RISK MANAGEMENT MAGAZINE

Vol. 18, Issue 3
September – December 2023

In collaboration with Wolters Kluwer

Direttore responsabile (Editor in chief): Maurizio Vallino – Condirettore (Managing Editor): Corrado Meglio
INDEPENDENT JOURNAL WITH NO PUBLIC FUNDING (Law 250/1990)
Poste Italiane - Spedizione in abbonamento postale – 70% aut. DCB / Genova nr. 569 anno 2005

IN THIS ISSUE

Scientific journal recognized by ANVUR and AIDEA

Risk Management Magazine
Volume 18, Issue 3, September - December 2023

Direttore Responsabile (Editor in Chief)
Maurizio Vallino (AIFIRM General Manager; BPER Banca, Italy)

Condirettore (Managing Editor)
Corrado Meglio (AIFIRM Vice President; Banca Credito Popolare, Italy)

Editorial Board
Giampaolo Gabbi - Chief Editor Business Economics Area (SDA Bocconi, Italy); Paolo Giudici - Chief Editor Statistical Economics Area (Università di Pavia, Italy); Daniel Ahelegbey (Università di Pavia, Italy); Raffaella Calabrese (University of Edimburgh, UK); Robert Eccles (Oxford University, UK); Franco Fiordelisi (University of Essex, UK); Pier Giuseppe Giribone (Università di Genova, Italy); Gulia Iori (City, University of London, UK); Richard M. Levich (New York University, USA); Michèle F. Sutter Rüdisser (University of San Gallen, Switzerland); Peter Schwendner (ZHAW Zurich University of Applied Sciences, Switzerland); Alessandra Tanda (Università di Pavia, Italy).

Scientific Committee
Arianna Agosto (Università di Pavia, Italy); Ruggero Bertelli (Università di Siena, Italy); Paola Bongini (Università Milano Bicocca, Italy); Anna Bottasso (Università di Genova, Italy); Marina Brogi (Università La Sapienza di Roma, Italy); Ottavio Caligaris (Università di Genova, Italy); Rosita Cocozza (Università Federico II di Napoli, Italy); Costanza Consolandi (Università di Siena, Italy); Simona Cosma (Università di Bologna, Italy); Paola Ferretti (Università di Venezia, Italy); Andrea Giacomelli (Università di Venezia, Italy); Adele Grassi (Vice Presidente APB, Italy); Mariantonietta Intonti (Università di Bari "Aldo Moro", Italy);Valentina Lagasio (Università La Sapienza di Roma, Italy); Duccio Martelli (Università di Perugia, Italy); Enrico Moretto (Università Milano Bicocca, Italy), Laura Nieri (Università di Genova, Italy); Adamaria Perrotta (UCD – University College Dublin, Ireland), Pasqualina Porretta (Università La Sapienza di Roma, Italy); Anna Grazia Quaranta (Università di Macerata, Italy); Enzo Scannella (Università di Palermo, Italy); Cristiana Schena (Università dell'Insubria, Italy); Giuseppe Torluccio (Università di Bologna, Italy).

Cartoonist: Silvano Gaggero

Ownership, Newsroom and Secretariat:
Associazione Italiana Financial Industry Risk Managers (AIFIRM), Via Sile 18, 20139 Milano

Registration number at Court of Milan n. 629 dated 10th September 2004

ISSN Print 2612-3665 – **ISSN Online** 2724-2153
DOI 10.47473/2016rrm
E-mail: risk.management.magazine@aifirm.it;

Tel. +39 329 138 0475

Printing
Algraphy S.n.c. - Passo Ponte Carrega 62-62r 16141 Genova

The authors bear sole responsibility for the opinions expressed in the articles.

MAILED TO AIFIRM SUBSCRIBERS WHO ARE RESIDENT IN ITALY AND DULY REGISTERED

Journal printed on 29th Decembre 2023

Peer review process on papers presented for publication

The papers that are presented to our magazine for publication are submitted anonymously to a double level of peer review.

The first level is a review of eligibility, implemented on the paper by the members of the Editorial Board, who assess the adequacy of the paper to the typical topics of the magazine.

The second level is a review of suitability for publication, implemented on the paper by two referees, selected within the Editorial Board, the Scientific Committee or externally among academics, scholars, experts on the subject who assess the content and the form.

Editorial regulation

"Risk Management Magazine" is the AIFIRM (Italian Association of Financial Industry Risk Managers) magazine, fully dedicated to risk management topics.

The organization includes the managing editor, a joint manager and an Editorial Board and a Scientific Committee composed by academics.

The magazine promotes the diffusion of all content related to risk management topics, from regulatory aspects, to organizational and technical issues and all articles will be examined with interest through the Scientific Council.

The papers shall be presented in Microsoft Word format, font Times New Roman 10 and shall have between 5.000 and 12.000 words; tables and graphs are welcome.

The bibliography shall be written in APA format and shall accuratel specify the sources.

An Abstract in English is required (less than 200 words) highlighting the Key words.

The authors bear sole responsibility for the opinions expressed in the articles.

The Statement on ethics and on unfair procedures in scientific publications can be found on our website www.aifirm.it.

Processo di referaggio degli articoli proposti per la pubblicazione

Gli articoli che sono proposti alla rivista per la pubblicazione sono sottoposti in forma anonima a due successivi livelli di referaggio.

Il primo livello di referaggio (di ammissibilità) viene effettuato sull'articolo dai membri dell'Editorial Board che ne valutano la congruità ai temi trattati dalla rivista.

Il secondo livello di referaggio (di pubblicabilità) viene effettuato sull'articolo da due referee scelti all'interno dell'Editorial Board, del Consiglio Scientifico o all'esterno tra accademici, ricercatori, esperti della materia, che ne valutano il contenuto e la forma.

Regolamento redazionale

"Risk Management Magazine" è il periodico di AIFIRM (Associazione Italiana Financial Industry Risk Managers) ed interamente dedicato ai temi del risk management.

La sua struttura organizzativa prevede, oltre al direttore responsabile, un condirettore, un Editorial Board e un Consiglio Scientifico formato da accademici.

La rivista favorisce la diffusione di tutti i contenuti afferenti i temi del risk management, dagli aspetti normativi a quelli organizzativi e alle technicalities e vaglierà con interesse, per mezzo del Comitato Scientifico, i contributi che le perverranno.

Gli articoli proposti dovranno pervenire in formato Microsoft Word carattere Times New Roman 10 ed avere un numero di parole compreso tra 5.000 e 12.000; è gradita la presenza di tabelle e grafici.

La bibliografia deve essere redatta in formato APA, specificando accuratamente le fonti.

Si richiede la predisposizione di un Abstract in lingua inglese (meno di 200 parole) con indicazione delle Key words.

Le opinioni espresse negli articoli impegnano unicamente la responsabilità dei rispettivi autori.

Sul sito www.aifirm.it è pubblicata la Dichiarazione sull'etica e sulle pratiche scorrette nella pubblicazione dei lavori scientifici.

Referees Year 2023

Arianna Agosto (University of Pavia), Ruggero Bertelli (University of Siena), Paola Bongini (University of Milan Bicocca), Ottavio Caligaris (University of Genoa), Pier Giuseppe Giribone (University of Genoa), Adele Grassi (Vice President of APB), Valentina Lagasio (Sapienza University), Duccio Martelli (University of Perugia), Enrico Moretto (University of Insubria), Laura Nieri (Università di Genova), Adamaria Perrotta (University College Dublin), Enzo Scannella (University of Palermo), Cristiana Schena (University of Insubria), Alessandra Tanda (University of Pavia),

Artificial Intelligence: new data and new models in credit risk management

Rossella Locatelli (University of Insubria), Giovanni Pepe (KPMG), Fabio Salis (Credit Agricole), Andrea Uselli (University of Insubria)

Corresponding author: Rossella Locatelli (rossella.locatelli@uninsubria.it)

Article submitted to double-blind peer review, received on 2nd October 2023 and accepted on 13th December 2023

Abstract

During the last decade, the increase in computational capacity, the consolidation of new data processing methodologies and the availability of access to new information concerning both individuals and organizations, aided by the widespread internet usage, has increased the development and implementation of artificial intelligence (AI) within companies. The application of AI techniques in the banking sector attracts wide interest as the extraction of information from data is inherent to banks. As matter of fact, for many years now models play a crucial role in several banks processes and are strictly regulated when they drive capital measurement processes. Among banks' risk models a special role is played by credit ones, as they manage the most relevant risk banks face and are often used in regulatory relevant processes. The new AI techniques, coupled with the usage of novel data, mostly unstructured ones related to borrowers' behaviors, allow for an improvement of the accuracy of credit risk models, that so far relied on structured internal and external data.

This paper takes inspiration from the Position Paper Aifirm 33/2022 and its English published translation (Locatelli, Pepe, Salis (eds), 2022.

The paper is focused on literature review regarding the most common AI models in use in credit risk management, also adding a regulatory perspective due to the specific regime banking models are subject when they are used for regulatory purposes.

Furthermore, the exploration of forthcoming challenges and future advancements considers a managerial perspective. It aims to uncover how credit risk managers can leverage the new AI toolbox and novel data to enhance the credit risk models' predictive power, without overlooking the intrinsic problems associated with the interpretability of the results.

Keywords: credit risk models, artificial intelligence, bank management, unconventional data

JEL Code: C00, C40, G21, G32, O30.

1. Introduction

The growing availability of increasingly digitalized data flows is changing the way companies can stand to gain a real competitive edge by harnessing these data and employing sophisticated processing techniques to cut costs and implement decision-making processes that are as data driven and automated as possible, also gaining in efficiency.

The banking and financial service industry is, by its very nature, specialized in generating and interpreting information, and is also characterized by managing financial and non-financial risks, by using a large amount of data that contribute to develop specific and advanced models for risks measurement.

The critical role of these models in risk management processes and the regulatory regime banks are subject to, make risk models the object of a strict supervision that inevitably affects banks' choices.

In this regard, banks must avoid being trapped by "traditional models" as the ability to harness developments in artificial intelligence (AI) constitutes a critical success factor in the competition with new market participants – Fintechs – whose proposition is often built around the usage of AI capabilities.

Hence, an effective deployment of AI for their day-to-day activities will play a special role in keeping banks up to the challenge of the market evolution. Lending and pricing decisions making are a first example, as well as monitoring exposures for risk management and measurement, and performance assessment.

Among others, credit risk continues to deserve a special attention and prioritization in banking management, due the critical role banking credit plays in European financial and economic systems, to support growth and investments. The "key" of a virtuous functioning of the system based on a proper credit assessment of the borrower, the process by which the lender evaluates the creditworthiness of the counterpart by quantifying the risk of loss and by evaluating the "cost" of this risk in terms of cost of lending. Since the development of credit scorings and ratings, credit decision-making has expanded to encompass risk-based pricing, the Know Your Customer (KYC) principle, independent evaluation standards, and internal credit scoring. This process involves multiple entities and relies on extensive datasets sourced from external applications, open banking data, and various third-party sources (Zhang et al., 2014).

The paper aims to conduct a literature review focusing on the prevalent AI models utilized in credit risk management within the banking sector. The paper evaluates the feasibility of implementing these models within current banking practices. It offers examples to understand significant challenges associated with integrating unstructured data. Additionally, it offers a focus on the regulatory perspective, emphasizing the significance of validating models when employed for regulatory purposes.

Section 2 is focused on new data, especially unstructured data, that are included in AI techniques. Sections 3 and 4 summarize the most common AI models and some application in credit risk management: nowadays financial institutions can have access to a huge

amount of new data, for example those coming from digital payments, that can help understanding borrowers' behaviours. To do so banks must upgrade their systems and capabilities to make robust credit decisions in a fast-changing context (Zensar Technologies, 2023).
Section 5 discusses forthcoming challenges and future advancements, to understand the ways credit risk managers can leverage the new AI toolbox and new data in order to enhance the predictive capacity of credit risk models. This examination also addresses the inherent challenges tied to interpreting the results.

2. New data in the context of credit risk models

As mentioned, in addition to the information that banks acquire during their business relationship with customers, intermediaries have traditionally been able to acquire additional information externally (i.e., macroeconomic data, credit reports stemming from the Central Bank credit register and information generated by other credit bureaus). The field of "additional" or "alternative" information has substantially expanded in the recent years, due to the progressive digitalization of the economy leading to businesses and individuals generating vast quantities of information most of which is unstructured and have both financial and non-financial source and content (Altman et al. (2010) for the relevance of soft and non-financial information in credit evaluation of microfirms).
We have several examples of such data. They include online information, news and stories, newspaper articles, and social media posts and comments, which can contribute to define individual's or company's "online reputation". Moreover, other data describe consumer habits and preferences, such as their daily financial transactions: the increasing penetration of debit and credit cards and electronic payments generates a large and new set of information not available up to now.
A major regulatory development, the second European Payment Services Directive (PSD2), then made this information more accessible and thus easier to be used for credit scoring purposes. The new Directive has made information arising from financial transactions much easier to use as, on one hand, it has required financial intermediaries to give third parties access to information that was, until then, exclusively the prerogative of such intermediaries while, on the other, it has made it possible for the same banks to use third-party information when the potential new customers allow to do so. Other data refer to digital footprints, geolocation, utility payments, and the usage of TLC services (Zensar Technologies, 2023).
The whole amount of these alternative data can make credit risk management and borrower's evaluation more effective. The larger the datasets in use, the more complete analysis supporting the credit score: a larger number of diverse data sources is becoming more and more now accessible to financial institutions with the purpose to make credit and financial evaluations and decisions more effective, and ultimately fairer.
Berg et al. (2018) demonstrated that digital footprint variables like email domain and type of device have a better chance than the traditional credit bureau scores at predicting the default of a borrower or late payments in loans' installments and also they argue that a lender that refers to both traditional credit scores and digital footprint of the borrower in its models will be able to make a better credit decision: credit access can increase without compromising the loan quality. The variables used to investigate the digital footprint of the applicants are usually the email domain, the type of device used for the connection, the time at which the contact happened. Taken together, these variables can be useful to investigate certain features of the credit applicants that otherwise would have not been examined (e.g. the type of device is easily associated with applicant's level of income).
In 2022 a study performed by Forrester Consulting on behalf of Experian (Experian, 2022) has highlighted the need to focus on leveraging alternative data and emerging technologies to improve the credit decisioning processes of an organization. More in detail, the surveyed lenders consider that data augmentation remains a top priority for credit risk assessment: 77% of lenders consider that the high or critical priority to leveraging open banking to access real-time data, while the percentage is about 75% both for leveraging new data sources (alternative data) and emerging technologies (AI).
In this context of greater availability of "alternative" data, the Covid-19 pandemic, increased the overall uncertainty regarding the macroeconomic conditions, reinforcing the need for banks expand their traditional datasets to include information updated in real time, or gathered from less consolidated sources. Such measures have been necessary to compensate for the loss of sensitivity suffered by "traditional" rating instruments during the pandemic. This extreme event impacted the "traditional" information used in credit rating systems (EBA, 2020a), as loan moratoria makes it more difficult to assess the evolution of a borrower's financial position and thus its ability to repay, crippling the rating systems' aptitude to correctly gauge the customers' riskiness, risk that could materialize once the various relief measured were lifted. This context therefore reinforced the need for intermediaries to expand their datasets and deepen their analyses of the data they already had, in order to identify, far enough in advance, any "zombie" borrowers, i.e. obligors surviving solely thanks to the relief measures and for which the banks needed to determine the most appropriate steps to take in order to mitigate credit risks[1].
In light of the above, the modelling of the counterparties' creditworthiness can therefore take advantage of the usage of alternative data in multiple manners.
The banks use financial transactional data to integrate and refresh company's financial statement. This approach prevents reliance solely on outdated financial information, which tends to lose its relevance swiftly during periods of regime change.

[1] Moreover, the priorities of the ECB Banking Supervision of 2021 include: "[…] banks' capacity to identify any deterioration in asset quality at an early stage and make timely and adequate provisions accordingly, but also on their capacity to continue taking the necessary actions to appropriately manage loan arrears and non-performing loans.", as reported at www.bankingsupervision.europa.eu.

Moreover, by combining bank performance data (i.e., internal, and external performance data) with information released by credit bureaus and "new information" has made it possible to raise the degree of the models' precision, thereby improving their contribution to pinpointing opportunities for development, as well as in the measurement of risks.
The availability of "high frequency" data, accessible in real time or nearly so, enables the construction of models that exhibit greater sensitivity than traditional ones to changes in borrower characteristics. This attribute makes them more forward-looking in nature.
Moreover, incorporation of alternative data can enrich the credit risk models: banks and financial institutions can increase the quality of the loans portfolio, in terms of lower probabilities of default, a higher efficiency in credit decisions leading to a reduction of non-performing loans and default rates. Recent research suggests that AI models can enhance the performance from 10% to 15%, leading to a significative decrease in credit losses from 40% to 20%. Additionally, these models can potentially contribute to a reduction of about 25% in the exposure towards risky customers (McKinsey, 2021b). Furthermore, advanced models can drive revenue growth by enhancing the customer experience (cross-selling), and even lead to higher operating margins by increasing operational efficiency, by easily automating or reengineering some processes thanks to the adoption of new platforms based on AI/ML tools.
The best practice for the optimal use of one's data in the development of cutting-edge credit risk management models lies in the ability to integrate various type of information available on the counterparties evaluated throughout multiple business processes: "structured" or "traditional data" meet a set of predetermined rules in terms of types; "semi-structured data", which contain semantic tags without having the typical structure associated with relational databased; lastly, "unstructured data" do not have a predefined model and cannot be organized into rows and columns (e.g., images, audio and video recordings, e-mails, spreadsheets).
When implementing AI models for credit risk management purposes, the primary challenge lies in handling unstructured data, since even in today's digital world most information available is unstructured. Different structuring techniques can be employed in the analysis of unstructured data: they mainly refer to text analysis and Natural Language Processing (NLP). These techniques utilize algorithms to extract natural language rules, rendering then the unstructured information comprehensible to computers.
The most important procedures include, but are not limited to, topic modelling, where tags are created for specific topics using key words through, for example, Latent Dirichlet Allocation (LDA) which identifies the key words to extract; part-of-speech-tagging, which entails recognizing the type of entity extracted (e.g., an individual, a company, a place), by means of named-entity recognition; sentiment analysis, employed to identify the sentiment, opinion, or conviction of a statement, from very negative to neutral to very positive.
A specific domain where these techniques are gaining notable relevance is the one of banks' transactional data. Transactional data have been used in a wide variety of applications in banking analytics since many years, but their use in the past has been limited to constructing simple aggregated features of structured information, such as calculating the average spending over recent months or determining the total income for the past year, etc. This approach has clear limitations in fully harnessing the potential value of transactional data. This is particularly relevant in the case of unexpected events, where the classical approach could fail to rapidly capture changes in individuals' behavior. For example, amid the Covid-19 pandemic transactional data were extremely useful to track the crisis, and extract insight over the change in mobility and expenditure across different income classes.
One of the steps needed to grasp the powerful information content of transactional data is the categorization, which involves the usage of Natural Language Processing tools allocating transactions to a commonsense "categories" based on the words found in the text elements. For example, once all transactions are gathered in a single repository, after a possible data integration step, those mentioning car-related expenses would be all allocated under the "car" category. To this aim the NPL algorithm will look for different words or acronyms referencing the concept of a car (in Italian "macchina", "autovettura", "assicurazione auto", "meccanico", etc).
A crucial step in the treatment of unstructured data consists in the cleansing of the latter. The main activities in the preparation and cleansing of a text include: the removal of special characters; the standardization of text; removal of "stop words" (e.g., articles, conjunctions, prepositions, commonly used verbs); removal of non-significant words or numbers; and stemming or lemmatization, which reduce an inflected word to its root form (e.g., "go", "went", and "going" would be mapped towards their stem "go").
As the amount of available data increases and the data become more complex, we must keep in mind that the relationship among the different explanatory variables and the target one is not linear. Hence, to effectively capture such increased complexity, the utilization of machine learning (ML) techniques becomes critical.

3. New models for credit risk management

Sometimes there is a sort of "overlap" between Machine Learning (ML) and Artificial Intelligence" (AI). While they are interconnected, AI encompasses a broader scope that involves the creation of IT systems aimed at executing tasks without human intervention. ML, on the other hand, is a subset of AI, that involves specific techniques allowing systems to learn and improve from data without being explicitly programmed.
AI encompasses various methods beyond ML, including expert systems, natural language processing, computer vision, and robotics, among others. These techniques enable AI systems to mimic cognitive functions, learn from experiences, and adapt to new information, extending beyond the specific algorithms used in ML.
Literature analyzed several techniques and tools. An excellent and complete analysis is provided by Breeden (2020), who focused on ML methods as the result of a combination of data structure, architecture, estimator, selection, or ensemble process. More in detail, among these key components, "architecture" represents the step in which the difference between traditional statistical methods (regression approaches, state transition models) and ML (convolutional networks, feed-forward networks, and recurrent neural

networks) comes to evidence. Moreover, according to Breeden (2020), ensemble modelling seems to have a relevant added value in credit risk, because of the typically limited datasets available, as it can combine the forecasts from different model types as they can capture different aspects of the data and they contribute to increase the effectiveness of the model. Ensemble models can be divided into homogenous methods, that combine multiple models of the same type, and heterogenous models where the "mix" aggregates any type of models. Among the first ones, bootstrapping aggregation (bagging), decision trees and random forests are probably the most well-known techniques, as well as boosting, i.e., a process of building subsequent models on the residuals of previous models (Breeden, 2020; Schapire, 2003). Nalić et al. (2020) analyzed the importance of hybridization and ensemble models, increasing the performance of ML algorithms.

Figure 1 contains a graphic representation of the different ways in which the two techniques work. Boosting considers the sequential estimation of the multiple decision trees, whereas Bagging, also known as Random Forests, accounts for multiple decision trees in a parallel fashion. It is worth mentioning that both ensemble techniques can be generalized for different underlying models.

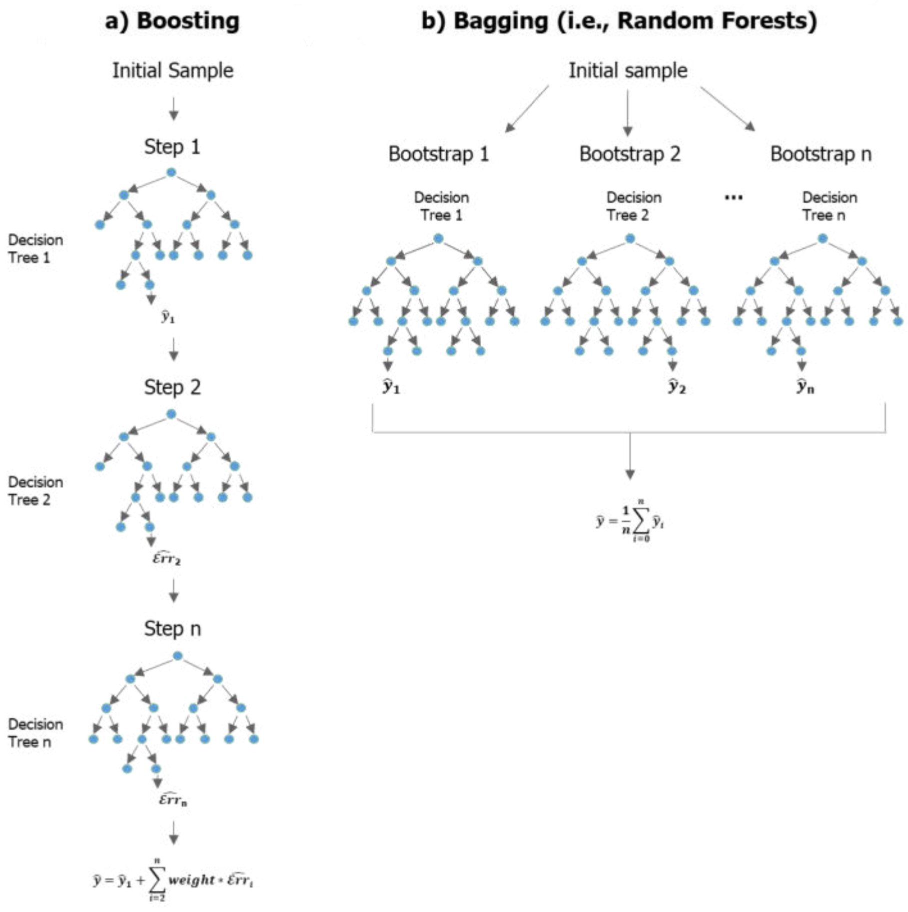


Figure 1: Two different model ensemble techniques. Boosting and Bagging

According to Breeden (2020), several applications of ML to credit risk are in use, following different specific purposes and employing specific explanatory variables. Notably, in credit scoring, decision trees, neural networks, and nearest neighbors worth a mention. More in detail, decision trees are based on an algorithm, whose logic can be interpreted and translated into "if-then" scenario, that follows the structure of a tree comprised of a series of nodes and branches. The nodes represent a macro-class of input variables, in which each node corresponds to a specific linear test function, which establishes the appropriate partitions of the input variables. The branches or, to be more precise, the arches, represent all the properties, or the splitting rules, which determine the path within the tree and, finally, the classifications. These properties/rules are defined in relation to the specific values assumed by the attribute identifying the parent node. Decision trees are very efficient when working with categorical variables, where the classification is expressed through explicit conditions, the algorithm itself determines the significant variables, and the presence of outliers does not excessively affect the result (typically the outlier data will be isolated in peripheral nodes).

Additionally, it's worth noting that random forest is a specific method based on decision trees.

This technique involves replicating the tree's estimation numerous times (often exceeding 1,000 iterations), utilizing only a single subset of the available variables in each iteration. This classification is based on a regression and classification algorithm that creates a forest of decision trees constructed out of multiple datasets, extracted through bootstrapping (random sampling that organizes the various nodes and divides them randomly). The more trees there are in the forest, the better the results of the model will be. It is important for there to be a low correlation between the models entered: in this way, each decision tree created takes independent decisions and thus reduces individual errors, constituting a protective barrier between one tree and another. Certain trees may generate distorted results, whereas others will generate correct results, leading the model in the right direction.

The algorithm named "Nearest neighbors" has the purpose to classify each observation by analyzing the "*k*" individual closest to it. These models learn from past examples (for example past lending experiences) to collect data and info for a "comparable" loan. The most challenging point is about the identification of "effective" comparable as the success of these methods is related to the uniformity of the distribution of the dataset. If data are scarce or extremely sparse, the reference to past experiences may add value in the quality of the analysis, overcoming interpolation or extrapolation.

Neural networks are often considered the last frontier in AI techniques applied to data modelling. These methodologies are often thought of as a "black box" in terms of the association between the input parameters and the output classification. Neural networks are based on a model that puts the explanatory variables in communication with the target variables through various layers of latent variables, referred to as "hidden layers", consisting of combinations of transformed input variables. A neural network is indeed an "adaptive" system capable of modifying its own structure (the nodes and interconnections) based on both external data and internal information that interconnect and move through the neural network during the learning and reasoning phase. Artificial neural networks are non-linear structures of statistical data organized as modelling tools. They receive external signals on a layer of nodes (i.e., the processor) and each of these "entry nodes" is connected to various internal nodes in the network which are typically organized into several levels so that each individual node can process the signals received, transmitting to the next levels the result of its processing (i.e., more sophisticated, detailed information). It is the hidden layers' task to use the features of the dataset to learn new features. In such cases, the network will use the features it has learned in a hidden layer to learn additional new features that are even more significant in the next layer, and so on until they are used for classification, regression, or clustering in the output layer. One of the most important positive elements of neural networks is that they work extremely well for both classification and regression in deterministic problems.

Furthermore, the classifications that they produce are very robust for noisy data and they are also capable of exploiting relationships between features that are difficult to identify using other modelling approaches. A neural network model may be highly accurate, reflecting the treatment of non-linear data (Giudici et al., 2023).

Neural networks are one of the most extensively tested methods for credit scoring, probably because they are flexible as they can be combined with other algorithms to manage hybrid ensembles.

A peculiar type of neural network is the autoencoder, which is developed to generate new data by compressing the input in a latent space, reconstructing then the output based on the information gathered. The great advantage brought by this, is that it identified observations that would otherwise lead to large estimation errors. In conclusion, tree-based models are extensively employed in the treatment of transactional data, often obtaining better performances with respect to their neural networks counterparts whose capability increases proportionally with respect to the extent of the underlying dataset. Moreover, tree-based models allow for a more manageable training than neural networks, being easier and less time consuming to train. However, it is worth noting that even though tree-based models can be deemed more directly interpretable than neural networks, the latter are considered more appropriate in the treatment of transactional data given its highly time-dependent nature.

4. Application of AI tools in credit risk management

As mentioned, within the lending life cycle, banks are relying increasingly on AI and analytics capabilities to add value in credit decisioning, as well as in other areas of their business (McKinsey, 2021a).
However, the primary challenge lies not only in developing methodologically and technologically sophisticated models but foremost in building them so that they produce results that can be more or less easily explained to the various users of the outputs of the credit risk management models in the various fields where they are used[2]. These areas include business strategies business strategies, where the models' output must be explained both to the customers and to the credit managers; the quantification of credit losses on loans within the portfolio; and determining parameters used in computing capital requirements, as banking regulators investigate carefully the process followed by banks to compute them[3].
As highlighted in its recent survey, the Institute for International Finance (IIF, 2019) confirmed the fact that one of the main areas of application of ML techniques is credit scoring. Specifically, the survey shows that while the use of these models for supervisory reporting purposes is somewhat limited by the need to implement simple and easily interpretable models, these limitations apply to a lesser extent when the same techniques are used for management purposes. One important aspect that concerns both the "supervisory

[2] Institute of International Finance (IIF) Briefing Note (June 2021) - Explainability is a trust issue. Regulators are concerned with levels of complexity that offer limited or no insight into how the model works (i.e., so called black boxes). They want to know how models, including AI and ML, reach decisions on extending or denying credit, whether FIs have appropriate risk controls in place, and the like. A challenge of using AI/ML models is often the lack of transparency, which is imperative to building trust with customers and other stakeholders. Banks have long had to explain their decision processes to improve confidence in the robustness of a model.

[3] See CRR 575/2013, paragraph n. 179: "The estimates shall be plausible and intuitive".

reporting" and "management reporting" applications, is the integration of the traditional data used for credit risk management (e.g., credit bureaus) with other innovative data, mainly based on transactions and "point-in-time" data and information, that can improve models' accuracy.

There are many potential applications for ML techniques applied to innovative, high-frequency data.

First, they may be used in advanced early warning systems (EWS), for a prompt identification of borrowers whose behavior "tends" to reflect a potential risk, with respect to a target event which is usually a precursor of a proper default classification (i.e., 30-days delinquency, classification to IFRS9 Stage 2). While traditional systems usually require indicators based on expert opinions, ML techniques are well suited to handling large quantities of high-frequency data/updates and hence make it possible to generate effective and efficient models, with the forward-looking feature needed for EWS.

Secondly, ML techniques make it possible to capture further in advance a borrower's likelihood to transition to migrate to another stage as per the IFRS staging allocation rules.

Moreover, the competitive advantage gained from the in-depth analysis of data and high-speed processing and decision making comes from the ability to gather core information (e.g., income, consumption, etc.) from non-traditional data sources and to extract patterns that are relevant for the purposes of monitoring credit performance over a medium-term horizon, more extended the one normally in use in risk monitoring process.

Different examples of applications suggest that the common underlying idea is the purpose of combining algorithmic intelligence, to handle large quantities of data and qualitative information by saving time and costs and increasing effectiveness, with human intelligence. This combination aims to optimize credit decisions and making the credit risk management more comprehensive, fulfilling both the compliance and operating goals.

Rhzioual Berrada et al. (2022) produced a brief literature review aiming to find how AI could get benefit to the bank industry in credit risk assessment and default prediction. The authors highlight that AI methods in use for commercial banks constitute a large set of applications, as the banking industry is one of the more advanced fields using AI and machine learning. According to their analysis, feature and data preprocessing are the most important factors to be managed with the purpose to optimize algorithms and quality of the results, in terms of accuracy and precision.

Theuri and Olukuru (2022) provide a literature review regarding the application of AI in banking: they highlight that the bulk of the research is focused on credit scoring techniques, especially in consumer credit as its large growth, with particular regard to classification and the proper application of algorithms. Credit risk assessment has also been experimented with using ML. For instance, Bussmann et al. (2021) applied AI model to peer-to-peer lending platforms, with the purpose to estimate borrowers' future behaviour. Wang et al. (2020) focused the comparative assessment of several classifiers involved in ML techniques for credit scoring.

Global financial crisis, COVID pandemic, credit crunch, climate and (specifically transition) risks and other major risky conditions and scenarios have, more in general, enhanced the attention to credit risk assessment and to predicting default risk. In particular, Moscatelli et al. (2020), via the application of ML ratings, highlighted the benefits for a lender, able to offer credit to safer and larger borrowers, so that it is possible to observe a decrease in its credit losses. Estimating ML ratings involves using machine learning algorithms to analyze historical data and generate credit ratings or scores applied to borrowers. By following different steps, the process includes data collection, the selection of most relevant variables and the application of the algorithm to test models. Once the models are validated, they are integrated into the lenders' credit risk management system. This integration enables the assessment of new loan applicants or existing borrowers, assigning them credit scores or ratings based on their financial history and other relevant data.

Another portion of research is focused on alternative data sources (Roeder 2021 on public and private companies; Monk et al. 2019), to be used in addition to established credit risk indicators for supporting credit risk management. As underlined, these data sources contain important signals to identify changes in borrowers' behaviours and in their financial conditions. An example may be provided by the previously mentioned transactional data, that are particularly apt to recognize transaction patterns usually lead to insolvency.

The literature review presented by Roeder (2021) highlights that, on the one side, as expected, contributions focused on ML algorithms normally use more complex models (e.g., random forests or neural networks) to improve the quality and the accuracy of predictions, and, on the other, most of the papers suggest that variables related to alternative data provide explanatory value or improve the prediction power. There is space, also in other domains, for alternative data sources, for example in the analysis of credit worthiness of SMEs, where the datasets of information are normally smaller, especially with regards to market-based metrics.

According to the diversification in methods, we can highlight several ML and AI applications in the broad area of credit risk in banking, into different operational contexts, even if some cautions are required before a large adoption (Breeden 2020).

In terms of applications, first of all – as mentioned – ML are applied for credit scoring purposes, going ahead of the traditional statistical methods (regression analysis, discriminant analysis). Pérez-Martín et al. (2018) highlight that the increase in financial operations has led to an increase in the size of datasets: the huge volume of data and information requires new data techniques and their simulation experiments – based on home equity loans and both on traditional statistical methods, parametric and non-parametric methods and AI tools – suggest that the effectiveness of a method and, in broader terms, the final judgment about its acceptance or not is also a consequence of the major purpose of the analysis (the minimization of the mean square error, the best computational efficiency, the best response time, etc.).

Modeling corporate defaults and bankruptcies is a similar problem, even if a lower amount of data and less standardized inputs. Barboza et al. (2017) tested ML models to bankruptcy prediction and concluded that in terms of accuracy ratios these methods are

better than the traditional models and, more recently, new studies applied similar techniques also to general default events, not only to bankruptcy. Nevertheless, ML models are not perfect, because the risk of misclassification, the more computational process time required and also the fact that extensive datasets are not common. Moreover, the author points out that additional research is needed with regards the use of macroeconomic variables as explanatory variables in ML models, by combining these external data with firm-specific data and information.
Van Thiel and van Raaij (2019) tested experiments about AI credit risk models for probability of default in consumer lending: their findings put in evidence as AI techniques performed better than traditional models. Default prediction improves in accuracy if lenders can access to effective and/or estimated income, spending data and create relations between economic/financial data and social media data.

4.1 The integration between traditional and AI models

The increase in the availability of data, new and sophisticated analytical techniques, and customers' rising expectations of a complete digital experience are key factors, that are increasingly crucial to market success of nowadays financial firms. The models adopted by banks to estimate the probability of default of their counterparties are also evolving to make most of these opportunities, by exploiting internal and proprietary data sources to the bank not yet used, new data sources including those outside the bank, and new ML algorithms.
As mentioned, innovative data, such as current account transaction or data extracted from social networks, are increasingly included in Credit Risk models through one or more specific ML modules operating along with "traditional" modules which generally cover established information areas (e.g., socio-demographic, financial, banking behavior, qualitative assessment). Such integration of traditional modules with ML modules led to the development of high-performance models capable of rapidly capturing phenomena and signals that traditional models find more difficult to intercept. Crosato et al. (2021), for instance, analyzed the use of online indicators from firms' websites to predict firms' default and borrowers' creditworthiness in the banking sector.
Some brief experiences, in which the traditional module has been integrated by an ML module employing alternative techniques and data, worth a mention.
The first example consists in the implementation of a transactional module capable of discriminating the riskiness of companies by looking at trends in the payment of salaries, correctly penalizing the score of those companies having irregular salary payments, whereas positively score those companies having seasonal business activities – e.g. agricultural enterprises - hence seasonal salary payments. Another practical case concerns the identification of counterparty risk during the Covid-19 crisis, in which the transactional score proved to be extremely reactive in identifying changes in the riskiness of borrowers: in a real-life case the transactional score remained stable during 2019, rapidly grew after February 2020, when the social restrictions were introduced and began falling after June 2020 when the restrictions were eased. A further practical application for a transactional module could be the implementation of an early warning system due to the long prediction horizon of transactional information, capable also of detecting signs of credit quality deterioration that had not occurred before. Another practical case involves employing a random forest algorithm to detect default within 12 months of loan origination. Such methodology yields more predictive results compared to the traditional logit model, despite possible problems in terms of interpretability of the results.
As already mentioned, stand-alone AI models may be used to supplement more traditional modelling techniques. Han et al. (2013) provided an example of this integration, in the case of employing traditional models to select the relevant drivers, which may then be reprocessed and combined in the final model using advanced algorithms. In this way, the traditional models are the first "filter" that excludes variables that are highly correlated or not meaningful from a statistics or credit perspective, whereas alternative methods are used to develop the model. Another way of using two techniques jointly could be the creation of a function integrating the output of a traditional model with the output of one or more ML models. With this approach, it is assumed that various algorithms can more accurately model the different aspects or sub-areas of the phenomenon in question and that their interaction might bring out the "best" of the various approaches.
A third possibility consists of integrating traditional and innovative models with an expert, rather than statistical, approach. For example, this could happen using a notching matrix, adjustments or overrides of the outputs of the traditional model, particular in those sub-areas where the predictive power is weakest. At the same time, there may be instances in which traditional models and alternative models lead to convergent predictions and other instances in which the predictions are divergent or inconclusive, enabling the analyst to focus the expert valuation on the later.
Moreover, AI models can be used to correct prediction errors committed with traditional models.
Integrating traditional models with alternative AI-based methodologies in one of the ways described above may offer a series of advantages over using a stand-alone AI model. First, this choice may be a gradual transition towards adopting more advanced methodologies, combining modelling techniques that are now consolidated in market practice with more innovative solutions, thereby preserving the interpretability of the model's outputs, especially for internal users of rating models like loan managers and credit analysts. The integration approach might also be the easiest solution in the medium term for banks that use internal models not only for management reporting purposes but also for supervisory reporting since, in this case, the models must be formally approved by the regulator. To this end, combining traditional models with AI models can guarantee more comparable results or make it possible to quantify and compare the extra performance of the AI modules with the model's overall discriminative capacity. Similarly, this approach leaves a certain degree of flexibility in its design, in that it does not preclude the possibility of "dismantling" the AI

component to apply, if necessary for the individual use cases, only traditional models. However, the decision to integrate these types of models may give rise to unknown variables. Certainly, this choice implies the need to estimate several models (both traditional and alternative) and entails the identification of the most appropriate integration methodologies, requiring inevitable extra efforts for risk management units, not to mention the potential complexities of the deployment phase when they are taken from the test environment and implemented in the bank's legacy systems.
Finally, the decision to apply an integration approach may, in certain cases, reflect a compromise that, for extremely non-linear phenomena that need to be modelled, could lead to a dull performance compared to a purely AI-based model.

4.2 Use of AI models for the validation of traditional models

Going beyond the integration, another relevant point concerns the use of AI models for validation and benchmarking of traditional models.
The banking regulation on one hand, and the supervisory practice on the other one, concur to set quite high standards to the validation practice banks have to adopt, especially for the most sensitive models, i.e. those used for computing capital charges or for setting the loan loss provisions. In these cases, validation teams must challenge internal models before they are put in production and afterwards, along their entire life cycle.
Such a regime calls for continuous advancement of the validation practice even making use of ML techniques to develop challenger models capable to single out the weaknesses of the more traditional ones.
In this regard examples exist of cases where ML models have been used to validate IRB models, i.e. models used for capital computation purposes, once authorized by the competent banking supervisory authority (Aifirm, 2022; Locatelli, Pepe, Salis (eds.) 2022). A first example pertains to the usage of ML techniques to identify any relevant bias in the ranking yielded by a corporate rating model. More in detail, the analysis was conducted by comparing the estimates resulting from the concerned model, based on traditional statistical methods (i.e. logistic regression), and those produced by ML approaches (decision trees, random forests, neural networks), leaving unchanged both the training set data (including the time horizon) and the explanatory variables. The results were assessed using established materiality thresholds of divergence and no significant differences arose, suggesting a negligible model risk.
A second example refers to a retail rating model, again developed using the traditional logistic approach. To this aim the validation team developed an alternative ML model, combining modules developed with traditional statistical techniques with other realized using ML techniques. In the latter case specific methodologies were employed to select and handle the explanatory variables, consistently with the ML techniques. It emerged a challenger rating model whose performances have been compared to those of the main model in-sample and out-of-sample/out-of-time. The comparison showed the accuracy of the alternative model, which was submitted to the regulatory validation in lieu of the original one.
ML models are also used for validating models used for management reporting purposes; even in this case different case studies can be briefly summarized. For example, the application of scoring models developed with ML techniques to challenge expert credit scoring and, more in detail, to measure the customer's riskiness over a short-term horizon (e.g., six months). The model was developed to provide a risk assessment of borrowers during Covid-19 period and is distinctly point-in-time. This entailed the use of a long list of variables fed by multiple sources, such as financial statements, transactional data, credit bureau data, industry data, etc. The model was employed to challenge the specific assessments provided by the bank's credit managers, assessing the borrowers' resilience after the impact of the pandemic.
Another example concerns the development of a ML model used to challenge the PD retail rating model used for computing the loan loss provisions according to the IFRS9 accounting principle. To this aim an artificial neural network model (of the deep learning variety) was developed to challenge the in-production rating model, estimated using a logistic regression. Despite utilizing the same set of variables for training both models, the neural network required a pre-processing phase to normalize the categorical variables.
The time horizon on which the two models have been trained was different, as in the case of the challenger model more recent data have been used. Therefore, the two models have been put in production to assess the concerned portfolio at the same date. For the purposes of the comparison, standard performance indicators (Cap Curve and Accuracy Ratio, Roc Curve and AUROC (AUC), Confusion matrix) have been used. It emerged that the in-production model shows performances just below those of the challenger model.
The negligible difference between the performance of the two models confirmed that the in-production model's got good predictive power and accuracy and, therefore, was suitable to be used to compute the expected credit losses in accordance with IFRS 9.
AI techniques used for validation, testing and benchmarking purposes have now become general operating practice to the same degree as ordinary validation techniques with respect to managerial models and boast excellent performance even though they are still continuously evolving. They are, however, less frequently used for models aimed for computing capital charges due to the greater regulatory constraints, although in recent years the regulator has incentivized their use to make the validation practices more challenging.

5. Evolution and new challenges in AI models

5.1 The financial services regulatory framework: some open questions

The regulatory landscape for credit risk and decisioning is becoming increasingly more complex. There are a huge set of laws and regulations that financial institutions are required to respect. First, we refer to specific bank regulation, for example Capital adequacy

requirements (since Basel II), but also credit provisioning and accounting (IFRS 9 and 17). Second, privacy rules (GDPR) and anti-discrimination acts are relevant (e.g. EU Artificial Intelligence Act and Consumer Credit Protection Act).
As highlighted previously, over the past decade, the use of alternative models and data have disrupted the existing credit risk processes within banks and lending institutions. Despite the evident benefits and enhancements, these advancements also bring forth several challenges, introducing new risks and potential drawbacks. As a consequence, there is an increasing scrutiny around these tools, leading to increased attention and monitoring from regulators. The framework is not homogeneous: regulatory requirements is largely still under development.
In November 2021, the European Banking Authority (EBA) released a discussion paper with the purpose to understand the challenges and opportunities coming from machine learning (ML) tools to be applied in the context of internal ratings-based (IRB) models to calculate regulatory capital for credit risk.
EBA recognizes that ML models are more complex than traditional modeling techniques such as regression analysis or simple decision trees, and often they lack in transparency, but they can count on large data availability and increased storing capacity. It's worthy to mention that some issues, such as data usage and explainability, are not entirely novel in managing IRB models, but advanced techniques may lead to new challenges or even to some "black box" models that may create difficulties in interpreting the results and in the understanding by the management functions (EBA, 2021). For these reasons, EBA provided a set of recommendations to favour an appropriate use of ML/AI techniques, the coexistence with other models and to contribute to improve the harmonization of rules about capital requirements across Europe.
According to the IIF 2019 report, ML finds its most common application in credit risk within the realms of credit decisions/pricing. However, its use in IRB models is more restricted, primarily serving as a complement to standard models. For example, ML techniques can provide "model validation", data quality analysis, selection and construction of explanatory variables. At the same time, it is worthy to highlight the trade-off between predictive power and interpretability. This trade-off also explain why the adoption of ML tools for managerial decisions in credit risk management is more straightforward *vis à vis* their application in areas subject to the regulatory scrutiny. This is due to the strict regime supervisors have imposed on these models, emphasizing the need for both predictive accuracy and comprehensibility.
EBA also recognizes that the use of ML techniques within the IRB framework might add value, if some recommendations about adequate monitoring, validation and explainability are ensured. First of all, all the stakeholders involved should have an appropriate level of knowledge of the model's functioning; financial institutions have to avoid unnecessary complexity in the modelling approach. Furthermore, the model has to be correctly interpreted and understood, especially in the case a human judgement is required, and regular updates need to be analyzed and justified, even though the framework for credit risk is typically stable.
A specific issue is posed with regard to "representativeness" and "data quality" issues, because of the direct relationship between models' performances amount and quality/representativeness of the training data. According to EBA's guidelines, financial institutions, must ensure sufficient data quality, especially in the case external and/or unstructured data are included in the models, with the purpose to guarantee accuracy, completeness and appropriateness of the data.
The European Banking Federation (EBF) on September 2021 released a position paper on the EC proposal for a regulation laying down harmonized rules on artificial intelligence (AI Act).
Among other comments, EBF poses special attention on the importance of insuring consistency in supervisory expectations and practices among different national competent authorities. A "level playing field" must be ensured: in the case the same AI application is used by different entities thar are supervised by different Authorities using misaligned approaches, this principle would be broken. "Same activity, same risks, same rules" principle must be guaranteed.

5.2 Explainability and interpretability of AI algorithms

Machine learning and AI present some unique challenges to application in credit risk and the growth in terms of "quality" of the methodologies and innovative applications has very been very significant in recent years.
The use of AI models generally produces performance results expressed, for example, in terms of accuracy, that tend to surpass those observed for classic models based on parametric estimates (e.g., logit/probit).
However, the accuracy can also cause greater methodological complexity, which asks higher attention on the interpretability and usability of AI models by the various stakeholders. This concern can be relevant both for "management" purposes, such as employing the model to support decision-making, and for supervisory reporting. Interpretability, in the context of decision-making processes highlights an important connection with the concept of robustness or stability, the ability of retaining good predictions also in case of unexpected situations.
Interpretability and stability are generally deemed crucial features associated with a model in the context of its use in decision-making processes, because while interpretability relates to the use of models in predictable conditions, stability provides information on how the model will behave in unexpected situations.
Accuracy and quality data are also crucial, but not sufficient. The guidelines established in the Fair Credit Reporting Act, concerning consumer protection and his privacy and security set that lenders cannot discriminate against protected classes and that consumers receive explanations in the case of denial of credit. So, the question is about the ability of ML techniques to avoid these discriminations, and other similar concerns in model risk management must be solved before a large adoption of these techniques (Breeden 2020).

This specific issue can be also linked to the “explainability”: the result of the credit worthiness’ analysis has to be understandable by the consumer, so that in the future he can make improvements in his financial situation to be “eligible” in the case of a future credit application.
One of the main objectives set forth in the regulatory framework on AI proposed on 21 April 2021 by the European Commission focuses on resolving and lessening problems related to the governance of risks deriving from AI applications. The proposal aimed to implement the development of an ecosystem of trust by proposing a legal framework for trustworthy AI, to increase confidence in AI-based solutions that can improve human well-being (European Commission 2021). As the presence of a number of risks, a series of indicators could be required to comply with specific trustworthy principles like accuracy, robustness, fairness, efficiency and explainability and making AI more trustworthy and, more in detail, precise, stable inclusive, efficient, and explainable (European Commission, 2021; Locatelli et al., 2022).
Based on the new requirements highlighted in the European Commission’s proposed regulatory framework on AI, current and future research will focus increasingly on the operating effectiveness of key principles for trustworthy.
Among others, a special focus on explainability or interpretability deserves a mention. As remarked by Hamon et al. (2020), AI is becoming a key technology in automated decision-making systems based on personal data, with a potential and significant impact on the fundamental rights of individuals. In this context, the General Data Protection Regulation (GDPR), in force since 2018, has introduced a set of rights that relate to the explainability of AI and the right of any individual to ask for an explanation of the decisions taken on her application.
As already mentioned, the use of AI models generally yields more accurate performance outcomes compared to traditional statistical models reliant on parametric estimates, such as logit/probit models.
In 2018, the European Commission published a document containing ethics guidelines for AI, prepared by a group of AI experts (European Commission, 2018). In this document, the European Commission described the explainability and interpretability of AI algorithms as fundamental requirements for trustworthy AI.
In June 2023, European Parliament adopted the final text and amendments on the proposal for harmonized rules on artificial intelligence (so called “Artificial Intelligence Act”). According to the final text, AI systems used to evaluate the credit score or creditworthiness of individual are required to be classified as “high-risk AI systems”, because the impact of these decisions on his/her financial resources, spending decisions and access to essential services (European Parliament, 2023). The proposal also highlights the importance of the risk of discrimination as a possible but unintentional “output” of the AI systems towards vulnerable individuals or groups.
More developments in research and practice are expected to tackle this risk, as the disparate impact can also be a subtle phenomenon, both in avoiding and in detecting it (Skanderson, 2021). The consequences of the applications of AI tools in the financial industry are still largely unknown and difficult to foresee. The progress in European regulation aiming to establish a uniform legal framework ant to regulate AI systems in accordance with the values and fundamental rights of the European Union is a crucial point (Visco, 2023).
Also, the European Central Bank (ECB) has demonstrated a clear awareness of the usefulness of AI approaches in credit risk management and it encouraged banks to use all the available data – especially unstructured data – to appreciate borrowers’ creditworthiness, also by including forward-looking elements and instant data in these analyses. This development is a required course of action for credit risk analysis but, at the same time, an open challenge for the next future is about the way by which the AI-based models that use unstructured data will replace traditional approaches and information or if they will supplement them, making them more predictive and responsive to context changes.
Since, most likely, the solution will vary depending on the purpose of the credit analyses, one clue might be the role that instant economics analyses are filling with respect to economic predictions. In this regard, a perception is emerging that while the approaches based on high-frequency data are extremely useful in rapidly capturing turnarounds in the economic cycle, they meet with more difficulties when they are used to precisely quantify the level of activity. In this field, traditional techniques maintain a competitive edge because of the higher quality of official statistics.
Similarly, it is plausible that traditional models are likely to maintain a place in credit risk analysis, as AI-based models progressively support them in fields where they are superior in connection with the type of information to be processed and the speed with which they make use of this information. A key element to bear in mind in this structural credit risk modelling change relates to all cases in which AI techniques are applied to aspects impacting human life. Preserving ethics and ensuring transparency remains a pivotal consideration across all AI applications.

References

Aifirm, Artificial Intelligence e credit risk. Possibili utilizzi di metodologie e dati alternativi nei sistemi interni di rating [it], Position Paper n. 33, 2022

Altman, E.I., Sabato, G., & Wilson, N., The value of non-financial information in small and medium-sized enterprise risk management, J. Credit Risk 6 (2), 2010

Anderson R.A., Credit Intelligence and Modelling: Many Paths Through the Forest of Credit Rating and Scoring. Oxford University Press, 2022

Angelini E., di Tollo G. and Roli A., A neural network approach for credit risk evaluation. The quarterly review of economics and finance, vol. 48, n. 4, pp. 733-755, 2008

Baesens B., Van Gestel T., Viaene S., Stepanova M., Suykens J. and Vanthienen J., Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, vol. 54, n. 6, 2003, pp. 627–635

Balogh, S., & Johnson, C., AI can help reduce inequity in credit access, but banks will have to trade off fairness for accuracy—For now. Insider.com, 2021

Barboza F., Kimura H., Altman E., Machine learning models and bankruptcy prediction, Expert Systems with Applications, Volume 83, 2017

Bequé A. and Lessmann S., Extreme learning machines for credit risk: An empirical evaluation. Expert Systems with Applications, 86, 2017

Berg T., Burg V., Gombović A. and Puri M., On the Rise of FinTechs – Credit Scoring using Digital Footprints, NBER Working Papers 24551, 2018

Breeden J., A Survey of Machine Learning in Credit Risk, 2020, DOI:10.13140/RG.2.2.14520.37121

Bussmann N., Giudici P., Marinelli D., and Papenbrock J., Explainable machine learning in credit risk management, Computational Economics, 2021, vol. 57, n. 1, pp. 203-216

Capital Requirements Regulation (CCR), Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms, 2013

Carmona, P., Climent, F., & Momparler, A., Predicting failure in the US banking sector, International Review of Economics and Finance, 61, 2019

Crosato, L., Domenech, L., & Liberati C., Predicting SME's default: Are their websites informative?, Economics Letters, 204, 2021

Crosato, L., Domenech, L., & Liberati C., Websites' data: a new asset for enhancing credit risk modeling, Annals of Operations Research, 2023

Deloitte, Artificial Intelligence for Credit Risk Management, April, 2020

Di Biasi P., Gnutti R., Resti A. and Vergari D., Machine learning per il rischio di credito: quale ruolo nei modelli regolamentari? [it], in Bancaria, n. 5/2022

EBA, Report on Big Data and Advanced Analytics, January, 2020a

EBA, First evidence on the use of moratoria and public guarantees in the EU banking sector, Thematic Note EBA/Rep/2020/31, November, 2020b

EBA, Discussion Paper on machine learning for IRB models, November, 2021

ESMA, Artificial intelligence in EU securities markets, February 2023

European Commission, Ethics Guidelines for Trustworthy AI, June 2018

European Commission, Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, April 2021

European Parliament, Artificial Intelligence Act, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Texts Adopted, June, 2023

Experian, Experian Study on Credit Decisioning and Alternative Data Use, 2022

Financial Stability Board, Artificial intelligence and machine learning in financial services. Market developments and financial stability implications, 1, 2017

FitchSolutions, Credit risk, data and AI. Managing spiralling demands and delivering value, 2022

Giudici P., Raffinetti E., Shapley-Lorenz eXplainable Artificial Intelligence, Expert Systems With Applications, Expert Systems With Applications, volume 167, issue 14, p 104, 2021, https://doi.org/10.1016/j.eswa.2020.114104

Giudici P., Centurelli M. and Turchetta S., Measuring AI SAFEty, 2023, available at https://ssrn.com/abstract=4450731

González-Fernández M. and González-Velasco C., An Alternative Approach to Predicting Bank Credit Risk in Europe with Google Data, Finance Research Letters, vol. 35, 2019

Gül S., Kabak Ö. and Topcu I., A Multiple Criteria Credit Rating Approach Utilizing Social Media Data, Data & Knowledge Engineering, vol. 116, 2018, pp. 80-99

Hamon, R., Junklewitz, H., Sanchez, I. Robustness and Explainability of Artificial Intelligence - From technical to policy solutions, EUR 30040, Publications Office of the European Union, Luxembourg, Luxembourg, 2020, ISBN 978-92-79-14660-5 (online), doi:10.2760/57493 (online), JRC119336

Herrmann H. and Masawi B., Three and a half decades of artificial intelligence in banking, financial services, and insurance: A systematic evolutionary review, Strategic Change, vol. 31, 2022, pp. 549-569

Institute of International Finance, Machine Learning in Credit Risk, 2019

Institute of International Finance, Explainability is a trust issue, Briefing Note, June, 2021

Khandani A. E., Kim A. J. and Lo A. W., Consumer credit risk models via machine learning algorithms, Journal of Banking & Finance, vol. 34, n. 11, pp. 2767-2787

Koyuncugil A.S. and Ozgulbas N., Financial early warning system model and data mining application for risk detection, Expert systems with Applications, vol. 39, n. 6, 2012

Lessmann S., Baesens B., Seow H.-V. and Thomas L. C., Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, vol. 247, n. 1, 2015, pp. 124-136

Locatelli R., Pepe G. and Salis F. (eds)., Artificial Intelligence and Credit Risk, Palgrave Macmillan, 2022

McKinsey & Company, AI-powered decision making for the bank of the future, March, 2021a

McKinsey & Company, Designing next-generation credit-decisioning model, December, 2021b

Mhlanga D., Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment, International Journal of Financial Studies, 2021, vol. 9., n. 39, https://doi.org/10.3390/ijfs9030039

Monk A., Prins M., and Rook D., Rethinking Alternative Data in Institutional Investment, The Journal of Financial Data Science, 2019, Winter, pp. 14-31

Moscatelli M., Parlapiano F., Narizzano S., and Viggiano G., Corporate default forecasting with machine learning, Expert Systems with Applications, 2020, vol. 161

Nalić, J., Martinović, G., & Žagar, D., New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers, Advanced Engineering Informatics, vol. 45, 2020

Odinet C.K., Fintech Credit and the Financial Risk of AI, in Cambridge Handbook of AI and Law (Kristin Johnson & Carla Reyes eds., 2022), University of Iowa Legal Studies Research Paper No. 2021-39, available at SSRN: https://ssrn.com/abstract=3917638

Onay C. and Öztürk E., A Review of Credit Scoring Research in the Age of Big Data, Journal of Financial Regulation and Compliance, vol. 26, n. 3, 2018, pp. 382-405

Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M., Big data techniques to measure credit banking risk in home equity loans. Journal of Business Research, 89, 2018

Imane Rhzioual Berrada I., Barramou F.Z. and Bachir Alami O., A review of Artificial Intelligence approach for credit risk assessment, 2022, "2nd International Conference on Artificial Intelligence and Signal Processing (AISP)", DOI: 10.1109/AISP53593.2022.9760655

Roeder J., Alternative Data for Credit Risk Management: An Analysis of the Current State of Research, 2021, BLED 2021 Proceedings

Safi R. and Lin Z., Using Non-Financial Data to Assess the Creditworthiness of Businesses in Online Trade, Proceedings of the 18th Pacific Asia Conf. on Information Systems, Chengdu, China, 2014

Schapire R.E., The boosting approach to machine learning: An overview, Nonlinear estimation and classification, 2003, pp. 149-171

Skanderson D.M., Managing Discrimination Risk of Machine Learning and AI Models, ABA Journal of Labor & Employment Law, vol. 2, 2021

Theuri J. and Olukuru J., The impact of Artificial Intelligence and how it is shaping Banking, Kenya Bankers Association (KBA) Centre for Research on Financial Markets and Policy, Working Paper Series, 2022, WPS/07/22

Van Thiel D. and Van Raaij W.F., Artificial intelligence credit risk prediction: An empirical study of analytical artificial intelligence tools for credit risk prediction in a digital era, Journal of Risk Management in Financial Institutions, 2019, vol. 12, n. 3, pp. 268-286

Visco I., Intervento del Governatore della Banca d'Italia [it], Italian Banking Association, Rome, July 5th, 2023

Wang, Y., Zhang, Y., Lu, Y., & Yu, X., A comparative assessment of credit risk model based on Machine Learning – A case study on bank loan data, Procedia Computer Science, vol. 174, 2020

West D., Neural network credit scoring models, Computers & Operations Research, vol. 27, n. 11-12, September, 2000, pp. 1131-1152

Yang S., Liu Z. and Wang X., News Sentiment, Credit Spreads, and Information Asymmetry, North American Journal of Economics & Finance, vol. 52, 2020

Zensar Technologies, The Case for Re-inventing the Credit Decisioning Approach

Zhang Z., Gao G., and Shi Y., Credit risk evaluation using multicriteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research, vol. 237, n. 1, pp. 335-348

Zhu, Y., Zhou, L., Xie, C., Wang, G.-J., & Nguyen, T. V., Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 2019

Modeling the interest rates term structure using Machine Learning: a Gaussian process regression approach

Alessio Delucchi (Avvale S.p.A) – Pier Giuseppe Giribone (University of Genoa – Department of Economics; BPER Banca – Financial Engineering)

Corresponding Author: piergiuseppe.giribone@bper.it

Article submitted to double-blind peer review, received on 2nd August 2023 and accepted on 2nd Decembre 2023

Abstract

The correct modeling of the interest rates term structure should definitely be considered an aspect of primary importance since the forward rates and the discount factors used in any financial and risk analysis are calculated from such structure. The turbulence of the markets in recent years, with negative interest rates followed by their recent substantial rise, the period of the COVID pandemic crisis, the political instabilities linked to the war between Ukraine and Russia have very often led to observe anomalies in the shape of the interest rate curve that are difficult to represent using traditional econometric models, to the point that researchers have to address this modeling problem using Machine Learning methodologies. The purpose of this study is to design a model selection heuristic which, starting from the traditional ones (Nelson-Siegel, Svensson and de Rezende-Ferreira) up to the Gaussian Process (GP) Regression, is able to define the best representation for a generic term structure. This approach has been tested over the past five years on term structures denominated in five different currencies: the Swiss Franc (CHF), the Euro (EUR), the British Pound (GBP), the Japanese Yen (JPY) and the U.S. Dollar (USD).

Key Words: Interest rates term structure, Nelson-Siegel model, Svensson model, de Rezende-Ferreira model, Gaussian process regression.

JEL code: C52-C53-C55-E43-E47

1. Introduction

The correct estimation of the interest rates term structure is of primary importance for financial analysts, risk managers, actuarial experts, and policy makers. Precisely to meet this specific need, a substantial scientific literature has developed aimed at its correct model representation.

One of the first advanced approaches is the smoothed bootstrap initially proposed by (Bliss and Fama, 1987). They proposed to derive zero rates from raw market data and then fit them to the data with a smooth and continuous curve.

To this end, numerous curve fitting spline methods have been employed: quadratic and cubic splines (McCulloch, 1971 and 1975), exponential splines (Vasicek and Fong, 1982), B-splines (Shea, 1984) and (Steeley, 1991), quartic maximum smoothness splines (Adams and Van Deventer, 1994) and penalty function-based splines (Fischer, Nychka and Zervos, 1994) and (Waggoner, 1997).

These approaches have been criticized by (Annaert et al., 2012) because they are characterized by unwanted economic properties as they are statistical techniques that do not incorporate micro-macro economic principles in their functioning.

(Seber and Wild, 2003) also highlight that these methodologies contribute to having a "black-box" type of interpretative effect and therefore should be avoided in contexts of standard financial markets free from turbulence.

(Nelson and Siegel, 1987), (Svensson 1994 and 1996) and, subsequently (de Rezende and Ferreira, 2013), approached the problem of obtaining a smooth bootstrap through nonlinear regression models able to fit reasonably well for different families of term structure shapes observed on the financial markets.

These models are parsimonious, consistent with the theoretical interpretation of the term structure suggested by (Litterman and Scheinkman, 1991) and held in high esteem by both academics and professionals.

Let us consider, for example, that the Nelson-Siegel model and the Svensson model are extensively used by central banks and monetary policy makers (Bank of International Settlements, 2005) and (European Central Bank, 2008).

It should be noted that non-linear models and in particular those which envisage the estimation of a large number of parameters can be subject to potential instability in the calibration phase, having to resort to a numerical optimization routine, which most times is constituted by an algorithm of local search for solutions, generally a quasi-Newtonian one such as L-BFGS (Nocedal, 1980) or a Direct Search one, as a simplex by (Nelder and Mead, 1965).

(Cairns and Pritchard, 2001) show that the estimates of the Nelson-Siegel model are very sensitive to the starting values used in the optimization. Moreover, time series of the estimated coefficients have been documented to be very unstable (Barrett, Gosnell and Heuson, 1995), (Fabozzi, Martellini and Priaulet, 2005), (Diebold and Li, 2006), (Gurkaynak, Sack and Wright, 2006), (de Pooter, 2007).

Finally, the standard errors on the estimated coefficients, though seldom reported, are too large (Annaert et al., 2012).

In addition to the potential technical-computational problems mentioned above, we should also consider the fact that a traditional econometric model assumes a priori the functional form according to which the data observed on the market should be explained.

This approach should be pursued whenever possible, typically during periods of stable, non-turbulent financial markets.

If we consider the anomalies that have recently characterized the financial markets, among which we mention: the issue of negative interest rates and their subsequent sharp rise, the current inflationary context and the ongoing war in Europe between Ukraine and Russia, then it could be considered as incongruous to use canonical econometric models to represent the interest rates term structure.

This aspect has already been highlighted by various studies in which a "bottom-up" approach was implemented, i.e., starting from the data, without making a priori hypotheses on the shape of the function of the interest rate curve, the problem was tackled with Machine Learning paradigms (Giribone, 2023).

Among the statistical methods pertaining to this family, the Radial Basis Function (RBF) Neural Network (Cafferata et al., 2019) and the feed-forward Artificial Neural Network (Caligaris and Giribone, 2015) are worth mentioning.
As has been reiterated in (Cafferata, Giribone and Resta, 2018) it should be emphasized that the use of Machine Learning methods and particularly those connected with Deep Learning should somehow be justified: it is unreasonable to adopt statistical methods which are more sophisticated than necessary if market conditions do not require it or if there are no available quotes.
We have chosen the Gaussian Process (GP) regression for performing this task because it is a methodology that is able to work with relatively little data available. It also inherited some theoretical principles common to diffusive processes (see third section of the paper) and it has been shown to produce good results in similar financial applications (Gonzalvez *et al.*, 2019).
The present study fits into this context and proposes a heuristic of choice between different models for the correct representation of the interest rates term structure. The proposed algorithm, starting from the simplest traditional econometric models (Nelson-Siegel and Svensson) and reaching the most complex ones (de Rezende-Ferreira), evaluates their performance in terms of goodness of fit (adjusted R^2) and estimation stability of the coefficients (analysis of confidence bands and outliers). Only if the traditional approaches are not in line with expectations, the heuristics would automatically implement a Machine Learning method: the approach proposed in this paper is a Gaussian process regression with an automatic selection of different kernels.
The selection heuristic was tested on interest rates term structures over the last five years for different currencies (USD, CHF, GBP, EUR, JPY), each characterized by different financial instruments from which the relative zero rates were derived (Deposits, Futures, FRAs and Swaps).
The following section summarizes the main features of the traditional econometric models for representing the interest rates term structure (Nelson-Siegel, Svensson and de Rezende-Ferreira). The third section illustrates the operating principles of a Gaussian Process Regression providing evidence of how an incorrect or unsatisfactory modeling reached with the previously mentioned approaches can be solved. The fourth section illustrates the operational logic of the model selection heuristics in detail and applies them to different case studies. The last section provides the statistics that emerged from the algorithm and draws the conclusions of the study.

2. Non-linear parametric models

(Nelson and Siegel, 1987) were the first to introduce a simple model for interest rates that also has a satisfactory predictive power both for short and very long maturities; these characteristics make it a still relevant approach both for scholars and professionals. Many researchers over the years further developed this model. Among the many contributions, the approaches presented here are those proposed by (Svensson, 1994) and (de Rezende and Ferreira, 2013) who, as will be explained, added additional terms to the '87 formula in order to have a better fit for the term structure under particular circumstances.

2.1 The Nelson Siegel model

A class of functions that generates the typical yield curve shapes is that associated with solutions to differential equations. The expectations theory of the term structure of interest rates provides heuristic motivation for investigating this class since, if spot rates are generated by differential equations, then forward rates, being forecasts, will be the solution to the equations discussed in (Nelson and Siegel, 1987). The researchers explored two cases:

- The instantaneous forward rate is the solution to a second order differential equation with real and unequal roots.
- The instantaneous forward rate is the solution to a second order differential equation with real and equal roots.

In the first case the instantaneous forward rate is defined as:

$$F(t) = \beta_0 + \beta_1 \cdot \exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \exp\left(-\frac{t}{\tau_2}\right) \quad (1)$$

However, tests made by Nelson and Siegel showed that fitting the model considering $F(t)$ as forward rates resulted in overparameterization and they explained this in two ways: firstly, they observed the fact that changing τ_1 and τ_2 caused almost no change to the fit obtained and, secondarily, using statistical software, the model described failed to satisfactorily converge to a robust solution.

For these reasons Nelson and Siegel studied the second case. The forward rate as a solution for a differential equation with equal roots is:

$$F(t) = \beta_0 + \beta_1 \cdot \exp\left(-\frac{t}{\tau}\right) + \beta_2 \left(\frac{t}{\tau}\right) \exp\left(-\frac{t}{\tau}\right) \quad (2)$$

Integrating the formula for the spot rate they obtained:

$$R(t) = \beta_0 + (\beta_1 + \beta_2) \cdot \frac{\tau\left[1-\exp\left(-\frac{t}{\tau}\right)\right]}{t} - \beta_2 \cdot \exp\left(-\frac{t}{\tau}\right) \quad (3)$$

Or in the more canonical form:

$$R(t) = \beta_0 + \beta_1 \cdot \frac{\tau\left[1-\exp\left(-\frac{t}{\tau}\right)\right]}{t} + \beta_2 \cdot \left[\frac{\tau\left[1-\exp\left(-\frac{t}{\tau}\right)\right]}{t} - \exp\left(-\frac{t}{\tau}\right)\right] \quad (4)$$

Nelson and Siegel show that function (4) is able to capture the typical shapes assumed by interest rate term structures. Looking more closely at the function and at the meaning of the coefficients, the three βs can be seen as the strength of the different term, with β_0 measuring the weight of the long-term rates, β_1 the weight of short-term rates and β_2 the weight of the medium-term rates. This kind of interpretation depends on the τ factor, that is considered as a time decay factor which affects mostly the short-term component, only mildly the mid-term one and does not influence the long-term part at all.

An example of this model at work is shown for two cases, one in which the model can satisfactorily explain the behaviour of interest rates and another one in which instead it fails to converge at all (Figure 1).

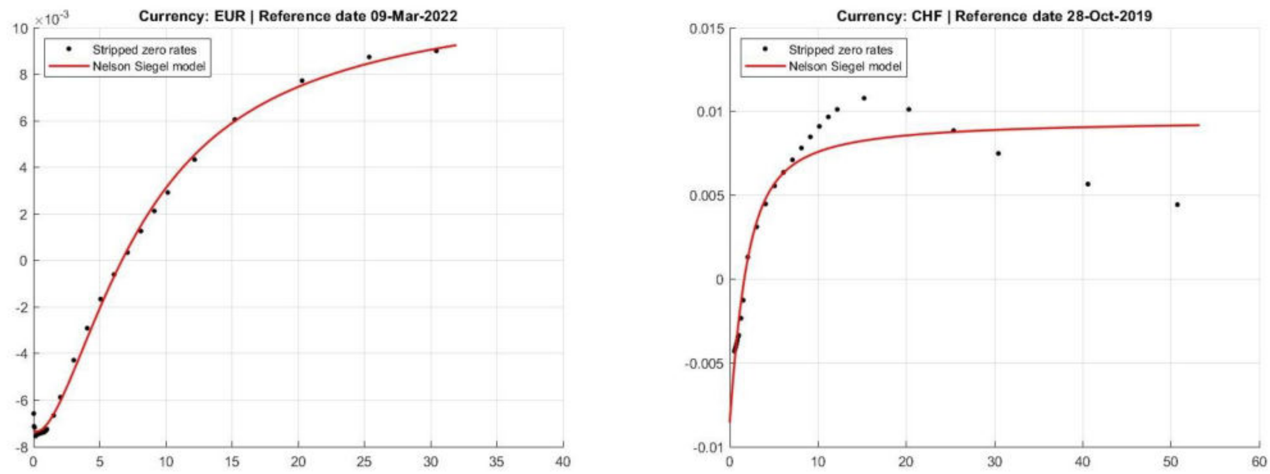


Figure 1: The Nelson Siegel model: good fitting versus poor fitting

2.2) The Svensson model

(Svensson, 1994) proposed a new version of the Nelson and Siegel model in which the author added a further term in order to catch a second hump and thus increase the flexibility of the model. The forward rate function for the Svensson model is as follows:

$$F(t) = \beta_0 + \beta_1 \cdot \exp\left(-\frac{t}{\tau_1}\right) + \beta_2\left(\frac{t}{\tau_1}\right)\exp\left(-\frac{t}{\tau_1}\right) + \beta_3\left(\frac{t}{\tau_2}\right)\exp\left(-\frac{t}{\tau_2}\right) \quad (5)$$

Applying the formula of the spot rate by integrating the forward rate, the new function defining the spot rate is:

$$R(t) = \beta_0 + \beta_1 \cdot \frac{\tau_1\left[1-\exp\left(-\frac{t}{\tau_1}\right)\right]}{t} + \beta_2 \cdot \left[\frac{\tau_1\left[1-\exp\left(-\frac{t}{\tau_1}\right)\right]}{t} - \exp\left(-\frac{t}{\tau_1}\right)\right] + \beta_3 \cdot \left[\frac{\tau_2\left[1-\exp\left(-\frac{t}{\tau_2}\right)\right]}{t} - \exp\left(-\frac{t}{\tau_2}\right)\right] \quad (6)$$

The fourth term, made of two parameters β_3 and τ_2 (which must be positive) is then able to capture a term structure which shape includes two humps. The interpretation of the other terms remains the same.
Svensson proves that the Nelson Siegel model goodness of fit is fulfilling in most cases, but sometimes, when the term structure shape proves to be more complex, the extended model can improve the fit in a significant way.

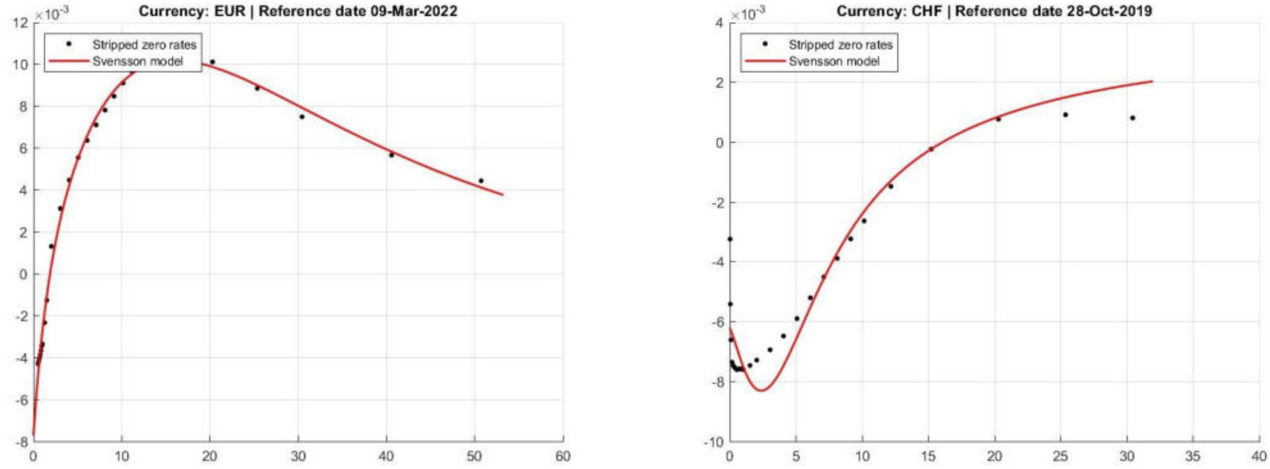


Figure 2: The Svensson model: good fitting versus poor fitting

2.3) The de Rezende and Ferreira model

(de Rezende and Ferreira, 2011) introduced a model that further develops the Nelson Siegel and the Svensson models, adding a fifth factor to address an additional need for flexibility. The new formulas for spot and forward rates are:

$$F(t) = \beta_0 + \beta_1 \cdot \exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \left(\frac{t}{\tau_1}\right) \exp\left(-\frac{t}{\tau_1}\right) + \beta_3 \left(\frac{t}{\tau_2}\right) \exp\left(-\frac{t}{\tau_2}\right) + \beta_4 \left(\frac{t}{\tau_3}\right) \exp\left(-\frac{t}{\tau_3}\right) \quad (7)$$

$$R(t) = \beta_0 + \beta_1 \frac{\tau_1\left[1-\exp\left(-\frac{t}{\tau_1}\right)\right]}{t} + \beta_2 \left[\frac{\tau_1\left[1-\exp\left(-\frac{t}{\tau_1}\right)\right]}{t} - \exp\left(-\frac{t}{\tau_1}\right)\right] + \beta_3 \left[\frac{\tau_2\left[1-\exp\left(-\frac{t}{\tau_2}\right)\right]}{t} - \exp\left(-\frac{t}{\tau_2}\right)\right] + $$
$$+ \beta_4 \left[\frac{\tau_3\left[1-\exp\left(-\frac{t}{\tau_3}\right)\right]}{t} - \exp\left(-\frac{t}{\tau_3}\right)\right] \quad (8)$$

We notice that the fifth term proposed recalls the one introduced by Svensson. The interpretation for this new term is that of a second slope of the curve (to catch a third hump), while the interpretation of the other terms remains unaltered.

This model is expected to work well in case of a very complex and twisted curve, so in those few cases in which the preceding models may fail to fit or tend to underfit.

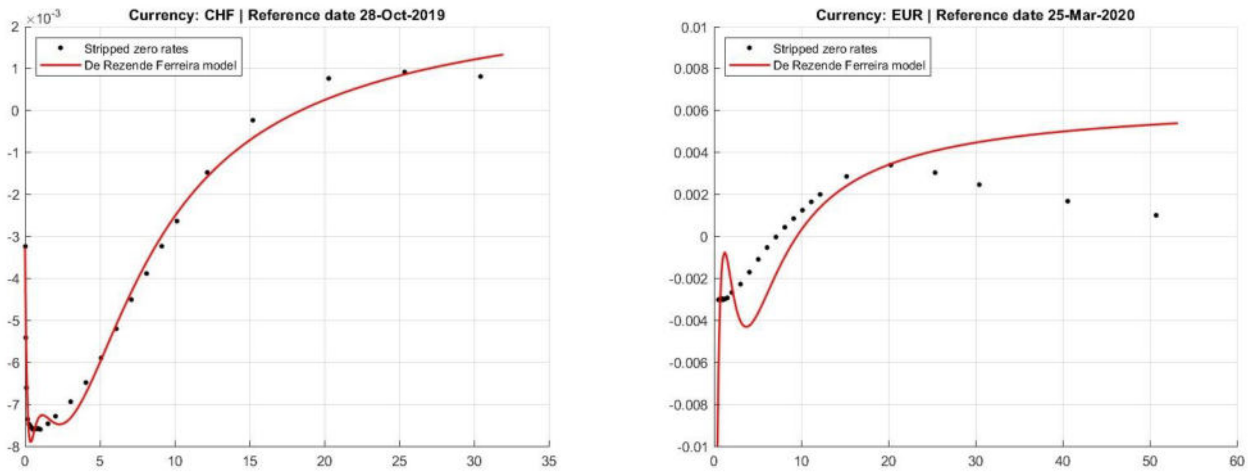


Figure 3: The de Rezende – Ferreira model: good fitting versus poor fitting

3. A Machine Learning approach through a Gaussian process regression

There are many ways to interpret GP regression models, and following the work of Rasmussen and Williams (2006) there are two main approaches to GPs:

-A weight-space view: the typical way of looking at a regression model, mostly focused on parameters.

-A function-space view: considering GP as a distribution over functions.

In both cases Bayesian Linear Regression (BLR) is involved and indeed we can think of GPs as a generalization of this peculiar type of regression.

BLR takes advantage of normal distribution properties (conditioning and marginalization) in order to analytically solve the regression problem. It is composed of three elements: Prior distribution, Likelihood and Posterior distribution.

3.1 The Weight-space view

Given the dataset: $D = \{(x_i, y_i) | i = 1, \cdots, n\}$, the linear regression model is:

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon \qquad y = f(\boldsymbol{x}) + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma_n^2) \quad (9)$$

Where $\boldsymbol{x}$ is the input vector, $\boldsymbol{w}$ is the vector of weights (i.e. parameters) of the linear model, f is the function value and y is the observed target value. We have assumed that the observed values y differ from the function values $f(\boldsymbol{x})$ by additive noise, which follows an independent, identically distributed Gaussian distribution with zero mean and variance σ_n^2.

This noise assumption, together with the model directly gives rise to the likelihood, the probability density of the observations given the parameters, which is factored over cases in the training set because of the independence assumption to give

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{w}) = \prod_{i=1}^{n} p(y_i | x_i, \boldsymbol{w}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(y_i - x_i^T \boldsymbol{w})^2}{2\sigma_n^2}\right) = \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_n^2} |\boldsymbol{y} - \boldsymbol{X}^T \boldsymbol{w}|^2\right) = \mathcal{N}(\boldsymbol{X}^T \boldsymbol{w}, \sigma_n^2 I) \quad (10)$$

Where $|\boldsymbol{z}|$ denotes the Euclidean length of vector $\boldsymbol{z}$, I is the identity matrix of dimension n. In the Bayesian formalism we need to specify a prior over the parameters, expressing our belief about the parameters before we look at the observations. We put a zero mean Gaussian prior with covariance matrix Σ_p on the weights: $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_p)$.

Through Bayes theorem it is possible to find the posterior distribution of the parameter:

$$Posterior = \frac{Likelihood \times Prior}{Marginal\ Likelihood} \quad (11)$$

Considering the particular case of $\boldsymbol{w}$:

$$p(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{X}) = \frac{p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{y}|\boldsymbol{X})} \quad (12)$$

The marginal likelihood as the name suggests is found through marginalization and it is independent from the parameters (the parameters are marginalized out). Analytically, it can be written as:

$$p(\boldsymbol{y}|\boldsymbol{X}) = \int p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{w})p(\boldsymbol{w})\, d\boldsymbol{w} \quad (13)$$

The posterior in Eq. (12) combines the likelihood and the prior, and captures everything we know about the parameters. Writing only the terms from the likelihood and prior which depend on the weights, we obtain:

$$P(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma_n^2}(\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{w})^T(\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{w})\right) \exp\left(-\frac{1}{2}\boldsymbol{w}^T\Sigma_p^{-1}\boldsymbol{w}\right) \propto \exp\left(-\frac{1}{2}(\boldsymbol{w} - \bar{\boldsymbol{w}})^T\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^T + \Sigma_p^{-1}\right)(\boldsymbol{w} - \bar{\boldsymbol{w}})\right) \quad (14)$$

Where $\bar{\boldsymbol{w}} = \sigma_n^{-2}\left(\sigma_n^{-2}\boldsymbol{X}\boldsymbol{X}^T + \Sigma_p^{-1}\right)^{-1}\boldsymbol{X}\boldsymbol{y}$ and we recognize the form of the posterior distribution as Gaussian with mean $\bar{\boldsymbol{w}}$ and covariance matrix A^{-1}:

$$p(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y}) \sim \mathcal{N}\left(\bar{\boldsymbol{w}} = \frac{1}{\sigma_n^2}A^{-1}\boldsymbol{X}\boldsymbol{y}, A^{-1}\right) \quad (15)$$

Where $A = \sigma_n^{-2}\boldsymbol{X}\boldsymbol{X}^T + \Sigma_p^{-1}$ (Rasmussen and Williams, 2006). Notice that for this model (and indeed for any Gaussian posterior) the mean of the posterior distribution $p(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y})$ is also its mode, which is called the maximum a posteriori (MAP) estimate of $\boldsymbol{w}$.

To make predictions for a test case we average over all possible parameter values, weighted by their posterior probability. Thus the predictive distribution for $f_* = f(\boldsymbol{x}_*)$ at $\boldsymbol{x}_*$ is given by averaging the output of all possible linear models with reference to the Gaussian posterior

$$p(f_*|\boldsymbol{x}_*, \boldsymbol{X}, y) = \int p(f_*|\boldsymbol{x}_*, \boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{X}, y)d\boldsymbol{w} = \mathcal{N}\left(\frac{1}{\sigma_n^2}\boldsymbol{x}_*^T A^{-1}\boldsymbol{X}y, \boldsymbol{x}_*^T A^{-1}\boldsymbol{x}_*\right) \quad (16)$$

The predictive distribution is again Gaussian, with a mean given by the posterior mean of the weights from Eq. (15) multiplied by the test input, as one would expect from symmetry considerations. The predictive variance is a quadratic form of the test input with the posterior covariance matrix, showing that the predictive uncertainties grow with the magnitude of the test input, as one would expect for a linear model.

The Bayesian linear model suffers from limited expressiveness. A simple idea to overcome this problem is to first project the inputs into some high dimensional space using a set of basis functions and then apply the linear model in this space instead of directly on the inputs themselves. As long as the projections are fixed functions (i.e. independent of the parameters $\boldsymbol{w}$) the model is still linear in the parameters, and therefore analytically tractable.

Specifically, we introduce the function $\phi(\boldsymbol{x})$ which maps D-dimensional input vector $\boldsymbol{x}$ into an N dimensional feature space. Further let the matrix $\Phi(\boldsymbol{X})$ be the aggregation of columns $\phi(\boldsymbol{x})$ for all cases in the training set. Now the model is

$$f(\boldsymbol{x}) = \phi(\boldsymbol{x})^T\boldsymbol{w} \quad (17)$$

Where the vector of parameters has length N. The analysis for this model is analogous to the standard linear model, except that everywhere $\Phi(\boldsymbol{X})$ is substituted for $\boldsymbol{X}$. Thus the predictive distribution becomes

$$f_*|\boldsymbol{x}_*, \boldsymbol{X}, y \sim \mathcal{N}\left(\frac{1}{\sigma_n^2}\phi(\boldsymbol{x}_*)^T A^{-1}\Phi y, \phi(\boldsymbol{x}_*)^T A^{-1}\phi(\boldsymbol{x}_*)\right) \quad (18)$$

With $\Phi = \Phi(\boldsymbol{X})$ and $A = \sigma_n^{-2}\Phi\Phi^T + \Sigma_p^{-1}$. To make predictions using Eq. (18) we need to invert the A matrix of size $N \times N$ which may not be convenient if N, the dimension of the feature space, is large. However, we can rewrite the equation in the following way:

$$f_*|\boldsymbol{x}_*, \boldsymbol{X}, y \sim \mathcal{N}\left(\phi_*^T\Sigma_p\Phi(K + \sigma_n^2 I)^{-1}y, \phi_*^T\Sigma_p\phi_* - \phi_*^T\Sigma_p\Phi(K + \sigma_n^2 I)^{-1}\Phi^T\Sigma_p\phi_*\right) \quad (19)$$

Where we have used the shorthand $\phi(\boldsymbol{x}_*) = \phi_*$ and defined $K = \Phi^T \Sigma_p \Phi$. To show this for the mean, first note that using the definitions of A and K we have $\sigma_n^{-2}\Phi(K + \sigma_n^2 I) = \sigma_n^{-2}\Phi(\Phi^T \Sigma_p \Phi + \sigma_n^2 I) = A\Sigma_p \Phi$.
Now multiplying through by A^{-1} from left and $(K + \sigma_n^2 I)^{-1}$ from the right gives $\sigma_n^{-2} A^{-1}\Phi = \Sigma_p \Phi (K + \sigma_n^2 I)^{-1}$, showing the equivalence of the mean expressions in Eq. (18) and Eq. (19). For the variance we use the matrix inversion lemma, setting $Z^{-1} = \Sigma_p$, $W^{-1} = \sigma_n^2 I$ and $V = U = \Phi$ therein.

3.2 The Function-space view

Gaussian Processes are defined as “a collection of random variables, any finite number of which have a joint Gaussian distribution” (Rasmussen and Williams, 2006).

A Gaussian process is completely specified by its mean function and covariance function. We define the mean function $m(\boldsymbol{x})$ and the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ of a real process $f(\boldsymbol{x})$ as:

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \quad (20)$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big]$$

And will write the Gaussian process as:

$$f(\boldsymbol{x}) \sim GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \quad (21)$$

A Gaussian process is defined as a collection of random variables. Thus, the definition automatically implies a consistency requirement, which is also sometimes known as the marginalization property. This property means that if the GP e.g. specifies $(y_1, y_2) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then it must also specify $y_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ where Σ_{11} is the relevant submatrix of Σ.

This can be seen as the prior over functions. According to this prior it is possible to define the joint distribution of f, the training outputs and f_* the test outputs:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \quad (22)$$

From this joint distribution, the predictive distribution can be derived, that is, as before, the conditional distribution of f given $\boldsymbol{x}_*, \boldsymbol{x}$ and y:

$$P(f_* | \boldsymbol{x}_*, \boldsymbol{x}, y) \sim N(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*^T) \quad (23)$$

That is exactly the same function defined in the weight-space view. Proceeding through these steps, the reason why GP regression is considered a generalization of BLR becomes clear: GP regression uses kernels instead of basis functions to find the families of the functions for regression.
Using kernels allows to define a very broad family of functions that basis functions alone could not handle. This makes GPs more flexible as it is still possible to implement the Bayesian update and reach a good posterior predictive fit.

3.3 Kernel functions and hyperparameters

Kernel functions control the model, they determine which kind of function is more or less likely to be sampled. The kernel is a function that measures how similar two inputs are and therefore it is quite clear why such functions are used to produce covariance matrices in GPs.

Let us suppose we have $\boldsymbol{x}$ and $\boldsymbol{x}'$; the kernel function is:

$$k(\boldsymbol{x}, \boldsymbol{x}' | \boldsymbol{\tau}) \quad (24)$$

$\boldsymbol{x}$ and $\boldsymbol{x}'$ can refer to any two objects, provided that we can measure similarity between them. $\boldsymbol{\tau}$ is a vector of hyperparameters used to tune the kernel function. The output of the kernel function will be a similarity measure, large and positive if the inputs are very similar, large and negative otherwise.

There are technical restrictions on which functions can be used as kernels, since the covariance matrix must be positive definite (the reason for this restriction is based on the (Mercer, 1909) theorem). The model comes down to which prior functions are likely to be sampled and this is dictated by the kernel, meaning that this aspect is fundamental. It is essential to decide what makes two x similar or dissimilar.

The idea is that GP will sample functions with close y values for $\boldsymbol{x}$ deemed similar by the kernel. In order to clarify this concept, consider the Squared exponential kernel function (or RBF function), one of the most used for Gaussian Processes:

$$k(\boldsymbol{x}, \boldsymbol{x}' | \boldsymbol{\tau}) = \sigma_f^2 \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right) + \sigma_n^2 \quad (25)$$

Hyperparameters in the squared exponential are as follows:

ℓ: length scale. It scales distances between the x. It means that if ℓ is a small value, most input pairs are considered different, and this implies that the sample functions can “wiggle” more rapidly. If ℓ is large on the other hand, most input pairs will be considered similar, leading to smoother sample functions. Intuitively, this function tells how far we have to go before things become virtually uncorrelated with one another.

σ_f^2: output scale (or signal variance) determines the scale of the y values. If σ_f^2 increases, the function spans a bigger part of the y axis; if it decreases, the opposite holds.

σ_n^2: noise variance. It is not a direct parameter of the kernel function, but it influences the likelihood function from which the optimal values of σ_f^2 and ℓ are found.

But how do we find the values for these parameters? As anticipated, the likelihood function plays a fundamental role. The τ vector determines how kernel measures similarity.

The idea is to select those hyperparameters that maximize the log likelihood of y after integrating out possible functions.

The idea is to optimize (i.e., find the maximum of):

$$\ln P(y|\boldsymbol{x},\boldsymbol{\tau},\sigma_n^2) = \ln \int p(y|f,\sigma_n^2)P(f|\boldsymbol{x},\boldsymbol{\tau})\,df \quad (26)$$

Computing $\ln P(y|\boldsymbol{x},\boldsymbol{\tau},\sigma_n^2)$ essentially means finding hyperparameters that improve the fitting of the data through a function f sampled by the prior, and this is done over an infinite number of sample functions.

Intuitively we want to pick the hyperparameters where the prior functions explain the data in the best way, thus hyperparameters for which the f_s fit the data well without conditioning.

This means that:

$$\ln P(y|\boldsymbol{x},\boldsymbol{\tau},\sigma_n^2) \sim \ln N(y|\boldsymbol{0},K(X,X)+\sigma_n^2 I) \quad (27)$$

The gradient of this function can be computed with respect to the hyperparameters. The probability function is differentiable, hence any algorithm based on gradient is able to obtain hyperparameters that maximize the probability. With GPs we can optimize a huge number of hyperparameters.

There are many types of kernels, and the designer can even decide to combine them by adding one to the other (sum of kernels means that two functions are sampled and then the sum of the two leads to the "final" kernel) or multiplying one to the other.

This allows to create more complicated models that can better explain data.

An application of this Gaussian Processes could be done in the fitting of the yield curves for which the parametric models in the previous chapter gave poor results. The results of the GP regression are reported in Figure 4.

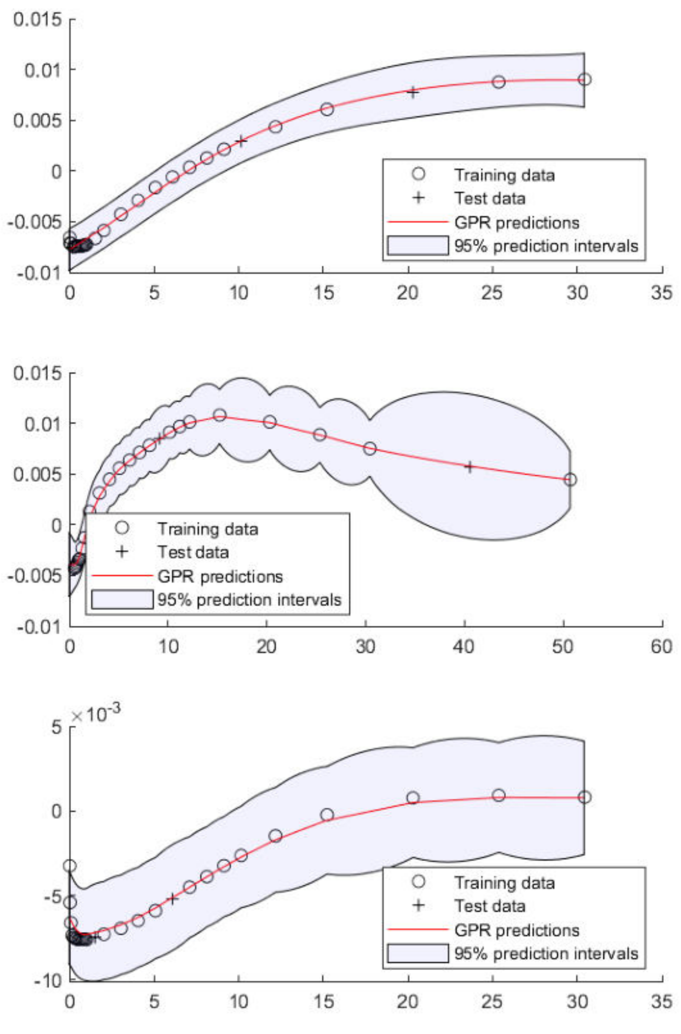


Figure 4: Gaussian Process applications for the previous poor fitting cases

The kernel functions considered in this study are:

- Squared Exponential: already discussed in this section.

- Exponential:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\frac{\|x-x'\|}{\ell}\right) \quad (28)$$

Both the Squared Exponential and the Exponential kernel functions work quite well for smoother functions, while in case of functions with kinks or local structures, other kernel functions, such as the (Matérn, 1960), perform better.

- Matérn 3/2:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{3}\|x-x'\|}{\ell}\right) \exp\left(-\frac{\sqrt{3}\|x-x'\|}{\ell}\right) \quad (29)$$

- Matérn 5/2:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{5}\|x-x'\|}{\ell} + \frac{5\|x-x'\|^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5}\|x-x'\|}{\ell}\right) \quad (30)$$

In short, the Matérn kernel functions tend to be more flexible as they are derived by considering a smoothness parameter with value 3/2 and 5/2 in the cases considered here. Higher values of the smoothness parameter result in smoother and more differentiable functions.

- Rational quadratic:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \frac{\|x-x'\|^2}{2\alpha\ell^2}\right)^{-\alpha} \quad (31)$$

Where α is a positive scale parameter.

The Rational quadratic kernel function incorporates a balance between short-range and long-range correlations. This is possible thanks to the scale parameter α. A higher α value results in a smoother function, capturing long-range correlations. Conversely, a lower α value leads to a rougher function that emphasizes short-range correlations. The idea is that as the value of α changes, a higher weight is assigned to a different section of the curve.

An Automatic Relevance Determination (ARD) version of all the previous kernels can be applied to any kernel function that has a length scale. This method, by introducing a separate length scale parameter for each input variable in the covariance function of the GP model, is a check for the relevance of the input variable.

When the length scale for a particular input variable is small, the GP model becomes more sensitive to variations in that variable, conversely, when the length scale is large, the GP model becomes less sensitive to variations in that variable.

The ARD kernel functions are then:

- ARD Squared exponential kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\sum_{i=1}^{n} \frac{\|x_i - x_i'\|^2}{2\ell_i^2}\right) \quad (32)$$

- ARD Exponential kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\sum_{i=1}^{n} \frac{\|x_i - x_i'\|}{\ell_i}\right) \quad (33)$$

- ARD Matérn 3/2:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \sqrt{3}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|}{\ell_i}\right) \exp\left(-\sqrt{3}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|}{\ell_i}\right) \quad (34)$$

- ARD Matérn 5/2:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \sqrt{5}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|}{\ell_i} + \frac{5}{3}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|^2}{\ell_i^2}\right) \exp\left(-\sqrt{5}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|}{\ell_i}\right) \quad (35)$$

- ARD Rational quadratic:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \frac{1}{2\alpha}\sum_{i=1}^{n} \frac{\|x_i - x_i'\|^2}{\ell_i^2}\right)^{-\alpha} \quad (36)$$

4. Case Study

The methodologies described previously will be used to model the interest rates term structure, considering different time periods, different currencies and for each currency different financial instruments (according to their market liquidity). The time range examined starts from 1st January 2018 and ends on 21st March 2023. Five currencies have been considered: the Swiss Franc (CHF), the Euro (EUR), the British Pound (GBP), the Japanese Yen (JPY) and the U.S. Dollar (USD).The instruments used to model the term structure for each currency are: swaps for CHF, GBP and JPY; deposits, forwards and swaps for EUR; deposits, futures and swaps for USD. The granulometry for each currency, with the corresponding terms and instruments, is shown in Table 1. Data used in this study can be considered in line with the best market practice given that they are retrieved from the Bloomberg® yield curves module. Zero rates and discount factors for each eligible date in the time span have been bootstrapped from market rates. The eligibility of dates depends on one criterion: the number of par rates available; if the missing rates are more than ten, the date is considered ineligible. In case the number of missing rates is less or equal to ten, the missing rates will be interpolated, and the zero rates will be computed. This selection process is further developed in the flow chart depicted in Figure 5.

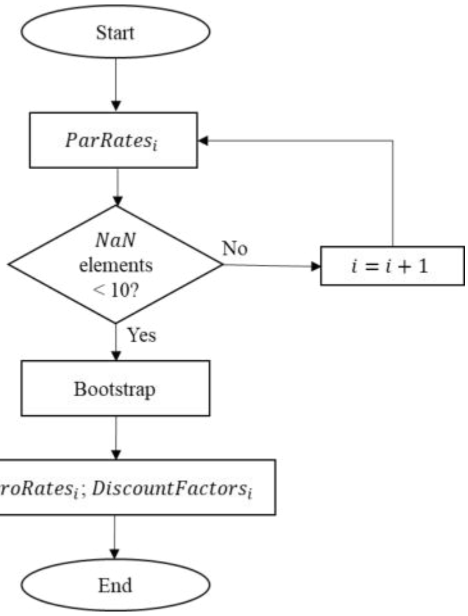


Figure 5: Selection process for the Zero rates stripping algorithm

Instruments per currency

Starting date: 01/01/2018						End date: 03/21/2023			
CHF		EUR		GBP		JPY		USD	
Term	Instrument	Term	Instrument	Term	Instrument	Term	Instrument	Term	Instrument
1W	Swap	6M	Deposit	1W	Swap	1W	Swap	3MO	Deposit
2W	Swap	FRA1X7	FRA	2W	Swap	2W	Swap	FUT_1	Futures
1MO	Swap	FRA2X8	FRA	1MO	Swap	3W	Swap	FUT_2	Futures
2MO	Swap	FRA3X9	FRA	2MO	Swap	1MO	Swap	FUT_3	Futures
3MO	Swap	FRA4X10	FRA	3MO	Swap	2MO	Swap	FUT_4	Futures
4MO	Swap	FRA5X11	FRA	4MO	Swap	3MO	Swap	FUT_5	Futures
5MO	Swap	FRA6X12	FRA	5MO	Swap	4MO	Swap	2Y	Swap
6MO	Swap	FRA9X15	FRA	6MO	Swap	5MO	Swap	3Y	Swap
7MO	Swap	FRA12X18	FRA	7MO	Swap	6MO	Swap	4Y	Swap
8MO	Swap	2Y	Swap	8MO	Swap	7MO	Swap	5Y	Swap
9MO	Swap	3Y	Swap	9MO	Swap	8MO	Swap	6Y	Swap
10MO	Swap	4Y	Swap	10MO	Swap	9MO	Swap	7Y	Swap
11MO	Swap	5Y	Swap	11MO	Swap	10MO	Swap	8Y	Swap
12MO	Swap	6Y	Swap	12MO	Swap	11MO	Swap	9Y	Swap
18MO	Swap	7Y	Swap	18MO	Swap	12MO	Swap	10Y	Swap
2Y	Swap	8Y	Swap	2Y	Swap	15MO	Swap	11Y	Swap
3Y	Swap	9Y	Swap	3Y	Swap	18MO	Swap	12Y	Swap
4Y	Swap	10Y	Swap	4Y	Swap	2Y	Swap	15Y	Swap
5Y	Swap	11Y	Swap	5Y	Swap	3Y	Swap	20Y	Swap
6Y	Swap	12Y	Swap	6Y	Swap	4Y	Swap	25Y	Swap
7Y	Swap	15Y	Swap	7Y	Swap	5Y	Swap	30Y	Swap
8Y	Swap	20Y	Swap	8Y	Swap	6Y	Swap	40Y	Swap
9Y	Swap	25Y	Swap	9Y	Swap	7Y	Swap	50Y	Swap
10Y	Swap	30Y	Swap	10Y	Swap	8Y	Swap		
12Y	Swap	40Y	Swap	12Y	Swap	9Y	Swap		
15Y	Swap	50Y	Swap	15Y	Swap	10Y	Swap		
20Y	Swap			20Y	Swap	11Y	Swap		
25Y	Swap			25Y	Swap	12Y	Swap		
30Y	Swap			30Y	Swap	15Y	Swap		
				40Y	Swap	20Y	Swap		
				50Y	Swap	25Y	Swap		
						30Y	Swap		
						35Y	Swap		
						40Y	Swap		

Table 1: Financial instruments used for bootstrap and Interest rates term structures granulometry

After this preliminary filter, the initial 1362 dates for each of the five currencies become 1349 for CHF, 1342 for EUR, 1326 for GBP, 1293 for JPY, and 1347 for USD.
The surfaces of all the stripped zero rates and discount factors, divided by currency, are reported from Figure 6 to Figure 10.

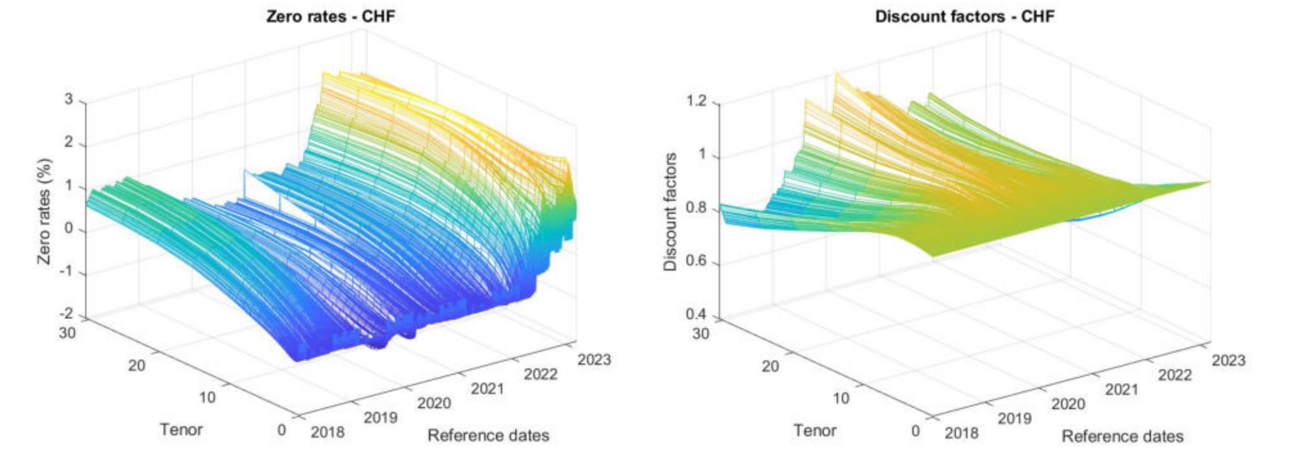


Figure 6: Term structure and Discount factors surface - CHF

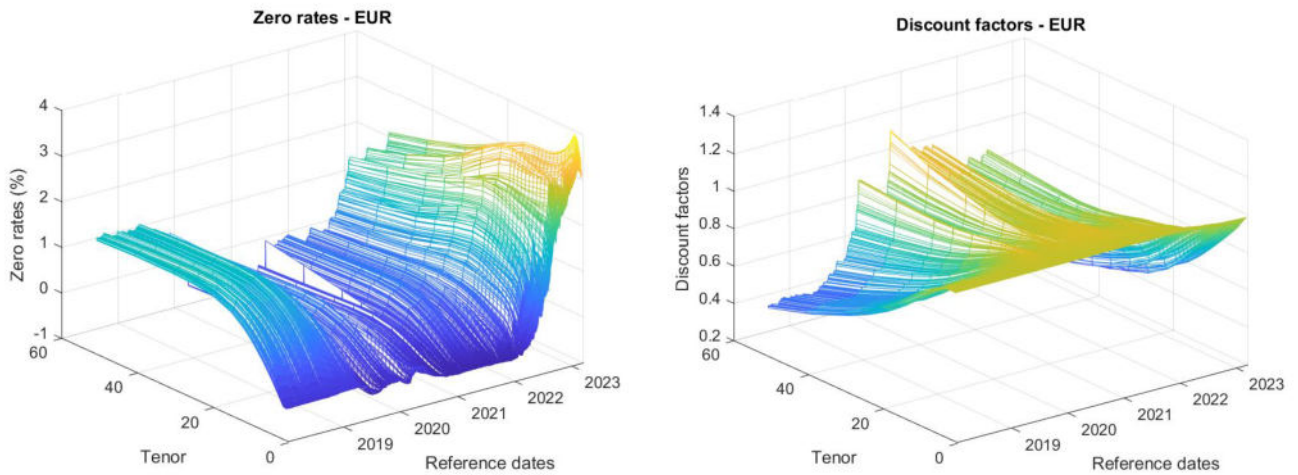


Figure 7: Term structure and Discount factors surface - EUR

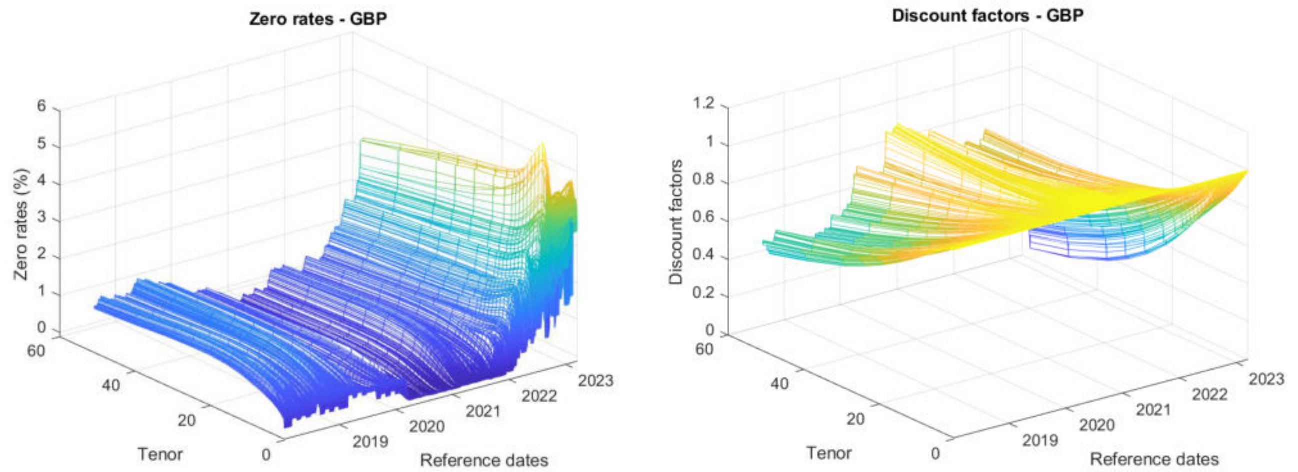


Figure 8: Term structure and Discount factors surface - GBP

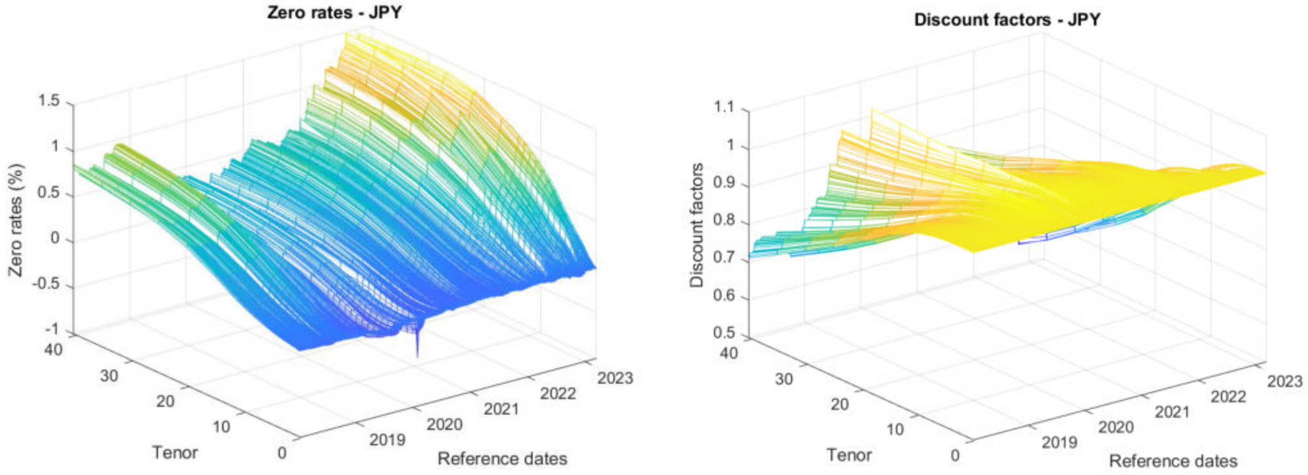


Figure 9: Term structure and Discount factors surface - JPY

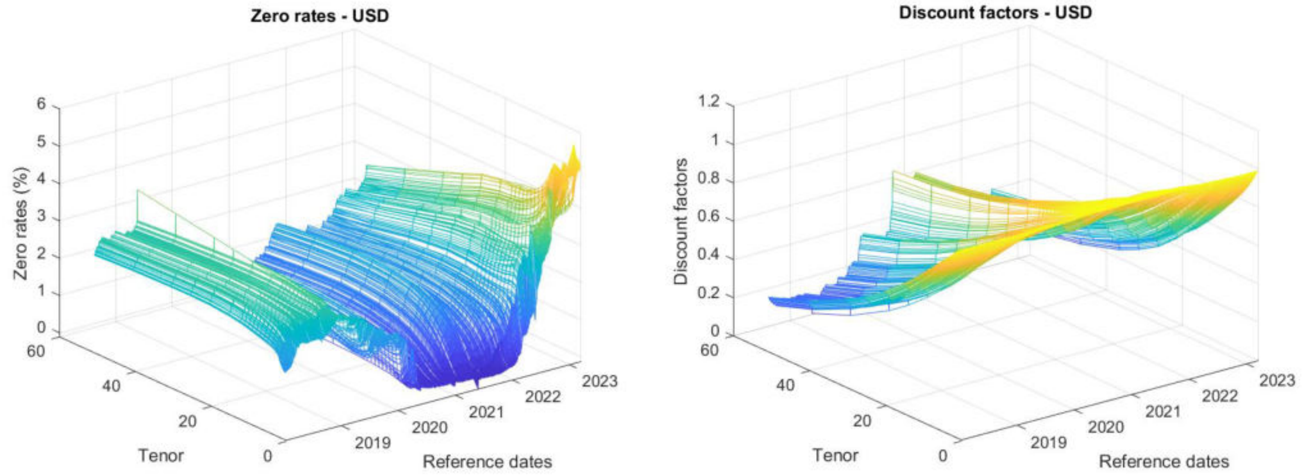


Figure 10: Term structure and Discount factors surface – USD

The term structure for each eligible date is then modelled following parsimonious criteria. The idea is to start from the simplest model and use more complex models only if certain criteria on the goodness of fit or stability of the results are not fulfilled.

This means that the first model that will be used is Nelson Siegel, then Svensson, then de Rezende and if none of the parametric models works satisfactorily, then the Gaussian Process Regression will be implemented.

The first discriminating criterion is the goodness of fit, measured through the adjusted R-square. The threshold is set to a value of 0.95, and below such value the model is rejected in favour of more complex (and flexible) models.

After checking the goodness of fit, the focus is on the stability of results. The selection for stability is conducted following two steps:

- Detection of unrealistically unstable results,

- Detection of outliers.

In the first step, the coefficients of the parametric models are analyzed observing their mean value and using a 95% confidence level. If a model's coefficient has an upper (lower) bound of the confidence band which is ten times higher (lower) than the mean value, then it is considered unrealistically unstable, and the model is discarded.

The second step, on the other hand, implies a more "classical" procedure of outlier detection. If the model's coefficient has an upper (lower) bound that is higher than the mean value plus two times the standard deviation, then the model is discarded. It is worth to highlight that we have implemented a robust check on the starting guesses related to the nonlinear least squares solver.

If statistical performances are not aligned with the previous criteria using the first random initial values, the algorithm automatically produces two other sets of values for the solver.

If none of these attempts works, then the heuristic will take into consideration a more complex parametric model. The process of model selection is further developed in the flow chart below (Figure 11).

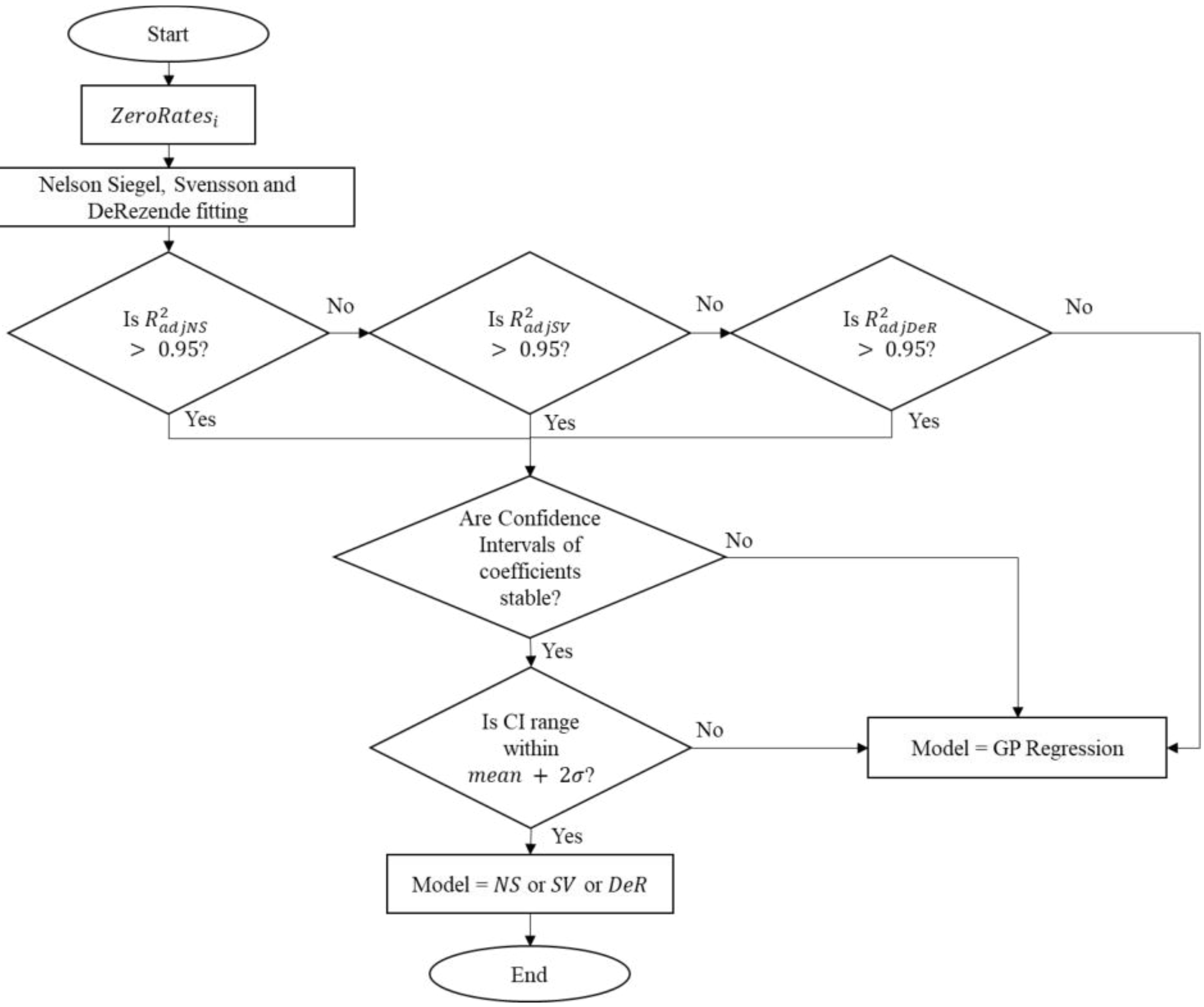
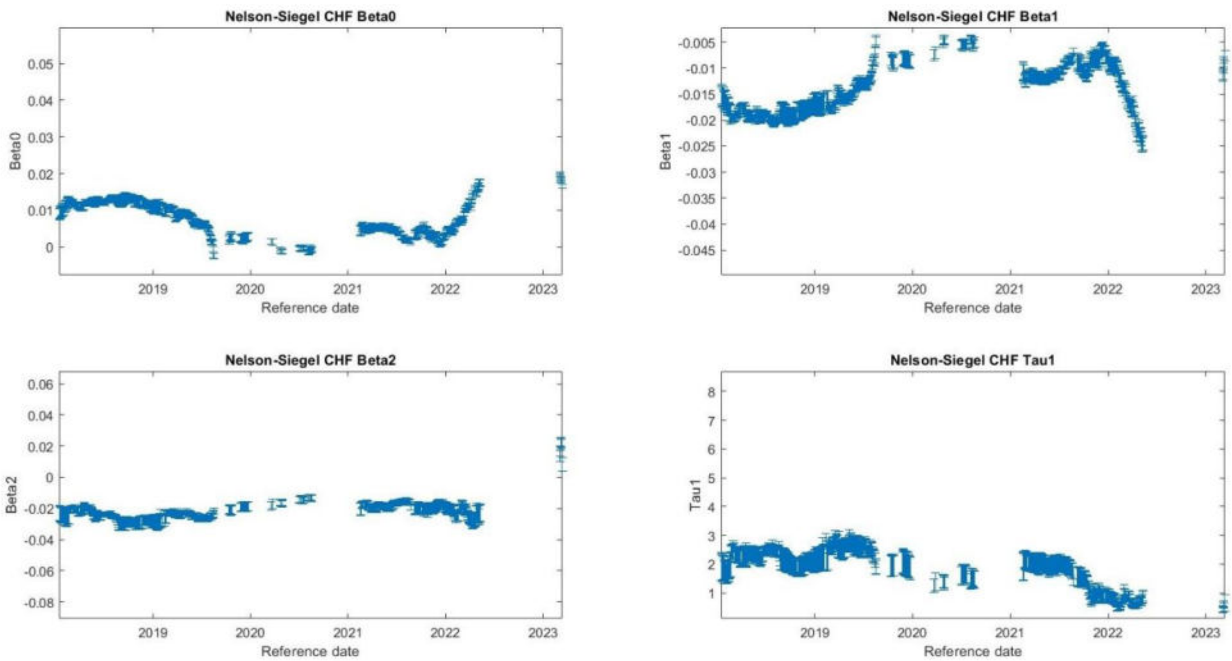


Figure 11: Heuristic for the model selection

The results of this selection are reported in the next Figures (12-16).



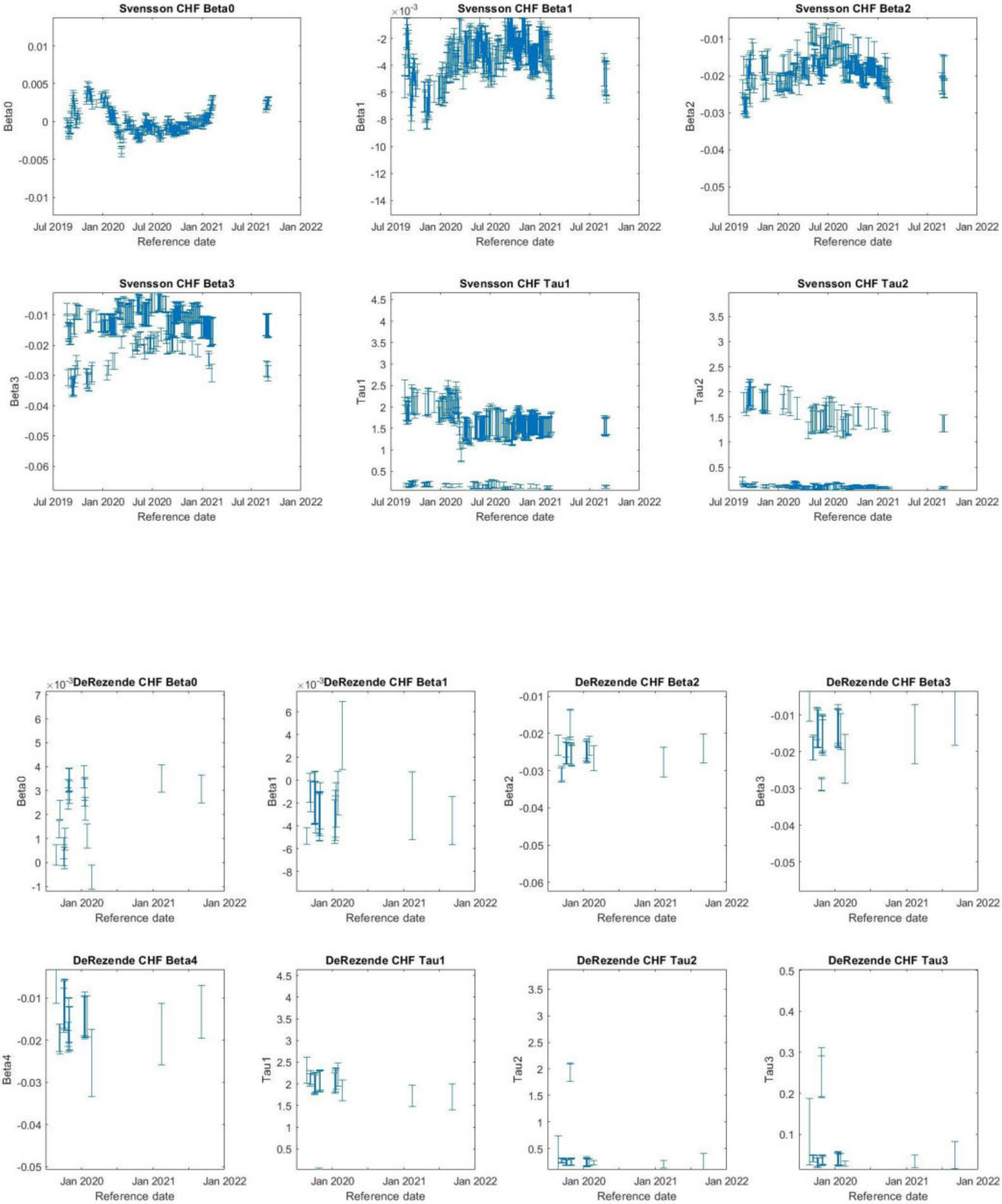


Figure 12: Nelson-Siegel, Svensson and de Rezende coefficients for the CHF Interest rates term structures

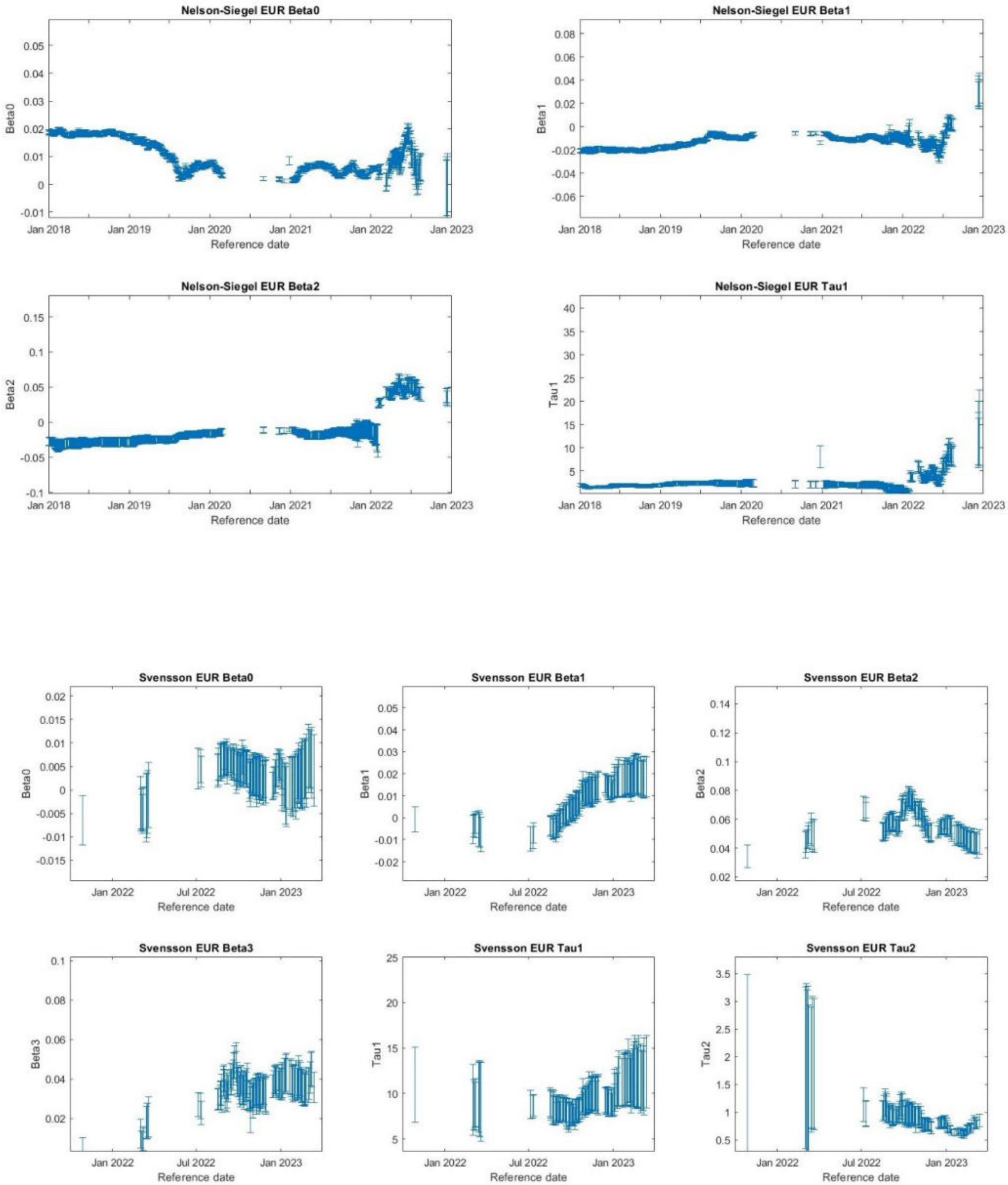
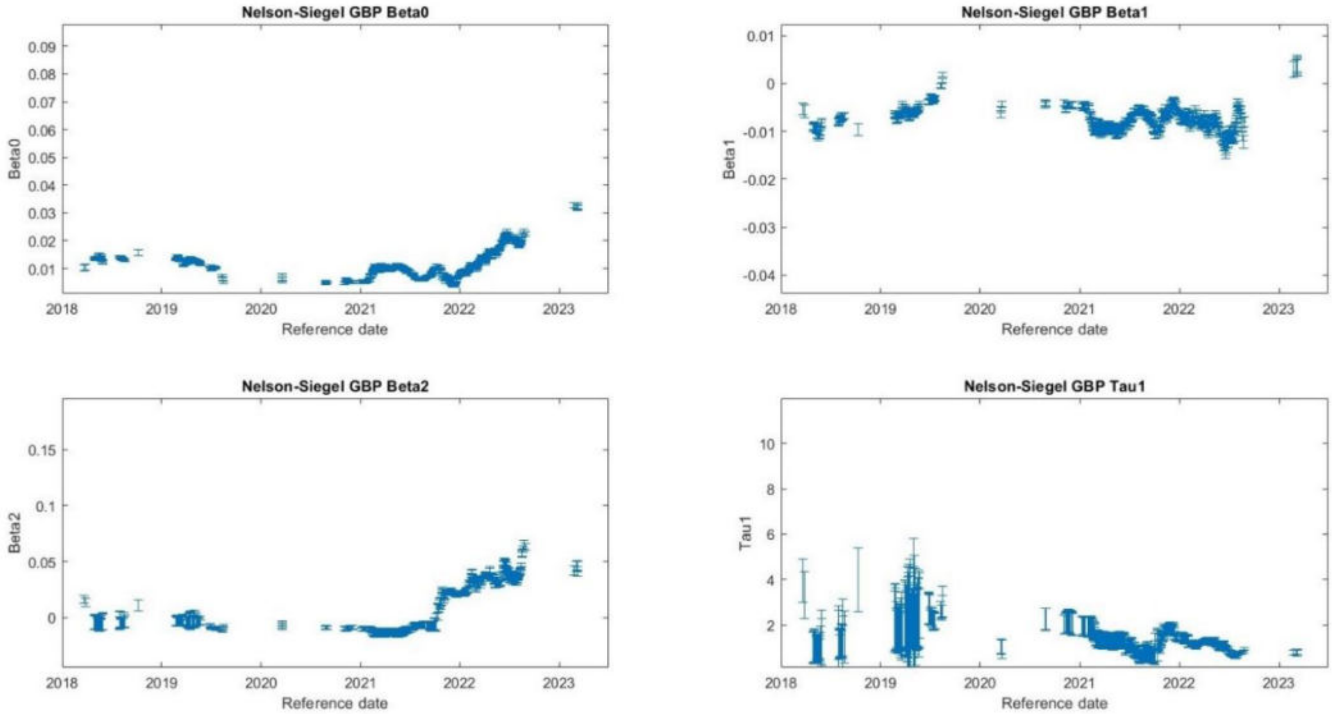
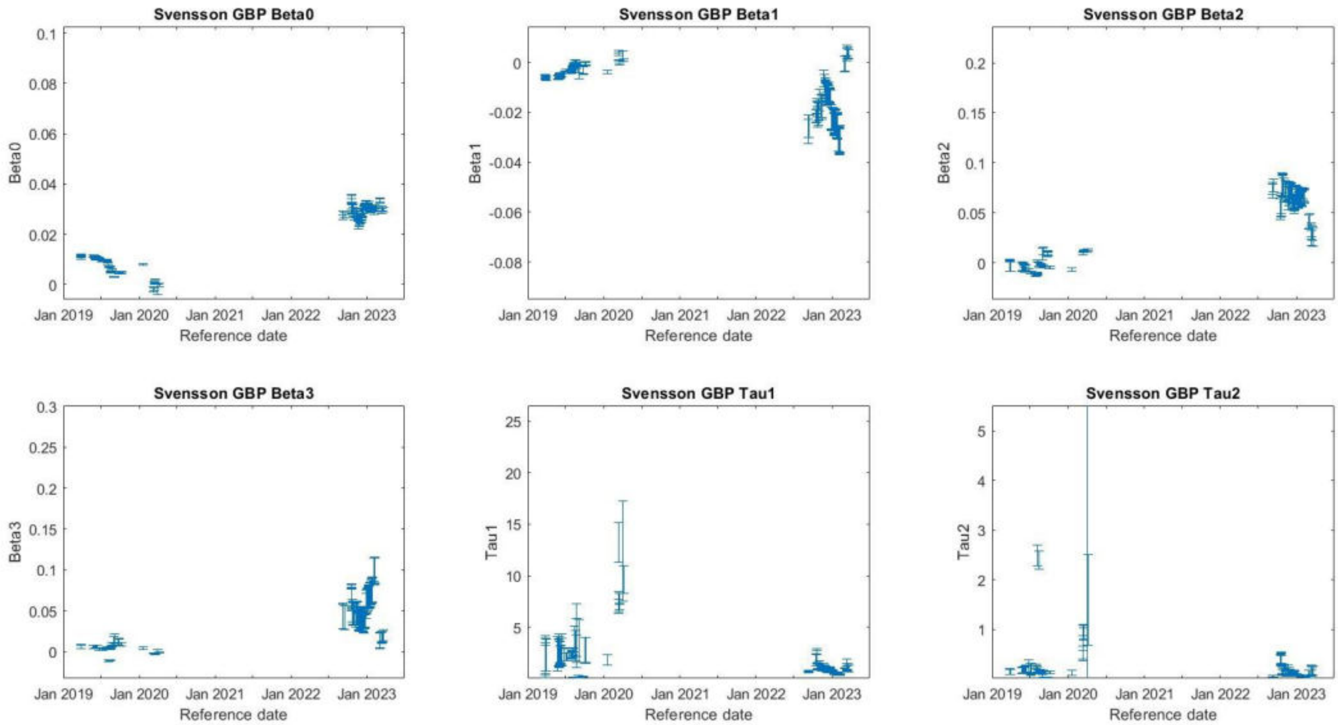


Figure 13: Nelson-Siegel and Svensson coefficients for the EUR Interest rates term structures

The results for the de Rezende model for the EUR currency are quite peculiar: for every estimation made with this model, the coefficients are deemed as unstable.


Nelson-Siegel GBP Beta0
Beta0
Reference date
Nelson-Siegel GBP Beta1
Beta1
Reference date
Nelson-Siegel GBP Beta2
Beta2
Reference date
Nelson-Siegel GBP Tau1
Tau1
Reference date


Svensson GBP Beta0
Beta0
Reference date
Svensson GBP Beta1
Beta1
Reference date
Svensson GBP Beta2
Beta2
Reference date
Svensson GBP Beta3
Beta3
Reference date
Svensson GBP Tau1
Tau1
Reference date
Svensson GBP Tau2
Tau2
Reference date

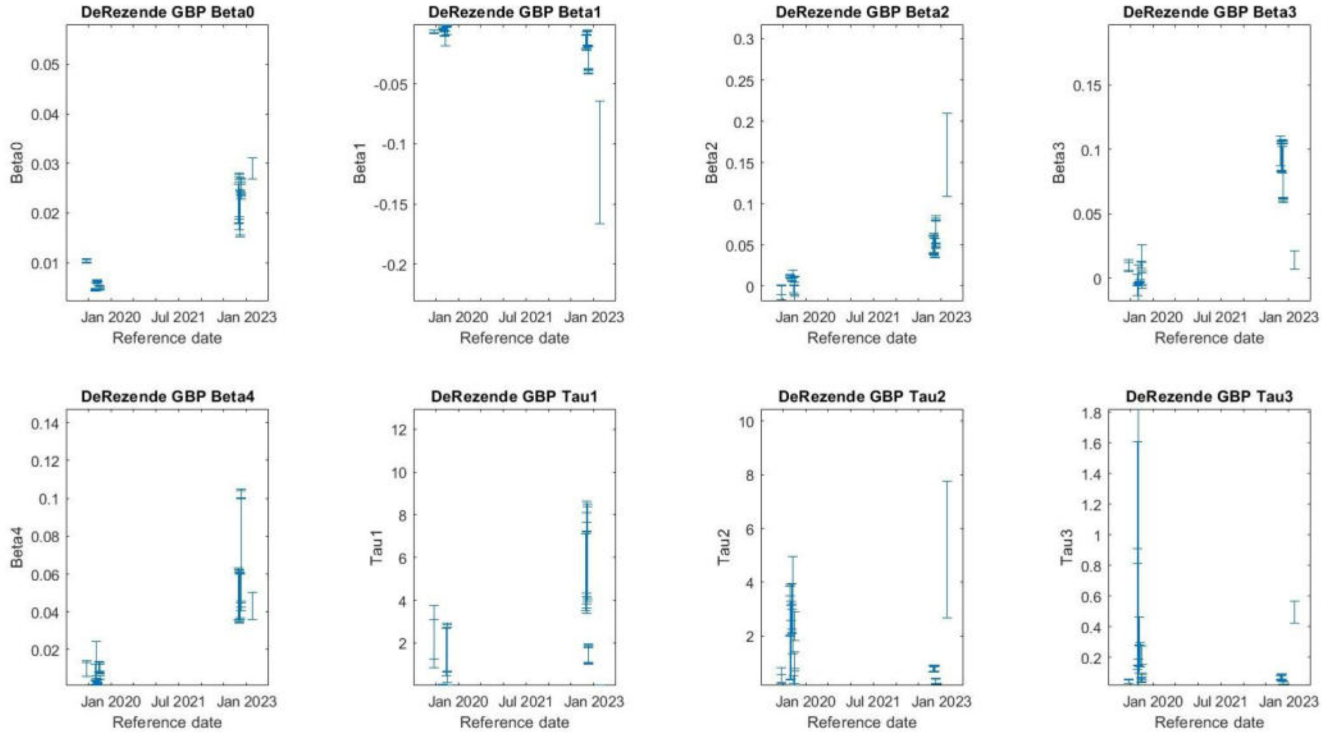


Figure 14: Nelson-Siegel, Svensson and de Rezende coefficients for the GBP Interest rates term structures

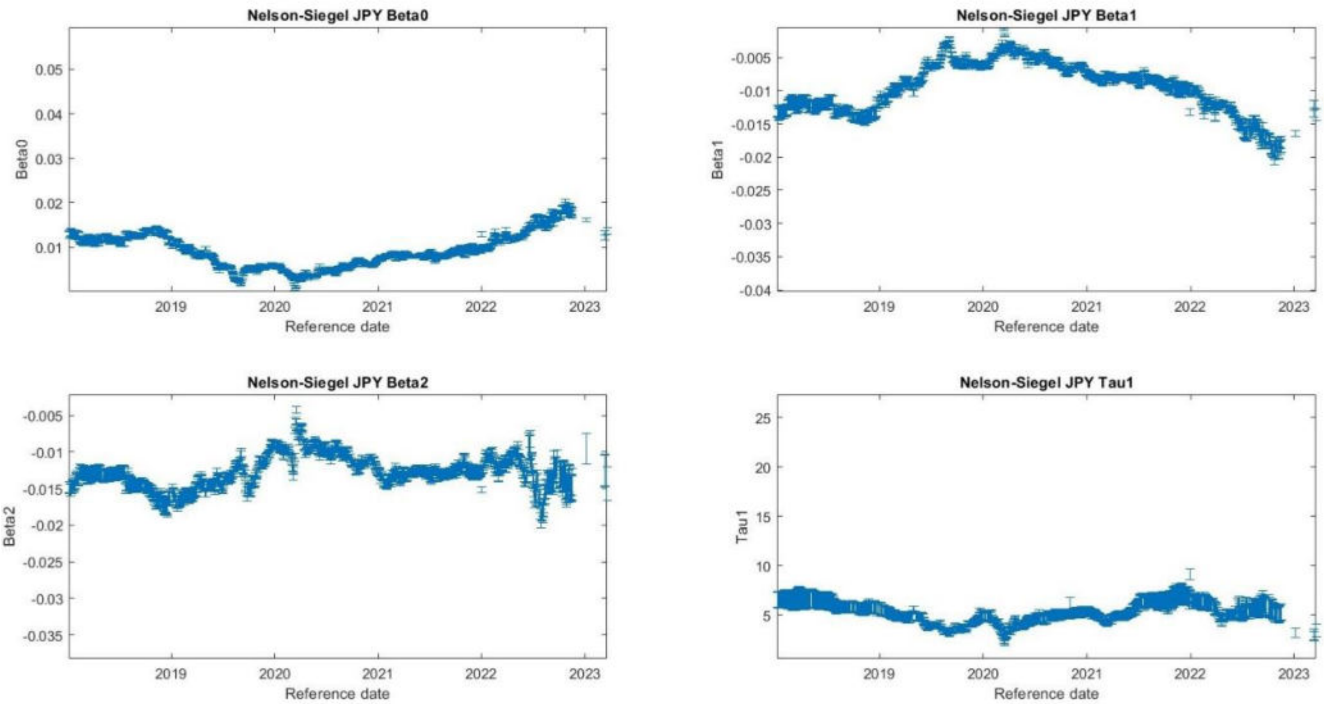


Figure 15: Nelson-Siegel coefficients for the JPY Interest rates term structures

The JPY is another interesting case, as after the procedures described above, basically only the Nelson Siegel model is used, with the GP regression used for the few points deemed as outliers.

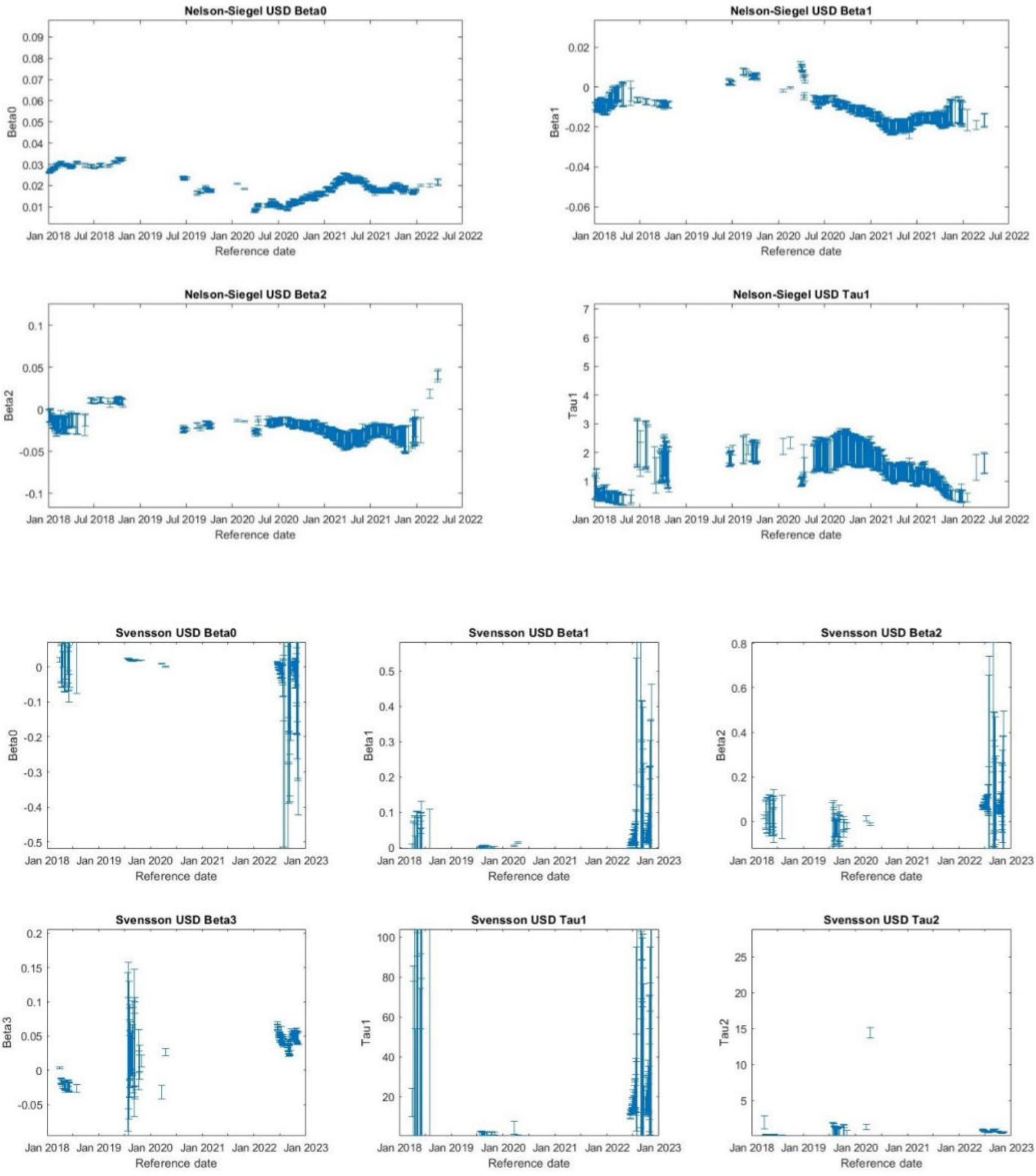


Figure 16: Nelson-Siegel and Svensson coefficients for the USD Interest rates term structures

As in the EUR case, for the USD currency, again, the results with the de Rezende model are too unstable.

After the selection process, the reference dates for which the modelling through parametric models did not give a satisfying result are represented through the Gaussian Process regression.
Given that the yield curves that failed to be modelled with the parametric approaches can assume very different shapes, the kernel choice is made automatically among a list of kernel functions, each one with peculiar characteristics that can be useful for modelling the different characteristics of the term structures. For a list of the kernel functions, see paragraph 3.

After the kernel choice, the model has been run and a 10-fold cross validation procedure has been implemented in order to check the performance of the model. There are no overfitting problems given that the MSE of the training and test sets almost match in all cases. As an example of the procedure conducted, the (10-fold) MSE of the term structure with reference date 8th February 2019, modelled through GP regression, has been plotted with respect to the iterations of the procedure (Figure 17).

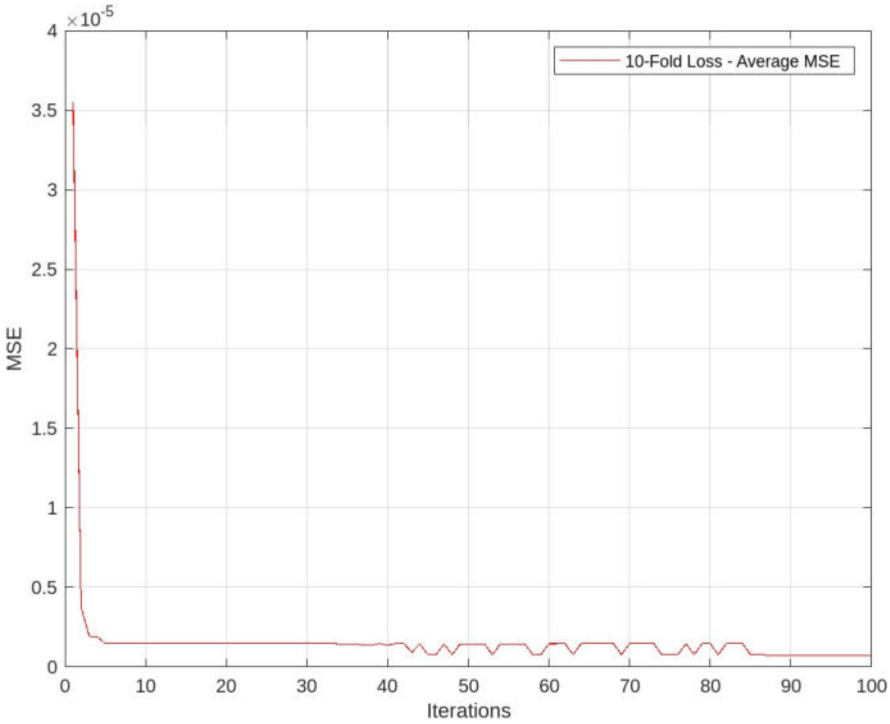


Figure 17: Coefficients for GBP Interest rates term structure

5. Main Results and Conclusions

The results computed on the parametric models are in line with evidence from other works. For example, (Svensson, 1996) and (de Rezende, 2013) in their cited works found that the (Nelson Siegel, 1987) model is the most applied, as is the case here.

Besides, the role played by Machine Learning should be highlighted. The number of models applied per currency and the overall results are displayed in Figure 18.

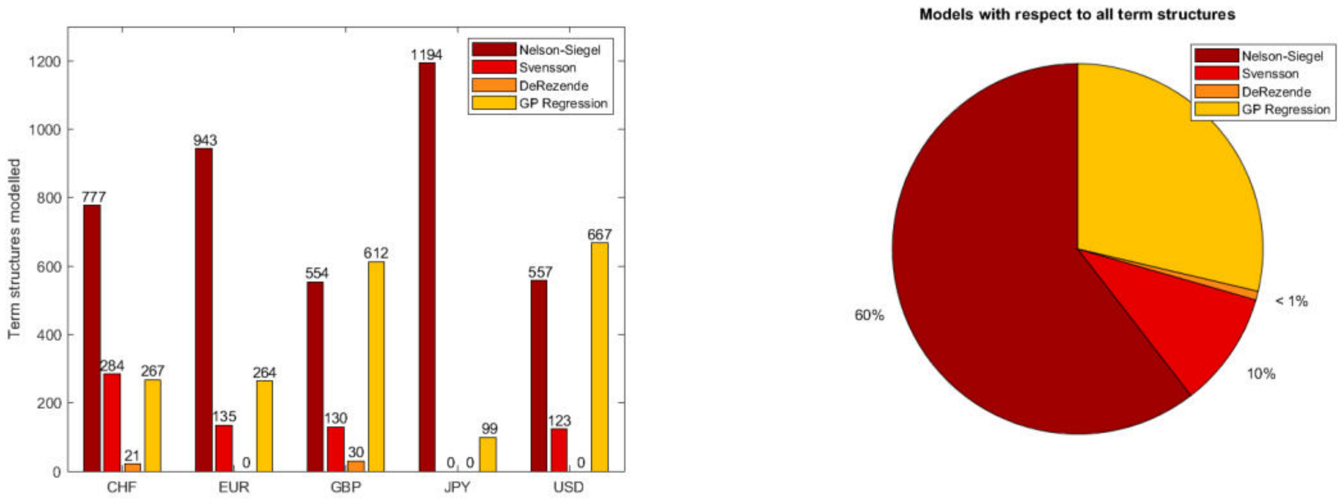


Figure 18: Models used per currency

As shown, most of the term structures have been modelled through the Nelson Siegel and the GP regression. The case of Japan is quite curious, as almost the entirety of the term structures over the reference period have been modelled through Nelson Siegel. This may also be explained observing the smoothness of the surface of the zero rates, due to a greater stability in interest rates compared to the other currencies. Another observation, deduced from Figure 18, is the similarity between USD and GBP and between EUR and CHF.

Overall, 29% of the models used are GP Regression models: it is a large figure. This may be because the reference period considers sub periods of strong turbulence in financial markets in general, and in particular related to monetary policies. This assessment seems to be confirmed if we analyse both the graphs of the coefficients and the zero rates surfaces in section 4: the period of the COVID-19 outbreak and the recent surge in inflation have led to high volatility, and it seems to have affected the capability of parametric

models to effectively model the term structure. This can be seen in the graphs (12-16) related to models' coefficients, in the white gaps (i.e. the time periods without the confidence intervals) in the first period of the pandemic and in the very last part of the reference period that coincided with high inflation and stricter monetary policies. As highlighted above, observing the graphs of coefficients (12-16), again, the case of Japan looks quite striking.
A less "qualitative" result concerns the kernel functions used in the GP regressions. As for the term structure modelling problem, Automatic Relevance Determination kernel functions appear to be better than their "standard" counterparts. Surprisingly, the Squared Exponential kernel function proves to be the less used kernel function for this kind of regression problem. These results are shown in the bar chart in Figure 19.

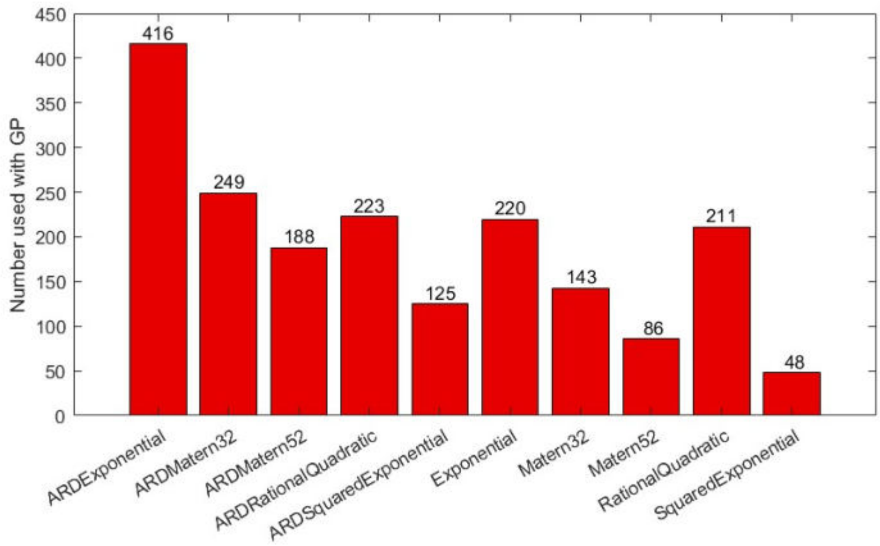


Figure 19: Kernels used with respect to GP models

In order to increase the level of clarity and explainability of the Gaussian Process output, the implementation of statistical methods similar to those applied in (Giudici and Raffinetti, 2023) and in (Giudici, Centurelli and Turchetta, 2024) should be considered for further research developments.
At the current state of this study, the gap between the prediction obtained from the Machine Learning method and the output computed by a spline interpolation of the zero rates points has been used as a potential control for the responses. If this gap were higher than a prefixed threshold, the anomaly would be reported to the analyst who can thus control and intervene in the choice of the method to be applied to efficiently solve the regression.
Another potential improvement to the model could be implemented combining different kernels by adding (and/or multiplying) them together to model the term structure, starting with those kernels that appeared as the most promising: a linear combination of ARD kernels. The reason for this suggestion is to increase the fitting potential of the regressive methodology as it would entail a higher level of adaptability compared to the one obtained from the implementation of a single kernel. The better the fitting is, the better the estimation of the discount factors and of the implied forwards rates are.

References

[1] Adams, K. J., Van Deventer D. R. (1994). “Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness”, Journal of Fixed Income, Vol. 4, No. 1, 52-62.
[2] Annaert, J., Claes, A. G. P., de Ceuster, M. J. K., Zhang, H. (2013). “Estimating the Yield Curve Using the Nelson-Siegel Model: A Ridge Regression Approach”. International Review of Economics & Finance, Vol. 27, 482-496.
[3] Bank of International Settlements (2005). “Zero-Coupon Yield Curves – Technical Documentation”, BIS Paper No. 25.
[4] Barrett, W. R., Gosnell, T. F. Jr., Heuson, A. J. (1995). “Yield Curve Shifts and the Selection of Immunization Strategies”, The Journal of Fixed Income, Vol. 5, No. 2, 53-64.
[5] Bliss, R., Fama, E. (1987). “The Information in Long-Maturity Forward Rates”. American Economic Review, Vol. 77, 680-692.
[6] Cafferata A., Giribone P. G., Neffelli M., Resta M. (2019). “Yield curve estimation under extreme conditions: do RBF networks perform better?” – Chapter 22 in book: "Neural Advances in Processing Nonlinear Dynamic Signals" – Springer.
[7] Cafferata A., Giribone P. G., Resta M. (2018). “Interest rates term structure models and their impact on actuarial forecasting” – QFW18: Quantitative Finance Workshop 2018 (UniRoma3 – Rome).
[8] Cairns, A. J. G., Pritchard D. J. (2001). “Stability of Descriptive Models for the Term Structure of Interest Rates with Applications to German Market Data”, British Actuarial Journal Vol. 7, Issue 3, 467-507.
[9] Caligaris O., Giribone P. G. (2015). “Modellizzare la curva dei rendimenti mediante metodologie di apprendimento artificiale: analisi e confronto prestazionale tra le tecniche regressive tradizionali e le reti neurali”, AIFIRM Magazine Vol. 10, N. 3.
[10] de Pooter, M. (2007). “Examining the Nelson-Siegel Class of Term Structure Models”, Tinbergen Institute Discussion Paper, IT 2007-043/4.
[11] De Rezende R.B., Ferreira M.S. (2013). “Modeling and Forecasting the Yield Curve by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach”, Journal of Forecasting Vol. 32, Issue 2, p. 111–123.
[12] Diebold, F. X., Li, C. (2006). “Forecasting the Term Structure of Government Bond Yields”, Journal of Econometrics, Vol. 130, Issue 2, 337-364.
[13] European Central Bank (2008). “The New Euro Area Yield Curves”, Monthly Bulletin, (February 2008), 95-103.
[14] Fabozzi, F. J., Martellini, L., Priaulet, P. (2005). “Predictability in the Shape of the Term Structure of Interest Rates”, The Journal of Fixed Income, Vol. 15, No. 1, 40-53.
[15] Fisher, M., Nychka, D., Zervos, D. (1994). “Fitting the Term Structure of Interest Rates with Smoothing Splines”, Finance and Economics Discussion Series, Federal Reserve Board.
[16] Giribone P. G. (2023). “Notes on Quantitative Financial Analysis”. AIFIRM Edizioni – Educational Book Series. ISBN: 979-12-80245-19-9.
[17] Giudici P., Centurelli M., Turchetta S. (2024). “Artificial Intelligence risk management”. Expert Systems with Applications, Vol. 235, ISSN 0957-4174.
[18] Giudici P., Raffinetti E. (2023). “SAFE Artificial Intelligence in finance”. Finance Research Letters, Vol. 56, ISSN 1544-6123.
[19] Gonzalvez J., Lezmi E., Roncalli T., Xu J. (2019). “Financial Applications of Gaussian Processes and Bayesian Optimization”. Capital Markets and Asset pricing eJournal. url: https://api.semanticscholar.org/CorpusID:159451782.
[20] Gurkaynak, R. S., Sack, B., Wright, J. H. (2007). “The U.S. Treasury Yield Curve: 1961 to the Present”, Journal of Monetary Economics Vol. 54, Issue 8, 2291-2304.
[21] Litterman, R. B., Scheinkman J. (1991). “Common Factors Affecting Bond Returns”. Journal of Fixed Income, Vol. 1, No. 1, 54-61.
[22] Matérn B. (1960). “Spatial Variation”. Meddelanden från Statens Skogsforskningsinstitut, Vol. 49, No. 5. Almänna Förlaget, Stockholm. Second Edition (1986), Springer-Verlag, Berlin.
[23] McCulloch (1971). “Measuring the Term Structure of Interest Rates”. The Journal of Business, Vol. 44, Issue 1, 19-31.
[24] McCulloch (1975). “The tax-adjusted yield curve”. The Journal of Finance, Vol. 30, Issue 3, 811-830.
[25] Mercer, J. (1909). “Functions of Positive and Negative Type, and Their Connection with the Theory of Integral Equations”. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 209, 415-446.
[26] Nelder J. A., Mead R. (1965). “A Simplex Method for Function Minimization”. The Computer Journal Vol. 7, Issue 4, 308 – 313.
[27] Nelson, C., Siegel, A. F. (1987). “Parsimonious Modeling of Yield Curves”, Journal of Business, Vol. 60, 473-489.
[28] Nocedal, J. (1980). “Updating Quasi-Newton Matrices with Limited Storage”. Mathematics of Computation Vol. 35, Issue 151, 773-782.
[29] Rasmussen C. E., Williams C. K. I. (2006). “Gaussian Processes for Machine Learning”. The MIT Press.
[30] Seber, G. A. F., Wild, C. J. (2003). “Nonlinear Regression”, Wiley Series in Probability and Statistics.
[31] Shea G. (1984). “Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations”. Journal of Financial and Quantitative Analysis, Vol. 19, Issue 3, 253-269.
[32] Steeley, J. M. (1991). “Estimating the Gilt-Edged Term Structure: Basis Splines and Confidence Intervals”, Journal of Business Finance & Accounting, Vol. 18, No. 4, 513-529.
[33] Svensson, L. E. O. (1994). “Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994”. IMF Working Paper, WP/94/114, 1-49.
[34] Svensson, L. E. O. (1996). “Estimating the Term Structure of Interest Rates for Monetary Policy Analysis”. Scandinavian Journal of Economics, Vol. 98 (1996), 163-183.
[35] Vasicek O. A., Fong H. G. (1982). “Term Structure Modeling Using Exponential Splines”. The Journal of Finance, Vol. 37, Issue 2, 339-348.
[36] Waggoner, D. F. (1997). “Spline methods for extracting interest rate curves from coupon bond prices”. Working Paper, No. 97-10, Federal Reserve Bank of Atlanta.

Data Analytics for Credit Risk Models in Retail Banking: a new era for the banking system

Adamaria Perrotta (UCD University College Dublin), Andrea Monaco (UCD University College Dublin), Georgios Bliatsios (UCD University College Dublin)

Corresponding author: Adamaria Perrottta (adamaria.perrotta@ucd.ie)

Article submitted to double-blind peer review, received on 21st November 2023 and accepted on 13th December 2023

Abstract
Given the nature of the lending industry and its importance for global economic stability, financial institutions have always been keen on estimating the risk profile of their clients. For this reason, in the last few years several sophisticated techniques for modelling credit risk have been developed and implemented. After the financial crisis of 2007-2008, credit risk management has been further expanded and has acquired significant regulatory importance. Specifically, Basel II and III Accords have strengthened the conditions that banks must fulfil to develop their own internal models for estimating the regulatory capital and expected losses. After motivating the importance of credit risk modelling in the banking sector, in this contribution we perform a review of the traditional statistical methods used for credit risk management. Then we focus on more recent techniques based on Machine Learning techniques, and we critically compare tradition and innovation in credit risk modelling. Finally, we present a case study addressing the main steps to practically develop and validate a Probability of Default model for risk prediction via Machine Learning Techniques.

Keywords: Credit Risk Management, Risk Prediction, Machine Learning, Loan Defaults.

1. Introduction

Estimating the risk profile of clients is one of the most important activities in the banking sector. The number and variety of methodological approaches to model credit risk in the banking system has strongly grown in the last 20 years, due to the increasing relevance that risk has for firms' business models. Indeed, a strategic, central role has been given to credit risk for the banking system by the Regulator. Therefore, large internal teams of specialists have been tasked with the development and validation of sophisticated credit risk models for default prediction subject to well established regulatory frameworks. Such frameworks are being created by Central Banks and other banking supervisory entities and are usually the product of forums where these authorities meet and exchange ideas. An example of such a forum is the *Basel Committee on Banking Supervision (BCBS)* which was formed in 1974.
In 1988, the BCBS issued the first Basel Capital Accord (Basel I) which had the main objective of ensuring that all financial institutions operating in an international environment hold enough cash reserves. In 1999, the same entity published several proposals for changes to the current regulations to reinforce and level out the risk management process. Such proposals heavily contributed to the New Capital Agreements, known as Basel II and Basel III accords (see (Bank for International Settlements, 2005a) and (Bank for International Settlements, 2011)). In (Bank for International Settlements, 2005a), the Committee created a robust risk management framework by introducing the two main approaches currently in us to determine the minimum regulatory capital requirements for credit risk.
The first approach is known as *Standardized Approach,* and it allows banks to use prescribed estimates for the key risk factors involved in the calculation of the Risk Weighted Assets (RWA). The second and more common approach is known as *Internal Rating Based (IRB) Approach* and it permits banks to use internally developed models for estimating the RWA. Of course, the IRB is subject to regulatory approval to be put in use. The IRB approach can be further divided into two sub-approaches; one referred to as *Foundation – IRB Approach (FIRB)* and the other one as *Advanced - IRB Approach (AIRB)*. Within the AIRB approach, banks can use their own models to estimate the likelihood that a credit obligation will not be met while the remaining risk factors are prescribed by the Regulator. In view of the AIRB approach, all risk factors can be estimated using internal models. Credit institutions which are allowed to apply the FIRB or AIRB usually combine quantitative and qualitative techniques in their estimation method to get the best risk predictions. The key credit risk factors for which a significant number of models is developed on a regular basis are the Probability of Default (PD), Exposure at Default (EAD) and Loss Given Default (LGD); they contribute to the estimate of the RWA (Bank for International Settlements, 2005b) and the Economic Capital (Bank for International Settlements, 2009). Furthermore, under the *IFRS 9 Regulatory Framework* (Bank for International Settlements, 2017) the same risk factors are computed via different estimation methods and used in the calculation of *Expected Credit Loss* (ECL) (EBA/GL/2017/06, 2017) and provision management.
In addition to the new Regulation, another driver for the growth of credit risk modelling is the increased capability of exploiting, collecting, and processing of large dataset of portfolio debtors. Such information allows us to detect the underlying credit risk dynamics which have been completely ignored until now. Obviously, since credit risk models are predominantly data driven, the underlying training dataset should be of the best possible quality to ensure the robustness of results and reliability of the estimate (European Central Bank, 2019). Moreover, since the availability and the distribution of the data can heavily influence the design of a model, it is important to ensure that a set of rules and policies regarding the collection, structure, cleansing, transformation, storage, and assessment of the available dataset is in place prior to commencing model development. In particular, these rules can offer guidance with respect to how to handle missing values, when and if exclusions should be applied, how to treat outliers and several other controls depending on the nature of the data.
In literature, a big number of classification algorithms for borrowers' creditworthiness assessment have been proposed, jointly with some comparison studies (see as an example (Baesen, et al., 2003) and (Lessmann, Baesens, Seow, & Thomas, 2015)). Assuming that a sufficiently large data set is in place, based on the nature of the risk factor and the regulation governing the model's design, a wide range of statistical tools is available for developing credit risk models. Such tools include, but are not limited to, Logistic Regression, Generalized Linear Classifiers, Random Forests, Time Series Analysis, Bayesian Inference, Artificial Neural Networks, and hybrid methods which combine various continuous and/or discrete probability distributions.
Once the risk model has been developed, its predictive ability and discriminatory power must be assessed and monitored on a regular basis. This can be done through various validation techniques which, depending on the nature of the risk factors and the development approach, might include the use of confidence intervals, root mean square error, confusion matrices, accuracy ratio, receiver operating

characteristic curve (Irwin & Irwin, 2012), Kendall's Tau and other statistical indicators. Additionally, the model performance can also be assessed by benchmarking its outputs against the ones produced by alternative models and/or against historical observations.
The purpose of this contribution is to critically review the most used model development and validation procedures for credit risk in retail banking. In Section 2 we focus on the impact of Regulator on Credit Risk Methodologies and briefly motivate the introduction of ML approaches for regulatory capital estimation. In Section 3 we review some of the traditional statistical methods to PD and LGD estimates. Then in Section 4 we focus on more recent studies based on Artificial Intelligence and Machine Learning techniques. We conclude the paper with a case study, presented in Section 5, showing some of the basic and practical steps performed in the financial industry to develop and validate a PD model via Machine Learning Techniques. Conclusions will be discussed in Section 6.

2. The impact of Regulator on Credit Risk Methodologies

As pointed out in the Introduction, within Basel's Accords the Regulator defined a general framework of assumptions and models to estimate credit risk.
The first set of requirements to estimate credit risk is the one known as Basel I Accord (1998): in such agreement, the Regulator stated the definition of *minimum capital requirement* in terms of *Risk weighted assets*. In Basel I framework, the regulatory capital was computed estimating unexpected portfolio losses, while credit risk evaluation was modelled through the combined effect of a systematic and an idiosyncratic default rate risk factor. The Regulator also required financial institutions to customize the credit risk model with respect to the risk profile of the portfolio's constituents, defining five possible categories: corporate, sovereign, bank, retail, equity. Finally, the Basel Committee introduced in credit risk modelling the so-called through cycle PD: the average default rate performance for a particular customer over an economic cycle.
Basel I approach has been heavily criticized for being insufficiently granular, so the Basel II Accord (2004) provided a more refined methodology proposing two possible approaches:

- *the Standard Approach*: it is based on a simple categorization of debtors, without considering their actual credit risks. This approach relies on external credit ratings.
- *the Internal Ratings-Based (IRB) approach*: within this approach, financial institutions are allowed to use internal models to compute the regulatory capital requirement for credit risk.

Basel II accords set the reference parameters (i.e., the parameters of the IRB approach) that any credit risk model must estimate to compute the minimum capital requirement. Thus, each institution had to independently define the levels of the PD, LGD and EAD, starting from its own score models. Basel II also reinforced the validation process setting up new rules:

- Banks must have a robust system to validate the accuracy and consistency of rating systems, processes, and the estimation of all relevant risk components.
- A bank must demonstrate to the Regulator that the internal validation process enables it to assess the performance of internal rating and risk estimation systems consistently and meaningfully.

Basel I and II methodological framework constitute a constraint that has strongly influenced the development of credit risk modelling since any possible change/improvement must be in line with the Regulation to be effectively implemented. The Basel framework mainly relies on the estimate of three key risk factors: PD, LGD and EAD since they constitute the main ingredients to compute the minimum capital requirement.
The traditional methodological approach to credit risk is based on the explicit modelling of these quantities. However, with increased computing capabilities and access to larger amounts of data, the current literature describes many improvements upon simpler traditional risk models, which are limited in their ability to incorporate big data. The availability of large data sources and a higher level of info details allows both to increase the accuracy of traditional models to better fulfil regulators' requirements and to apply Machine Learning techniques to analyze and forecast credit risk. For this reason, there is increased interest in the development of machine learning models that can capture nonlinear relationships between variables of a dataset (Sadhwani, Giesecke, & Sirignano, 2021). However, ML models have their drawbacks. They can be computationally expensive and difficult to interpret; moreover, they can bring to ethical tradeoff (Lee & Floridi, 2021). In addition to that, ML approaches are distant from those covered by the current regulatory framework. This is the reason why the switch from traditional approaches to the full use of ML algorithms to risk decision-making is still far from the implementation.

3. Traditional Approaches to Credit Risk Analysis

3.1 Traditional Approaches to PD Models

The large variety of PD models used by banks to estimate credit risk can be classified into two main categories: *Structural Models* and *Reduced Form Models* (Duffie & Singleton, 2012). *Structural Models* estimate the probability of default starting from assumptions about the value of its assets and liabilities. The basic idea is that a company defaults if the value of its assets is less than the debt of the company. Conversely, the *Reduced Form Models* assume default as an external cause regardless of the value of the assets/liabilities of the company. In this section we will present an overview of the principal Structural and Reduced Form Models used in credit risk management.

3.2 Structural Models

The class of structural models identifies the family of models that describes the dynamics of credit event with the structure of a company's debt: the event of default occurs when the market value of a company's assets $A(t)$ falls below a threshold set by the nominal value of the debt $D(t)$. According to this model, the default event will depend on the dynamics of the assets $A(t)$ and the dynamics of the default barrier level represented by debt $D(t)$. The default event occurs at the time τ when $A(\tau) < D(\tau)$ and in this case

the equity value is equal to zero. Therefore, within the structural models setting the default event is translated into a barrier event estimation problem: we should estimate the probability of the default event in the time horizon [0, T], where T is the maturity of the debt, that is $P[A(\tau) < D(\tau) \mid \tau \in [0,T]]$.
The first example of structural model is the well-known Merton model (Merton, 1974). The basic assumptions of the Merton model are:

- *A(t)* follows a log-normal stochastic distribution.
- The default is evaluated on a fixed time horizon T.
- *A(t)* has both an equity component *E(t)* and a debt component *D(t)*: *A(t) = E(t) + D(t)*.
- The debt *D(t)* is represented by a zero-coupon bond with maturity T.
- the default can occur only at maturity T of the debt, and it happens if *A(T) <D(T).*

The Merton model can be easily applied if we agree on its strong hypotheses. This model provides in fact an explicit formula for risky bonds (Merton, 1974). Such formula can be used to estimate the PD of the firm as well the credit spread structure of the risky bond. However, the model's "oversimplified" hypotheses are responsible for some limitations in its use. Firstly, the Merton model simplifies the liabilities' structure assuming the debt as a single zero-coupon bond. This hypothesis allows the model to assess the occurrence of the default only at maturity T. Then, the "oversimplified" dynamics for the asset liability is responsible for a zero value of credit spread at very short maturities. Such a prediction is not confirmed by market observations. In conclusion, it is more than clear that some improvements to the Merton model are needed to get a more realistic estimate of PD.
Under this light, there are different models that extend the Merton model. Each of these models improves the original Merton framework by removing some assumptions; in the following we report the most representative ones.

- **The Black-Cox model** (Black & Cox, 1976). This model extends the Merton model by assuming that the default can occur before the maturity of the liability T. According to this model, the default barrier is a function of time *H (t)* of exponential type. This model belongs to the larger family of *first hitting time models*.
- **The KMV model** (Bharath & Shumway, 2006). The KMV model improves the Merton model exploiting historical data on defaults. This model is based on historical estimation of the probability of default, allowing to simulate the default over different time horizons. A big advantage of the KMV approach is the high accuracy of PD estimates in case of extreme events simulation. This great performance is due to the historical calibration of the model.
- **The Credit Grade Model** (Finger, et al., 2002). This model extends the structural models introducing a stochastic dynamic to the debt/barrier. The model in fact attributes a log-normal stochastic dynamics to both the asset *A(t)* and the recovery rate *RR(t)*[1], resulting in a stochastic dynamic also for the debt *D(t).* The model also allows us to estimate the survival probabilities in closed form therefore one can calibrate the model on the level of market spreads to market data. This model, as well as the KMV, has a level of accuracy in the simulation of extreme events much more accurate than the Merton model.

Despite such attempts to potentiate the Merton Model performances, the main drawbacks of this class of model are still present. Those models require to estimate the firm's asset value, which is non-observable. Moreover, structural-form models cannot incorporate credit-rating changes that occur quite frequently for default-risky corporate debt. These limitations can be partially removed by adopting a *Reduced Form* approach to model PD.

3.3 Reduced Form Models

The class of Reduced Form models, also known as *intensity models*, describes the process of default through a minimum set of hypotheses. The Reduced Form models are particularly suited to model credit spreads and their basic formulation makes them easy to be calibrated on corporate bond data or Credit Default Swaps (CDS)[2]. Given a time interval $[t, t + \Delta t)$ and a default time τ, the probability that time τ falls within of the interval $[t, t + \Delta t)$ is equal to $\lambda(t)\Delta t$, where $\lambda(t)$ is generally named *hazard* or *intensity rate.* Since Reduced Form models assume that PD ~ $\lambda(t)\Delta t$, then the default event follows a Poisson process. Moreover, since the probability of default is completely specified by the function $\lambda(t)\Delta t$, the Reduced Form models differ between each other in relation to different hypotheses made on $\lambda(t)$. We can consider three sub-classes of Reduced Form models:

- **The Time Homogeneous models**: the intensity of default $\lambda(t)$ is deterministic and constant over time. Therefore, the dynamics of credit spreads results constant and deterministic. Within this sub-class it is immediate to infer the survival probability within a fixed horizon from the level of market spreads CDS. Its main limitation is that Time Homogeneous models are not able to describe the complex structure of credit spread, implicit in quoted instruments.
- **The Time Variant models**: this sub-class of Reduced Form models is based on the hypothesis of deterministic intensity but variable over time. Very often this family of models is specified in terms of cumulative intensity. This sub-class allows to reproduce, with more accuracy respect to time homogeneous models, credit spread dynamics implicit in quoted financial instruments.

[1] Recovery rate is the extent to which principal and accrued interest on defaulted debt can be recovered, expressed as a percentage of face value. The recovery rate enables an estimate to be made of LGD = 1 – RR.
[2] A credit default swap (CDS) is a financial derivative that allows an investor to "swap" or offset his or her credit risk with that of another investor.

- **The Stochastic Intensity models**: the intensity of default $\lambda(t)$ is stochastic (as an example, a CIR process). These models allow us to consider the market volatility of credit spreads. The default intensity dispersion can be obtained from credit spread options market prices or from CDS time series analysis.

The Reduced Form models family was originated by Jarrow and Turnbull (see (Jarrow & Turnbull, 1992) and (Jarrow & Turnbull, 1995)), and Duffie and Singleton (Duffie & Singleton, 1999). This family of models has been intensively studied in literature (see (Duffie & Singleton, 2012) and (Elliott, Jeanblanc, & Yor, 2000)). Some authors point out an intrinsic connection between Structural and Reduced Form models; Reduced Form models are intended as Structural Models in a different information filtrations: structural models are based on the firm's management information, while Reduced Form models are based on the information available on the market see (Duffie & Lando, 2001) and (Jarrow & Protter, 2004)).

3.4 Traditional Approaches to LGD models

Reduced Form models introduce separate explicit assumptions on the dynamic of the PD and the recovery rate RR. Despite these two variables - PD and RR - are mostly independently modelled, in literature they did not receive the same scientific attention. There are few studies on RR dynamics compared to the PD ones. The reason for this asymmetry relies on the basic assumption of credit risk models. Traditionally, credit risk models assume that RR depends on individual features like collateral or seniority and is not influenced by systematic factors; therefore, it has been considered as independent of PD.
Most recently, Basel Regulation has pointed out the relevance of this indicator, so RR modelling has attracted the interest of analysts and researchers. According to the Basel Accords, in fact, the recovery rate is used in the capital requirement formulas in a linear way. Its estimate is therefore crucial to model LGD with high accuracy. Moreover, the possibility provided by the IRB approach to set the LGD values tailored to bank's portfolios has increased research works on RR. To build LGD models some phenomenological peculiarities need to be considered. In particular, the observed historical distributions of RR are bimodal: the recovery rates are concentrated either in high values (around 70-80%) or low ones (around 20-30%). Moreover, RR values are strictly dependent on the industrial sector of obligors: tangible asset-intensive industries, especially utilities, have higher recovery rates than service sector firms, with some exceptions such as high tech and telecom. In the remaining part of this Section we will provide an overview of the different approaches to estimate LGD and of the prevalent literature available for modelling purposes.

3.4.1 LGD Estimate

There are three main different approaches to computing LGD: Market LGD, Workout LGD, Implied Market LGD.

- **The Market LGD**: This approach is based on market sentiment. Even once the default occurs, defaulted bonds and loans can still be traded on the market. Therefore, their prices include investor expectations on the entire recovery process: capital recovery of the restructuring costs and the related uncertainty.
- **The Workout LGD**: This approach is based on the historical analysis of defaulted loans to predict future values of LGD rates. All cash flows generated by the recovery process, all recoveries, as well as all costs are taken into account in the period ranging from the day of the credit event to the final recovery. These cash flows must be discounted; however, it is not obvious which discount rate needs to be applied, in case of debt restructuring actions via the issuance of risky assets such as equity or warrants.
- **The Implied Market LGD**: This approach is based on the analysis of market prices of bonds or loans before default. The spread between a loan-specific interest rate and the risk-free interest rate is equal to the expected loss thus the LGD value is deduced from the ratio between the spread and default probability.

In the case of bonds, the estimate of the LGD is less straightforward. The spread above risk-free rate is an indicator of the risk premium demanded by investors to price PD and LGD, as well as liquidity premiums.

3.4.2 LGD Modelling

Due to the complex phenomenology underlying the recovery process as well as the existence of three different approaches to compute LGD, there are several methods available in literature to model recovery dynamics. Here we briefly describe the most used ones.
In Frye's structural framework it is addressed that the value of LGD is affected by the PD level and therefore it cannot be modelled independently from it (see (Frye, 2000a) and (Frye, 2000b)). Such studies had a huge impact on regulatory practice since they suggested changes in guidelines of Basel accords. Indeed, in the model proposed by Frye in (Frye, 2000a) and (Frye, 2000b), defaults are driven by a single systematic factor: the correlation between the RR and PD, which derives from their mutual dependence on the systematic factor. However, the distribution of RR is shown to be different in high-default periods from low-default ones. The negative correlation between default rates and RR relies on the dependence of loans collateral from systematic factor. If the economy experiences a recession, then the default rate increases while the value of collateral decreases as well as the associated RR.
In a similar way to Frye's approach, the Jarrow model (Jarrow R. A., 2001) correlate both RR and PD to the state of the macroeconomy. The model introduces the liquidity premium and equity prices in a calibration procedure where RR and PD are explicitly separated.
Jokivuolle and Peura in (Jokivuolle & Peura, 2000) propose a model where the collateral drives the recovery dynamics. According to such a model, the collateral value is correlated with the PD while the credit event is triggered by the total asset value.
Tasche in (Tasche D. , 2004) introduces a single risk factor model, where LGD volatilities can be statistically estimated. Moreover, the model considers defaulted obligors. This model allows us to compute the capital charges according to Basel definition with an accurate numerical approximation, and it is one of the reasons why it is heavily implemented in banking practice.
It is important to underline that all these models consider the different characteristics of the obligor by segmenting the portfolio in terms of default period, loan-to-value ratio, customer type, credit score, etc.; these factors may be different according to the portfolio under analysis. Obviously, the setup of LGD models based on obligors' information benefits from the increases of available

information on obligors. This is the reason why in the last few years the use of ML techniques for LGD modelling, such as Regressions and Neural Networks approaches, has hugely increased. Consequently, several research studies have been conducted to address the validity of these approaches to PD and LGD modelling.

4. Machine Learning Methodologies to Credit Risk Analysis: an overview

4.1 Main Applications in Credit Risk: pros and cons

In recent years, the possibility of collecting and storing big datasets as well as the increase of computer performances provided the opportunity to increase the use of ML techniques in finance. Indeed, while conventional econometric methods fail to exploit nonlinear relations between features and hidden information deduced by unstructured data sources, ML models allow detecting peculiar patterns and dependencies from these new datasets. Thus, ML techniques promise to make data analysis for managing financial risk more efficiently. On the other side, this gain is not a "free lunch". Large financial datasets are in fact characterized by increased noise and nonlinear patterns, so they infer significant statistical challenges to modelling.
Financial institutions have developed new systems based on ML to drive expert decisions in the domain of credit risk modelling. The most common use of ML in credit risk is in credit decisions/pricing, followed by credit monitoring and collections. These techniques are mainly used by large firms due to their benefits of scale, access to data and large resources. The most active sector of ML applications is risk management, compliance, customer engagement and credit. In particular, the main applications ML techniques in credit risk management ca be synthesized as follow:

- Model validation: ML models are used as a benchmark to the standard model for capital requirements calculation.
- Data improvements: ML techniques can be used to improve data quality allowing us to clean, pre-process and analyze rich datasets.
- Variable selection: ML techniques allow to detect explanatory variables with useful predictive capacities within large datasets.
- Risk differentiation: compared to traditional PD model, ML models increase the differentiation of risks while computing the probability of default.

One of the main advantages of using ML for risk analysis is the improvement of risk differentiation. ML techniques may increase the discriminatory power of model allowing to identify risk drivers with better accuracy compared to traditional models. Therefore, ML models are used to optimize portfolio segmentation and take data-driven decisions. Moreover, ML can be used in a hybrid modality, confirming the selection of data features used in traditional models.
While ML techniques improve quantitative performances of credit models, the choice of appropriate economic theories and assumptions supporting the ML methodology could constitute a challenge. Indeed, applying IRB models requires understanding and interpreting underlying model dynamics, which is not always so straight: this could obviously prevent the use of ML techniques, despite their great performances.
Additionally, new ML models allow us to base credit score predictions on a broader range of variables than those traditionally included in the classic statistical models (Sadok, Sakka, & El Hadi El Maknouz, 2022). However, financial analysts must understand whether predictions based on big datasets could make credit available to individuals or companies that were previously considered ineligible using traditional methods. This is a very important item for driving right decisions, since a ML application could cause unintended negative consequences for the banking system, causing economic instability if not implemented with care (Eitel-Porter, 2021). Finally, it is important to investigate how the use of ML in credit risk models affects issues surrounding ethics choices, bias and discrimination in the lending market, because this could heavily contribute to housing inequality and racial disparities among minorities (Zou & Khern-am-nuai, 2022).

4.2 A literature review of ML in Credit Risk

For completeness and critical understanding of this research work, we have performed an extensive literature review to select the principal ML techniques used for credit risk assessment. In the following we present the outcomes of such selection. From now on, we will use the word "feature" to refer to the variables of a financial dataset. In literature, the word "feature" is equivalent to "factor" or "variable" or "characteristic". Features are the basic building blocks of datasets. The quality of the features in a dataset has a major impact on the quality of the insights one will gain, especially when ML algorithms are employed.

- **Linear Regression and Logistic Regression**: given a set of observations (i.e., the dependent variable or regressand) and some feature map of explanatory variables (i.e., regressor), linear regression set the best parameter configuration able to minimize residual errors of a linear objective function. Linear regression takes continuous inputs (e.g., profit margin, efficiency ratio, cash ratio, debt ratio, earnings per share, etc.) and outputs a continuous variable. Logistic regression uses a logit function to model extend linear regression to a binary dependent variable. These methodologies are used to forecast a company's financial distress (credit scoring) and then its PD level, taking continuous input related to the company profile (see (Altman, 1968), (Orgler, 1970), (West, 2000)).
- **Artificial Neural Network**: this technique is inspired by the functioning of the human brain. Individual neurons are modelled by logistic regression therefore the neural network is a multiple layer infrastructure that connects logistic regression classifiers. These algorithms are designed to recognize patterns and make predictions involving a large number of parameters. Main application of this methodology is the credit risk analysis and forecasting of borrower's credit profile dynamics (see (Hsieh, 2005), (Abdou, Pointon, & Elmasry, 2008), (Angelini, Tollo, & Roil, 2008), (Pang & Gong, 2009)).

- **Support Vector Machine**: this method allows to model nonlinear classification problems. The technique identifies hyper-plane multidimensional surfaces to separate classes of dataset. The boundary decision made by SVM is accurate compared to other techniques. However, due to the use of a kernel function it is not easy to attribute each prediction to an individual variable. SVM in credit applications is a supervised learning methodology that analyze data to perform credit risk scoring (see (Baesen, et al., 2003), (Huang, Chen, Hsu, Chen, & Wu, 2004), (Schebesch & Stecking, 2005), (Shin, Lee, & Kim, 2005)).
- **K-Dimensional Tree**: this technique allows solving classification problems using a binary tree structure (i.e., a sequence of nodes and branches), which sub-divides dataset by a hyperplane. Most of the effort to apply the method relies on the setting of the tree structure. It is quite flexible, allowing us to handle any probability distribution and non-linearity in the model. The usability of the method relies on a rich out-of-sample dataset, in fact one of the main limits of the technique is the risk of data over fitting. The method is often used to classify the credit profile of a company (Breiman L. , 2001).
- **Decision Tree**: the method allows to implement "if-and-else" questions to solve a specific classification problem. The Decision Tree has a binary tree structure like a K-D Tree, the model prediction is obtained through a sequence of nodes and branches that allow to easily interpret the decision-making logic. In a decision tree the size of the tree can grow to adapt to the complexity of the classification problem. The methodology is applied to forecast company profile conditioned to info and events using categorical and continuous inputs (Galindo & Tamayo, 2000).
- **Random Forest**: this technique can be classified as an ensemble learning method based on multiple decision trees. The decision trees are randomly generated in an iterative mode allowing to obtain a forest. The classification result is defined as the class selected by most of the trees. Random forests use the same training set to train multiple decision trees allowing to reduce the variance of the output. This method as the previous mentioned ones allows to classify credit profile of a company (Breiman L. , 2001) .
- **Boosting**: this technique can be classified as an ensemble method to reduce bias and variance in supervised learning. The method is based on the exploitation of ensemble characteristics of models to increase accuracy of forecast/description performances compared to the use of a single methodology: convert weak learners to strong ones. Boosting algorithms iteratively train single classifiers with respect to a distribution adding them to a final strong classifier. Once added the classifiers are re-weighted to increase accuracy. There are many boosting algorithms, these mainly differ by method of weighting training data (see (Amin, Islam, & Murase, 2009), (Nanni & Lumini, 2009), (Yu & Wang, 2009)).

5. Probability of Default: Elements of Model Development and Validation Within the Framework of Machine Learning

In this section we present a case study showing the development and validation process of a Probability of Default PD model via ML techniques. Before focusing on a specific ML method, we want to describe the main operative steps to perform in *any* ML method for risk management. This would facilitate the understanding of the case study and will shed light on how ML algorithms are changing the model development and validation in the banking sector.

Applying ML techniques in credit risk results in performing a multi-step process, where the different stages are:

1. Pre-Processing of the dataset: this step refers to any type of processing performed on raw data to prepare it for another data processing procedure.
2. Features Selection: in this step the analyst selects the most relevant variables to forecast the PD.
3. Data Modelling: in this step the analyst chooses the ML model to implement and the specific dataset to calibrate the model parameters.
4. Model Validation: once a model has been calibrated and implemented, the model performances must be tested in different market regimes.
5. Model Deployment: If the model has passed steps 1-4, it goes to production. At this stage potential operating risks must be managed.

As already pointed out in the previous sections, ML algorithms relies on large datasets; in particular, usually three different samples are introduced to set, validate, and finally test the model:

1. Training Sample: this dataset is used to calibrate the model parameters (step 3 above).
2. Validation Sample: this dataset is used to determine the values of the parameters of the model (step 4 above).
3. Test Sample: this dataset is used to measure the performance of the model (step 4 above).

In the rest of the paper, we present a case study showing how to implement steps 1-5 above to develop a PD model via ML techniques. To this aim, we built an artificial dataset of 1,000 home mortgages with 25 variables of account, demographic, and sociological nature (see Table 9 in the Appendix section for details).
The dataset was created in a way to mimic a real word database: we couldn't avail ourselves of a real one since banks don't provide actual obliger's information due to data protection regulations unless the research has been conducted with the financial institution.

To make the case study as authentic as possible, the variable names follow the format presented in the ECB's Loan-Level data templates[3].
Among the above-mentioned ML methods for credit risk assessment, we decided to compare a linear with a non-linear classifier: Logistic Regression (LR) vs k-nearest neighbor (k-NN). We employed Logistic Regression (LR) since on the one hand, the target variable **Default Indicator** (see Table 9) is of a binary nature and, on the other hand, LR is a well-established, simple to interpret and the most widely used approach in the banking sector. Among the non-linear classifiers, we have referred to k-NN since it is a relatively simple approach that is completely nonparametric. There are only two choices a user must make: the number of neighbors k and the distance metric to be used. We referred to the most common choice of distance metrics, i.e., the Euclidean distance. Validation techniques for measuring the predictive strength of the model are also being discussed.
We begin this section by introducing the raw data, the preprocessing of the dataset and the feature selection procedure which will lead us to the final dataset used for developing the model (Section 5.1). Then, we discuss the model fitting process and present the resulting model coefficients (Section 5.2). Although the calibration of the rating system process is out of scope from our case study given that we are developing the model on a simulated dataset, we will describe some of the basic methodological steps which are involved in this process. We conclude this section with the description of the validation procedure aiming to measure the performance of the model (Section 5.3).

5.1 Data Preprocessing and Features Selection

In this section we perform an initial data analysis and cleaning of the raw data as well as a features selection to get the final set of features for developing the PD model. In the Appendix – Table 9 - the final set features have been highlighted in bold. To address the borrower's default prediction problem, we define our target variable as Default Indicator, a numeric binary variable, describing whether the obligor is in default (1) or not (0). As already mentioned above, the dataset is the most important component of any ML model since the robustness of its output is strongly related to the quality of the inputs used for training. Therefore, several quantitative and qualitative checks are required to assess whether the dataset is fit for purpose. Thus, a preliminary step to derive useful insights regarding the distribution of the data and potential relationships among the variables is a *visual data exploration*. For example, in our dataset there is a downward trend regarding the number of defaulted borrowers with respect to the variable "Age" (see Figure 2). Therefore, one can infer that default rate and age appear to be negatively correlated and such relationship is expected to be reflected in the model, assuming that these variables are chosen for development purposes.
The next step is to control potential missing data and the presence of outliers among the variables. It is worth noting that there can be a few valid reasons for missing values for some variable and therefore when such a phenomenon is present, qualitative judgement should be employed. In case there is more than 5% of missing data for a given variable a decision whether to replace or keep the missing values needs to be made. For the features having less than 5% of missing rate, we decided to perform data imputation by substituting the nulls with the most frequent value of each feature (see (Siddiqi, 2017), (Joenssen & Bankhofer, 2012)). In case of replacement, a different choice is made in relation to the type of the variable: the mode, the mean or the median of the sample are potential candidates for missing numeric data while for nominal data types, the most frequently occurring values could be used (Siddiqi, 2017). When missing rates are higher than 30%, the variable has been dropped from the modelling dataset (Siddiqi, 2017).

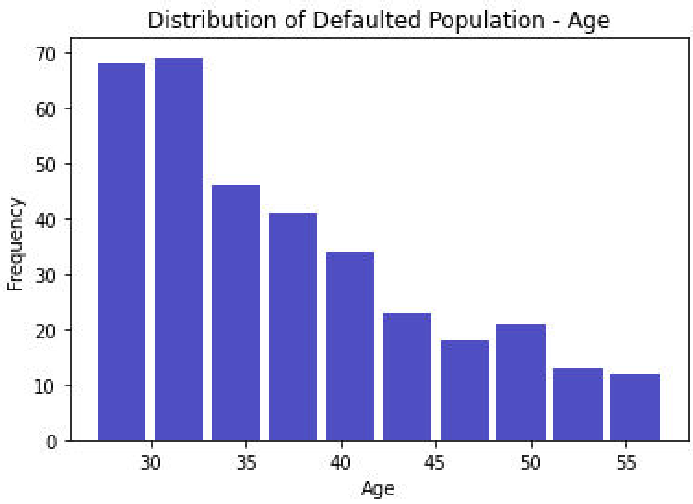


Figure 2: Distribution of Defaulted Population - Age

For what concerns the numerical features, the presence of outliers, i.e., observations that are significantly different from most of the data points in a set, can have negative impact on model training.
Therefore, their detection and treatment are important as well. There are several methods for dealing with outliers. In our case study, we now discuss the *Z-scores* and *Inter Quartile Range (IQR)* fences, also known as Tukey's fences.
Assuming a set of observations $X = \{x_1, \dots x_n\}$ has mean $\bar{x}$ and standard deviation σ, the Z-score of the observation $i \in \{1, \dots n\}$ is defined as:

$$Z_i = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

[3] https://www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html

By construction, the distribution of the Z-scores is standard normal *N(0,1)*. Within the Z-scores approach, the common practice is to classify as outliers those observations with absolute value greater than three.
The IQR approach measures the dispersion in the dataset by calculating the difference between the third (Q_3) and the first (Q_1) quartile of the data's distribution. The formula for calculating the Tukey's fences is given by:

$$[Q_1 - kIQR, Q_3 + kIQR] \quad (2)$$

where $IQR = Q_3 - Q_1$ and k is a positive constant. In our case study, we refer to the IQR method to detect outliers and we choose k=1.5, since this is the most common selection in banking practice. Once detected, there are several approaches for treating outliers including deleting, capping/flooring, replacing, or transforming (e.g., logarithmic transformation) the data. In our example, we chose to floor and cap the outliers in each case by using the lower and upper bounds of the Tukey's fences given in eq. (2). In Table 1 we present the number of available data, missing and outlier rate per variable in our dataset. As discussed at the beginning of this section, the expected total number of observations per variable is 1,000; however, as can be seen in the following table, this is not always the case.

Feature Name	**Number of Observations**	**Missing Rate**	**Outliers Rate**
Additional Loans	1,000	0%	N/A
Age	1,000	0%	0%
Application Date	1,000	0%	N/A
Application ID	1,000	0%	N/A
Burreau Score Value	1,000	0%	0%
Current Interest Rate Index	1,000	0%	N/A
Default Indicator	1,000	0%	0%
Employment Contract Type	997	0%	N/A
First Time Buyer	540	46%	N/A
Foreign National	688	31%	N/A
Interest Rate	1,000	0%	0.7%
Interest Rate Type	1,000	0%	N/A
Loan Term	955	5%	0%
Loan To Value	1,000	0%	0%
Marital Status	958	4%	N/A
Number of Debtors	1,000	0%	0%
Payment Schedule	987	1%	N/A
Post Code	668	33%	N/A
Primary Income	1,000	0%	0%
Principal	1,000	0%	1.5%
Property Rating	1,000	0%	N/A
Property Type	1,000	0%	N/A
Savings Size	961	4%	N/A
Secondary Income	236	76%	0%
Secondary Income Index	1,000	0%	N/A

Table 1: Data Quality

From Table 1 we observe that the variable "Secondary Income" has a missing rate of 76%. As discussed earlier, variables with such a high percentage of missing values should be excluded from the training sample. However, for this one this shouldn't be the case since from Table 9 we can infer that there is a direct one-to-one relationship between the features "Secondary Income", and "Secondary Income Index".
Additionally, in Figure 3 the variable "Secondary Income Index" has no missing values and out of the 1,000 total entries, 236 are equal to "Y". Thus, this implies that the true missing rate of "Secondary Income" is in fact zero.

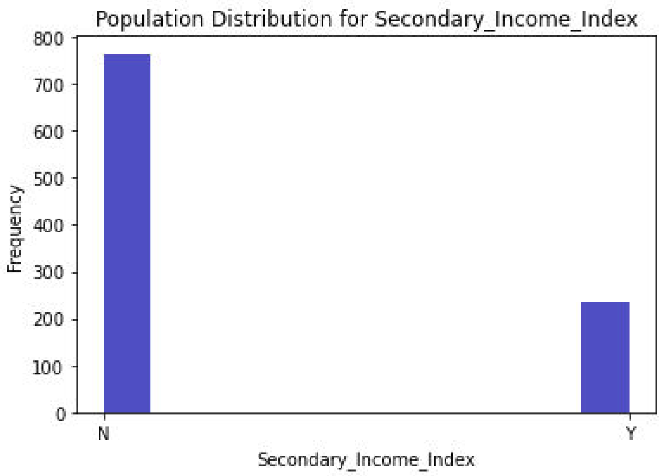


Figure 3: Population Distribution for Secondary Income Index

Table 9 (in bold) provides the final list of candidate variables. The variables "First Time Buyer", "Foreign National" and "Post Code" have been dropped from the list as they have a missing rate greater than 30% and any outliers for the remaining variables have been treated as discussed above. Moreover, the variable "Application ID" has also been removed as it bears no real information since it represents the unique identifier of the borrower in the dataset. Furthermore, our earlier qualitative analysis showed that the variables "Secondary Income" and "Secondary Income Index" are equivalent. Therefore, we do not include both in the analysis since in the case they are chosen for model development, it will have an unwanted impact on the regression coefficients and will generate misleading results. For this reason and without significant loss of generality, in our case study we chose the numeric variable "Secondary Income" for further analysis. Once the quality of the dataset has been verified, the next step is to identify which of the variables have sufficient predictive power to be considered for PD modelling purposes. In our approach, we initially determine the set of variables which we will use to fit the LR model as described in the following and then we further refine this list by considering: the level of statistical significance of each variable once LR has been fitted and the presence of multicollinearity among the variables. We underline that we chose Logistic Regression (LR) as our preferred modeling method because: it is well suited for such type of PD problems since it limits the output in the (0,1) space (the target is of binary type), it is straight forward to implement, it has low computational cost, and it is widely used in the banking sector for such type of problems. Let $A = \{feature_1, \dots feature_n\}$ be the set of all the predictive features, then the logistic regression equation, in terms of the logarithm of the odds, is given by:

$$log(odds) = log\left(\frac{PD}{1-PD}\right) = \alpha + \sum_{i=1}^{n} \beta_i \, feature_i, \qquad PD \in (0,1), \ \ (3).$$

The regression coefficients α and β_i are estimated through the non-linear least square method.
LR can be applied using either the raw data of each feature or by grouping it into a number of bins/categories, therefore converting all variables to categorical ones. The latter approach is usually preferred due to the benefits it offers including but not limited to:

- Reducing the number of categories in the case of discrete variables and thus reducing computational time.
- Ensuring monotonicity in terms of predictor variables given their linear relationship with the dependent variable.
- Providing better understanding of the behavior of each predictor variable.

In our case study, we adopt a quantile-based grouping approach and then we implement the Weight of Evidence Encoding (WoE) (Baesens, Roesch, & Harald, 2016). WoE is a univariate measure describing the relationship between a predictor variable and the dependent one. Assuming that we have consolidated the values of a feature into m distinct bins (or equivalently "buckets" or "groups"), the WoE is given by:

$$WoE_i^{feature} = ln\left(\frac{Distribution\ of\ Performing\ cases}{Distribution\ of\ Defaulted\ cases}\right), i \in \{1,2, \dots m\} \quad (4)$$

Once the WoE for all predictor variables has been calculated, eq. (3) becomes:

$$log(odds) = log\left(\frac{PD}{1-PD}\right) = \alpha + \sum_{i=1}^{n} \beta_i \, WoE^{feature_i}, \qquad PD \in (0,1). \quad (5)$$

In our dataset we decided to group the continuous variables into 10 bins while we use the set of distinct values of each of the categorical ones as per our initial segmentation. Then, we calculate the Weight of Evidence per bin per feature as defined by eq. (4). Next, we further adjust the bins until strict monotonicity with respect to WoE is achieved. It should be noted however that fine tuning of the bins should always be performed in a careful manner and in line with Subject Matter Expert input since crude enforcement of such an approach could sometimes lead to a biased model behavior. In addition to improving model performance, this encoding process allows for smooth resolution of the missing values problem since, when applicable, a separate bin can be created to group them together. Following the fine tuning of the bins, for each feature we computed the Information Value (IV); it is measure of the predictive strength of a feature, given by the following formula:

$$IV_{feature} = \sum_{i=1}^{m} (Distribution\ of\ Performing\ Cases_i - Distribution\ of\ Defaulted\ Cases_i) WoE_i^{feature} \quad (6).$$

Referring to the above formula, a variable is classified as weak, medium or strong predictor based on the following widely used bandwidths:

- IV < 0.1: Weak Predictor.
- 0.1 ≤ IV < 0.3: Medium Predictor.
- IV ≥ 0.3: Strong Predictor[4]

In our case study, a variable is included into the list of candidates for modelling if and only if its IV is greater than or equal to 0.1.
We conclude this paragraph providing an example of how to calculate the WoE and IV for the variable "Burreau Score Value" described in Table 9.
As mentioned earlier, we start our approach by creating ten bins of equal size each containing 100 observations each and then we calculate the number of performing and defaulted obligors per bin (Default Indicator = 0). Then, we compute the distribution of both performing and defaulted population segments.
This is done by dividing the number of obligors per bin per segment (performing/defaulted) by the total number of obligors belonging to the same segment. Finally, the WoE per bin is calculated using eq. (4).

Bin	Observation	Lower Bound	Upper Bound	Performing Cases	Defaulted Cases	Dist. Perf. Cases	Dist. Def. Cases	WoE
1	100	500	539	59	41	7.52%	19.07%	-0.9311
2	100	539	585	61	39	7.77%	18.14%	-0.8477
3	100	586	629	70	30	8.92%	13.95%	-0.4477
4	100	631	670	79	21	10.06%	9.77%	0.0299
5	100	670	695	80	20	10.19%	9.30%	0.0912
6	100	695	719	89	11	11.34%	5.12%	0.7957
7	100	719	743	82	18	10.45%	8.37%	0.2213
8	100	744	769	85	15	10.83%	6.98%	0.4396
9	100	769	804	91	9	11.59%	4.19%	1.0186
10	100	804	900	89	11	11.34%	5.12%	0.7957

Table 2: WoE for the variable "Burreau Score Value", Bins = 10

As can be seen from Table 2, a reversal trend is observed between bins 6 and 7 and bins 9 and 10. Furthermore, the WoE for bins 1 to 3 has the same sign while bins 4 and 5 have similar value. Thus, it would make sense to group together bins 1 - 3, 4 - 5, 6 - 7 and 8 - 10. In our case, this can be done by generating the respective quartiles for this variable.

Bin	Observation	Lower Bound	Upper Bound	Performing Cases	Defaulted Cases	Dist. Perf. Cases	Dist. Def. Cases	WoE
1	250	500	607	157	93	20.0%	43.26%	-0.771
2	250	607	695	192	58	24.5%	26.98%	-0.098
3	250	695	753	214	36	27.3%	16.74%	0.487
4	250	754	900	222	28	28.3%	13.02%	0.775

Table 3: WoE for the variable "Burreau Score Value", Bins = 4

[4] When a variable has IV greater than 0.5 usually is considered as a suspiciously strong predictor. To verify the validity of such an outcome, it is suggested that further data investigation is conducted coupled with qualitative analysis regarding its impact and importance in the modelling process.

From Table 3 we can see that when 4 bins are used, the WoE for this variable is strictly monotonous as opposed to the case of 10 bins (see Table 2) implying better predictive performance. Finally, if we use the values Table 3 in equation (6) we have that $IV_{\text{Burreau Score Value}} = 0.35 > 0.1$, therefore "Burreau Score Value" is a candidate variable for modelling. In addition, since we implemented WoE encoding, each unique value of this variable is mapped to one and only one bucket as defined in Table 3.
Table 4 below provides the final number of bins along with the calculated IV for each of the final variables in our dataset. Remember that the IV cut-off rate for including a variable in the analysis has been set to 0.1 meaning that from now on we focus only *on the first eight variables* from the list below[5]. Finally, when possible, we have created a separate bin for the missing values allowing for smooth encoding process.

Feature Name	**IV**	**# Bins**
Loan To Value	*0.716827*	*4*
Additional Loans	*0.626395*	*2*
Secondary Income	*0.490542*	*3*
Age	*0.394578*	*5*
Burreau Score Value	*0.351431*	*4*
Number of Debtors	*0.259039*	*3*
Savings Size	*0.157953*	*3*
Property Rating	*0.104202*	*4*
Application Date	0.097471	4
Principal	0.057218	5
Primary Income	0.043727	5
Property Type	0.028451	4
Interest Rate	0.026268	5
Interest Rate Type	0.026173	2
Marital Status	0.020673	4
Loan Term	0.013827	4
Payment Schedule	0.007772	3
Current Interest Rate Index	0.003292	3
Employment Contract Type	0.000378	4

Table 4: Information Value

5.2 Model Development

The aim of this case study is to address borrowers' default prediction problem via estimating the Probability of Default (PD) of each individual loan. The approach we propose in this article consists of a *regression problem,* i.e., we aim to predict the value of a depended variable (the PD) via modeling its relationship with one or more independent variables (the features). In this section we show how to fit a *Logistic Regression model* on the final list of candidate variables subject to the two selection criteria mentioned in paragraph 5.1 (the level of statistical significance of each variable once LR has been fitted and the presence of multicollinearity among the variables). Specifically, a variable will remain in our model if it is statistically significant, i.e. p-value ≤ 0.05 and its Variance Inflation Factor (VIF) is less than 5 as this is a commonly used cut-off threshold in banking practice (see (Ron Johnston, Jones, & Manley, 2018)). The VIF measures the degree of multicollinearity in our multiple regression model. In general, multicollinearity means that two or more variables of a set can linearly approximate some other variables from the same set. If a high degree of multicollinearity is present,

[5] We note that for the two variables having IV greater than 0.5 no data quality issues have been raised. Furthermore, in view of Table 9, it is evident that they contain significant qualitative information for our modelling purposes.

there could be various negative effects such as making the estimates of the regression coefficients unstable or inflating their standard errors. This would imply a bias in the interpretation of the relationship between a predictor variable and a dependent variable. The formula for calculating the VIF is given by:

$$VIF_{variable_i} = \frac{1}{1 - R^2_{variable_i}} \quad (7)$$

where $R^2_{variable_i}$ is the R - squared value resulting from regressing the i^{th} variable against all other available variables. We underline that in our case we have implemented the WoE encoding. This approach shifts the focus from inferring relationships between the independent variables and the target one towards enhancing the predictive power of the model. Therefore, the interpretation of the regression coefficients is not as straightforward as it would be had we used the actual variable values.
Table 5 below provides the estimated coefficients via LR and the respective p-values for each of our eight candidate variables.

#	Feature Name	Coefficient	Standard Error	p-Value
0	Constant	-1.4571	0.1682	<0.001
1	Loan To Value (WoE)	-1.5686	0.1847	< 0.001
2	Additional Loans (WoE)	-1.1834	0.1512	< 0.001
3	Secondary Income (WoE)	-1.8103	0.4623	0.0001
4	Age (WoE)	-0.4895	0.2466	0.0472
5	Burreau Score Value (WoE)	-0.9739	0.2878	0.0007
6	Number of Debtors (WoE)	**0.8648**	**0.4907**	**0.078**
7	Savings Size (WoE)	2.3549	0.8501	0.0056
8	Property Rating (WoE)	-1.0258	0.3531	0.0037

Table 5: LR Model Coefficients (Initial fit)

The variable "Number of Debtors" is not statistically significant given that its p-value is outside the tolerance (p-value ≤ 0.05). Therefore, it will be dropped from the list and the model will be refit. The outcomes of the updated model after removing the variable "Number of Debtors" are reported in Table 6, where all the variables are statistically significant and have VIF < 5.

#	Feature Name	Coefficient	Standard Error	p-Value	VIF
0	Constant	-1.6259	0.1421	<0.001	-
1	Loan To Value (WoE)	-1.5382	0.1827	< 0.001	1.104557
2	Additional Loans (WoE)	-1.1786	0.1506	< 0.001	1.051574
3	Secondary Income (WoE)	-1.8208	0.4684	< 0.001	2.921492
4	Age (WoE)	-0.4856	0.2464	0.0487	1.923134
5	Burreau Score Value (WoE)	-0.9582	0.2864	< 0.001	1.945571
6	Savings Size (WoE)	2.3085	0.8596	0.0072	2.8061
7	Property Rating (WoE)	-1.0139	0.3585	0.0041	1.206845

Table 6: LR Model Coefficients (Final)

At this stage, we compute the PD for each individual borrower using equation (5). Once we have identified the features that have a significant predictive power for the LR, we decided to further investigate the default prediction problem using also a non-linear classifier. Classifiers using non-linear algorithm are as an example the support vector machine, artificial neural network, k-nearest neighbour, naïve Bayes, random forest, and Bayesian network. As mentioned above, in this paper we decided to compare LR with k-NN since this approach is computationally affordable and completely nonparametric. The basic principle behind this method is that a given instance within a data set will generally exist in proximity with other instances sharing similar properties. Hence, additional information about an instance can be obtained by observing other instances that are close to it, that is, the Nearest Neighbours (NNs).

If the instances within a data set are tagged with a classification label, then the class of a new instance can be determined by observing the classes of its NNs. The advantage of nearest-neighbour classification is its simplicity. There are only two choices the modeler must make: the number of neighbours k and the distance metric to be used. We decided to use the Euclidean distance as measure. we have implemented the k-NN algorithm for k =3, 5, 7 using the Euclidean distance (Sun & Huang, 2010).
As mentioned in the introduction, the calibration of the model is out of the scope of this article; despite that, we want to conclude this paragraph describing the main steps of the calibration procedure, since it would be the final task to perform in building a PD model before validating it.
In real world banking practice, retail portfolios of medium size usually contain tens of thousands of borrowers. Therefore, the risk profile at portfolio level is easier to understand if the obligors are classified into grades and a PD - representative of the population – is assigned to each grade. The calibration sample usually contains hundreds of thousands of data while the observation period extends to more than 10 years.
This process is commonly referred to as calibration of a rating system and it is usually performed in a hybrid manner where statistical tools and expert judgement are combined in such way that:

1. All grades are homogeneous, meaning that obligors of similar risk profile are grouped together.
2. All grades have distinct risk characteristics and differ from one another.
3. All grades have a significant number of borrowers.
4. The default rate across the grades displays a monotonous behaviour.
5. The PD assigned to each grade is representative of the long run average of the portfolio's default rate.

5.3 Model Validation

In this section we introduce the concept of Model Validation and its relevance for every financial institution. We will briefly introduce the most used statistical tools used in this process and we apply them to our case study. Model validation is a well-regulated process with the main goal to verify whether a model developed for assessing any financial risk (in our case credit risk) addresses in a satisfactory and appropriate manner the business needs and its design objectives. Therefore, all financial institutions have in place large and independent teams tasked with ensuring that every model used for risk management is quantitatively and qualitatively sound and fully compliant with the latest regulatory requirements. For this purpose, the data used for model development, the model design, the model documentation, and the policy framework on which the model was built are subject to investigation and validation. In other words, no model will go live in a production system without having successfully passed the validation process. The process of validating a newly developed model is referred to as "Initial Validation". Once the model has been deployed to production, it is then subject to periodic reviews to ensure that its performance remains fit for purpose and is still aligned with the current regulatory framework, which is frequently updated. The process of periodically assessing the quality of a live model is referred to as "Periodic Validation". If a model fails some periodic validation review, for example due to changes in the macro-economic environment, then depending on the severity of the identified issues either it will have to be re-trained by incorporating the latest available data or completely re-built. Given that most models are mainly statistically developed, the validation process has a strong quantitative component. This is especially true for PD models for credit risk assessment. In the quantitative analysis, the most prominent areas of interest are to verify whether the population used during model development is representative of the current population and whether the model has good predictive ability and discriminatory power.
In literature, there are two types of samples used for validation purposes: the *in-time* sample and the *out-of-time* sample. The *in-time* sample randomly reserves a portion, usually from the 20% to the 30%, of the dataset used for the model development to the validation. This means that, the model is trained on the remaining percentage (from 70% to 80%) and validated on the reserved sample. The *out-of-time* sample contains information that are outside the time frame of the data used for model development. Indeed, the validation is performed on a later dataset than that on which the model has been fitted. To check the representativeness of the training dataset upon the *out-of-time* validation sample, the Population Stability Index PSI is used. The PSI is an index that measures how much a variable has shifted over time and is used to monitor applicability of a statistical model to the current population. However, PSI statistic intends to capture any significant shifts in the distribution of the development sample over time; since our model is not subject to calibration and it is based on a simulated dataset, those two tests are not applicable to our case study. For this reason, we have applied an *in-time* validation with 70 / 30 split, meaning that 70% of the dataset was used for model development and 30% for validation purposes.
The discriminatory power of a PD model measures its ability to differentiate well between defaulted and non-defaulted borrowers. The most popular tools available for this purpose are the Receiver Operating Characteristic (ROC) curve, the Area Under the ROC Curve (AUC), the Cumulative Accuracy Profile (CAP) and the Accuracy Ratio (AR). We decided to use the Area Under the Curve (AUC) and the Accuracy Ratio (AR), since they are robust performance measures for a large number of classifiers (including LR) and they are extensively used in the literature related to ML methods applied to credit risk (see for example (Tasche D. , 2008), (Tang & Chi, 2005), (Fantazzini & Figini, 2009), (Kruppa, Schwarz, Arminger, & Ziegler, 2013), (Addo, Guegan, & Hassani, 2018)). The AUC is a performance measurement statistic describing the strength of the classifier in terms of assigning a lower PD to a true random performing observation than a true random defaulted observation. In general, the performance of a classifier like the AUC can be described through a confusion matrix of the following form:

		Observed	
		Default	**Performing**
Predicted	**Default**	True Positive	False Positive
	Performing	False Negative	True Negative

Table 7: Confusion Matrix

Here, True Positive (TP) represents the number of obligors that the model classified as in default and were actual defaults, False Positive (FP) represents the number of borrowers that the model classified as in default but were in performing status, False Negative (FN) represents the number of obligors that the model classified as performing but were in default status and finally, True Negative (TN) represents the number of obligors that the model classified as performing and were in performing status. Starting from the values contained in the confusion matrix, one can calculate the following two rates:

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP + FN} \quad (8)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN} \quad (9)$$

Based on this set-up, the Receiver Operating Characteristic (ROC) curve is then defined as the set of all points (FPR, TPR) across all possible cut-off probability thresholds, i.e., the level representing the probability of a true prediction. Notice that all points along the diagonal line represent a model for which the TPR is equal to the FPR for each cut-off threshold, thus, a random model with no substantial discriminatory power. Furthermore, points (0,0) and (1,1) imply cut-off thresholds of 1 and 0 respectively. Simply put, in the first case all points have been classified as in performing status meaning that TP= FP =0 therefore, (FPR, $TPR) = (0,0). In the second case, all points have been classified as in default status meaning that FN=TN=0 therefore (FPR, TPR) = (1,1). Once the ROC curve has been generated, the AUC statistic is then simply defined as the area between the FPR axes and the ROC curve. The discriminatory power of a model is defined as follows:

- AUC = 0.5: No Substantial Discrimination.
- 0.5 < AUC < 0.7: Weak Discrimination.
- 0.7 ≤ AUC < 0.8: Moderate Discrimination.
- AUC ≥ 0.8 High Discrimination.

In our case study, we consider the results obtained in Table 6 and we compute the ROC curve in case of LR and the k-NN algorithm for k =3, 5, 7 (see Figure 4 and Figure 5 below).

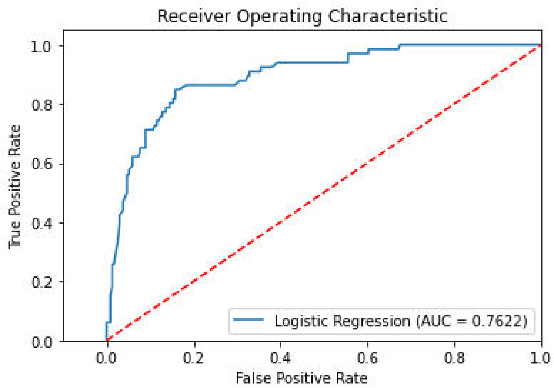


Figure 4: Logistic Regression, 7 Features, AUC

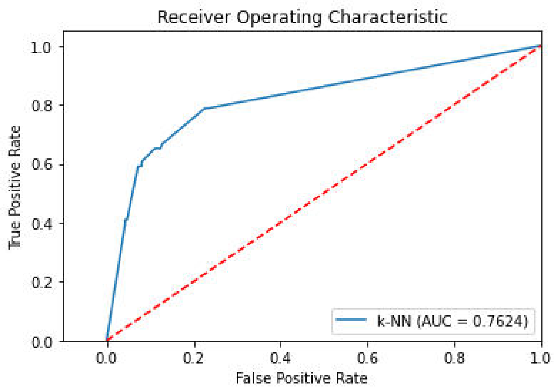


Figure 5: k-NN, k = 3, 7 Features, AUC

Table 8 shows the AUC values in case of LR and the k-NN algorithm for k =3, 5, 7. Both models have a moderate discrimination power, the AUC difference between LR and k-NN is negligible for k = 3, while the LR algorithm performs better than the k-NN when k becomes larger (k = 5, 7).

Number of Features	LR	3NN	5NN	7NN
7	0.7622	0.7624	0.7213	0.7494

Table 8: Model Comparison – AUC – 7 features

Another popular tool used for measuring the discriminatory power of a model is the Cumulative Accuracy Profile (CAP) (see (Tasche D. , 2008)). This curve differs from the ROC one in the sense that instead of plotting the TPR versus the FPR across all possible regression thresholds, it represents the cumulative percentage of default rate against the corresponding cumulative population percentage. When evaluating a PD model via CAP, it is assumed that the population has been sorted in descending order PD wise. Furthermore, similarly to the ROC curve, the diagonal line described by the equation y=x represents a random model with no substantial discriminatory power, while the perfect model is described by a CAP curve in which the maximum response rate is achieved at the lowest possible end of the population's distribution. As an example, Figure 6 displays three different models of poor, medium and high discriminatory power. Note that all three of them are bounded by the random and perfect model CAP curves.

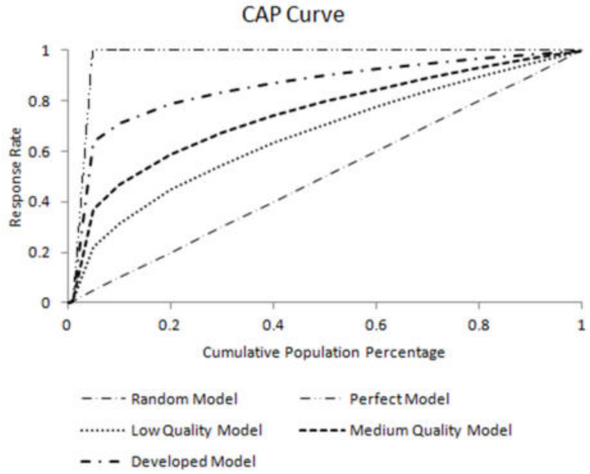


Figure 6: Sample CAP Curve

Starting from the CAP curve, the discriminatory power of a model can be estimated using the Accuracy Ratio (AR) which is defined as the quotient of the area between the CAP curves of the developed and the random model over the area defined by the CAP curves of the perfect and the random model. Specifically, if A denotes the area between the developed and the random model and B the area between the perfect and the developed model:

$$AR = \frac{A}{A+B} \quad (10)$$

In a successful model, the AR ranges as $0 \leq AR \leq 1$, and the higher the value is, the stronger the discriminatory power of the model is. Observe that equation 10 is equivalent to AR = 2AUC – 1. In our case study, since AUC ranges from 0. 7213 to 0.7624, then AR ranges from 0.4426 to 0.5248, confirming the moderate predictive power of the PD model.

Typically, once the model validation is completed, in the event of a positive outcome the model can be either enabled to the production system or can continue to be used as is until the next periodic review. However, if this is not true then depending on the severity of the recommendations that have been called out by the validation team, the model will be subject to several corrective actions spanning from minor adjustments to some of its components up to full re-built. To this point, it should be clear that model development and validation are two distinct and independent functions both serving the same objective, i.e., to ensure that all models used by a financial institution are of the best possible quality.

6. Conclusions

In this paper we have performed a critical review of the most used model development and validation procedures for credit risk in retail banking. After having described the impact of Regulator on Credit Risk Methodologies and briefly motivated the introduction of ML approaches for regulatory capital estimation, we have reviewed some of the traditional statistical methods to PD and LGD estimates. Then we have focused on more recent studies based on Machine Learning techniques. As next step, we have constructed a case study showing how to develop and validate a PD model via a Machine Learning Techniques. We have used a simulated dataset to show the data pre-processing, cleansing and the application of the Weight of Evidence Encoding, a powerful technique well suited for Logistic Regression problems since it introduces a monotonic relationship between the target variable and the predictors. Finally, we have compared the LR model against non-linear classifiers, the k-NN algorithm (for k= 3, 5 & 7) and we measured the predictive power of both models using the AUC and AR. The results indicate that on the one hand, the LR classifier coupled with WoE performed in a similar way to the k-NN for k =3, while it outperformed LR for k=5 and k = 7. Given the interpretability and simplicity that the LR method offers coupled with less computational effort and complexity as opposed to non-parametric machine learning models, we conclude that the choice of an LR model leveraging the WoE technique is very well suited and produced good predictive power in comparison to k-NN.

Machine learning methods are of major importance in the retail banking sector since most, if not all, of the models used for risk management make us of these techniques. As Science and Technology advance year over year and at a rapid pace, as well as the availability of large dataset is increasing, financial institutions are constantly strengthening their decision-making processes by developing sophisticated algorithms which can consume large pools of data in a very short time and are in line with the current regulation as well. Since banks play a very important role in the global economy, better decision-making means better risk management which in turn means greater financial stability. It is always worth remembering that despite the increase of ML techniques and availability of large dataset, risk management remains human activity and effective risk management would not be possible without strong critical reflection and experts' judgement at every step of the process.

Appendix

The following table provides the feature name, description and information regarding whether it is included or not in the model development phase. The variable names follow the format presented in the ECB's Loan-Level data templates. In bold you can find the features that have been included in the development process.

#	Feature Name	Description	Type
1	**Additional Loans**	**Nominal variable, describing whether the obligor has additional loans. Values: Y/N**	**Account Type**
2	**Application Date**	**Ordinal variable, describing the date of the application. Values between January 2010 to December 2019.**	**Account Type**
3	Application ID	Nominal variable, describing the obligor's unique identifier.	Account Type
4	**Burreau Score Value**	**Numeric variable (Integer) describing the obligor's credit score. Values between 500 to 900.**	**Account Type**
5	**Current Interest Rate Index**	**Nominal variable, describing the index on which the mortgage was written on. Values: Euribor_3M, Euribor_6M, No_Index.**	**Account Type**
6	***Default Indicator (Target Variable)***	***Numeric variable, describing whether the obligor is in default or not. Values: 1 (Default), 0 (Performing).***	***Account Type***
7	First Time Buyer	Nominal variable, describing whether the obligor is a first-time buyer. Values: Y /N.	Account Type
8	**Interest Rate**	**Numeric variable (float), describing the interest rate applied to the mortgage. All interest rates have been rounded to the first decimal place. Values from 4.3% to 7.5%.**	**Account Type**
9	**Interest Rate Type**	**Nominal variable, describing whether the interest rate is fixed or floating. Values: Fixed / Floating.**	**Account Type**
10	**Loan Term**	**Numeric variable (Integer), describing the maturity of the mortgage. Values: 20, 25 & 30.**	**Account Type**
11	**Loan To Value**	**Numeric variable (float), describing the percentage of the value of the mortgage with respect to the value of the property. Values (rounded to the second decimal place) range from 77% to 85%.**	**Account Type**
12	**Number of Debtors**	**Numeric variable (Integer), describing the number of obligors on which the mortgage is written on. Values 1, 2 or 3.**	**Account Type**
13	**Property Type**	**Nominal variable, describing the purpose for which the mortgage was given. Values (1) Holiday/second home (2) non-owner-occupied/buy-to-let (3) Other (4) Owner-occupied.**	**Account Type**
14	**Payment Schedule**	**Nominal variable, describing the interest payment schedule. Values: 3M / 6M.**	**Account Type**
15	**Principal**	**Numeric variable (Integer), describing the face value of the mortgage. Values (rounded to thousands) from 144,000 to 648,000.**	**Account Type**
16	**Property Rating**	**Ordinal variable, describing the quality of the property. Values (from best to worst): (1) CAT_1 (2) CAT_2 (3) CAT_3 (4) CAT_4**	**Account Type**
17	Secondary Income Index	Nominal variable, describing whether a secondary income exists. Values: Y / N	Account Type
18	Foreign National	Nominal variable, describing whether the obligor is of foreign nationality. Values: Y / N	Demographic Type
19	Post Code	Nominal variable, describing the zip code of the obligor's address.	Demographic Type
20	**Age**	**Numeric variable (Integer), describing the obligor's age at application date. Values from 27 to 57.**	**Sociological Type**
21	**Marital Status**	**Nominal variable, describing the marital status of the obligor. Values: M/C (Married or Cohabitating), D (Divorced), S (Single).**	**Sociological Type**
22	**Primary Income**	**Numeric variable (Integer), describing the obligor's annual salary. Values (rounded to thousand): 36,000 to 95,000**	**Sociological Type**
23	**Saving Size**	**Ordinal variable, describing the obligor's saving status. Values: Above 50k / Below 50k.**	**Sociological Type**
24	**Secondary Income**	**Numeric variable (Integer), describing the size of the obligor's annual secondary income. Values (rounded to thousand) from 4,000 to 15,000.**	**Sociological Type**
25	**Employment Contract Type**	**Nominal variable, describing the obligor's occupation type. Values: (1) Fixed Term Contract (2) Permanent Contract (3) Self-Employed**	**Sociological Type**

Table 9: Variable Description

Bibliography

EBA/GL/2017/06. (2017). *Guidelines on credit institutions' credit risk management practices and accounting for expected credit losses.*

Abdou, H., Pointon, J., & Elmasry, A. (2008). Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking. *J. Expert Systems with Applications, 35(3)*, 1275-1292.

Addo, P., Guegan, D., & Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 08/WP/2018*.

Altman, I. E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*, 589-611.

Amin, M. F., Islam, M. M., & Murase, K. (2009). Ensemble of single-layered complex-valued neural networks for classification tasks. *Neurocomputing, 72*, 2227-2234.

Angelini, E., Tollo, G. D., & Roil, A. (2008). A Neural Network Approach for Credit Risk Evaluation. *The Quarterly Review of Economics and Finance, 48(4)*, 733-755.

Baesen, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Van-Thienen, J. (2003). Benchmarking state-of-art classification algorithm for credit scoring. *Journal of the Operational Research Society, 54*, 627-635.

Baesens, B., Roesch, D., & Harald, S. (2016). *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS.* Wiley.

Bank for International Settlements. (2005a). *Basel Committee on Banking Supervision, International convergence of capital measurement and capital standards: a revised framework.* Basel, Switzerland.

Bank for International Settlements. (2005b). *An Explanatory note on Basel II IRB Risk Weight Functions.* Basel, Switzerland.

Bank for International Settlements. (2009). *Range of practices and issues in economic capital frameworks.* Basel, Switzerland.

Bank for International Settlements. (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems.* Basel, Switzerland.

Bank for International Settlements. (2017). *IFRS 9 and expected loss provisioning - Executive Summary.* Basel, Switzerland.

Bharath, S. T., & Shumway, T. (2006). Forecasting default with the KMV-Merton model. *AFA 2006 Boston Meetings Paper*.

Black, F., & Cox, J. C. (1976). Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance 31(2)*, 351-367.

Breiman, L. (2001). Random Forest. *Machine Learning, 45(1)*, 5-32.

Duffie, D., & Lando, D. (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica, 69(3)*, 633-664.

Duffie, D., & Singleton, K. (1999). Modelling term structure of defaultable bonds. *Rev. Financial Stud., 12*, 687-720.

Duffie, D., & Singleton, K. (2012). *Credit risk: pricing, measurement, and management.* Princeton University Press.

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. *AI Ethics 1*, 73-80.

Elliott, R. J., Jeanblanc, M., & Yor, M. (2000). On models of default risk. *Mathematical Finance, 10(2)*, 179-195.

European Central Bank. (2019). *ECB Guide to Internal Models.*

Fantazzini, D., & Figini, S. (2009). Random Survival Forests Models for SME Credit Risk Measurement. *Methodology and Computing in Applied Probability volume, 11*, 29-45.

Finger, C., Finkelstein, V., Pan, G., Lardy, J., Ta, T., & Tierney, J. (2002). *Credit Grades. Technical Document.* . New York: Riskmetrics Group.

Frye, J. (2000a). Collateral Damage. *Risk, April, 13(4)*, 91-94.

Frye, J. (2000b). Collateral damage detected. *Federal Reserve Bank of Chicago Working Paper, Emerging Issues Series, October*, 1-14.

Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Computational Economics, 15(1)*, 107-143.

Hsieh, N. (2005). Hybrid mining approach in design of credit scoring model. *Expert Systems with Applications, 28*, 655-665.

Huang, Z., Chen, H. C., Hsu, J., Chen, W. H., & Wu, S. (2004). Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support System, 37(4)*, 543-558.

Irwin, R. J., & Irwin, T. C. (2012). *Appraising Credit Ratings: Does the CAP Fit Better than the ROC?* IMF Working Paper, 12/122.

Jarrow, R. A. (2001). Default parameter estimation using market prices. *Financial Analysts Journal, 57(5)*, 75-92.

Jarrow, R., & Protter, P. (2004). Structural versus reduced-form models: A new information based perspective. *J. Investment Management, 2(2)*, 34-43.

Jarrow, R., & Turnbull, S. (1992). Credit risk: Drawing the analogy. *Risk Magazine, 5(9)*, 51-56.

Jarrow, R., & Turnbull, S. (1995). Pricing Derivatives on financial securities subject to credit risk. *The Journal of Finance, 50(1)*, 53-85.

Joenssen, D., & Bankhofer, U. (2012). Hot Deck Methods for Imputing Missing Data. *Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science. 7376*, pp. 63-75. Berlin: Springer.

Jokivuolle, E., & Peura, S. (2000). A model for estimating recovery rates and collateral haircuts for bank loans. *Bank of Finland Research Discussion Paper, 2*.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications, 40(13)*, 5125-5131.

Lee, M., & Floridi, L. (2021). Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs. *Minds & Machines*, 165-191.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance 29(2)*, 449-470.

Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 36*, 3028-3033.

Orgler, Y. (1970). A credit scoring model for commercial loans. *J. Money Credit Bank., 2*, 435-445.

Pang, S., & Gong, J. (2009). C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. *Systems Engineering - Theory and Practice, 29(12)*, 94-104.

Ron Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality and Quantity, 52*, 1957–1976.

Sadhwani, A., Giesecke, K., & Sirignano, J. (2021). Deep Learning for Mortgage Risk. *Journal of Financial Econometrics*, 313-368.

Sadok, H., Sakka, F., & El Hadi El Maknouz, M. (2022). Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance, 10(1)*.

Schebesch, K. B., & Stecking, R. (2005). Support vector machine for classifying and describing credit applicants: Detecting typical and critical regions. *Journal of the Operational Research Society, 56*, 1082-1088.

Shin, K. S., Lee, T. S., & Kim, H. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28(1)*, 127-135.

Siddiqi, N. (2017). *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards, 2nd Edition.* Wiley.

Sun, S., & Huang, R. (2010). An adaptive k-nearest neighbor algorithm. *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, 91-94.

Sun, S., & Huang, R. (n.d.). Sun, An adaptive k-nearest neighbor algorithm. *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, 91-94.

Tang, T., & Chi, L. (2005). Predicting multilateral trade credit risks: comparisons of Logit and FuzzyLogic models using ROC curve analysis. *Expert Systems with Applications, 28(3)*, 547-556.

Tasche, D. (2004). The single risk factor approach to capital charges in case of correlated loss given default rates. *arXiv preprint cond-mat/0402390*.

Tasche, D. (2008). Validation of internal rating systems and PD estimates. *The Analytics of Risk Model Validation*, 169-196.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27(11-12)*, 1131-1152.

Yu, L., & Wang, S. L. (2009). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring. *European Journal of Operational Research, 195*, 942-959.

Zou, L., & Khern-am-nuai, W. (2022). AI and housing discrimination: the case of mortgage applications. *AI Ethics*.

Operational Risk framework and Standardised Measurement Approach (SMA)

Paolo Fabris (Avantage Reply); Alessandro Leoni (Avantage Reply); Ilaria Marfella (Avantage Reply)

Abstract

On December 2017, the Basel Committee published the "Basel III: Finalising post-crisis reforms" (also known as Basel IV) that introduces the Standardised Measurement Approach (SMA) to define the Pillar I operational risk capital requirement that is foreseen to entry into force on the 1st of January 2025, replacing all the existing approaches.
This approach not only introduces a new method to be used to calculate the operational risk capital requirement but details several updates that have to be applied to the main components of the framework such as Governance, Loss Data Collection and Risk Self-Assessment.
With the entry into force of the SMA, banks have the chance to fully re-think their operational risk Management Framework (ORMF) integrating the different components and making it more efficient and effective in terms of data governance, process management and reporting.
This paper describes the SMA methodology to be implemented to calculate the Pillar I operational risk capital requirement and provides an overview of the expected impact on the different components of the ORMF of the banks.

Keywords

Operational Risk, Risk Management, Risk Management Framework, Regulatory Capital Requirement, Standardised Measurement Approach, SMA, Basel III Reform, Business Indicator Component, Internal Loss Multiplier, Loss Data Collection, Risk and Control Self-Assessment, Operational Risk Management Framework, Capital Requirement Regulation

Introduction

Operational risk represents the second risk category in terms of capital requirement across the EU banking sector and, also considering the increase of Operational Risk Management Framework (ORMF) complexity within the banking sector, it has become a priority on Regulator agenda.
Since the existing operational risk framework failed to capture the fact that operational risk is often associated with other types of risk and since some weaknesses regarding governance, risk data aggregation capabilities and reporting practices emerged, supervisors required banks to adopt an integrated risk management approach through the introduction of the SMA as specified by the "Basel III: Finalising post-crisis reforms".
This paper provides an overview of how, considering the entry into force of the revised operational risk framework foreseen for the 1st of January 2025, banks have the opportunity to perform an overall refresh of their operational risk related processes in order to be compliant with the new framework and to introduce efficiency and effectiveness in the key processes and procedures related to Loss Data Collection (LDC), Data Governance, Risk Self-Assessment, Monitoring and Reporting.

Operational risk in a nutshell

Operational risk has become an increasingly important type of risk to be managed by financial institutions. The Basel Committee has introduced a series of substantial reforms to the international regulatory framework mainly addressing the strengthening of the capital adequacy scheme.
Operational risk is the second most important RWA (Risk-Weighted Asset) component representing, on average, 9,5% of total RWAs.

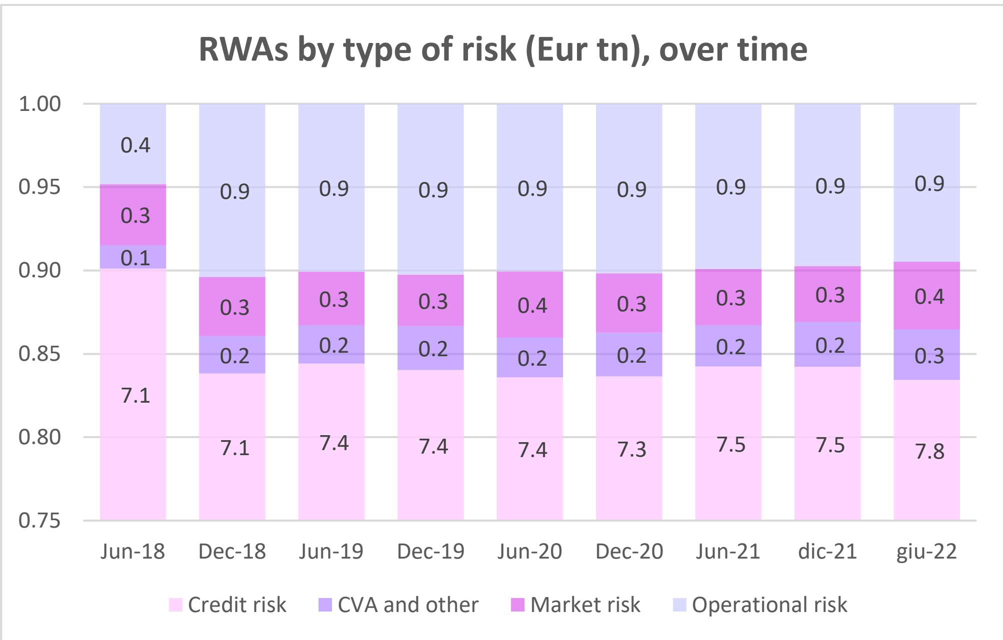


Figure 1: RWA by type of risk.

Source: Risk Assessment of the European Banking Authority (December 2022) - Supervisory reporting data

In their responses to the Risk Assessment Questionnaire, banks identified the main drivers that may affect their activities leading to operational losses and, under the SMA, triggering an increase of the operational risk capital requirement.
The graph reported below shows that **Cyber risk and Data security** are the most prominent drivers of increased operational risk, followed by Conduct, Legal and Fraud risks.
However, also **IT failures and organizational change** are considered as relevant drivers of operational risk since organizational change risks arise when institutions further adapt their organizational setups to a digital environment.
Risks stemming from sophisticated and organised cyber-attacks with potentially big impact as well as other ICT-related incidents are very high if we consider the complexity and interconnectedness of ICT systems, both owned by banks and those dependent on third-party providers[1].
In addition, the relevance of **Conduct and Legal risk** has significantly increased compared to last year. Concerns about past misconduct behavior, such as breaches of sanctions, redress for mis-selling, fines associated with financial crime and misconduct continue to uphold and add to operational risks.
Beyond reputational damage for the banks concerned, misconduct costs have been substantive and added to challenges to attain sustainable profits. **Risk of fraud** continues to increase in banks' perceptions, especially considering that they are relying on digital and remote solutions to perform their daily operations, to deliver their services to customers, and to conduct business.
These have resulted in an enhanced exposure and vulnerability to frauds and to increasingly sophisticated cyber-attacks.
In the future, the operational risk outlook continues to be high.
Economic and geopolitical uncertainty coupled with a high level of cyber risks will contribute to maintaining a high level of operational risk.
The incentives to circumvent the sanctions may potentially provide opportunities for the emergence of new types of misconduct.
It is therefore important to strengthen the monitoring of business operations and therefore the related operational risk.

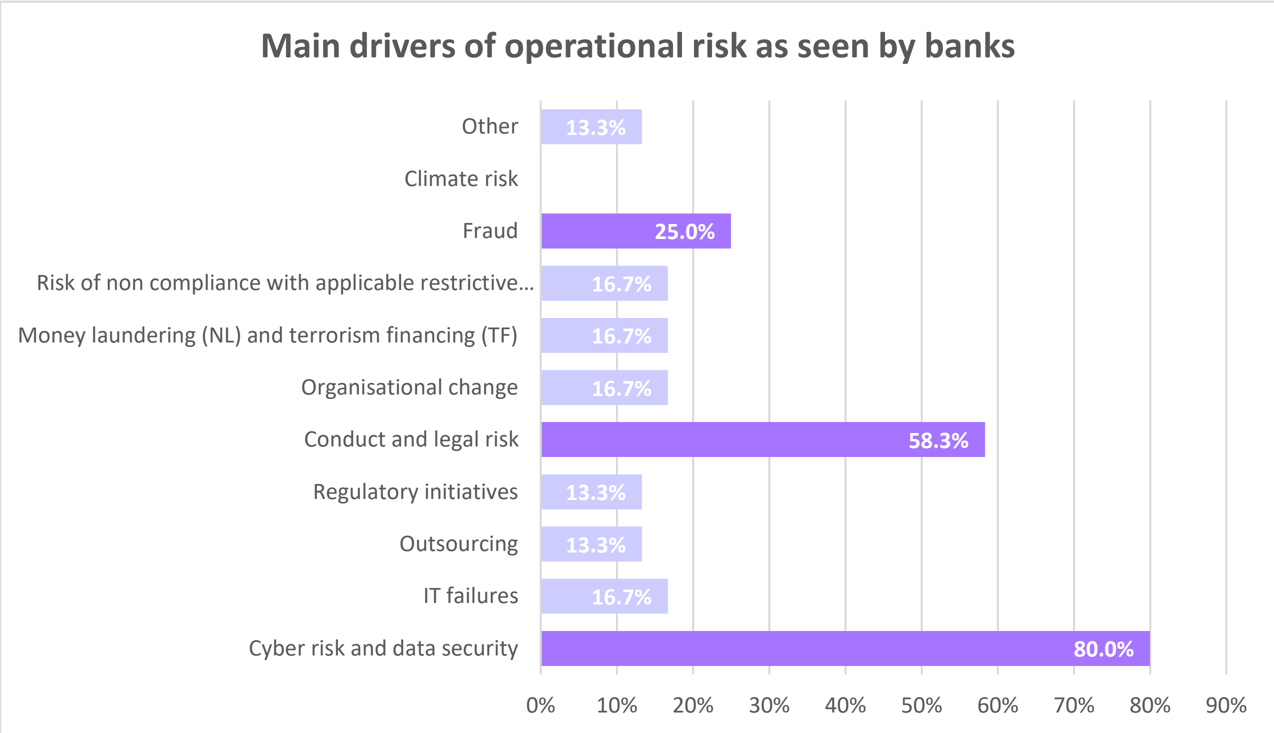


Figure 2: Main driver of operational risk as seen by banks. Source: Risk Assessment of the European Banking System (December 2022) - Reporting data

Source: Risk Assessment of the European Banking Authority (December 2022) - RAQ for banks

The Basel Committee highlights that a failure to fully implement appropriate operational risk identification and management practices may result in direct and material financial losses, or reputational and consequential losses, and could lead to a systemic impact on other banks, customers, counterparties and the financial system.
In recent years, the amount of gross losses has been quite stable. Despite the Covid-19 pandemic, the gross losses have decreased in 2021 at the lowest value of the last 10 years.
Going forward, phishing attempts and other types of cyber-attacks are expected to become more common. Moreover, the evolution of the products and activities performed by banks, such as participation in virtual currency transactions where the identities of the individuals involved are not fully transparent, may expose banks to additional risks (cases related to money laundering, terrorist financing and sanctions due to non-compliance have increased in recent years).
Additionally, an inadequate management of environmental, social or governance (ESG) factors might increase the reputational costs.

[1] Digital Operational Resilience Act - Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector

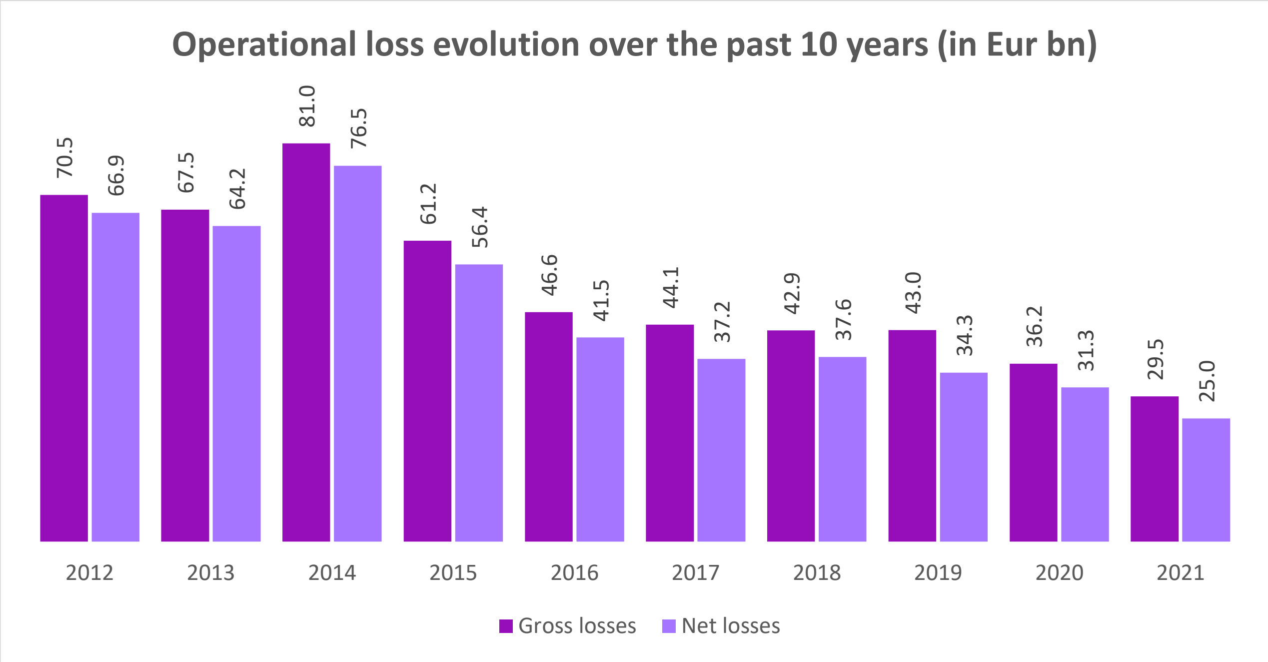


Figure 3: Operational loss evolution over the past 10 years. Source: Basel Committee on Banking Supervision – Basel III Monitoring report February 2023

Since a poor management of these types of risk could lead to a materialization of new losses, supervisors require banks to stay vigilant in times of economic turmoil and uncertainty and strengthen their monitoring of operational risks.
The improvement in operational risk capital adequacy is one of the most central elements of the global regulatory response to enhance credibility in the EU banking sector.

Regulatory framework

The Basel III framework implemented through the Regulation EU 575/13 (Capital Requirement Regulation - CRR) has foreseen three different approaches, with growing complexity and requirements to be fulfilled, to be used to calculate the operational risk capital requirement:

- **Basic Indicator Approach** - the own funds requirement for operational risk is equal to 15 % of the average over three years of the relevant indicator (proxy of the gross income)
- **Standard Approach** - banks activities are divided into eight business lines with different weighting factor: corporate finance, trading and sales, retail banking, commercial banking, payments and settlements, agency services, asset management and retail brokerage. The own funds requirement for operational risk is calculated as the average over three years of the sum of the annual own fund requirements across all business lines
- **Advanced Measurement Approach** – the capital requirement is quantified through empirical models created by the banks considering internal loss data, external data, scenario and business environment analysis, or internal control factors.

With the aim to reduce possible inconsistencies in the calculation of capital requirements by institutions, in 2016 the Basel Committee on Banking Supervision (BCBS), proposed the Standardised Measurement Approach (SMA).
The SMA substitutes all the existing approaches for the calculation of the Pillar I operational risk capital requirement. The final guidance has been included in the “Basel III: Finalising post-crisis reforms” published on December 2017[2].
In August 2020, the BCBS published a Consultative Document on the Revisions to the Principles for the Sound Management of Operational Risk (PSMOR)[3] with the aim to provide a guidance on the management of operational risk.
The document emphasizes the importance of an efficient Operational Risk Management Framework and encompasses five areas where the banks are expected to follow the best practices: Governance, Risk Management Environment, Information and Technology, Business Continuity Planning and the Role of Disclosure.
The BCBS pointed out that it is necessary to integrate all the specific components into the Operational Risk Management Framework avoiding to consider each element as stand alone.
Indeed, an effective and efficient management of the operational risk allows banks to continue to carry out their operations despite the occurrence of outstanding events.

Capital Requirement calculation under SMA

The Operational Risk Capital Requirement (OCR) can be summarised as follows:

$$OCR = BIC \times ILM$$

[2] Basel III: Finalising post-crisis reforms (bis.org)
[3] Revisions to the Principles for the Sound Management of Operational Risk (bis.org)

Thus, according to the new standardised approach for operational risk, a bank's operational risk capital requirement is based on two components:

1. the ***Business Indicator Component*** (**BIC)**, calculated as the Business Indicator (BI), a balance sheet metric, multiplied by marginal coefficients that depend on the BI amount.

Bucket	BI range (in € bn)	BI Marginal Coefficient
1	≤ 1	12%
2	1 < BI ≤ 30	15%
3	> 30	18%

Table 1: SMA marginal coefficient

2. *the* ***Internal Loss Multiplier*** (**ILM**), which is a scaling factor that considers the amount of the operational losses recorded by the bank on a 10-years horizon. For banks in bucket 1 (i.e., with BI ≤ EUR 1 billion), the ILM is equal to 1, therefore internal loss data does not affect the capital calculation.

More in detail, as per "Basel III: Finalising post-crisis reforms":

- the Business Indicator is defined as the sum of three components:

    1. interest, leases and dividend component (ILDC) calculated as:

       ILDC = Min [<u>Abs (Interest Income – Interest Expense)</u>; 2,25% * <u>Interest Earning Assets</u>] + <u>Dividend Income</u>

    2. services component (SC) calculated as:

       SC = Max [<u>Other Operating Income</u>; <u>Other Operating Expense</u>] + Max [<u>Fee Income</u>; <u>Fee Expense</u>]

    3. financial component (FC) calculated as:

       FC = <u>Abs (Net P&L Trading Book)</u> + <u>Abs (Net P&L Banking Book)</u>

  where the underlined factors indicate a 3-years average

- the Internal Loss Multiplier is calculated as:

$$ILM = \ln\left(\exp(1) - 1 + \left(\frac{LC}{BIC}\right)^{0,8}\right)$$

  where the Loss Component (LC) is equal to 15 times average annual operational risk losses incurred over the previous 10 years.

It is necessary to highlight that the standards set by the Basel III Reform allow Jurisdictions to disregard historical losses for the calculation of operational risk capital for all relevant institutions. For the calculation of the minimum own fund requirements, in order to ensure a level playing field within the Union and to simplify the calculation of operational risk capital, those discretions have been exercised by the European Commission in a harmonised manner by disregarding historical operational loss data for all institutions[4].

However, the European Central Bank (ECB) "*considers that taking into account the loss history of an institution would entail more risk sensitivity and loss coverage of capital requirements, addressing the divergence of risk profiles of institutions in highly sensitive issues such as conduct risk, money laundering or cyber incidents, and would provide greater incentives for institutions to improve their operational risk management. The ECB would therefore favor an implementation where the internal loss multiplier is determined by historical losses incurred by the institution and gradually introduced*"[5].

[4] Regulation of the European Parliament and of the Council amending Regulation (EU) No 575/2013 as regards requirements for credit risk, credit valuation adjustment risk, operational risk, market risk and the output floor

[5] European Central Bank - Opinion of the European Central Bank of 24 March 2022 on a proposal for amendments to Regulation (EU) No 575/2013 of the European Parliament and of the Council as regards requirements for credit risk, credit valuation adjustment risk, operational risk, market risk and the output floor.

Also, national Regulator like Bank of Italy considers that, in a phase in which technological developments and the use of outsourcing are leading to an increase in operational risks, disregarding historical operational loss data reduces the sensitivity of the methodological approach to risk and could weaken the incentives for banks to adopt virtuous behavior[6]. Therefore, since ECB already introduced the idea of a gradual introduction of the Internal Loss Multiplier, banks should develop an efficient Loss Data Collection system that allows to historicize all the relevant information regarding the operational loss and perform simulation on the impact of the adoption of the ILM on the capital requirement.

Comparison between SMA and current approaches

The existing methods that banks can apply in order to determine the operational risk capital requirement generate distortions and do not allow to the stakeholder to correctly compare the exposure of the banks to the operational risk.
The SMA allows to balance:

- **simplicity** – lowering the complexity of the calculation of the capital requirement;
- **comparability** – granted by the fact that all institutions will use the same method;
- **risk sensitivity** – due to the introduction of the Business Indicator Component which considers the business model and the operational losses data recorded by the bank in a time horizon of 10 years.

EVALUATION CRITERIA	EXISTING APPROACHES	SMA
Simplicity	The Advanced Measurement Approach (AMA) involves complex modeling and data retrieval. The Traditional Standard Approach (TSA) requires to develop a procedure to allocate the Relevant Indicator in the different Business Line.	The calculation of the capital requirement using the SMA is performed through a simple algorithm. The SMA does not require the usage of external data nor scenario analysis.
Comparability	It is difficult to compare the operational risk capital requirement due to the lack of homogeneity of: - the Advanced Measurement Approach (AMA) models developed by the banks; - the procedure developed by the banks to allocate specific items in the Business Line using the TSA.	Greater comparability, granted by the application of the same algorithm to all banks, allows the regulator to identify and respond to potential systemic issues.
Risk Sensitivity	For banks that use Basic Indicator Approach (BIA) and TSA there is no correlation between the operational risk capital requirement and the operational losses occurred.	The Internal Loss Multiplier (ILM) component introduces a link between the operational risk capital requirement and the operational losses, increasing the risk sensitiveness.

Table 2: Comparison AS-IS vs SMA. Source: Reply elaboration

Main impact adopting SMA

With the entry into force of the Standardised Measurement Approach, all banks are required to update the Operational Risk Management Framework considering the impact of the new approach. In particular, the following impacts have to be considered:

- modelling of the capital requirement calculation as per the SMA
- development of an efficient and effective Loss Data Collection process including the building up of an historical losses database
- update of the internal processes and procedures related to the operational risk

Main Impacts	Cost/Effort		
	Governance	IT/application	Processes
Requirement calculation algorithm	Low	Medium	Medium
LDC	High	High	High
Update processes and procedure	Medium	Low	High

Table 3: SMA Impacts. Source: Reply elaboration

[6] La recente proposta della Commissione europea di modifica delle regole prudenziali per le banche: un quadro d'insieme e una prima valutazione - Intervento di Paolo Angelini (Vice Direttore Generale della Banca d'Italia) - Comitato Esecutivo dell'Associazione bancaria Italiana - Roma, 19 gennaio 2022

On the 26 September 2023, the Basel Committee on Banking Supervision published the Basel III Monitoring Report.[7] The analysis shows the change in Tier 1 Minimum Requirement Capital (MRC) due to the revisions to the operational risk standards.
More in details, the results of the analysis summarized in the table below highlights the specific impact (in percentage) of Internal Loss Multiplier (ILM) and Business Indicator Component (BIC) on Minimum Requirement Capital (MRC). The analysis emphasises that:

- **for AMA banks**, the business-driven BIC increased from 66,45% to 79,49% of the 2017 operational risk MRC. Although the loss component decreases similarly as the BIC increases**, due to the logarithm feature of the ILM, the final MRC of the new Standardised Approach (SA) is still increasing by almost 10% over the past six years**. According to the analysis, if these banks would use the Basic Indicator Approach (hypothetical BIA) instead of the AMA, the current MRC in 2022 would be 65,87%, i.e., about 40% lower than the current reported operational risk MRC.
- **for the non-AMA Group 1 banks (**that have Tier 1 capital of more than €3 billion)**[8], the hypothetical BIA is about 10,5% higher than current MRC**, which indicates that this cluster uses a less conservative approach to measure their risk exposure and benefit from the use of the current indicator-based approaches of ASA or SA.

	Group 1 AMA banks					Group 1 non-AMA banks				
	20k LC (lhs)	Hypothetical BIA	BIC	20k new SA	Reported operational risk MRC	20k LC (lhs)	Hypothetical BIA	BIC	20k new SA	Reported operational risk MRC
Years	Per cent of 2017 op risk MRC					Per cent of 2017 op risk MRC				
2017	244,34	58,01	66,45	93,92	100,00	165,75	107,85	118,58	107,50	100,00
2018	248,64	58,86	67,25	95,34	104,10	183,30	112,02	123,62	113,88	104,76
2019	245,12	60,50	69,39	97,03	108,19	208,00	116,08	129,02	120,11	107,08
2020	223,66	58,71	67,98	93,41	93,45	204,52	113,64	128,95	118,12	107,37
2021	211,40	62,22	73,41	97,96	99,27	214,37	126,63	144,92	129,88	119,03
2022	201,34	65,87	79,49	102,25	105,70	200,56	133,15	153,64	135,91	122,68

Table 4: Impacts of Internal Loss Multiplier and Business Indicator Component on Minimum Requirement Capital. Source: Basel Committee on Banking Supervision – Basel III Monitoring report September 2023 – Statistical annex

A different analysis shows the changes in operational risk MRC due to the use of SMA approach and considering the application of two national discretions:

- to set the internal loss multiplier equal to one and hence base capital requirements for operational risk solely on the business indicator component for all banks in a jurisdiction; and
- to have Bucket 1 banks (with BI ≤ EUR 1 billion)[9] measure their ILM using their loss history, rather than apply ILM = 1 to all Bucket 1 banks.

Changes in MRC for operational risk In per cent			
	With chosen approach	ILM = 1	20K 10Y
Group 1 banks	1.8	0.5	9.7
of which: Europe	25.9	21.3	58.3
of which: Americas	-3.1	-23.8	-3.1
of which: RW[10]	-11.0	24.1	-11.1
of which: G-SIBs[11]	-7.2	-10.2	0.0
Group 2 banks	6.8	10.0	20.7

Table 5: Changes in MRC for operational risk. Source: Basel Committee on Banking Supervision - Basel III Monitoring Report (September 2023)

[7] Basel Committee on Banking Supervision – Basel III Monitoring report September 2023
[8] Group 1 banks are those that have Tier 1 capital of more than €3 billion and are internationally active. All other banks are considered Group 2 banks
[9] Institutions have been divided into three buckets based on the BI thresholds. Bucket 1 consists of institutions with BI ≤ EUR 1 billion, bucket 2 consists of institutions with EUR 1 billion < BI ≤ EUR 30 billion and bucket 3 consists of institutions with BI > EUR 30 billion.
[10] RW: rest of the world
[11] G-SIBs: Global systemically important banks

It is possible to see that the final operational risk framework generates an aggregate small increase in operational risk MRC of approximately 1.8% for all Group 1 banks while an increase of 6.8% for the Group 2 banks[12]. In addition, at a regional level, Europe faces a significant increase of around 26%. However, if all banks used the less risk-sensitive BI component only ("ILM=1"), the operational risk MRC for Group 1 banks would slightly increase by 0.5%. If all Group 1 banks applied the ILM based on the average losses above €20,000 of the past 10 years ("20k 10Y"), the impact would be 9.7%. The comparison between ILM=1 and ILM 20k, on a regional level, shows that the MRC in Europe (delta of 37 percentage points) is the most affected by the revision to ORMF.

Operational Risk Management Framework (ORMF) components and SMA provisions

In 2014, the BCBS performed a review of the implementation of the Principles for the Sound Management of Operational Risk and in March 2021 published its revisions of these principles. The BCBS' purpose was to provide banks with guidance to facilitate the implementation of the principles in their operational risk management framework and ensure consistency with the new operational risk framework foreseen by the Basel III reforms. All the following Principles reflect sound practices relevant to all banks.

- **GOVERNANCE**

  For the purposes of correct management of operational risk, it is important to identify and integrate the roles and responsibilities of the governance structures (operational risk management functions, business units and support functions) that will be involved in the process of calculating the regulatory capital for operational risk.
  SMA PROVISION: an effective governance is a key component of the operational risk management framework in order to reduce the operational losses through an efficient decisional process that allows to manage the identified potential risks.

- **LOSS DATA COLLETION**

  In order to correctly manage the operational risk, it is important that banks' internal operational losses data respect the standards in terms of the soundness, quality and integrity of the data.
  SMA IMPACT: banks have to build up a database considering the general and specific criteria on loss data identification, collection and treatment. ILM calculation has to be based on a 10-year observation period and has to be linked to bank's current business activities, technological processes and risk management procedures. Moreover, the criteria used to identify and manage the operational losses have to be well documented.

- **RISK CONTROL SELF-ASSESSMENT**

  Banks should perform qualitative and quantitative self-assessments of their operational risks. These self-assessments are important to identify and understand the underlying causes of risks to which banks are most exposed (including "relevant risks") and to evaluate the effectiveness of controls that have been put in place to monitor such risks. Moreover, through this tool banks can carry out a forward-looking evaluation of potential risks to which could be exposed (for example, due to the offering of new services) and can give to the Top Management an overview of risks that bank is facing in order to take properly corrective actions.
  SMA IMPACT: an effective risk control self-assessment is essential to prevent unexpected internal operational losses from materialising. Since, banks' internal operational losses have become part of operational risk capital requirement, the risk assessment is important to identify risks to which banks could be exposed in the future, in order to implement controls related to those risks and, in this way, avoid that some occurring events could translate into substantial financial losses (increasing the bank's ILM).
  Moreover, while implementing the SMA, banks should evaluate to develop a continuous self-assessment process which would allow the monitoring and the timely analysis of the identified risks.

- **KEY RISK INDICATORS**

  In order to be always aware of and monitor own exposure evolution to operational risk on a daily basis, banks should develop and integrate metrics that provide early warning signs of trends, potential risks and vulnerabilities within the different areas of the bank. The KRIs, as a preventive measure, allow the timely implementation of possible mitigation measures following the worsening of the bank's level of risk.
  SMA IMPACT: well-defined key risk indicators will ensure the identification and the monitoring of threats and vulnerabilities related to banks' current operations, and provide a warning signal of the current and emerging risks. In other words, key risk indicators allow the banks to be aware of the risks that may arise before that such risks materialise in financial losses (increasing the bank's ILM).

- **RISK MITITGATION STRATEGY**

  Banks should have a strong control environment that uses appropriate risk mitigation and/or transfer strategies in those circumstances where internal controls do not adequately address risk. Management may complement controls by seeking to transfer risk to another party (such as through insurance) by carefully considering the extent to which risk mitigation tools actually reduce the risk involved. However, since risk transfer is an imperfect substitute for sound risk management controls and programs, banks should view risk transfer tools as complementary to, rather than substitute for, sound internal control of operational risk.

[12] Group 1 banks are those that have Tier 1 capital of more than €3 billion and are internationally active. All other banks are considered Group 2 banks

SMA IMPACT: for banks in buckets 1 and 2, in addition to the optimisation of the risk transfer procedures, it is necessary to develop an effective risk mitigation strategy to cut down the level of losses that might occur due to the materialising of an identified risk.

- **REPORTING SYSTEM**

Banks should have adequate reporting mechanisms in place to support proactive operational risk management. To effectively manage risk, the right information needs to be presented to the right people at the right time. Therefore, it is essential to identify a reporting model for the production and communication of summary and aggregated information, as well as detailed information, on the risks identified and assessed to allow the board of directors and senior management to make effective risk decisions.
SMA IMPACT: an efficient reporting system allows the management to be aware of the current and emerging risk in a timely manner enabling to establish a risk mitigation strategy aimed to correctly manage these risks and reduce the operational losses.

- **APPLICATION SOLUTION**

In order to respond to regulatory changes and managerial needs, banks should have application solutions to:
- strengthen the control system and continuously intercept the evolution of the risk profile;
- facilitate the monitoring of risk mitigation interventions from point of view of Continuous Assessment; and
- guarantee the traceability of the data and the replicability of the processes.

Thus, automated processes introduce risks that need to be addressed through robust technology governance and infrastructure risk management programs.
SMA IMPACT: with the introduction of the SMA, will be even more important to have an integrated tool that allows to manage all the data and processes related to the Operational Risk Management Framework (i.e., Loss Data Collection, Requirement Calculation, Risk Self-Assessment, Reporting, etc.).

Conclusion

In accordance with the final Basel III reforms package, the current approaches to Pillar 1 minimum capital requirements for operational risk are being replaced by the new Standardised Method Approach starting from January 1, 2025.
Considering that all banks within the European Union will have to move on the new method for calculating the regulatory requirement, it is necessary to point out that several qualitative requirements must be met in order to reduce inconsistencies in the capital requirement calculation process and in the supervisory reporting related to operational risk. Despite some differences that depend on the inherent characteristics of a bank, clear strategies supervised by management, a strong operational risk culture, effective risk management and internal reporting, and contingency planning are all crucial elements for building up and maintaining a sound operational risk framework for banks of any size and scope. In other words, **an efficient and consistent operational risk framework is essential for the proper assessment, prevention and mitigation of operational risk.**

The link between MiFID and Risk Appetite Framework as an application of best practices for wealth management and the entire value chain of the financial industry

Gianluca Macchia (Studio Tridente-Macchia); Emanuele De Angelis (Banca Generali S.p.A); Michele Vitagliano (Università di Bari)

Abstract

After a short review of the MiFID regulations and the RAF, the paper identifies the link between them which allows to mitigate a balance sheet risk sustained by the financial intermediary and, at the same time, to improve its stability and value creation, through a maximization of customer loyalty. The client's attitude towards risk can be summarized in these terms: "I don't like risk, but I like to win". Thus, a three-dimensional approach towards expected utility is suggested for estimating risk tolerance: risk aversion, loss aversion and reflection. In addition, a definition of the client's financial objectives is required, combined with greater disclosure - which allows the construction of a financial statement – to attest that risk-taking is indeed a luxury, as indicated by the metrics of the discretionary wealth ratio, and therefore, of the Standard of Living Risk (SLR). The next step consists in the determination of a set of portfolios along the efficient frontier where risk is represented by the expected shortfall, the determination of which belongs to a Generalized Extreme Value Theory logic. The client's objectives are described in terms of probability of success, where the latter is a function of an initial endowment, a potential positive contribution of financial resources over time as well as an expected return level. The above is expressed through a practical case that envisages the determination of a set of EGPF portfolios and the identification of the specific portfolio, obtained as a solution to a static and dynamic optimization problem, where the objectives have been formalized through the calculation of the associated utility.

1. Introduzione

L'operato del Regolatore europeo, in merito all'offerta di prodotti finanziari, può essere sintetizzato attraverso un insieme di concetti fondamentali. Più nello specifico, in un contesto di Risk Management, tali concetti chiave sono riassumibili nella relazione tra tolleranza al rischio (risk tolerance) e propensione al rischio (risk appetite). Negli ultimi anni, a seguito dell'attenzione dedicata alla tolleranza e alla propensione al rischio degli investitori, si è assistito all'introduzione del questionario MiFID, declinando questi concetti in una profilatura del cliente. Tale strumento si configura come il principale mezzo per acquisire informazioni dettagliate creando un legame tra conoscenza ed esperienza del cliente dei mercati finanziari e la sua propensione al rischio. Quasi contemporaneamente, in risposta alla Grande Crisi Finanziaria del 2007-2009, l'introduzione del Risk Appetite Framework (RAF) ha consentito agli istituti finanziari di dotarsi di pratiche volte a far interagire la misurazione degli obiettivi di rischio con la solidità patrimoniale così da migliorare, in ultima istanza, la stabilità del sistema finanziario nel suo complesso. Contrariamente a quanto possa emergere da un primo superficiale confronto, diviene fondamentale integrare la conoscenza e la profilatura del cliente acquisita tramite MiFID in un quadro più ampio che governa i principi cardine della rischiosità degli Istituti di credito: la definizione del Risk Appetite Framework non può non includere le informazioni richieste ed estrapolate dal questionario MiFID. Il processo di definizione della governance di rischio e del capitale risulta limitata qualora non sia pienamente armonizzata con l'evoluzione del business che, di concerto alle componenti di ricavo (gli asset), lavorano nella stesura dei piani industriali. L'attenzione al cliente, alle sue priorità ed alla propensione al rischio, rappresenta uno dei temi più complicati e più impattanti nella definizione di un sistema volto alla crescita sostenibile raggiungibile togliendo l'elemento aleatorio e volatile delle fonti di finanziamento che impattano la redditività stessa. Le ragioni di cui sopra sono alla base della ricerca che si vuole analizzare in questo lavoro, ovvero: l'intermediario finanziario potrebbe trarre notevoli vantaggi nella mitigazione del balance sheet risk se riuscisse a comprendere approfonditamente l'atteggiamento "reale" del cliente nei confronti del rischio, adottando, in fase di definizione degli obiettivi di investimento, un approccio che fa leva ed integra le informazioni richieste dal questionario MiFID. Oltre a ciò, una più approfondita comprensione delle necessità e degli atteggiamenti del cliente permetterebbe all'intermediario finanziario di conseguire una maggiore fidelizzazione con conseguenti benefici tangibili in termini di stabilità e crescita aziendale. Per conseguire tali obiettivi, è necessario, tuttavia, abbandonare la concezione tradizionale della finanza che considera l'investitore preoccupato solo del rapporto rischio e rendimento del portafoglio. Il concetto di rendimento e rischio deve essere unito al concetto di "utilità" valutata attraverso tre dimensioni: risk aversion, loss aversion e reflection mediante una modalità "lottery-style". Inoltre, si evidenzia come una maggior conoscenza del cliente rifletta una maggior probabilità di fidelizzazione che implica un maggiore livello di fiducia e trasparenza riguardo la gestione della sua completa situazione patrimoniale, economica e finanziaria, che consente di determinare una misura di discretionary wealth ratio e quindi di standard living risk (SLR). Il metodo proposto in questo lavoro si caratterizza come innovativo in quanto, utilizzerà un set di portafogli d'investimento posizionati sulla frontiera efficiente, considerando il rischio in termini di expected shortfall legata ad una distribuzione Frechét o Gumbel, appartenenti alla teoria delle extreme value (EVT). Tale scelta è dettata dalla maggiore resilienza di questa distribuzione rispetto a quanto utilizzato nelle attuali modalità che legano la determinazione della frontiera ottimale all'utilizzo di una distribuzione normale o lognormale approcciando il mercato in ottica "prudenziale". Inoltre, il raggiungimento degli obiettivi del cliente, in una versione realistica di multi-obiettivo, è espresso in termini di probabilità di successo, collegata direttamente a una dotazione iniziale, a contributi temporali ed a una misura di rendimento del portafoglio nel tempo. L'approccio metodologico proposto in questo lavoro viene declinato su un caso pratico in cui, a seconda dei livelli di risk aversion, loss aversion e reflection, si perviene alle diverse caratterizzazioni dei portafogli d'investimento (EGPF), ottenute attraverso la risoluzione di un problema di ottimizzazione statica finalizzato al raggiungimento, in base ai relativi livelli di importanza, degli obiettivi definiti con il cliente ed espressi mediante una funzione di utilità associata.

2. MIFID II e la futura MIFID III

I cambiamenti della regolamentazione finanziaria hanno cercato di seguire l'innovazione finanziaria che è giunta sui mercati e quindi sugli investitori, in particolare quelli retail. Come sappiamo tutto nasce nel 2007 quando la direttiva 2004/39/CE, meglio nota come MiFID, che intende armonizzare le diverse normative nazionali europee oltre a migliorare il sistema di tutela degli investitori. In particolare, va a dettare nuove regole relativamente alla condotta degli intermediari finanziari, classificando la clientela tra retail, qualificata e istituzionale, introducendo per la prima volta regole sui conflitti d'interesse ed inquadrando la consulenza finanziaria tra le attività riservate così come introduce l'obbligo da parte degli intermediari di comunicare la presenza d'inducements. Tutto ciò per arginare un'asimmetria informativa tra intermediari finanziari e clientela retail. Vengono introdotti per la prima volta concetti come adeguatezza ed appropriatezza, con la prima che fa leva sulla conoscenza ed esperienza in materia d'investimenti (1), situazione finanziaria – reddituale e patrimoniale (2), obiettivi d'investimento (3) e propensione al rischio (4). Appropriatezza che, invece, fa leva su un ridotto numero di elementi da parte dell'investitore. Rimane, tuttavia, ammessa un'operatività in execution only. La consulenza finanziaria viene definita attraverso tre livelli di servizio che si distinguono attraverso regole differenti di comportamento per l'intermediario che è tenuto a rispettare in luogo delle raccomandazioni fornite al cliente. Approvata nel 2014 con la direttiva 2014/65/CE ed entrata in vigore nel gennaio 2018, la direttiva MiFID II ha come scopo quella di migliorare la protezione dell'investitore attraverso un approccio maggiormente incentrato sulla trasparenza degli intermediari finanziari. In particolare, la nuova normativa detta i principali cambiamenti:

- **Product Governance & Target market**: viene previsto un controllo ex-ante gravante sia sulle società di gestione del risparmio (SGR) che devono definire un target market.
- **Requisiti professionali richiesti allo Staff**: per garantire un elevato *standard* di professionalità nel miglior interesse dei clienti, vengono declinati i requisiti professionali tali da assicurare il possesso di una qualifica adeguata ed una congrua esperienza.
- **Inducements**: nella prestazione del servizio di consulenza "su base non indipendente", si prevede che gli inducements debbano superare quanto previsto dal test di ammissibilità, che ne misura il beneficio tangibile per il cliente, in una logica di miglioramento della qualità del servizio adeguate.
- **Adeguatezza**: viene ristretto il campo dei prodotti finanziari per i quali sia ammesso il servizio "execution only"
- **Costi**: devono essere indicati al cliente in modo aggregato e regolare.

Per quest'ultimo punto, negli ultimi anni la normativa si è in parte focalizzata su una miglior trasparenza dei costi attraverso informative ex-ante ed ex-post mediante documentazioni dedicate che nella versione ex-post prende il nome di rendiconto MiFID dei costi. Parte del risultato, lato cliente, è il questionario MiFiD che cerca di ottemperare la definizione del profilo di rischio e quindi del livello di propensione e tolleranza al rischio, tenuto conto degli obiettivi finanziari che in tal sede è possibile desumere. Nell'attesa poi della Mifid III, il 25 maggio 2023 la Commissione europea ha pubblicato la Retail Investment Strategy che modifica il regolamento sui prodotti di investimento al dettaglio e assicurativi preassemblati (PRIIP). Essa pone l'attenzione certamente su alcuni temi, che saranno oggetto della normativa che verrà. In particolare, si prevede per i distributori di segnalare come essi debbano identificare e quantificare tutti i costi di distribuzione e valutare se quest'ultimi siano giustificati e proporzionati. Se un prodotto, insieme ai costi di distribuzione, si discosta dal benchmark, l'impresa sarà autorizzata a distribuirlo solo se ulteriori test e valutazioni stabiliranno che i costi e gli oneri sono comunque giustificati e proporzionati. Nel caso opposto, ossia se i distributori non potessero giustificare tali costi, dovrebbero prendere in considerazione la possibilità di ricorrere a fornitori alternativi. In regime di "consulenza finanziaria", le proposte d'investimento devono prevedere la raccomandazione di un prodotto più efficiente in termini di costi tra quelli identificati come adatti al cliente e che offrono caratteristiche simili. Devono inoltre raccomandare, tra la gamma di prodotti individuati come idonei, uno o più prodotti privi di caratteristiche aggiuntive che non siano necessarie per raggiungere gli obiettivi di investimento del cliente e che comportino costi aggiuntivi, dove i criteri saranno definiti attraverso successivi atti delegati. Relativamente agli inducements la decisione della Commissione Europea è stata quella di proporre un "approccio graduale" tale da permettere ai distributori di modificare i propri sistemi e ridurre al minimo i costi di un tale cambiamento. Si prevede, dopo un periodo di tre anni dall'adozione di tali misure, una valutazione sugli effetti degli incentivi con eventuali misure legislative che potrebbero portare ad un loro divieto. Inoltre, si afferma come l'adeguatezza e l'appropriatezza debbano essere considerati a livello di portafogli complessivo del cliente. Ma può essere evitata, e quindi i consulenti finanziari non necessitano d'informazioni sulle conoscenze ed esperienze del cliente o sulla composizione del suo portafoglio esistente, laddove siano interessati solo prodotti ben diversificati, non complessi ed efficienti in termini di costi. Quindi si mantiene l'adeguatezza solo nel caso di prodotti complessi.

3. Risk Appetite Framework

Come sappiamo, la governance interna è una delle principali priorità del Single Supervisory Mechanism (SSM) e uno degli elementi chiave del Supervisory Review and Evaluation Process (SREP). In un ambiente che si modifica – con particolare riferimento alle dimensioni economiche, finanziarie, competitive e soprattutto normative – è fondamentale considerare il proprio profilo di rischio complessivo in un'ottica di sostenibilità del modello, l'attenzione è sempre più spostata sulla capacità di implementare una sana governance. In poche parole, ciò vuol dire implementare le best practices relative al Risk Appetite Framework. Tale impostazione la troviamo rispettata in tutte le banche, che tuttavia vedono considerato un approccio proporzionato per tener conto delle dimensioni, del modello aziendale e del livello di complessità delle stesse. L'SSM pone alte aspettative nei confronti dei board delle banche, dove quest'ultimi hanno il compito di contestare, approvare e supervisionare l'attuazione da parte del management degli obiettivi strategici, della governance e della cultura aziendale. Per questo motivo, il board dovrebbe includere una prospettiva di rischio nelle discussioni strategiche e dimostrare un'efficace supervisione delle funzioni di rischio e controllo, e per questo motivo dovrebbe essere fortemente coinvolto nel processo di validazione e monitoraggio del Risk Appetite Framework. L'SSM si aspetta inoltre che le banche sviluppino e istituiscano un RAF completo, che dovrebbe aiutarle a rafforzare la consapevolezza del rischio e a promuovere un'adeguata cultura del rischio. Come prerequisito per una sana gestione del rischio, il RAF dovrebbe definire il livello di tolleranza al rischio che l'ente è disposto ad assumere in relazione sia ai rischi finanziari che a quelli non finanziari. I parametri e i limiti di rischio dovrebbero essere implementati in modo coerente all'interno delle entità e delle linee di business e dovrebbero essere monitorati e segnalati regolarmente al consiglio. Il RAF dovrebbe inoltre rimanere allineato al piano aziendale, allo sviluppo della strategia, alla pianificazione del capitale e della liquidità e ai piani di remunerazione degli istituti finanziari. L'introduzione del RAF in Italia è avvenuta con il 15° aggiornamento della circolare 263/06 della Banca d'Italia[1] e a livello europeo con l'emanazione da parte del Financial Stability Board (FSB)[2]. Esso altro non è che un documento strategico che ne identifica il processo di gestione del rischio, permettendo la creazione di

[1] "Nuove Disposizioni di Vigilanza Prudenzia le per le Banche" (Titolo V, (Titolo V, capitolo 7-"Il Sistema dei Controlli Interni"), emanato in data 2 Luglio 2013

[2] "Thematic Review on Risk Governance" del 12 Febbraio 2013 e "Principles for an effective Risk Appetite Framework" in del 18 Novembre 2013

una situazione di trasparenza e di coerenza tra i rischi accettati e gli obiettivi che s'intendono perseguire. Ciò, raccordandosi con gli altri processi esistenti (ICAAP, ILAAP, Capital Allocation, Strategic Plan, Recovery Plan e Remuneration Framework). Scopo dei rischi inclusi nel RAF dovrebbe essere la loro comprensibilità.

Figura 1: Fasi Risk Appetite Framework

Rispetto a documenti come lo Strategic Plan, il Risk Appetite è incentrato sui risk limits e policy mentre il secondo sarà incentrato sugli obiettivi di performance. In conformità con i principi del Financial Stability Board[3], l'istituzione di un RAF efficace è uno strumento strategico per rafforzare una forte cultura del rischio negli istituti finanziari, a sua volta fondamentale per un sound risk management. L'evoluzione normativa in materia di RAF ha spinto gli intermediari a rivedere ed aggiornare i processi aziendali così come da richiesta esplicita del regulator. Lapproccio ad oggi implementato è il Top-Down approach.

Figura 2: Top-Down Approach

Fonte: Ernst & Young "Risk Appetite Framework (RAF) - Basilea 3: risks and supervision 2014 " – Materiali Convegno Giugno 2014

Importanti sono poi i meccanismi di collaborazione tra il Chief Financial Officer e Chief Risk Officer in ottica RAF. In particolare, durante la fase di approvazione della struttura del risk indicator framework, il board fissa ex-ante gli obiettivi in termini d'indicatori come, ad esempio regulatory capital ratio. Tale fase avvia poi quella di pianificazione strategica ed operativa che sosterrà lo sviluppo del financial statement a livello patrimoniale e reddituale. Per il CFO porterà alla determinazione del piano industriale mentre per il CRO sarà utile per prendere visione delle ipotesi di proiezione degli aggregati del financial statement fornendo, allo stesso tempo, le proxy di rischio per andare poi a quantificare le misure di rischio incluse nel risk indicator framework[4], che si suddividono tra primari, complementari ed operativi. Da ultimo il CRO deve verificare che sia sempre rispettata l'adeguatezza patrimoniale oltre ad una sostenibilità del RAF stesso. Durante la fase di monitoraggio del Risk Appetite di breve e di lungo periodo si osservano flussi informativi – almeno su base trimestrale – relativi agli esiti delle verifiche funzionali al monitoraggio del RAF. Questi hanno ad oggetto indicatori come PdV, MINT, MOL per andare a calcolare gli assorbimenti di capitale oltre ad un monitoraggio dei livelli di risk appetite, risk tolerance, risk limits/early warnings.

4. Relazione Mifid II e RAF: un quadro integrato per la gestione del rischio lungo tutta la filiera cliente - intermediario

Certamente, ad una prima vista, i due approcci – Mifid da un lato e RAF dall'altro – servono scopi diversi: da un lato gli intermediari – banche in testa – e dall'altro gli investitori retail in massima parte. Ad ogni modo, ad un'analisi più attenta tali approcci condividono non solo alcuni concetti di fondo – gestione del rischio connesso ad obiettivi definiti – ma possiamo osservarne il fatto che un miglioramento della prima permette una miglior determinazione della risk-aversion oltre a poter meglio gestire i rischi dell'intermediario alla luce del balance sheet risk che può manifestarsi in particolari situazioni di stress del mercato finanziario che conducano a withdrawals importanti. A tal proposito verrà effettuata, non solo un'analisi comparativa di quanto già visto in ottica regolamentare ma si analizzeranno importanti questioni legate a: misure di rischio e correlation risk, assumptions lato risk

---

[3] Principles for an Effective Risk Appetite Framework, FSB, November 2013

[4] Dove nei primari abbiamo l'adeguatezza patrimoniale regolamentare (CET 1 Ratio, TIer 1 Ratio e Total Capital Ratio), liquidità (LCR e NSFR), leva finanziaria (Leverage Ratio), Performance Risk Adjusted (EVA, RAROC, RARORAC), ICAAP (Capital Adequacy Ratios). Nei complementari ritroviamo le diverse declinazioni di risk weighted assets a livello credito, market e controparte. Lato operativo è necessario guardare ai livelli di concentrazione del portafoglio lato geografico/rating/settori d'esposizione oltre a limiti VaR e limiti di perdite registrate per finire con la volatilità del NOPAT e la variazione del costo del capitale.

preferences/utilità attesa nell'intento di mostrare, a livello operativo, nell'ottica di servizio nell'ambito del private banking/wealth management. Tutto ciò, senza dimenticare come il miglioramento delle informazioni legate ai clienti vadano ad impattare positivamente in una dimensione che possiamo definire di servizio di wealth management, con tutto quello che consegue in termini di miglioramento del servizio e quindi di percezione del valore lato cliente.

4.1 Metodologie VaR: limiti, sfide e possibili risposte.

Un punto focale, considerando la qualità/quantità dei dati acquisibili dal questionario mifid, è il misalignment, in termini di risk aversion, tra il risultato ottenuto attraverso il questionario Mifid II ed il suo vero valore. Ciò espone inevitabilmente l'intermediario ad un balance sheet risk sino a comprometterne la propria stabilità finanziaria oltre ad una creazione di valore, in virtù di possibili withdrawals degli assets da parte dei clienti che possono accadere durante fasi di downturn economico-finanziario. Collegata a ciò, si rileva inoltre un aspetto relativo al vantaggio competitivo dell'intermediario stesso, identificabile in termini di valore del servizio di consulenza finanziaria che beneficerebbe molto di un allineamento della risk-aversion con gli obiettivi finanziari del cliente, e tenuto conto dei fattori di rischio da lui sostenuti, al fine di minimizzare bias comportamentali. Al fine di poter ottenere tale obiettivo è necessaria avere una maggiore conoscenza del cliente per poterne determinare la più adeguata funzione di utilità. Drivers di quest'ultima potranno poi essere sottoposti a stress test al fine di poterne analizzare gli eventuali impatti negativi. La conoscenza del cliente deve portare alla redazione di un financial statement corrente e previsionale, che incorpori gli obiettivi finanziari del cliente, ove quest'ultimi siano oggetto di adeguate analisi finanziarie. Accanto a ciò, si ritiene importante da un lato l'uso di modelli VaR, e non solo, che dimostrino una maggiore efficacia durante fasi complesse di mercato, e dall'altro, l'implementazione di una dashboard ragionata di misure risk-adjusted performance e di drawdown[5] che permettano di raggiungere i più elevati standard operativi in termini di market risk & wealth management. Sintetizzando quanto è possibile incontrare nell'ambito dell'offering attuale per i servizi di gestione di portafoglio per la clientela privata – private banking e wealth management – le metodologie di gestione del rischio utilizzate vedono l'uso di processi che, facendo perno sulla regolamentazione attuale, si affidano a metodologie d'identificazione del rischio di mercato che tendono a disattendere, in chiave ex-post, l'affidabilità promessa ex-ante soprattutto in presenza di fasi ribassiste aventi un impatto non trascurabile. Si fa riferimento a metodologie come quella della simulazione storica, eventualmente corretta attraverso aggiustamenti che conosciamo come age e weighted volatility, all'interno dell'insieme di tecniche note come EWMA[6]. Risultati simili si riscontrano con l'uso della metodologia parametrica, che, può prevedere una modalità di aggiustamento attraverso l'espansione di Cornish-Fisher[7]. In un'analisi dei pro e contro per l'investitore si rileva che: 1) La metodologia di simulazione storica – spesso effettuata con un frame giornaliero ed un periodo temporale di breve-medio periodo, mostra certamente una semplicità implementativa benché tenda a sottostimare il rischio nelle fasi di mercato rialzista (tail risk estremamente pronunciato) e riduca fortemente il budget risk – e quindi la possibilità di sfruttare in termini di rendimento potenziale futuro – non potendo prendere esposizioni in linea con i premi al rischio di una o più asset classes fin quando l'evento – market meltdown – non è uscito dal calcolo della media mobile. Nel caso della metodologia parametrica il tema è rappresentato dal ricorso all'ipotesi di normalità dei rendimenti. Tema a parte è l'utilizzo delle tecniche Monte-Carlo che permettono di poter eliminare ogni riferimento ma al tempo stesso richiedono maggiore complessità computazionale, laddove l'interesse è rivolto all'analisi dei risk factors[8]. Ciò premesso, le risposte a cui si giunge sono certamente nel senso di considerare una metodologia che cerca di superarne i limiti. Avendo a mente ciò è altresì interessante andare a considerare misure che siano rivolte al correlation risk come espressione poi delle misure VaR ed ES. Procedendo per gradi il primo aspetto è quello di ricorrere a tecniche appartenenti all'Extreme Value Theory. Elemento fondamentale della volatilità di portafoglio e VaR è certamente il correlation risk, le cui pairwise correlation devono essere considerate time-varying pena l'inaffidabilità dei modelli utilizzati. A tal fine è fondamentale utilizzare metriche che ci permettano di poter gestire in maniera efficace tale fonte di rischio per poter rispettare il mandato di rischio del cliente. Una possibile risposta a ciò è la Extreme Value Theory, che permette di determinare i livelli di rischio dell'investitore che siano decisamente più resilienti rispetto a situazioni negative di mercato. Ciò, in quanto una perdita di valore è generalmente molto costosa, sia per l'investitore che, potenzialmente, per l'intermediario che deve gestirne gli effetti, anche in termini di un balance sheet risk. Le peculiarità, che ne testimoniano una maggior resilienza rispetto ad un VaR storico o parametrico, sono rappresentate dalla loro struttura. Senza entrare troppo nei particolari è tuttavia opportuno dire quanto segue. Se l'attenzione deve essere posta maggiormente sui rendimenti negativi di un portafoglio/assets è cruciale identificare per il risk manager gli eventi estremi, che possiamo identificare in questo caso con cali del mercato azionario, ma non solo, di un'entità importante. A tal fine si devono assumere alcune distribuzioni di probabilità che sono differenti dalla normale, ampiamente utilizzata in finanza. Un approccio per la stima dei parametri è rappresentato dal teorema di Fisher-Tippett per il quale, l'aumento dimensionale del campione permette ai valori estremi una convergenza verso una distribuzione conosciuta come Generalized extreme value (GEV).

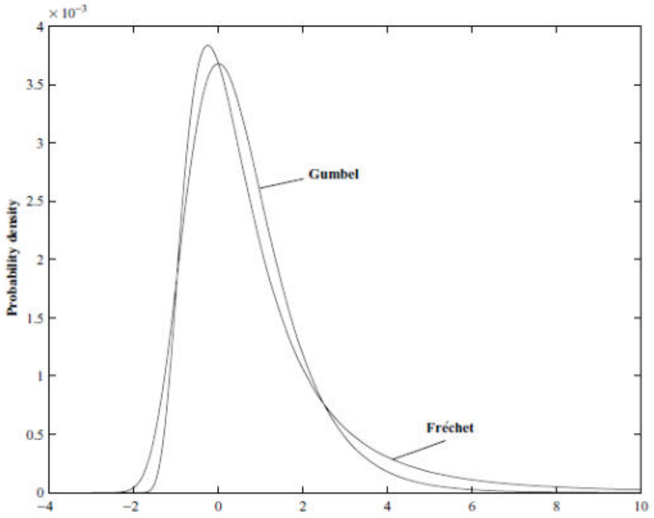
$$\boldsymbol{H_{\xi,\mu,\sigma}} = \begin{cases} \boldsymbol{exp}\left[-\left(\mathbf{1}+\boldsymbol{\xi} * \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] & \boldsymbol{\xi \neq 0} \\ & \boldsymbol{se} \\ \boldsymbol{exp}\left[-\,\boldsymbol{exp}\left(\frac{x-\mu}{\sigma}\right)\right] & \boldsymbol{\xi = 0} \end{cases} \qquad \text{[Eq. 1]}$$

Dove x soddisfa la condizione $1+\frac{\xi(x-\mu)}{\sigma} > 0$. Questa distribuzione ha tre parametri:
μ: $location\ parameter, che\ misura\ la\ central\ tendency\ di\ M_n$
σ: $scale\ parameter, che\ misura\ la\ dispersione\ di\ M_n$
ξ: $tail\ index, che\ misura\ la\ shape\ della\ tail$

---

[5] Tali misure riguardano sia il modo d'intendere la performance del portafoglio in un'ottica money weighted return (MWR) che in termini di rendimento aggiustato per il rischio (RAP).
[6] Exponential Weighted Moving Average
[7] Ottiene un miglioramento della skewness e kurtoris
[8] Risk factors time-varying di cui interessa la struttura di correlazione nel tempo

Figura 3: Frechèt & Gumbel a confronto



Fonte: Gunter Meissner "Correlation Risk Modeling and Management" 2014 Wiley

Dei tre casi generali conosciuti[9] della GEV distribution, interessa analizzare il caso in cui:

- $\xi > 0$, la GEV diventa una Frechet distibution e le tails sono pesanti (heavy) come nel caso della t-distribution e pareto distribution
- $\xi = 0$, la GEV diventa una Gumblet distibution e le tails sono piccole (light) come nel caso della normal e lognormal distribution

Nei mercati finanziari esistono solo la prima e la seconda situazione descritta e quindi sono importanti le distribuzioni di Frechét e quella di Gumbel e le loro rispettive procedure di stima. In merito, tre sono le modalità base per effettuare tale scelta:

1) I ricercatori sono fiduciosi di una parent distribution
2) Il ricercatore applica una statistica test per non rigettare l'ipotesi $\xi = 0$. In questo modo il ricercatore usa l'assunzione che $\xi = 0$
3) Il ricercatore potrebbe desiderare essere conservativo ed assumere che $\xi < 0$ per evitare un model risk

Tabella 1: Declinazioni Generalised EVT VaR

FRECHÉT	GUMBEL
$\boldsymbol{\mu - \frac{\sigma_n}{\xi_n}\left[1 - (-n\,ln(p))^{-\xi_n}\right]}$ [Eq. 2]	$\boldsymbol{\mu_n - \sigma_n \ln\,[-n\ \ln(p)]}$ [Eq. 3]

Da notare che con una shape compresa tra 0,2 e 0,4 le tails spariscono più lentamente rispetto alla normale. In questo caso il delta normal approach ed il metodo della simulazione storica sottostimano il rischio mostrando code più grasse. Importante è la calibrazione dei modelli osservati, che possono essere stimati con più modalità[10]. Soluzione alternativa, sempre all'interno della letteratura EV, è la Peaks-Over-Threshold Approach (POT). Quest'ultima richiede pochi parametri rispetto a quelli richiesti per la EVT. Essa fornisce il modo naturale per modellizzare valori che sono più grandi o più piccoli del livello soglia. In questo modo, la POT corrisponde alla GEV theory. La POT inizia con il definire una variabile casuale X come perdita. Si definisce u come soglia per valori positivi e la distribuzione delle excess losses oltre la soglia u come:

$$\boldsymbol{F_u(x) = P[X - u \le x | \mathrm{X} > \mathrm{u}] = \frac{F_u(x+u) - F_u(x)}{1 - F(u)}} \qquad \textbf{[Eq. 4]}$$

per x > 0. Ciò fornisce la probabilità che una perdita superi la soglia u al massimo fissata in x. La distribuzione di X stessa può essere trasformata nelle forme più comunamente utilizzate come la normale, log-normale, t, ecc.. Ad ogni modo, come u diventa grande, il teomena Gnedenko-Pickands-Balkema-deHaan (GPBdH) indica che la distribuzione $\boldsymbol{F_u(x)}$ converge ad una generalised Pareto distribution, data da:

$$\boldsymbol{G_{\xi,\beta}(x) = \begin{cases} 1 - \left[\left(1 + \frac{\xi x}{\beta}\right)^{\frac{1}{\xi}}\right] & \xi = 0 \\ 1 - \left[-\frac{\xi}{\beta}\right] & \xi = 0 \end{cases}} \qquad \text{[Eq.5]}$$

Generalized Pareto Distribution mostra una curva che va sotto la distribuzione normale prima della coda. Essa poi si muove sopra la distribuzione normale fino a che raggiunge la coda estrema. La Generalised Pareto Distribution poi fornisce un'approssimazione lineare della coda. Uno degli obiettivi di usare la POT è calcolare il VaR, da quale poter determinare l'Expected Shortfall, anche chiamato come Conditional VaR:

[9] $\xi < 0$, la Generalised EVT diventa una Weibull distibution e le tails sono ancor più piccole (lighter) rispetto alla normal distribution. Tale situazione in accade mai in finanza.

[10] Measuring Market Risk, 2th edition Kevin Dowd

$$VaR = \mu + \frac{\beta}{\xi}\left\{\frac{n}{N_u}[1-(p)]^{-\xi} - 1\right\} \quad \textbf{[Eq. 6]}$$
$$Expected\ Shortfall = \frac{VaR}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \quad \textbf{[Eq. 7]}$$

Interessante poi osservare un'applicazione della multivariata EVT, al fine di mostrare come i valori estremi possano essere interdipendenti.

4.2: Utilità attesa e schemi di preferenze di portafoglio alternativi nel wealth management

Tema centrale in una "client's perspective" è rappresentato innanzitutto dall'acquisizione delle informazioni in un ampio spettro da parte del cliente. Ciò permette di: poter implementare soluzioni d'investimento maggiormente aderenti alla reale determinazione della funzione d'utilità del cliente che permettano il raggiungimento dei suoi obiettivi finanziari (1), minimizzare il balance sheet risk dell'intermediario, riducendone più possibile i withdrawals dei clienti in contesti di meltdown finanziari ed in termini ampi anche un eventuale rischio reputazionale (2). Nella finanza classica, che risulta essere ad oggi pesantemente impiegata, si utilizza una tradizionale funzione d'utilità – chiamata power utility – dove la risk aversion è rappresentata dal solo parametro γ:

$$\boldsymbol{U = 2 - W^{(1-\gamma)}} \quad \text{[Eq. 8]}$$

dove W è la variazione di ricchezza in un singolo periodo 1 + r, r è il singolo periodo rendimento del portafoglio e γ è parametrizzato tra 1,5 e 12. Tale funzione soddisfa due elementi chiave: 1) più ricchezza è sempre preferita (derivata prima dell'utilità è sempre positiva) 2) il beneficio dell'extra ricchezza svanisce all'aumentare della ricchezza posseduta (derivata seconda dell'utilità è sempre negativa). Tale funzione è approssimata bene dall'approccio media-varianza per bassi livelli di risk aversion o distribuzioni di rendimenti normali. Ad ogni modo, nell'affermare che nessun investitore razionale – anche in presenza di una bassa risk aversion – è agnostico rispetto ad una moderata skew, la power utility function necessita di considerare ulteriori elementi. Prova ne è il fatto che durante la GCF dove molti clienti ed advisors avrebbero certamente preferito evitare di avere portafogli particolarmente esposti ad una skew negativa andando ad utilizzare soluzioni tattiche che avrebbero certamente minimizzato effetti negativi in termini di withdrawals e redditività. Negli anni, si è fatta strada la PT (Kahneman & Taversky 2004) che cerca di essere la risposta alla frase molto spesso detta dagli investitori "I don't like risk, but I live to win". Tale teoria prevede un processo in cui si massimizzi l'utilità attesa dell'investitore mentre si minimizzano gli effetti dell'estimation error. A livello pratico, essa prevede di utilizzare le reali preferenze degli investitori, considerandolo irrazionale. Ciò ci porta a considerare funzioni di utilità che incorporino ulteriori fattori. Versione "ultima" della funzione di utilità, che ci permette di poter analizzare la problematica, è la S – Shape Utility function che considera, ovviamente i parametri della kinked utility function[11], integrandoli. Risultato è l'aver tre "dimensioni" delle preferenze di rischio: γ: che misura la risk aversion λ: che misura la loss aversion, che si nota quando un investitore è decisamente più sensibile ad una perdita che ad un gain φ: che misura la reflection, che si ha quando gli investitori sono avversi al rischio solo quando si tratta di gain e cercano il rischio solo quando tratta di perdite. In formule:

$$\boldsymbol{U} = \begin{cases} \boldsymbol{2 - W^{(1-\gamma)}\ \ per\, r \geq 0} \\ \boldsymbol{2 - \lambda W^{(1-\gamma)}\ \ per\, r < 0, \varphi = 0} \\ \boldsymbol{2 - \lambda(2-W)^{(1-\gamma)}\ \ per\, r < 0,\ \varphi = 1} \end{cases} \quad \text{[Eq. 9]}$$

Risultato: la derivata seconda non è più negativa per qualsiasi rendimento e la preferenza dei momenti sarà determinata da forze concorrenti della risk aversion, loss aversion e reflection. In termini di asset allocation, si passa da una modern portfolio theory dove gli input sono i rendimenti, le varianze e covarianze che attraverso un ottimizzatore con fissate regole di minimizzazione delle varianze attese e massimizzazione dei rendimenti attesi, ad una impostazione dove la massimizzazione dei rendimenti sono basati sull'utilità attesa. Quest'ultimi sono il risultato di un possibile risultato congiunto di tutti gli assets, ossia dei loro rendimenti[12]. Tale approccio non considera la creazione di una frontiera efficiente in virtù del fatto dell'evidenza dei tre parametri dell'utilità definiti dall'investitore che determinano un solo portafoglio appropriato. Misurare la risk preference dell'investitore vuol dire specificare il profilo della funzione di utilità in tre dimensioni precedentemente richiamate, calibrandole attraverso un questionario lottery-style (Holt & Laury, 2002). La determinazione della risk aversion, γ, così come per le altre è possibile stimarla attraverso l'utilizzo di un questionario. In questo caso vengono poste 5 domande all'investitore, con le quali sarà possibile determinarne anche la curvatura della power utility function.

Tabella 2: Questionario propensione al rischio del cliente per determinazione risk aversion

Domanda 1: tra le seguenti due opzioni, quali preferisci?

Scelta A: 50% di vincere 2€ ed il 50% di vincere 1,60€
Scelta B: 50% di vincere 3,85€ ed il 50% di vincere 0,10€

Domanda 2: tra le seguenti due opzioni, quali preferisci?

Scelta A: 60% di vincere 2€ ed il 40% di vincere 1,60€
Scelta B: 60% di vincere 3,85€ ed il 40% di vincere 0,10€

Domanda 3: tra le seguenti due opzioni, quali preferisci?

Scelta A: 70% di vincere 2€ ed il 30% di vincere 1,60€
Scelta B: 70% di vincere 3,85€ ed il 30% di vincere 0,10€

Domanda 4: tra le seguenti due opzioni, quali preferisci?

Scelta A: 80% di vincere 2€ ed il 20% di vincere 1,60€
Scelta B: 80% di vincere 3,85€ ed il 20% di vincere 0,10€

---

[11] Presenta lo stesso profilo della power utility quando i rendimenti sono positivi. Quando i rendimenti sono negativi l'utilità cade ad un ratio λ preferendo una skew positiva quando $\lambda > 1$.

[12] A livello iniziale è possibile utilizzare rendimenti storici, avente una frequenza mensile. "Modern Asset Allocation for Wealth Management"

Domanda 5: tra le seguenti due opzioni, quali preferisci?

Scelta A: 90% di vincere 2€ ed il 10% di vincere 1,60€
Scelta B: 90% di vincere 3,85€ ed il 10% di vincere 0,10€

La modalità con cui però è possibile assegnare ad un investitore una precisa risk aversion passano generalmente attraverso l'utilizzo di tecniche di self-assessment della risk tolerance, experience con rischio, etc. Dal punto di vista accademico della progettazione dei questionari, la capacità di un test di misurare la caratteristica prevista è nota come "validità" (Grable, 2017) che tuttavia rimane difficile da misurare. La determinazione della loss aversion, λ – che varia tra 1 (no loss aversion) a 3 (massima loss aversion), ci permette di determinare la curvatura della power function utility, avviene ricorrendo ad un questionario[13] costruito nella seguente modalità:

Tabella 3: Questionario propensione al rischio del cliente per determinazione loss averion

Domanda 1: viene lanciata una moneta equilibrata. Con testa perdi 3 € e con croce vinci 6 €

Scelta A: Accetti di giocare la lotteria
Scelta B: Rifiuti di giocare la lotteria

Domanda 2: viene lanciata una moneta equilibrata. Con testa perdi 4 € e con croce vinci 6 €

Scelta A: Accetti di giocare la lotteria
Scelta B: Rifiuti di giocare la lotteria

Domanda 3 viene lanciata una moneta equilibrata. Con testa perdi 5 € e con croce vinci 6 €

Scelta A: Accetti di giocare la lotteria
Scelta B: Rifiuti di giocare la lotteria

Domanda 4: viene lanciata una moneta equilibrata. Con testa perdi 6 € e con croce vinci 6 €

Scelta A: Accetti di giocare la lotteria
Scelta B: Rifiuti di giocare la lotteria

Dove rifiutare la prima domanda comporterà, dovendo necessariamente rifiutare le successive, di ottenere il valore più elevato della variabile. Diversamente, accettare tutte le domande comporterà un suo valore più contenuto.
Per quanto attiene poi la misurazione della reflection, φ, si sottolinea come tale variabile può assumere un valore binario tra 0 ed 1, contrariamente alla loss aversion e risk aversion. In questo caso, l'investitore viene posto davanti a due domande identiche, dove la piccola differenza è costituita dalla parola guadagno, presente nella prima domanda, e perdita, presente in quella successiva:

Tabella 4: Questionario propensione al rischio del cliente per determinazione reflection

Domanda 1: tra le seguenti due opzioni, quali preferisci?

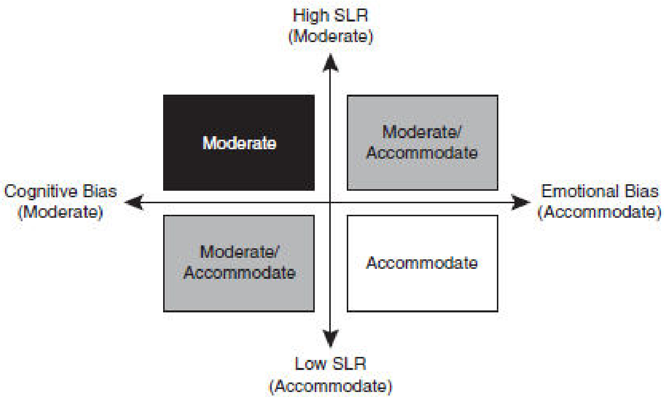
Scelta A: Sono certo di guadagnare 20 €
Scelta B: Ho 1/3 delle possibilità di poter guadagnare 60€

Domanda 2: tra le seguenti due opzioni, quali preferisci?

Scelta A: Sono certo di perdere 20 €
Scelta B: Ho 1/3 delle possibilità di poter perdere 60€

Nell'includere gli obiettivi finanziari dell'investitore, si applica l'idea del goal based wealth management che vuol dire effettuare un override moderato dei tre parametri della funzione di utilità. Al di là di moderare i behavioral bias come SLR aumenta, si raccomanda l'adozione di un framework che raccomandi agli advisor di essere meno disposti a compromessi rispetto agli emotional bias[14] e cognitive bias[15].

Figura 4: Standard Living of Risk & behavioral bias



Fonte: Modern Asset Allocation for Wealth Management

[13] Scelte possibili è tra il questionario multi-dimensionale qualitativo (Grable, Lytton 1999) o attraverso una soluzione che espliciti le preferenze individuali (Guillemette, Yao, James, 2015)
[14] Guidati da risposte nel questionario da risposte subconscie difficili da rettificare s
[15] Guidati da una mancanza d'informazione Esempi sono il self-control bias e l'endowent bias

L’elemento chiave è che quando sono presenti obiettivi finanziari, acquistare rischio dall’essere una preferenza personale diviene un lusso. Ciò è facilmente intuibile una volta che viene introdotta una misura di discretionary wealth relativa che chiamiamo SLR (Wilcox, Horvitz DiBartolomeo, 2006) formalmente rappresentata come:

$$\boldsymbol{SLR} = \mathbf{100\%} - \frac{\boldsymbol{Discretionary\ Wealth}}{\boldsymbol{Total\ Assets}} \qquad \text{[Eq. 10]}$$

dove il discretionary wealth è la differenza il totale attività ed il totale delle passività dell’investitore desumibile da un balance sheet model. Importante è l’interpretazione di tale valore. Innanzitutto, esso può essere negativo o positivo. In tal caso, prima di parlare di risk taking o no risk taking, è opportuno portarsi in area positiva attraverso modificazioni delle attività dell’investitore come allungamento del periodo temporale per ottenere l’obiettivo e/o riduzione del medesimo. Portato tale valore in territorio positivo, ci possiamo trovare nella situazione dove l’investitore mostra un basso livello di risk-taking nel caso di un basso livello di discretionary wealth ratio che inserito nella formula precedente ci restituisce un valore dell’SLR alto. Ovviamente, solo nel caso opposto si avrà un livello elevato di risk taking. Ciò sottolinea l’importanza della misura relativa di discretionary wealth al posto di un suo valore assoluto dove è importante però andare a determinare lato assets, il present value dell’employment capital less comsuption e lato liabitities, il present value del retirement capital.

Tabella 5: Balance Sheet Model dell’investitore

ATTIVITA'	PASSIVITA'
Liquidità **Portafoglio investito (securities)** Quota esposizione azionaria Quota esposizione obbligazionaria/credito Quota esposizione investimenti alternativi (private equity) **Attività immobiliari** Prima casa ed altri asset immobiliari Assets immobiliari **Valore attuale pensione pubblica** **Valore attuale Human Capital**	**Debiti fiscali a breve termine** **Debiti bancari a breve termine** **Mutuo per ristrutturazione immobiliare** **Valore attuale obiettivo tenore vita desiderato** **Equity o surplus capital**
Totale Attività	Totale Passività + Equity

DISCRETIONARY WEALTH
DISCRETIONARY WEALTH = ATTIVITA' - PASSIVITA'
SLR = 1 - (DISCRETIONARY WEALTH/TOTALE ATTIVITA')

Asset immobiliari: investimenti effettuati direttamente sul comparto immobiliare (case ed altre proprietà)
Valore attuale pensione pubblica: ammontare totale del montante contributivo ottenuto attraverso attualizzazione della rendita pensionistica individuale
Valore attuale Human Capital: capacità dell’investitore di produrre reddito attraverso l’attività professionale. Si ottiene attraverso l’attualizzazione dei flussi di reddito attesi ad un tasso di attualizzazione “adjusted”, ossia risk free rate + un premio per l’incertezza dello scenario ipotizzato.
Valore attuale “obiettivo tenore di vita desiderato”: obiettivo finanziario determinato dal cliente rispetto ai propri desiderata. Tale valore rappresenta la liability da ottenere nel momento in cui il cliente andrà in pensione
Capital o surplus capital: differenza tra le attività e le passività.

In una rappresentazione delle tre dimensioni e degli obiettivi finanziari fissati, si specifica il profilo della funzione di utilità e sue “correzioni” rimanendo coerenti con la misura SLR e quindi di discretionary wealth ratio per definire la composizione dei diversi portafogli, che considererà il livello di VaR e Conditional VaR tollerabili.

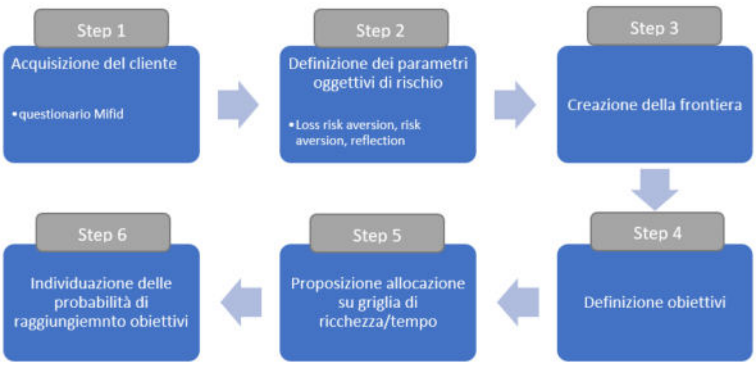
4.3 Dagli obiettivi ai rischi: Un caso pratico

La profilatura di un cliente è notevolmente influenzata non solo dalla ricchezza economica che possiede, ma sicuramente anche dagli obiettivi[16] reali che vuole raggiungere negli anni. Uno dei lavori del banker è quello di far conciliare questi due aspetti la cui causalità è ben chiara, ma può spingere a comportamenti irrazionali e reazioni incontrollate in fasi di mercato non favorevoli perdendo di vista la tutela del patrimonio. Definito il profilo di rischio del cliente è necessario successivamente focalizzarsi sugli obiettivi finanziari, ricordando che, erroneamente, si tende spesso a definire strategie di investimento non coinvolgendo gli obiettivi dell’investitore all’interno del processo di allocazione. Quanto proposto dalla teoria moderna del GBWM (Goal based wealth management) si occupa della definizione concreta di obiettivi a medio e lungo termine e successivamente della strategia di investimento utilizzata per raggiungerli. Introducendo un nuovo quadro di riferimento per la gestione patrimoniale basata sugli obiettivi, coerente con la modern portfolio theory oltre che con gli sviluppi della finanza comportamentale, il rischio è inteso come probabilità di non raggiungere i propri obiettivi, e non solo come la deviazione standard dei portafogli degli investitori. A tal scopo risulta essere utile l'approccio suggerito in cui, derivato il profilo di rischio, si instaura un dialogo continuativo con il cliente definendone i suoi obiettivi, l’orizzonte temporale, i possibili costi $c_k(t)$ distribuiti lungo il periodo di investimento per raggiungerli e dell’importanza attribuita (cioè

[16] Nella trattazione il concetto di “obiettivi” è riferito a beni materiali/immateriali che l’investitore è interessato ad acquisire nel corso del tempo. Gli obiettivi saranno associati al processo di investimento idoneo per ottenerli.

“l’utilità[17]”) che ognuno di questi ha per l’investitore. A riguardo è fondamentale il paper di Markowitz (1952) nella definizione della nozione di “frontiera efficiente” come l’insieme dei migliori portafogli possibili che rappresentano il compromesso ottimale tra rischio e rendimento. Rimanere sulla frontiera efficiente riduce al minimo anche il rischio GBWM poiché, per un livello fissato di volatilità, desideriamo il rendimento atteso ottimale per raggiungere gli obiettivi. Pertanto, idealmente, consideriamo solo i portafogli sulla frontiera efficiente, sebbene l’algoritmo funzioni altrettanto efficacemente se siamo limitati a un insieme di portafogli che non si trovano sulla frontiera. Il risultato di Markowitz per quanto riguarda le allocazioni ottimali del portafoglio è stato esteso e volto a massimizzare l'utilità della ricchezza patrimoniale dell'investitore W (Merton (1969); Merton (1971)). L'ostacolo principale all'attuazione pratica di questo approccio è la determinazione di una funzione di utilità, che è specifica dell'investitore. La funzione di utilità dell’investitore può essere in contrasto con la tradizionale teoria della funzione di utilità, come mostrato, nella Prospect Theory (Kahneman e Tversky (1979)) al pari della practices da seguire per incorporare in modo appropriato le considerazioni comportamentali nell'ottimizzazione. Per fare ciò, creiamo inizialmente un'approssimazione potenzialmente molto grossolana per l'utilità (preferenza) assegnata a ciascun obiettivo. L'algoritmo ottimizza il valore atteso della somma delle utilità derivanti dagli obiettivi raggiunti e, attraverso una strategia di ottimizzazione, calcola le probabilità corrispondenti di raggiungere ciascun obiettivo. L'output dell'algoritmo è una tabella che raccorda obiettivi con le probabilità ottimizzate di raggiungimento per ognuno di essi. Il fatto che il risultato sia espresso in termini di probabilità, anziché di utilità attesa, è la chiave per comunicare in modo efficace con gli investitori. La ricerca comportamentale mostra che gli investitori comprendono meglio il concetto di probabilità di raggiungere un obiettivo rispetto alla maggior parte dei termini finanziari comunemente utilizzati dai gestori patrimoniali. Gli investitori sono in difficoltà nel comprendere il significato diretto di un’associazione obiettivo/utilità: non risulta intuitivo. È molto complesso far associare ad un investitore il suo livello di priorità di un obiettivo ad un valore quantificabile, rappresentato dall’utilità. Lo scopo della trattazione è quello di dialogare con l’investitore, trasformando le diverse assegnazioni di utilità ai vari obiettivi in corrispondenti livelli probabilità di raggiungerli. Più specificamente, se l’investitore ritiene di voler aumentare la probabilità di raggiungere uno specifico obiettivo lo farà con la consapevolezza che abbasserà la probabilità di raggiungere la maggior parte degli altri obiettivi. Se desiderano aumentare la probabilità di raggiungere tutti i loro obiettivi, possono decidere di aggiungere capitali ai loro investimenti ($injection, I_i$) in qualsiasi periodo di tempo specifico o insieme di periodi di tempo, oppure possono espandere la gamma di strategie di portafoglio di investimento a loro disposizione. L'investitore può quindi utilizzare l'algoritmo per ricalcolare rapidamente le nuove probabilità ottimizzate per il raggiungimento di ogni obiettivo. Questa procedura può essere ripetuta quante volte si desidera. Questo processo di loop ci consente di determinare direttamente le vere preferenze dell'investitore tra i propri obiettivi, invece di indovinarli o presumerli. Questo approccio fornisce la corretta estensione per comprendere il rischio GBWM nel contesto di molteplici obiettivi concorrenti. Non è più una probabilità unica. Ora è una raccolta di probabilità che competono tra loro, e questo rischio complessivo è minimizzato dalla completa soddisfazione del maggior numero possibile di obiettivi dell'investitore, ponderati in base alla loro importanza per l’investitore. Notiamo che poiché utilizziamo la programmazione dinamica (un processo *backward*), la natura di ciascun obiettivo non può dipendere dalla realizzazione di obiettivi precedenti, perché, in questo caso, si tratterebbe di un fenomeno *forward* che va avanti nel tempo. Ciò costituisce la limitazione più rilevante del metodo in questo articolo. Ad esempio, se un investitore ha l'obiettivo di acquistare un'auto in un dato momento, possono acquistare l'auto più bella che hanno considerato (obiettivo completo) oppure auto meno belle (obiettivi parziali) o non acquistare l'auto. Ma se dovessero decidere di non acquistare l’auto, non possono spostare l'obiettivo dell'auto a un anno successivo con il metodo descritto in questo documento. Riassumiamo il processo di creazione del profilo di rischio fino alla gestione degli obiettivi con questo diagramma:

Figura 5: Diagramma di processo wealth management



All’interno dello step 5 “Proposizione allocazione su griglia Ricchezza/tempo”, individueremo due tipi di soluzioni: una soluzione “dinamica” (GWBM) in relazione all’evoluzione del tempo e della ricchezza ottenuta ed una soluzione “statica”, proposta dalla frontiera EGBW dove i portafogli suggeriti verranno mantenuti lungo l’arco temporale dell’investimento; entrambe le soluzioni, aventi gli stessi input, individueranno la probabilità di raggiungimento degli obiettivi indicati.

4.3.1 GWBM: pilastri e risultati dell’ottimizzazione dinamica degli obiettivi

L’algoritmo offre una soluzione “dinamica” al problema di massimizzazione delle probabilità di raggiungimento degli obiettivi: determinando ex-ante il plateau di portafogli su cui il cliente può investire[18], ad ogni condizione di mercato che affronteremo, il calcolo

[17] Non importa capire come definire l’utilità di un obiettivo. Importante è capire come questa utilità sia un numero che indichi la priorità di un obiettivo e la sua importanza relativamente ad altri. Nella trattazione completa infatti è possibile non solo individuare degli obiettivi “All or nothing”, cioè singoli in diversi istanti temporali, ma anche di obiettivi “Partial goals” in cui per un istante temporale il cliente può essere chiamato a decidere quale spesa sostenere in base alla ricchezza del momento.”
[18] Nel proseguo della trattazione si tornerà sulla definizione dei portafogli ammissibili per il cliente data la sua rischiosità.

ci offrirà su quale portafoglio dovrà investire l'investitore. Per semplicità, qualora avessimo ad esempio un solo obiettivo G da raggiungere al tempo T stiamo cercando:

$$\max_{\{A(0),A(1)\dots A(T-1)\}} Pr(W(T) > G) \qquad \text{[Eq. 11]}$$

di massimizzare la probabilità, attraverso la proposizione di una serie di allocazioni nel tempo $(A(0), A(1)..A(T-1))$, che permetteranno alla scadenza T di avere la ricchezza disponibile W(T) per porter soddisfare l'obiettivo prefissato.
Per risolvere questo problema con la programmazione dinamica dobbiamo impostarne gli input e lo faremo svolgendo un caso pratico per facilitarne la comprensione.

- *Orizzonte temporale di investimento*

Consideriamo periodi di tempo $t = 0,1,\dots,T$ con un intervallo di h anni tra loro. Quindi se h = 0,25, allora $t = 4$ corrisponde a un anno da oggi, $t = 0$. Il periodo di tempo finale, $t = T$ per il portafoglio può corrispondere o meno alla data di fine investimento prevista per l'investitore.

- *Definire gli obiettivi dell'investimento*

Più recentemente, il settore Wealth Management ha sostenuto l'uso di una gestione patrimoniale basata sugli obiettivi (GBWM), dove gli investitori scelgono i portafogli di investimento sulla frontiera efficiente di Markowitz con l'obiettivo di massimizzare la probabilità di raggiungere i propri obiettivi finanziari piuttosto che massimizzare l'utilità o ottimizzare direttamente il proprio trade-off tra volatilità e rendimento atteso. Una vasta letteratura ha incorporato ed esteso l'ottimizzazione efficiente del portafoglio in base alla media-varianza, sia in contesti statici che dinamici, ed è ampiamente utilizzata nella pianificazione pensionistica. Nonostante questo, siamo soliti considerare la sola volatilità del rendimento del portafoglio come unica misura di "rischio" e non la possibilità che un investitore non raggiunga i propri obiettivi; sicuramente diminuendo la volatilità può aumentare la possibilità che l'investitore non raggiunga i propri obiettivi. Proprio quando la ricchezza nel portafoglio è insufficiente per raggiungere tutti gli obiettivi, tuttavia, l'investitore avrà bisogno di un modo per (ri)ponderare i propri obiettivi per istruire l'algoritmo su quali obiettivi dare priorità massimizzando l'utilità attesa funzione di quanti sono stati raggiunti nel tempo. Attenzione però che in questa accezione, l'utilità è un numero soggettivo come la priorità che l'investitore esprime. All'interno dell'esempio, ci siederemo davanti ad un investitore il cui problema economico riguarda il raggiungimento di due tra i più comuni obiettivi: l'acquisto di un'autovettura tra 5 anni e il pagamento dell'iscrizione universitaria dei propri figli tra 10 anni. Arricchiamo l'esempio supponendo che per i due obiettivi esista una "seconda scelta", chiamata "obiettivo parziale", su cui l'algoritmo ripiegherà qualora si presentasse la necessità che, al tempo in cui esiste l'esborso, non si disponga della liquidità necessaria per ottenere quello ritenuto principale o che questi obiettivi non portino ad una soluzione ottimale in termine di utilità. Esprimiamo gli obiettivi[19] in questa tabella mettendo in ordine per tempo (crescente) per tempo e per costo(crescente):

Tabella 6: Costi ed utilità degli obiettivi dell'investitore

Tempo	*5*		*5*		*10*		*10*	
Costo	$c_{1,5}$	120.000,00 €	$c_{2,5}$	130.000,00 €	$c_{1,10}$	160.000,00 €	$c_{2,10}$	180.000,00 €
Utilità	$u_{1,5}$	1.000	$u_{2,5}$	3.000	$u_{1,10}$	5.000	$u_{2,10}$	10.000

Verranno richiamati nella trattazione gli elementi della tabella con la dicitura $c_{k,t}$ dove il pedice *t* si riferisce al tempo in cui insiste il costo e con *k* alla tipologia del costo stesso. Ad esempio, al tempo (t=5) si notano due costi per un unico obiettivo (acquisto di una autovettura di diverso valore):

- Per l'acquisto dell'autovettura si rappresentano due costi $c_{1,5} = 120.000$ € (quella meno performante), $c_{2,5} = 130.000$ € (quella più performate). Il pedice k=1 si riferisce al primo e unico obiettivo (acquisto autovettura) al tempo 5.
- Per il pagamento dell'iscrizione ad una prestigiosa Università americana invece i costi $c_{1,10} = 160.000$ € e $c_{2,10} = 180.000$ €.

Utilizzeremo la stessa nomenclatura relativamente ai due pedici (*k* obiettivo e *t* tempo) per esprimere l'utilità ($u_{k,t}$).
Per un istante temporale *t* definiremo con $K_{max}(t)$ il numero massimo di costi sostenibili nell'istante temporale di riferimento (ad esempio, per il tempo 5 si ottiene $K_{max}(5) = 2$).
Abbiamo altresì preso in considerazione il parametro inflazione (posizionata mediamente al 2%) che impatterà la spesa/costo prevista negli anni.

- *Portfolio Evolution and Portfolio Investment Strategies*

Definito un orizzonte temporale di investimento, verrà proposta un'allocazione ottimale posizionata nella frontiera al fine di poter massimizzare sia il valore atteso dell'utilità dell'investitore che altresì la probabilità di raggiungere gli obiettivi scelti in funzione del livello di ricchezza raggiunto nel tempo di osservazione.

[19] In questo caso si è scelto di creare appositamente un problema sfruttando due obiettivi tra i più comuni che si possono affrontare nella vita. Naturalmente l'algoritmo è versatile per *n* obiettivi in tempi distinti: questo presuppone logicamente avere a disposizione una macchina abbastanza performante per raggiungere la soluzione del problema in tempi rapidi. Questo problema che viene presentato ha ottenuto la sua soluzione in 3 secondi (processore Core i5).

In altre parole, ad ogni istante temporale, si guarderà al livello di ricchezza raggiunto e si sceglierà un'allocazione consona tra quelle proposte, che possa massimizzare la probabilità di raggiungere gli obiettivi. L'interazione con l'investitore parte proprio dalla prima fase, cioè quella di definizione degli obiettivi da raggiungere posizionati nell'arco temporale dell'investimento. Il dialogo sarà favorito nella misura in cui l'oggetto del discorso non saranno solo misure di rischio a sé stanti, ma verranno incorporate all'interno del processo di massimizzazione delle probabilità di raggiungimento di obiettivi economici reali. Non di rado si ravvisa la necessità di trasformare discorsi complicati (risk based) in misure di probabilità comprensibili a tutti. Ma andiamo con ordine.
In questa trattazione si semplifica e si restringe il subset delle asset classes disponibili per l'investitore non sottovalutando assolutamente tutto il processo di investimento e di proposizione del portafoglio (e di gestione): questa fase però è subordinata alla conoscenza del grado di "avversione" al rischio dell'investitore derivante dal questionario. Rispetto a quanto previsto dalla letteratura, partendo dalle risposte del questionario qualitativo proposto nei capitoli precedenti, si è proceduto alla mappatura dei parametri di avversione al rischio del cliente, di loss aversion e di risk tolerance derivanti dalla definizione della "Power Utility" ritenendola più idonea a descrivere la percezione al rischio degli investitori. Accostando i rendimenti storici della asset classes (R) alla funzione di utilità scelta ($\boldsymbol{U}_{(\boldsymbol{\gamma},\boldsymbol{\lambda},\boldsymbol{\varphi})}$) per ogni combinazione di $risk\ aversion\ \gamma, loss\ aversion\ \lambda, reflection\ \varphi$, si ottengono i vettori di allocazione **w*** secondo l'equazione:

$$\boldsymbol{w}^{*} = \boldsymbol{max}\ \boldsymbol{E}[\boldsymbol{U}_{(\boldsymbol{\gamma},\boldsymbol{\lambda},\boldsymbol{\varphi})}(\boldsymbol{R})] \qquad \text{[Eq.11]}$$

Dove **w*** risultano essere i pesi allocabili nelle asset classes rischiose (equity) rispetto a quelle meno rischiose (governativi, liquidità). Si riporta in tabella (Tabella 2: Questionario propensione al rischio del clientei risultati ottenuti.

Tabella 7: Portafoglio che ottimizza l'utilità dell'investitore dati i parametri($\boldsymbol{risk\ aversion\ \gamma, loss\ aversion\ \lambda, reflection\ \varphi}$)

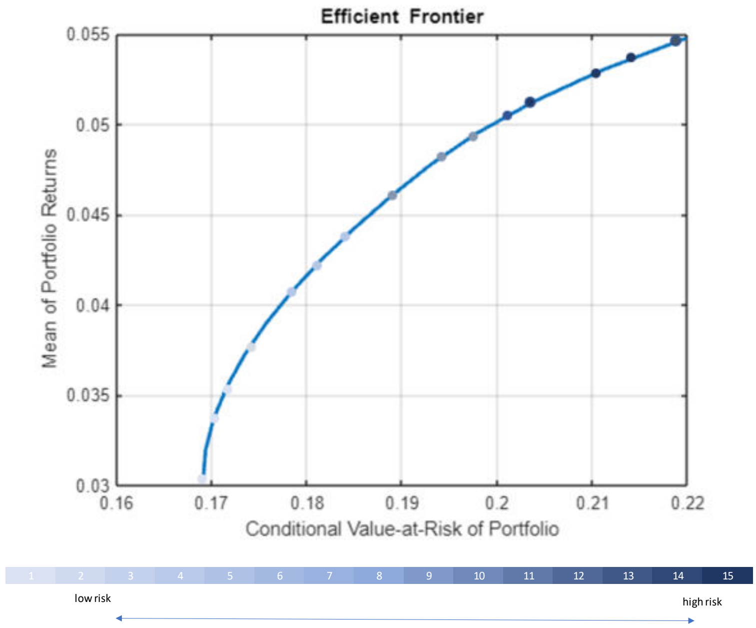
Reflection φ=0

Loss aversion mapping	Asset class	Risk aversion mapping			
		γ=3	**γ=6**	**γ=9**	**γ=12**
λ=1	**Equity World**	54,3%	42,7%	36,9%	33,4%
	Equity Emu	45,7%	42,3%	40,4%	39,5%
	Global Gov. (1-5y)	0,0%	15,0%	22,8%	27,1%
	Liq.	0,0%	0,0%	0,0%	0,0%
		γ=3	**γ=6**	**γ=9**	**γ=12**
λ=2	**Equity World**	47,4%	34,8%	35,3%	32,0%
	Equity Emu	43,1%	36,5%	39,9%	38,4%
	Global Gov. (1-5y)	9,3%	21,6%	24,1%	27,4%
	Liq.	0,2%	7,2%	0,7%	2,2%
		γ=3	**γ=6**	**γ=9**	**γ=12**
λ=3	**Equity World**	41,4%	33,5%	31,4%	31,8%
	Equity Emu	40,4%	36,6%	35,9%	38,3%
	Global Gov. (1-5y)	16,9%	22,1%	25,1%	27,6%
	Liq.	1,3%	7,8%	7,7%	2,3%

Questa tabella descrive la molteplicità degli investitori aventi un profilo di rischio che viene descritto dai parametri $risk\ aversion\ \gamma, loss\ aversion\ \lambda, reflection\ \varphi$. Si suppone quindi che, una volta svolto il questionario (Tabella 2: Questionario propensione al rischio del cliente l'investitore sia stato classificato con parametri ($\gamma = 6, \lambda = 1, \varphi = 0$): l'investitore in oggetto, idealmente, può supportare un buon grado di rischio sul portafoglio, avendo un orizzonte di investimento medio lungo termine (gli obiettivi sono a 10 anni) e con una buona capacità di gestire gli eventi negativi del mercato (loss aversion minima) avendo però attenzione al rischio complessivo dell'investimento (risk aversion media, pari a 6). La ponderazione di portafoglio che ne massimizza l'utilità, dato il suo profilo di rischio, è una combinazione tra Equity (circa 85%) e i titoli governativi di breve-medio termine 15%. Questo portafoglio "risk profile" descrive alla perfezione il massimo grado di rischio che il cliente può sostenere in base alla sua profilatura: a parità di input nessun altro portafoglio con rischiosità maggiore può essere associato a questo profilo di rischio.
Individuato il portafoglio "risk profile" dell'investitore, è necessario individuare l'insieme di portafogli ammissibili per il cliente (il suo basket di portafogli) che saranno scelti in base alla ricchezza raggiunta e alla massimizzazione dell'utilità (e della probabilità di ottenimento degli obiettivi). Viene per prima cosa associata una metrica che descriva quantitativamente il concetto di "rischio" per il portafoglio "risk profile" così da individuare un insieme di portafogli ammissibile per l'investitore.
L'obiettivo è quello di proporre dei portafogli con un dato livello di *downside risk* coerente con la rischiosità ottenuta in sede di compilazione di questionario. Questa nuova frontiera avrà come input i parametri di descrizione qualitativa del rischio (risk aversion, loss aversion e reflection) e come output una serie di portafogli il cui Conditional Value at Risk[20], CvaR, massimo (fatto sul 5% della coda sx) è quello indicato dall'ultimo portafoglio sulla frontiera disegnato in blu scuro utilizzando i parametri sopra elencati. Ogni portafoglio sulla frontiera rappresenta il portafoglio che massimizza il rendimento minimizzando la rischiosità che il cliente dovrà gestire qualora allocasse su un portafoglio all'interno della frontiera[21]. Su questa frontiera scegliamo allora i 15 portafogli utili per il proseguo dell'esercizio, dal più rischioso (blu scuro) al meno rischioso (blu chiaro).

[20] La coda della distribuzione, in ottica di CVaR, è stata modellizzata secondo una Frechet utilizzando l'evidenza empirica dei periodi di stress posizionando il parametro shape equity $\xi = 0.3$.
[21] Per l'ottenimento della frontiera efficiente abbiamo utilizzato il software MatLab "Optimization Toolbox version 9.1 (R2021a)".

Figura 6: Frontiera Efficiente Conditional VaR



- *La ricchezza*

Elemento di estrema importanza per la soluzione al problema GBWM con obiettivi multipli è la definizione della ricchezza in ogni istante futuro dell'orizzonte di investimento. In questa trattazione possiamo interpretare la "ricchezza" come il valore attuale del portafoglio investito nel periodo di tempo considerato. Poiché ovviamente la ricchezza non è "statica" definiamo una scala di valori che contenga i_{max} valori di ricchezza dove i_{max} può essere qualsiasi numero desiderato (anche se si suggeriscono valori non troppo bassi perché possono causare imprecisioni e né valori troppo grandi o inutilmente lenti per lo svolgimento del problema). I valori della ricchezza i_{max} saranno definiti i limiti di ricchezza W_{min} e W_{max}. Iniziamo approssimando W_{min}e W_{max} annotandoli come valori di ricchezza minimo/massimo ragionevolmente raggiungibili per un investitore. Poiché abbiamo scelto di utilizzare il moto browniano geometrico per l'evoluzione del portafoglio modello in questo articolo, definiamo la dinamica della ricchezza nel tempo partendo dall' iniziale W(0) che evolverà come:

$$\boldsymbol{W}(\mathbf{t}) = \boldsymbol{W}(\mathbf{0}) * \boldsymbol{exp}\left[\left(\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\sigma}^2\right)\boldsymbol{t} + \boldsymbol{\sigma}\sqrt{\boldsymbol{t}}\boldsymbol{Z}\right] \qquad \text{[Eq.13]}$$

dove Z, in questo caso, è una variabile casuale normale standard e $\boldsymbol{\mu}$ è il rendimento atteso della ricchezza.

Per definire i limiti W_{min} e W_{max} occorre considerare anche i costi $c_k(t)$ e le entrate $I(t)$ previsti per ogni istante futuro (t) al per il raggiungimento degli obiettivi[22]. Otteniamo allora:

$$\overline{W_{min}} = min_{t\epsilon(0,1,2,....,T)}\left[W(0)e^{\left(\mu_{min}-\frac{\sigma_{max}^2}{2}\right)t-3\sigma_{max}\sqrt{t}} + \sum_{\tau=0}^{t}\left(I(\tau) - c_{k_{max}(\tau)}(\tau)\right)e^{\left(\mu_{min}-\frac{\sigma_{max}^2}{2}\right)t-3\sigma_{max}\sqrt{t}}\right] \qquad \text{[Eq. 14]}$$

$$\overline{W_{max}} = max_{t\epsilon(0,1,2,....,T)}\left[W(0)e^{\left(\mu_{min}-\frac{\sigma_{max}^2}{2}\right)t-3\sigma_{max}\sqrt{t}} + \sum_{\tau=0}^{t}\left(I(\tau) - c_{k_{max}(\tau)}(\tau)\right)e^{\left(\mu_{min}-\frac{\sigma_{max}^2}{2}\right)t-3\sigma_{max}\sqrt{t}}\right] \qquad \text{[Eq. 15]}$$

Dove i valori μ_{min} e σ^2_{max} sono rappresentativi dei valori rendimento/rischio ottenuti dai portafogli individuati all'interno della frontiera di cui al precedente paragrafo.
Possiamo a questo punto stimare la scala della ricchezza equi distanziando, su base logaritmica, i livelli ottenuti. Ottenendo alla fine un insieme di valori possibili:
Definiamo in ultimo la scala della ricchezza equi distanziando su base logaritmica i livelli ottenuti. Ottenendo alla fine una scala di valori possibili: $W_{min} < W_0 < W_{max}$
A questo punto, per risolvere il problema dinamico, abbiamo bisogno di costruire le probabilità condizionate corrispondenti ad ogni livello di ricchezza: per ogni istante temporale la probabilità condizionata di trovarsi nello stato di ricchezza $W_j(t+1)$ provenendo da $W_i(t)$ assumendo, se esistono, i costi $c_k(t)$ e le entrate $I(t)$ attribuibili al tempo t per il raggiungimento degli obiettivi sarà data da:

$$\overline{\boldsymbol{q}}\left(\boldsymbol{W_j}(\boldsymbol{t}+\mathbf{1})\middle|\boldsymbol{W_i}(\boldsymbol{t}), \boldsymbol{c_k}(\boldsymbol{t}), \boldsymbol{\mu_l}\right) = \boldsymbol{\phi}\left(\frac{\mathbf{1}}{\boldsymbol{\sigma_l}\sqrt{\boldsymbol{h}}}\left(\boldsymbol{ln}\left(\frac{\boldsymbol{W_j}(\boldsymbol{t}+\mathbf{1})}{\boldsymbol{W_i}(\boldsymbol{t})+\boldsymbol{I}(\boldsymbol{t})-\boldsymbol{c_k}(\boldsymbol{t})}\right) - \left(\boldsymbol{\mu_l} - \frac{\boldsymbol{\sigma_l^2}}{\mathbf{2}}\right)\boldsymbol{h}\right)\right) \qquad \text{[Eq. 16]}$$

La probabilità condizionata è stimata assumendo $\phi(z)$ il valore della funzione di densità di probabilità della variabile normale standard z. Sarà una funzione delle variabili sopra citate e del portafoglio di frontiera scelto da cui si deriveranno il rendimento atteso e la

[22] All'interno dell'esempio riportato non risulta alcun inflow. Verrà solo ipotizzato un inflow periodico alla fine della trattazione.

deviazione standard. Si otterrà una matrice $i_{max} x\ i_{max}$ in cui ogni riga riporterà le probabilità condizionate $\left(W_j(t+1)\middle|W_i(t), c_k(t), \mu_l\right)$ che, comunque, occorrerà normalizzarle per poter ottenerne la somma 1.

$$\bar{\boldsymbol{q}}\left(\boldsymbol{W_j(t+1)}\middle|\boldsymbol{W_i(t)}, \boldsymbol{c_k(t)}, \boldsymbol{\mu_l}\right) = \frac{\bar{q}\left(\boldsymbol{W_j(t+1)}\middle|\boldsymbol{W_i(t)}, \boldsymbol{c_k(t)}, \boldsymbol{\mu_l}\right)}{\sum_{J=1}^{I_{max}} \bar{q}\left(\boldsymbol{W_J(t+1)}\middle|\boldsymbol{W_\iota(t)}, \boldsymbol{c_k(t)}, \boldsymbol{\mu_l}\right)} \qquad \text{[Eq. 17]}$$

Introduciamo, in ultimo che all'interno della trattazione simuleremo che l'investitore abbia un investimento iniziale pari a $W_0 = 100.000\ Euro$.

4.3.2 Il problema dinamico

Possiamo tornare al problema di massimizzazione della probabilità di raggiungimento degli obiettivi. Dati i livelli di ricchezza simulati, l'orizzonte di investimento, i costi e le utilità attribuite agli obiettivi e i portafogli selezionati sulla frontiera, introduciamo la funzione "valore" che è funzione della ricchezza (e quindi dell'utilità) dell'investitore. Attraverso un processo di stima "backward" è possibile sfruttare l'equazione proposta da Bellman al fine di definire quali obiettivi è possibile raggiungere, quali eventualmente sostituire e quali risultano impossibili raggiungere. La funzione valore V, al tempo t + 1 è la somma attesa più alta possibile delle utilità da obiettivi raggiunti nei periodi $t+1, t+2..T-1$ più l'utilità dell'eccesso di ricchezza che otteniamo in t = T. Il termine "più alto possibile" significa che in ogni momento e punto della griglia della ricchezza, verrà stimata la funzione valore prendendo in considerazione le componenti di costo/utilità $\boldsymbol{c_{k,t}}\ e\ \boldsymbol{u_{k,t}}$ e tutti i possibili portafogli di investimento previsti l[23]. Il processo si definisce "backward" perché se l'obiettivo è determinare ogni valore della funzione valore V per ogni istante temporale e per ogni livello di griglia di ricchezza, partiamo dal periodo finale T, dove sappiamo con certezza che

$$\boldsymbol{V(W_i(T))} = \boldsymbol{U(W_i(T))} \qquad \text{[Eq. 18]}$$

Nel nostro esempio, sappiamo che al tempo T,

$$\boldsymbol{V\left(W_i(T)\right)} = \begin{cases} \boldsymbol{U\left(W_i(T)\right) = 10.000\ se\ \ W_i(T) \geq 180.000} \\ \boldsymbol{U\left(W_i(T)\right) = 5.000\ \ se\ \ 160.000 \leq W_i(T) < 180.000} \end{cases} \qquad \text{[Eq. 19]}$$

Se non otteniamo una ricchezza al tempo T maggiore di Euro 160.000, posizioniamo l'utilità $U\left(W_i(T)\right) = 0$
Poiché conosciamo le probabilità di transizione stimiamo la seguente equazione di Bellman per il problema dinamico e risolviamola per ciascun istante temporale, livello di costo e portafoglio scelto.

$$\boldsymbol{V(W_i(t))} = \boldsymbol{\max_{k,l}}\left[\boldsymbol{u_k(t)} + \textstyle\sum_{J=1}^{imax} \boldsymbol{V(W_j(t+1))} * \ \boldsymbol{q(W_j(t+1)|W_i(t), c_k(t)}, \mu_l)\right] \qquad \text{[Eq. 20]}$$

Il processo di ottimizzazione dinamica dei due obiettivi produrrà una matrice dove nell'asse delle ordinate ci saranno i livelli di ricchezza (in migliaia di Euro) derivanti tempo per tempo dal processo stocastico e nell'asse delle ascisse ci sarà il tempo. Il colore di ogni quadratino indicherà ad ogni istante temporale il portafoglio scelto all'interno della frontiera che massimizza l'utilità attesa dell'investitore; la probabilità di realizzare gli obiettivi è anch'essa massimizzata e deriva dalla funzione valore: ognuno di questi quadratini si porta dietro quindi le informazioni del portafoglio consigliato utile per colpire gli obiettivi futuri dato il livello di ricchezza raggiunto. Più il livello di ricchezza nell'istante futuro posseduto dal cliente sarà alto più il portafoglio ribilanciato secondo le caratteristiche del cliente tenderà ad essere meno rischioso poiché già disporrà della ricchezza necessaria a centrare gli obiettivi. Nel grafico i portafogli più rischiosi sono quelli con un colore blu scuro.
Il grafico sopra aiuterà il Consulente nel dialogo con l'investitore perché dimostrerà come possa evolvere la rischiosità del suo investimento in base al livello di ricchezza raggiunto al tempo e gli obiettivi segnalati.
In ogni istante temporale è possibile definire la probabilità di raggiungere tutti gli obiettivi (inclusi quelli parziali) e qualsiasi tipo di combinazione degli stessi dato il livello di ricchezza e il portafoglio di investimento che ne massimizzi questa probabilità.
Per determinare la distribuzione di probabilità della ricchezza dell'investitore nei tempi futuri, vengono determinate le probabilità di transizione e le informazioni sulla strategia ottimale, $k_{i,t}$ e $l_{i,t}$ dall'equazione di Bellman per far evolvere la distribuzione di probabilità in avanti nel tempo, a partire da t = 0, quindi t = 1, e terminando con t = T -1.
Più nel dettaglio, al t=0 definiamo i_0 così da ottenere W_{i_0} risultante la ricchezza iniziale W(0) e, quindi, $p(W_{i_0}(0)) = 1$. Settiamo le restanti probabilità al tempo 0 come $p(W_i(0)) = 0$ per ogni $i \neq i_0$.

$$\boldsymbol{p(W_j(t+1))} = \textstyle\sum_{J=1}^{imax} \boldsymbol{q(W_j(t+1)|W_i(t), c_{k_{i,t}}(t)}, \mu_{l_{i,t}}) * \boldsymbol{p(W_i(t))} \qquad \text{[Eq. 21]}$$

Impostiamo quindi t = 1 nell'equazione sopra e la eseguiamo nuovamente per ogni $j = 1,2...i_{max}$ continuando in questo modo fino ad arrivare al caso t = T -1. Questo restituisce la distribuzione di probabilità per ogni punto della griglia della ricchezza in ogni periodo di tempo del portafoglio.
Sviluppando tutta l'ottimizzazione dinamica è possibile all'istante zero e con un livello di ricchezza $W(\boldsymbol{0})$ indicato dal cliente individuare che il raggiungimento dell'obiettivo in maniera probabilistica, ad ogni tempo.
Considerata la rischiosità iniziale, la ricchezza investita ($W(\boldsymbol{0}) = \boldsymbol{100.000\ Euro}$), il cliente selezionato ha dunque le probabilità[24] di raggiungere i 4 obiettivi due e l'obiettivo uno indicate dalla seguente tabella.

---

[23] Con la dimensione di massimizzazione "l" intendiamo la coppia di (media,varianza) di uno dei portafogli ottenuti nella frontiera efficiente CVaR funzione dei parametri risk aversion, reflection e loss aversion del cliente trattati nei paragrafi precedenti. Facciamo attenzione che la matrice $V(W_i(t))$ che otterremo sarà multidimensionale ($WxWxCx$μ) da cui vorremmo andare a scegliere il valore maggiore di $V(W_i(t))$ per ogni tempo dati i costi e la coppia di media varianza del portafoglio.

[24] Una volta calcolata questa distribuzione di probabilità della ricchezza, possiamo determinare la probabilità di raggiungere qualsiasi obiettivo specifico, completo o parziale, in un dato momento t: In ogni momento t per ogni $k = 1,2...k_{max}(t)$, sommiamo $p(W_i(t))$ su ogni i dove $W_{i,t} = k$. Questo dà la probabilità che venga scelta ciascuna componente k nei vettori costo/utilità. Una volta che questo è noto, i vettori costo/utilità vengono

Figura 7: Andamento della ricchezza investita

Wealth/ Time	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10
74,88	15	15	15	15	15	15	15	15	15	15
79,34	15	15	15	15	15	15	15	15	15	15
84,07	15	14	11	15	15	15	15	15	15	15
89,07	15	11	9	11	15	15	15	15	15	15
94,38	15	9	7	9	15	15	15	15	15	15
100,00	15	8	6	7	15	15	15	15	15	15
105,96	15	7	5	6	11	15	15	15	15	15
112,27	15	6	4	5	9	11	15	15	15	15
118,95	15	5	3	4	7	9	15	15	15	15
126,04	15	4	2	3	6	7	15	15	15	15
133,55	15	8	6	2	5	6	11	15	15	15
141,50	15	10	7	6	4	5	9	11	11	11
149,93	15	12	9	7	3	4	7	9	9	9
158,86	15	10	11	9	2	3	6	7	7	7
168,32	15	10	10	11	6	2	5	6	6	6
178,34	12	8	10	10	7	6	4	5	5	5
188,97	10	8	8	10	9	7	3	4	5	4
200,22	10	8	8	8	11	9	2	3	5	3
212,15	10	8	8	8	13	11	6	2	6	2
224,78	8	6	8	8	15	13	7	6	8	6
238,17	8	6	6	8	15	15	9	7	8	10
252,36	8	6	6	6	14	15	11	9	8	12
267,39	8	5	6	6	13	14	13	11	8	15
283,31	8	4	5	6	13	13	15	13	13	15
300,19	8	3	4	6	12	13	15	15	13	15
318,07	8	3	3	5	11	12	14	15	12	12
337,01	6	3	3	4	11	11	13	14	8	10
357,09	6	3	3	3	10	11	12	13	6	8
378,35	6	3	3	3	10	10	10	8	5	8
400,89	6	3	3	3	10	10	10	6	5	6
424,77	5	3	3	3	10	10	8	5	5	5
450,07	4	3	3	3	10	10	6	5	4	5
476,87	3	2	3	3	9	10	5	5	4	5
505,28	2	2	2	3	8	9	5	4	3	4
535,37	2	2	2	3	7	8	5	4	3	4
567,26	2	2	2	2	6	7	4	3	3	4
601,04	2	2	2	2	6	6	4	3	3	4
636,84	1	2	2	2	5	6	3	3	2	3
674,77	1	1	2	2	5	5	3	3	2	3
714,96	1	1	1	2	4	5	3	2	2	3
757,55	1	1	1	2	4	4	3	2	2	3
802,67	1	1	1	1	4	4	2	2	1	3
850,48	1	1	1	1	4	4	2	2	1	2
901,13	1	1	1	1	3	4	2	1	1	1
954,80	1	1	1	1	3	3	2	1	1	1
1011,67	1	1	1	1	3	2	1	1	1	1
1071,93	1	1	1	1	2	2	1	1	1	1
1067,65	1	1	1	1	2	2	1	1	1	1

Tabella 8: Probabilità associate ad obiettivi

$u_{1,5}$	$u_{2,5}$	$u_{1,10}$	$u_{2,10}$	P{fullfill t=5 goal,c 120.000}	P{fullfill t=5 goal,c 130.000}}	P{fullfill t=10 goal,c 160.000}}	P{fullfill t=10 goal,c 180.000}}
1.000	5.000	3.000	10.000	8,1%	15,1%	11,62%	34,56%

Possiamo dedurre che dato il rischio massimo sopportabile dall'investitore (da cui derivano gli asset mix di portafogli) e data la ricchezza iniziale unita alle priorità (utilità) attribuite ex-ante, risulterà un'alta probabilità di ottenere l'università migliore, con conseguente riduzione delle probabilità di acquisto della macchina in entrambi gli obiettivi. Potremmo ottenere il valore di utilità totale attesa ottimale $E[u] = V(W_{i0}(0)) = 4642$ dell'investitore. Tale risultato è uguale al prodotto scalare del vettore delle utilità con il vettore della probabilità. Qualora l'investitore non fosse soddisfatto del grado di raggiungimento degli obiettivi in ottica probabilistica, è possibile incrementare la probabilità di raggiungimento dell'obiettivo desiderato in 3 differenti modi non per forza alternativi:

a) Concordare una utilità maggiore all'obiettivo desiderato
b) Aumentare il capitale iniziale W(0) = 100.000 Euro
c) Introdurre del denaro (inflows) periodico all'interno degli investimenti

In breve, riposizionando le utilità come indicato nella tabella seguente, troviamo aumentate le probabilità (a parità di capitale investito iniziale) di ottenere obiettivi differenti, ma purtroppo, data la rischiosità che il cliente può sostenere, non possiamo ottenere risultati migliori se non con l'aumento del capitale iniziale.

Tabella 9: Sensitivity su probabilità di accadimento

$u_{1,5}$	$u_{2,5}$	$u_{1,10}$	$u_{2,10}$	P{fullfill t=5 goal,c 120.000}	P{fullfill t=5 goal,c 130.000}}	P{fullfill t=10 goal,c 160.000}}	P{fullfill t=10 goal,c 180.000}}
1.000	5.000	3.000	10.000	8,1%	15,1%	11,62%	34,56%
1.000	2.000	100	20.000	6,2%	3,15%	7,10%	38,90%

Nella tabella sopra si è condotto un ulteriore esempio sulla formulazione dei 2 obiettivi (4 se consideriamo i 2 parziali) ipotizzando un'allocazione dell'utilità che il cliente, ex-ante, attribuisce agli obiettivi. Qualora il cliente attribuisca un po' più di priorità (in termini di utilità) all'obiettivo di una Università più prestigiosa sovrappesandola rispetto a quella "meno prestigiosa", otteniamo un innalzamento della probabilità di raggiungimento di questo obiettivo con una riduzione sostanziale delle probabilità di realizzo rispetto all'acquisto di una macchina performante. E' logico che, qualora l'investitore non fosse ancora soddisfatto della probabilità ottenuta, occorrerà allora rivedere il mix di utilità attribuito o di aumentare gli investimenti iniziali anche in ottica di conoscenza della sua ricchezza patrimoniale.

---

ricollegati ai loro obiettivi originari e alla probabilità che ogni obiettivo totale o parziale sarà raggiunto è determinato sommando i componenti connesse a quell'obiettivo totale o parziale.

4.3.3. Problema di ottimizzazione statica: l'Efficient Global Probability Frontier (EGPF)

Proviamo adesso a semplificare l'esercizio precedente introducendo un problema "statico" basato sulla definizione di obiettivi ed investimenti coerenti con il rischio del cliente. Dall'esercizio precedente abbiamo appreso come il cliente può lavorare sulla "utilità" di ogni obiettivo, influenzando la probabilità di raggiungimento degli stessi. Qualora ad esempio il cliente avesse due soli obiettivi e nessuna scelta "parziale" (siamo nel caso di due obiettivi "All or nothing"), potremmo trovarci a gestire una situazione come quella della tabella in basso.

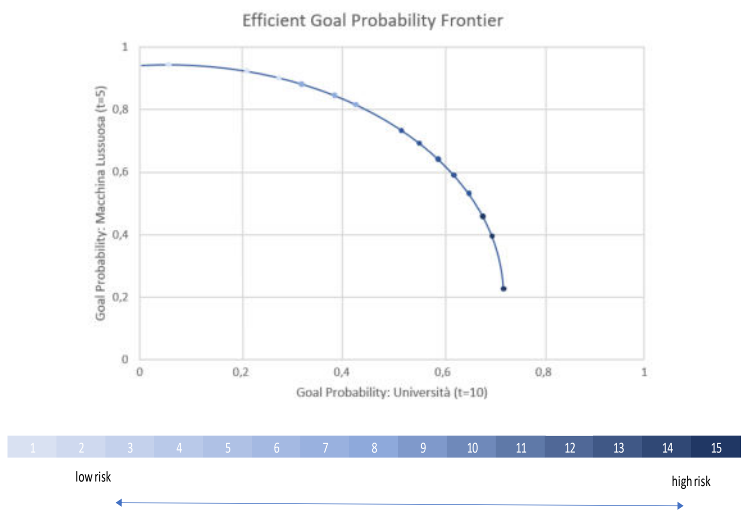
Tabella 10: Obiettivi su ottimizzazione statica

Obiettivi

Tempo	*t=5*	*t=10*
Costo	120.000,00 €	180.000,00 €
Utilità	1.000	10.000

In questo caso abbiamo deciso di semplificare ponendo un obiettivo a t=5 (cambiare l'autovettura) e a t=10 (pagamento delle tasse Universitarie). Non ci sono "seconde scelte". Poniamo anche come p_1 la probabilità di raggiungimento dell'obiettivo al t=5 (acquisto nuova autovettura) e p_2 probabilità di raggiungimento dell'obiettivo al t=10 (pagamento retta Universitaria). Troviamo abbastanza difficile poter parlare al cliente in termini di utilità per ogni obiettivo, ma possiamo però semplificare questo passaggio sviluppando una frontiera che possa riportare le probabilità di raggiungimento degli obiettivi e il portafoglio che realisticamente aiuta a raggiungerli. Ipotizzando di essere al tempo iniziale (t=0) e di non poter cambiare l'allocazione per i restanti anni, possiamo reiterando l'algoritmo visto nella sezione precedente ottenere, per ogni portafoglio tra i 15 proposti e per più funzioni di utilità, una "frontiera Pareto efficiente" che riporterà le coppie di probabilità (p_1 e p_2) ottenibili con l'investimento iniziale e con livelli di utilità differenti. Per ogni coppia è possibile ottenere il portafoglio di investimento che, utilizzato e mantenuto nel tempo, possa raggiungere quella probabilità di ottenimento degli obiettivi. Vige sempre la convenzione che il portafoglio più scuro è quello più rischioso secondo la frontiera individuata nella figura in cui viene presentata la l'EGPF, Efficient Goal Probability Frontier.

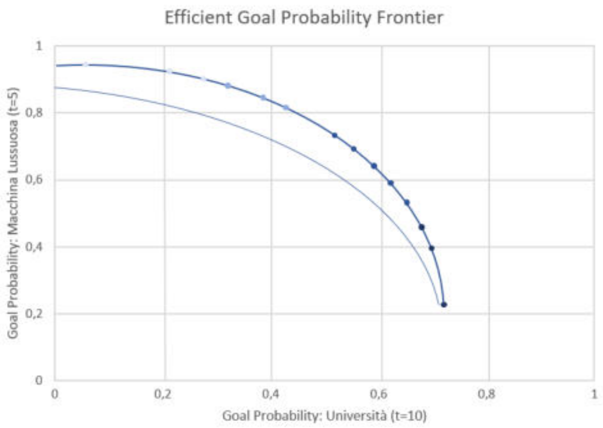
Figura 8: Efficient Goal Probability Frontier



Appuntiamo le seguenti proprietà della probabilità di goal della frontiera efficiente (EGPF):

1. Per due obiettivi, l'EGPF è una curva. Per n gol, il EGPF è un'iper-superficie (n - 1) dimensionale.
2. Ogni punto sull'EGPF è mappato su un insieme di pesi di utilità che riflettono la relativa importanza di ciascun obiettivo per l'investitore.
3. Dietro ogni punto dell'EGPF ci sono due strategie ottimizzate:a) una strategia di investimento ottimizzata e b) una strategia ottimizzata per raggiungere gli obiettivi. Ciò che è fondamentale qui è che l'EGPF sia dichiarato rigorosamente in termini di probabilità, non in termini di utilità, così l'investitore è in grado di specificare le proprie preferenze. Dalla parte degli investitori è molto più facile comprendere l'idea di poter scegliere l'allocazione partendo dalla probabilità di realizzazione degli obiettivi. Se l'investitore può dedicare più denaro all'investimento iniziale, possiamo trovare l'investimento iniziale minimo necessario per ottenere il punto di probabilità dell'obiettivo desiderato. Questo consente all'investitore di non spendere né in eccesso né in difetto per ottenere le probabilità di goal desiderate. In alternativa, l'investitore potrebbe non essere in grado di accedere inizialmente ad ulteriore denaro, ma potrebbe invece essere in grado di contribuire nel tempo. In questo caso, il metodo per determinare l'investimento iniziale minimo aggiuntivo può essere facilmente modificato per determinare invece gli ingressi minimi monetari nel tempo al fine di ottenere la desiderata probabilità sull'obiettivo.
Ipotizzando infatti un inflow periodico del 2% sul capitale iniziale (può essere equivalente agli incassi dei dividendi sugli investimenti e, in qualche modo, abbattendo l'inflazione stimata in esercizio), potremmo ottenere una nuova EGPF dove infatti si verifica l'aumento delle probabilità di ottenimento dei due obiettivi. Aumentano allora le probabilità di ottenere l'obiettivo 1 (la macchina lussuosa) e le probabilità legate al pagamento della retta diminuendo la rischiosità del portafoglio iniziale stimata per le coppie di probabilità ricercate.

Figura 9: Inflows su EGPF



5. Conclusioni

Il legame tra questionario Mifid e RAF, contrariamente a quanto si possa pensare, è particolarmente importante per l'intermediario finanziario, specializzato nel servizio di wealth management per clientela private banking, HNWI e UHNWI. Questo risulta sia in un'ottica di gestione del rischio del balance sheet risk che nell'ottica di una value proposition di livello del servizio stesso di wealth management, in cui il cliente benefici di un processo di gestione del rischio integrato nel wealth planning e nelle scelte di asset management che mostri in maniera trasparente il raggiungimento degli obiettivi finanziari definiti nel tempo e relative probabilità di possibile successo. Soprattutto è particolarmente importante per i clienti del segmento private banking ed HNWI – estendibile a quella UHNWI, dove la frase "I don't like risk, but i like win" trova un forte fondamento al fine di esprimere la funzione di utilità attesa e nel sostenere il concetto che il risk-taking è un lusso così come identificato dal livello della misura dell'SLR complementare a quella della discretionary wealth ratio in cui peso importante è espresso dal valore attuale dello human capital e del valore attuale degli obiettivi finanziari qualitativamente e quantitativamente identificati. È necessario quindi richiedere una maggiore quantità e qualità dei dati del cliente rispetto a quelli richiesti dal questionario mifid e loro costante monitoraggio nel tempo che permettono di poter modificare quanto fin a quel momento fissato in termini di obiettivi finanziari. Output parziale è costituito da un financial statement del cliente che possa definire non solo la misura di SLR ma che permetta, in ogni momento, di poter accertare profili di stabilità, solvibilità e liquidità del cliente. Punto iniziale è la determinazione delle tre dimensioni dell'utilità attesa, risk aversion, loss aversion e reflection, attraverso opportuni schemi d'indagine (lottery-style technique) al fine di determinare i pesi del portafoglio d'investimento ove gli obiettivi finanziari sono legati ad un valore d'utilità e quindi siano ordinabili per livello di priorità. In una logica reverse engineering è possibile chiedere al cliente le probabilità di successo dell'obiettivo finanziario. A questo punto, a seconda del valore dell'SLR, del profilo di rischio determinato dalle tre dimensioni dell'utilità attesa e dagli obiettivi sarà possibile poter consigliare al cliente non solo il portafoglio d'investimento coerente che esprima le probabilità di successo lato obiettivi finanziari ma anche le possibili azioni in termini di financial statement, incluse attività di ottimizzazione fiscale e non solo laddove possibili. Tutto ciò assieme all'implementazione delle best practices in termini di metodologie e di misurazione del rischio (VaR & Conditional VaR), incluse quelle relative al correlation risk. Quanto sinora affermato è coerente con l'idea di determinare un risk appetite framework "personale" che in termini di risk appetite statement sia time varying a seconda dello svilupparsi della propria vita finanziaria. Ciò, in parte gestito da relationship manager che dovranno necessariamente potenziare le loro capacità tecniche e relazionali al fine di poter ottenere quel know-how necessario.

Bibliografia

R. Bellman "On the Theory of Dynamic Programming" Proceedings of the National Academy of Sciences, 1952

D. M. Berns, PhD "Modern Asset Allocation for Wealth Management" Wiley, 2020

J. Brunel "Goals-Based Wealth Management: An Integrated and Practical Approach to Changing the Structure of Wealth Advisory Practices", Wiley, 2015

A.Consiglio, F. Cocco, S. Zenios "Scenario optimization asset and liability modelling for individual investors. Annals of Operations Research 152, Springer, 2007

Ernst & Young "Risk Appetite Framework - Basilea 3: risks and supervision 2014 " – Materiali Convegno, 2014

ECB Banking Supervision "SSM Supervision on governance and risk appetite", 2016

S.R. Das, D. Ostrov, A. Radhakrishnan, D. Srivastav "Dynamic Optimization for Multi-Goals Wealth Management" Journal of Banking and Finance, 2021

S.R. Das, D. Ostrov, A. Radhakrishnan, D. Srivastav "Dynamic Portfolio Allocation in Goals-Based Wealth Management" Computational Management Science, 2020

S.R. Das, D. Ostrov, A. Radhakrishnan e D. Srivastav "Efficient Goal Probabilities: A New Frontier", Journal of Investment Management, 2023

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk, 1979

S.R. Das, H. Markowitz, J. Scheid, and M. Statman "Portfolio Optimization with Mental Accounts", Journal of Financial and Quantitative Analysis, 2010

S.R. Das, D. Ostrov, A. Radhakrishnan, and D. Srivastav "Goals-Based Wealth Management: A New Approach", Journal of Investment Management, 2018

K. Dowd, "Measuring Market Risk" 2nd Ed. , Wiley, 2005

P.S. Hagan, D. Kumar, A.S. Lesniewski, D.E. Woodward "Managing Smile Risk" Wilmott Magazine, 2002

P. Jorion "Value at Risk: the new benchmark for managing financial risk" 3rd Ed, 2007

Markowitz, H. H. "Portfolio Selection", Journal of Finance, 1952

G. Meissner "Correlation Risk Modeling and Management" Wiley, 2014

R.T. Rockafellar, S. Uryasev "Optimization of Conditional Value-at-Risk" Journal of Risk, Spring, 2000

Risk Management Magazine
Volume 18, Issue 3, September – December 2023

Direttore Responsabile (Editor in Chief)
Maurizio Vallino

Condirettore (Managing Editor)
Corrado Meglio

Editorial Board
Giampaolo Gabbi - Chief Editor Business Economics Area (SDA Bocconi); Paolo Giudici - Chief Editor Statistical Economics Area (Università di Pavia); Daniel Ahelegbey (Università di Pavia); Raffaella Calabrese (University of Edimburgh); Robert Eccles (Oxford University); Franco Fiordelisi (University of Essex); Pier Giuseppe Giribone (Università di Genova); Gulia Iori (London City University); Richard M. Levich (New York University); Michèle F. Sutter Rüdisser (University of San Gallen); Peter Schwendner (ZHAW Zurich University of Applied Sciences); Alessandra Tanda (Università di Pavia).

Scientific Committee
Arianna Agosto (Università di Pavia, Italy); Ruggero Bertelli (Università di Siena, Italy); Paola Bongini (Università Milano Bicocca, Italy); Anna Bottasso (Università di Genova, Italy); Marina Brogi (Università La Sapienza di Roma, Italy); Ottavio Caligaris (Università di Genova, Italy); Rosita Cocozza (Università Federico II di Napoli, Italy); Costanza Consolandi (Università di Siena, Italy); Simona Cosma (Università di Bologna, Italy); Paola Ferretti (Università di Venezia, Italy); Andrea Giacomelli (Università di Venezia, Italy); Adele Grassi (Vice Presidente APB, Italy); Mariantonietta Intonti (Università di Bari "Aldo Moro", Italy); Valentina Lagasio (Università La Sapienza di Roma, Italy); Duccio Martelli (Università di Perugia, Italy); Enrico Moretto (Università Milano Bicocca, Italy), Laura Nieri (Università di Genova, Italy); Adamaria Perrotta (UCD – University College Dublin, Ireland), Pasqualina Porretta (Università La Sapienza di Roma, Italy); Anna Grazia Quaranta (Università di Macerata, Italy); Enzo Scannella (Università di Palermo, Italy); Cristiana Schena (Università dell'Insubria, Italy); Giuseppe Torluccio (Università di Bologna, Italy).

Cartoonist: Silvano Gaggero

Ownership, Newsroom and Secretariat:
Associazione Italiana Financial Industry Risk Managers (AIFIRM), Via Sile 18, 20139 Milano

Registration number at Court of Milan n. 629 dated 10th September 2004

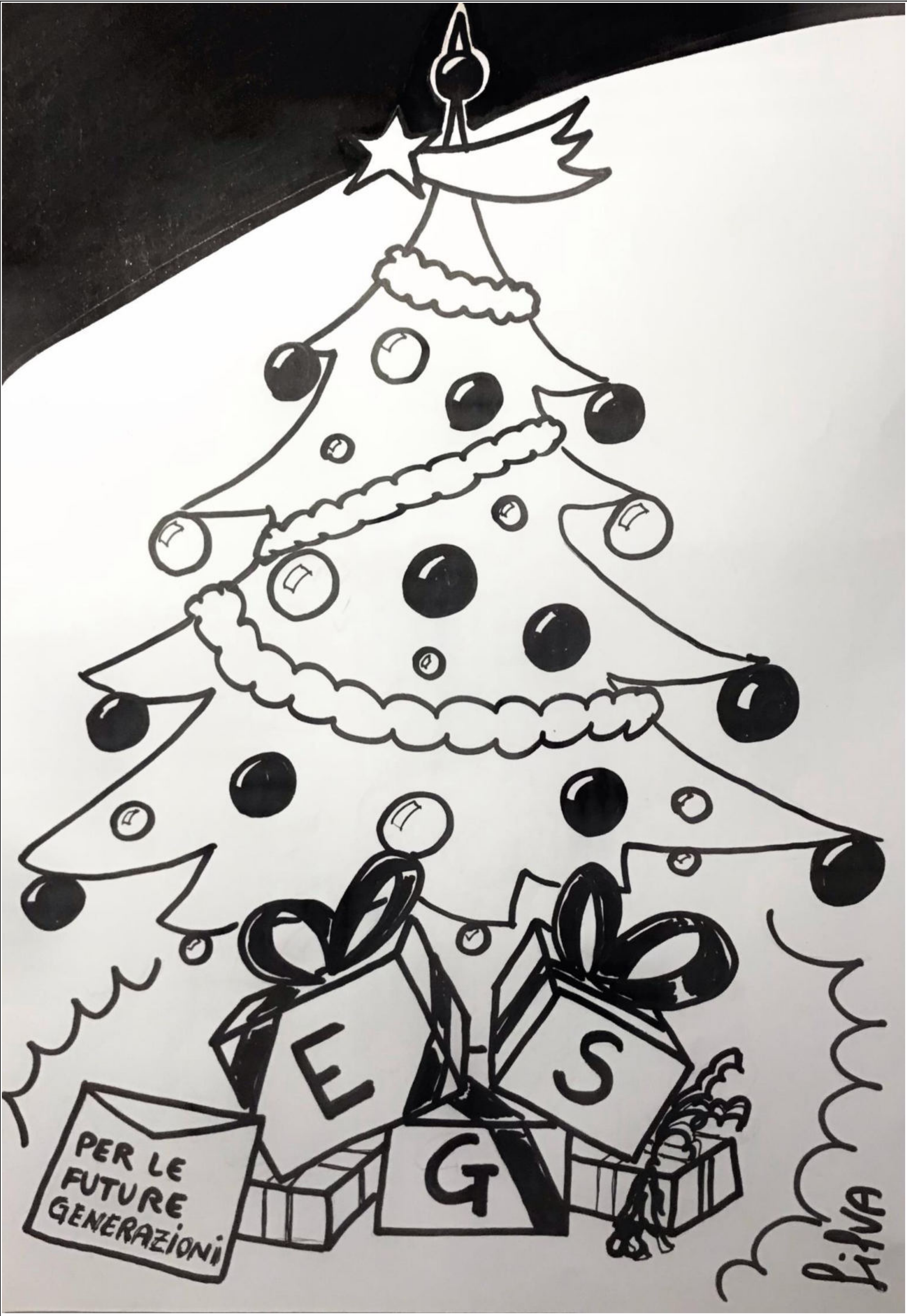
ISSN Print 2612-3665 – **ISSN Online** 2724-2153
DOI 10.47473/2016rrm

E-mail: risk.management.magazine@aifirm.it; Tel. +39 329 138 0475

Printing
Algraphy S.n.c. - Passo Ponte Carrega 62-62r 16141 Genova

Authors bear sole responsibility for the opinions expressed in their articles.

MAILED TO AIFIRM SUBSCRIBERS WHO ARE RESIDENT IN ITALY AND DULY REGISTERED


E
S
G
PER LE
FUTURE
GENERAZIONI

Wolters Kluwer

When you have to be right

Naviga l'attuale contesto normativo e prendi decisioni chiare e rapide.

Soddisfare i requisiti normativi, gestire le operazioni finanziarie, affrontare ogni singolo rischio ed elaborare una strategia integrata per rischio e finanza, sono tutti aspetti su cui lavoriamo insieme a piu' di 800 banche.

Le nostre soluzioni consentono una gestione integrata del rischio, dalla contabilita' alle segnalazioni di vigilanza, agevolando gli adeguamenti normativi e consentendo la condivisione dei dati certificati in maniera trasversale. Ciò consente ai vari dipartimenti di lavorare insieme per ottenere una visione unica e prendere solide decisioni strategiche.

Wolters Kluwer FRR è pronta a facilitare la vostra gestione operativa dei rischi nel rispetto delle normative. La nostra soluzione integrata e modulare ha ricevuto numerosi riconoscimenti da enti indipendenti del settore. La rivista Risk ci ha nominati "Sistema di segnalazioni di vigilanza dell'anno" per tre anni consecutivi e la Chartis RiskTech100 ci ha indicati come il fornitore numero 1 in due diverse categorie: segnalazioni di vigilanza e rischio di liquidità.

RiskTech 100 2022

Scopri la nostra soluzione finanziaria OneSumX su
www.wolterskluwer.com