -kt) + \sigma_\delta \exp -kt \int_0^t (\exp(ks)\sqrt{\delta_s} E[dW_s] \tag{17}$$

$$= \theta(1 - \exp(-kt)) + \exp(-kt)\delta_0. \tag{18}$$

Discretizing this to look at the time point k conditional on the information at time point $k-1$ rather than at 0 we get our desired expression to use for the CLS estimation as:

$$E[\delta_k|F_{k-1}] = \theta(1 - \exp(-k\Delta t)) + \exp(-k\Delta t)\delta_{k-1}. \tag{19}$$

From this we can easily derive the CLS expression to be minimized by:

$$\sum_{k=1}^{n}(\delta_k - \theta(1 - \exp(-k\Delta t)) + \exp(-k\Delta t)\delta_{k-1})^2. \tag{20}$$

This expression can then be numerically minimized to find estimates for k_i and θ_i for each fund, where the subscript i denotes that the parameter has been estimated with the data from fund i. The used k and θ estimates are then simply the arithmetic mean of k_i and θ_i.
Since δ is not directly observed by us it must be approximated for the contributions. Using equation (2) we have that:

$$\hat{\delta_t} = \frac{(-\log(\frac{C-D_{t_k}}{C})) - (-\log(\frac{C-D_{t_{k-1}}}{C}))}{\Delta t} \tag{21}$$

where we have approximated the derivative by a finite difference approximation and D_{t_k} is, as before, the cumulated contributions up to time t_k.

3.2.2 Estimating the volatility of the drawdown rate

The conditional variance of δ_t also has an analytic expression. We already have an expression with (17) for the first moment of (15) so we only need to find the expression for the second moment to calculate the variance. We see that:

$$Var[\delta_t] = E[(\delta_t)^2] - E[(\delta_t)]^2 \tag{22}$$

$$= E[(\exp(kt)\delta_0 + \theta(1 - \exp(-kt)) + \sigma_\delta \exp(-kt)\int_0^t (\exp(ks)\sqrt{\delta_s}dW_s)^2] \tag{23}$$

$$-(\theta(1 - \exp(-k\Delta t)) + \exp(-k\Delta t)\delta_{k-1})^2. \tag{24}$$

The first expression term under the expression, when expanding the factor, will be taken out by the $E[(\delta_t)]^2$ term leaving:

$$2(\exp(-kt)\delta_0 + \theta(1-\exp -kt))\sigma_\delta \exp(-kt)E[\sigma_\delta \exp(-kt)\int_0^t (\exp(ks)\sqrt{\delta_s}dW_s)] \tag{25}$$

$$+\sigma_\delta^2 \exp(-2kt)E[\int_0^t (\exp(ks)\sqrt{\delta_s}dW_s)] \tag{26}$$

$$= 2(\exp(-kt)\delta_0 + \theta(1-\exp -kt))\sigma_\delta \exp -ktE[\sigma_\delta \exp(-kt)\int_0^t (\exp(ks)\sqrt{\delta_s}dW_s)] \tag{27}$$

$$+\sigma_\delta^2 \exp(-2kt)E[\int_0^t (\exp(2ks)\delta_s dW_s)] \tag{28}$$

$$= \sigma_\delta^2 \exp(-2kt)\int_0^t (\exp(2ks)E[\delta_s]dW_s). \tag{29}$$

We can now insert our expression for $E[\delta_s]$ from (17) in (29) which yields after some simplifications:

$$\frac{\sigma_\delta}{k}\delta_0(\exp{(-kt)} - \exp{(-2kt)}) + \frac{\theta\sigma_\delta^2}{2k}(1 - \exp{-kt})^2. \tag{30}$$

As with the expectation this can easily be discretized to denote the time point k conditional on the information at time point $k-1$ which finally yields our desired expression as:

$$Var[\delta_k|F_{k-1}] = \sigma^2(\nu_0 + \nu_1\delta_{k-1}) \tag{31}$$

where $\nu_0 = \frac{\theta}{2k}(1 - \exp{(-k\Delta t)})^2$ and $\nu_1 = \frac{1}{k}(\exp{(-k\Delta t)} - \exp{(-2k\Delta t)})$). According to [6] an estimator for σ_i^2 inspired by standard linear regression is:

$$\hat{\sigma_i}^2 = \frac{1}{n}\sum_{k=1}^{n}\frac{\delta_k - \theta(1 - \exp{(-k\Delta t)}) + \exp{(-k\Delta t)}\delta_{k-1}}{\nu_0 + \nu_1\delta_{k-1}}, \tag{32}$$

where ν_0 and ν_1 are evaluated at $\hat{\alpha}$ and $\hat{k}$. As with $\hat{\alpha}$ and $\hat{k}$ an estimator for σ is found by taking the arithmetic mean of the σ_i of each of the funds.

3.3 Calibrating the distributions

In this section we outline how to calibrate the parameters going into the process modelling the distributions.

3.3.1 Estimating the long run mean of the fund

For the distributions there are three parameters that need to be estimated: m, the long run mean of the funds, α, the coefficient of reversion to the fund multiple, and σ_p the volatility of the distributions.
For m an unbiased estimator, since the mean of the fund multiple will (even with the additional augmentations to the model) tend to m, is obtained simply by $\sum_{i=1}^{M} P_T^i$ where with the notation from section 3.2 that P_T^i is the cumulative distributions of fund i at the end of the lifetime of the fund T.

3.3.2 Estimating the coefficient of mean-reversion of the distributions

To estimate α we again make use of the CLS method and seek to find the value $\hat{\alpha}$ that minimizes:

$$\sum_{k=1}^{n}(P_k^i - E([P_k^i|F_{k-1}])^2 \tag{33}$$

for each fund i. A reasonable estimate for $\hat{\alpha}$ is then obtained as the average of these estimated $\hat{\alpha_i}$ for each fund i. The conditional expectation of P_k has already been defined in equation (10) as

$$E_s[M_t] = m - (m - M_s)\exp\left(\alpha(t^2 - s^2)\frac{1}{2}\right). \tag{34}$$

Inserting this into (13), after discretizing t and s and with $m = \hat{m}$, $M_s = P_{k-1}$, $M_t = P_k$ we immediately find the sum to be minimized as:

$$\sum_{k=1}^{n}(P_k^i - (\hat{m} - \hat{m} - P_{k-1}^i)\exp\left(\alpha(t_k^2 - t_{k-1}^2)\frac{1}{2}\right)))^2. \tag{35}$$

3.3.3 Estimating the volatility of the distributions

To estimate σ_p we make use of our processes for the log capital distributions in (4) and first estimate the variance during each time interval $[t_{k-1}, t_k]$ by using the standard formula for variance $Var(X) = E[X^2] - E[X]^2$ to get:

$$\sigma_k^2 = \log\left(\frac{1}{M}\sum_{i=1}^{M}((\Delta t p_k^i)^2\right) - 2 * \log\left(\frac{1}{M}\sum_{i=1}^{M}((\Delta t p_k^i)\right) \tag{36}$$

where $\Delta t = t_k - t_{k-1}$, p_k is the instantaneous distribution at time k, and the expectation has been taken as the arithmetic mean over the n different funds. The variance for the entire time period is now calculated by taking the sum of the σ_k weighted by the average distribution at each k, i.e $w_k = \frac{1}{M}\sum_{i=1}^{M}(p_k^i)$. Finally the volatility of the distributions can be defined as:

$$\sigma_p = \sqrt{\sum_{k=1}^{n} w_k(\sigma_k)^2}. \tag{37}$$

Note that we only observe P, the cumulative distributions, and not p, the instantaneous distributions, hence p_k in the equations above will have to be substituted by the numerical derivative of P as $p_k = \frac{P_{k+1} - P_k}{\delta t_k}$.

3.4 Confidence intervals

In order to find confidence intervals for the estimated parameters we may use the method of bootstrapping [7]. To find bootstrapped statistics of our parameters we simply draw with replacement a sample of the cash flows M times from our M funds and compute an estimate of the parameter from this new sample of funds. This can be repeated as many times as necessary to produce a set of bootstrapped estimates for the parameters. A confidence interval is then constructed by taking the quantities (e.g. of the 95% interval of the set).
Because of the numerical minimization being done to find estimates of some of the parameters this method is very time consuming but thankfully it only needs to be done once.

3.5 Simulating paths

After finding estimates for the parameters the future cash flows can be simulated by discretizing the equations (4),(14) and (15). It is straightforward to simulate the paths of the cashflows many times by drawing consecutively from a normal distribution. From these simulations we can derive Monte-Carlo statistics of the cash flows paths such as the Expected path, gotten by taking the arithmetic mean of the path at each time step and appealing to the law of large numbers. Similarity we may obtain estimates for the quantiles of the paths by ordering the results for each simulation at each time step and taking the kth order statistic of the sample where $k = floor(p * M)$ with p and M being the desired quantile and the number of simulations respectively and $floor$ denoting the function rounding to the closest int smaller than the given value [24].

3.5.1 Feller condition

To ensure that the process for δ_t in equation (4) stays strictly positive the following condition, sometimes known as the feller condition, see [8], needs to be fulfilled:

$$2k\theta > \sigma^2. \tag{38}$$

When discretizing the equation however this is not enough to guarantee the process stays above zero, which is a necessary condition for our model. A common solution, employed for example in [6], which is employed in this analysis, is to substitute $\sqrt{\delta_t}$ with $\sqrt{max(\delta_t, 0)}$ to ensure the process stays well behaved.

4 Backtesting

Two primary mathematical methods will be used to access how well the method models real life data.

4.1 Coefficient of determination

The coeficcient of determination, commonly denoted as R^2, is generally described as the proportion of variance in the dependant variable that can be explained by an independent variable. In our situation the dependant variable will be the vector of observed cash flows and the independent variable the expected value of the cash flows produced by the model.
The coefficient of determination is the calculated as

$$R^2 = a - \frac{SS_{res}}{SS_{tot}} \tag{39}$$

where

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \tag{40}$$

with y_i being the individual observed data and $\bar{y}$ the mean, and

$$SS_{res} = \sum_i (y_i - x_i)^2 \tag{41}$$

with x_i is the predicted value. [14]
The coefficient of determination will in this situation serve as a measure of how well the model describes the observed cash-flows on average, in order to judge how well the model captures the variance observed in the cash flows of private equity funds we will analyse the quantiles of the cash flows, detailed below.

4.2 Backtesting quantiles using binomial distribution hypothesis testing

In this thesis we will produce quantiles based on simulations of the cash flows, as a reality check to see if these quantile results are reasonable we can run a backtest of the quantiles using the methodology below [15].
A simple hypothesis test to check at what confidence level the hypothesis that a model is correct can be rejected is to run a backtest and check the number of exceedances against the expected amount. This is a common idea employed when it comes to backtesting qunatiles in a VaR model, which is very similar to the way it's employed in this thesis [20,21]. The way this works is as follows:

- Observe n past outcomes thought to come from specific distributions.
- Find the quantiles of each of the distributions for a confidence level p.

- Count the number of cases k where the the outcome is more extreme than that of the p quantile.

- Pose the null hypothesis that the model is correct, i.e. $H_0 : p = p^*$ where p is the probability we believe is true, and p^* is the true probability of an exceedance.

- Use the binomial distribution to find probabilities $P(X < x_1)$ and $P(X > x_2)$ so that $P(X < x_1)+P(X > x_2) = \epsilon$ where ϵ is the value that is closest to, but still lower than, a specified confidence interval α (often set to 0.05 or 0.01). Many intervals will satisfy this criteria so we generally seek the one that satisfies $P(X < x_1) \approx P(X > x_2) \approx \frac{\epsilon}{2}$ Using our notation above the calculation for the probabilities are:

$$P(X > x_2) = \sum_{i=x_2+1}^{\infty} \binom{n}{i} p^i p^{(n-i})\tag{42}$$

  and

$$P(X < x_1) = \sum_{i=0}^{x_1-1} \binom{n}{i} p^i p^{(n-i})\tag{43}$$

  [18]

- For a given level of α, e.g. 0.05, we then find the corresponding values of x_1 and x_2. If we have that $k \notin [x_1, x_2]$ we would reject the null hypothesis $H_0 : p = p^*$ at the ϵ significance level.

4.2.1 Power of a binomial test

A way to gauge if you have enough trials in your test for you to trust the results is to find the power for the test statistic. The power, denoted $1 - \beta$, is the probability of observing an outcome from a binomial distribution with the parameter p^* given the α confidence interval constructed with the value p. A value for the power is therefor given by:

$$1 - \beta = 1 - (P(X \leq x_2) - P(X \leq x_1 - 1))\tag{44}$$

Using this you can determine if you have enough trials in your test to trust your method, this of course depends on your tolerance for different values of p^* than p potentially being admitted. You can also use this method to back out the needed number of trials to assure that you have a high enough power level for a given value of p^*. [18,22]

5 Illustrating the output of the model

Before applying the model using historical data we first want to verify that the method produces reasonable results to get an idea of what to expect when performing the analysis on real data, as well as to convince ourselves that the method is implemented correctly.
To this end mock cumulated contributions and distributions are simulated by assuming parameters for the processes and simulating paths as described in section 2.5. The parameters used for simulating the values and the corresponding calibrated parameters are shown in table 1. The simulation was done at first with 100 time steps and 10^6 simulations to ensure that the granularity of the simulation process wouldn't be the cause of any poor results when performing the calibration and visualising the results. The produced cumulated contributions and distributions, from the simulation using our assumed parameters, along with the corresponding J-Curve is shown in figures 1,2 and 3 respectively. As we can see from the shape the method seems to produce results that are in line with our expectations. Looking at table 1 we see that the calibration method seems to work well for most parameters with only σ_δ producing dubious results. This is because of the different methodology used for this parameter compared to the other ones, namely it being estimated as a mean of the volatility at each time step. As a result the accuracy of this parameter estimation seems to converge much slower to the true value as the number of time steps are increased, as the number of time steps are increased the value of σ_δ moves closer to 1.5.
To get a better understanding of the number of funds needed to produce reliable values we again perform the calibration using more realistic values for the time steps and the number of simulations. It is mentioned in section 1.1 that the time period for funds are in the magnitude of 10 years, assuming that we would model the time period in 6-month increments that gives us a value of around $n = 30$ for the number of time steps, allowing for longer than average fund lifespans. The calibrations are then done using 10,20 and 100 simulations (representing available past funds used for calibration) and the results are shown in tables 8,9 and 10. In the tables we show the 0.05 and 0.95 bootstrap quantiles for each value with the calibrated output in between. We observe that with only 10 funds we are liable to get quite poor results with the true value for some of the parameters falling well outside the confidence interval. With 20 funds (which we will later see is the number of funds we will be working with) the results are noticeably better but still suffer from some of the same inadequacies as with 10. For 100 simulated funds we get very stable and reliable results so this would represent a more optimal scenario for carrying out the analysis.
It must also be noted that the parameters below are calculated using the simulated drawdown rate and the instantaneous distributions directly, when using their respective estimations from the observed cumulated contributions and distributions the results for the volatilities are less accurate. One reason for this could be because of the numerical differentiation amplifying the noise from the random components in the simulated data.

6 Initial analysis and cleaning of the data

The data available for the analysis is a flat file of historical transactions for private equity dating back to the start of 2009 provided by The Institution, a division of one of the leading banks in Sweden working with investments in private equity. After removing duplicate and uncategorized transactions 47.000 rows remained. In order to use the data it first has to be transformed from raw transactions into cumulated contributions and distributions, the primary work on this was done in a relational database management system using a standard query language [17].

6.1 Limitations of the data

Unfortunately no meta data was available for the funds in the file and as such a lot of necessary information was missing and had to be estimated from the transactions. The missing data and how it was estimated is outlined below.

- No initial committed capital available. Since no data on committed capital was available this number had to be estimated as the sum of the contributions. It has been observed [13] that often times a fraction of the commited capital remains undrawn, so this is an issue. One could take the approach of calculating the committed capital as a fraction of the total cumulated contributions, but as no data at all is available on what a suitable fraction could be to use it was decided to not go with that idea.

- No end time date is available. Because of this we only looked at funds with an end-date no later than in 2019 and used the final transaction date as the ending time date of the fund.

- No investment period available. Because of this we forewent the concept all together and assume that the investment period runs the entire duration of the fund. This poses no notable strain on the model.

- No redemption data available. This is potentially a large issue. A redemption is a re-callable distribution, i.e. a distribution made that will increase the remaining capital. Since the transactions don't specify which distributions are re-callable, if any are, we have to make the assumption that none exist. Because of this it is possible that our estimates for the committed capital become inflated but with no data available for us to estimate how much of the distributions are typically recallable we have no basis to make the adjustment.

6.2 results of the data analysis

After grouping the transactions together and discarding still open investments or otherwise unusable funds we are left with 20 funds. This is a fairly low amount of funds and is mainly due to the fact that a lot of investments had to be excluded as they were not yet closed. Using data for still open funds for calibration purpose could be considered but as several figures are calculated from the transactions themselves, e.g. committed capital, this seems dangerous to attempt.

After calculating the cumulative contributions, distributions and net cash position on a monthly basis some key figures were derived for each fund. Table 2 shows the min, average and max amount of some of the key figures across the funds. Notable is that the committed capital seems to be in the magnitude of 10^8 SEK and the average fund multiple is 1.38. The average life cycle of the fund seems to be about 8 years but this is likely not indicative of funds in general as the limited time scale we are looking at will exclude a lot of longer lived funds as they are not yet closed. In the same way our estimate of M may not be indicative of funds in general since it is natural to assume that longer lived funds will produce a larger M.

6.3 Visualising the data

In figures 4, 5 and 6 the average value of the cumulated contributions, distributions and net cash position (cumulated distributions - contributions) at each time point for the data are plotted.

From the plots we can clearly see the expected J-curve behaviour in the net cash positions. Interesting to note is that the cumulative contributions seem to ramp up first after an initial period of lower contributions after the start date of the fund. This can be observed even more clearly by looking at the instantaneous capital drawdowns, given by equation (3) (after discretizing) plotted in figure 7. There we see that a disticnct increase in the speed of contributions seems to happen around year 3. This runs contrary to the assumptions of the model [2] where we model the speed of the contributions to be directly proportional to the remaining commitment (see equation (1)). To see if this behaviour is due to large abnormalities in just a few of the funds or if it is a constant trait through all of them we plot the above mentioned estimated instantaneous capital drawdowns for each of the individual funds and show the results in figure 13. Here we can see clearly that apart from one or two funds we always have a period of lower activity in drawdowns followed by a quick ramping up a few years into the investment period.

7 Calibration analysis and results

In order to perform the calibrations the data has been transformed into two matrices of cumulated contributions and distributions respectively. In order to make the funds directly comparable with each other the contributions and distributions have all been scaled by the commited capital for each fund and the cumulated contributions and distributions have been summed on a monthly basis with each start time of the fund being set to $t = 1$. As the fund lifetimes are of different lengths the size of the matrix is set as the max life time of the funds, 117 months (see table 2). For ther rest of the funds the cumulated contributions and distributions remain unchanged after the end date of the fund in the matrices.
Because we need to calculate δ_k for each month (see equation (21)) we run into issues in the numerical procedures if the time scale is too short since the estimates for δ_k will drop to zero when data is too sparse. This issue seems to arise because of the small amount of data we have to work with being too spread out as a result of using small time periods. Dividing the data into 6-month buckets instead of 1-month ones seems to solve this issue and produce more stable results. As such we have 20 funds with 22 data points each.

7.1 Results

The results of the calibration are detailed in the tables 3 and 4 below for the contributions and the distributions. The numbers in brackets below the estimates are the 95% confidence intervals for the calibration calculated with the bootstrap procedure outlined in section 2.4.
It can be seen from the bootstrapped confidence intervals that k in particular likely varies a lot from fund to fund and that, even under the assumption that our chosen funds can be adequately described by our model, we would need a lot larger sample than 20 to be able to put any real degree of confidence in our estimated parameter. The variance of the contributions also seems very high compared to θ which in turn doesn't seems as dependant on the sample size. All the parameters going into modelling the distributions appear to be quite stable as well.

8 Validating the model and results

We will evaluate the model on two different metrics, first we will determine the coefficient of determination for the expected cash flows predicted by the model for both the contributions and distributions. This will tell us if the expected value output of our model coincides with the average actual outcomes of the cash flows, or if there is a systematic error.
Secondly we will backtest the quantiles of the method. This will tell us if the variability observed in the cash flows can be well modeled by our method.

8.1 Coefficient of determination

In order to gauge how well the model fits the observed cash flows on average we calculate the coefficient of determination (also know as R^2) between the expected outcome of the calibrated model, easily obtained by taking the mean of a Monte-Carlo simulation (how to do this is outlined in section 3.5), and the actual observed cumulated contributions and distributions. The results for the cumulated contributions, distributions and the net cash position is shown in table 5 below. In figure 8 below the the actual cash flows and those outputted by the model are shown.

We can see both from the plot and the R^2 values that the model for the contributions doesn't seem to fit the data very well at all. This appears to be because, as previously noted, there seems to be a pattern in the data of the bulk of the contributions not taking place at the very start of the funds lifetime but rather lagged by a few years. Evidently the types of private equity observed here exhibit patterns not taken into account by the model.
Looking at the distributions instead we see a much more promising fit, there is some under- and overestimation occurring in the start and end of the time periods but overall the fit looks quite nice and considering the small amount of data being used some discrepency can be expected. The R^2 value is also very high, this should be expected as the calibration method basically serves to maximize R^2.

8.2 Back-testing the model

The coefficient of determination calculated above tells us a bit on how well the model fits the data on average. Beyond this we are interesting in knowing if the variability produced by the method accurately reflects that which can be observed. In section 5 we produced simulations based on calibrated parameters in order to produce "quantile paths" of the cash flows, one of the goals of this thesis was to answer the question asking if it is possible to determine with the method how much capital you ought to have on hand in order to better manage your liquidity. One way of determining such a number would be to look at the quantile of the cash flow distributions (up to, say, 6 months in the future), in this way you get a figure similar to value at risk where you expect that in 19/20

cases (if the confidence level is set to 0.95) you will be liquid enough to answer the calls. In order to determine if the quantiles the model outputs we can backtest the values against real data to measure the amount of exceedances, the theory behind this approach is outlined in section 4.2
To perform the backtest we iterate in 6 month intervals from the first startdate of a fund in 2009 until the final data points in our sample at 2020. At each time interval we simulate M sample paths 6 months forward for each of the funds that are active at the time, using as input the cumulated contributions/distributions for each fund up to time t_k, where t_k is the start time of the interval being looked at, in other words we are always looking at the quantile one time step ahead conditional on the actual observed contributions/distributions up to that point. We then look at the 0.05 quantile of the sum of the simulated cash flows and compare it with the sum of the actual cash flows observed in that interval. This is done for each time interval with the portfolio of active investments. In this way we get a form of value at risk estimate for the entire portfolio of investments that the manager might be looking at. The result of the backtest is shown in tables 6 and 7 for the contributions and distributions respectively and the simulated cumulated contributions and distributions are shown in figures 9 and 10 as a reference to the backtesting results, note that these are simulations based on information at t_0 and average over all the funds while the backtests are conditional on information at t_k on an individual investment level so they are not directly comparable.

8.2.1 Backtesting the contributions

The results for the contributions are shown in table 6. Note that the table shows only one single investment. The result would be the same regardless and this way it becomes more clear where the issue lies. The calculated quantiles for the cumulated contributions are uniformly 1 (all capital called upon) almost immediately, looking at figure 9 we can see clearly why this is so: the 0.95 qunatile is close to the entire committed capital even when all the capial is undrawn. The table has been cut short after index 6 as no additional information can be gained past this stage. Because of the obvious poor fit of the contributions not much stock should be put into the analysis of the quantiles.

8.2.2 Backtesting the distributions

From looking at the plots in figure 10 we're led to believe that the quantiles of the distributions ought to be more tractable than those of the cumulated contributions. Looking at the results for the distributions shown in table 7 we see that we have one exceedance on the lower quantile out of 22 trials. We can see that $P_{binom}(X \geq 1) = 67.6\%$, where X is the number of exceedances, meaning that this is a very common result under the assumption that the model is correct. Constructing a binomial test as outlined in section 4.2 using $\alpha = 0.05$ we find that $\epsilon = P(X < 0) + P(X > 3) = 0.02218$, meaning that we have $k = 1, x_1 = 0$ and $x_2 = 3$, and since $k \in [x_1, x_2]$ we can not reject the null

hypothesis that the model is correct at the ϵ significance level.
Note that the outcomes are not strictly increasing in table 7. This is because there are funds of different start and end periods in the portfolio over the 11 year period being looked at. As such some investments will enter and exit the portfolio as time goes on changing the conditional probability distribution of cumulated distributions.
The test above is a good one because it illustrates the usefulness of the model as it deals with the investments on a portfolio scale, because of the low number of trials however the test does not have a lot of statistical power and the probability of a type-II error is large [20]. It is therefor suitable to also test the quantiles on an individual fund level to get a better gauge of how well our method performs.
Using again time periods of 6 months we find that for 308 total trials we have 18 exceedances. This translates to $P_{binom}(X \geq 18) = 28.22\%$ under the binomial distribution, again indicating a very common result.
Determining x_1 and x_2 according to the method described in section 4.2 we find that $P(X < 9) = 0.0274$ and $P(X > 23) = 0.0223$ leading to $\epsilon = 0.0497$. Since $k = 18$ and $[x_1, x_2] = [9, 23]$ we have that $k \in [x_1, x_2]$ and we can not reject the null hypothesis that our model is true at an ϵ significance level.
To illustrate the higher statistical power displayed by the second binomial test we construct power plots [20] using the methodology introduced in section 4.2.1 for the two situations shown in figures 11 and 12. The plots show on the y-axis the probability of drawing a correct conclusion from the hypothesis tests, i.e. not encountering a type 2 error, for different values of p^* (shown along the x-axis). We see from figure 12 that the probability of incorrectly not rejecting the null hypothesis at the 5% confidence level if $p^* > 0.1$ is very low. In fact we have that the power level of the test is 92% for a p^* value of 0.1 which seems adequate for our purposes as we already have strong reasons to believe that $p^* = p$ from the derivation of the method.
It is worth noting that without the introduction of the conditional long run mean of the fund (outlines in section 2.2.1) there would be no exceedances at all and the model would've had to be rejected (at least as a means of calculating quantiles). This could be expected as without the augmentation the cumulated distributions are pulled aggressively to the (then set as constant) long run mean of the fund drowning out any variance.

9 Conclusions

In this section we will critically examine the results of the thesis and evaluate the results. We will also comment on some of the empirical results of the data and examine the usefulness of the model, judging the modelling of the contributions and distributions separately.

9.1 Contributions

It is clear from this study that, at least for the private equity investments used to calibrate the model, a different method is needed to adequately model the contributions. We have seen from analysing the cumulated contributions and the instantaneous speed of the contributions that there appears to be a lag of around 3 years before the contributions start to ramp up in speed, this can be inferred to be the primary reason that the model fails to predict the contributions with any accuracy. As such it must be determined that the model for the contributions as they are now is not useful for the purposes posed by this thesis and that the current method needs to be heavily revamped or built upon in order to be useful. We touch upon potential reasons for the lag in the contributions in section 9.2.
In short it must then be stated that the analysis put forward in [1,2,3] does not prove very useful for modelling the contributions observed in this thesis in a meaningful way. Judging by the reasonable results obtained in the paper by Buchner, Axel[1]; Kaserer, Christoph[2] and Wagner, Niklas [3] that run contrary to those seen here it seems likely that the differences in result obtained are because of differences in the data used. As described in section 6 we are unable to determine much of the specifics of the investments, it is possible that the contracts for the underlying funds used as data in [1,2,3] are specified differently than those for the funds used in this thesis leading to a difference in the behaviour of the cash flows.

9.2 Distributions

From comparing plots of the real and simulated data, the R^2 test and back-testing the quantiles there is reason be believe that the model for the distributions could be used to answer some of the questions posed in section 1.3, especially as a guideline in managing reinvesting the distributed capital. Calculating quantiles of the cumulated distributions of an entire portfolio of private equity investments as done in section 6.2.2 could be valuable information when it comes to managing liquidity risk, especially if the managing institution is reliant on consistent distributions from its investments in order to meet other cash demands.

10 Discussion

In this section we will discuss the shortcoming of the analysis and suggest improvement for further research.

10.1 Potential reasons for the lag in contributions

According to industry professionals dealing closely with private equity investments at The Institution the sort of lag observed in the cumulated contributions is not entirely unexpected. An example given was that if you have an investment period of 4 years it would be more expected to see large amounts of contributions happen in years 3 and 4 rather than in years 1 and 2. One reason suggested for this that is physiological in nature is that the fund manager would be more selective in his investments during the first years, knowing that he still has a lot more time to make well informed analysed decisions. You also wouldn't want your first investment to be a bad one purely for the reason that it doesn't look very good. During the last years of the investment period there would instead be more focus on making the most out of the committed capital, and the fund manager is less selective in his investments, thus leading to an increase in the speed of contributions as observed in figure 7 [12]. With this in mind we can further draw the conclusion that the assumptions of the model are not in line with the behaviour of the data for the contributions, and that further analysis is needed in order to model the contributions using stochastic processes.

10.2 Potential improvements

The biggest area of improvement would be to find a way for the model to describe the observed lag in the contributions. Apart from this the one of the biggest potential errors in the analysis comes from the sparse amount of data. In order to better calibrate the method you would need data from many more funds, see section 5 for more on this. There was also no way to distinguish between different types of investments in private equity, e.g. if the investment is in a buyout or start-up [16], so it was not possible to analyse the potential differences between how the cash flows behave across different types of private equity investments. If more data were available one could perhaps find in which classes of investments the model performs well and where it is a poor fit. The short time span of where data for funds was available also means that we had to exclude longer lived funds which introduced bias in our selection. More information about the limitation of the data are mentioned in section 6.1.

Another improvement could be in the additional augmentation to the model outlined in section 3.2.1. While the augmentation appears to have improved the performance of the method by better modelling the variance (see section 8.2.2) it is not obvious that this is the only choice. An additional complexity that could be added would be to fit another coefficient into equation (15) that would need to be calibrated. In this way we have more degrees of freedom in modelling the variance at the cost of a less simple method.

References

[1] Buchner, Axel; Kaserer, Christoph; Wagner, Niklas *Stochastic modeling of private equity: an equilibrium based approach to fund valuation (2006).*

[2] Buchner, Axel; Kaserer, Christoph; Wagner, Niklas *Private Equity Funds: Valuation, Systematic Risk and Illiquidity (2016).*

[3] Buchner, Axel *Risk Management for Private Equity Funds (2017).*

[4] Cox, John Ingersoll, Jonathan Ross, Stephen. *A Theory of the Term Structure of InterestRates. JSTOR, 1985..*

[5] Lawrence A. Klimko and Paul I. Nelson *On Conditional Least Squares Estimation for Stochastic Processes 1978*

[6] Ludger Overbeck and Tobias Rydén *https://www.jstor.org/stable/3532742?read-now=1seq=6page$_s$can$_t$ab$_c$ontents*

[7] *https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18$_0$5S14$_R$eading24.pdf*

[8] Hansjörg Albrecher , Philipp Mayer , Wim Schoutens , Jurgen Tistaert *The Little Heston Trap (2006)*

[9] *http://pubs.sciepub.com/jfe/5/1/4/index.html*

[10] *Inside Private Equity: The Professional Investor's Handbook*

[11] Metrick, Andrew, and Ayako Yasuda. *"The economics of private equity funds."Review of Financial Studies (2010)*

[12] *Discussions with industry professionals at The Institution managing private equity investments*

[13] Paul Gompers and Josh Lerner *The Venture Capital Revolution*

[14] Montgtomery, Peck Vinning *Introduction to Linear Regression Analysis, 2012*

[15] Hervé Abdi1 *The Binomial Distribution The Binomial and Sign Tests*

[16] *http://www.privco.com/knowledge-bank/private-equity-and-venture-capital*

[17] Connolly, Thomas M.; Begg, Carolyn E. (2014). Database Systems *A Practical Approach to Design Implementation and Management Pearson*

[18] *https://www.value-at-risk.net/backtesting-coverage-tests/*

[19] *https://explorable.com/type-i-error*

[20] Gyln A. Holton *Value at Risk: Theory and Practice*

[21] Hult, Lindskog, Hammarlid and Rehn *Risk and Portfolio Analysis: Principles and Methods, Springer*

[22] *https://sites.calvin.edu/scofield/courses/m343/F15/handouts/binomialTestPower.pdf*

[23] *https://financetrain.com/private-equity-venture-capital-leveraged-buyouts-and-exit-strategies/*

[24] Paul Glasserman *Monte Carlo Methods in Financial Engineering*

11 Tables and Figures

Table 1: Calibration results for the model validation

	k	θ	σ_δ	m	α	σ_P
Parameter value used	2	1	1.5	1.5	0.03	1
Calibrated value	1.988	1.037	1.660	1.496	0.0302	1.0004

Table 2: Average, min and max values of key figures for the funds

	Min	Average	Max
Numer of transaction	102	336	887
M (fund multilple)	1.015	1.374	1.988
Lifetime of fund (in quarters)	17	31	39
Commited Capital	1.47e+7	3.65e+8	5.65e+9

Table 3: Calibration results for the contributions

	Calibrated result	Bootstrap 95% CI
θ	4.7377	[4.21, 5.05]
k	1.7397	[1.29, 7.19]
σ_δ	24.7999	[22.72, 26.67]

Table 4: Calibration results for the Distribution

	Calibrated result	Bootstrap 95% CI
M	1.3537	[1.26, 1.45]
α	0.0597	[0.0486, 0.0696]
σ	1.2847	[1.156, 1.326]

Table 5: Coefficient of determination for the model estimated cumulated contributions and distributions

	Contributions	Distributions
R^2	0.2504	0.9896

Table 6: Backtest of exceedances for 1 fund of the contributions normalized by commited capital

t_k	P=0.05 quantile	Observed outcome	Exceedance
1	0.8944	0.0230	0
2	0.9942	0.0790	0
3	0.9996	0.1587	0
4	1	0.223	0
5	1	0.2957	0
6	1	0.355	0
...	...	...	0

Table 7: Backtest of exceedances for distributions. Values are in cash times a factor of 10^9

t_k	P=0.05 quantile	Observed outcome	Exceedance
1	0.0000	0.0374	0
2	0.0552	0.0732	0
3	0.0844	0.1161	0
4	0.1368	0.1854	0
5	0.2118	0.2351	0
6	0.2552	0.3170	0
7	0.3445	0.3818	0
8	0.4246	0.5015	0
9	0.5769	0.7215	0
10	0.8742	0.9777	0
11	1.0883	1.1614	0
12	1.1555	1.3738	0
13	1.1757	1.3428	0
14	1.3298	1.3971	0
15	1.4546	1.5209	0
16	0.9509	1.0160	0
17	0.8688	0.8634	1
18	0.6609	0.8337	0
19	0.4179	0.4231	0
20	0.3391	0.3769	0
21	0.1711	0.1723	0
22	0.1723	0.1856	0

Table 8: Calibration results for the model validation using 10 simulations

	0.05 quantile	mean	0.95 quantile
θ	0.9196	1.0890	1.2936
k	1.0131	1.2816	1.5589
σ_δ	0.8738	0.9570	1.0483
m	1.3606	1.4273	1.5069
α	0.0244	0.0291	0.0339
σ_p	0.7891	0.8539	0.9059

Table 9: Calibration results for the model validation using 20 simulations

	0.05 quantile	mean	0.95 quantile
θ	0.9542	1.0072	1.0586
k	1.6299	2.2696	2.9144
σ_δ	1.0955	1.2310	1.3673
m	1.4940	1.5265	1.5580
α	0.0288	0.0310	0.0329
σ_p	0.8237	0.9266	0.9403

Table 10: Calibration results for the model validation using 100 simulations

	0.05 quantile	mean	0.95 quantile
θ	0.9858	1.0289	1.0756
k	1.6796	1.9761	2.0676
σ_δ	1.0908	1.1124	1.1329
m	1.4914	1.4981	1.5049
α	0.0298	0.0302	0.0306
σ_p	0.9653	0.9820	1.0382

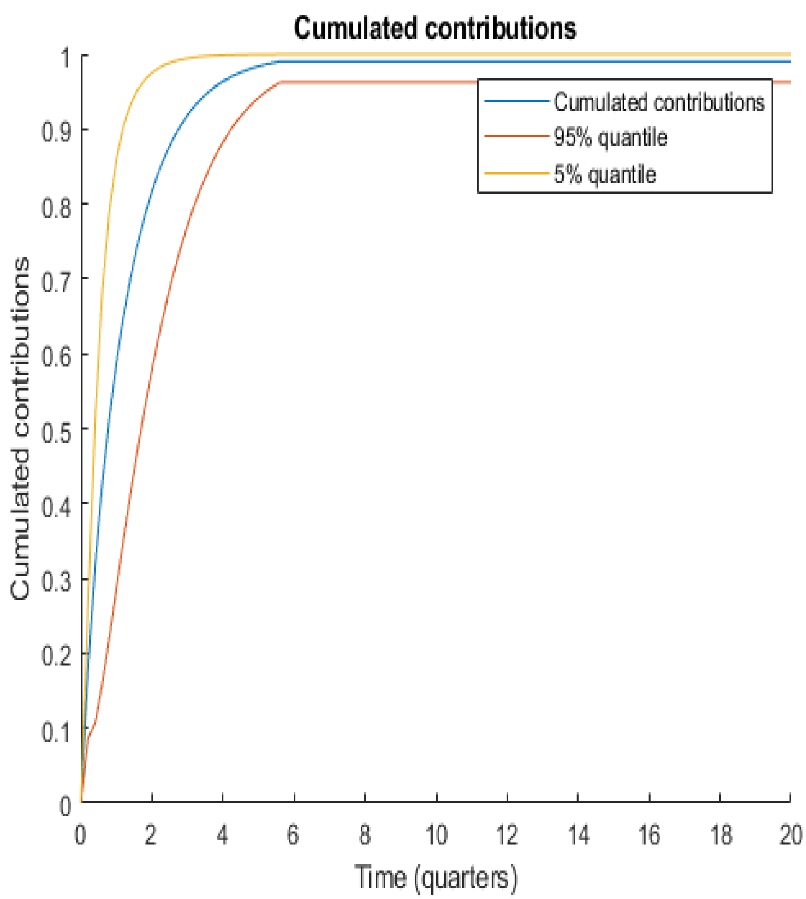

Figure 1: A plot of the cumulated contributions for the model validation analysis

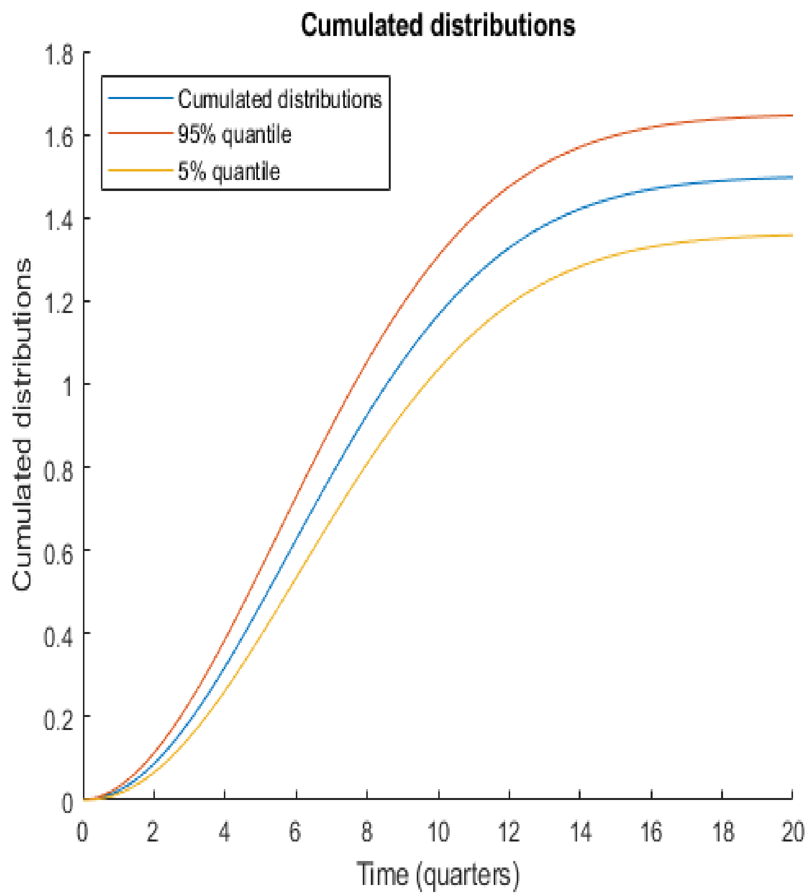

Figure 2: A plot of the cumulated distributions for the model validation analysis

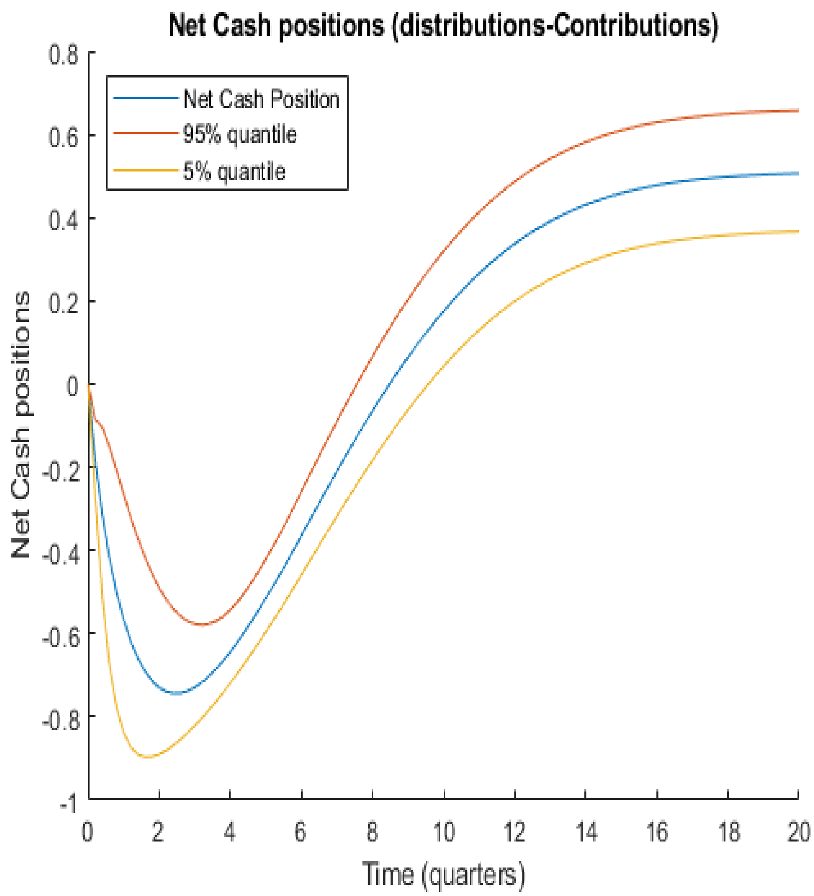

Figure 3: A plot of the net cash position for the model validation analysis

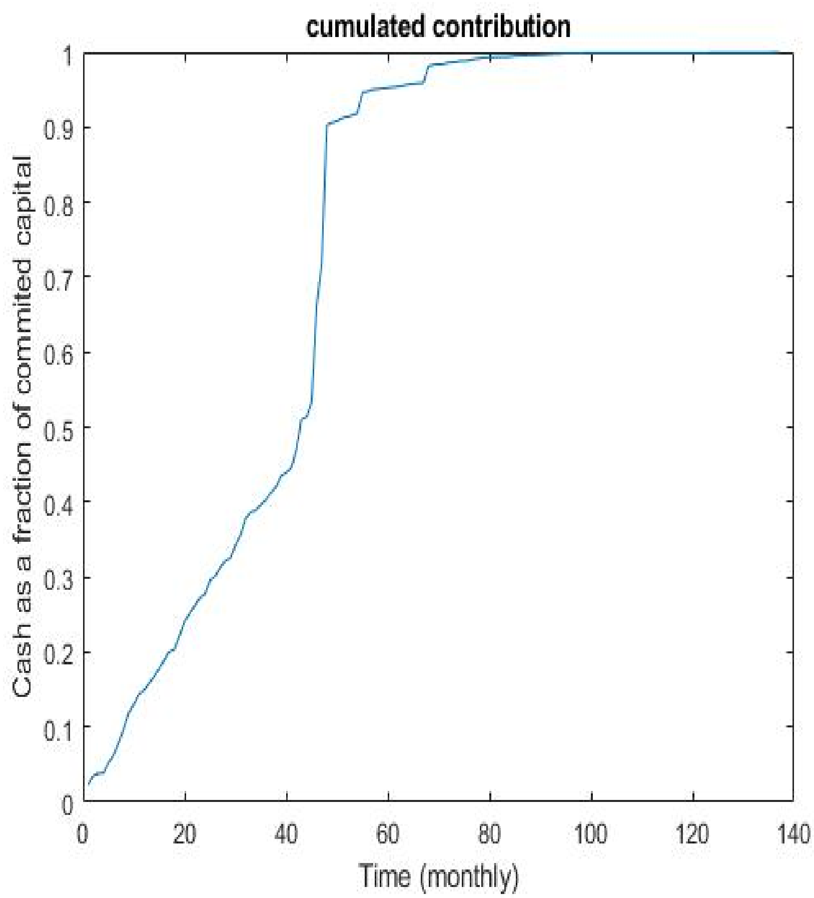

Figure 4: A plot of the cumulated contributions averaged across all the funds for the data

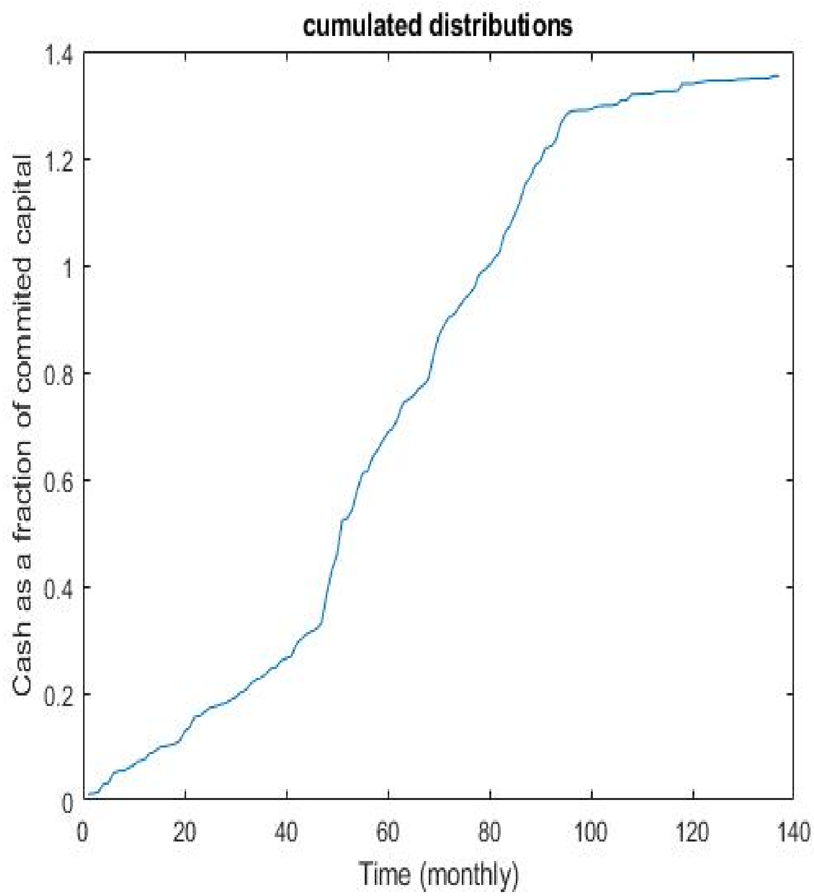

Figure 5: A plot of the cumulated distributions for the data

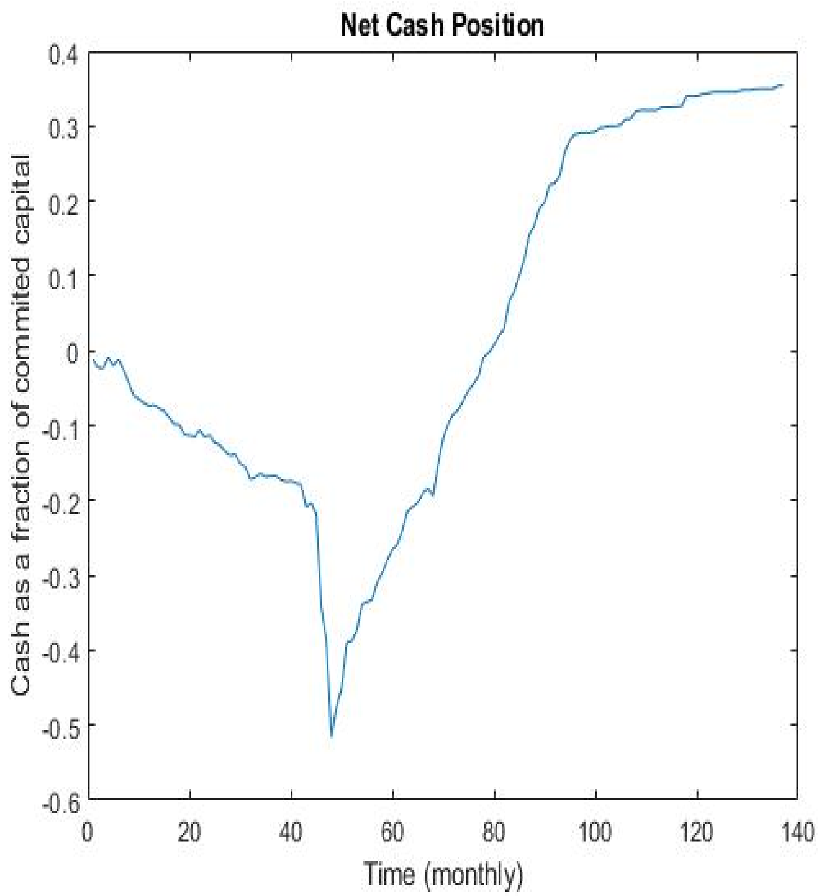

Figure 6: A plot of the net cash position for the data

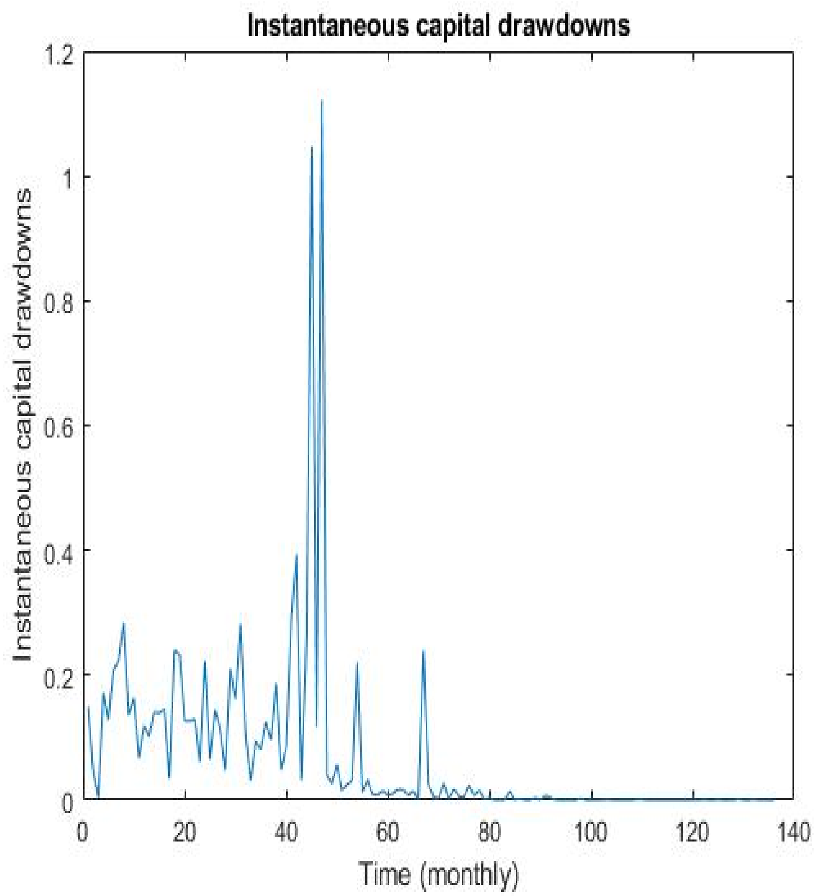

Figure 7: A plot of the instantaneous capital drawdowns for the data

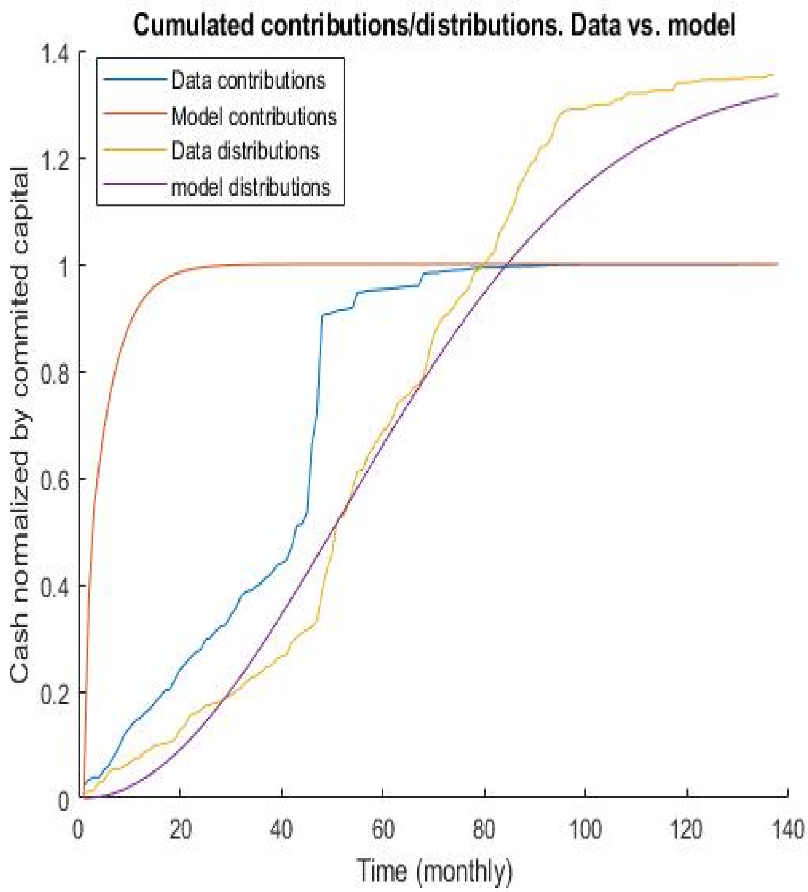

Figure 8: A plot of the average cash flows predicted by the model compared to the actual average cash flows observed

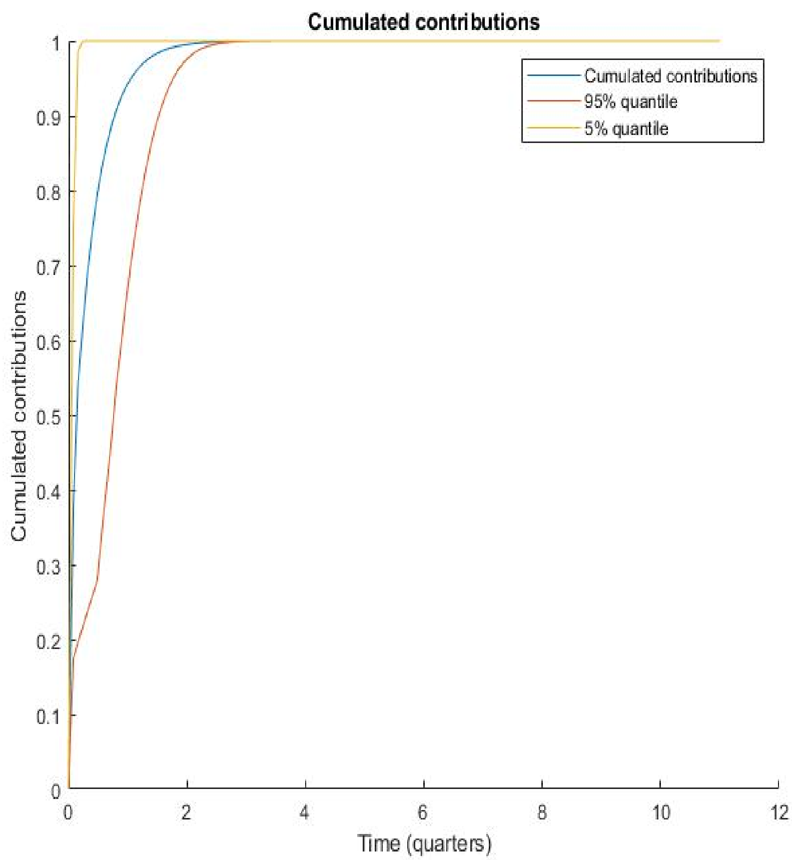

Figure 9: A plot of the cumulated contributions simulated with the calibrated parameters

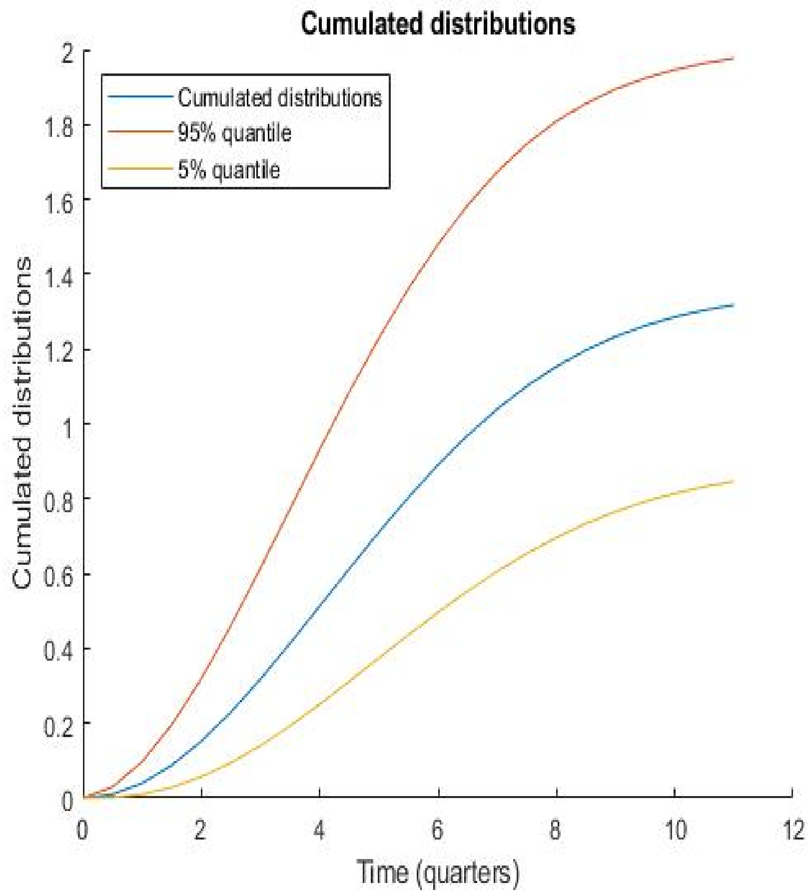

Figure 10: A plot of the cumulated contributions simulated with the distributions parameters

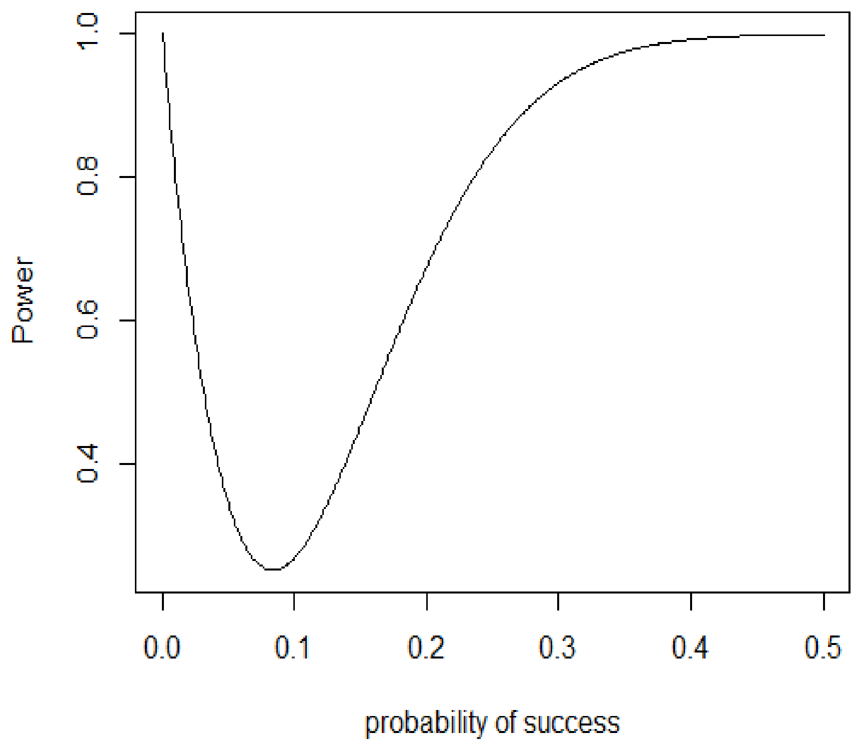

Figure 11: A plot showing the power (1 - β) as a function of different values of p^* for the binomial test with 22 trials and p value of 0.05

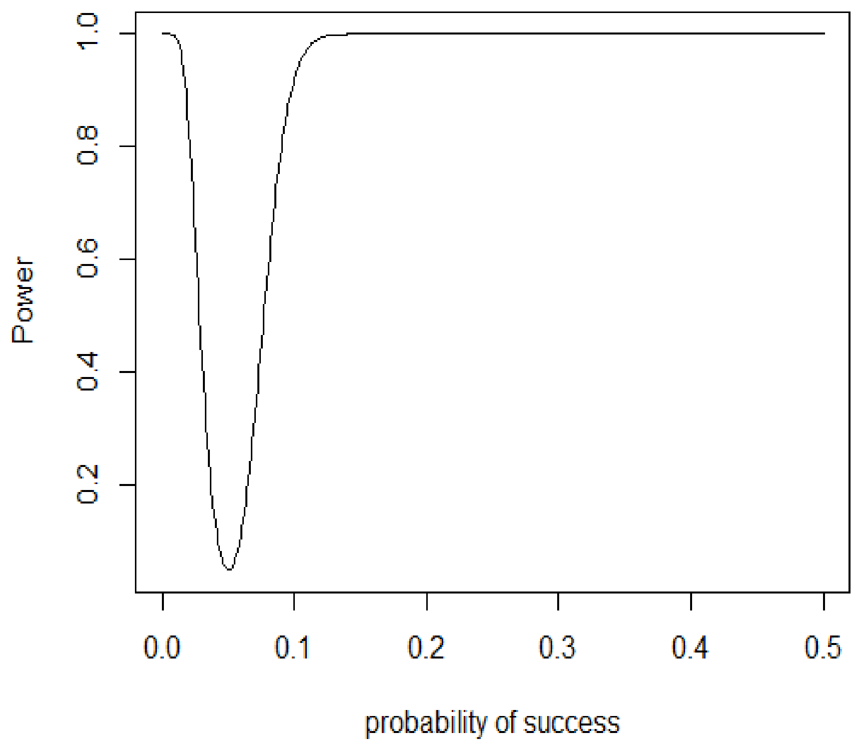

Figure 12: A plot showing the power (1 - β) as a function of different values of p^* for the binomial test with 308 trials and p value of 0.05

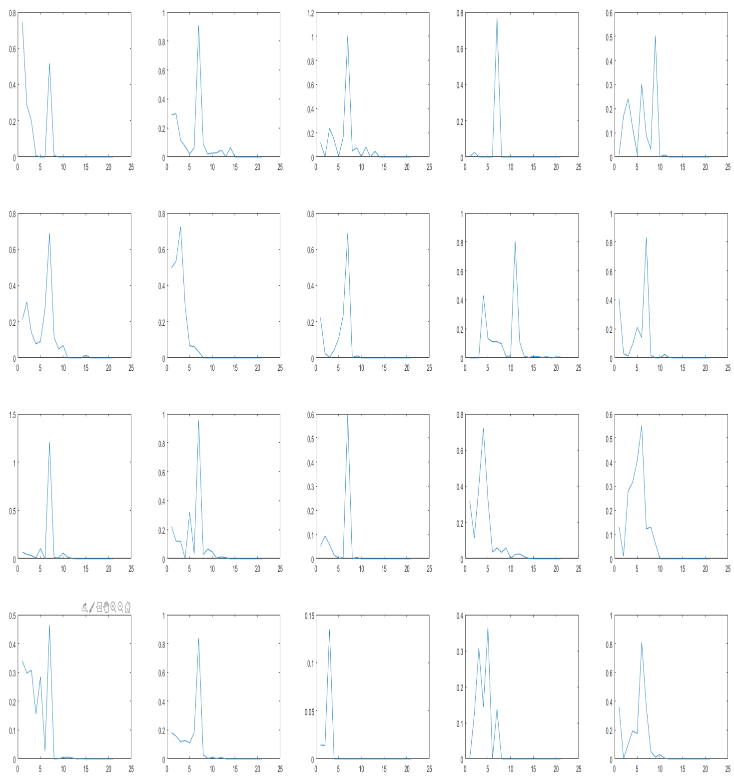

Figure 13: A set of plots showing the estimated instantaneous capital drawdowns for each used fund in the analysis

TRITA -SCI-GRU 2020:062