POLITECNICO DI TORINO

Master's degree
in Mathematical Engineering

Master's degree thesis

Large portfolios credit risk analysis with LT-Archimedean copulas and application to a case of securitised ABS

Supervisor
Prof. Semeraro Patrizia

Candidate
Gallo Vittorio

Company Supervisors
Ph.D. Giovannini Diego Pierluigi
Mazzone Alessandra
Chieppa Valentina

Academic year 2022-2023

Acknowledgements

Intraprendere questo percorso accademico è stata un'esperienza arricchente e trasformativa e sono grato a tutte le persone e le situazioni che ne hanno reso possibile la realizzazione. Sono profondamente grato per la resilienza che questo percorso mi ha richiesto: le sfide che ho incontrato lungo il cammino sono state lezioni preziose, che mi hanno spinto a crescere ed ad adattarmi, alimentando ulteriormente la mia determinazione a continuare.

Un sincero ringraziamento va alla mia relatrice, Prof. Patrizia Semeraro, per il suo continuo sostegno e per la sua infinita pazienza durante tutto il percorso di tesi.

Vorrei ringraziare tutti i supervisori aziendali di Intesa Sanpaolo S.p.A., Valentina, Alessandra e Diego, che mi hanno dato l'opportunità di applicare i miei studi a domini del mondo reale e che mi hanno seguito costantemente per ogni dubbio o curiosità.

Un ringraziamento speciale va a tutti i colleghi universitari, che mi hanno spinto e aiutato a raggiungere importanti traguardi accademici. Non dimenticherò mai i giorni passati a creare importanti progetti (così come quelli spesi a chiederci perché il nostro codice non funzionasse!), ma soprattutto ricorderò tutte le risate e la felicità che abbiamo condiviso in questi anni.

Desidero riservare un ringraziamento particolare a tutti i miei amici più cari, che sono stati i pilastri della mia forza, sostenendomi sia nelle gioie che nelle sfide di questo percorso. Tra di loro, desidero esprimere la mia profonda gratitudine a Tommaso e Giulia, che sono stati solidi punti di riferimento su cui contare e su cui spesso rivolgere il mio carattere testardo.

La mia gratitudine più profonda è riservata a mia madre, Michela, che è stata al mio fianco in ogni occasione e rappresenta per me una fonte di ispirazione. Le parole non saranno mai sufficienti per esprimere la profondità della mia gratitudine per tutto ciò che hai fatto per me. Un profondo ringraziamento va a mio padre, Paolo, per avermi inculcato i valori di dare sempre il massimo in ogni situazione e di non arrendersi mai. I tuoi consigli rimarranno per sempre nella mia mente.

Grazie a tutti, dal profondo del mio cuore, per essere stati parte di questo memorabile percorso.

Vittorio

Contents

List of Figures

List of Tables

List of Algorithms

Introduction

Credit risk, also known as counterparty risk or default risk, is a fundamental concept in the field of finance, especially in the quantitative-related section of this macro-topic. It refers to the possibility of default of a counterpart, called the borrower, that fails to meet the pre-specified financial obligations by not repaying a loan or not meeting any other contractual obligation. The analysis of this type of risk is a critical consideration for banks, financial institutions, investors, and businesses in the procedure of evaluating probabilities of default, pricing financial instruments, and calibration of mathematical models to evaluate correctly the instruments they are proposing or purchasing. To manage the previously explained type of risk in a useful and precise way, a lot of mathematical models have been developed during the last years, especially following the advent of necessary regulations in the financial sector. For example, the Basel III regulatory framework outlines regulatory capital requirements based on quantitative credit risk measures. Quantitative credit risk analysis provides a systematic and data-driven approach to understanding, managing, and mitigating credit risk and is widely approached in the securitisation processes of, for example, CDO, MBS, and other Asset Backed Securities.

In this thesis, following an initial exposition of the economic significance associated with the subject matter and the formulation of procedures governing the analysis of many financial assets, we undertake an examination of a novel quantitative approach for credit risk assessment, based on the article [3] authored by *Cui H. et al.*, which aims to elucidate the asymptotic behavior of the probability of large losses and the Expected Shortfall of a credit portfolio. This project involves the introduction and the application of the LT-Archimedean Copulas, specific mathematical instruments well-suited for discerning the inter-dependencies and risk associations among obligors. Subsequent to the rigorous specification of these mathematical formulations and the creation of a rare-event simulation algorithm, the analytical procedure is applied on an ABS loan data tape provided by *Intesa Sanpaolo S.p.A.* to find the pool's loss tail probabilities and the relative CVaR. Specifically, from the provided dataset two important parameters were esteemed and computed for correctly evaluating the interdependence strength of the pool and its quality. The quantification of interdependence strength among the constituent assets within the pool is achieved via an adapted implementation of Kendall's Tau concordance measure. Concurrently, the appraisal of the credit quality of individual obligors is conducted through a novel methodology grounded in spread-based analysis. Following an exhaustive sensitivity analysis, the loss tail probabilities will serve as the basis for pricing the security that is sourced from the data. Subsequently, the resulting price will be juxtaposed with the pricing outcomes generated through current market practices, enabling to assessment of the reliability and precision of this innovative quantitative approach.

Chapter 1

Economic background

The fixed-income market, often referred to as the bond market, is a segment of the financial markets primarily focused on the issuance, trading, and valuation of fixed-income securities. These securities are characterized by their contractual obligation to pay a predetermined stream of cash flows to investors over a specified period, for example, government or corporate bonds. It plays a central role in capital allocation, risk management, and investment strategies, making it a cornerstone of the broader financial landscape. In the new century, this market has undergone significant transformations and adaptations in response to changing economic conditions, technological advancements, and regulatory developments. Fixed-income securities have remained a fundamental component of global financial markets, serving as a key instrument for raising capital and managing risk. To manage the various risks associated with fixed-income securities, numerous credit-based financial instruments have been conceived and put into practice. The ones we will analyze in this thesis, both descriptively and quantitatively, are Asset-Backed Securities, securities similar to bonds with the difference that the notes are repaid through the cash flows generated by the underlying pool's assets themselves.

1.1 Asset-Backed Securities

Asset-Backed Securities (ABS) are a broad class of financial instruments backed by pools of underlying assets, or receivables, which can include various types of loans, such as auto loans, credit card debts, mortgages, or even revenue from specific projects. The history of ABS in the fixed-income market can be traced back to the latter half of the XX century. The concept began to take shape in the 1960s and 1970s as financial institutions, mostly banks, sought new ways to manage and diversify risk. As a consequence, in the subsequent decades, the whole financial scenario saw a significant expansion in the investments related to this type of securities, which brought a "boom" in the early years of the next century. This trend created a sort of "financial bubble", which exploded due to the famous *subprime crisis* in 2008, widely enhanced by the wrong and misleading pricing and construction of a wide range of fixed income instruments, mostly the ones related to residential mortgages (RMBS). In the aftermath of the financial crisis, regulators and policymakers implemented a series of reforms to address the vulnerabilities in the ABS market, which changed and enhanced the quality of the already present regulations in the Basel regulatory framework, as better explained in Section 1.4.

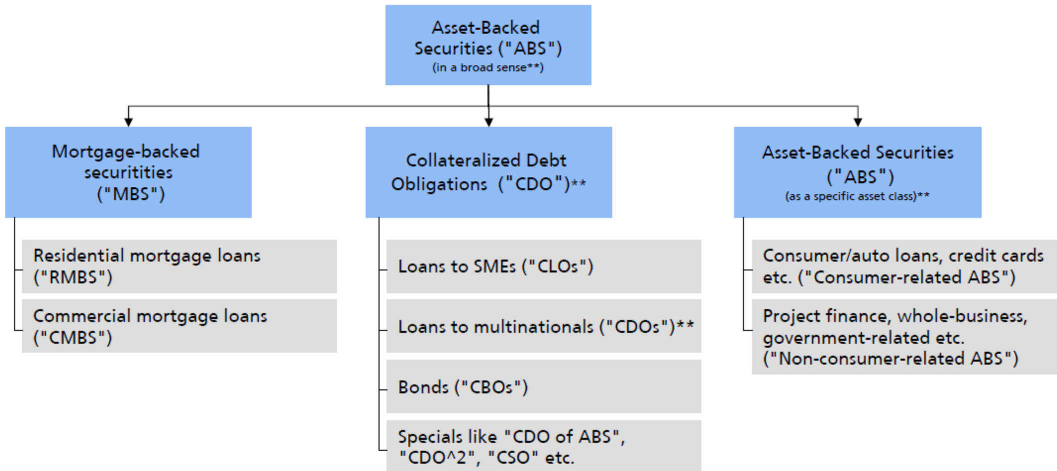


Figure 1.1: Main asset classes in securitisation.

As it can be seen in Figure 1.1, ABS involves a large band of flexibility in terms of asset classes and structures, ranging from the ones based on residential mortgages to the synthetic ones. Assigning these credit instruments offers notable advantages to the banks. They enhance liquidity by transforming illiquid assets, like mortgages and loans, into market-traded securities, which are inherently more liquid. Moreover, through this mechanism, a portion of the credit risk associated with these securities is shifted to investors, who are compensated for assuming it.

But why are these financial instruments a solid and convenient way to invest? First of all, they allow an investor to range over new types of financial instruments, which are extremely diversified thanks to their *granularity* characteristic[1]. Secondly, these instruments offer a wide range of risk classes, which can come to meet the necessities of each risk profile. Concurrently, from the point of view of the issuer, the *securitisation process* helps to shift part of the credit risk to investors, contributing to hedging procedures and acting as a sort of insurance towards the underlying lenders.

1.2 Securitisation process

The financial structure of an ABS is created following a *securitisation process*, such as a procedure that involves different financial entities which contribute to transforming illiquid assets into liquid ones[2]. The main structure of the process is briefly explained in Figure 1.2.

[1]In a real-world scenario, many of these instruments, especially the ones made by consumer receivables, are constructed on hundreds or thousands of fixed income underlying assets. As a consequence, despite the risk of insolvency associated with each customer, the probability of default of all the underlying ones is exceedingly remote. At the same time many others, for example, CMBS, are made up of few elements particularly because of the huge exposition each of them takes.

[2]In Italy, the securitisation procedure is regulated by the law L. 130/99, entered into force on May 29, 1999. Many upgrades have been made to this law, each of them listed in the act cited in [8].

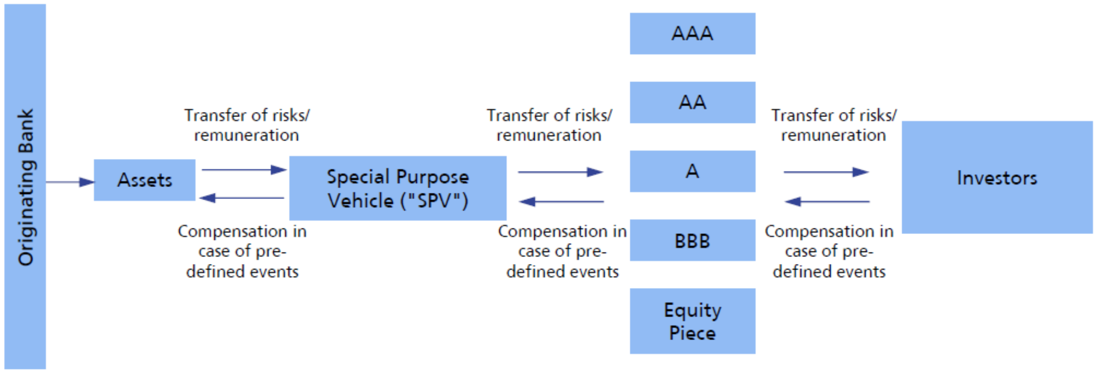


Figure 1.2: Basic securitisation structure.

To outline the securitisation process is important to introduce the main transaction parties:

- **Originator**: the starting owner of the receivables, for instance, banks, corporates, or governments, which collects sets of fixed-income underlying, such as loans or mortgages, and ensembles them to create a huge illiquid pool.
- **Special Purpose Vehicle (SPV)**: also called Special Purpose Entity or Bankruptcy-remote entity, is a subsidiary created *ad hoc* by a parent company to isolate the financial risk. Alternatively, the SPV may be a holding company for the securitisation of debt. Its objective in the procedure is to sever the connection between the pool and the originator, thereby mitigating counterparty risk and also accounting for tax-related considerations.
- **Rating agencies**: is a company that asses the credit quality of other companies or governments. Its aim in this process is to validate the credit structure of the financial instruments proposed by the SPV. Examples of this type of companies are Moody's Investors Service, Standard & Poor's Corporation, and Fitch Ratings.

Besides the parties listed above, a lot of other participants are involved in this process, which can include for instance a Trustee, a Servicer, a Swap Counterparty, and many others. Ultimately, when the financial instrument is made available on the market, potential investors, which could include entities like funds, corporations, or other banks, have the option to purchase it. From their perspective, the advantages of investing in such a financial instrument are associated with the potential for diversification within a novel class of financial instruments.

The securitisation process involves careful specifications, including structuring, credit enhancement, and investor communication to create securities with varying levels of risk and returns. First of all, the originator selects loans, mortgages, or similar, which contractually generate cash flows over time, then ”transfers” them to the SPV: this passage is structured as a sale, with legal ownership of the assets passing from the originator to the SPV. In this starting phase, the securitisation process is additionally equipped with quality-improving specifications, commonly known as *credit enhancement* methods.

1.2.1 Credit enhancement

These techniques are mainly used to diminish the risk of insolvency of the underlying pool's assets, meanwhile trying to make the purchase of such securities more attractive. The choice of credit enhancement methods depends on the specific transaction, the nature of the underlying pool's assets involved, and the desired risk-return profile for investors. The most relevant techniques exploitable in this context are:

- **Overcollateralization**: loans or mortgages are usually backed by tangible assets or collateral. In this case, the originator may allocate more collateral than required to cover the debt's principal and interest payments, providing an additional cushion against potential defaults.
- **Cash reserve**: is a layer of protection that has to be exhausted before noteholders have to bear any losses; it can act as a buffer against payment shortfalls.
- **Excess spread**: mathematically, is the difference between the yield on each underlying pool's asset and the interest paid by these notes[3]. It is an effective measure of how an ABS is risky; specifically the higher an excess spread is, the higher the risk of default associated with the debt instrument. It can also be expressed by calculating the difference between interest payments derived from borrowers and the weighted average coupon paid on the securitized notes.
- **Credit derivatives**: many financial credit-based derivatives can be used in the form of insurance on these notes. Famous examples are the *Credit Default Swaps (CDS)*, such as derivatives used to transfer credit risk from the issuer to a third party. CDS contracts are primarily used to hedge against the risk of default or a decline in the creditworthiness of a specific issuer; in this case, the sensation of large losses of a debt-related instrument may cause the investor to stipulate a CDS contract.
- **Subordination or credit tranching**: is a technique used for structuring different classes of notes, known as tranches, subordinating the re-payments of the high-risky ones to the low-risky ones. In such a way, the riskier tranches bear the largest quantity of losses and serve as a "cushion" for the higher-quality ones. Specifically, each tranche has to bear losses only if all its subordinated ones are fully exhausted by losses.

The previously named tranches have different credit ratings given by credit rating agencies, which usually base their response on both quantitative and qualitative analysis. Indeed, tranching allows for the allocation of risk and return to different classes of investors. Tranches are structured such that the highest-rated tranches receive priority in cash flows and have the lowest risk of default, while lower-rated tranches offer higher potential returns, but also bear more risk. Specifically, the tranches are named:

1. **Senior tranche**: is the highest-ranked[4] and safest segment of an ABS structure. From an investor's point of view, it is more attractive when one's financial profile is risk-averse.
2. **Mezzanine tranche**: the middle-level tranche offers the investors higher yields when compared to the previous one. Specifically, they have a lower rating than the senior tranche, typically AA or below.
3. **Equity tranche**: also known as the junior or subordinated tranche, is the riskiest tranche involved in this type of security. This tranche is the most remunerative one in terms of yield as well as the one that involves the highest risk. It is often unrated.

[3] Additional senior costs, such as rating fees and legal costs, must be added to this spread.

[4] In some cases, the could also be a higher tranche called *Super-senior tranche*, which is the most secure one among all the others.

As explained in the list above, the three sections differ both in risk quality and yield, generating different investment opportunities. Following this subdivision, the risk associated with each tranche is linked peculiarly to its relative yield, reflecting the risk-premium associated with a potential investment, as can be seen in Figure 1.3. Specifically, the *discounted cash-flow* methodology is used to establish the prices and the fair interest rates to associate with each tranche. Moreover, this categorization helps to satisfy the financial interests of different profiles, from those seeking safety and stability to those seeking higher yields with greater risk.

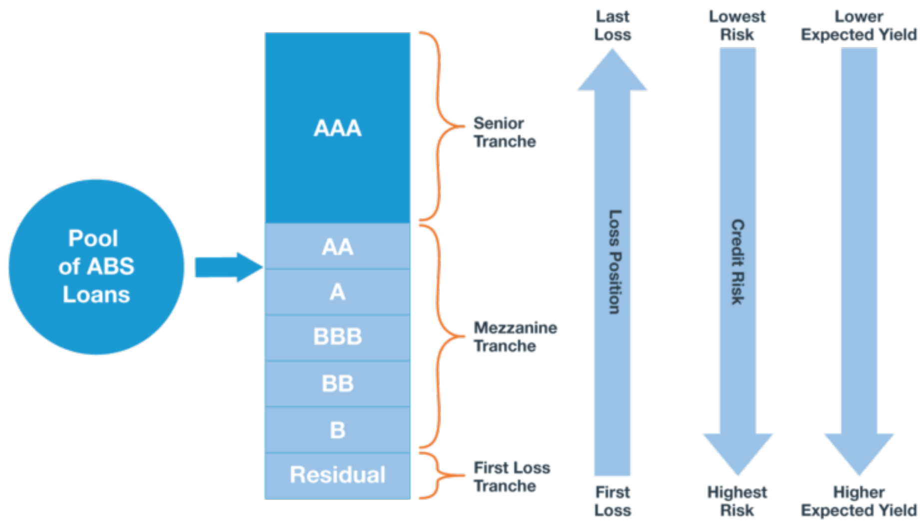


Figure 1.3: Structure of an ABS divided in tranches, outlining the associated evolution of expected losses, credit risk and yield.

For further in-depth settings, some important specifications have to be made on the repayment structure. The payoff of a general Asset-Backed Security is regulated by a so-called *waterfall structure*, such as a repayment rule that involves sequential and subordinate payments, from which the appellation waterfall. Specifically, the repayments are scheduled following a well-defined form of protection. The main one is the *Sequential repayment approach*, according to which a tranche receives all principal payments until it is fully paid off before the next tranche receives any principal. Subsequently comes the *Pro-Rata repayment approach*, in which tranches receive principal payments proportionally, ensuring that most of them receive a share of principal payments simultaneously. Despite that, in any case, the second structure cannot be referred to as the equity tranche. In a practical real-world context, this last *form of protection* is associated with the advent of predetermined trigger factors, such as the raising of the Cumulative Default Ratio (CDR). In such cases, when any trigger event comes to light, a *switch* between the first and the second repayment structures happens.

In the final phase of the securitisation process, after the segmentation into tranches and the formulation of the debt derivative instrument, the issuance is introduced to the market, transforming it into an investable, therefore liquid, asset. Investors on this type of asset receive principal and interest on tranches while paying a lower price in the acquisition. Consequently, they assume a spectrum of risks inherent to this debt instrument. Hence, to build successful investment strategies involving these instruments, is mandatory to make technical considerations on the risks associated and subsequently decide how to manage them.

1.3 Risks in the ABS market

In the re-payment process of an Asset-backed Security, the underlying pool's assets generate interests and principal payments, as well as potential losses may occur when the obligors of those receivables do not serve their obligations, for instance, payments are not delivered following the pre-specified contractual agreements. Hence, investors in this type of financial instrument have to bear numerous and different types of risk arising from either the structure, the underlying assets of the securitised portfolio, or the contractual rules that determine the consequences investors have to deal with. The primary categories of risk that investors encounter when engaging in fixed-income securities investments, as delineated in the literature [4] and [14], include:

- **Credit risk**: this risk profile directly mirrors the creditworthiness of issuers or obligors, encompassing several facets, including *default risk*, which is the probability of an issuer failing to meet its predetermined obligations. *Credit-spread risk*, which entails the necessity of a higher interest rate spread due to a perceived increase in default risk. *Downgrade risk*, which involves the perception of a reduced credit quality rating by the rating agencies mentioned in Section 1.1.
- **Interest-rate risk**: split up into level-risk and yield-curve risk, is the risk associated with adverse changes in the values of the interest rates.
- **Reinvestment risk**: the cash flows originating from fixed-income security are typically presumed to undergo reinvestment, making the variability in returns a crucial factor that significantly impacts these types of obligations.
- **Call/prepayment risk**: as certain bonds, specifically designated as *callable bonds*, incorporate an embedded option that grants the issuer the prerogative to redeem all or a portion of the bond issue before its scheduled maturity date, it is pertinent to acknowledge that the acquisition of such obligations from an investor's standpoint entails a different type of risk. Specifically, when observing the external market conditions, if the interest rates are lowering, there is a high probability of an option exercise of these types of bonds. Characteristics of callability are well outlined in the contract.
- **Inflation, or purchasing-power, risk**: This risk arises when the purchasing power of money decreases. The variation of the inflation rate may diminish the reinvestment opportunity of the generated cash flows too.
- **Liquidity risk**: by measuring the liquidity of a debt instrument with the bid-ask spread, this risk is correlated to the possibility of selling a debt instrument at a lower price rather than at its true value.
- **Exchange-rate, or currency, risk**: when the purchase of a security is carried out in a state with a foreign currency, the exchange rate has a particular impact in evaluating the quality

of an investment. Cash flows may be affected by exchange rate inaccuracies, thus their value may be lowered both from a profit and a reinvestment perspective.

- **Volatility risk**: the expected volatility of interest rates may affect the value of a debt instrument, especially for callable bonds. A good measure of this type of risk is offered by multi-factor risk models, and is referred to as *Vega*: it represents the price variation of a security concerning its implicit volatility.
- **Commingling risk**: fixed-income cash-flows derived from borrowers are typically collected by the Servicer, so they can be mixed in other internal accounts not securitised. In case of insolvency of the Servicer, these cash-flows may undergo in its bankruptcy estate.
- **Set-off risk**: the risk associated with a set-off operation by a borrower about the securitised loan.

During the ownership phase of a debt instrument, various other categories of risk may arise, albeit in a predominantly qualitative context. For instance, **Political or legal risk** refers to alterations in the legal framework governing a bond, consequently impacting its inherent characteristics such as maturity, callability, tax implications, and so forth. Additionally, macroeconomic dynamics and factors may affect the quality of a fixed-income investment. **Event and Sector risks** are the ones that refer to the possible happening of critical events that can affect a specific investment. Simultaneously, the likelihood of unfavorable divergent movements within specific market sectors should also be taken into account as a risk factor when acquiring a debt instrument.

Although not all of the risks described in the list above are quantifiable, the primary risk factors associated with any bond index and portfolio, for example, credit risk and interest-rate risk, can be quantified within coherent evaluations. *Credit risk* can be quantified through probabilistic models, based on credit scoring, feature importance, and interdependence structures. *Interest-rate risk* can be evaluated by observing yield curves and predicting the temporal evolution of the referred rates, for example taking into consideration and calculating risk-reflecting measures, commonly known as *credit spreads*. Taking into consideration the totality of these risks is a fundamental prerequisite for credit institutions, such as banks. To ensure that, during the last decades many regulatory agreements have been emitted and internationally acknowledged.

1.4 Regulations in the fixed-income market

Another important topic to be outlined is the regulation conditions present in the *Basel agreements*. The aim of the *Basel Accords* is to enhance the stability and integrity of the global banking system by establishing minimum capital and risk management standards for banks. They are widely adopted by countries and financial institutions around the world, although implementation may vary from one jurisdiction to another. Specifically, *Basel I*, the first one outlined in 1988, introduced risk-weighted assets and an 8% minimum Capital Adequacy Ratio (CAR)[5], mainly focusing on credit risk. *Basel II*, defined in 2004, expanded on *Basel I*, adding operational and market risk, allowing internal risk models and requiring regulatory approval. *Basel III* responded to the 2008 financial crisis with stricter capital and liquidity rules, higher CAR, and buffers against losses, addressing for instance leverage risks. In conclusion, *Basel IV*, which has not been established yet, continues to refine *Basel III*, aiming for consistent and comparable risk-weighted asset calculations

[5]CAR is a measure that quantifies how much capital a bank has available, reported as a percentage relative to the bank's credit exposures subjected to risk-weighting.

across banks, mostly focusing on a quantitative point of view. Summing up, the final objective is to evaluate in the most possible correct way the risks associated with the fixed-income instruments, thus characterizing and analyzing them quantitatively. As a consequence, in this thesis, we will specifically concentrate on the quantitative analysis of the credit risk of bond portfolios or fixed-income financial instruments based on the credit quality of the underlying assets. Furthermore, the most emphasis will be put on the ABS instruments, which can be treated as collections of fixed-income securities, each with the possibility of default.

Chapter 2

Mathematical model

Since the onset of the XXI century, the field of quantitative finance analysis and research has witnessed numerous theoretical advancements. Given the substantial shifts in market dynamics and conditions, the emergence of more sophisticated market participants, and the growing complexity of investable assets, most of the research has focused on the quantitative pricing of corporate debts. However, many of these models have failed to describe real-world phenomena. In the rigorous quantitative analysis of fixed-income instruments, encompassing both pricing and risk evaluation, it becomes virtually imperative to identify and quantify the entirety of the risks to which an investor is susceptible. The mathematical modeling of these underlying factors serves as a significant tool to enhance the precision of characterizing the attributes of the financial instruments under examination.

In this chapter we will provide a concise overview of the key models used in quantitative credit risk assessment, subsequently, we will concentrate on a new approach that deepens into the probability estimation of large losses arising from defaults in a credit portfolio, using the *Archimedean copulas*. For the sake of organization and clarity, all the relative MATLAB codes can be found in Appendix A.

2.1 Credit Risk quantitative modeling

Credit risk quantitative modeling is a specialized domain within quantitative finance that employs mathematical and statistical techniques to assess and quantify the credit risk associated with forms of debt instruments. It revolves primarily around the risk that the issuer of a bond may default on its interest payments or, in the worst case, fail to repay the principal amount at maturity. This risk is contingent on the creditworthiness of the issuer and a multitude of other variables, rendering its comprehensive evaluation an indispensable facet within the quantitative landscape of finance. As it holds paramount importance and impact, over the years a lot of derivative instruments have been created and priced, for example, ABS, making the fixed-income market a crucial character in global finance. This specialized topic is particularly essential for financial institutions, investors, and portfolio managers who need to evaluate the creditworthiness of borrowers or issuers of debt securities.

As explained in [4], the literature highlights three quantitative approaches to analyzing credit risk:

- **The structural approach**: the basis goes back to Black and Scholes [2] and Merton [13], relying on the fact that corporate liabilities are contingent claims on the assets of a firm. According to this approach, a firm defaults if its assets are insufficient according to some measure. It can be based on a *classical approach*, which involves only the final non-default threshold, or on a *first passage approach*, which characterizes the default of the asset directly when its market value goes at any time below a pre-specified threshold. As it can be inferred from the specifications above, these models are also known as *threshold models*. The lack of concreteness of these models hides in the unique source of uncertainty given by the asset's market value, for example letting space to "pre-default" events.
- **The reduced-form approach**: built upon a negligible level of information about an asset's market value, this approach defines the dynamics of default in an exogenous manner through a default rate λ, also referred to as *intensity*, which can evolve through time. When following this approach, it's noteworthy that there is a one-to-one relation between the intensity of defaults and the corresponding default probabilities. Reduced-form models provide a flexible framework for modeling the dynamics of multiple-issuer credit risk. However, calibration of the model to market variables is not trivial because of the scarcity of default data and the need to model a large number of parameters simultaneously.
- **Incomplete-information approach**: combines the two previous approaches by selecting their most noticeable features. On one side, it picks the intuitiveness of the structural approach, while on the other side picks the tractability and the empirical characteristic of the reduced-form approach, as explained in [5]. What distinguishes this approach from a traditional first-passage time model is that it assumes that investors do not know the value of the default barrier, so it can be modeled by the presence of uncertainty.

The primary objective of these models, whether in theoretical or empirical contexts, is to accurately establish the probabilities of default associated with fixed-income securities. This determination serves as the basis for establishing fair prices or spreads applicable to debt-derivative instruments. In the particular process of default probability estimation, it is important to take into consideration a dependence between defaults among different firms or assets in pre-determined pools. To substantiate this claim, empirical evidence has revealed that financial institutions are affected by common macroeconomic dynamics, such as systematic or sector-specific factors. Therefore, obligors tend to exhibit stronger dependence in stressed scenarios, consequently resulting in more likely simultaneous defaults.

To elucidate an approach grounded in mutual dependence among obligors, certain foundational topics have to be made explicit. This entails a specific emphasis on the mathematical modeling of a dependence structure that manifests robust interdependence, even within the tails of the underlying random variables. To correctly model those financial random variables, it is also necessary that the tails should manifest non-negligible probabilities.

2.2 Copulas

As well explained in the literature [15], copulas are mathematical functions that join or "couple" multivariate distribution functions to their one-dimensional marginals. Copulas find applications in various fields, including hydrology for modeling multivariate rainfall patterns, insurance for modeling joint claims distributions, and finance. Copulas in finance aims to provide a robust framework for modeling and understanding the dependence relationships between financial variables, in the case of this thesis the obligors' default. In mathematical details, a copula function, or better a *n-copula*, is defined as follows.

Definition 1. *An n-dimensional copula is a function denoted as $C : \boldsymbol{I}^n \to \boldsymbol{I}$ such as:*

1. *Its domain is $\boldsymbol{I}^n$, where $\boldsymbol{I}$ represents the continuous real interval $[0,1]$, and its codomain is $\boldsymbol{I}$.*
2. *For every n-dimensional vector $\boldsymbol{u}$ in $\boldsymbol{I}$, the copula function applied in $\boldsymbol{u}$ equals to 0 if at least one component of the previous vector is 0. At the same time, if all coordinates of $\boldsymbol{u}$ are equal to 1 except for u_k, it holds that $C(\boldsymbol{u}) = u_k$.*
3. *For all the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\boldsymbol{I}^n$ such that $\boldsymbol{a} \leq \boldsymbol{b}$ element-wise, C has to be n-increasing. A n-copula is n-increasing if $V_C([a,b]) \geq 0$, where V_C is the volume of the hyper-rectangle defined as:*

$$V_C([\boldsymbol{a},\boldsymbol{b}]) = \Delta_{\boldsymbol{a}}^{\boldsymbol{b}} C(\boldsymbol{t}) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} ... \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} C(\boldsymbol{t})$$

and where:

$$\Delta_{a_k}^{b_k} C(\boldsymbol{t}) = C(t_1, ..., t_{k-1}, b_k, t_{k+1}, ..., t_n) - C(t_1, ..., t_{k-1}, a_k, t_{k+1}, ..., t_n), \ \ \forall k = 1...n$$

It can be further proved that for any copula with $n \geq 3$, each k-margin is a k-copula with $2 \geq k \geq n$.

The link between copula functions and n-dimensional distribution functions is established by the *Sklar's n-dimensional Theorem*:

Sklar's n-dimensional Theorem. *Let H be an n-dimensional distribution function with marginals $F_1, F_2, ..., F_n$. Then there exists an n-copula $\boldsymbol{C}$ such that $\forall \boldsymbol{x} \in \bar{\mathbf{R}}^n$:*

$$H(\boldsymbol{x}) = H(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \tag{2.1}$$

In addition to that, if each marginal cumulative distribution function is continuous then the copula is unique, otherwise is uniquely determined on the Cartesian product of the ranges of each marginal.

From the Theorem 2.2 it follows that a copula structure is the restriction to the hypercube of a cumulative distribution function of a vector with uniform marginals. As a consequence, it is also possible to provide an equivalent stochastic representation for copulas using random vectors with uniform marginals. We recall below the main families of copulas used to describe dependence structures in applications.

- **Independence copula**: to formalize the concept of independence among n distinct random variables, it is necessary to introduce a function that not only encompasses no links between marginals but also upholds the essential prerequisites of a copula. This notion finds a well-suited representation in the form of a multiplicative function:

$$C(u_1, u_2, ..., u_n) = u_1 \cdot u_2 \cdot ... \cdot u_n \tag{2.2}$$

- **Fréchet-Hoeffding bounds**: as explained in [15], each copula is bounded by two particular functions, called FH-bounds. The lower bound, defined as $\mathbf{W}^n(\mathbf{u}) = \max(u_1 + ...u_n - n + 1, 0)$, represents almost surely the strictly countermonotonicity linkage between one random variable and the other $n-1$. The upper bound, denoted as $\mathbf{M}^n(\mathbf{u}) = \min(u_1, ..., u_n)$, represents almost surely the strictly comonotonicity linkage between one random variable and the other $n-1$. It is important to notice that the lower bound $\mathbf{W}^n$ fails to be a copula for $n \geq 2$ but is the best possible choice for a range estimation.

- **Elliptical copulas**: they derive their name from their association with elliptical distributions, particularly the multivariate Gaussian distribution, which serves as a cornerstone for this category of copulas. Common examples of elliptical copulas include the *Gaussian copula*, the *t-copula*, and the *Clayton copula*. Each family has its parameterization and characteristics, making it suitable for specific types of data and dependence scenarios. They are particularly valuable when the assumption of multivariate normality is reasonable or when focusing on symmetric dependence patterns. However, they may not be well-suited for capturing more complex or non-linear relationships between variables, such as extreme dependence in the tails of the distributions.
- **Archimedean copulas**: Archimedean copulas are a versatile class of copula functions that simplify the modeling of multivariate dependencies by reducing them to a one-parameter family based on a generator function. They are chosen when the dependence structure between random variables can be described using a single, well-defined function. In addition, these copulas can model tail dependence, which is the probability of extreme events occurring simultaneously in multiple variables. This is crucial not only in risk assessment and reliability analysis but also in the context of environmental science.

In this work, we use Archimedean copulas to model extreme events, formally introduced in the next section.

2.3 Archimedean copulas

Archimedean copulas, introduced by C. H. Ling in 1965 [10], are named after the ancient Greek mathematician Archimedes due to the mathematical properties of its elements, which are closely tied to his work. In particular, a 2-dimensional Archimedean copula behaves like a binary operation on the interval **I**, making it possible to define an axiom similar to the *Archimedean axiom for positive real numbers*. Archimedean copulas are particularly useful for simplifying the modeling of multivariate dependencies by reducing them to a single-parameter family of copula functions. They allow modeling dependence in arbitrarily high dimensions with only one parameter, denoted in this thesis as α, governing the strength of dependence. In the context of rare events, they can effectively model tail dependence, which is the likelihood of extreme events in one variable being associated with extreme events in another variable. An Archimedean copula is formally defined as follows.

Definition 2. *An n-dimensional Archimedean copula is defined by:*

$$C^n(\boldsymbol{u}) = \phi^{[-1]}(\sum_{i=1}^{n} \phi(u_i)) \tag{2.3}$$

where ϕ is called the generator of the Archimedean copula and has to satisfy the following conditions:

1. *ϕ has to be a continuous and strictly decreasing function from $\boldsymbol{I}$ to $[0, \infty]$.*
2. *$\phi(1) = 0$.*

3. *The pseudo-inverse of ϕ, called $\phi^{[-1]} : [0,\infty] \rightarrow \boldsymbol{I}$ is defined as:*

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t < \phi(0), \\ 0 & \phi(0) \leq t \leq \infty. \end{cases}$$

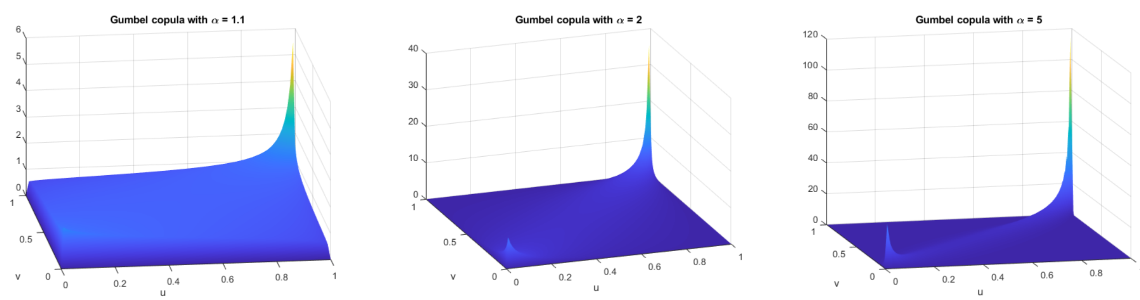
   As a specification, if $\phi(0) = \infty$, then $\phi^{[-1]} = \phi^{-1}$.

4. *The inverse of the generator ϕ^{-1} is completely monotonic on $[0,\infty]$, such as it has to satisfy:*

$$(-1)^k \frac{d^k}{dt^k}\phi^{-1}(t) \geq 0, \qquad k = 0,1,2,...$$

   It means that the inverse of the function we are taking into consideration has to be positive, decreasing, convex, and so on according to the mathematical constraint above[1]*.*

In the practical application of copulas to model specific scenarios, a fundamental requisite is the selection of an appropriate generator function. For this purpose, a lot of copulas have been formulated: among them stand out notable instances, including the Clayton copula, the Ali-Mikhail-Haq copula, and the Gumbel-Hougaard copula. All these copulas fall under the umbrella of one-parameter Archimedean copulas, wherein the specification of the intensity dependence parameter, denoted as α, plays a pivotal role in determining several distinctive and characteristic properties of the copula.



(a) (b) (c)

Figure 2.1: Bivariate Gumbel Archimedean copula representations, varying the dependence parameter α.

As it is noticeable in Figure 2.1, by varying the dependence parameter α, in the corners of the representations the surface gets sharper, underlying a more evident relation of dependence between variables.

A method for constructing Archimedean copulas involves leveraging the concept of convex sums, which give rise to copulas derived from Laplace-Stieltjes transforms of distribution functions. In particular define the *convex sum*, or *mixture*, as:

$$H(\mathbf{u}) = \int_0^\infty \left(\prod_{i=1}^n F_i^\theta(u_i) \right) d\Lambda(\theta) \tag{2.4}$$

[1]In the 2-dimensional case this condition is not explicit, but is a consequence of the fact that ϕ^{-1} is positive and strictly decreasing.

further assume that $\Lambda(0) = 0$ and let:

$$\psi(t) = \int_0^\infty e^{-\theta t} d\Lambda(\theta) \tag{2.5}$$

It is possible to notice that by selecting $F_i(u_i) = \exp(-\psi^{-1}(u_i)),\ \forall i = 1, ..., n$ the expression in Equation 2.4 becomes $H(\mathbf{u}) = \psi\left(\sum_{i=1}^n \psi^{-1}(u_i)\right)$. Furthermore, since $\psi^{-1}(1) = 0$, its marginals are uniform and hence it is possible to see that a source of generators of Archimedean copulas consists of *inverses of Laplace transforms*. Those particular types of copulas are also known as *LT-Archimedean copulas*.

Definition 3. *An Archimedean copula whose generator ϕ is the inverse of the LS-transform of a generic mixing distribution Λ is called an LT-Archimedean copula.*

Moreover, note that for a generic element i, reminding that an Archimedean copula generator ϕ is a decreasing function, it is possible to define the i^{th}*-marginal conditional probability* as follows.

Definition 4. *Let i be a generic index of the random vector $\boldsymbol{U} = (U_1, U_2, ..., U_n)$ defining an LT-Archimedean copula structure. Then, the i^{th}-marginal conditional probability is defined as:*

$$p_i(v) = \mathbb{P}(U_i \leq u_i | \Lambda = v) \tag{2.6}$$

where Λ is the mixing variable of the LT-Archimedean copula.

To have a reformulation of the joint distribution, from the specifications in Equation 2.4 it is noticeable that the LT-Archimedean copula admits the following stochastic representation.

Proposition 1. *The joint distribution of the random vector $\boldsymbol{U} = (U_1, U_2, ..., U_n)$ defined by:*

$$\boldsymbol{U} = \left(\phi^{-1}\left(\frac{R_1}{\Lambda}\right), ..., \phi^{-1}\left(\frac{R_n}{\Lambda}\right)\right) \tag{2.7}$$

where $R_i,\ i = 1, ..., n$ are independent exponential standard random variables, ϕ^{-1} is the LS-transform function and Λ is a random variable, is an LT-Archimedean copula with generator ϕ and mixing variable Λ.

Proof. Thanks to the relationship between an LT-Archimedean copula generator ϕ and its mixing variable Λ, as proved in [12], it is possible to provide to the vector $\mathbf{U}$ a precise stochastic representation. Specifically, note that conditioning on $\Lambda = v$, the variables in the vector made explicit in Equation 2.7 are independent. Following Definition 2.6, in this case, the i-th marginal conditional probability can be reformulated as:

$$\begin{aligned} p_i(v) = \mathbb{P}(U_i \leq u_i | \Lambda = v) = \mathbb{P}\left(\phi^{-1}\left(\frac{R_i}{\Lambda}\right) \leq u_i | \Lambda = v\right) = \\ = \mathbb{P}\left(\frac{R_i}{\Lambda} \geq \phi(u_i) | \Lambda = v\right) = \mathbb{P}(R_i \geq \Lambda\phi(u_i) | \Lambda = v) = e^{-v\phi(u_i)} \end{aligned} \tag{2.8}$$

From the result presented above, also remember that the inverse of the generator of an LT-Archimedean copula ϕ^{-1} is set to be the LS-transform of the mixing variable Λ, as previously made explicit in Equation 2.5, is possible to make evident the underlying LT-Archimedean copula structure:

$$
\begin{aligned}
P\left(U_1 \leqslant u_1, \ldots, U_n \leqslant u_n\right) &= \int_0^{\infty} P\left(U_1 \leqslant u_1, \ldots, U_n \leqslant u_n \mid \Lambda = v\right) \mathrm{d}\Lambda(v) \\
&= \int_0^{\infty} \prod_{i=1}^{n} p_i(v) \mathrm{d}\Lambda(v) \\
&= \int_0^{\infty} \exp\left(-v\left(\phi\left(u_1\right)+\cdots+\phi\left(u_n\right)\right)\right) \mathrm{d}\Lambda(v) \\
&= \phi^{-1}\left(\phi\left(u_1\right)+\cdots+\phi\left(u_n\right)\right)
\end{aligned}
\tag{2.9}
$$

□

As a necessity for the analytical manipulation of formulas and the next specifications, it is extremely useful to find the formulation of the mixing variable Λ in an explicit form. Mathematically arguing, the aim is to find a random variable whose LS-transform is in the form of the inverse of the generator function of the pre-specified copula. As the article [17] explains, two important random variables can be associated in the previous way to the Clayton and Gumbel copulas:

	Generator $\phi(t)$	**Mixing distribution Λ**
Clayton	$t^{-\alpha} - 1$	$\Gamma(\frac{1}{\alpha}, 1)$
Gumbel	$(-ln(t))^{\alpha}$	$St(\frac{1}{\alpha}, 1, (\cos(\frac{\pi}{2\alpha}))^{\alpha}, 0)$

Table 2.1: Table of the associations between Archimedean copulas, generators, and associated mixing distribution.

In our approach, we focus on Gumbel's. In this case, the mixing variable is the *One-sided standard Stable distribution* whose parameters are explained in table 2.1. For a general stable distribution, the density can be written as:

$$f_{\Lambda}(x; \alpha, \beta, c, \mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t; \alpha, \beta, c, \mu) e^{-ixt} dt \tag{2.10}$$

where:

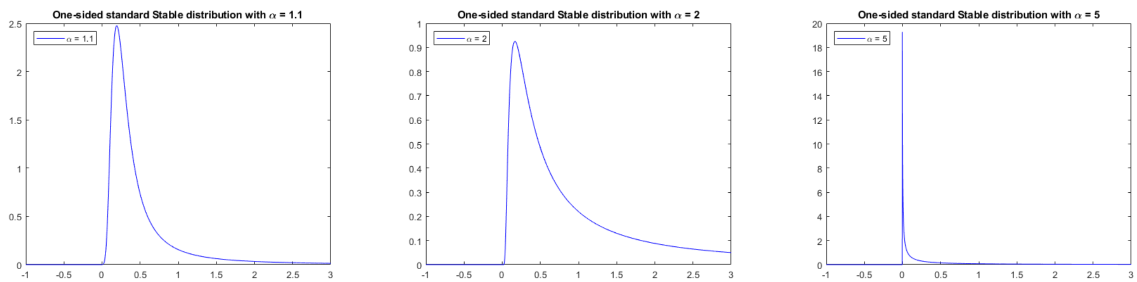
$$\varphi(t; \alpha, \beta, c, \mu) = \exp\left(it\mu - |ct|^{\alpha}(1 - i\beta \operatorname{sgn}(t)\Phi(t))\right), \tag{2.11}$$

$$\Phi(t) = \begin{cases} (|ct|^{1-\alpha} - 1)\tan(\frac{\pi\alpha}{2}) & \alpha \neq 1, \\ -\frac{2}{\pi}\ln(|ct|) & \alpha = 1. \end{cases} \tag{2.12}$$

In the treated specific case, after defining $q = \exp(\frac{-i\pi}{2\alpha})$, the density becomes:

$$
\begin{aligned}
f_\Lambda(x; \frac{1}{\alpha}, 1, (\cos(\frac{\pi}{2\alpha}))^\alpha, 0) = L_\alpha(x) &= \frac{1}{\pi}\Re\left[\int_{-\infty}^{\infty} e^{itx} e^{-q|t|^{\frac{1}{\alpha}}} dt\right] \\
&= \frac{2}{\pi}\int_0^\infty e^{-\operatorname{Re}(q)t^{\frac{1}{\alpha}}} \sin(tx)\sin\left(-\operatorname{Im}(q)t^{\frac{1}{\alpha}}\right) dt
\end{aligned}
\tag{2.13}
$$

The Stable density in Equation 2.13 is not in closed form, but can be evaluated in any point x using numerical methods, for example, the Gauss-Legendre Gaussian quadrature formula. In this case, MATLAB computes the probability density function using the direct integration method, as explained in Stable Distribution by MATLAB [9], with some precautions regarding the parameter $\frac{1}{\alpha}$.



(a) (b) (c)

Figure 2.2: Probability density function of a One-sided standard Stable random variable, varying the values of the interdependence α.

As it is possible to infer from each of the sub-figures in Figure 2.2, by varying the value of the dependence parameter α the probability density function gets sharper near zero, reaching higher values, but maintaining its heavy tailed characteristic.

We conclude this section by observing that Archimedean copulas represent a powerful tool in the world of statistics and probability theory, offering a concise and elegant means to model complex dependence structures among random variables. However, while the utilization of Archimedean copulas offers numerous advantages over alternative approaches, such as their straightforward simulation procedures and analytical tractability, these copulas come with inherent limitations. Specifically, all the k-margins of an Archimedean copula are identical, and typically only one, or at best two, parameters can limit the nature of the dependence structure presented by this family. Moreover, the intrinsic symmetry inherent in Archimedean copulas may not always align seamlessly with the asymmetrical realities of real-world scenarios. In the following section, we further introduce the risk measures used in our approach.

2.4 Risk measures

In the context of credit risk assessment, the main objective is to quantify large loss probabilities or to evaluate, through a precise metric, the risk exposure in determinate cases. As a consequence, many risk measures have been devised during the last decades, which have often led to long debates on their quantitative correctness and applicability. For this purpose, it is necessary to define the concept of loss.

Definition 5. *Consider a portfolio made up of n different assets whose default can be expressed as Bernoulli random variables $X_1, X_2, ..., X_n$. The loss of the portfolio can be defined as:*

$$L_n = \sum_{i=1}^{n} c_i X_i \tag{2.14}$$

where $c_i \ \ \forall i = 1, ..., n$ are the different exposures of each asset in the portfolio.

In the context of tail loss probability estimation, an important measure of risk has to be taken into consideration: the *Value at Risk (VaR)*. This statistical risk measure quantifies the maximum potential loss an investment or portfolio could experience over a specified time horizon at a given confidence level. In practical details, it provides an estimate of the worst-case loss within a defined probability.

Definition 6. *Let L_n be the random variable relative to the loss of a portfolio expressed as 2.14 and F_{L_n} its cumulative distribution function. The Value at Risk (VaR) at level γ of the portfolio loss is defined as:*

$$VaR_\gamma(L_n) = \inf\{x \in \mathbb{R} : F_{L_n}(x) \geq 1 - \gamma\} \tag{2.15}$$

It is important to notice that if the random variable relative to the portfolio loss is continuous, the VaR can be also expressed as $VaR_\gamma(L_n) := F_{L_n}^{(-1)}(1-\gamma)$, where $F_{L_n}^{(-1)}$ denotes the inverse function of the cumulative distribution function of the loss random variable L_n.

In this thesis, we are interested in estimating analytically the probability of large losses, better explicit as $\mathbb{P}(L_n > nb)$, where b is a settable number indicating the total risk exposition assumable on average. This consideration expresses a perfect link with the VaR measure: following Definition 6, the probability value expressed as $\mathbb{P}(L_n > nb)$ symbolizes the level $1 - \gamma$.

The VaR risk measure is widely used in many finance sectors, such as risk management, finance reporting, and financial control, but also finds space in other fields, for instance, environmental sciences, insurance, and healthcare. However, it is a controversial risk management tool principally because relatively considers extreme events, which often happen, and concurrently, in a quantitative way, does not result in a coherent risk measure. To manage that, some other risk-evaluating measures have been developed, one of which has a special relation with VaR: the *Conditional Value at Risk (CVaR)*. CVaR, also known as *Expected Shortfall*, goes a step further than VaR. While VaR provides the expected loss at a specific confidence level, CVaR gives the expected loss given that the loss exceeds the VaR. In essence, it provides a measure of the average loss that would be incurred if the portfolio performs worse than the VaR. In the context of risk assessment, CVaR is a coherent risk measure, such as satisfying the properties of monotonicity, sub-additivity, homogeneity, and translational invariance beyond the one of normalization. Furthermore, it offers a more comprehensive insight into the tail risk and the potential severity of losses in the worst-case scenarios, making it a valuable tool for risk management and decision-making in financial and investment contexts.

Definition 7. *Let L_n be the random variable relative to the loss of a portfolio as expressed in 2.14 and F_{L_n} its cumulative distribution function. Furthermore, let γ be the minimum percentage level of exposure and $VaR_\gamma(L_n)$ the relative Value at Risk of the loss. If the random variable L_n that defines the loss is continuous, CVaR can be defined as:*

$$CVaR_\gamma(L_n) = \mathbb{E}[L_n|L_n \geq VaR_\gamma(L_n)] = \frac{1}{\gamma}\int_{VaR_\gamma(L_n)}^{+\infty} x dF_{L_n}(x) \tag{2.16}$$

As inferable in Equation 2.16, CVaR pays more attention to the tails of the loss random variable compared to VaR because it considers not just the probability of extreme events occurring but also the severity of the losses in those extreme events. This focus on the tails of the distribution makes CVaR a more robust and conservative risk measure, especially when dealing with the potential impacts of extreme events or tail risk. In real-world scenarios, to overcome the shortcomings of VaR, any asymptotic analysis is mainly conducted on the second outlined risk measure. In more mathematical details, it is noticeable that Equation 2.16 can be also expressed as:

$$CVaR_{\mathbb{P}(L_n>nb)}(L_n) = \mathbb{E}\left[L_n \mid L_n > nb\right] = nb + n\frac{\int_b^{+\infty} \mathbb{P}\left(L_n > nx\right) \mathrm{d}x}{\mathbb{P}\left(L_n > nb\right)} \tag{2.17}$$

where b is a settable number indicating the total risk exposition assumable on average. Moreover, especially in the context of simulation, the requirement for a systematic revision of the numerator in Equation 2.17 becomes apparent. Consequently, a potential reconfiguration is elucidated in the subsequent Proposition.

Proposition 2. *Let CVaR be defined as in Equation 2.17. Then, the Expected Shortfall formula can be also explained as:*

$$CVaR_{\mathbb{P}(L_n>nb)}(L_n) = \mathbb{E}\left[L_n \mid L_n > nb\right] = nb + \frac{\mathbb{E}\left[(L_n - nb)_+\right]}{\mathbb{P}\left(L_n > nb\right)} \tag{2.18}$$

where the function $(\cdot)_+$ is the maximum between zero and its argument.

Proof. The reformulation in Equation 2.18 of the CVaR holds thanks to the following specifications. Define $a := \sum_{i=1}^n c_i$ and f_{L_n} as the probability density function of the total loss. Subsequently, denote as $F_{L_n}(x) = \mathbb{P}(L_n \leq x)$ the cumulative distribution function of the total loss and as $\bar{F}_{L_n}(x) = \mathbb{P}(L_n > x)$ the survival function of the total loss. As a consequence, the analysis of the numerator of the fraction in the right part of the equation results in:

$$\mathbb{E}\left[(L_n - nb)_+\right] = \int_0^a [x-nb]_+ f_{L_n}(x)dx = \int_{nb}^a (x-nb) f_{L_n}(x)dx =$$

$$= \int_{nb}^a x f_{L_n}(x)dx - \int_{nb}^a nb f_{L_n}(x)dx = [xF_{L_n}(x)]_{nb}^a - \int_{nb}^a F_{L_n}(x)dx - nb\left[F_{L_n}(x)\right]_{nb}^a =$$

$$= a - nb \cdot F_{L_n}(nb) - \int_{nb}^a F_{L_n}(x)dx - nb + nb \cdot F_{L_n}(nb) = (a-nb) - \int_{nb}^a F_{L_n}(x)dx =$$

$$= (a - nb) - \int_{nb}^{a} (1 - \bar{F}_{L_n}(x))\, dx = \int_{nb}^{a} \bar{F}_{L_n}(x) dx = n \cdot \int_{b}^{\frac{a}{n}} \bar{F}_{L_n}(ny) dy = n \cdot \int_{b}^{\frac{\sum_{i=1}^{n} c_i}{n}} \mathbb{P}\left(L_n > ny\right) dy$$

which gives the same formulation as in Equation 2.17.

□

Since VaR and CVaR look at the distribution of the loss' tail, it is important to model the asymptotic behavior of $P(L_n > nb)$. In the next section, we introduce a class of functions with a convenient asymptotic behavior that applies to Archimedean copulas.

2.5 Regular Varying functions

Regular varying functions are a category of functions that exhibit a particular asymptotic behavior. These functions have broad-ranging applications across various domains: within the realm of probability theory, they serve as instrumental tools for modeling heavy-tailed distributions, a critical endeavor in comprehending sporadic events. Furthermore, these functions are employed in extreme value theory to model the tails of probability distributions, which is crucial in fields like hydrology, climatology, and insurance.

Intuitively, a positive Lebesgue measurable function f on $(0, \infty)$ is a regularly varying function (RV) if it behaves as a power law function near infinity. Regular varying functions are known for their slow growth or decay as their argument becomes large. The characteristic of heavy-tailedness is regulated by a parameter, in the case of this thesis called α, and assumed equal to the level of dependence in the Archimedean copulas described in Sub-section 2.3.

Definition 8. *A function f is said to be regularly varying at infinity with index α, generally $\in \mathbb{R}$, but in this case $\in [1, \infty)$, if for $x > 0$:*

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^{\alpha} \tag{2.19}$$

Moreover, every $f \in RV_{\alpha}$ can be expressed as $f(x) = x^{\alpha} l(x)$, where $l \in SV$ (*Slowly varying functions*: a special case in which $\alpha = 0$).

Over the course of the previous century, a substantial ensemble of theoretical frameworks and precise definitions has been established, offering a valuable compendium for this subject matter, as referenced in [1]. Among these, a particularly significant milestone is the *Karamata's Tauberian theorem*, which holds significant importance in the demonstration of specific concepts enunciated in this approach and relative to Karamata's theory of regular variation.

Karamata's Tauberian Theorem. *Let $\alpha > -1$, $l \in SV$ and $c \in \mathbb{R}$. Let U be a non-decreasing right-continuous function on $\mathbb{R}$ with $U(x) = 0$ for all $x < 0$. The following statements are equivalent:*

$$U(x) \sim \frac{c x^{\alpha} l(x)}{\Gamma(1 + \alpha)}, \quad x \to \infty \tag{2.20}$$

$$\hat{U}(s) \sim c s^{-\alpha} l(\frac{1}{s}), \quad s \to 0^{+} \tag{2.21}$$

where $\hat{U}(s)$ is the Laplace-Stieltjes transform of the function U:

$$\hat{U}(s) := \int_{-\infty}^{\infty} e^{-sx} dU(x) = \int_{[0,\infty)} e^{-sx} dU(x) \tag{2.22}$$

Is important to notice that the function U can represent the cumulative distribution function of a random variable, thus rendering Equation 2.22 the Laplace transform of the probability density function of the same random variable.

In the context of a credit risk model in which the interdependence relationship among assets is structured as an LT-Archimedean copula, it is necessary to outline that the parameter α indicates at the same time the interdependence between obligors and regulates the heavy-tailedness property of the systematic factor Λ. Moreover, it is assumed a strict relationship between the random variable Λ and the copula generator function, as better explained in Sub-section 2.3 in the context of LT-Archimedean copulas. For further analysis, thanks to Karamata's Tauberian theorem adapted to this precise situation, it is possible to show that the generator of the associated Archimedean copula is regularly varying to 1 with index $-\alpha$, as better explained and proved by the following Proposition.

Proposition 3. *Let $F_\Lambda(v)$ be the cumulative distribution function of the mixing variable Λ outlined in equation 2.5 and $\bar{F}_\Lambda(v) = 1 - F_\Lambda(v)$ its survival function. Furthermore, assume that $\bar{F}_\Lambda(v)$ is a regular varying function with index $-\frac{1}{\alpha}$, as its associated variable has to be heavy-tailed. Then, the generator ϕ_Λ of the LT-Archimedan copula associated with the mixing variable Λ is regularly varying to 1 with index $-\alpha$.*

Proof. As a consequence of the assumptions of the proposition, it can be explained that:

$$\bar{F}_\Lambda(v) \sim kv^{-\frac{1}{\alpha}}, \quad v \to +\infty$$

with $k > 0$ and:

$$F_\Lambda(v) \sim 1 - kv^{-\frac{1}{\alpha}}, \quad v \to +\infty$$

For the Karamata's Tauberian Theorem 2.5, satisfied the condition $-\frac{1}{\alpha} > -1$, it results that:

$$\phi_\Lambda^{-1}(s) \sim 1 - \hat{k}s^{\frac{1}{\alpha}}, \quad s \to 0^+$$

with $\hat{k} > 0$. The meaning of the result above is that the inverse of the generator ϕ_Λ^{-1} and the cumulative distribution function have a sort of mirror behavior w.r.t. 0^+ and ∞.
Now, by making the change of variable $s \to \phi_\Lambda(1 - \frac{1}{t})$, the relation above becomes:

$$\phi_\Lambda^{-1}(\phi_\Lambda(1 - \frac{1}{t})) \sim 1 - \hat{k}(\phi_\Lambda(1 - \frac{1}{t}))^{\frac{1}{\alpha}}, \quad t \to +\infty$$

Specifically, the aforementioned outcome is accurate in the limit as t approaches infinity due to the condition $\phi_\Lambda(1) = 0$, given by the second generator's property of an Archimedean copula. To uphold this condition, the unique feasible recourse is to let t asymptotically approach infinity. As the asymptotic equivalence is commutative, consequently to the previous relation results:

$$1 - \hat{k}(\phi_\Lambda(1 - \frac{1}{t}))^{\frac{1}{\alpha}} \sim 1 - \frac{1}{t} \quad t \to \infty$$

$$(\phi_\Lambda(1-\frac{1}{t}))^{\frac{1}{\alpha}} \sim \frac{1}{\bar{k}t} \quad t \to \infty$$

$$\phi_\Lambda(1-\frac{1}{t}) \sim \tilde{k}t^{-\alpha} \quad t \to \infty$$

with $\tilde{k} > 0$. From that it is inferable that $\phi_\Lambda \in RV_{-\alpha}$.

□

Within the domain of credit risk analysis, the establishment of these mathematical structures serves the purpose of laying the groundwork for a robust and reliable model. First and foremost, the incorporation of a copula structure enables the precise technical representation for modeling the interrelationships among variables. Secondly, the imperative need for a risk measure to assess potential extreme scenarios and outcomes holds significant relevance, especially in practical applications. Lastly, the recognition of the existence of rare events necessitates the development of a model that accounts for the heavy-tail property. It is through the incorporation of these model assumptions that a comprehensive credit risk model can be constructed.

2.6 Credit risk model

To well introduce this approach it is necessary to make explicit the main objectives set in a credit risk management scenario. First of all, many financial institutions' exposure to credit risk is not just confined to a single obligor, but to a large portfolio of multiple obligors, or in other cases underlying debt instruments. Secondly, the default of a single obligor is not mandatory independent from other defaults, so a sort of correlation structure is needed to model in a strongly dependent way the succession of default events, for example in a stressed economic scenario. To accomplish the two previous assumptions, the approach of this thesis follows a *threshold model*: in this scenario a single default occurs when a certain variable, usually the firm's debt, exceeds a pre-specified threshold. To hold the previous assumptions, a certain correlation structure is needed to model dependence among defaults. Indeed, when market conditions worsen, simultaneous defaults occur with non-negligible probability. To account for that, classical copula structures, like the Gaussian one, cannot make evident the dependence of extreme events. As explained in Sub-section 2.3, the family of Archimedean copulas fits very well in this context; in particular the Gumbel one has been chosen for simulation purposes, but the model also accounts for interchangeability among other Archimedean copulas.

First of all, it is necessary to consider n different obligors. Following the threshold model, each obligor defaults if its latent variable X_i exceeds a pre-specified threshold x_i. In particular, denoting as c_i the risk exposure, the total loss can be defined as:

$$L_n = \sum_{i=1}^{n} c_i \mathbb{1}_{\{X_i > x_i\}} \tag{2.23}$$

The loss function presented in Equation 2.23 can also expressed in terms of copulas, using an equivalent threshold model. Initially, it can be noted that there is a pre-specified link between marginal default probabilities and portfolio default probabilities: since $U_i = F_i(X_i)$ and

$p_i = \bar{F}_i(x_i), \ \forall i = 1, ..., n$, the threshold models $(X_i, x_i)_{1 \le i \le n}$ and $(U_i, 1 - p_i)_{1 \le i \le n}$ are equivalent threshold models[2] as explained in [12]. Consecutively, it has to be noted that a single obligor default can be correlated in a specific way with the default of other obligors. Mathematically, this fact exhibits the presence of an underlying copula structure, which is better explained in Definition 1. The underlying correlation structure can be furthermore reformulated in other terms: thanks to Sklar's n-dimensional Theorem 2.2, the joint distribution $(X_1, ..., X_n) \sim F(x_1, ..., x_n)$ can be expressed as $(U_1, ..., U_n) \sim C(u_1, ..., u_n)$, where F is the joint cumulative distribution function and C is the associated underlying copula. Furthermore, to reflect the theory of diversification, the probability of large portfolio losses should diminish as the number of obligors increases. Hence, each p_i, which represents the singular probability of having a debt larger than x_i, can be expressed as $l_i f_n$, where:

- f_n: is an n-dependent function converging to zero as n approaches infinity. It reflects the quality of the portfolio and the diversification effect when the number of obligors increases. It also provides mathematical convenience in the analytical calculations.
- l_i: strictly positive constants associated with the credit quality of each obligor and accounting for variations among them. To characterize potential heterogeneity, some restrictions are needed: the set $\{(c_i, l_i) : i \geq 1\}$ takes values in a finite set W and, denoted by n_i the occurrences of each element in the set, it is assumed that each fraction $\frac{n_i}{n}$ converges to a positive value w_i.

Furthermore, to model coherently the interdependencies among obligors, this approach introduces the usage of Archimedean copulas in the credit risk estimation, in particular the LT-Archimedean family introduced in Definition 3. Following this fact, also for analytical purposes, it is possible to specify the vector $\mathbf{U} = (U_1, U_2, ..., U_n)$ with an alternative stochastic representation, as well explained in Proposition 1. As a consequence of its proof, particularly from Equation 2.8, it is inferable that large portfolio losses are essentially determined by the outcome of the mixing distribution Λ and by the associated LT-Archimedean copula generator ϕ. Following the previous specifications, the modified loss function becomes:

$$L_n = \sum_{i=1}^{n} c_i \mathbb{1}_{\{U_i > 1 - l_i f_n\}} \tag{2.24}$$

where the joint vector $\mathbf{U}$ of the obligor's variables U_i can be reformulated as in Equation 2.7.

For the forthcoming computations, it is imperative to maintain a clear focus on the ultimate objective of the approach. To elucidate, the primary goal is to discern the asymptotic characteristics of the tail probabilities associated with the loss function L_n for the pool, as delineated in Equation 2.24, particularly as the number of obligors approaches infinity. In light of the compelling need to systematically model the interdependencies among these obligors, it becomes necessary to establish an underlying copula structure, with a specific emphasis on selecting a copula from the LT-Archimedean family, as defined in Definition 3. Specifically, it is noteworthy to underscore that the generator ϕ is instrumental in defining this chosen copula. Given the presence of this foundational Archimedean copula structure, the reformulation of the vector $\mathbf{U}$, as stipulated in Equation 2.7, offers a constructive means to analytically explore the asymptotic characteristics of

[2] In particular, the default event $\{X_i > x_i\}$ is equivalent to $\{U_i > 1 - p_i\} \ \ \forall i = 1, ..., n$ for a generic obligor i. Indeed, as $U_i = F_i(X_i)$, if the event $\{X_i > x_i\}$ occurs: $U_i = F_i(X_i) > F_i(x_i) = 1 - \bar{F}_i(x_i) = 1 - p_i$ since F_i, as a cumulative distribution function, is an increasing monotone function.

loss tail probabilities. This analysis takes into account the inclusion of the mixing variable Λ, introduced in Equation 2.4, which represents the unique exogenous source of market stress. As a mathematical modeling consequence, its survival function must belong to the family of the regular varying functions, defined in Definition 8, with index $-\frac{1}{\alpha}$. In addition, Proposition 3 provides a means to explicitly elucidate the unique heavy-tailed characteristic of the generator ϕ associated with the mixing variable, which is precisely related to the model's interdependence parameter α. With these foundational suppositions in place, and given the assumption of a monotonic probability density function for the mixing variable Λ, it is possible to initiate explicit calculations about the asymptotic properties of loss tail probabilities.

Taking into consideration the reformulation of the i^{th}-marginal conditional probability $p_i(v)$ as defined in Definition 4, it is possible to define the i^{th}*-modified marginal conditional probability* as:

$$\begin{aligned}\hat{p}_i(v) &:= \mathbb{P}\left(U_i > 1 - l_i f_n \mid \Lambda = \frac{v}{\phi\,(1-f_n)}\right) = \\ &= 1 - \mathbb{P}\left(U_i \leq 1 - l_i f_n | \Lambda = \frac{v}{\phi(1-f_n)}\right) = \\ &= 1 - p_i\left(\frac{v}{\phi(1-f_n)}\right) = 1 - e^{-v\frac{\phi(1-l_i f_n)}{\phi(1-f_n)}}\end{aligned} \tag{2.25}$$

As this approach aims to study asymptotic behaviors, we need to observe the limit when the number of obligors tends to infinity [3]. Therefore, the function resulting from the asymptotic limit of the i^{th}-modified conditional marginal probability outlined in Equation 2.25, as the number of obligors grows to infinity, can be defined as:

$$\tilde{p}_i(v) := \lim_{n\to\infty} \hat{p}_i(v) = 1 - \exp(-v l_i^{\alpha}) \tag{2.26}$$

Furthermore, by defining the function $r(v)$ as:

$$r(v) := \sum_{j\leq|W|} c_j w_j \tilde{p}_i(v) = \sum_{j\leq|W|} c_j w_j (1 - \exp(-v l_j^{\alpha})), \quad n \to \infty \tag{2.27}$$

it follows that using Kolmogorov's strong law of large numbers, the conditioned distribution of the portfolio loss converges almost surely to:

$$\frac{L_n | \Lambda = \frac{v}{\phi(1-f_n)}}{n} \to r(v), \quad n \to \infty \tag{2.28}$$

It is imperative to emphasize that the function $r(v)$ exhibits strict monotonicity concerning the variable v and possesses an upper bound denoted as $\bar{c} = \sum_{j\leq|W|} c_j w_j$. This upper bound can be interpreted as the ultimate average loss incurred when all obligors default. For analytical purposes, denote as v^* the unique solution in $(0, \bar{c})$ to $r(v) = b$: it represents the threshold value for the random variable Λ so that the average portfolio loss is less than b. At this point, the asymptotic behavior of the loss tail probability can be analytically explained.

[3]Observe that: $\lim_{n\to\infty} \frac{\phi(1-l_i f_n)}{\phi(1-f_n)} \sim \frac{(l_i f_n)^{-\alpha}}{f_n^{-\alpha}} = l_i^{-\alpha}$

Proposition 4. *Let L_n be the loss function of a portfolio made up of n obligors as defined in Definition 2.24 and let f_n be its quality function. Further assume that $\exp(-n\beta) = o(f_n)$, $\beta > 0$ and that the interdependence parameter α is greater than 1. It follows that, for a fixed value of the threshold parameter b, the probability of a portfolio large loss as the number of obligors approaches infinity results in:*

$$\mathbb{P}(L_n > nb) \sim f_n \frac{(v^*)^{-\frac{1}{\alpha}}}{\Gamma(1-\frac{1}{\alpha})}, \quad n \to \infty \tag{2.29}$$

where v^* is the unique solution of the equation $r(v) = b$ and $\Gamma(\cdot)$ is the Gamma function.

The exact proof of this result is formally explained in the main article [3].

Some specifications about the result of Proposition 4 have to be made:

- The asymptotic probability is mostly dictated by the parameters α and f_n. The larger the value of α, the more obligors tend to default simultaneously; at the same time, the probability of large losses is directly proportional to the sequence f_n, symbolizing the fact that with a higher credit quality of the pool, the probability of large losses diminishes. In particular, by choosing different values for f_n, portfolios will have different credit ratings. For example, a portfolio equipped with $f_n = \frac{1}{n}$ consists of high-quality obligors, while a portfolio equipped with $f_n = \frac{1}{\ln^{\frac{1}{2}}(n)}$ has more risky ones.
- Another hidden role is played by the constants l_i, implicitly present in the specification of v^*. They reflect the single quality of each obligor.

Besides that, for a correct evaluation and possible investment comparisons, a more comprehensive measure of risk has to be explicit, possibly in a closed analytical form. As already said at the beginning of Section 2.4, to have a correct measure of the risk associated with a precise portfolio or pool of assets it is useful to explicit the Conditional Value at Risk. From the reformulation of the CVaR presented in Equation 7, it is evident a possible explicit reformulation within the closed-form probability sharp asymptotic presented in Equation 2.29:

$$CVaR_{\mathbb{P}(L_n > nb)}(L_n) = \mathbb{E}\left[L_n \mid L_n > nb\right] \sim nb + n\frac{\int_{v^*}^{\infty} r'(v) v^{-1/\alpha} \mathrm{d}v}{(v^*)^{-1/\alpha}}, \quad n \to \infty. \tag{2.30}$$

where $r'(v)$ is the first derivative of the function $r(v)$, defined in Equation 2.27, with respect to the variable v and v^* is the unique solution of the equation $r(v) = b$. The explicit result in Equation 2.30 shows that asymptotically CVaR is linearly proportional to the number of obligors in the pool.

From the perspective of defining a rare-event simulation algorithm based on the asymptotic analytical findings, the subsequent section will highlight the outcomes and comparisons of the simulations and their numerical counterparts.

2.7 Simulation and numerical results

The correctness of the asymptotic formulas obtained in Section 2.6 inspires the necessity of creating a rare-event simulation algorithm. For this purpose, we have developed the code for an asymptotic simulation method for the estimation of the loss tail probability and the CVaR, which is reported in Appendix A Section A.2. Furthermore, we compared its results with the numerical ones, obtained by the formulas in Section 2.6.

Since the context of this thesis involves the probability estimation of rare events, naive simulation approaches, like the Monte Carlo one, are notoriously inefficient and can provide ambiguous results. To solve this problem, variance reduction rare event simulations are involved, in particular within the *Conditional Monte Carlo* algorithm. As demonstrated in the article [3], this estimator has a bounded relative error as well as it reflects algorithmic efficiency and variance reduction.

In addition to the general simulation, the most important parameters of the approach have been varied to exhibit its robustness. By varying the parameter α it is possible to examine the variation of the loss tail probability in the case of different interdependence intensities: extreme independence is linked to the value 1 as well as its growth to infinity symbolizes the increase in the interdependence strength. At the same time, by varying the parameter b is possible to analyze the changes in the probabilities concerning the relative threshold of failure. Furthermore, by varying the function f_n it is possible to modify the quality of the considered pool of assets, with a specific consideration of its number of obligors n. These simulation-related variation studies have been performed and commented on in the next Section.

2.7.1 Loss tail probability estimation

To find the probability of large losses arising from rare events, two ways are generally used: numerically, such as following analytical calculations, or via simulation. About the numerical results, the estimation of the loss tail probability has been derived by substituting the specific values of v^*, α, n, and b into Equation 2.30. Additionally, the computation of the integral contained in the denominator of the equation's second term was performed numerically. This integral computation was executed through the utilization of the *integral* command within MATLAB, employing global adaptive quadrature and default error tolerances. Conversely, regarding the simulation results, this sub-section proposes a variance reduction algorithm to simulate efficiently and coherently the loss tail probability, which is particularly necessary to compute the CVaR. Furthermore, the results that arise from the simulation process are compared to the ones obtained numerically.

As can be seen in the following pseudocode, the probability simulation algorithm is articulated into four steps that follow the previously explained approach.

Algorithm 1 Conditional Monte Carlo Algorithm (CMC)

1: Generate n independent standard exponential random variables $R_i,\ i = 1, ..., n$.
2: For each obligor, transform R_i to $O_i = \frac{R_i}{\phi(1-l_i f_n)},\ i = 1, ..., n$.
3: Order the obtained values of O, then find $O_{(k)}$, such as the value of O_i which index satisfies $k = \min\{l : \sum_{i=1}^{l} c_{(i)} > nb\}$.
4: Return the CMC estimator:

$$S(\mathbf{R}) := \mathbb{P}(L_n > nb|\mathbf{R}) = \mathbb{P}(\Lambda > O_{(k)}|\mathbf{R}) = \frac{\sum_{j=1}^{nrep} \mathbb{1}_{\{|\Lambda_j > O_{(k)}|\mathbf{R}\}}}{nrep}$$

where $nrep$ is the number of repetitions of the simulation process.

Empirical results and comparisons are shown in Tables 2.2, 2.3 and 2.4, where the analysis is conducted by changing respectively the values of α, n and b. The credit quality function is set as $f_n = \frac{1}{n}$ and the analysis is conducted following an homogeneous weighting ($c_i = c = 1$, $l_i = 0.5\ \forall i = 1, ..., n$). Furthermore, for reproducibility purposes, the seed of the RNG used is set to 300829.

$\mathbb{P}(L_n > nb)$	**CMC**	**Numerical results**
$\alpha = 1.1$	$6.3000 \cdot 10^{-5}$	$6.1756 \cdot 10^{-5}$
$\alpha = 2$	$4.8500 \cdot 10^{-4}$	$4.4472 \cdot 10^{-4}$
$\alpha = 5$	$7.8900 \cdot 10^{-4}$	$7.8096 \cdot 10^{-4}$

Table 2.2: Table of the differences in the probability estimation between simulation and numerical results, after fixing the number of obligors to $n = 500$ and $b = 0.8$.

$\mathbb{P}(L_n > nb)$	**CMC**	**Numerical results**
$n = 100$	$2.2530 \cdot 10^{-3}$	$2.2236 \cdot 10^{-3}$
$n = 500$	$4.8500 \cdot 10^{-4}$	$4.4472 \cdot 10^{-4}$
$n = 1000$	$2.3100 \cdot 10^{-4}$	$2.2236 \cdot 10^{-4}$

Table 2.3: Table of the differences in the probability estimation between simulation's and numerical results, after fixing $\alpha = 2$ and $b = 0.8$.

$\mathbb{P}(L_n > nb)$	**CMC**	**Numerical results**
$b = 0.3$	$1.0160 \cdot 10^{-3}$	$9.4469 \cdot 10^{-4}$
$b = 0.5$	$7.3200 \cdot 10^{-4}$	$6.7766 \cdot 10^{-4}$
$b = 0.8$	$4.8500 \cdot 10^{-4}$	$4.4472 \cdot 10^{-4}$

Table 2.4: Table of the differences in the probability estimation between simulation's and numerical results, after fixing the number of obligors to $n = 500$ and $\alpha = 2$.

In addition to the proximity of the two measurements, a concise examination of the outcomes reveals a discernible trend: as the quantity of obligors rises, there is a notable reduction in the likelihood of encountering significant losses. That explicitly manifests the diversification effect of the portfolio. Concurrently, by increasing the dependence parameter α, the probability increases too. Furthermore, an observable effect emerges when increasing the loss-threshold parameter b by the underlying logic, leading to a corresponding decrease in the probability.

$\mathbb{P}(L_n > nb)$	**CMC**	**Numerical results**
$f_n = \frac{1}{\ln(n)}$	$9.9304 \cdot 10^{-2}$	$8.9197 \cdot 10^{-2}$
$f_n = \frac{1}{\sqrt{n}}$	$1.0107 \cdot 10^{-2}$	$9.9443 \cdot 10^{-3}$
$f_n = \frac{1}{n}$	$4.8500 \cdot 10^{-4}$	$4.4472 \cdot 10^{-4}$

Table 2.5: Table of the differences in the probability estimation between simulation's and numerical results, after fixing the number of obligors to $n = 500$, $\alpha = 2$ and $b = 0.8$.

Another noteworthy observation pertains to the examination of different credit quality functions, the simulation outcomes of which are juxtaposed with their numerical counterparts in Table 2.5. Beginning with the first entry in the table, it is conceivable that these "features" exhibit varying probability outcomes, with the initial one aligning more closely with the characteristics of a *deteriorated pool*, due to its higher probability results. All the MATLAB codes for the homogeneous simulation and the general one can be found in Appendix A, Sections A.1, A.2 and A.4.

2.7.2 Simulation of the Expected Shortfall and results comparison

Further considerations can be extended to the estimation of the *Expected Shortfall*, utilizing both the asymptotic and simulation approaches. Specifically, in this instance, the Conditional Monte Carlo algorithm is employed to assess not only the probability of encountering significant losses but also the expected value of losses surpassing the threshold denoted as nb. In this context, the supplementary step to be incorporated into the CMC algorithm for estimating the probability of significant losses involves computing the difference between the estimated loss in a given scenario and the threshold nb, specifically only in cases where the scenario's loss exceeds said threshold. Moreover, to compute the integral present in Equation 2.30, where obtaining a closed-form solution is not feasible due to its complexity, a numerical technique is employed. The resolution involves utilizing the MATLAB command *integral* to approximate the integral's value within a specified tolerance level. All the simulations are carried out with $\alpha = 2$, $b = 0.8$, $f_n = \frac{1}{n}$ and the usual homogeneous weighting.

CVaR	**CMC**	**Numerical results**	**Relative discrepancy**
$n = 100$	96.1820	96.3684	$1.9378 \cdot 10^{-3}$
$n = 500$	483.6165	481.8418	$3.6696 \cdot 10^{-3}$
$n = 1000$	964.4459	963.6836	$7.9034 \cdot 10^{-4}$
$n = 5000$	4815.8611	4818.4182	$5.3098 \cdot 10^{-4}$

Table 2.6: Table of the differences in the CVaR estimation between simulation's and numerical results, after fixing $\alpha = 2$ and $b = 0.8$. In the last column, it is possible to observe a relative discrepancy between the two results.

Table 2.6 illustrates the concordance between the two estimates, with the relative discrepancy, defined as:

$$rel\_disc = \frac{|CMC - Numerical\ \ results|}{CMC} \tag{2.31}$$

diminishing as the portfolio's obligor count increases, despite in the first line but probably as a consequence of the seed of the RNG used. This observation underlines that increasing the number of obligors enhances the accuracy of the analytical estimate, highlighting a convergence trend. All the MATLAB codes for the homogeneous simulation and the general one can be found in Appendix A, Sections A.3, and A.4. To be clear, another approach that can be followed in this case is the *Importance Sampling* one, an important variance reduction technique that ensures more precision in the valuation of expected values in the case of extreme events simulation.

In a real-world context, the objective is to identify certain attributes of a portfolio and integrate them into the previously elucidated methodology. The specific parameters of interest encompass the interdependence index α and the values l_i for each obligor in the pool. These parameters can be calibrated using the current pool's data, historical behaviors, correlation with some other financial instruments or deduced through credit ratings issued by Credit Rating agencies. Consequently, to have a rigorous evaluation, the methodology will be implemented in the subsequent chapter using a *loan data tape* generously provided by *Intesa Sanpaolo S.p.A.*

Chapter 3

Application to a case of securitised ABS

To establish an efficacious testing protocol for the mathematical methodology previously delineated, this chapter undertakes an exhaustive examination utilizing a loan data tape provided by *Intesa Sanpaolo S.p.A.* More precisely, the primary objective of this chapter is to deliver a comprehensive dataset analysis, with a subsequent focus on the precise computation, estimation, and selection of critical model parameters. The outcomes of this analysis will be compared to the current market practice benchmarks, which will be consecutively briefly explained. The loan data tape is updated to the first day of September 2023. It is important to notice that each feature concerning the monetary value of a loan is expressed in euros (€), while the ones regarding coupons or rates are expressed in percentages and not as decimal numbers.

3.1 Loan data tape description

To well present the totality of the data and to characterize the utility of each feature, some dataset statistics are necessary. The original loan data tape taken into analysis is a pool made up of 4139 different loans, each of them identified by an *ID number* for privacy reasons and well presented in the first column. Besides that, the dataset involves other 19 features, each of them providing a useful characteristic of any loan presented. Since the beginning, it is important to notice that the information given by each feature differs a lot, therefore in the next steps, there will be preprocessing procedures to analyze coherently these data. As a consequence, it is coherent to list and explain all these features.

1. **Amount Loan**: is the total amount of the loan given by the bank to an obligor. This feature involves the following data distribution and statistics:

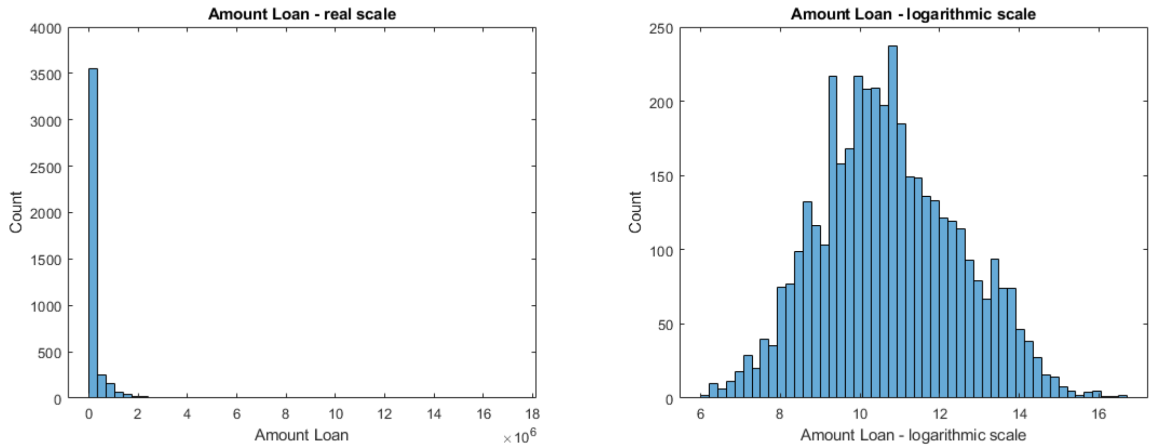


Figure 3.1: Data distribution of the Amount Loan feature. On the left, data are presented in a normal scale, while on the right data are presented in a logarithmic scale for better visualization and comprehension.

Minimum	**Mean μ**	**Maximum**	**Mode**	**Standard deviation σ**	**Coefficient of variation $\frac{\sigma}{\mu}$**
478.00	222724.76	17268366.80	5000.00	700591.14	3.15

Table 3.1: Statistics of the feature Amount Loan.

As it is inferable from the previous figures and tables, the amounts present in this feature of the data tape are very heterogeneous: starting from a minimum loan valued 478€, the loan with the maximum money request presents a value of nearly 17 million €. Aside from that, the most frequent amount is 5000 €. In conclusion, to fully describe this attribute it is important to notice that, concerning its measure scale, the standard deviation is a relatively high value.

2. **Current Amount Loans**: represents the quantitative of money that each borrower has to pay updated since the current date (September 1, 2023). Specifically, it represents the exposition parameter c_i for each obligor i, as previously presented in Equation 2.23. Concurrently, a clarification has to be made: since some types of loans might have a *recovery rate* R after default, in an ABS context this exposition parameter can be scaled by a factor of $1 - R$ to better quantify the real exposure that an investor is facing. This feature involves the following data distribution and statistics:

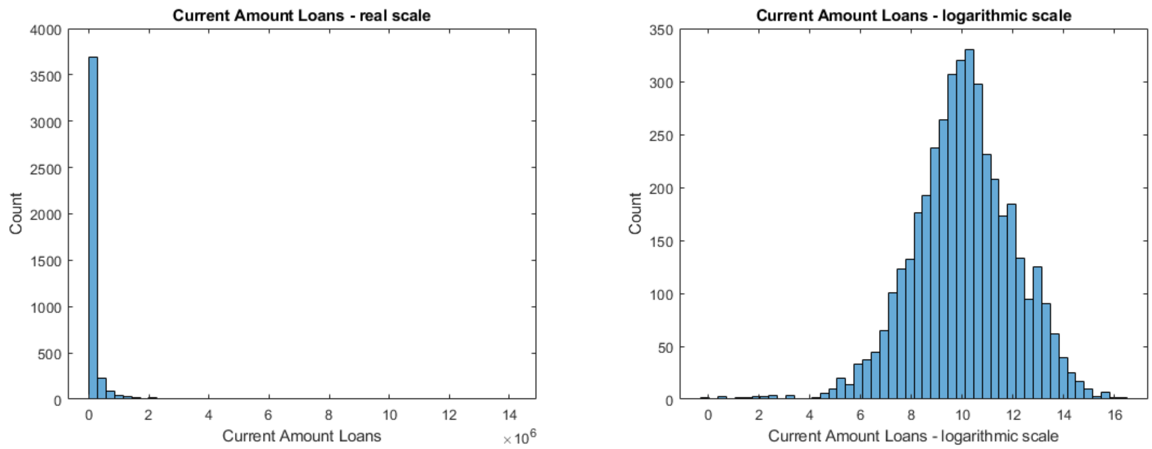


Figure 3.2: Data distribution of the Current Amount Loans feature. On the left, data are presented in a normal scale, while on the right data are presented in a logarithmic scale for better visualization and comprehension.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
0.80	144195.58	14199218.66	3175.56	543329.88	3.77

Table 3.2: Statistics of the feature Current Amount Loans.

As discernible from the preceding figures and tables, in this case, the feature exhibits pronounced heterogeneity. Ranging from a minimum loan amount of 0.80 €to a peak value nearing 14 million €, this feature spans a wide spectrum of values. In summation, a comprehensive characterization of this attribute warrants acknowledgment of the relatively elevated standard deviation concerning its scale of measurement.

3. **Gross Coupon**: this feature refers to the gross coupon given to each borrower concerning the original amount of the requested loan. This feature involves the following data distribution and statistics:

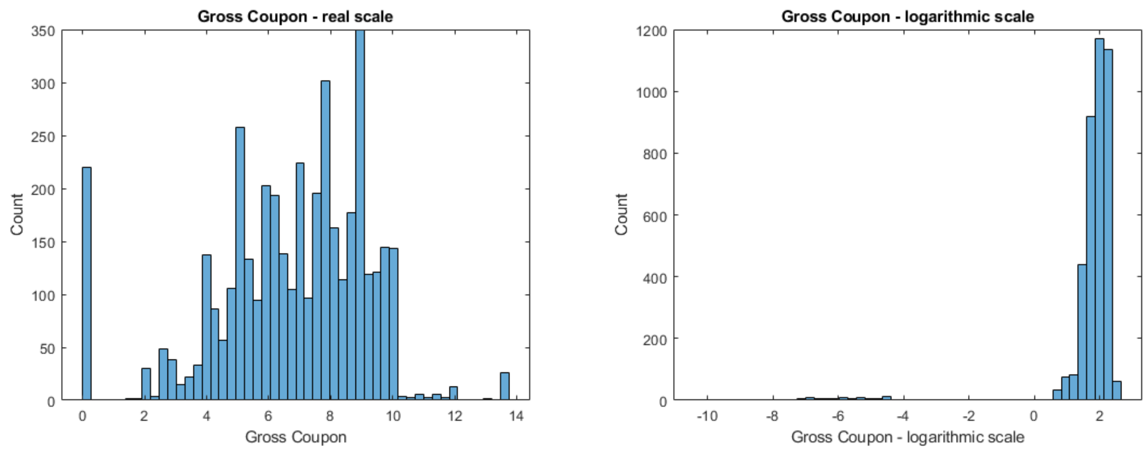


Figure 3.3: Data distribution of the Gross Coupon feature. On the left, data are presented in a normal scale, while on the right data are presented in a logarithmic scale for better visualization and comprehension.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
0	6.70	13.75	0	2.54	0.38

Table 3.3: Statistics of the feature Gross Coupon.

As it is noticeable from the table above, this feature ranges in the interval $[0, 13.75]$, concentrating the main frequency on the value 0: this represents that many loans in this pool do not have any interest rate fixed on them.

4. **Rate type**: the type of interest rate applied to each loan. In this data tape, each loan is equipped with a *fixed rate*.

5. **Original Maturity Term**: as each loan is contractually defined as an end date for the payments, this feature represents the maturity term of the loans in the pool. All the data represented in the following figure and table are expressed in *months*. This feature involves the following data distribution and statistics:

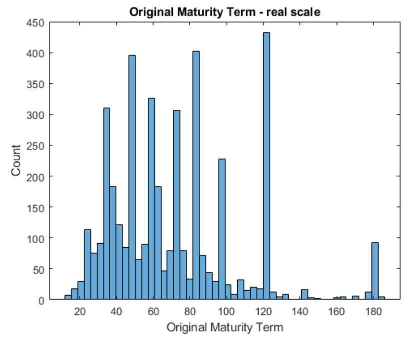


Figure 3.4: Data distribution of the Original Maturity Term feature.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
12	71.56	186	120	34.55	0.48

Table 3.4: Statistics of the feature Original Maturity Term.

With specific regard to the table above, it is noticeable that every loan has a duration higher than one year, with a maximum of fifteen years and a half. However, the most frequent duration is ten years, quite distant from the average duration presented in the second column.

6. **Remaining Term**: concerning the month of September 2023, this attribute reflects the remaining number of months until the date of maturity of each loan. This feature involves the following data distribution and statistics:

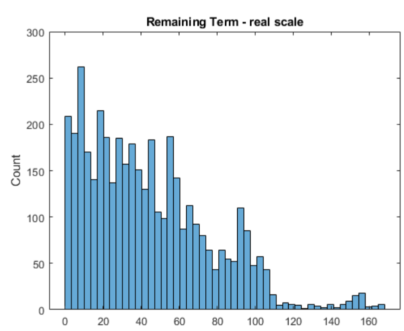


Figure 3.5: Data distribution of the Remaining Term feature.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
0	44	168	11	32.77	0.74

Table 3.5: Statistics of the feature Remaining Term.

As it is inferable from the table above, some loans have already reached the maturity term, while others will still be active, unless the default of the corresponding obligor, for 14 years.

7. **IsBaloon**: represents if the payment of the loan will be done in its integrity at maturity or if it is amortized during the months of "life" of the loan. In this specific case, every loan is amortized over months.

8. **Age**: represents the months passed since the loan has been emitted. It is a measure of the longevity of the loan. This feature involves the following data distribution and statistics:

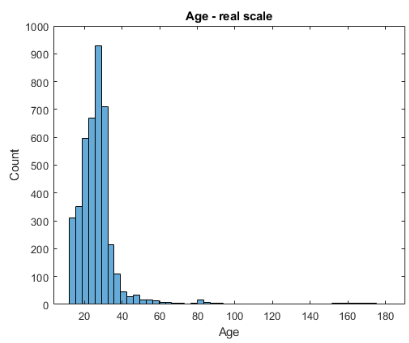


Figure 3.6: Data distribution of the Age feature.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
12	27.60	182	26	15.93	0.58

Table 3.6: Statistics of the feature Age.

Following the statistics of the data presented above, each loan presented has been issued at least one year ago, while some loan has a "life" longer than 15 years.

9. **Issue Date**: this feature refers to the date on which the loan has been emitted. It is of particular importance because often part of the interest rate associated with a loan is a consequence of the market conditions at that precise time or at the same time could depend on some other time-linked financial rates. This feature involves the following data distribution and statistics:

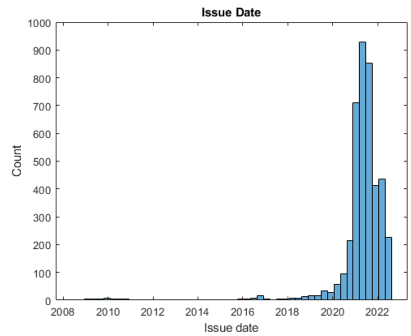


Figure 3.7: Data distribution of the Issue Date feature. On the left, data are presented in a normal scale, while on the right data are presented in a logarithmic scale for better visualization and comprehension.

First date	**Mean date**	**Last date**	**Most frequent date**
06/01/2008	04/13/2021	08/01/2022	06/01/2021

Table 3.7: Statistics of the feature Issue Date. The format of the date is MM/DD/YYYY.

As shown by the previous table some loans have been issued many years ago, with the first one dating back to June 2008. Despite that, the majority of these loans have been emitted in the last two years, and at the same time, 97% of all of them have been emitted after October 2018.

10. **Next Paydate**: represents the date on which the payment of each loan will be carried out. In this specific case, each next loan payment will be carried out on October 1, 2023.

11. **Frequency**: the frequency of payments of each loan. In this case, every loan has a monthly payment frequency.

12. **Delinquency**: this attribute reflects the presence of some lack of payments that happened during the life of a loan. In particular, it involves four different categories: *Current*, *30 Days*, *60 Days* or *90+ Days*.

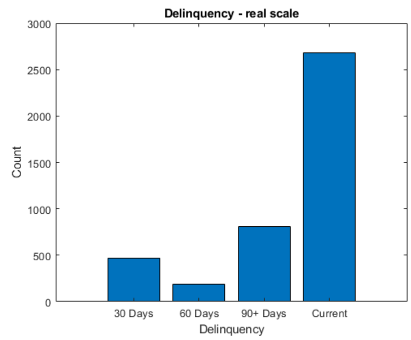


Figure 3.8: Data distribution of the Delinquency feature.

The majority of the loans in the pool (2677 out of 4139, such as almost 65% of the total) have no lack of payments as described in the barplot above. This represents the fact that the total pool has a good quality.

13. **Months Delinquency**: this attribute reflects the number of days passed from the last delinquency for each loan with the attribute *Delinquency* different from the category *Current*.

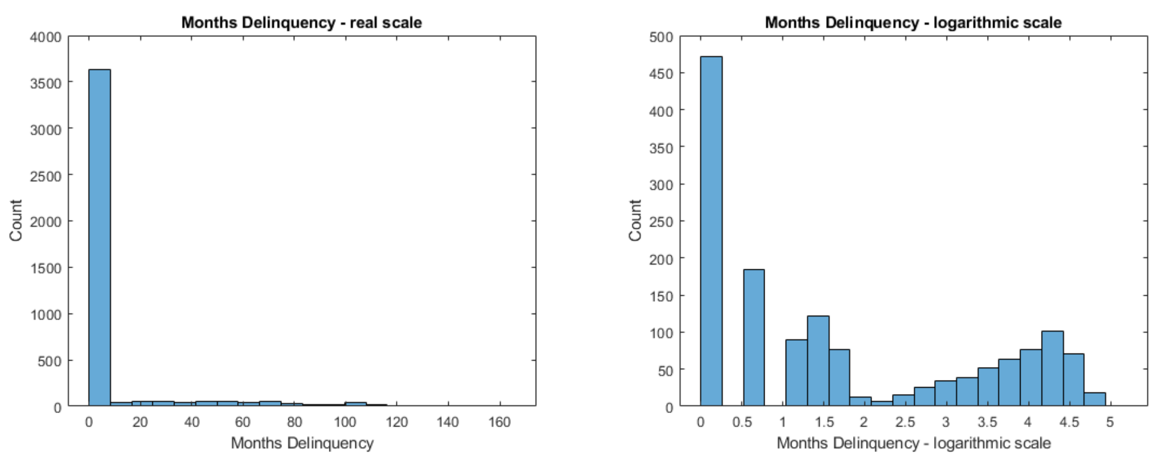


Figure 3.9: Data distribution of the Months Delinquency feature. On the left, data are presented in a normal scale, while on the right data are presented in a logarithmic scale for better visualization and comprehension.

As it is inferable from above, the majority of the loans have no delay in the payments, while some of them have a delay in the order of 165.

Minimum	Mean μ	Maximum	Mode	Standard deviation σ	Coefficient of variation $\frac{\sigma}{\mu}$
0	7.19	165	0	20.53	2.86

Table 3.8: Statistics of the feature Months Delinquency.

14. **Maturity date**: represents the maturity date of each loan, both for the already and not finished ones.

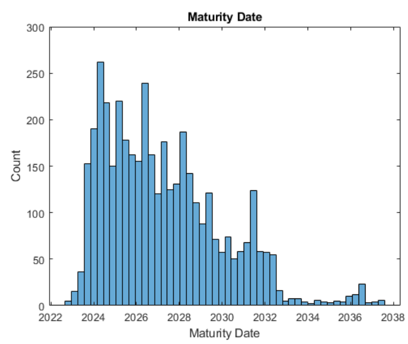


Figure 3.10: Data distribution of the Maturity Date feature.

First date	Mean date	Last date	Most frequent date
09/01/2022	03/31/2027	08/10/2037	07/01/2024

Table 3.9: Statistics of the feature Maturity Date. The format of the date is MM/DD/YYYY.

The table above outlines the fact that some loan is already over, while others, in the absence of obligors defaults, will persist until 2037.

15. **Amortization Schedule**: the payment scheduling of each loan after the current date.

16. **Direct Deal Group**: represents the type of loan requested by each obligor. It involves five different categories: *Finalized*, *Furniture*, *New Vehicles*, *Personal Loans* and *Used Vehicles*.

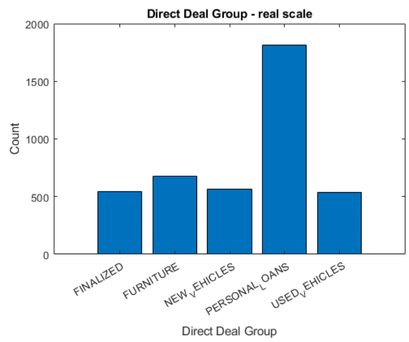


Figure 3.11: Data distribution of the Direct Deal Group feature.

As presented in the barplot above, almost half of the loans (1815) are referred to as personal loans, while the other four types are almost uniformly distributed

17. **Group Collateral Type**: defines the collateral type associated to each loan. In this specific case, each loan is associated with the class *Personal/Auto Loans*.

18. **Defaulted**: this feature represents whether a loan is defaulted or not during its ”life”, such as if the loan has been interrupted by the bank.

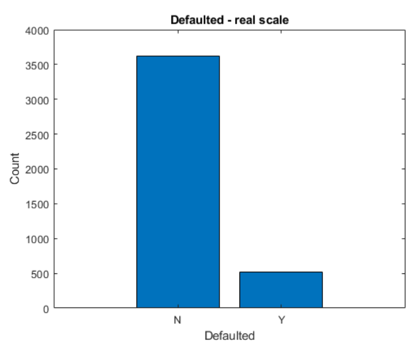


Figure 3.12: Data distribution of the Defaulted feature.

The barplot above represents that the quality of the pool is high, with only 514 loans that have defaulted (such as almost 12%) during their life.

19. **Related Deal Type**: defines if the loan is for consumers or commercial purposes. In this specific case, each loan is associated with the class *Consumer Loan*.

As succinctly delineated in the preceding list, the attributes encompass a diverse array of data types, including numerical, categorical, and datetime formats. To enable effective data analysis, the non-numerical variables have been subjected to conversion procedures, as will be better explained in the following sections. The datetime format has been harmonized using a *date-ordering* approach, while the categorical format has undergone *one-hot encoding*. Furthermore, to ensure the reliability of subsequent analyses, non-informative variables have been systematically pruned from the dataset. In summation, following this comprehensive pre-processing phase, the dataset now comprises 20 distinct variables, which will be harnessed for subsequent statistical assessments and parameter estimation.

3.2 Parameters computation

To derive meaningful insights from the dataset within the context of the methodology expounded in Chapter 2, it is imperative to engage in data manipulation. This necessary process focuses on the computation or estimation of two critical variables: the parameter denoted as α and the vector represented by l. In particular, it is important to remember the loss formulation made explicit in Equation 2.24, for which the predefined underlying LT-Archimedean copula structure is outlined in Equation 2.7. In this formulation, the parameter α has multiple roles: it reflects the overall interdependence among obligors in the pool as well as it outlines the heavy-tailedness property of the generator of the mixing variable Λ. Moreover, each element of the vector l reflects the quality of the related loan. The subsequent part of this section is dedicated to elucidating the intricacies of the parameters computation or estimation procedures for these aforementioned variables, alongside a comprehensive exposition of the results stemming from the original dataset. This comprehensive treatment seeks to provide a deeper understanding of the interplay between data manipulation and the analysis of the vector l and the parameter α.

The computation of the vector l is necessary to give each obligor in the pool a sort of *credit measure* that manifests its singular quality, based on both the provided database and another financial instrument. Particularly, this subsection aims to estimate each obligor's quality based on the level of interest rate given to each loan concerning another interest rate measure, such as the corresponding *Swap interest rate*. After that, these values will be spanned into a proper range to overcome the necessities of this thesis' credit risk model. To well define a quality measure of each loan in the pool, is necessary to evaluate the attributed riskiness given, in this case, by the bank. To this extent, is therefore necessary to introduce an important financial interest rate instrument: the *Swap interest rate*. The Swap interest rate, often referred to simply as the *Swap rate*, is a fundamental financial benchmark used in the realm of interest rate derivatives and fixed-income markets. It represents the fixed interest rate that two counterparties agree to exchange periodically over a specified period. Swap rates are instrumental in managing interest rate risk, as they enable entities to hedge against fluctuations in interest rates. As such, the swap interest rate is a cornerstone in the world of finance, serving as a valuable tool for risk management, credit risk assessment, and financial product creation. Furthermore, these interest rates are always related to maturity, and as any other interest rate instrument are therefore related to their issuing date. For instance, figure 3.13 shows the past evolution over time of the 2-year swap interest rate compared to the 5-year and the 15-year ones.

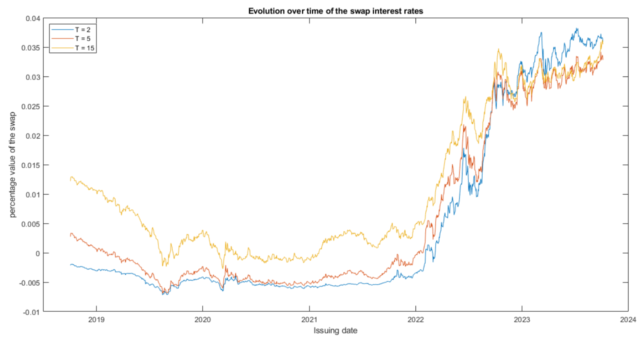


Figure 3.13: Evolution over time of the Swap interest rates values. In blue the maturity is 2 years, in red the maturity is 5 years and in yellow is 15 years.

As a particular specification, is necessary to note that these interest rates can go under the zero threshold, symbolizing low general riskiness in that period. In more mathematical details, the swap rate is the fixed interest rate that one party agrees to pay in a plain vanilla interest rate swap: it is essentially the fixed rate that makes the two sets of cash flows, fixed and floating, have the same present value. It can be mathematically formulated as:

$$S = \frac{\sum_{i=1}^{n} \text{ Floating Payment }_{i} \cdot \zeta\left(t_i, r\right)}{\sum_{i=1}^{n} \text{ Year Fraction }_{i} \cdot \zeta\left(t_i, r\right)} \tag{3.1}$$

where n is the total number of payments [1], *Fixed Payment* represents cash-flow associated to the the fixed rate exchanged at each payment date, *Year Fraction* represents the year fraction payment associated to the floating rate exchanged at each payment date, and $\zeta\left(t_i, r\right)$ is the discounting factor used to discount cash flows to their present value at each payment date t_i with respect to the benchmark rate r. The relevance of this financial instrument in the computation of the vector l stems from the prevalent practice of pricing loans in tandem with such financial instruments. The interest rate values associated with these instruments are established as pivotal benchmarks to gauge an initial risk threshold. Consequently, in the context of this analysis, they are employed as foundational components for delineating the credit quality associated with the loans within each asset pool.

To well analyze and associate these financial instruments we took into consideration the Swap interest rates issued in the last five years, with a maturity spanning from one year to fifteen. With this explanatory foundation in place, the subsequent aspect that warrants attention pertains to the *Gross Coupon* feature, which delineates the percentage of payments allocated to interest. As briefly explained in the features list present in the previous section, this attribute assumes percentage values in the range $[0, 13.75]$, outlining a great heterogeneity among different loans. As a consequence, to conduct a reliable analysis, all the loans with the feature Gross Coupon value equal to zero have been removed from the dataset[2]. This decision stems from the fact that generally consumer loans are granted with interest rates relatively higher than zero.

[1] Generally, payments are homogeneously spanned through the year and mainly on a monthly basis. In this last case it results $t_i = \frac{i}{12}$.

[2] These loans were approximately 130, so their elimination doesn't vary the overall analysis.

Following these last specifications, the next part of the analysis is related to the calculation of the credit spread of each loan, which is defined for each loan as:

$$spread = GC - Swap(ID,\ OMT) \tag{3.2}$$

where GC is the percentage value of the gross coupon, ID is the issuing date, OMT is the original maturity term and $Swap$ is the swap rate corresponding to the relative issuing date and maturity term. The calculation of this value yields a spectrum of diverse outcomes, which, in the context of this analysis, are linked to an indicator of the inherent quality of each loan. Nonetheless, the following points need to be addressed.

- It is essential to recognize that the value of each swap rate undergoes daily fluctuations, resulting in recurring oscillations in its valuation.

- The maturity terms of individual loans do not invariably align with whole numbers of years; quite often, they manifest as fractions thereof. Consequently, achieving an exact match between the maturity term of each loan and the corresponding Swap rate can be challenging. To overcome this challenge, the approach adopted involves a criterion where each loan is associated with the Swap rate having an annual maturity term higher than one year if the loan's maturity term, expressed in annual terms, exceeds the threshold of 0.5 years between two consecutive years. Conversely, if the loan's annualized maturity term falls below this threshold, it is associated with the swap rate corresponding to one year less than its annual maturity.

- Given that the entire set of swap rates under consideration originates from a five-year time frame, certain loans issued before this period lack a coherent association with swap rates. To address this challenge, our approach entails allocating the most recent available Swap rate to such loans in cases where they exhibit no delinquency or defaults. Conversely, for loans that are in default, the spread assigned to them is the highest among all cases. For loans with no defaults but some days of delinquency, their associated spread is calculated as $\frac{5}{6}$ of the maximum spread. This choice is grounded in the recognition that older issues may carry significantly greater risk than newer ones, necessitating a consistent approach to associating them with potential failures.

- Given that the dataset may contain issuing dates that do not coincide with business days, it becomes impractical to identify an exact match for assigning swap rates to these loans. Consequently, for such cases, the approach involves associating the nearest available rate, considering the proximity of the issuing date.

After all these specifications, a coherent spread measure for all the loans can be computed, but a further post-processing operation has to be made. This consequence arises from the presence of some negative spread values, which are in contrast to the general meaning of this instrument. Specifically, the creation of this instrument must play a role in assessing *positive* risks and not *negative* ones, which financially arguing will be suffered by the issuer. As a consequence, the post-processing action needed is the removal of these loans from the total pool. This procedure makes the dataset remain with 3972 different loans, which Gross coupon, spread, and associated Swap rate values can be seen in figure 3.14.

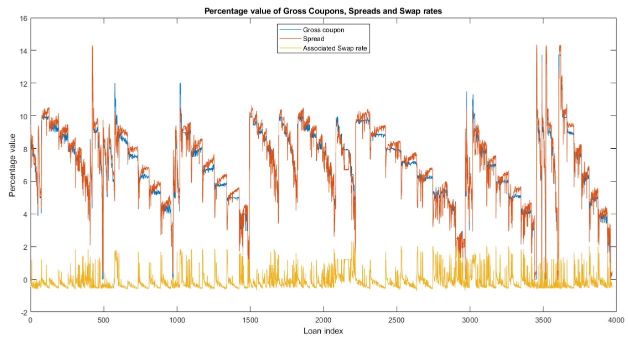


Figure 3.14: Percentages values within the spread calculation. In blue is the Gross Coupon, in yellow is the Swap rate, and in blue is the Spread.

As it is noticeable from Figure 3.14, the associated Swap rates are quite lower than the Gross coupon values, with the further characteristic that they sometimes go under the zero threshold. All this procedure is well defined in the corresponding code present in Appendix A, Section A.6, and is briefly explained by the following pseudocode.

Algorithm 2 Spreads-computing algorithm

```
Compute the Years to Maturity of each loan i.
Compute the Issuing date of each loan i.
Search for the related Swap rates for each loan i.
if The Issuing date is not present then:
    Assign the nearest Swap rate concerning the Issuing date.
else
    Assign to each loan the relative Swap rate.
end if
Calculate the spread for each loan as in Equation 3.2.
if The Issuing date is older than the first one present in the database then:
    if The loan is defaulted then:
        Assign the maximum spread.
    else if The loan is not defaulted, but there are delinquency days then:
        Assign 5/6 of the maximum spread.
    else
        Assign the nearest spread concerning the issuing date.
    end if
end if
```

Consequently, for the quality risk measuring part, there is a necessity to link the singular values of the spread to the ones of the vector l. The first specification to be made at this point stems from an empirical consideration: to obtain a coherent result for the loss tail probabilities and the CVaR, the vector l has to be well calibrated. Specifically, with this large quantity of loans, it has been empirically shown that each value must not get lower than the value of 0.79. This observation follows the fact that when calculating the value of v* in the original model, there is the necessity to find the zero of a function numerically. Therefore, each element of the vector l

represents an important factor in this evaluation. When the values considered for each element of l are too low or too high, the zero-computing function, most of the time, leads to numerical errors or early stopping of the calculation. As a consequence, we empirically observed that if each element is under the value 0.79 the algorithm will not reach convergence, while if each element exceeds the unitary threshold, it may achieve convergence, but results will not be coherent within the application. Starting from the previous considerations, to obtain robust observations, it results that each value must be spanned into the range $[0.79, 1.00)$. For this purpose, in the l-computing function, we have formulated many different methods to "translate" the relevance of the already calculated spreads into a precise value present in the pre-specified range. To achieve coherence concerning the model specified in Chapter 2 Section2.6, especially concerning Equation 2.24, it is necessary to make evident that high values for l_i reflect low-quality loans, while low values reflect good quality ones. This necessity justifies some choices made in the following list of methods.

1. **Uniform**: this method associates to each element of the vector l the value 0.80.

2. **Random**: this method associates to each element of the vector l a random value between 0.80 and 1.00 (excluded).

3. **Classical Swaps**: this method represents a pioneering approach, establishing a direct correspondence between the inherent significance of each spread value and a separate value within the interval $[0.80, 1.00)$. Upon the computation of the spread for each obligor, this technique assigns to each element a value derived from the application of one of the following functions: the *Min-Max function*, the *Sigmoid function* or the *Rectified Linear Unit (ReLU) function*. These functions are meticulously crafted to systematically transform each spread value into the predetermined target range. In particular:

   - Min-Max function:

$$MM(spread_i) := 0.8 + 0.19 \cdot \frac{spread_i - \max(spread)}{\max(spread) - \min(spread)} \tag{3.3}$$

   - Sigmoid function:

$$S(spread_i) := 0.8 + 0.19 \cdot \frac{1}{1 + \exp(-spread_i)} \tag{3.4}$$

   - ReLu function:

$$R(spread_i) := 0.8 + 0.19 \cdot \frac{\max(spread_i - 1, 0)}{\max(spread)} \tag{3.5}$$

4. **Swaps plus defaults**: this method integrates and refines the concepts of the preceding approach. Primarily, it designates the maximum value, such as 0.99, to each previously defaulted loan, while introducing a distinct methodology for non-defaulted loans that takes into account the duration of delinquencies. For precise elucidation, it employs a linear combination of the *Arctangent function* applied to the count of days in delinquency and the *Sigmoid function* applied to the spreads. This combination is meticulously configured to map values into the aforementioned range, and its formulation is as follows:

$$SPD(spread_i, daysDEL_i) := 0.79 + \frac{1}{5\pi}\arctan(daysDEL_i) + \frac{0.1}{1+\exp(-spread_i)} \tag{3.6}$$

where *daysDEL* is the number of days of delinquency for each obligor. As it can be noted, this function spans loans with low spread and no days of delinquency into small values, while loans with low quality are spanned into high values.

Thanks to these specifications, it is conclusively possible to compute the vector l in different ways, whose results are expressed and compared in Table 3.10[3].

Method	**Minimum**	**Maximum**	**Mean μ**	**Standard deviation σ**
Uniform	0.80	0.80	0.80	0.00
Random	0.80	0.99	0.82	0.05
Classical Swaps (Min-Max)	0.80	0.99	0.89	0.03
Classical Swaps (Sigmoid)	0.89	0.99	0.98	0.01
Classical Swaps (ReLU)	0.80	0.97	0.88	0.03
Swaps plus defaults	0.84	0.99	0.92	0.04

Table 3.10: Results comparison for the l-computing methods.

As it is inferable from the table above, the most relevant results, in terms of coherence and variability, are obtained by the *Swaps plus defaults* method. As a consequence, this method will be used in the final computation of the results. Nevertheless, a good alternative method might be the *Classical Swaps - ReLU* one, which involves coherence of the results and variability. All the code for this part can be found in Appendix A, Section A.5.

After the computation of the vector l, the remaining parameter to be estimated is the interdependence parameter α. To accurately estimate this parameter, it is imperative to comprehend its underlying significance. As expounded in Chapter 2, parameter α serves as a multifaceted indicator, encapsulating both the heavy-tailed nature of the mixing variable Λ and the intricate web of interrelationships that bind obligors within the analyzed asset pool. Notably, from this juncture onward, it becomes indispensable to underscore that, in the context of the copula framework, this

[3]For reproducibility purposes, the associated seed for the RNG is 300829.

parameter must capture the overarching interdependencies among all obligors within the pool. This important characteristic lays the foundation for the forthcoming mathematical considerations regarding the dataset. In particular, this section will describe a method that estimates the value of the considered parameter initially for a single pair of variables, then it will be generalized for the totality of them.

To meaningfully delve into the inference of this parameter, it is imperative to establish a-priori that its admissible value range extends from 1 to $+\infty$, encompassing the spectrum from mutual independence to complete comonotonicity. This range inherently signifies that the behavior of obligors is profoundly linked to the magnitude of this parameter, specifically about the mathematical formulation of the one-parameter generator inherent to the Archimedean copula. To elucidate this point further it is possible to analyze the first column of the Table 2.1: as the parameter α progressively approaches infinity, a noteworthy trend emerges. This trend is characterized by the various obligors' features, symbolized by distinct variables, converging toward a uniform output pattern within the predefined copula structure, thereby accentuating the paramount influence of α on the emergent dynamics. To further enter into details within the estimation method, it is necessary to introduce a measure of concordance: *Kendall's* τ. This measure, often referred to as *Kendall's Rank Correlation Coefficient*, is a statistical measure used to assess the strength and direction of association between two random variables. Specifically, it quantifies the degree of concordance or discordance in the rankings of pairs of data points. Furthermore, it's a non-parametric measure of association, meaning it doesn't rely on assumptions about the distribution of the data and is robust to outliers. It can take values ranging from -1 to 1, well reflecting the characteristic of concordance or discordance. In the case of two general random variables X_1, X_2 and an independent copy of them X_1', X_2', it can be explicitly defined as:

$$\tau_{1,2} = \mathbb{E}\left[\operatorname{sign}\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right)\right)\right] \tag{3.7}$$

where $sign(\cdot)$ is the sign function. As outlined before, to reflect a measure of concordance, it involves the sign of a multiplication, measuring the concordant or discordant behavior of two random variables. A positive value for τ indicates a positive association, suggesting that higher values of one variable tend to be associated with higher values of the other and vice versa, while a negative value indicates a negative association, meaning higher values of one variable tend to be associated with lower values of the other. As well explained in [6], when dealing with data or statistical calculations, Kendall's τ has an obvious estimator, referred to as the *Sample version of Kendall's* τ. Based on the random sample $X_i = (X_{i,1}, X_{i,2}), i \in \{1, ..., n\}$, it is given by:

$$\hat{\tau}_{n,1,2} = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \operatorname{sign}\left(\left(X_{i_1 1} - X_{i_2 1}\right)\left(X_{i_1 2} - X_{i_2 2}\right)\right) \tag{3.8}$$

Moreover, it is possible to extract a measure of concordance between two variables, even though they are not normalized in the unitary range. Specifically, to lead back any variable to the range $[0, 1]$, it is possible to *rank* observations and then normalize their ranks concerning the total number of outcomes, to reconstruct a pseudo cumulative distribution function. Besides that fact, it is necessary to find a useful link to connect the measure of concordance with the estimation of the parameter α of the copula. To reconcile this fact, it is noticeable that when dealing with Archimedean copulas the measure τ can be rewritten in a closed analytical form concerning the generator ϕ of the copula. Specifically, the reformulation can be expressed as:

$$\begin{aligned}\tau =& \mathbb{P}[(X_1 - X_1')(X_2 - X_2') > 0] - \mathbb{P}[(X_1 - X_1')(X_2 - X_2') < 0] = \\ &= 4\mathbb{E}[C(U_1, U_2)] - 1 = \\ &= 1 + 4\int_0^1 \frac{\phi(t)}{\phi(t)'} dt\end{aligned} \tag{3.9}$$

where $\phi(t)'$ is the first derivative of the copula generator concerning the independent variable t and $C(\cdot, \cdot)$ is selected as the copula function. It is also important to notice that when focusing on the generator of the one-parameter copulas, it is evident that is also a function of the interdependence parameter α, as previously explained in Table 2.1. In the specific context of this thesis, two families of Archimedean copulas have been taken into consideration: the Gumbel and the Clayton families. For the first one, the measure of concordance τ results:

$$\tau(\alpha)_{Gumbel} = 1 + 4\int_0^1 \frac{\phi(t)}{\phi(t)\prime} = 1 + 4\int_0^1 \frac{tln(t)}{\alpha} dt = \frac{\alpha - 1}{\alpha} \tag{3.10}$$

while for the second one:

$$\tau(\alpha)_{Clayton} = 1 + 4\int_0^1 \frac{\phi(t)}{\phi(t)\prime} = 1 + 4\int_0^1 \frac{t^{\alpha+1} - t}{\alpha} dt = \frac{\alpha}{\alpha + 2} \tag{3.11}$$

For coherence purposes, it can be noted that the value of τ in each of the two cases is spanned in the correct range of definition. As a consequence, starting from the last calculations, the value of the interdependence parameter α can be retrieved from the inverse formula relative to $\tau(\alpha)$. Particularly, after the estimation of $\tau(\alpha)$ concerning the formula in Equation 3.8, the parameter α can be obtained in a specific way using the inverse formula related to the relative Archimedean copula. Nonetheless, when dealing with datasets comprising more than two variables, as commonly encountered in real-world datasets, the preceding computations are unsuitable, given their intrinsic limitation to a bivariate context. Consequently, Kojadinovic and Yan, as delineated in their article [7], introduced an unbiased estimator capable of accurately appraising the parameter α across the entire dataset. This method entails the utilization of the mean pairwise estimation of the interdependence parameter α, initially derived through the inverse of Equation 3.8. This estimator can be succinctly articulated as follows:

$$\begin{aligned}\hat{\alpha}_n =& \binom{d}{2}^{-1} \sum_{1 \le j_1 < j_2 \le d} \tau^{-1}\left(\hat{\tau}_{n,j_1 j_2}\right) = \\ =& \binom{d}{2}^{-1} \sum_{1 \le j_1 < j_2 \le d} \tau^{-1}\left(\binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \operatorname{sign}\left(\left(X_{i_1 j_1} - X_{i_2 j_1}\right)\left(X_{i_1 j_2} - X_{i_2 j_2}\right)\right)\right)\end{aligned} \tag{3.12}$$

where τ^{-1} is the inverse of Kendall's Tau function expressed in the function of the interdependence parameter, n is the number of features of the dataset and d is the number of obligors in the pool. Furthermore, it is possible to explicit the binomial factor as $\frac{x(x-1)}{2}$ for a generic number x. Thanks to this reformulation it is possible to estimate the value of α for a generic, but coherent, dataset. In the empirical analysis of the dataset provided, the first thing to do is the pre-processing of the variables: since the dataset also involves noninformative variables, there is a necessity to maintain only the informative ones[4]. After that, datetime attributes will be transformed into whole numbers using the MATLAB function *datenum*, which coherently transforms these data into numerical ones in an ordered way. At last, the categorical attributes will be transformed with the one-hot encoding technique. At this point, is important to notice that the dataset has dimension $n \times d_{enc}$, where n is the number of obligors in the pool, while d_{enc} is the number of features obtained after the encoding procedures. Consolidated this last observation, the estimation of both τ and α can be performed. Table 3.11 shows the results obtained for the two families of copulas.

n	d_{enc}	α_{Gumbel}	$\alpha_{Clayton}$
3972	20	1.5891	1.1782

Table 3.11: Results relative to the estimation of the parameter α.

All the code about the computation of this parameter has been incorporated into Appendix A Section A.7, necessitating certain important clarifications. Firstly, the scarcity of the informative features within the dataset might engender more uncertainty in this parameter's estimation; thus, enhancing the precision of the preceding analysis may be achievable by expanding the consideration of additional variables. Secondly, it merits acknowledgment that alternative methods for estimating this parameter have been well-documented, as expounded in the article [6]. These alternative approaches encompass other types of concordance measures, such as *Spearman's* ρ or *Blomqvist's* β, alongside likelihood-maximization techniques. Lastly, the previous way to calculate the interdependence parameter does not take into consideration variations in the spreads, which in real-world scenarios characterize extremely well the riskiness characteristic of an ABS. Notwithstanding this alternative, this method is a good trade-off between execution time and precision. Furthermore, it is expedient to recognize that the computational time involved is on the order of approximately 7 seconds, rendering it particularly efficient, especially when contrasted with alternative methodologies that necessitate certain assumptions that may not be tenable within the present analytical context. Conclusively, it has to be outlined that the presence of a single parameter in the copula limits the interdependence modeling possibilities. It is a consequence of the fact that the parameter α is an all-encompassing interdependence parameter, thus not considering single relations between different obligors or groups of them. Regarding the code formulation, aside from the previously elucidated estimation approach we have inserted the possibility to choose which value of α has to be taken into consideration. This necessity arises to make more fluent the computation in the case of stress-test simulations or the case of unique variations in other parameters. As a consequence of the esteems presented above, to obtain important results we have chosen the value of the parameter related to the Gumbel copula, which will be taken into consideration in the estimation of the loss-tail probabilities and of the CVaR.

[4]Specifically, at the beginning of the following analysis, the only variables taken into considerations are *Amount Loan, Current Amount Loans, Gross Coupon, Original Maturity Term, Remaining Term, Age, Issue Date, Delinquency, Months Delinquency, Maturity Date, Direct Deal Group, and Defaulted*

3.3 Results

In this section, the most important results regarding the application of the model presented in Chapter 2, Section 2.6, will be presented, analyzed, and compared. The parameters computed and estimated in the previous section will be used in the model to obtain the loss tail probabilities, the value of CVaR, and the value of v*. The remaining parameters to be chosen are:

- **The pool's quality function** f_n: in the following experiments the function that will be used is $f_n = \frac{1}{\sqrt{n}}$, while other functions could be tested, as presented in Table 2.5. This decision is based on the fact that, since all these loans are consumer loans their quality is relatively medium.
- **The threshold parameter b**: in the following calculations, the value of the relative threshold parameter b has been chosen to represent 70% of the total value of the underlying loans, divided by the total number of loans in the pools. Furthermore, in additional analysis, it will be varied to observe the behavior of the loss tail probability following this change.

To clarify the results obtained in the previous section, is important to remember that the parameter α has been estimated to be 1.5891, following the reformulation given by the Gumbel's Archimedean copula. Moreover, the quality vector l has been computed with the *Swaps plus defaults* method. Based on these specifications, the results obtained by this thesis' method are shown in Table 3.12.

$\mathbb{P}(L_n > nb)$	**CVaR**	**v***	**Execution time**
$5.2553 \cdot 10^{-3}$	542029965.6250	1.4413	111.15 seconds

Table 3.12: Final results relative to the loan data tape.

Several pertinent observations arise from these outcomes. Initially, the estimated probability is notably low, a consequence of the assumption that occurrences of such substantial loss magnitudes are exceedingly rare. Moreover, the Conditional Value at Risk exhibits a robust estimate of the expected loss in the event of a default, nearing approximately 92% of the aggregate current pool value. In conclusion, the value of v* aligns cohesively with the probability definition, explicitly articulated in Equation 2.29, and the computational time remains under the two-minute threshold.

3.4 Sensitivity analysis

Based on the findings elucidated in the previous section, distinct patterns emerge, especially when altering selected parameters. To gain a precise understanding of the tail behavior in loss tail probabilities and CVaR, a pivotal step involves generating a probability curve in response to variations in the threshold parameter b. This step is performed by making b assume different percentage values between 10% and 99%.

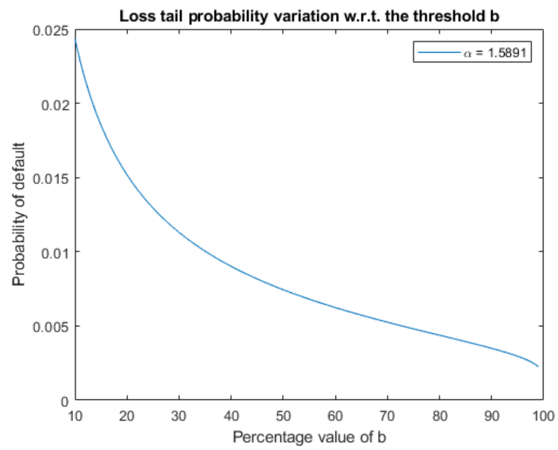


Figure 3.15: Loss tail probability with respect to the variation of the threshold parameter b.

Consistent with the model's assumptions, the value of the probability decreases as the threshold parameter b increases. Moreover, the trend does not reflect any particular linear behavior concerning this variation. In real-world scenarios, non-linear patterns are frequently observed, underscoring the need for comprehensive calculations to establish a robust scenario. Furthermore, the same analysis has been conducted on the behavior of the Expected Shortfall.

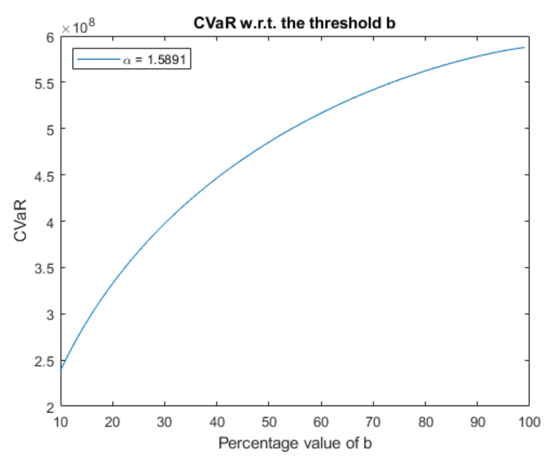


Figure 3.16: CVaR with respect to the variation of the threshold parameter b.

As can be inferred by the chart in Figure 3.16, the CVaR becomes higher as the value of b grows, also this time with non-linear behavior. Moreover, to demonstrate the model's robust generalization concerning the obligors' interdependence, one can analyze the changes in the loss tail probability and Conditional Value at Risk concerning variations in the parameter α, while maintaining the fixed threshold parameter $b = 70\%$.

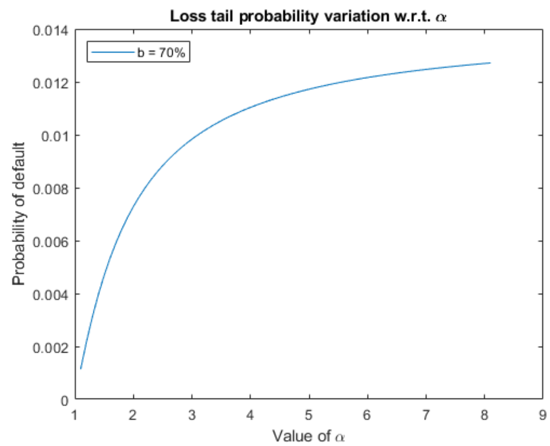


Figure 3.17: Loss tail probability variation with respect to the variation of the parameter α.

As it can be seen in Figure 3.17, as the parameter α increases the associated loss tail probability gets higher, manifesting in another particular way the coherence of the model concerning real-world assumptions.

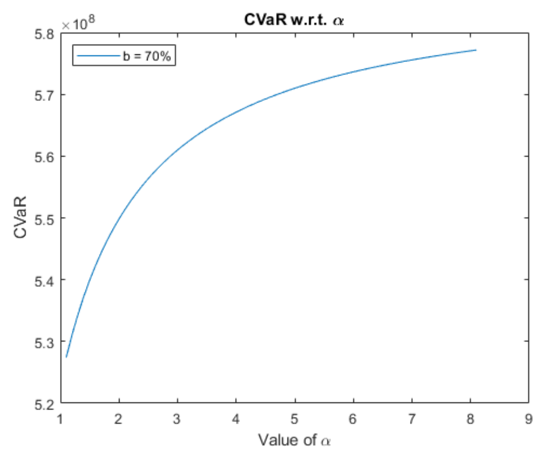


Figure 3.18: CVaR variation with respect to the variation of the parameter α.

Also in the scenario depicted in Figure 3.18, it is evident that the expected shortfall value rises in tandem with an increase in the interdependence parameter α, illustrating the non-linear linkages concerning the outcomes of this particular stress test.

At this point, the necessity of exhibiting the overall robustness of the approach arises. To generalize the quality of the pool, some other tests have been performed with respect to the variation of the corresponding quality function. The results can be seen in the following table.

	$\frac{1}{n}$	$\frac{1}{\sqrt{n}}$	$\frac{1}{\ln(n)}$	$\frac{Defaulted\_quantity}{Total\_quantity}$
$\mathbb{P}(L_n > nb)$	$8.3386 \cdot 10^{-5}$	$5.2553 \cdot 10^{-3}$	$1.1505 \cdot 10^{-1}$	$8.5072 \cdot 10^{-3}$

Table 3.13: Probability results after the variation of the pool's quality function.

To manifest the importance of the quality function, it is inferable from Table 3.13 that by reducing its magnitude the probability of large losses increases, thus manifesting its correlation with the overall quality of the pool. Moreover, in the last column is possible to analyze a data-based approach: by defining the quality of the pool as the percentage of the current amount of the defaulted loans inside[5] with respect to the current total quantity, it is possible to obtain a data-driven esteem of its quality. In this case, the probability results take into evidence a possible middle overall quality characteristic of the pool. It is noteworthy that the Conditional Value at Risk and the v* values do not exhibit analytical dependency on the quality function. This observation prompts a dual perspective. On one hand, it presents a favorable aspect wherein the individual components of the vector l serve as quality checkers, thereby enhancing the model's robustness. Conversely, the absence of interdependence between these values and changes in the quality function precludes an unequivocal assessment of specific quantities based on overall quality considerations.

The only parameter remaining to be varied to assess the robustness of the approach is the parameter α. Despite it can be esteemed directly by the data, the necessity of doing stress-tests by varying this parameter arises. To better show the efficacy of these tests, 2-D and 3-D plots are subsequently shown.

[5] If there are no already defaulted loans in the pool, the function is replaced by $\frac{1}{n}$.

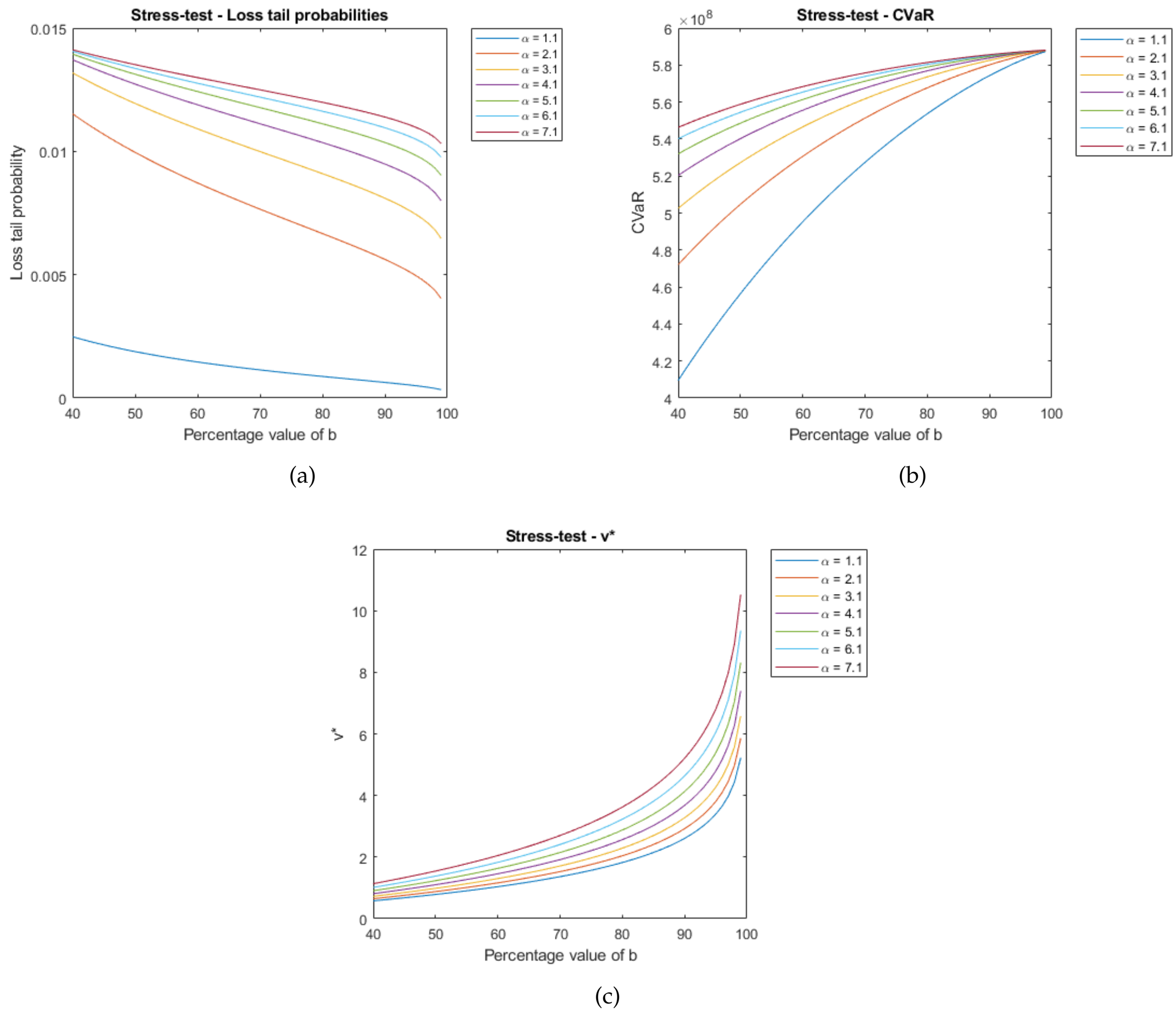
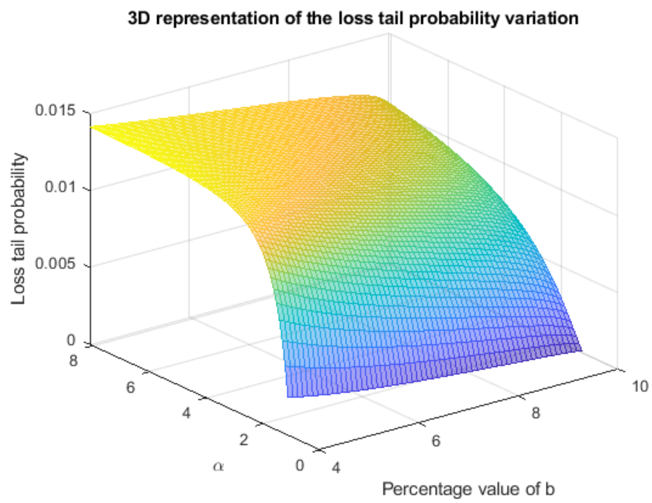
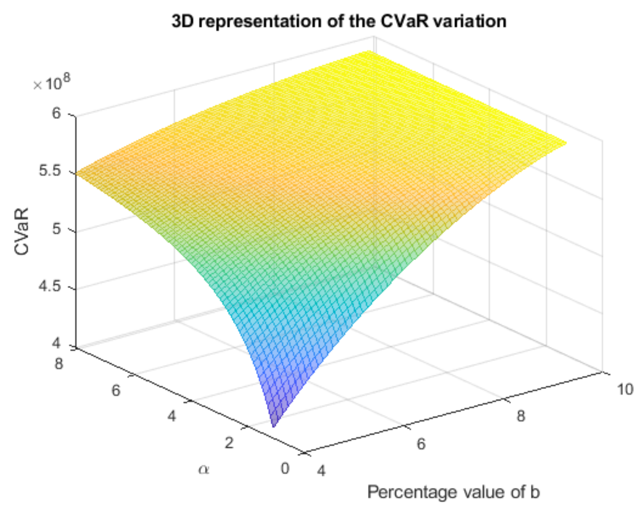


Figure 3.19: Loss tail probabilities, CVaR and v* by varying α and b - 2D representation.

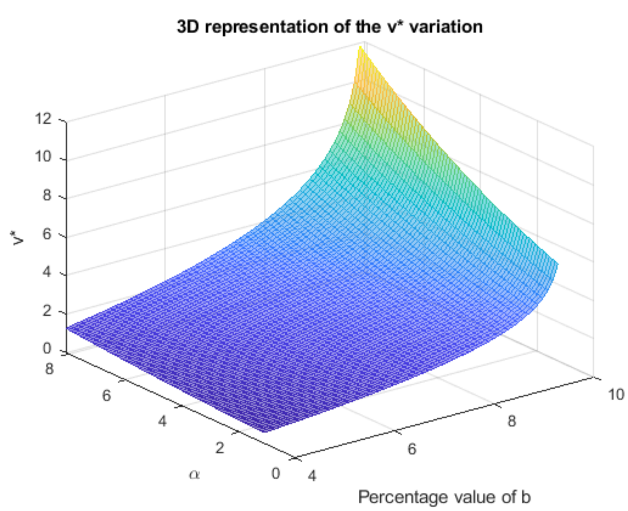
As depicted in Figure 3.19, each of the three variables exhibits distinctive responses to variations in both parameters. In Sub-figure 3.19(a), exemplified by the loss tail probability outcomes, a discernible trend emerges: an increase in the interdependence parameter α corresponds to elevated probabilities, aligning consistently with the fundamental concept at the base of the approach's definition. Sub-figures 3.19(b) and 3.19(c) reveal that the behavior of these variables about alterations in the parameter α aligns with real-world expectations. Specifically, heightened interdependence results in progressively higher loss given default values.



(a)



(b)



(c)

Figure 3.20: Loss tail probabilities, CVaR and v* by varying α and b - 3D representation.

Within the three sub-figures presented in Figure 3.20, the behavior of the three examined variables aligns harmoniously with the model's specifications and the fundamental principles of credit risk analysis, thus manifesting its robustness peculiarity. Notably, the probability of large losses increases concomitantly with a growth in the interdependence parameter α and a reduction in the threshold parameter b, signifying that a heightened obligor interdependence within the structure is associated with an elevated risk of failure. Furthermore, it is evident that when α assumes a low value, indicating a lack of obligor interdependence, the probability of substantial losses diminishes, demonstrating the detachment of defaults within the total pool. Moreover, the analysis of the Conditional Value at Risk mirrors the model's underlying assumptions, supporting the fact that a higher probability of large losses corresponds to increased losses in the event of default. At last, the systematic factor v* exhibits an upward trajectory as both analyzed parameters increase, indicating a tendency for a broader spectrum of issuer failures. It is noteworthy that this factor assumes higher values particularly when b assumes elevated values.

3.5 Pricing and comparison with the market practice results

The prevailing market pricing procedure for Asset-Backed Securities involves a methodical approach. As elucidated in Section 1.1, ABS encompasses a diverse spectrum of credit derivative instruments, giving rise to nuanced pricing considerations contingent upon the specific asset class. In the context of this thesis, our focus is on the subset encompassing consumer and personal loans, as delineated in the right portion of Figure 1.1. Within the pricing protocol for these instruments, financial analysts amalgamate disparate financial modeling techniques to unearth pertinent insights. The principal objective is the determination, for a given pool, of the *prepayment rate* and *credit spread* for each tranche. To be precise, the credit spread measure serves as a metric characterizing the inherent risk level of a financial instrument, denoted in basis points. As explained in the literature [16] and [11], a different number of such measures exist and are commonly known with the previously cited name since they attempt to measure the return of the credit asset relative to some higher credit quality benchmark. For a comprehensive assessment of the aforementioned financial metric, various characteristics of the pool are subjected to rigorous analysis. These attributes include the issuance nation, along with an array of additional factors, encompassing the issuance date, loan type, instrument rating, and, most notably, the tranche seniority requiring evaluation. Typically, these market data are sourced through management software. Furthermore, all the historical data associated with the loan pool assume a crucial role within the pricing framework. However, in the specific scenario of a recently introduced ABS within the tradable market, there exists no historical data about prepayment rates. Consequently, rating agencies provide a range for this parameter, grounded in historical analyses of analogous instruments. Given all these specifications, the overall pricing process of an ABS involves the pricing of each tranche. At last, the specific market practice pricing procedure of an Asset-Backed Security follows the discounted cash-flow method. In particular, without the possibility of having the certainty of deterministic cash flows deriving from the underlying loans, it is necessary to model them using their expected value, obviously discounted by the appropriate discounting factor. It is further necessary to explain that each tranche of the total pool could have a different price as well as a different interest rate premium.

The general discounted cash-flow formula for each tranche t can therefore be expressed as:

$$P_t = \frac{\sum_{i=N_0^t}^{N_f^t} \zeta\left(\frac{i}{m}, r + \text{spread}\right) \cdot \mathbb{E}[C_i]}{V_t} \cdot 100 \tag{3.13}$$

where the apex t is the tranche index, N_0^t is the starting index of refunding of each tranche, N_f^t is the final index of refunding of each tranche, V_t is the outstanding of each tranche, C_i is the stochastic cash-flow at time i, *spread* is the evaluated credit spread for the ABS and $\zeta\left(\frac{i}{m}, r + \text{spread}\right)$ is the discounting factor evaluated at the year fraction $\frac{i}{m}$ with respect to the interest rate r and the associated credit spread. To discount each cash flow coherently, the interest rate r considered for the discounting factor is assumed to be the EURIBOR with a maturity of one month, fixed at the date of the first of September, 2023. Moreover, its functional form is assumed to be the simple discrete discounting:

$$\zeta\left(\frac{i}{m}, r + \text{spread}\right) = \frac{1}{1 + \frac{r+\text{spread}}{m} \cdot i} \tag{3.14}$$

with the further specification that the payments are related to a monthly basis, such as $m = 12$. Following these specifications, the most important part of the pricing procedure is to establish in a coherent way how to model the expected cash flows at each time. In addition to the estimate of the probability of large and different losses of the pool and its prepayment rate, there is a necessity to establish the temporary distribution of defaults. This need follows from the intention to understand more precisely when large losses occur and therefore give a more time-accurate pricing. In real-world scenarios, the two default functional forms most used involve a decreasing structure, as it is assumed that a debtor who has been able to repay the majority of the loan manages to complete the last payments, or a uniform structure, for which the estimated total defaults occur uniformly throughout the duration of the loan. In some cases, losses are also assumed to happen at the first possible time. Moreover, it is necessary to specify that repayments happen every month. Starting from all the previous specifications, the cash quantity available at each month i, which determines the stochastic cash-flows, can be modeled in formulas as:

$$\mathbb{E}[C_i] = CF_i - CD_i + R_{i-1} \tag{3.15}$$

where CF_i is the total cash flow that theoretically is generated by the underlying loans, which always theoretically has to be paid by obligors at time i, CD_i is the cumulative gross default, such as how much money has been lost since the issue of the security, and R_i is the recovered amount of money at time i. Each expected cash-flow $\mathbb{E}[C_i]$ is composed of a capital part q_i and an interest part r_i, which are highlighted in the loan data tape for each loan. Moreover, the theoretical cash flows CF_i could be increased by prepayments, thus symbolizing the necessity of searching for prepayment rates.

To best specify each element of Equation 3.13, it is imperative to notice that the values of N_0^t and N_f^t are not deterministic, because they depend on the time in which each tranche is fully exhausted by capital payments. As a consequence, they have to be evaluated numerically by setting a system of two equations:

$$\sum_{i=N_t^0}^{N_t^f} q_i = V_t, \qquad t \in S, M \tag{3.16}$$

It is possible to notice that each of the previous equations is not linear, thus they cannot be solved in closed form, but only numerically. To do that, the usage of an exhaustive search algorithm to find the extremes of the summation is necessary. Moreover, assuming that $N_0^S = 1$, notice that the previously named research cannot be carried out in parallel, but only in series. Specifically, the search can continue only when the final period of refunding of the senior tranche is found, and so on with the other tranche extremes. Following all the previous model peculiarities, the price of each tranche can be evaluated. The results that will be shown consecutively are referred only to the senior tranche because it is the most important part of the ABS. Besides these specifications, the accurate pricing of the instrument under consideration in this thesis necessitates a comprehensive set of model characterizations. Firstly, among the various waterfall structures discussed in Sub-section 1.2.1, the *Sequential repayment* structure has been chosen. This structure dictates that payments in each tranche are subordinate to the fulfillment of obligations in the superior tranches. This defensive mechanism significantly mitigates losses for the senior tranche by prioritizing repayments. As a consequence, a substantial shortfall is required to trigger any loss resulting from loan defaults within the pool, thereby necessitating the derivation of tail loss probabilities. Furthermore, several other credit enhancement mechanisms have been implemented. For instance, this Asset-Backed Security incorporates overcollateralization of the loans, alongside the establishment of a cash reserve within the pool to further mitigate potential losses and fortify the lower tranches. Furthermore, an excess spread corresponding to the weighted interest rate of 7.0% is associated with the pool. This credit enhancement method serves as "money trapping" for the instrument, such as it helps investors bear fewer losses during the life of the instrument. For specific calculations, the coupon given to this tranche is equal to the one-month EURIBOR rate plus 70 basis points. Thanks to the calculations given by the management software, the market practice result at the date of September 29, 2023, shows a price of 99.80.

The procedural framework we propose in this last part of the section unfolds as follows. Leveraging the credit risk model expounded in Section 2.6, especially concentrating on Equation 2.29, we facilitate the derivation of a vector comprising probability quantiles[6]. This is achieved by systematically varying the threshold parameter, denoted as b, spanning the entire spectrum of singular percentage values between 10% and 99% concerning the outstanding of the underlying pool. The resultant vector allows us to deduce singular quantile probabilities by taking the difference between consecutive elements of the previously named vector. To better specify, the useful values that have to be taken into consideration for the parameter b are the percentage values of loss that could impact the tranche to be priced; in our case, the corresponding value of the relative tranche-impact threshold is $b^* = 47\%$. In addition to these quantile probabilities, we can compute loss threshold exposures for the designated values of b, represented as the product $n \cdot b$, where n denotes the number of obligors in the pool as elucidated in Table 3.11. Furthermore, drawing from

[6]For the calculation of these quantiles, the quality of the pool is set as *medium quality*, corresponding to the function $f_n = \frac{1}{\sqrt{n}}$, while the parameter α and the vector l are computed as explained in Section 3.2

the payment structure of the pool as defined by its monthly cash flows, we can compute theoretical cash flows and subject them to the appropriate discounting methodology. To this extent, it is noteworthy that the maximum payment date T corresponds to the point at which, assuming no losses, the tranche attains full payment. Thanks to these specifications, the *nonhazardous* price can be computed using the discounted cash-flow method:

$$P_{nh} = \sum_{i=1}^{T} \zeta\left(\frac{i}{12}, r\right) \cdot C_i \tag{3.17}$$

where the discounting factor is expressed as in the first point of the list presented in this section and T is the maximum payment date. The nonhazardous price value, in the case of the analyzed pool, results to be 100.59. After that, our proposed methodology aims to find the *no-loss* price, such as the risk-free price scaled by a factor corresponding to the probability of a loss lower than the tranche-impact threshold:

$$P_{nl} = P_{nh} \cdot \mathbb{P}(L_n < nb^*) = P_{nh} \cdot (1 - \mathbb{P}(L_n > nb^*)) \tag{3.18}$$

where b^* is the relative tranche impact threshold. In our case, this value turns out to be 99.75. At this juncture, for each value of b and, correspondingly, for each threshold $n \cdot b$, the anticipated cash flows can be meticulously assessed. Specifically, assuming a temporal framework where losses are presumed to manifest at the earliest conceivable opportunity, cash flows will be meticulously adjusted to the extent that they affect the senior tranche. In our case, the current affecting value is in the order of $2.70 \cdot 10^8$, signifying that each loss surpassing this threshold will induce a sequential reduction in the cumulative cash flows. Following this computation, each of these cash-flows is subjected to pricing using the formula denoted in Equation 3.17, thereby yielding 53 distinct prices, each denoted as $P_j, \quad j = 1, ..., 53$. Conclusively, to derive an estimate of the issue's price, we augment the price computed through Equation 3.18 by the summation of the product of each cash flow price P_j, previously determined, multiplied by the corresponding quantile probability. The ultimate price is subsequently evaluated through the following formula:

$$P_{fin} = P_{nl} + \sum_{j=1}^{53} P_j \cdot \left(\mathbb{P}(L_n > nb_{j+46}) - \mathbb{P}(L_n > nb_{j+47})\right) \tag{3.19}$$

where b_j is the threshold parameter indicating the $j^t h$-percentage level. Note that this computation is similar to a discretization of the CVaR formula expressed in Equation 2.17, relative to the price of the analyzed issue. The application of this formula to the ABS analyzed in this thesis leads to a final price of 100.10. As can be finally observed, this price is very similar to the one obtained by the market practice procedure, with a relative discrepancy of only $3.0060 \cdot 10^{-3}$.

Conclusions

This thesis has presented an innovative approach for estimating loss tail probabilities and the CVaR risk measure for a generic loan pool, which are further applied to a real-world loan data tape provided by Intesa Sanpaolo S.p.A. By employing a stochastic representation of an LT-Archimedean copula, the interdependence structure of the pool's loss function was effectively modeled. This modeling approach addresses the need to accurately represent rare events, which are of paramount importance in real-world scenarios. Consequently, this approach finds analytical esteems of the asymptotic probability of large losses deriving from the pool as well as an asymptotic estimation of CVaR. Based on the model's assumptions, we also developed a variance reduction algorithm, called Conditional Monte Carlo. Consecutively, the approach was rigorously tested using real-world data. The dataset played a pivotal role in evaluating the interdependence correlation parameter and the esteem of the pool's overall quality. As a result, we successfully derived a robust probabilistic structure for loss tail probabilities, which is integral to the pricing of the pool itself. It is important to note that the explained approach while achieving good results in terms of pricing, does make certain qualitative assumptions in the process. The significance of the qualitative assumptions in the pricing procedure underscores the inadequacy of solely determining a singular probability of large losses or accurately estimating the timing of defaults. Instead, a holistic analysis that takes both factors into account could be crucial and of certain interest in future research. The utilization of advanced mathematical techniques, such as the LT-Archimedean copula structure, which is increasingly gaining prominence in the realm of finance, has proven to be a valuable and innovative tool for addressing the complex task of pricing a real Asset-Backed Security. This complexity justifies the adoption of such mathematical techniques, especially when compared to other financial instruments.

Appendix A

MATLAB Codes

A.1 Asymptotic estimate's code - special homogeneous case

```matlab
%% Asymptotic probability estimate
% This code follows the approach presented in this thesis in the homogeneous case
format long e
% Setting the seed of the RNG for reproducibility
rng(300829)

% Settable values
n = 500; % Number of different obligors in the pool
nrep = 10000; % Number of repetitions
alpha = 2; % Interdependence coefficient (Estimable parameter)
b = 0.5; % Percentage of maximum loss
c = ones(n,1); % Weights of the various assets
l = 0.5*ones(n,1); % Values of l_i, reflecting the quality of each obligor
f_n = @(x) 1/x; % Decreasing function simbolizing the quality of the pool

% Calculating vstar as in the article
vstar = l^(-alpha)*log(c/(c-b));

% Calculating the probability of large losses with the explicit analytic
% formula.
if alpha == 1
    prob_asymp = f_n(n)*vstar^(-1/alpha);
else
    prob_asymp = f_n(n)*vstar^(-1/alpha)/gamma(1-1/alpha);
end

fprintf('The estimated probability (with correlation) with n = %d and alpha = %d is %0.7d. \n', n, alpha, prob_asymp);
```

A.2 Conditional Monte Carlo - Algorithm for large losses' probability estimation

```
%% Conditional Monte Carlo estimator for portfolio losses
% Rare-event variance reduction technique used to estimate large losses'
% probabilities in the context of a credit portfolio.
% To conduct the simulations it is possible to change the values
% of the variables in the part of 'Settable values'.
clear all
close all
clc
format long e
% Setting the seed of the RNG for reproducibility
rng(300829)

% Settable values
n = 500; % Number of different obligors in the pool
nrep = 10000; % Number of repetitions
alpha = 2; % Interdependence coefficient (Estimable parameter)
b = 0.5; % Percentage of maximum loss
c = ones(n,1); % Weights of the various assets
l = 0.5*ones(n,1); % Values of l_i, reflecting the quality of each obligor
f_n = @(x) 1/x; % Decreasing function simbolizing the quality of the pool

% Useful functions
phi = @(x, alpha) (-log(x)).^alpha; % Copula generator
phival = phi(ones(n,1)-f_n(n)*l, alpha);
cs = (cos(pi/(2*alpha)))^(alpha);

% Making the probability distribution of the one-sided standard stable rv.
% IMPORTANT: the value of 'delta' has to be set different whenever the
% value of alpha is changed, because the distribution has to start in 0.
% alpha = 1.1 --> 'delta' = 0.5
% alpha = 2 --> 'delta' = 0.25
% alpha = 5 --> 'delta' = 0.25
stable = makedist('Stable', 'alpha', 1/alpha, 'beta', 1, 'gam', cs, 'delta',
    0.5);

% For plotting purposes of the one-sided stable standard rv
x = -1:.001:3;
pdf1 = pdf(stable,x);
figure
plot(x,pdf1,'b-');
title("One-sided standard Stable distribution with \alpha = " + alpha)
legend("\alpha =" + alpha,'Location','northwest')

%% Simulation's part start
Vtot = zeros(nrep,1);
```

```
    for step = 1:nrep

        % First phase: simulating from independent standard exponential rvs and
        % from the one-sided stable rv
        V = random(stable, 1, 1); % Simulating from the one-sided stable rv
        R = exprnd(1, n, 1); % Standard: lambda = 1

        % Second phase: trasformation of the exponential outcomes in the vector O
        O = R./phival;

        % Third.one phase: ranking the values of O
        [ordO, indexes] = sort(O, 'descend');
        tot = 0;

        % Calculating the correct index
        for i=1:n
            if tot >= n*b
               break
            else
                tot = tot + c(indexes == i);
            end
        end

        i = n*b;
        reqO = ordO(n-i+1);
        Vtot(step) = V > reqO;

    end

    % Third.two phase: estimating the probabilty of a loss greater than n*b
    prob_est = sum(Vtot)/nrep;
    fprintf('The estimated probability of a loss larger than %d is %0.8f. \n', n*b,
       prob_est);
```

A.3 Conditional Monte Carlo - Algorithm for Expected Shortfall in the homogeneous case

```
%% CMC: Estimation of the expected shortfall
% Rare-event variance reduction technique used to estimate the CVaR
% in the context of a credit portfolio.
% To conduct the previous simulations it is possible to change the values
% of the variables in the part of 'Settable values'.
clear all
close all
clc
format long e
% Setting the seed of the RNG for reproducibility
rng(300829)
% Settable values
n = 1000; % Number of different obligors in the portfolio
nrep = 1000000; % Number of repetitions
alpha = 2; % Dependence coefficient (Estimable parameter)
b = 0.8; % Percentage of maximum loss
c = 1; % Weights of the various assets
l = 0.5; % Values of l_i, simbolizing variations effect among obligors
f_n = @(x) 1/x; % Decreasing function simbolizing the quality of the portfolio

% Useful functions
phi = @(x, a) (-log(x)).^a; % Copula generator
phival = phi(ones(n,1)-f_n(n)*l, alpha);
cs = (cos(pi/(2*alpha)))^(alpha);
rv = @(v) 0.5^(alpha) * exp(-v.*(0.5^(alpha))).*v.^(-1/alpha);

% Making the distribution of the one-sided standard stable rv.
% IMPORTANT: the value of 'delta' has to be set different whenever the
% value of alpha is changed, because the distribution has to start in 0.
% alpha = 1.1 --> 'delta' = 0.5
% alpha = 2 --> 'delta' = 0.25
% alpha = 5 --> 'delta' = 0.25
stable = makedist('Stable', 'alpha', 1/alpha, 'beta', 1, 'gam', cs, 'delta',
    0.25);


%% Simulation start
Vtot = zeros(nrep,1);
Loss = zeros(nrep,1);

for step = 1:nrep

    % First phase: simulating from independent standard exponential rvs and
    % from the one-sided stable rv
    V = random(stable, 1, 1); % Simulating from the one-sided stable rv
```

```
        R = exprnd(1, n, 1); % Standard: lambda = 1

        % Second phase: trasformation of the exponential outcomes in the vector O
        O = R./phival;

        % Third.one phase: ranking the values of O
        [ordO, indexes] = sort(O, 'descend');
        tot = 0;

        i = n*b;
        reqO = ordO(n-i+1);
        Vtot(step) = V > reqO;
        Loss(step) = sum(V>O);

    end

    % Third.two phase: estimating the probabilty of a loss greater than n*b
    prob_est = sum(Vtot)/nrep;
    fprintf('The estimated probability of a loss larger than %d is %0.8f. \n', n*b,
        prob_est);

    % Fourth phase: estimating the expectedvalue of losses larger than n*b
    expectedvalue = sum(max(Loss - n*b*ones(nrep,1), 0))/nrep;

    % Calculating the CVaR
    CVaR = n*b + expectedvalue/prob_est;
    fprintf('The estimated CVaR is %0.4f. \n', CVaR);

    %% Asymptotic results
    % Calculating vstar as in the article
    vstar = l^(-alpha)*log(c/(c-b));

    % Calculating the probability of large losses with the explicit analytic
    % formula.
    if alpha == 1
        prob_asymp = f_n(n)*vstar^(-1/alpha);
    else
        prob_asymp = f_n(n)*vstar^(-1/alpha)/gamma(1-1/alpha);
    end

    % Calculating the asymptotic value of CVaR
    asyCVaR = n*b + n*integral(rv, vstar, +Inf)/(vstar^(-1/alpha));
    fprintf('The asymptotic CVaR is %0.4f. \n', asyCVaR);

    %Calculating the relative discrepancy
    reldisc = abs(asyCVaR - CVaR)/CVaR;
    fprintf('The relative discrepancy is %0.8f. \n', reldisc);
```

A.4 General computing function - loss tail probability, CVaR and v*

```
%% General function for estimating portfolio losses' characteristics
% Function that estimates the probability of large losses of a given bond
% portfolio as well as the relative CVaR.

% INPUTS
% n : number of obligors in the pool
% nrep: number of repetition for the Simulative method
% alpha: iterdependence parameter (heavy tailedness parameter). It admits
%        values in [1,Inf).
% b : threshold for the loss
% c: vector of the expositions of the loss
% l: quality vector of the obligors
% f_n: quality function of the pool
% copula: copula function to be used in the Exact calculation. It admits
%         the strings "Gumbel" and "Clayton"
% seed: seed of the RNG for reproducibilty of the results
% method: method to be used. It admits the strings "Simulative" or "Exact"

% OUTPUTS
% prob: probability of a default larger than n*b
% CVaR: expected shortfall
% vstar: threshold value of the systematic factor V. Only computed when
%        method == "Exact"


function [prob, CVaR, vstar] = archest(n, nrep, alpha, b, c, l, f_n, copula,
   seed, method)

% Setting the seed of the RNG for reproducibility
rng(seed)

% Copula function to be used
if copula == "Gumbel"

    % Gumbel's copula generator
    phi = @(x, a) (-log(x)).^a;
    cs = (cos(pi/(2*alpha)))^(alpha);
    if alpha < 2
        delta = 0.5;
    else
        delta = 0.25;
    end
    % Mixing variable: one-sided Stable standard random variable
    varb = makedist('Stable', 'alpha', 1/alpha, 'beta', 1, 'gam', cs, 'delta',
       delta);

```

```
    elseif copula == "Clayton"

        %Clayton's copula generator
        phi = @(x, a) x.^(-a) - 1;
        % Mixing variable: Gamma randoma variable
        varb = makedist('Gamma', 'a', 1/alpha, 'b',  1);

    end

    % Simulative method
    if method == "Simulative"

        % Initializations
        Vtot = zeros(nrep,1);
        Loss = zeros(nrep,1);
        phival = phi(ones(1,n)-f_n(n)*l, alpha);

        % Simulation process
        for step = 1:nrep

            V = random(varb, 1, 1);
            R = exprnd(1, 1, n);
            O = R./phival;

            % Ordering the obtained values
            [ordO, indexes] = sort(O, 'descend');
            tot = 0;

            % Finding the ranked threshold value
            for i=1:n
                if tot >= n*b
                   break
                else
                    tot = tot + c(indexes == i);
                end
            end
            reqO = ordO(n-i+1);

            % Evaluating the exceeding of the threshold
            Vtot(step) = V > reqO;

            % Evaluating the total loss
            Loss(step) = (V > O)*c';
        end

        % Computing the probability of default
        prob = sum(Vtot)/nrep;
```

```
        fprintf('The estimated probability of a loss larger than %d is %0.8f. \n', n*b, prob);

        % Computing teh excpected shortfall
        expectedvalue = sum(max(Loss - n*b*ones(nrep,1), 0))/nrep;
        CVaR = n*b + expectedvalue/prob;
        fprintf('The estimated CVaR is %0.4f. \n', CVaR);

        % For computational purposes, it is necessary to explicit vstar
        % in another way
        vstar = "ONLY for Exact method";


    % Exact asymptotic method
    elseif method == "Exact"

        % Computing the frequency vector
        w = zeros(1,n);
        vals = [c;l];
        couples = zeros(2,n);
        count = 1;

        % Finding the couples then the frequency of each couple
        for i = 1:n
            if vals(1,i) == 0
                continue
            else
                ind1 = find(vals(1,:) == vals(1,i));
                ind2 = find(vals(2,:) == vals(2,i));
                int = intersect(ind1,ind2);
                couples(:,count) = vals(:,int(1,1));
                w(count) = size(int,2)/n;
                count = count + 1;
                vals(1,int) = 0;
            end
        end
        couples = couples(:,1:count-1);
        w = w(1:count-1);

        % Expliciting some useful functions:

        % r(v) function:
        rv = @(v) sum(couples(1,:).* w.* (1-exp(-v.*(couples(2,:).^alpha))));

        % Scaled r(v) function used to calculate the value of vstar numerically
        rvminusb = @(v) rv(v) - b;
        vstar = fzero(rvminusb, 1);

```

```
        % Derivative of r(v) multiplied by v^(-1/alpha) (used to calculate
        % numerically the integral.
        rvprime = @(v) (sum((couples(1,:).*w.*(couples(2,:).^alpha))*(exp(-v'*(
           couples(2,:).^alpha)))',1)).*v.^(-1/alpha);

        % Computing the asymptotic probability of losses larger than a
        % threshold
        if alpha == 1
            prob = f_n(n)*vstar^(-1/alpha);
        else
            prob = f_n(n)*vstar^(-1/alpha)/gamma(1-1/alpha);
        end
        fprintf('The asymptotic probability of a loss larger than %d is %0.8f. \n',
           n*b, prob);

        % Computing the expected shortfall
        CVaR = n*b + n*integral(rvprime, vstar, +Inf)/(vstar^(-1/alpha));
        fprintf('The asymptotic CVaR is %0.4f. \n', CVaR);

    end
```

A.5 l-computing function

```
    %% l computing function
% Function that computes the  obligor's quality vector
% IMPORTANT: the vector must have each value under 1 and greater than 0.8

% INPUTS
% - BD: Basedati loan data tape
% - SW: Tassiswap dataset
% - method: method to be used to compute l. The choices are "Uniform",
%           "Random", "ClassicalSwaps" and "Swapsplusdefaults".
% - n: number of obligors in the pool
% - arg: function to be used in the "Classicalswaps" method. The choices
%        are "Minmax", "Sigmoid", "ReLu".
% OUTPUT
% - l: obligor's quality function.

function [l, spread] = compute_l(BD, SW, method, n, arg)

% Uniform assignment
if method == "Uniform"

    l = 0.8*ones(1,n);
    spread = l;

% Pseudo-Random assignment
elseif method == "Random"

    rng(300829)
    l = rand(1,n);
    l(l < 0.8) = 0.8;
    l(l > 0.99) = 0.99;
    spread = l;

% Assignment using functions and swap rates
elseif method == "ClassicalSwaps"

    % Recalling the classical_swaps function
    % IMPORTANT: requires high computational time for the calculation
    spread = classical_swaps(BD, SW);

    if arg == "Minmax"
        l = 0.8 + 0.19*(spread'-min(spread))/(max(spread)-min(spread));
    elseif arg == "Sigmoid"
        l = 0.8 + 0.19*(1./(1+exp(-spread')));
    elseif arg == "ReLu"
        l = 0.8 + 0.19*max((spread'-1), 0)/max(spread);
    end
```

```

% Assignment using Swap rates and defaults of single obligors
elseif method == "Swapsplusdefaults"

    % Recalling the classical_swaps function
    % IMPORTANT: requires high computational time for the calculation
    spread = classical_swaps(BD, SW);

    l = zeros(1,n);
    % High value because they are already defaulted
    l(BD.Defaulted == "Y") = 0.99;

    % Rescaling w.r.t. the number of defaults during the issue's life
    MD = BD.MonthsDelinq';
    condN = BD.Defaulted == "N";
    l(condN) = (1/(5*pi) * (atan(MD(condN)))) + ...
               0.1*(1./(1+exp(-spread(condN)'))) + 0.79;
end
```

A.6 Spreads-computing function

```
%% Spreads computing function
% Function that calculates the spread of each loan
% INPUTS
% - BD: Basedati loan data tape
% - SW: Tassiswap dataset
% OUTPUT
% - spread: vector of spreads of each obligor in the pool

function [spread] = classical_swaps(BD, SW)

% 'BD' important data
c = BD.AmountLoan;
T = BD.OrigMaturityTerm;
Coup = BD.GrossCoupon;
Issues = BD.IssueDate;
MD = BD.MonthsDelinq;
Def = BD.Defaulted;

% 'SW' important data
BuD = SW.BUSINESS_DATE;
Val = SW.VALUE;

% Computing the years to maturity of each obligor
ytm = zeros(4139,1);
fl = floor(T/12);
ce = ceil(T/12);
```

```matlab
cond1 = ((T/12 - floor(T/12)) < 0.5) + (T/12 >= 15) > 0;
cond2 = ((T/12 - floor(T/12)) >= 0.5) + (T/12 < 15) == 2;
ytm(cond1) = fl(cond1); % Because ceil gives back 16 years loans too
ytm(cond2) = ce(cond2);

% Computing the relative indexes
ytmindex = zeros(length(c), 1);
for i = 1:length(T)
    if ytm(i) == 1
        ytmindex(i) = 10;
    elseif ytm(i) < 12
        ytmindex(i) = ytm(i)-2;
    else
        ytmindex(i) = ytm(i)-1;
    end
end

% Searching for the correlated swap rates
corr_swap = zeros(length(T), 1);
for i = 1:length(T)
    [y,m,d] = split(between(Issues(i),...
                  BuD((ytmindex(i)*1307 + 1):((ytmindex(i)+1)*1307))),...
                  {'years', 'months', 'days'});
    timedistance = 365*abs(y) + 30*abs(m) + abs(d);
    [~, diffind] = min(timedistance);
    corr_swap(i) = Val(diffind + ytmindex(i)*1307);
end

% IMPORTANT: some issues cannot be correctly classified because their issue
% date is not present in the Swap table. Then the attribution is performed
% as follows:
% - if the issue is defaulted the spread assigned is the maximum of all the
% other spreads;
% - if the issue has MonthsDelinq ~= 0 the spread assigned is 5/6 of the
% maximum spread.
% Last date available in the Swap table
mindate = datetime(2018,10,1);
old_issues = Issues < mindate;
old_defaulted = old_issues & (Def == Y);
MDnotzero = (MD ~= 0) & (Def == N) & old_issues;
% Multiplying by 100 to obtain percetages
corr_rate = corr_swap*100;
% Computing the spread
spread = Coup - corr_rate;
max_spread = max(spread);
spread(old_defaulted) = max_spread;
spread(MDnotzero) = 5/6 * max_spread;
```

A.7 α-computing function

```
%% Alpha computing function
% Function that computes the interdependence parameter alpha, between
% obligors
% INPUTS
% - U: uniform matrix with dimensions (w, h)
% - w: number of variales of U
% - h: number of obligors
% - method: method to be used. Available selections are "Tau" and "Select"
% - copula: copula function to be used
% OUTPUT
% - alpha: interdependence parameter

function [alpha] = compute_alpha(U,w,h, method, copula)
% Kendall's Tau method
if method == "Tau"

    if copula == "Gumbel"
        % Gumbel's copula parameter w.r.t. Kendall's Tau
        theta = @(x) 1./(1-x);
    elseif copula == "Clayton"
        % Clayton's copula parameter w.r.t. Kendall's Tau
        theta = @(x) 2*x./(1-x);
    end

    alpha = 0;
    for j1 = 1:h
        for j2 = j1+1:h
            tot = 0;
            for i1 = 1:w
                for i2 = i1+1:w
                    tot = tot + sign((U(i1,j1) - U(i2,j1))*(U(i1,j2)-U(i2,j2)));
                end
            end
            tau = (1/nchoosek(w,2))*tot;
            alpha = alpha + theta(tau);
        end
    end
    alpha = (1/nchoosek(h,2))*alpha;
end

% With this method to the user is asked to put as input the desired value
% IMPORTANT: if the value is high (e.g. alpha = 50) computation of prob and
% CVaR might fail due to numerical problems.
if method == "Select"
    alpha = input("Insert the value of the interdependence parameter alpha: \n\n");
end
```

Bibliography

[1] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*. Cambridge university press, 1987.

[2] Fischer Black and Myron Scholes. *The Pricing of Options and Corporate Liabilities*. Journal of Political Economy, 1973.

[3] Hengxin Cui, Ken Seng Tan, and Yang Fan. Portfolio credit risk with Archimedean copulas: asymptotic analysis and efficient simulation. *Annals of Operations Research*, 2022.

[4] Frank J. Fabozzi. *The handbook of fixed income securities*. Mc Graw Hill, 2021.

[5] Kay Giesecke and Lisa Goldberg. *Forecasting Default in the Face of Uncertainty*. Journal of Derivatives, 2004.

[6] Marius Hofert, Martin Mächler, and A. J. McNeil. Estimators for Archimedean copulas in high dimensions. *Journal of Statisical Software*, 2010.

[7] I. Kojadinovic and J. Yan. Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *Journal of Statisical Software*, 2010.

[8] Disposizioni sulla cartolarizzazione dei crediti. `https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1999-04-30;130`.

[9] Stable Distribution by MATLAB. `https://www.mathworks.com/help/stats/stable-distribution.html`.

[10] C. H. Ling. *Representation of associative functions*. Publicationes Mathematicae Debrecen, 1965.

[11] Richard Martin J. The credit spread curve. Fundamental concepts, fitting, par-adjusted spread, and expected return. *ArXiv*, 2022.

[12] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques and tools*. Princeton university press, 2015.

[13] Robert C. Merton. *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*. Journal of Finance, 1974.

[14] Helge Munkel. *Asset-Backed Securities: It's as easy as this! A Practical Factbook*. Global Markets Research, 2006.

[15] Roger B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

[16] Dominic O'Kane and Saurav Sen. *Credit Spreads Explained*. Fixed Income Qantitative Credit Research, 2004.

[17] Wang Xiaoping, Gao Huimin, Zeng Jianchao, and Ye Baolin. Estimation of Distribution Algorithms based on Nested Archimedean Copulas. *Procedia Engineering*, 2011.