RISK MANAGEMENT MAGAZINE

Vol. 17, Issue 2
May – August 2022

In collaboration with Wolters Kluwer

Direttore responsabile (Editor in chief): Maurizio Vallino – Condirettore (Managing Editor): Corrado Meglio
INDEPENDENT MAGAZINE 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 17, Issue 2, May - August 2022

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. +39 389 6946315

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 25th August 2022

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.

Overlaps between minimum requirements and capital buffers: the usability of the combined buffer requirement for Italian banks

Wanda Cornacchia (Directorate General for Economics, Statistics and Research, Bank of Italy) and Giulio Guerra (Directorate General for Financial Supervision and Regulation, Bank of Italy).

The current EU capital regulation[1] requires that banks comply with two main frameworks at the same time: one for prudential purposes, the other for resolution purposes.

The first one includes both a risk-weighted requirement (RW) and a leverage ratio requirement (LR). Similarly, the resolution framework, which ensures that banks have enough loss-absorbing and recapitalization capacity through a Minimum Requirement of Eligible Liabilities (MREL), is based on two ratios that are to be met in parallel: the MREL as a percentage of risk weighted assets (MREL-RW) and the MREL as a percentage of the total exposure measure used for the purpose of the leverage ratio (MREL-LR).

According to the EU regulation, the CBR is only required on top of the two risk-weighted requirements (RW and MREL-RW).

This asymmetry implies that the same Common Equity Tier 1 (CET1) capital can be used simultaneously to satisfy the CBR in one framework and a minimum requirement in another framework. In these cases, we talk about overlaps, which make it impossible to use (in whole or in part) the CBR to absorb losses without violating a minimum requirement.[2]

The term 'buffer usability' refers to banks' ability to use the CBR without breaching any minimum requirements. In the event of overlaps, a bank would not use (all or part of) the CBR even when allowed to do so because such a use would lead to a breach of a minimum requirement.

This issue, in turn, can undermine the decision of macroprudential authorities to release part of the CBR[3], i.e. to draw it down in order to allow banks to support the economy in bad times. Usability should not be confused with the unwillingness/reluctance of banks to use the buffers because of disincentives of various kinds (for instance, stigma effects due to financial market reactions or maximum distributable amount restrictions[4]).

A comprehensive measure of the overlaps (and hence CBR usability) can only be obtained by jointly comparing the use of the CET1 capital in each of the regulatory frameworks in place. That is why a comprehensive methodological approach is taken to measure the usability of the CBR.[5]

This approach differs from the one recently adopted by the ESRB[6], as it provides a broad overview of the actual usability of the CBR, considering that this is required not only on top of the RW requirement but also in addition to the MREL-RW requirement. Should the MREL-RW requirement prove to be higher than the RW one, the CBR may be more usable than it would be using the approach based solely on the RW requirement.

In panel (a) of Figure 1, a hypothetical bank is characterized by a binding MREL-LR. In this example, the MREL-LR is the requirement that absorbs most CET1 capital, fully overlaps with the CBR in the RW framework and partially overlaps with the CBR in the MREL-RW framework.

The CBR usability is therefore reduced to 50 per cent (1.5 per cent of RWAs instead of 3.0 per cent, as shown by the last bar in panel (a) of Figure 1).

In this case, if we only consider the CBR stacked on top the RW framework, the buffer usability would be equal to 0 per cent (i.e. the CBR usability from the interaction between the RW and MREL-LR).

A discrepancy between the approach focusing on the RW framework alone and our comprehensive approach can also occur when a bank is characterized by a binding MREL-RW requirement, if the MREL-LR requirement met with CET1 capital is higher than the RW requirement.

In panel (b) of Figure 1, the CET1 absorbed by the MREL-RW and MREL-LR is higher than the CET1 absorbed by the RW requirement. The CBR usability would be equal to 0 per cent if we only compared the interaction between the RW and MREL-LR requirements. With the full comparison proposed in this article, the CBR usability is instead 100 per cent, since the CBR in the MREL-RW framework does not overlap with any other requirement.

---

1 The 'banking package' comprises the Capital Requirements Directive and Regulation (CRD V/CRR II), the Bank Recovery and Resolution Directive (BRRD II) and the Single Resolution Mechanism Regulation (SRMR II).

2 The consequences of the violations are proportionate to their seriousness: those relative to the minimum requirements can lead to the declaration of failure (or risk of failure) of the bank and eventually to liquidation or resolution procedures; those relative to the CBR can lead to, among other things, limits on the distribution of dividends.

3 The CBR includes both releasable and non-releasable buffers by macroprudential authorities: for example, the CCyB is releasable while the CCoB is not. Anyway, all buffers included in the CBR are usable.

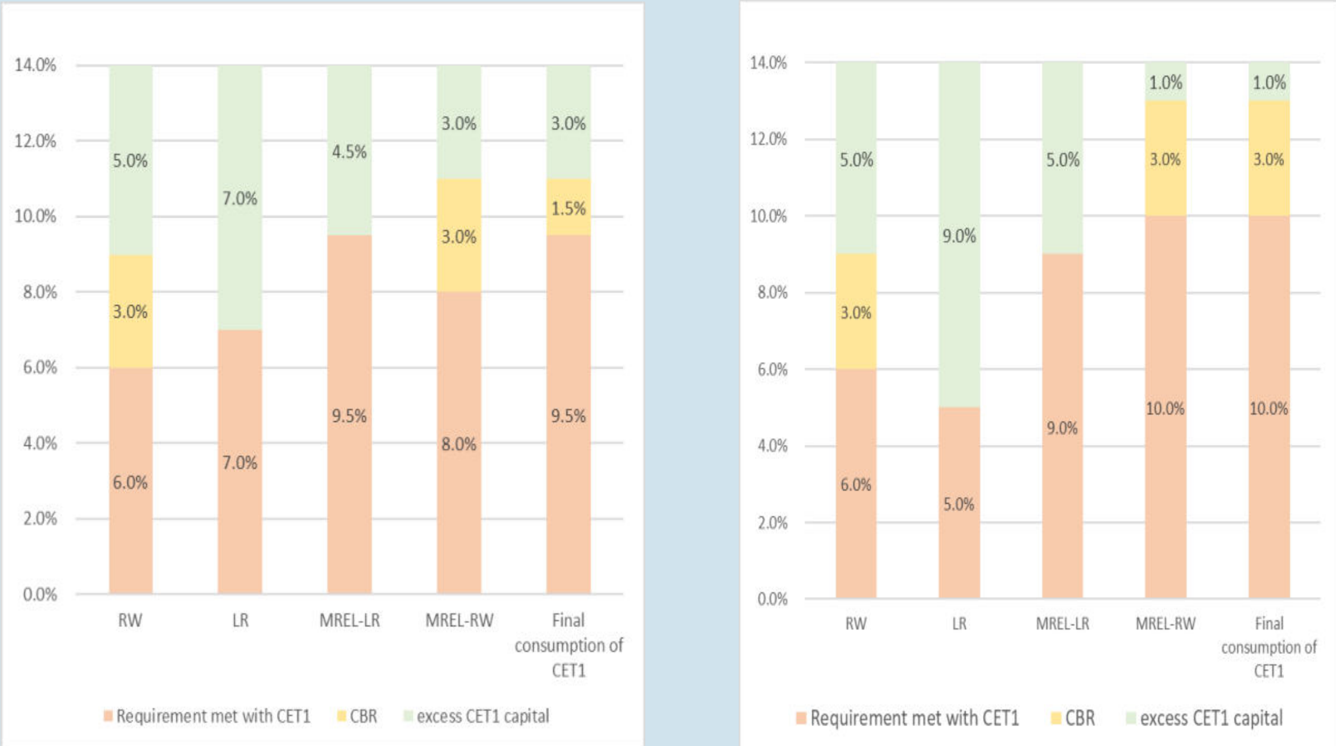
4 Article.141 of CRD IV introduced the concept of the Maximum Distributable Amount (MDA), which requires supervisory authorities to automatically restrict earnings distribution in the event of a CBR breach. Similarly, the MREL-MDA (M-MDA) is imposed by resolution authorities, though with more discretion and no automaticity.

5 For more details on the proposed methodology, see W. Cornacchia and G. Guerra, Overlaps between minimum requirements and capital buffers: the case of Italian banks, Notes on Financial Stability and Supervision – Bank of Italy, No.30, June 2022.

6 For further details, see the ESRB, Report of the Analytical Task Force on the overlap between capital buffers and minimum requirements, December 2021.

Figure 1: Final effects of the overlaps
(per cent of RWAs)

(a) **MREL-LR binding**	*(b)* **MREL-RW binding**



Note: Each panel shows the bank's CET1 ratio in per cent of RWAs. In particular, the histograms indicate how CET1 is used in the different frameworks: RW, LR and MREL in its weighted and leverage based dimensions (MREL-RW and MREL-LR). The notion of a stacking order defines the sequence in which different CET1 capital layers absorb losses. In the risk-weighted capital framework (the RW bar) excess capital covers losses first (green), followed by the CBR (yellow) and minimum requirement (pink). The distinction originates from the different consequences of any breach. The CBR is only required on top of the two risk-weighted requirements (RW and MREL-RW).

Table 1 highlights the contribution of the comprehensive approach being proposed by applying it to actual data for Italian banks. Around one fourth of the Italian banks are constrained by one of the leverage-based requirements (LR, MREL-LR, TLAC-LR): in such cases, an overlap occurs and reduces the CBR usability. The table shows the difference in the CBR's usability between the proposed approach and the RW approach. As illustrated in the two previous examples, applying the comprehensive approach reveals a significantly higher CBR usability: the CBR usability increases from 26.7 to 73.6 per cent for the whole Italian banking system. This improvement is driven by the Italian banks subject to MREL requirements (11 out of a total of 150 banks, which account for 80 per cent of total system assets), whose CBR usability increases from 10.8 to 69.0 per cent. Indeed, when the MREL-RW requirement met with CET1 is higher than the RW one (as in panel (b) of Figure 1), the CBR is more usable than is apparent from the approach based solely on the RW requirement. This explains why, by considering the regulatory requirements of the resolution framework as well, the usability of the CBR increases.

Table 1: CBR usability of Italian banks
(per cent of CBR; data as of December 2020)

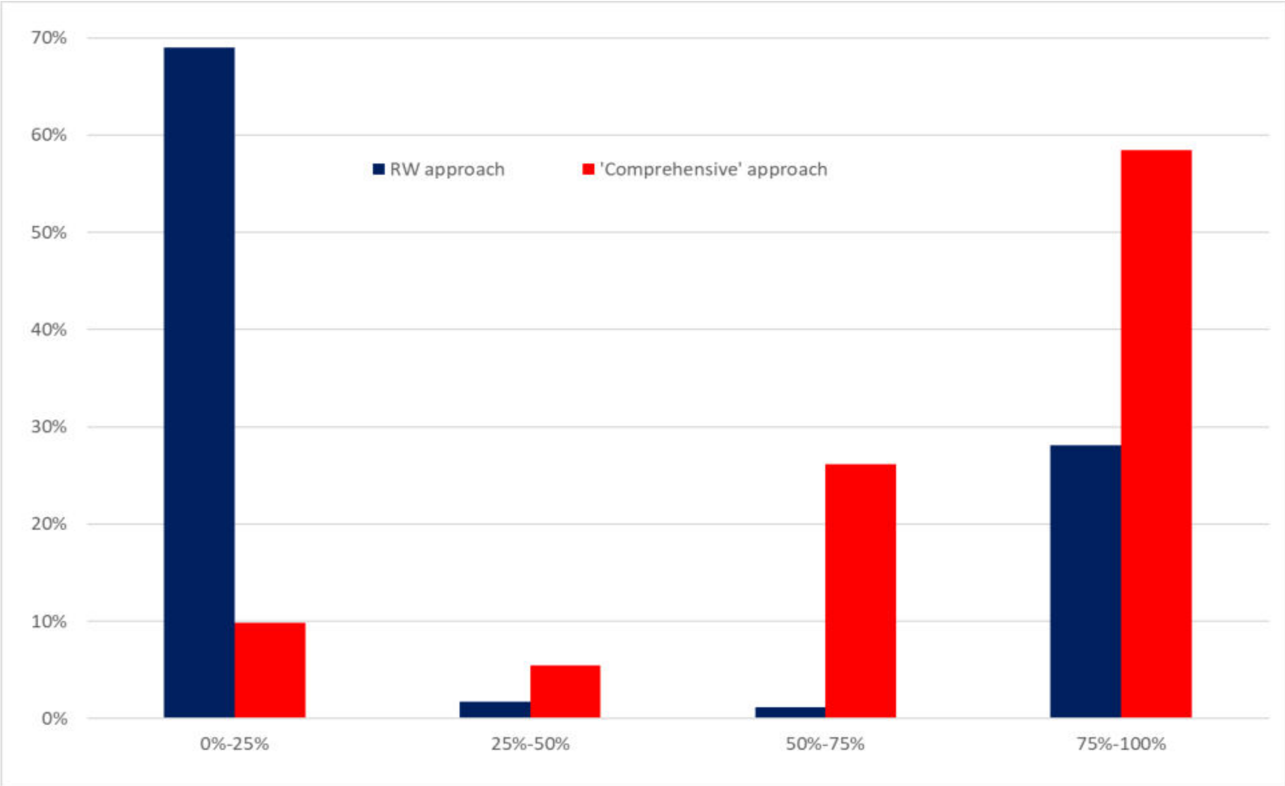
	RW approach	Comprehensive approach
Banks with MREL requirements (11 banks)	10.8	69.0
Whole system (150 banks)	26.7	73.6

Source: Supervisory and resolution reporting.

Note: The assumption of closing the shortfalls applies in each column. In our calculations, the P2R is included in the minimum RW requirements.

The difference between the two approaches to measuring the CBR's usability also emerges from the distribution of the risk-weighted assets (RWAs) of Italian banks by buckets of CBR usability (see the figure 2). In particular, based on the RW approach, almost 70 per cent of the banking system RWAs are attributable to banks with a very limited CBR usability (between 0 and 25 per cent). According to the comprehensive approach, instead, 85 per cent of RWAs are attributable to banks with a medium/high CBR usability (above 50 per cent).

Figure 2: Distribution of Italian banking system RWAs according to CBR usability
(share of banking system RWAs by bucket of CBR usability; data as of December 2020)



Source: Supervisory and resolution reporting.

Machine Learning for Credit risk: three successful Case Histories

Paolo Di Biasi (Intesa Sanpaolo), Rita Gnutti (Intesa Sanpaolo), Andrea Resti (Università Bocconi and senior advisor CRIF), Daniele Vergari (CRIF[1])

Abstract

As the financial services landscape witnesses an unprecedented change, banks can use machine learning ("ML") to expand their databases through alternative sources providing unstructured and semi-structured information, such as transaction data and digital footprint data. However, ML algorithms also suffer from several potential shortcomings, as they may overfit sample data and prove unstable over time, they may quickly become obsolete and need re-estimation, and they may prove hard to interpret. This paper joins the debate on ML in banks by providing three case studies that highlight the benefits of machine learning, while showing how its drawbacks can be minimised: a rating model developed within the IRB framework, a challenger model used to validate a bank's main model for retail PDs, and an early warning system based on transaction data.

1. Foreword

The use of machine learning ("ML") models in banks has raised considerable interest and sparked a lively debate, among both scholars and practitioners. As the financial services landscape witnesses an unprecedented change (due, e.g., to the digitalisation of credit processes, open banking regulations, competition from non-bank players and more prescriptive regulations), banks can use ML to expand their databases through alternative sources providing unstructured and semi-structured information, such as transaction data and digital footprint data. Additionally, ML can manage multi-dimensional data by automatically selecting meaningful features and creating meta-variables that summarise the most relevant information, thereby improving the performance of internal scoring/rating systems. On the other hand, ML algorithms also suffer from several potential shortcomings: first, they may overfit sample data and prove unstable over time (meaning that they perform poorly when applied to new data); second, they may quickly become obsolete and need recalibration/re-estimation to keep attaining high performance standards; third, quality checks for unstructured data (e.g. natural language) may prove more complex to run than they are for structured data sources; finally, ML models may prove hard to interpret, meaning that – although they lead to overall correct results – the factors driving their outcomes may prove hard to pinpoint (or may vary sharply between individuals, making it hard for users to identify a consistent pattern).

Over the last 20 years, ML has become increasingly popular with both banks and supervisors. A bibliographic search for the keywords "machine learning" and "bank" (or "banking", or "supervision") finds 41 relevant articles and two book chapters published between 2000 and 2021 (Guerra and Castelli, 2021), with a significant acceleration after 2016 (see Figure 1)[2].

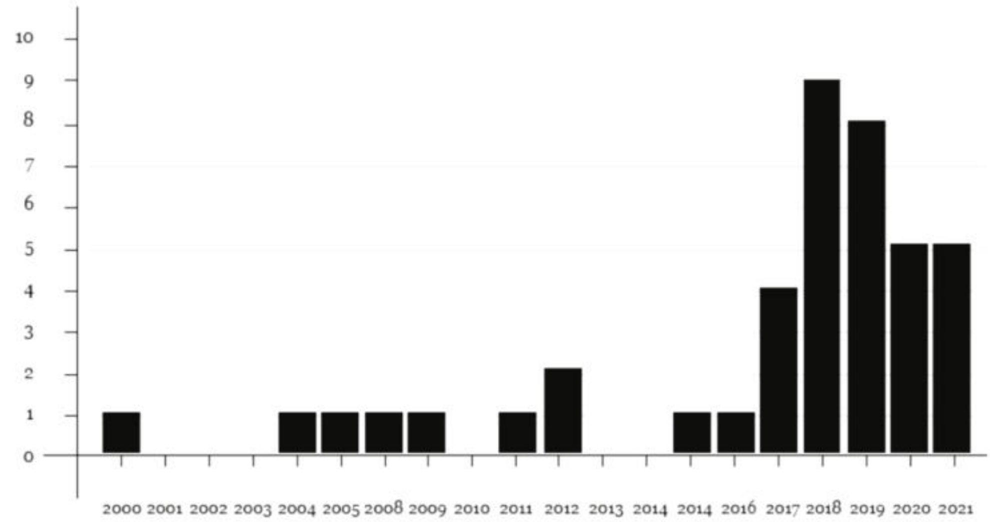


Figure 1 - Articles published in 2000-2021 referencing ML and banks/banking/supervision (source: Guerra and Castelli, 2021)

[1] This article draws on a paper (Di Biasi et al., 2022), written with Angelo Basile, Fiorella Bernabei, Cristina Caprara, Dario Cavarero, Mattia Marigliano, Roberta Ranaldi e Marco Vignolo (https://www.crif.it/ricerche-e-pubblicazioni/altre-risorse-e-ricerche/2022/marzo/l-applicazione-di-tecniche-di-ml-al-credit-risk-management-e-ai-modelli-irb/). We gratefully acknowledge comments from Simone Casellina (European Banking Authority), Francesco Cannata (Bank of Italy), as well as from various participants in a presentation to the EBA held on 17 March 2022.

[2] A similar survey, focusing on deep learning applications to banking and finance in 2014-2018, found 40 academic articles dealing mainly with stock market prediction and trading, while credit risk models only account for 12.5% of the published studies (Huang et al., 2020). Further surveys include (Leo et al., 2019) and (Rundo et al., 2019). The former looks at applications of ML in the management of banking risks (credit, market, operational and liquidity) and finds that ML usage doesn't appear commensurate with the importance of risk management and ML studies when considered separately; the latter looks at ML usage in the field of quantitative finance (including comparative studies about the effectiveness of ML-based systems), showing that innovative methods often outperform traditional approaches.

More recently, ML has been widely used by financial institutions worldwide. Out of 60 banks surveyed by the IIF in 2019[3], 25 were using ML models in a production setting (compared to 23 one year before), and another 27 were engaged in pilot projects (up from 12 in 2018). Credit scoring and early warning systems were by far the most common areas of application. As for 2020, there is evidence that investments in ML techniques by banks were not negatively impacted by Covid-19; indeed half of the UK-based institutions surveyed by the Bank of England (Bholat et al., 2020) expected an increase in the importance of ML and data science for future operations as a result of the pandemic.
Having implemented ML models in banks for several years, we would like to join the debate providing three case studies that highlight the benefits of machine learning, while showing how its drawbacks can be minimised. Table 1 provides a synopsis of the main characteristics of each project.

Table 1 – Synopsis of the three case studies discussed in this paper

Case study	*Benefits*	*Data and algorithms*	*Interpretability techniques*	*Main challenges*
A new IRB model aimed at retail SMEs	Provides customers with a full digital experience, focusing on high digitalisation standards and profiting from PSD2 and the GDPR. Deals with unprecedented conditions, e.g. due to Covid-19	Transaction data, credit cards, POS, and web sentiment. Decision trees, random forests and gradient boosting	Partial dependence plots, individual conditional expectation, LIME, and SHAP	Hiring new staff, investing in IT, reviewing regulations, deploying an ad-hoc validation framework
A challenger model to validate the IRB model for retail PDs	Performs initial validation to get supervisory clearance, and ongoing validation to promptly highlight any issues emerging from the bank's model	Innovative transaction data based on borrowers' current accounts. Decision trees with extreme gradient boosting ("XGB")	Feature importance analysis	Project timing and the choice of a trade-off between a full challenge and the need to ensure comparability
An early warning system based on transaction data	Ability to use transaction data to identify new information patterns	Individual transactions combined with pre-existing information. Random forests, XGB and neural networks	Feature importance analysis, SHAP, LIME and OptiLIME CRIF	Feature selection, model development and interpretability

In the remainder of this paper, each project will be described in detail: we start with two projects that relate to rating systems used for regulatory purposes, then we move to a "managerial" model focused on early warning systems.

2. Case 1: using ML for a new IRB model aimed at retail SMEs

The data available on SMEs (small and medium-sized enterprises) has recently experienced a sharp increase as new sources have emerged, providing value added to improve risk management models. Against this backdrop, a large bank decided to update its IRB model for the retail SME portfolio, comprising about 500,000 businesses with a turnover of up to €2.5 million and a credit exposure of less than €1 million. These are typically medium-sized limited liability companies, partnerships and sole traders/entrepreneurs. The bank launched a "Smart Lending" project, to provide retail SMEs with a full ("end-to-end") digital experience. Focusing on high digitalisation standards, the project aimed to profit from the evolution of the technological and regulatory framework (as PSD2 and GDPR, regulating data and customer protection, opened the door to the possibility of receiving new data for an all-round customer assessment). The rating model had to be available online and in real time, without easing risk management standards. ML algorithms provided a tool to improve the customer journey as well as rating performance, while also accounting for extraordinary events such as the Covid-19 pandemic.
An ML component can be introduced into credit risk models in two ways: by replacing previous models or by complementing them through new algorithms. In this case, ML was used to add new data sources to traditional ones, enhancing the breadth and accuracy of the pre-existing approaches. Accordingly, a new IRB model was developed that uses (see Figure 2):

- *traditional modules* to process information already used by the rating model, such as financial statements, external data (e.g., central credit registers like CRIF) and internal performance indicators;
- *modules using new data sources*, both internal and external to the bank, processed through either ML or traditional algorithms.

[3] See (Institute of International Finance, 2019).

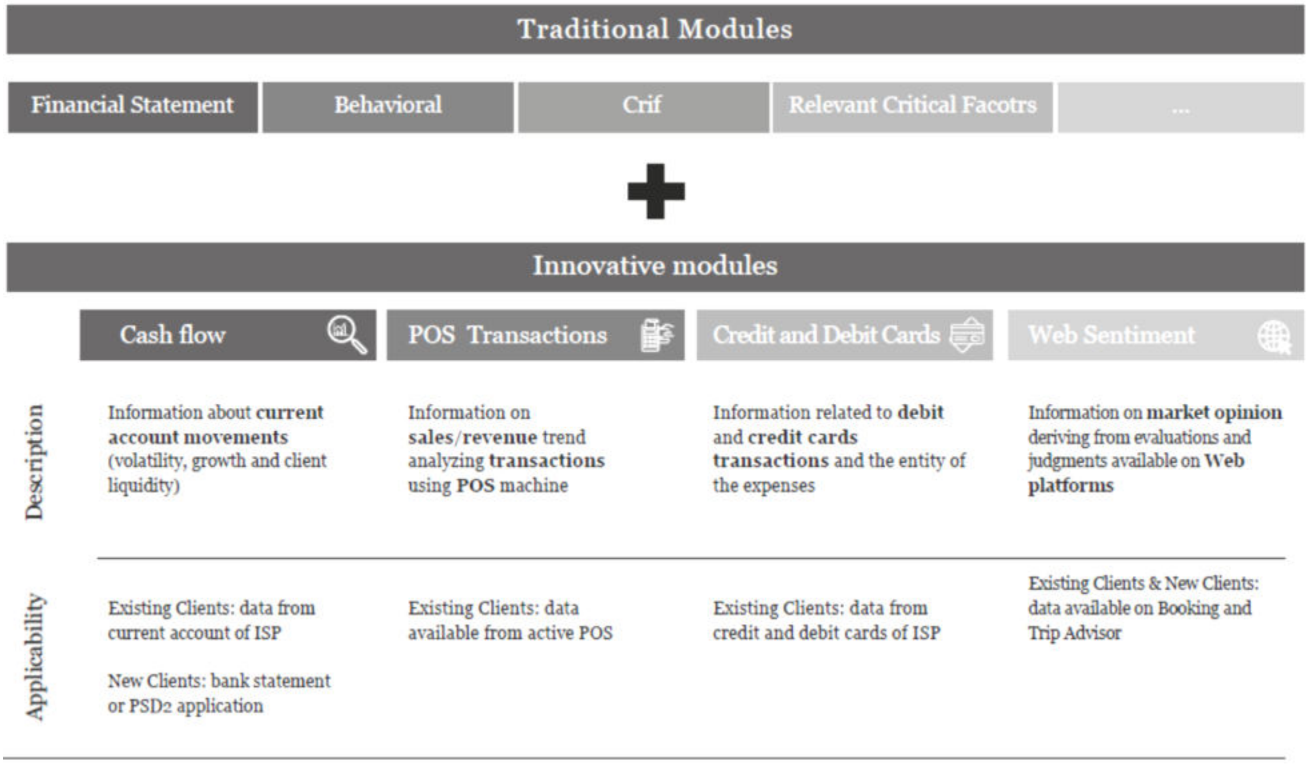


Figure 2 – Data sources of the new model for retail SMEs

New data sources include transaction data (from the bank's own current accounts and from other banks, available through PSD2), point-of-sale transaction data, payment card data, web sentiment data captured from web pages. Among these sources, transaction data is the most challenging, but also the best performing. It allows recent key information to be collected, and transforms the constraints imposed by PSD2 (forcing the bank to transfer information to third parties upon request) into an opportunity.

The algorithms selected for the ML component were[4]: decision trees, random forests and gradient boosting. Deep learning algorithms were not used, given their complexity: instead, a gradual approach was chosen, using algorithms that were more advanced than – while at the same time comparable to – logistic regression. As a benchmark to gauge their performance (and to improve interpretability) a traditional logistic regression was also run on the same input data.
Interpretability remained a key aspect throughout the estimation process. The main techniques deployed (besides traditional models used as a benchmark) were Partial Dependence Plots ("PDP"), Individual Conditional Expectation ("ICE"), LIME (Local Interpretable Model-agnostic Explanation, a technique that identifies the features that contribute most to an individual classification through a local approximation performed on slightly modified versions of the original observation[5]), and SHAP (SHapley Additive exPlanations, a relatively recent approach combining features from LIME and Shapley)[6].

The most effective methodology proved to be SHAP. It assigns a marginal contribution to each variable (feature) considering its possible interactions with other variables: for each combination of variables, the change in PD is measured, providing a basis to compute the relative weight of each feature. Figure 3 provides a sample SHAP plot, showing the range and impact of each variable by order of importance. Each point in the plot is a Shapley value (measured on the x axis) for a feature (as listed on the *y* axis, in decreasing order of importance) and an instance. The colours represent the value of a feature from low to high. Feature labels have been redacted for confidentiality reasons.

---

[4] See (European Banking Authority, 2020) for a brief discussion, and(Breeden, 2021) for a taxonomy of ML algorithms applied to credit risk.

[5] See (Ribeiro et al., 2016). An optimised version of LIME, tackling the instability problems that undermine its reliability, was proposed by (Visani et al., 2020): under this new approach ("OptiLIME CRIF"), stability is maximised for any chosen level of "adherence", i.e. similarity to the original ML model.

[6] Shapley values (a measure of how much each feature contributes to a prediction, based on a large number of comparisons between pairs of alternative feature sets); the Shapley value can "split" an individual prediction among all contributing features, providing a full explanation of why a given applicant has received a specific credit score. As noted in (Molnr, 2019), this can make it preferable in situations where the law provides customers with a "right to explanations". A thorough discussion of Shapley values is provided, e.g. in (Giudici and Raffinetti, 2021), which also introduces an extension of the original Shapley approach ("Shapley-Lorenz") based on Lorenz decompositions. LIME, Shapley and SHAP are known as *local* interpretation techniques, as they analyse how individual model predictions change when altering input data. They are mostly used to produce a visual representation that reflects the contribution of each feature (explanatory variable) to a single-point forecast, assigning a "weighted" importance to the characteristics that most affect the output generated by the model (e.g. default probability). Conversely, *global* interpretation techniques are aimed at understanding the relationship between each main feature (explanatory variable) of a model and its target variable in a more "traditional" way.

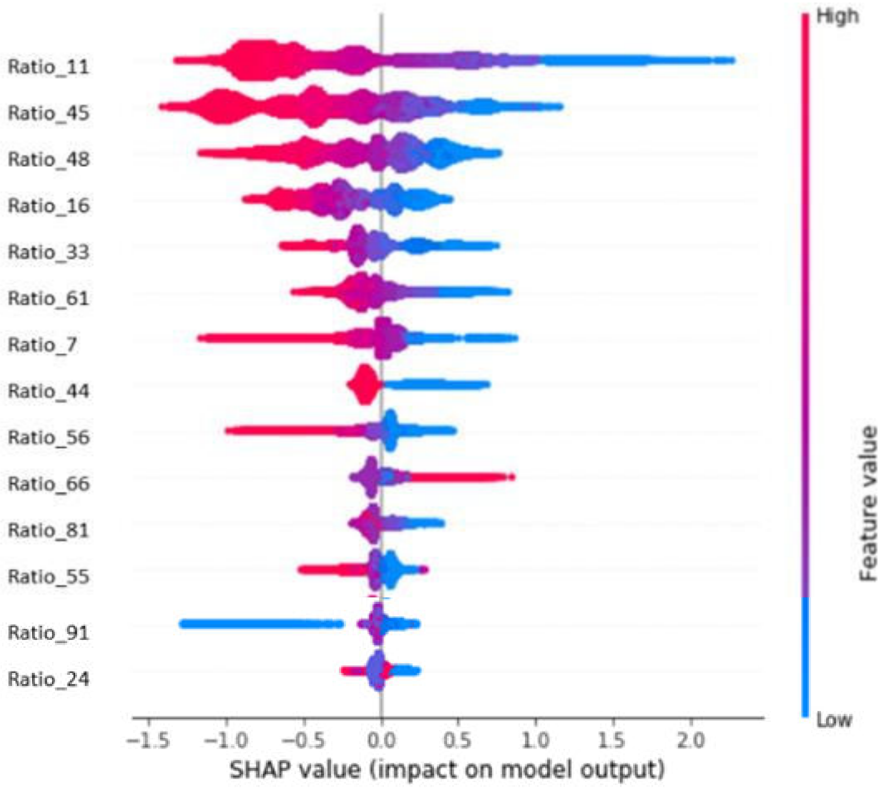


Figure 3 – Cash inflows for retail SMEs ("RS") estimated using transaction data

ML models showed the best performance (accuracy ratio) and a good out-of-sample stability. To assess overall performance, a benchmark model was created that used neither innovative sources, nor new algorithms. Keeping that model as the base case, marginal increases in the accuracy ratio were measured: new data sources processed through traditional methods led to a 5% increase; a further 5% increase emerged when the same sources were processed through ML. If ML had also been used to integrate the various modules in Figure 2, that would have led to a further 1% increase; that option, however, was not considered in the final model, as the improvements achieved in the previous steps were already satisfactory.

The new module using transaction data also proved very responsive in identifying changes in PD drivers following the Covid-19 outbreak. Indeed, the score based on transaction data remained stable in 2019, and quickly worsened during lockdown, capturing information that traditional default-forecasting models could not use in full. An example of its determinants is provided in Figure 4, showing cash inflows estimated from transaction data for different borrower groups (including industries that were deemed especially vulnerable to the drop in demand associated with lockdown regulations).

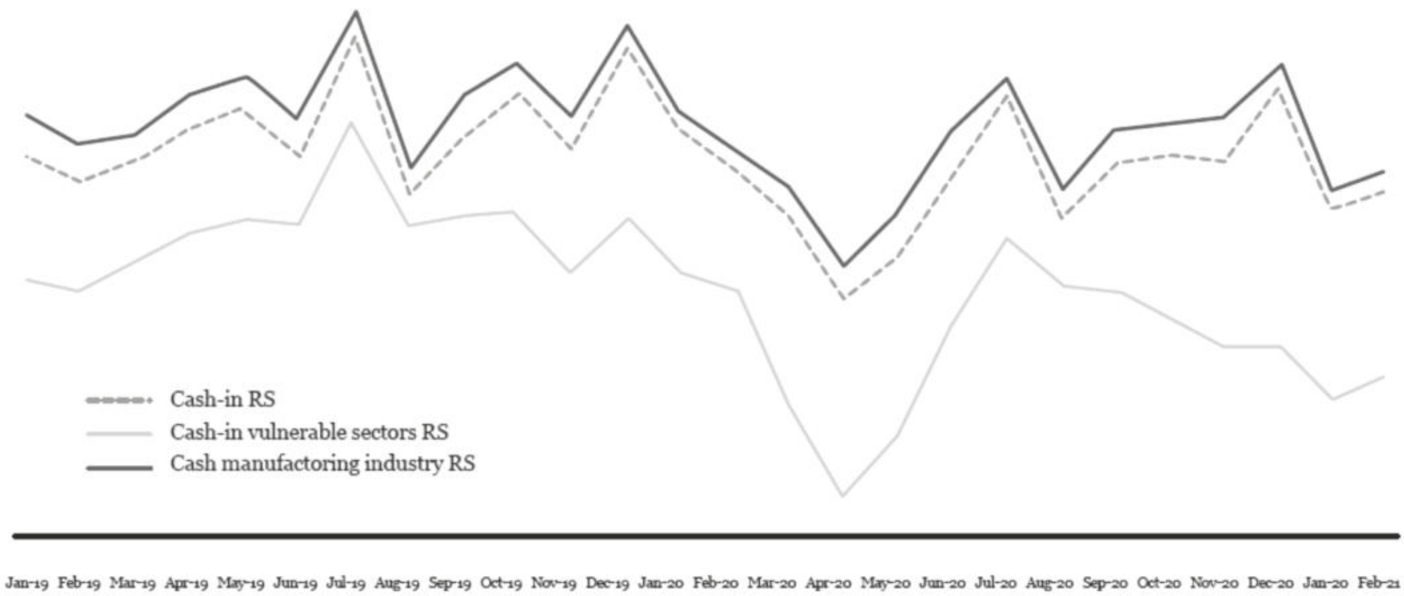


Figure 4 – Cash inflows for retail SMEs ("RS") estimated using transaction data

The ML module initially led to challenges that were (and to an extent are still being) addressed:

- the need to hire new experienced staff in a variety of fields: big data management and related programming languages, machine learning algorithms with a focus on transparency and interpretability, ability to provide adequate documentation, and knowledge of the underlying banking processes;
- a significant investment in IT, in both the estimation and implementation phases, in order to have an infrastructure that could assess machine learning algorithms on estimation samples and replicate them in a production environment;
- a review of the regulatory framework to check that the model was in line with the requirements imposed under the IRB approach. Special attention was paid to Article 179 a) of the CRR ("an institution's own estimates of the risk parameters PD, LGD, conversion factor and EL shall incorporate all relevant data, information and methods. […] The estimates shall be plausible and intuitive and shall be based on the material drivers of the respective risk parameters") and especially to the need to guarantee "plausible and intuitive" estimates. This meant that the bank had to be able to explain the model results, using the techniques (like LIME and SHAP) discussed above;
- finally, ML algorithms required an ad-hoc validation framework, taking into account their specificities (e.g., the optimization of hyperparameters[7]).

The new model allows for a fully digital lending process with ratings computed online, in real time and automatically, leading to greater efficiency and superior credit risk monitoring. The excellent performance of the new ML algorithms, using data retrieved in a fully automated way, allowed the rating to be produced without having to ask the borrower for financial statements/tax forms, which can weigh on the customer journey while becoming quickly obsolete. The model was validated by the ECB in May 2021 and is currently used to compute regulatory capital against credit risk.

3. Case 2: a challenger model to validate the regulatory model for retail PDs

A large European bank had recently changed its IRB model for estimating the PD of retail customers, in order to accommodate the EBA Guidelines on risk parameters and the new definition of default. Once the new model ("the bank's model") had been developed by the risk management department (using a combination of traditional and ML techniques), it had to be assessed by the validation unit as prescribed by the relevant regulations. The bank's pre-existing validation framework had to be reinforced by adding a "challenger" model, based on the same technology as the bank's model (including ML-based modules), to be used for *initial validation* (assessing alternatives to the bank's model obtained by changing/stressing some choices) and for *ongoing validation* (promptly highlighting any issues, including performance drops, in the bank's model).
The bank's model relies on the calculation of an "integrated score" incorporating several intermediate scores generated by specific modules (using, for example, personal data, CRIF's credit bureau data and scores, as well as other socio-demographic information). The challenger model had to focus on a few key modules that were considered especially relevant to the bank model's results. Namely:

- the *financial assets ("AFI")* module, which assesses the borrower's financial assets (including the current account balance) and represents a measure of the borrower's potential wealth;
- the *behavioural* module, focusing on the sub-module that evaluates a borrower's behaviour on the basis of that borrower's credit exposure and financial position with the bank;
- the *mortgage module,* which looks at the products owned by customers who also have a residential mortgage;
- the *cash flow* module, which uses current account data at a transaction level in order to assess the borrower's cash flow management (volatility, growth and liquidity levels) and to identify potential warnings, financial tensions and other income/expense flows. Here, both the bank's model and the challenger model used ML;

The challenger model experimented with many options, e.g. by changing variable categorisations and reducing cross-module correlations, testing different time frames for the indicators, applying different criteria when setting up the estimation sample, adding new indicators to the list of candidate variables. As concerns ML, the challenger model tested alternative solutions for hyperparameters, as well as simplified models based on different materiality thresholds.
By doing so, alternative results were generated for the four "challenged" modules, which were then combined (using the same methodology as in the bank's official model) with the results provided by "un-challenged" modules.

[7] Hyperparameters are parameters whose values control the structure and the learning process of an ML model, thereby determining the number and values of its final parameters. An example of hyperparameters could be the number of nodes and layers in a neural network (whereas its parameters are the weights used in the functions propagating information across nodes). An example of ML-based validation approaches is provided in Case 2 (§3).

The most interesting module was the cash flow module, where innovative transaction data was used, based on borrowers' current account movements.
The module is only applicable to customers using their current account as their "main" account, that is, for day-to-day transactions[8] (excluding accounts that are seldom used and do not correspond to the customers' real habits).
The input data was updated monthly and contained all the transactions taking place on a daily basis. The database aggregated both current accounts held with the bank and, with the explicit consent of the customer under PSD2, current accounts held with other banks[9].
The challenge activity was performed primarily through the tuning of the hyperparameters of the *extreme gradient boosting* XGB algorithm, the same one used in the bank's model[10].
To improve the comparability of the results, XGB was preferred to alternative techniques, such as random forests and deep neural networks.
The tuning activity focused on the following hyperparameters: learning rate, number of estimators (trees), maximum depth for each estimator, minimum child weight, column sample by tree, subsample, and gamma.
While most hyperparameters were already present in the bank's model, two of them (column sample by tree and subsample) were added by the validation unit. By sampling the features and the observations for any given tree, the risk of overfitting should be reduced.
A selection of the challenger models tested, as well as a comparison with the bank's model, is shown in Table 2[11]. Looking at the table, Challenger 5 ("Ch. 5") emerges as the challenger model with the highest overall accuracy, while Challenger 3 is the one with the best test sample accuracy. However, all challengers are characterized by a fairly stable performance in the test sample.

Table 2 - Challenging the hyperparameters (illustrative data)

			Challenger Model - candidates				
		Bank's Model	**Ch1**	**Ch2**	**Ch3**	**Ch4**	**Ch5**
Hyperparameters	Learning rate	0.1	0.1	0.1	0.1	0.1	0.1
	No. estimators	200	300	300	300	300	500
	Max. depth	5	5	5	5	5	5
	Min. child weight	10	10	5	5	5	5
	Column sample by tree	-	-	-	-	0.6	0.6
	Subsample	-	-	-	0.8	0.6	0.6
	Gamma	0.8	0.8	0.8	0.8	0.8	0.8
Results – AR (Gini Index)	Development sample	77.94%	80.06%	80.52%	80.63%	79.87%	82.83%
	Test sample	72.54%	72.69%	72.59%	72.78%	72.48%	72.46%
	Overall Sample	76.86%	78.58%	78.58%	79.06%	78.39%	80.76%

To gain a better insight into the underlying logic of the models, we computed the marginal contribution of each variable.
These scores quantify the relative importance of each variable when a model makes a prediction, allowing the most important ones to be identified.
The marginal contributions of all features can then be shown in decreasing order on a graph for each candidate model (Figure 5 provides an example).

[8] To be considered the "main" account, a current account has to meet at least one of these two criteria: i) it is used for receiving the customer's main source of income (salary, pension or welfare payments, self-employed income, housing rents or alimony); and/or ii) the customer receives money transfers or deposits cash on a regular basis.

[9] Keyword searches and tagging were performed on raw data, looking at different transaction types and their descriptions. All indicators were computed for a maximum of 12 months, merging all accounts to reflect the customer's situation as closely as possible. The list of candidate indicators was built by averaging, summing or detecting status changes over the last quarter, semester and year (e.g. the average salary was computed over the last 3, 6 and 12 months).

[10] XGB belongs to the category of ensemble models, as it is composed of a series of weak learners or decision trees, which are built upon in order to generate one final strong learner.

[11] For confidentiality reasons, the table only reports illustrative data.

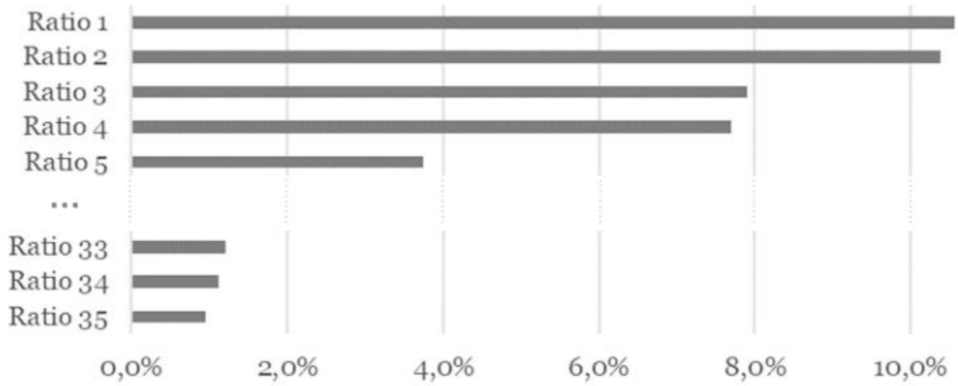


Figure 5 – Importance value graph of the first 35 indicators (illustrative data)

Based on such scores, five different cut-off values were considered: no cut-off, 0.7%, 0.8%, 0.9% and 1%.
For each cut-off strategy, the accuracy ratio ("AR") in the train sample, test sample and total sample were computed. Then, the average contributions to the AR of the variables included in each strategy were computed, and the strategy with the maximum average contribution was selected (see the illustrative example in Table 3).

Table 3 – Model selection (illustrative)

		#Variables				
		A	B	C	D	E
	Sample	90	40	35	30	28
AR	Train sample	80.63%	78.14%	77.87%	76.98%	76.64%
	Test sample	72.78%	71.59%	71.29%	70.53%	70.29%
	Total sample	79.06%	76.83%	76.56%	75.69%	75.37%
Avg. Contribution to AR of inserted variables	Train sample	0.05%	0.05%	0.15%	0.17%	0.18%
	Test sample	0.03%	0.05%	0.13%	0.12%	0.11%
	Total sample	0.05%	0.05%	0.14%	0.16%	0.15%

Criteria used to choose the final challenger model were: a balanced performance in both the test and the train sample, a lower number of variables, a set of features that were more intuitive and explainable to the model's users.

The final decision also considered the results obtained with alternative ML techniques (e.g., random forest), looking at common features and how their relative importance varied across models.

The challenger model enabled the validation unit to perform benchmarking of the bank's official model, looking at:

- the performance of all modules subject to challenge and their final accuracy;
- the different weight of the modules once they were combined into the integrated score;
- the change in the PD master scale when moving from the bank's model to the challenger model;
- the differences and special characteristics in the PD distributions (as shown in Figure 6).

The results of the challenger model were almost comparable to the bank's model: the differences in terms of rating were mostly within one notch; the increase in the accuracy ratio was almost immaterial and corroborated the choices made by the risk management department.

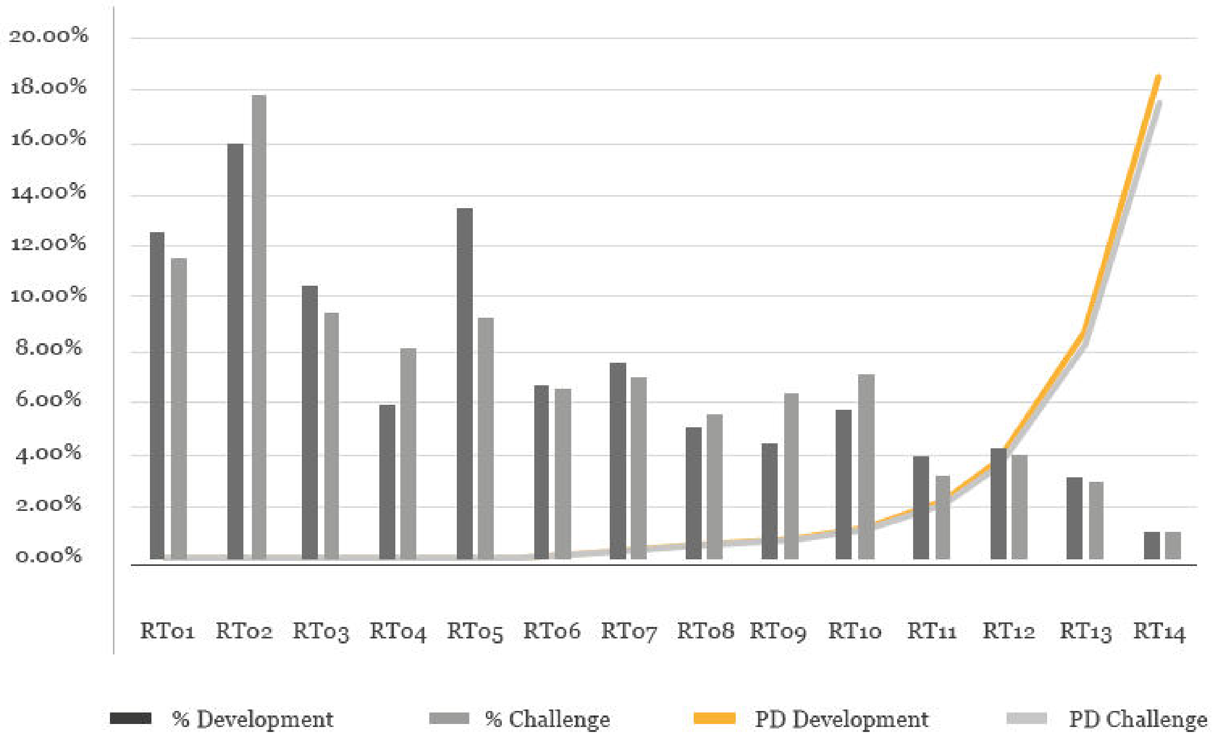


Figure 6 –Relative weight and average PD associated with different rating grades (bank's model vs. challenger model)

The main issues met when developing the challenger model were the following:

- project timing. Before the challenger model can be developed, the validation unit must learn about the bank's model and perform standard validation tests on it, in order to identify areas where the need for benchmarking is stronger. All this must take place in the period between the bank's "pre-application" to gain supervisory recognition of the new model and its final "application";
- the trade-off between the wish for a full challenge, which would completely overturn the framework adopted in the bank's model, and the need to ensure comparability of the results between the challenger model and the official one. As shown above, this quest for comparability means that the basic methodological choices made in the bank's model were kept unchanged in the challenge activity.

In short, the main issues faced were related to "how" the model had to be challenged through independent validation; we chose to act on the list of candidate variables and on the "judgmental choices" made by the risk management department when estimating the bank's model, taking into account only statistical evidence.

4. Case 3: using transaction data for early warning purposes

This case involves an Italian significant institution, a traditional retail bank serving both consumers and businesses and operating domestically. The project was part of a roadmap for the improvement of its early warning systems. The objective was two-fold: responding to supervisory requirements and adopting a cutting-edge early warning tool, with a view to improving risk management, including in a post-Covid-19 context.
The bank's pre-existing early warning model was based on a combination of different types of traditional data (e.g., socio-demographic information, internal and external credit-related signals, financial data, etc.). Traditional modelling approaches (including logistic regression) had proved very effective in discriminating risk and were generally accepted and understood by key users within the bank (e.g. credit analysts) and external stakeholders (e.g. supervisors).
However, due to technological and methodological advances, new transaction-level data was available and ready to use, which could be an important source of information if properly managed: thousands of new indicators could be developed for millions of customers. Where traditional approaches may not be able to fully exploit this data, one would expect machine learning models to identify new information patterns, including in the event of highly non-linear relationships between customer behaviour and credit risk.
The pre-existing solution only used behavioural information at an aggregated level, e.g. by looking at the number of transactions recorded in the last *n* days, at the number of days elapsed since the last transaction took place, or at the average credit amount used over a certain period of time. The bank was now interested in using "atomistic" transaction data that could potentially highlight

specific behaviours: for instance, instead of just considering the number or transactions, one may also look at the nature of the individual items (e.g., payment of instalments and taxes, purchase of goods or services, settlement of invoices, bank transfers from the parent company, etc.), and their mix and evolution over time.
The aim of the project was to enrich the pre-existing early warning solution through the development of several transaction-level modules (each one for a different customer segment) and their integration with the other components of the system (e.g. modules monitoring central credit registry data, etc.).
Leveraging more granular, powerful and faster-reacting transaction-level information, the new solution was expected to better support the bank in managing credit risk in a more challenging economic environment. The data for the ML module consisted of individual transactions for the entire customer base (both individuals and businesses, 24 months of daily transactions, more than 10 million transactions per month).
Each record included the sign, amount, and label ("description") accompanying each transaction.
The model development involved two steps. First, transactions were categorised and turned into customer-level indicators. Second, such indicators were merged with pre-existing information and the early warning system was developed.
Regarding the first step (developing customer-level indicators), transactions were categorised by means of machine learning algorithms integrated with NLP (Natural Language Processing) techniques.
These algorithms "learned" how to interpret current account transactions by reading their description, then assigned them to multiple classes of activities (for example: accounting services, legal fees, purchases of raw materials or services, interest payments, etc.). The algorithms predicted the most likely class of activity and assigned a "reliability" score to the categorisation in order to help identify (and potentially discard) weak outcomes. The process involved human supervision of the machine learning algorithms, aimed at introducing further calibrations if necessary. Once transactions were categorised, they were summarized into 20,000 customer-level indicators, to facilitate data handling, model governance and interpretability.
As far as the second step is concerned (developing the actual early warning system), ML methodologies were applied for both feature selection (i.e. to create a "short list" of relevant variables) and multivariate estimation[12]. The techniques used included three different model classes: random forests (RFs), extreme gradient boosting (XGB) and neural networks (NNs).

For each class, model estimation was performed through the following steps:

- hyperparameter definition (e.g. definition of the minimum number of items in a leaf) by means of a random search engine. This provided a higher configuration space for estimating models than a grid search (Bergstra and Bengio, 2012), thus achieving better results. Table 4 provides an example of how model performance can be affected by changing certain hyperparameters and leaving everything else unchanged;

Table 4 - Random Search Approach in a Random Forest Model

Random Search approach				
No. estimators	min. samples split	min. samples leaf	max. depth	Accuracy
200	**7**	**8**	**8**	**78.28%**
...	...	...	...	...
100	20	50	8	77.76%
...	...	...	...	
200	5	40	8	76.86%

- cross validation: in order to avoid overfitting, cross validation was performed through validation strategies ranging from 3-fold to over 10-fold. In the case shown in Table 4, a 5-fold cross validation strategy was applied to each configuration (in this case, over 1,000 different configurations of hyperparameters were tested). For each model configuration, the accuracy was defined by averaging the Area Under the Curve (AUC) of the 5 test samples defined in each model run;

[12] Although ML methodologies would also be applicable to the management of missing values, the latter were managed through traditional methodologies.

- performance evaluation: all alternative implementations within a class were assessed on the basis of different performance measures, including AUC, Gini index, confusion matrix, model reactivity, etc. (the average result for each cross-validation iteration was considered). Table 5 provides an illustrative example based on data for natural persons. Neural networks ("NNs") show the best performance (Gini index above 80%) and are the richest model in terms of the number of features and statistical units. However, random forests yield a higher TPCR (True Positive Classification Rate) while allowing greater model explainability (e.g., due to a lower number of variables).

Table 5 – Example of performance comparison for different model classes for natural persons

		GINI	TPCR	TNCR	Accuracy
RF	Development	75.2%	71.3%	**87.6%**	79.4%
	OOT	73.1%	72.0%	**85.0%**	78.5%
XGB	Development	74.6%	90.1%	66.5%	78.3%
	OOT	74.9%	89.9%	64.4%	77.1%
NNs	Development	**84.1%**	85.6%	83.3%	**84.4%**
	OOT	**82.3%**	84.8%	82.5%	**83.7%**

All the inputs to an ML model should be individually interpretable, as well as their impact on the target variable. One should, in principle, be able to explain all the determinants of each individual forecast. To achieve such a goal, both global and local interpretability techniques were used.
As concerns the former, feature importance analysis was used, generating visual outputs like in the example provided in Figure 7. It is worth noting that variables 6 and 25 would have been discarded by traditional short-listing methodologies, since their "information value" is below 0.01[13]: unlike linear models, ML models seem to capture relationships that are weak but relevant.

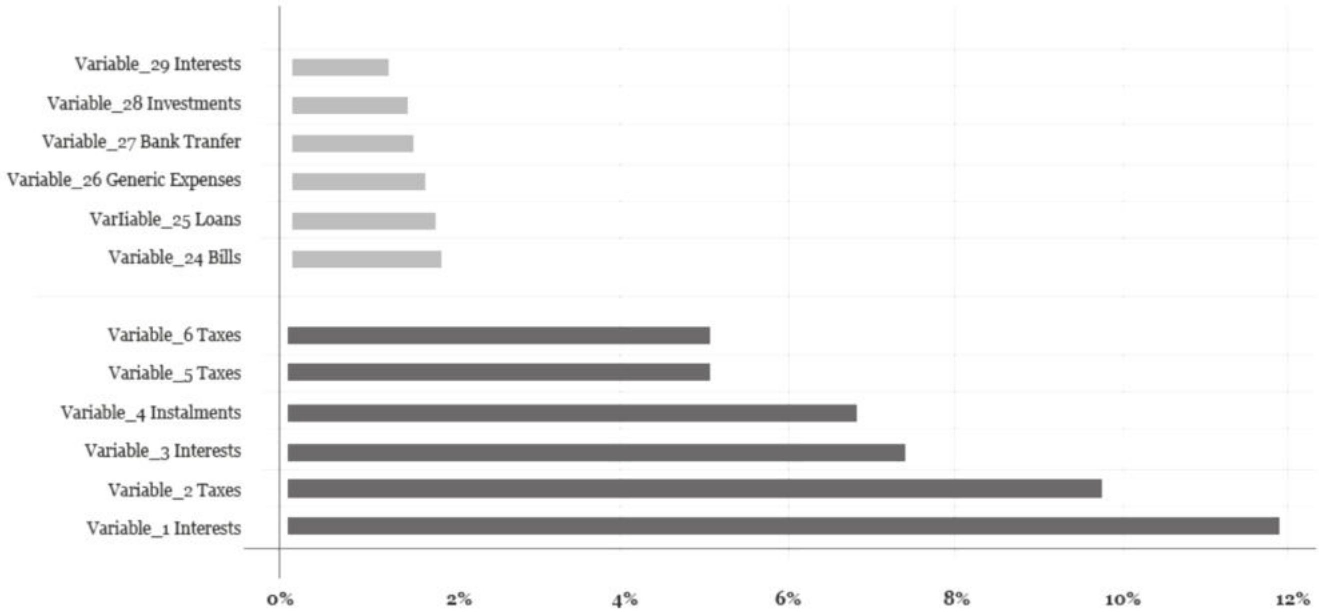


Figure 7 – Sample feature importance chart

[13] See, for example, (Zdravevski et al., 2011) for an introduction to information value and weight of evidence.

As for local interpretability techniques, Shapley values were used to identify variables with a stronger impact on PDs; LIME and OptiLIME-CRIF were also tested: both approaches led to very similar results.
The main benefits of the project can be summarised as follows:

- on the one hand, ML-based models increase accuracy. While such an increase is not dramatic (due to the fact that the bank already had in place a thoroughly tested system), it leads to a greater ability to classify high-risk customers, allowing pre-emptive risk management. On average, transaction data modules help increase overall performance, in terms of Gini index, by 6 percentage points, for both individuals and businesses;
- on the other hand, the bank has acquired a categorisation engine trained on proprietary data (and producing a reliability measure of each generated output), as well as a development lab to create, tune, test and deploy ML-based models.

While ML methodologies can reach outcomes that would otherwise be unattainable, they also raise additional challenges related to data management (i.e. the treatment of missing values and the choice of appropriate short-listing techniques), the governance of the overall modelling process and the interpretability of the final model.
In terms of *data management*, in order to skim the list of candidate variables (over 20,000) and make computations easier and faster, the following combination of traditional and ML steps was used:

- preliminary reduction: variables were excluded whenever they showed a high percentage of missing values, a high concentration of values on few levels or an extremely high correlation with other features;
- generation of new variables starting from those that had passed the first selection step (e.g. by taking the min, max, mean, trends, the standard deviation, etc.), leading to a new "long list" of about 2,000 variables;
- creation of a short list by identifying and removing the features that were not relevant enough in a multivariate model (low feature importance).

Careful feature selection (e.g. using multiple feature selection methods to avoid discarding relevant variables[14]) led to multiple advantages in terms of lower training time, lower risk of overfitting (as less redundant data means less noise, making biased decisions less likely) and greater performance.

Regarding *model development*, the hyperparameters' definition was optimised by means of a random search algorithm (testing several randomly selected combinations of hyperparameters) and cross-validation. Random search, as opposed to grid search, proved paramount in making the project manageable in terms of training time and costs[15].
Finally, as far as model *explainablility* is concerned, human oversight and interpretability were a major concern throughout the project, as they are key to gaining the trust of all stakeholders, ranging from internal functions to supervisors. Global and local interpretability techniques have helped address this goal in a satisfactory way.
Despite the presence of a well-established body of literature on machine learning, the project was carried out using a mixed approach, balancing the computational power of ML and the need for human expert control throughout the estimation process. For instance, missing value management and short-listing also leveraged traditional techniques, thus reducing the risk of counterintuitive solutions generated by ML alone.
The risk of relying on weak outcomes (which may perform badly on new data) was controlled through diversification: in both the modelling and the interpretability steps, several approaches were used, challenging one another. In the modelling step, three different classes were tested in order to select the most performing one (while understanding the power of each one); in the interpretability step, again, different approaches with a varying degree of complexity were adopted. The ultimate result is a compromise between total flexibility and the need to ensure an appropriate level of understanding by stakeholders.

5. Final remarks

ML is still expanding strongly in terms of methodological refinements and innovative applications but can be regarded as a well-established technology when it comes to its key characteristics, weaknesses, and strengths. Such a consensus, among researchers and practitioners, on what ML can and cannot do, should pave the way for a set of standards to be followed in the development, validation and supervision of ML-based models in banks.
ML is not a homogeneous body of results: indeed, it is a wide-ranging label used to encompass a set of techniques that are extremely diverse and should not be treated equally. This is especially true when it comes to interpretability and the risk of creating "black boxes" that are deployed without an appropriate level of awareness and oversight. While deep learning approaches are undoubtedly

[14] For instance, for natural persons, both random forest and gradient boosting-based selections were applied to discard 80% of variables (as opposed to discarding 85% by using random forest alone, and 90% if one were to look only at gradient boosting).
[15] Alternatives to random search include manual search (where several combinations of parameters are defined based on human judgement) and grid search (where different combinations of hyperparameters are chosen on a value grid, and parameters are then optimally chosen in a *neighbourhood*). While grid search requires the testing of every possible combination of hyperparameters, random search allows calibration of the number of search iterations on the basis of time/resource constraints. According to (Ribeiro et al., 2016), if the close-to-optimal region of hyperparameters occupies at least 5% of the grid surface, then random search with a fixed number of trials is highly likely to find it.

prone to such a risk, techniques like those used in our case studies (including, for example, decision trees, random forests, XGB) have reached full maturity and allow model developers to deal with transparency and avoid overfitting.
When it comes to transparency, however, the global and local interpretability techniques shown in our case histories should not be seen as a target, but rather as a means to facilitate the dialogue with model users. A continuous interaction with stakeholders (including the bank's business-oriented functions, its middle management and board of directors) is key to ensuring that all implications of a new algorithm are fully understood before it becomes part of an institution's risk management toolbox. Pilots, dashboards, and "explainers" should not be seen as mere sweeteners, given out to smoothen model acceptance, but rather as a fundamental step in model development, a recipe for greater robustness and a source of mutual enrichment for model engineers and users.
Finally, it should be borne in mind that slowing down innovation is not an option. While new ML-based models clearly involve risks and weaknesses that must be carefully addressed, inaction has its own costs and dangers. Discouraging the use of innovative models and data sources may result in banks competing with non-bank entities with one arm tied behind their back; increasing the gap between the models used for internal risk management purposes and those validated under the IRB approach may undermine the latter's credibility and, in the long run, prove detrimental to effective bank supervision.

References

- Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research 13, 281-305.
- Bholat, D., Gharbawi, M., Thew, O., 2020. The impact of Covid on machine learning and data science in UK banking. Bank of England Quarterly Bulletin.
- Breeden, J., 2021. A survey of machine learning in credit risk. JCR. https://doi.org/10.21314/JCR.2021.008
- Di Biasi, P., Gnutti, R., Resti, A., Vergari, D., Basile, A., Bernabei, F., Caprara, C., Cavarero, D., Marigliano, M., Ranaldi, R., Vignolo, M., 2022. Machine Learning for Credit Risk Management and IRB Models: Lessons from success Case Histories (A joint paper by Intesa Sanpaolo and CRIF). Intesa Sanpaolo - CRIF, Milano-Bologna.
- European Banking Authority, 2020. EBA Report on Big Data and Advanced Analytics (No. EBA/REP/2020/01). European Banking Authority, Paris.
- Giudici, P., Raffinetti, E., 2021. Shapley-Lorenz eXplainable Artificial Intelligence. Expert Systems with Applications 167, 114104. https://doi.org/10.1016/j.eswa.2020.114104
- Guerra, P., Castelli, M., 2021. Machine Learning Applied to Banking Supervision a Literature Review. Risks 9, 136. https://doi.org/10.3390/risks9070136
- Huang, J., Chai, J., Cho, S., 2020. Deep learning in finance and banking: A literature review and classification. Front. Bus. Res. China 14, 13. https://doi.org/10.1186/s11782-020-00082-6
- Institute of International Finance, 2019. Machine Learning in Credit Risk (Summary report). Washington D.C.
- Leo, M., Sharma, S., Maddulety, K., 2019. Machine Learning in Banking Risk Management: A Literature Review. Risks 7. https://doi.org/10.3390/risks7010029
- Molnar, C., 2019. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier.
- Rundo, F., Trenta, F., di Stallo, A.L., Battiato, S., 2019. Machine Learning for Quantitative Finance Applications: A Survey. Applied Sciences 9. https://doi.org/10.3390/app9245574
- Visani, G., Bagli, E., Chesani, F., 2020. OptiLIME: Optimized LIME Explanations for Diagnostic Computer Algorithms.
- Zdravevski, E., Lameski, P., Kulakov, A., 2011. Weight of evidence as a tool for attribute transformation in the preprocessing stage of supervised learning algorithms, in: The 2011 International Joint Conference on Neural Networks. pp. 181-188. https://doi.org/10.1109/IJCNN.2011.6033219

Beyond VaR and Expected Shortfall: The Stress Testing/Scenario Analysis approach for protecting the investors in the post-Covid19 era.

Gianluca Macchia (CIIA®, Independent Investment & Risk Analyst)

Abstract

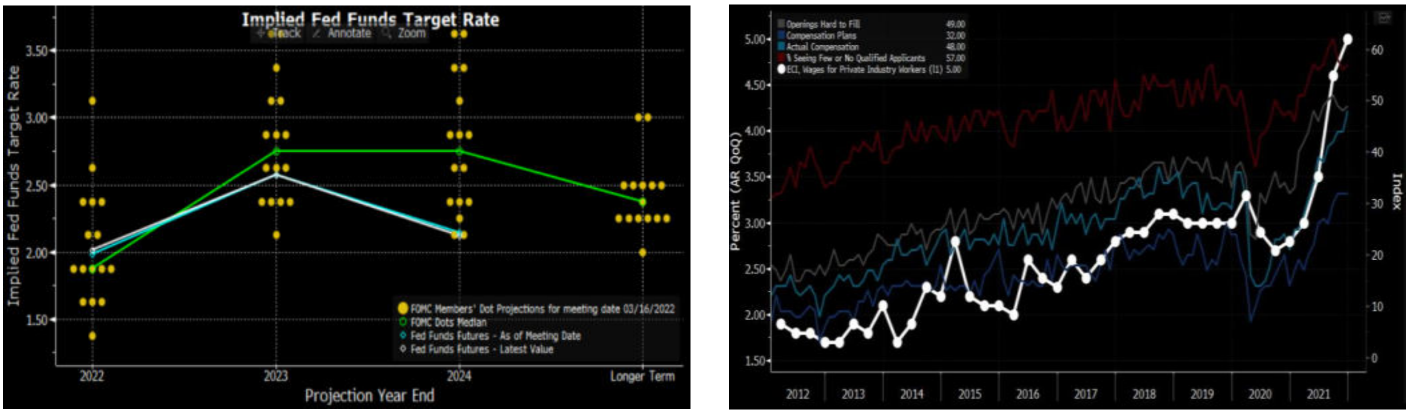
With political and economic scenarios changing at an ever faster pace, it is necessary to understand the potential effects on asset prices. Today, the topic of rising inflation in the US as well as in the Eurozone, although still considered temporary by central banks, confronts us with the "unexpected risk" of a deviation from the baseline scenario. This implies the risk of having an aggressive monetary policy in the US, in a restrictive direction, therefore harmful to the financial markets. In this context, the question arises: is it possible to contemplate these events beforehand and act in good time? The answer is Yes and good risk management practices are important, using stress testing / scenario analysis techniques to accompany risk measures such as VaR and Expected Shortfall. Implementing this concept, through the implementation of stress test / scenario analysis - Bloomberg Economics Forecast Models® and Bloomberg Factor Models® - the present work seeks to consider plausible adverse scenarios that may arise and to assess the related impacts in terms of portfolio. The final aim is to improve the information set for the investor, allowing him to avoid potential market falls, as far as possible, that could prevent him from achieving his investment objectives.

1. Introduzione

Spinta da una policy fiscale e monetaria eccezionalmente accomodante ed aiutata certamente da un contesto economico e geopolitico favorevole a livello internazionale, l'inflazione è il tema che in parte possiamo definire il "rischio inatteso" che potrebbe vanificare, almeno in parte, una crescita economica importante ai fini di una sostenibilità del debito governativo.
Lato risk management, ciò deve ricordare l'importanza di affiancare alle misure di VaR ed Expected Shortfall, misure di rischio come lo stess testing e di scenario analysis riuscendo nell'intento d'integrare con logiche forward looking un atteggiamento verso il rischio che altrimenti sarebbe legato al comportamento passato, troppo distante da quello odierno.
La creazione di scenari avversi è fondamentale per comprendere l'impatto sul portafoglio di determinati eventi migliorando il set informativo verso l'investitore che potrà tenerne conto all'interno delle logiche di costruzione del portafoglio d'investimento ed al tempo stesso delle strategie di gestione del rischio.
Ciò, attraverso l'implementazione di due modelli, Bloomberg Economics Forecast Models® e Bloomberg Factor Models®, cerca di considerare plausibili scenari avversi che possano manifestarsi e cosa essi comportino a livello macroeconomico prima e dei mercati finanziari poi per le diverse esposizioni, mostrando inevitabilmente delle differenze rispetto alle metodologie VaR ed ES.
Tale attività deve rappresentare un presidio irrinunciabile in un contesto di wealth management essendo parte di un processo d'investimento avente come finalità unica quella di permettere il raggiungimento degli obiettivi finanziari di ogni investitore migliorandone, al tempo stesso, la protezione da eventi inattesi.
Il contributo proposto cerca di comprendere come poter implementare schemi operativi di risk management nel mondo Wealth Management, guardando sia ad aspetti operativi che di risk reporting.
Obiettivo del presente lavoro è quello di comprendere come taluni scenari avversi, appositamente studiati, possano costituire delle deviazioni più o meno marcate da uno scenario economico baseline e come ciò possa riflettersi sul valore degli assets detenuti in portafoglio e quindi possano migliorare il set informativo a disposizione dell'investitore.

2. Analisi del contesto macroeconomico

Gli effetti della guerra in Ucraina stanno producendo una situazione di maggior volatilità specialmente sui mercati europei rispetto a quelli americani che hanno recentemente toccato livelli rispettivamente di 44,29 per il VDax e di 36,45 per il Vix Us. Risulta essere estremamente interessante osservare il decoupling delle monetary policy tra la Fed e la Bce con la prima che conferma il percorso di tightening che dovrebbe portare i tassi a fine 2022 ad un valore dell'1,75%, spinta da un'inflazione che nel breve periodo non accenna a ridursi anche a causa di tensioni geopolitiche - 7,9% la rilevazione a Febbraio 2022 del CPI (6.4% per l'inflazione core) – e da un miglioramento progressivo della domanda sostenuta dal mercato del lavoro e dall'eliminazione progressiva delle restrizioni Covid-19. Lato Bce, con il recente update relativamente alle previsioni di crescita dell'Eurozona che vede la crescita ridursi dello 0,5% nel 2022, trainata da una riduzione della consumer spending dell'1,3% ed un'inflazione che si porta nel 2022 al 5,1% dal 3,2%, va avanti il programma di tapering – che vede confermato la sua fine a giugno - assieme ad una maggiore flessibilità in termini di aumento dei tassi con il mercato che stima attualmente un incremento di 40 bps per il 2022. La Fed, dal canto suo, vede nell'inflazione il problema macroeconomico più rilevante.
Tuttavia, le recenti tensioni geopolitiche hanno indotto il chairman Powell se non altro a posticipare una strategia "shock and awe" avendo alzato a marzo di soli 25 bps rispetto ai 50 bps attesi dal mercato. Ad ogni modo, la flessibilità della Fed anche in considerazione dei dots di marzo 2022 – con un policy rate target fissato a non meno del 2.5% nel 2023, potrebbe essere accantonata laddove i dati dell'inflazione, come inizierebbe a mostrare l'Employment Cost Index (+4.2%) e la sua versione relativa ai lavoratori dell'industria privata (+5%), mostrassero una persistenza del fenomeno.
Importante sarà la pubblicazione ad inizio aprile del documento della riunione del FOMC in cui saranno esplicitate le previsioni economiche dalle quali sarà possibile comprendere quanto la Fed potrebbe trovarsi nella situazione di dover ricorrere ad una maggiore stretta monetaria.



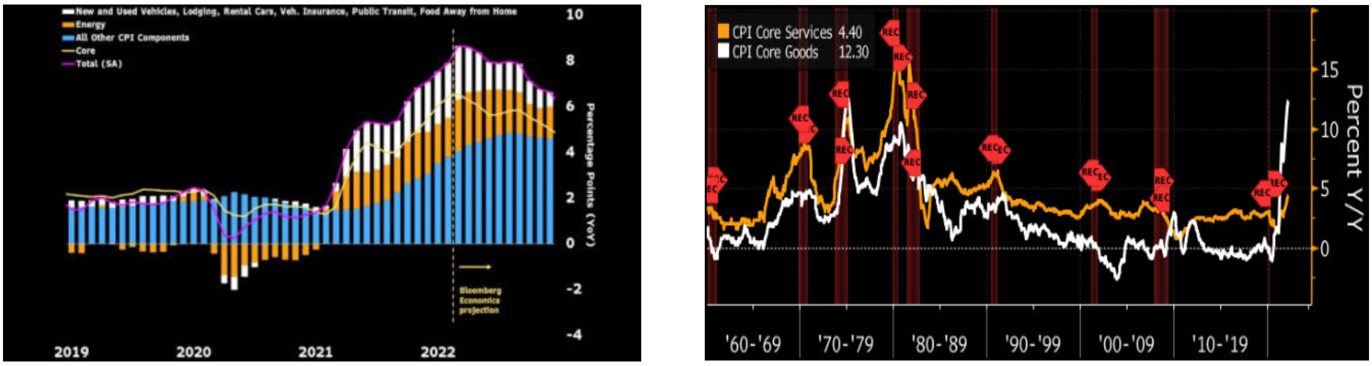
Fonte: Bloomberg LP

Fonte: BLS, NFIB

Fig. 1, Dots FOMC e Employment Cost Index (ECI)

Con una stima del CPI per il 4° trimestre 2022 del 6,8% YoY, in moderato calo dopo il picco dell'8,% registrato durante la primavera, le preoccupazioni della Fed sono legate 1) ai riflessi dei maggiori costi energetici provocati dalla guerra in Ucraina, 2) ad una più elevata inflazione dei servizi, 3) ai colli di bottiglia delle catene di approvvigionamento che non solo non si risolvono positivamente ma che tendono a peggiorare a causa di lockdown locali nella provincia del Guangdong.

Con la pubblicazione da parte del FOMC delle previsioni economiche sarà possibile capire quanto la Fed sia dietro la curva e quindi la politica monetaria possa, in virtù dei temi fin qui espressi, ad un'eventuale strategia "shock and awe" che nella riunione del 16 marzo 2022 è stata probabilmente solo rinviata.



Fonte: Bloomberg Economics, BLS

Fonte: BLS

Fig. 2, Trend e composizione del CPI 2022

Lato mercati continuano le difficoltà lato growth factor, assieme ad un'incertezza che si protrae lato value factor, così come continuano a perdere terreno i prezzi da parte delle scadenze a medio/lungo termine sulle principali yield curve governative.
Il tutto mentre materie prime, industriali e agricole in testa, continuano a scambiare su prezzi molto elevati.

Ciò detto, la preoccupazione degli investitori è in riferimento agli effetti che tale situazione potrà avere nel medio termine sulla crescita, al fine di valutare la possibilità di prepararsi a scenari alternativi rispetto a quelli di ripresa pre-guerra.
Il presente lavoro ha quindi come obiettivo, attraverso la creazione di un adverse scenario, di comprenderne l'impatto sul portafoglio e, di riflesso, come sia opportuno gestire non solo la fase attuale ma quella successiva nei diversi scenari al fine di evitare perdite rilevanti nel corso del 2022.

In termini operativi, si utilizzano i seguenti modelli: Bloomberg Economics Forecast Models® e Bloomberg Factor Models®, e per quest'ultimo il modello Bloomberg MAC2®, aventi 1) la finalità di effettuare un'analisi delle deviazioni dal baseline scenario in termini di GDP growth, inflazione e risposta di politica monetaria a causa di scenari sfavorevoli attraverso stress testing dei relativi drivers, 2) analizzare tali risultati in termini d'impatto sui portafogli.

3. Caratteristiche dei portafogli d'investimento

L'analisi dei portafogli, cosi come fotografata alla data del 18 marzo 2022, mostra come portafogli d'investimento che presentano profili di rischio non troppo dissimili possano avere reazioni differenti davanti a scenari avversi grazie ad una struttura strategica che vede il primo – Unconstrained Risk Control Portfolio - avere una componente rilevante in strategie liquid alternative rispetto ad un secondo – Moderate Risk Portfolio - che vede solo un esposizione direzionale alle diverse asset class.

CODICE BLOOMBERG	STRATEGIA	PESI (%)
BLEAI2E LX Equity	LONG SHORT EUROPE	4
BNSEBIA LX Equity	MONETARIO	7
BUBVD2E LX Equity	US VALUE	14
DITPDFC LX Equity	EUROPE DIVIDEND	11,5
DNCABIE LX Equity	CREDIT ABS RETURN	1,5
ECN FP Equity	EUROPE VALUE ESG LOW CARBON	7
HEUAPPI LX Equity	MULTI STRATEGY	3,5
JPGMSIA LX Equity	MULTI STRATEGY	9
MGGMECA LX Equity	GLOBAL MACRO	4
NOMAPEE LX Equity	MULTI STRATEGY	9
UBFIRII LX Equity	BOND CHINA	24
UDVD LN Equity	US DIVIDEND ARISTOCRATS	3
UQLTD IM Equity	US QUALITY	2,5

Tabella 1, Composizione Unconstrained Risk Control Portfolio, Dati al 18 Marzo 2022. Fonte Bloomberg L.P®

CODICE BLOOMBERG	STRATEGIA	PESI (%)
AXAEIGE LX	AXA WF-FRM EUROPE-E	25
AXWHACU LX	AXA WORLD-GLBL H/Y BON-ACUSD	3
AXEURCE LX	AXA WRLD FD-EUR SUS CR-ECEUR	23
FLEFPCC LX	JPM-PACIFIC EQTY-D USD ACC	4
LOAVPAU LX	LO FDS-ASIA VALUE BOND-USDPA	3
NNAACEE LX	NORDEA I SIC-NA STR EQ-EEUR	12,5
RATTRXU LX	RAM LX TACT II GLBL BNDTR-XN	8
VGREMEB LX	VF-MTX SUST EM MK L-BUSD	9
VONUVC2 LX	VONTOBEL-US EQUITY - C	12,5

Tabella 2, Composizione Moderate Risk Portfolio, Dati al 18 Marzo 2022. Fonte Bloomberg L.P®

Unconstrained Risk Control Portfolio*:* portafoglio che mostra una volatilità annualizzata del 7,90% ed un VaR ed Expected Shortfall trimestrali calcolati in modalità Monte Carlo rispettivamente pari a 6,28 e 8,41. In termini di composizione il portafoglio presenta un peso dell'equity del 60,15% mentre la componente fixed income e quella alternative sono rispettivamente al 31% ed al 24%. Pari al 7% il livello della liquidità tattica.

La contribuzione al rischio lorda da parte dell'equity è pari al 82,27% in cui il peso degli US è il 50% dovuto ad un'esposizione allo US Dividend e US Value mentre si ha una posizione netta negativa su US Growth. Europa ed Asia pesano rispettivamente per il 19% ed il 12%.

In termini settoriali il portafoglio si caratterizza per avere una posizione lunga su energy assieme ai settori pro-ciclici così come porta avanti una posizione negativa netta su consumer staples, health care ed utilities. Lato fixed income il portafoglio è esposto alla slope yield curve europea.

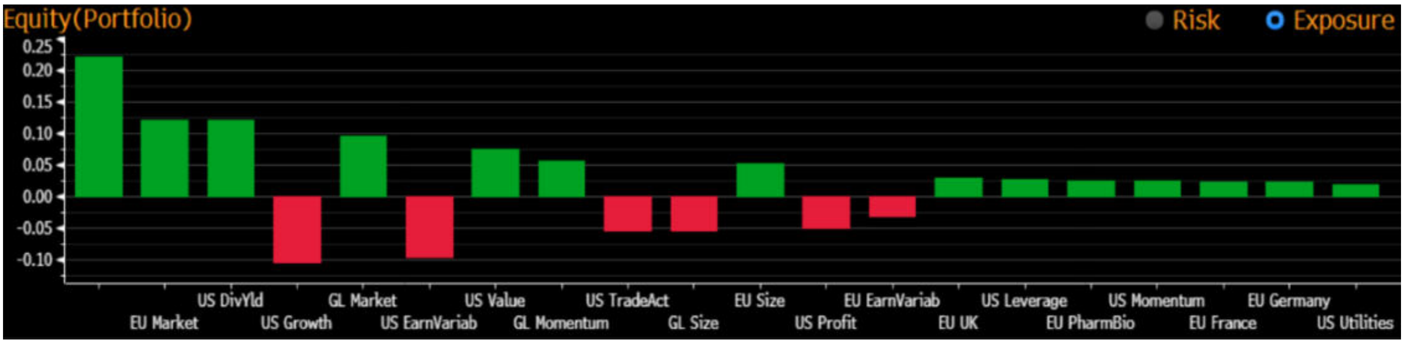


Fig. 3, Unconstrained Risk Control Portfolio Equity Contribution Risk. Fonte: Bloomberg L.P®

Moderate Risk Portfolio: In termini di composizione il portafoglio presenta un peso dell'equity del 60,15% mentre la componente fixed income è pari ad un 37,85% mentre a livello di rischio il portafoglio mostra una volatilità annualizzata del 9,09% ed un VaR ed Expected Shortfall rispettivamente pari a 6,73 e 9,13.

- *Equity*: Oltre ad un dividend yield del 2,29%, la componente equity mostra, a livello geografico, un'esposizione maggiormente focalizzata sull'Europa (33,53%) rispetto agli US al 29,51% mentre l'Asia si ferma ad un 10,94%. A livello settoriale accanto ai settori pro-ciclici come financials (14,83%) e consumer discretionary (10,53%) sono mantenuti pesi rilevanti su tecnologia (21,33%) e consumer staples (10,30%). In termini fattoriali viene confermata la struttura di portafoglio descritta.
- *Bond & Credit*: portafoglio esposto sostanzialmente al credito europeo – banche – ed in misura inferiore a quello US mentre a livello governativo l'esposizione all'Eurozona – focalizzata per lo più sulla slope ed a livello paese sull'Italia - è accompagnata da quella in Asia.

Fig. 4, Moderate Risk Portfolio Equity Contribution Risk Fonte: Bloomberg L.P®

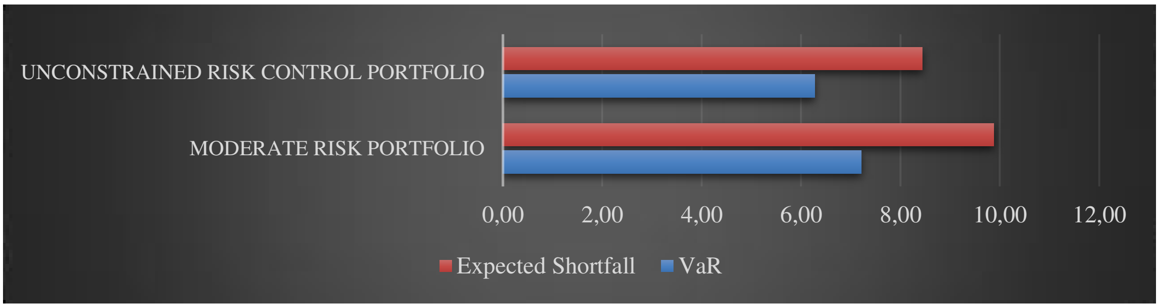


Fig. 5, VaR ed Expected Shortfall pre-scenario "Inflation Risk + Geo Risk" Fonte: Dati Bloomberg L.P®

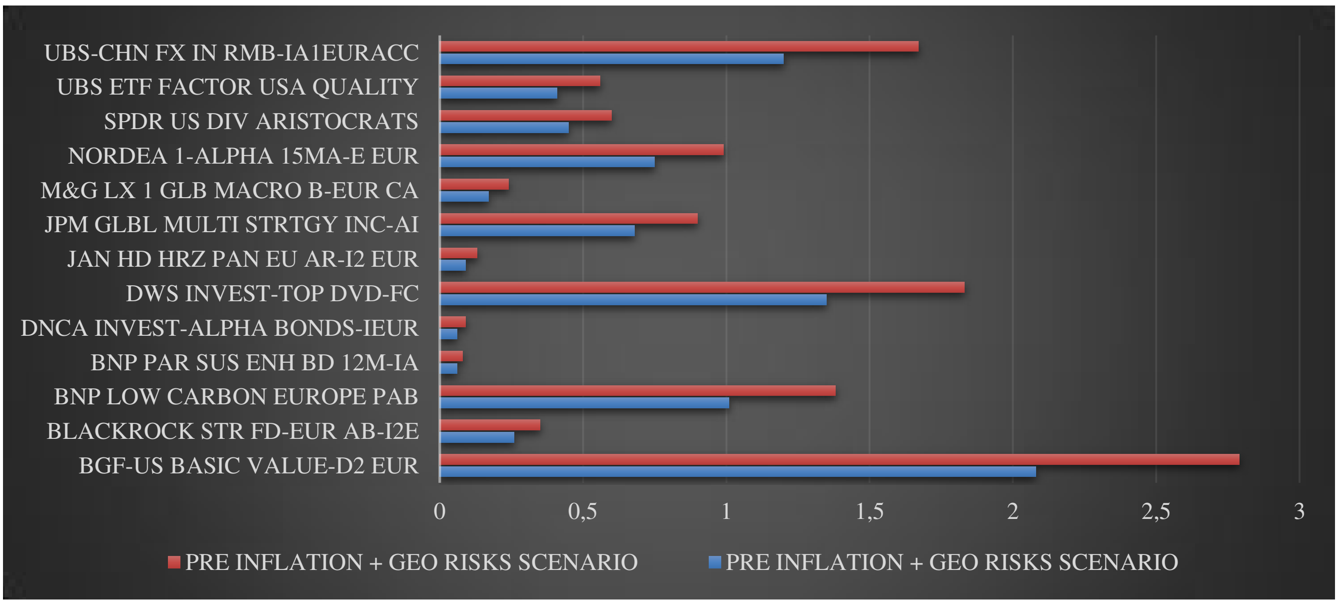


Fig 6 Unconstrained Risk Control Portfolio Contribution Var ed Expected Shortfall Fonte: Dati Bloomberg LP®

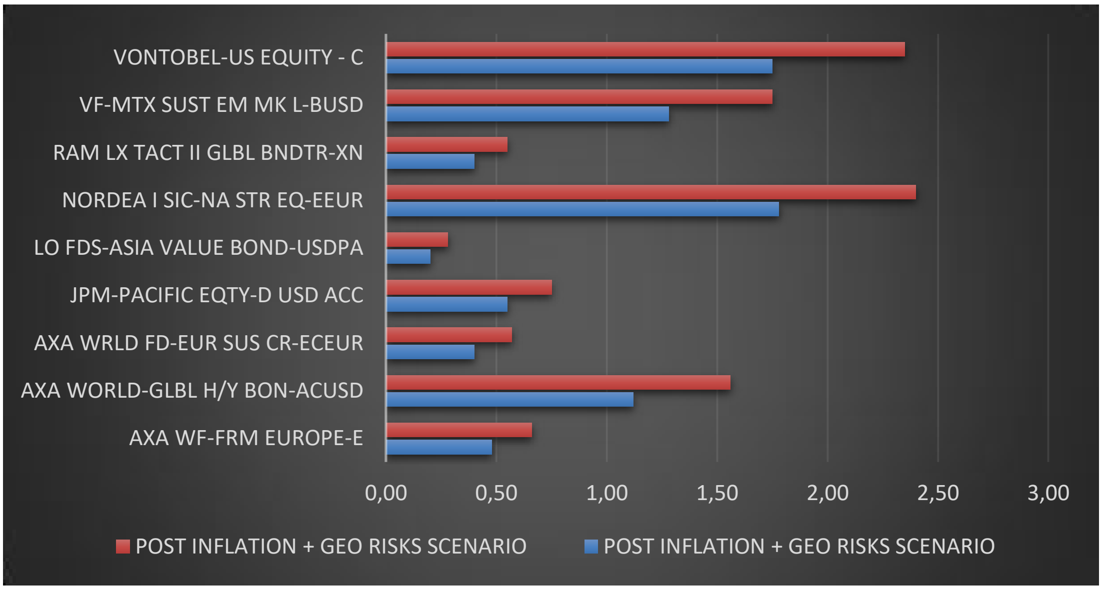
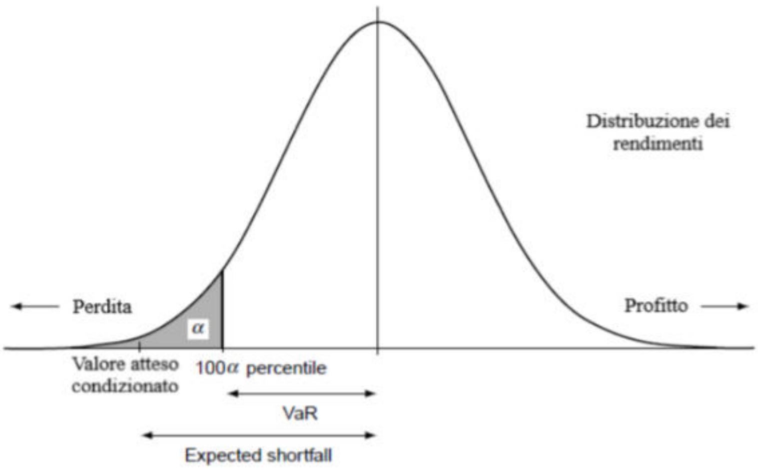


Fig 7, Moderate Risk Portfolio Contribution VaR & Expected Shortfall Fonte: Dati Bloomberg L.P®

4. Breve review delle misure di rischio

L'attività di asset management è indissolubilmente legata ad un'esposizione al rischio. La volatilità è certamente la misura di rischio più popolare, tuttavia, il suo problema principale è dettato dal fatto che non si preoccupa della direzione del prezzo. Altre due sono le misure di rischio particolarmente conosciute: il Valute-at-Risk (VaR) e l'Expected Shortfall (ES). L'importanza di tali misure nel portfolio management è legato alle loro caratteristiche teoriche ossia al fatto di essere misure coerenti di rischio. Se il VaR risponde affermativamente alle tre proprietà su quattro (i.e. monotonicità, traslazione in varianza ed omogeneità positiva), l'Expected Shortfall permette di soddisfare anche la subadditività permettendo al portafoglio di poter esprimere un rischio inferiore alla somma di due portafogli.
La proprietà della subadditività prima richiamata permette all'ES di mostrare, a parità di holding period e del livello di confidenza, il suo livello di convessità. Ciò vuol dire che l'Expected Shortfall dimostra di essere maggiormente appropriato rispetto al VaR nella risoluzione problemi di ottimizzazione del portafoglio.
L' Expected Shortfall è simile al VaR in quanto entrambi forniscono una misura di rischio comune attraverso diverse posizioni. ES può essere implementata nel determinare il calcolo delle perdite allo stesso modo in cui il VaR è implementato come misura del rischio, ed entrambi tenere adeguatamente conto delle correlazioni.



Fonte: Hull, John C. "*Risk Management and Financial Institution – 5th Edition*" Wiley (2018)

Fig. 8, VaR ed Expected Shortfall

Come vedremo successivamente, a parte il caso dei modelli proprietari, diverse sono le modalità con cui è possibile calcolare il VaR e l'Expected Shortfall: il metodo della simulazione storica, il modello parametrico ed il metodo basato simulazione Montecarlo.
Per quanto attiene alla prima possiamo affermare che rappresenta certamente la modalità più semplice da applicare in quanto considera la volatilità storica. All'interno di tale tipologia è importante l'approccio del Delta-Normal Approach secondo la quale VaR ed ES sono determinati come segue:

$$VaR = \mu_p + z\sigma_p \quad [1] \qquad ES = \mu_p + \sigma_p \frac{e^{-\frac{z^2}{2}}}{(1-x)\sqrt{2\pi}} \quad [2]$$

Dove x è il livello di confidenza, z è il punto sulla distribuzione normale avente un probabiltà x di essere superata. I limiti del delta-normal approach si manifestano quando sono presenti nel portafoglio prodotti non lineari nel loro payoff (come nel caso dei certificates). Ciò porta ad utilizzare la forma quadratica che permette di considerare oltre al delta anche il gamma:

$$\Delta P = \delta \Delta S + \frac{1}{2}\gamma \Delta S^2 \quad [3]$$

Argomento fondamentale è poi rappresento dalla stima della volatilità. I metodi più comuni per il calcolo sono quelli della simulazione storica e della volatilità implicita. All'interno della prima tipologia si annoverano l'Exponential Weighted Moving Average approach (EWMA) ed il GARCH (1,1). Nel primo caso si considerano le osservazioni via via più lontane da quelle correnti assegnandogli un peso più contenuto applicando una legge esponenziale. Utilizzando tale approccio, la varianza condizionale viene stimata utilizzando la seguente formula:

$$\sigma_n^2 = w_0 r_{n-1}^2 + w_0 \lambda r_{n-2}^2 + w_0 \lambda^2 r_{n-3}^3 + \cdots w_0 \lambda^t r_n^m \quad [4]$$

Dove n rappresenta il numero delle osservazioni utilizzate per stimare la volatilità.
Tale modello viene semplificato rispetto a due periodi di dati (n-1 ed n-2). Il modello in questione supera due problematiche: 1) se il numero delle osservazioni è molto elevata, ma la volatilità è ciclica, il modello potrebbe sovrastimare la volatilità; 2) se i rendimenti sono molto positivi o molto negativi per un lungo periodo passato potrebbe influire eccessivamente sul modello.
Argomento centrale in tale modello è la stima del decay factor (λ) che viene comunemente impostato ad un livello di 0,94 per i rendimenti giornalieri e 0,97 per quelli mensili così come utilizzati da parte di JP Morgan nel modello CreditMetrics. Alti valori di λ potranno minimizzare l'effetto dei rendimenti percentuali giornalieri, mentre i valori bassi di λ tenderanno ad incrementare l'effetto dei rendimenti percentuali giornalieri sulla stima della volatilità corrente. Il grande vantaggio di tale metodo è rappresentato dal fatto di richiedere pochi dati per stimare la volatilità.
Il secondo, il modello Garch (1,1), è utilizzato dagli analisti per predire la time-varying volatility. Un modello Garch (1,1) non solo incorpora stime più recenti della varianza e del rendimento al quadrato, ma considera anche una variabile che tiene conto del livello di lungo periodo della varianza. Il termine (1,1) si riferisce al peso attribuito al rendimento al quadrato ed a quello da attribuire alla varianza. Il miglior modo per descrivere un modello Garch (1,1) è quello di osservarne la formula che ne rappresenta la modalità di determinazione della varianza:

$$\sigma_n^2 = \omega + \alpha r_{n-1}^2 + \beta \sigma_{n-1}^2 \quad [5]$$

Dove:

- α *ponderazione del rendimento del periodo precedente*
- β *ponderazione della volatilità della stima precedente*
- ω *ponderazione della varianza di lungo termine pari a* γV_L
- V_L *varianza media di lungo termine, ossia* $\frac{\omega}{1-\alpha-\beta}$

Si ricorda che l'EWMA non è altro che un caso speciale del GARCH (1,1) dove $\omega = 0$; $\alpha = 1 - \lambda$ e $\beta = \lambda$. Similmente all'EWMA, β rappresenta l'exponential decay rate. Rispetto a quest'ultimo, il modello GARCH (1,1) aggiunge ulteriori informazioni assegnando anche una ponderazione alla stima della varianza media di lungo periodo. Un'ulteriore caratteristica della stima GARCH (1,1) è l'assunto implicito che la varianza tende a convergere ad un livello medio di lungo periodo, caratteristica mean-reverting quantomai importante nell'effettuare il pricing di titoli derivati come le opzioni scritte su commodities e tassi d'interesse.
Seconda modalità di determinazione del VaR e dell'Expected Shortfall è rappresentato dall'*approccio parametrico*. Tale metodo è basato sulla matrice varianze/covarianze[1] dove s'ipotizza che tutti i fattori di mercato seguano una distribuzione normale. Benché tale metodo sia particolarmente utilizzato, i suoi limiti sono: 1) non considerare le fat tails e la skewness, in particolare quella negativa, 2) assumere una dipendenza lineare dai risk factors. Relativamente alla stima della volatilità, il metodo parametrico assume una correlazione costante per i due drivers.
Per quest'ultime si rende altresì opportuno utilizzare "versioni" che siano meno legate ad eventi passati, tendendo sempre più verso soluzioni forward-looking. In ultima analisi, ciò è rappresentato da misure di stress test basati su scenari avversi debitamente modellati che cercano di comprendere, stante relazioni fattoriali tra i diversi assets, quale possa essere l'impatto sul portafoglio.
Ciò detto, si effettua una breve review delle misure di rischio maggiormente utilizzate puntualizzando in particolare sui punti di forza e di debolezza di ognuna ed identificando le best practices necessarie da implementare.
L'approccio Monte Carlo, utilizzato per modellizzare problemi complessi o per stimare variabili quando il campione è troppo ridotto, vede il suo grado di affidabilità legato al concetto del sampling error quantificato dall'errore standard del vero valore atteso. Tale valore è determinato stimando $\frac{s}{\sqrt{n}}$ e può essere ridotto:

- Aumentando il numero delle simulazioni (n) vedendo se al tempo stesso otteniamo una stabilizzazione della deviazione standard. In questo modo miglioriamo l'accuracy della simulazione;
- Utilizzare il metodo delle *variabili antitetiche* che permettono di avere una dispersione inferiore dell'output di simulazione in quanto le variabili generate sono correlate negativamente.

[1] Al fine di poter avere una rappresentazione adeguata del rischio è fondamentale utilizzare i risk factors

- Utilizzare la *variabile di controllo*, che rappresenta una metodologia largamente utilizzata per ridurre il sampling error nella simulazione Monte Carlo. Questo implica di rimpiazzare la variabile x che presenta proprietà non conosciute con una variabile y che invece presenta proprietà conosciute. La nuova variabile che si genera, $x^* = y + (\hat{x} - \hat{y})$, avrà un sampling error più piccolo di quello iniziale a patto che la statistica di controllo e la statistica d'interesse risultino essere altamente correlate. Questo porta ad avere un risultato della simulazione Monte Carlo migliore in quanto l'utilizzo della variabile di controllo permette una minor varianza rispetto alla situazione di partenza

All'interno di tale "famiglia" di tecniche annoveriamo il metodo Bootstrapping che utilizza dati storici invece di random data da una distribuzione di probabilità evitando di modellizzare direttamente i dati osservati così come non ne effettua alcuna assunzione sulla distribuzione dei dati.
Un vantaggio dell'approccio Monte Carlo è che indirizza più fattori di rischio assumendo una distribuzione sottostante e modellando le correlazioni tra i risk factors. Per effettuare le simulazioni, il risk manager dovrà solamente fornire i parametri per la media e deviazione standard e presupporre una distribuzione normale dei payoff. Altro vantaggio significativo è che la simulazione Monte Carlo può presupporre qualsiasi tipo di distribuzione purché sia possibile determinare le correlazioni tra i fattori di rischio. Lo svantaggio principale è rappresentato dalla lentezza del processo dovuto ad aspetti computazionali. L'approccio Monte Carlo è in genere utilizzato per grandi portafogli e richiede tempo.
Come sappiamo la storia non si ripete mai esattamente e per questo che l'attività di anticipazione dei unknow risks è da ritenersi quantomai complessa. Per passare ad una logica risk forward looking è opportuno implementare un processo di stress testing e scenario analysis andando a lavorare su una serie di drivers economici e finanziari che possano costituirne un framework operativo. A questo punto, è fondamentale per creare scenari estremi ma plausibili è utile ispirarsi, a livello metodologico, a quanto presente in letteratura. A riguardo certamente la metodologia proposta a Kupiec (1999) nota come conditional stress testing dove si regrediscono le variabili periferiche rispetto a quelle core. In questo modo le previsioni possono essere incorporate all'interno dello stress test. Tecnica quest'ultima migliorata da Kim e Finger (2000) che ipotizzano un certo comportamento tra le variabili core e quelle non core nella loro "broken arrow" stress test. Ad ogni modo, una buona regola è quella di osservare costantemente l'ambiente economico al fine di comprenderne le eventuali situazioni di stress future che non sono considerate attualmente.

5. Metodologia utilizzata

La metodologia utilizzata nel presente studio prevede l'utilizzo di due modelli: il Bloomberg Economics Forecast Models® e Bloomberg Factor Models®. L'idea è quella di considerare l'impatto di scenari avversi sui diversi drivers economici e finanziari ottenendo una serie di output lato inflazione, risposta monetaria della banca centrale e GDP growth. Tali risultati sono fondamentali per poterne stimare gli impatti sulle diverse asset classes del portafoglio fornendo un miglior set informativo all'investitore. In altri termini si cerca di comprendere come deviazioni dal baseline scenario impattino potenzialmente e negativamente sulla ricchezza dell'investitore e possano mettere a rischio l'ottenimento dell'obiettivo precedentemente determinato.
Lato macroeconomico, i modelli utilizzati per esaminare shock inattesi sono inclusi nei Bloomberg Economics Forecast Models® appartenenti alla famiglia dei semi-structural general equilibrium models[2] stimati attraverso metodi bayesiani. Un tale approccio permette una buona mimica dei modelli utilizzati dalle rispettive banche centrali – FRB/US[3], BASE/EUROZONE[4] e COMPASS/UK[5] – permettendo di stimare l'impatto degli shock endogeni ed esogeni in relazione a grandezze quali GDP growth, inflazione e risposta di politica monetaria. Caratteristica fondamentale di tali modelli sono: 1) la plausibilità empirica; 2) la coerenza rispetto ai modelli utilizzati dalle banche centrali; 3) la comparabilità tra i diversi paesi; 4) la tracciabilità. L'approccio del modello è una versione intermedia tra un tradizionale neo-keynesiano ed uno che vede una minor presenza di restrizioni teoriche volendo dare maggiore importanza alle osservazioni.
Come è possibile osservare dalla [6] i parametri $\alpha_y \alpha_r \alpha_q \, e \, \alpha_f$ e f determinano rispettivamente il grado di comportamento retrospettivo, l'elasticità del tasso di interesse della domanda, la reattività della produzione ai cambiamenti del tasso di cambio e nel moltiplicatore dell'impatto fiscale. L'effetto delle variazioni dei prezzi del petrolio, incertezza e globale la domanda è data da $\alpha_o, \alpha_v \, e \, \alpha_{yg}$.

$$y_t = \alpha_y y_t + (1 - \alpha_y) y_{t+1_{|t}} - \alpha_r r_{t-j} - \alpha_q \Delta q_{t-j} + \alpha_f \Delta f_t - \alpha_o \Delta o_{t-j} - \alpha_v v_t + \alpha_{yg} y_t^{glo} + \varepsilon_t^y \quad [6]$$

Se si assume che tutti gli agenti economici agiscano in modo razionale e guardino al futuro, si registrerebbe una forte variazione dei prezzi al momento dello shock. Mentre alcuni settori potrebbero essere maggiormente reattivi all'aumento dei costi rispetto ad altri, il divario tra la rigidità generale del mercato del lavoro e la retribuzione è coerente con tali salti raramente osservati nella pratica a livello dell'intera economia.
Relativamente ai tassi d'interesse, il gap del tasso di interesse reale, r_t è dato dal tasso di finanziamento effettivo dell'intera economia, i_e – a sua volta determinato come somma tra il risk free rate, i, ed il credit spread (cs), meno l'inflazione attesa $(p_{t+1|t})$. L'evidenza empirica mostra come gli aggiustamenti avvengono più lentamente rispetto a quanto può essere previsto in un modello forward-looking senza la presenza di vincoli creditizi. Un'assunzione estremamente importante che governa la risposta dell'economia a shock di politica monetaria è il tasso d'interesse e l'elasticità della domanda. Nel modello quest'ultimo viene identificato come il tasso di rifinanziamento marginale medio reale per l'intera economia.

---

[2] Modelli progettati per l'implementazione di un'approfondita attività d'analisi da parte dei policy makers. Modelli che si distinguono dai modelli DSGE grazie alla capacità di alternare ipotesi alternative sulla formazione delle aspettative degli agenti economici
[3] F. Brayton, T. Laubach e D. Reifschneider "The FRB/US Model: A Tool for Macroeconomic Policy Analysis" Board of Governon of the Federal Reserve System Aprile 2014
[4] E. Angelini, N. Bokan, K. Christoffel, M. Ciccarelli, S. Zimic "Introducing ECB-BASE: The blueprint of the new ECB semi-structural model for the euro area" Working Paper Series n°2315, Settembre 2019
[5] S. Domit, F. Monti e A. Sokol "A Bayesian VAR benchmark for COMPASS" Staff Working Paper n°583 Bank of England, 22 Gennaio 2016

L'impatto di una policy fiscale meno accomodante è catturato dal moltiplicatore fiscale. Un taglio della government spending comporterebbe una riduzione del GDP inferiore ad un impatto 1:1 per effetto di una riduzione delle importazioni.
Considerando ciò, per l'Eurozona e per UK si fissa un valore di 0.6 che riflette la valutazione effettuata da chi ha generato il modello e la composizione media delle misure adottate, da un lato per il consolidamento delle finanze pubbliche e, dall'altro, per stimolare l'economia. Per gli US viene utilizzato un valore leggermente più elevato (0.8) che riflette un giudizio sulla variata funzione di reazione della Fed che incorpora un obiettivo d'inflazione sotto la media. Lo sforzo di bilancio viene catturato dalla variazione del livello del bilancio strutturale, f, che segue un processo autoregressivo del tipo:

$$f_t = \gamma_f f_{t-1} + \varepsilon_t^f \qquad [7]$$

Nel modello la variabile inflazione viene considera avere un'origine domestica e/o estera. Se si assume che siano le pressioni sul costo reale marginale a guidare l'andamento dei prezzi -così come illustrato in Galì e Gertler (1999), e che queste siano ben rappresentate dall'output gap, la curva di Phillips mette in relazione l'inflazione generata a livello nazionale, π^D, con un suo lag rispetto a se stesso ed alle aspettative, l'output gap ed ad uno shock stocastico, $\varepsilon_t^{\pi D}$. I parametri β_π e β_y determinano il grado di comportamento previsionale nel setting dei prezzi e nella sensibilità dell'inflazione a rallentare. Al fine di catturare al meglio la sua inerzia osservata, si specifica come la persistenza nei movimenti dell'output gap o nelle variazioni delle aspettative di inflazione potrebbero produrre un percorso inerziale per l'inflazione, ma lascia aperta la possibilità a grandi salti. Ciò implicherebbe il fatto che sia l'inflazione a guidare l' output gap benché i dati suggeriscano l'opposto. Evidenza empirica e saper economico ci suggeriscono come sia la politica monetaria ad influenzare i prezzi benché con un lag temporale mai istantaneo come identificato in Rudd, Whelan (2007). Nel modello si presume che la policy monetaria in UK, US e nell'Eurozona siano credibili ma soprattutto stabili nel periodo di determinazione delle stime. Le aspettative di lungo termine sono legate all'inflazione target ($\pi_{t-1|t} = \bar{\pi}_t$). Relativamente agli shock sulle aspettative d'inflazione si utilizza un semplice processo autoregressivo.

La persistenza dell'inflazione interna, destagionalizzata e calcolata su base trimestrale – è inferiore a quello dell'output gap. Inoltre, essa nella sua versione headline inflation è ancora più bassa, grazie all'effetto positivo ma temporaneo degli shock dei prezzi all'importazione come pure su quelli energetici. La pendenza della curva di Phillips viene determinata attraverso il livello di sensibilità dell'inflazione domestica rispetto all'output gap. Da sottolineare come la headline inflation riflette sia dinamiche interne che esterne legate alle variazioni nel tasso di cambio, prezzo del petrolio ed eventuali shock aggiuntivi.
Relativamente al comportamento della banca centrale, si assume che il tasso di policy, i, segua una forward-looking Taylor rule. La funzione relativa vede come parametri i tassi d'interesse lagged di 1 periodo, l'expected output gap, una deviazione attesa dal livello target dell'inflazione interna, un aumento del credit risk attraverso il credit spread e il tasso di finanziamento assieme, ovviamente a un elemento di shock stocastico.

$$i_t = \delta_i i_{t-1} + (1-\delta_i)\delta_i y_{t-1|t} + (1-\delta_i)\delta_{\pi^D}\left(\pi^D_{t-1|t}\right) + (1-\delta_i)\bar{\pi}_t - \delta_{cs}\Delta \mathrm{cs}_t + \varepsilon_t^i \qquad [8]$$

Rispetto ai coefficienti della Taylor rule vengono fissati, in linea con i modelli delle diverse banche centrali, valori vicini a 0,8. L'includere l'output gap e l'inflazione nella funzione di reazione delle banche centrali è tra l'altro coerente con l'obiettivo di stabilizzare l'inflazione al valore target del 2%.
Approccio che si è dimostrato quantomai valido nel recente passato. Sia la crisi finanziaria che la pandemia hanno dimostrato che le banche centrali non restano a guardare quando si manifestano distorsioni del mercato del credito. Quando un tale fenomeno si verifica i mercati valutano rapidamente una policy più accomodante che normalmente accade da lì a poco. Per questo motivo s'inserisce nel modello la variazione dei credit spread nella funzione di reazione della banca centrale. Il grado con cui quest'ultima accomoda o compensa l'iniziale aumento negli spread è dato da τ_{cs}. Credit spread che sono considerati come una variabile esogena rispetto al modello che segue un processo autoregressivo:

$$cs_t = \gamma^{cs} cs_{t-1} + \varepsilon_t^{cs} \quad [9]$$

Da ultimo, si osserva come il modello utilizzi un 5y risk free rate. Dopo la stima si procede in un'attività di mapping tra le variazioni dei 5y borrowing costs rispetto alle altre scadenze a seconda degli strumenti utilizzati dal monetary policy makers: policy rate, forward guidance o asset purchases. Per calibrare tale mapping viene utilizzata la tecnica proposta da Altavilla et al (2019), Weale e Wieladek (2016) e Kim et al (2020). Come documentato poi da Altavilla et al (2019) e Lane (2019) tali strumenti hanno effetti diversi sulla term structure governativa. Infatti uno shock per il policy rate colpisce principalmente la parte a breve della yield curve, mentre l'effetto svanisce gradualmente tanto più ci si porti su scadenze via via più lunghe. Le variazioni nella forward guidance producono un hump-shaped effect sulla yield curve, aumentando il segmento 2-5y senza produrre alcun impatto in termini di policy rate ed avendone un esaurimento graduale fino alla scadenza 10y. Da ultimo, l'attività della banca centrale che si focalizza sugli asset purchases avrà l'effetto di appiattire la yield curve, influenzando maggiormente i tassi a lunga scadenza rispetto a quelli a breve.
Dinamiche di breve termine del tasso di cambio reale sono guidate dalla real uncovered interest parity mentre a medio termine esso è ancorato alla purchasing power parity (PPP) verso la quale il tasso di cambio converge in maniera graduale ad una velocità determinata dal parametro ψ_q. Quest'ultima è basata sull'idea generale espressa dalla letteratura che la prima parte dell'effetto dello shock sia compreso in n periodo compreso tra i 3 ed i 5 anni – Rogoff (1996). Questo approccio consente al tasso di cambio di variare, seppur in una modalità attenuata per un miglior adattamento empirico, in presenza di uno shift delle aspettative sui tassi d'interesse.

A livello di stima, tutti e tre i modelli utilizzati dalle rispettive banche centrali (FED, BCE e BOE) – ossia l'FRB/US oltre al BASE/Eurozone e COMPASS/UK - utilizzano metodi bayesiani. Il funzionamento dei modelli descritti fin qui si basa su un periodo che parte dal primo trimestre 2000 fino al quarto trimestre 2019.
Lato portafoglio, e quindi in relazione al Bloomberg Factor Models® si rinvia a quanto illustrato in G. Macchia (2021). Tale modello tipo fa riferimento ad un'analisi delle factor exposures del portafoglio al fine di comprenderne al meglio gli effetti sulle diverse asset class e quindi di osservarne le capacità di diversificazione delle stesse nel tempo.

6. Scenario

Lo scenario avverso che viene considerato è quello di un prolungamento della guerra che potrebbe imporre un'interruzione significativa dell'approvvigionamento energetico con un deciso maggior costo energetico per l'Eurozona visti i livelli di dipendenza nel medio termine.

In particolare, lo scenario implica:

- un significativo e stabile maggior prezzo del petrolio (150 dollari al barile) e del gas (300 euro al mwh) per tutto il 2022 con picchi durante il 2° e 3° trimestre 2022 e successivamente un progressivo ma lento ritorno alla normalità entro la fine dell'anno;
- una flessione della domanda globale (-1,15% nel 1Q 2022 e -2,40% nel 2Q2022) dettata sostanzialmente da una riduzione della consumer spending;
- un forte aumento della volatilità sui mercati finanziari – VDaX non inferiore ad un livello di 50 nel 2° trimestre del 2022.

Effetti di un tale scenario - solo sfiorato dal mercato in data 7 marzo 2022 con il prezzo del gas e del petrolio che sono giunti rispettivamente oltre i 250 euro per MW/h e di 125 dollari al barile – sono uno scenario di staglazione nel 2022.

In particolare, si ravvisano nei seguenti impatti:

- Un aumento del 3% YoY dell'Head CPI
- Una riduzione dell'1,55% YoY da parte del GDP
- Un aumento della disoccupazione dell'1,31% YoY per il 2022 e dell'1,30% YoY per il 2023.

Per la Bce si tratterebbe di ritornare ad una policy pienamente accomodante fino a tutto il 2023, con un eventuale rialzo dei tassi fissato a non prima del 1H 2024. Attesa una riduzione della yield governative curve – con un riduzione di 25 bps circa sulla scadenza 10y ed associata aumento della slope.

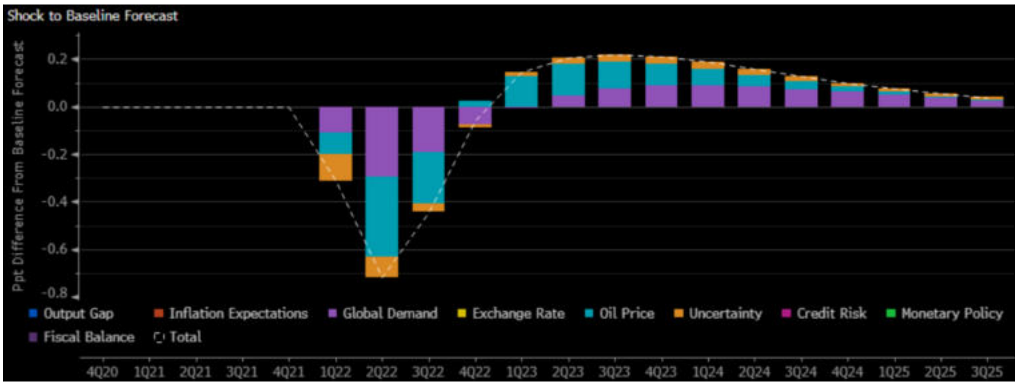


Fig. 9A, Stagflation Scenario impact on GDP growth

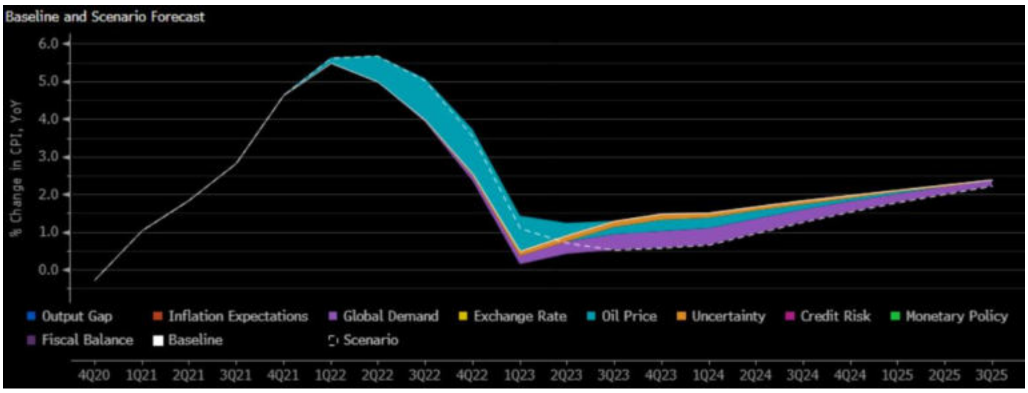


Fig. 9B, Stagflation Scenario impact on Inflation

Fig. 9C, Stagflation Scenario impact on Inflation

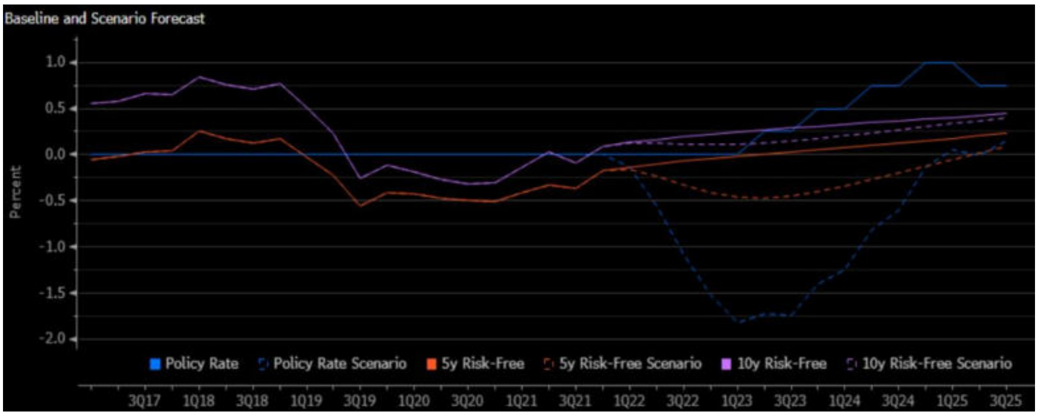


Fig. 9D, Stagflation Scenario impact on policy rate

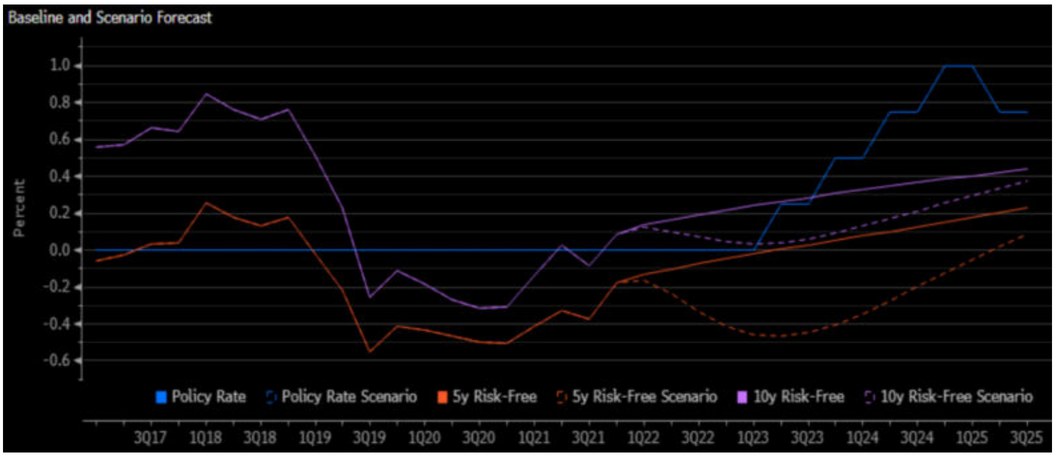


Fig. 9E, Stagflation Scenario on forward guidance

Per quanto riguarda invece l'economia americana, lo scenario avverso riguarda invece un'accelerazione del rialzo dei tassi nel corso del biennio 2022-2023 oltre ad una riduzione del balance sheet per circa 2,5 bn $ nel corso del triennio 2022-2024. Effetti di un tale scenario, che porterebbe ad un upward shift dell'intera yield curve governativa, comporterebbe un maggior repricing degli investimenti rischiosi oltre ad un deciso accorciamento del business cycle.

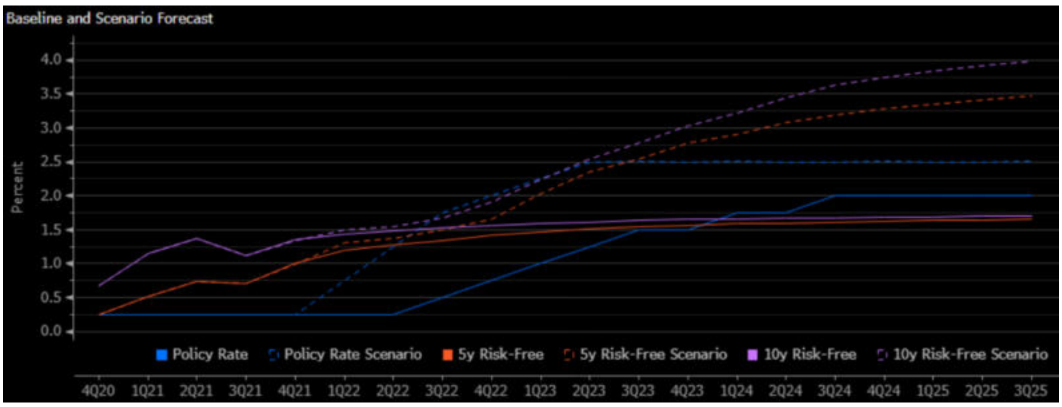


Fig. 10A, Alternative Shock and Awe Scenario per l'economia US

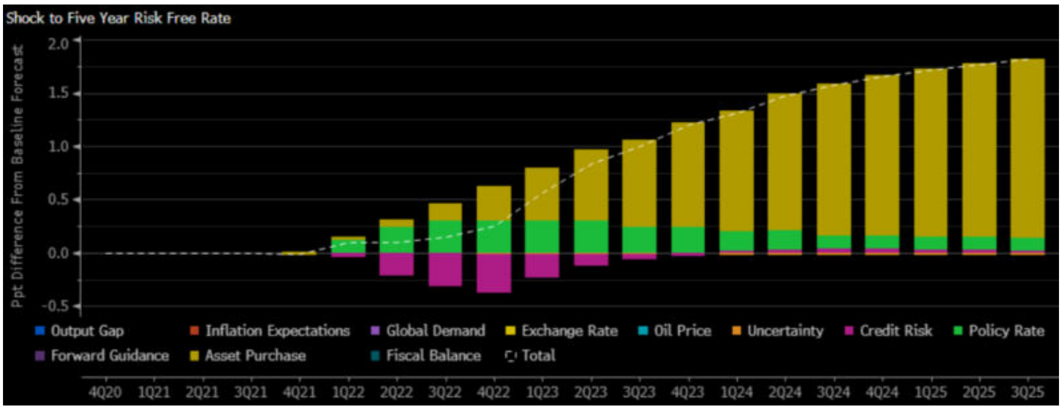


Fig. 10B, Alternative Shock and Awe Scenario per l'economia US

Grazie alle informazioni ottenute in termini d'impatto macroeconomico è possibile creare scenari in ambito finanziario.
Attraverso l'applicativo PORT di Bloomberg ed i Bloomberg Factor Models® possiamo identificare lo scenario che possa accogliere le informazioni precedentemente generate al fine di comprendere l'impatto sui diversi portafogli d'investimento e loro asset class.

In particolare, si considera:

- Una riduzione delle yield curve governativa dell'Eurozona con una moderata *bull flattening movement.* Riduzione di 25 bps del 10y associata ad un aumento della *slope* con tassi a breve termine che potrebbero portarsi marcatamente in area negativa (-1,5%)
- Fed Funds Rates a 150 bps con 5 aumenti da 25 bps per il 2022 per la yield curve governative US con riduzione slope (bear flattening)
- Petrolio a $150
- Aumento marcato della volatilità negli US (+60%) e Eurozone (+60%)
- +15% Usd/Eur
- +25% Gold

7. I risultati

Lo scenario avverso ha effetti diversi sui due portafogli: unconstrained risk control portfolio ed il moderate risk portfolio.
Benché tutti e due mostrino un rendimento positivo a seguito allo shock proposto, differente è la sua magnitudo (4,93% rispetto allo 0,81%).
Ciò è spiegato in parte da scelte a livello tattico e strategico, laddove nel primo caso è presente una componente liquid alternative (31%) che non è presente invece nel secondo portafoglio.

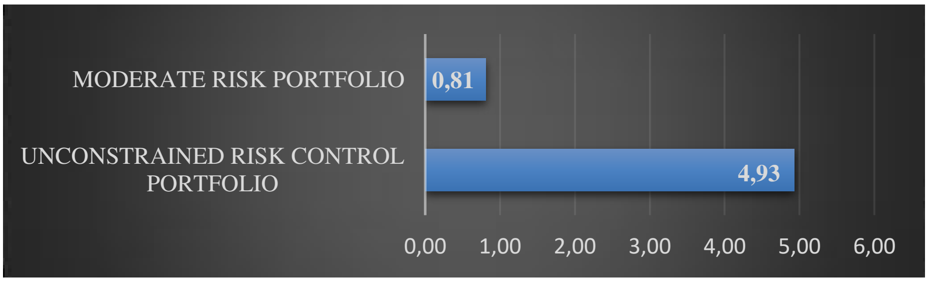


Fig. 11, Portfolio: Impatto Scenario Inflation Risk + Geo Risk

Per quanto riguarda l'Unconstrained Risk Control Portfolio, risulta essere pienamente nelle attese una performance negativa post scenario delle strategie di portafoglio che maggiormente sono esposte ad un quality factor mentre lo stesso non si ravvisa a livello value e per il dividend che mostrano di essere a maggior agio in contesti di mercato complessi in cui la Fed sarebbe obbligata ad una politica monetaria restrittiva in linea con i tassi impliciti nei Fed Funds Futures.
Si sottolinea la positiva contribuzione alla performance da parte delle strategie liquid alternative ed in particolare per quelle obbligazionarie.
In linea con il profilo rischio/rendimento è la performance attesa dalla strategia absolute return equity.
L'analisi della performance per la componente equity dei due portafogli conferma differenti pesi in riferimento all'esposizione in real estate, utilities ed industrial ma soprattutto ad una maggiore esposizione all'energy.
Il ricorso a strategie alternative, e quindi ad un'esposizione settoriale netta differente, è probabilmente da imputare le differenti performance per il settore materials, consumer discretionary e communication services.

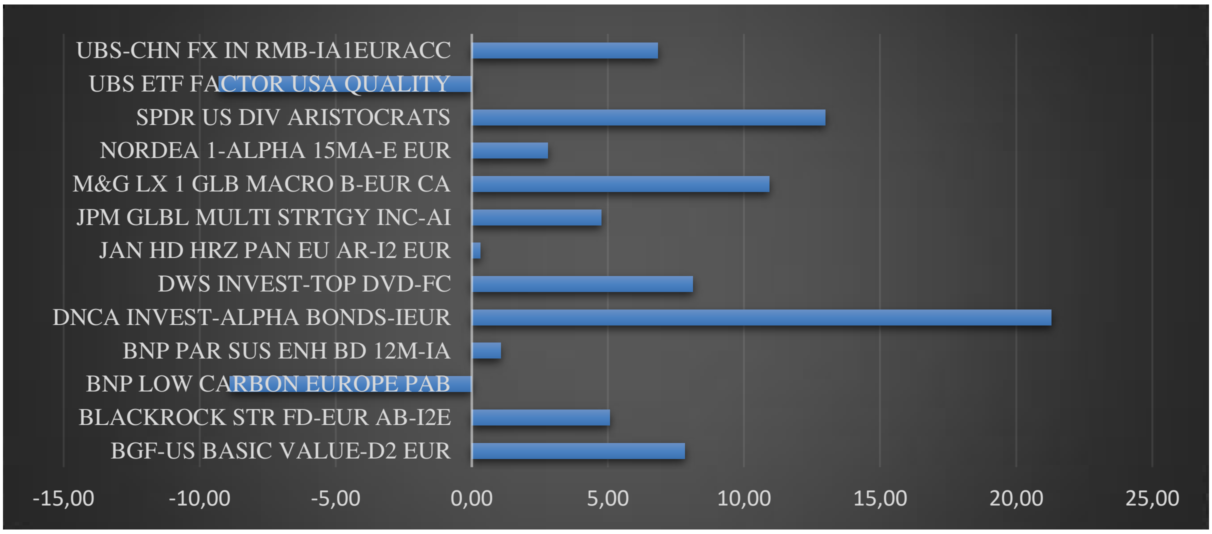


Fig.12A, Unconstrained Risk Control Portfolio: Impatto Scenario Inflation Risk + Geo Risk

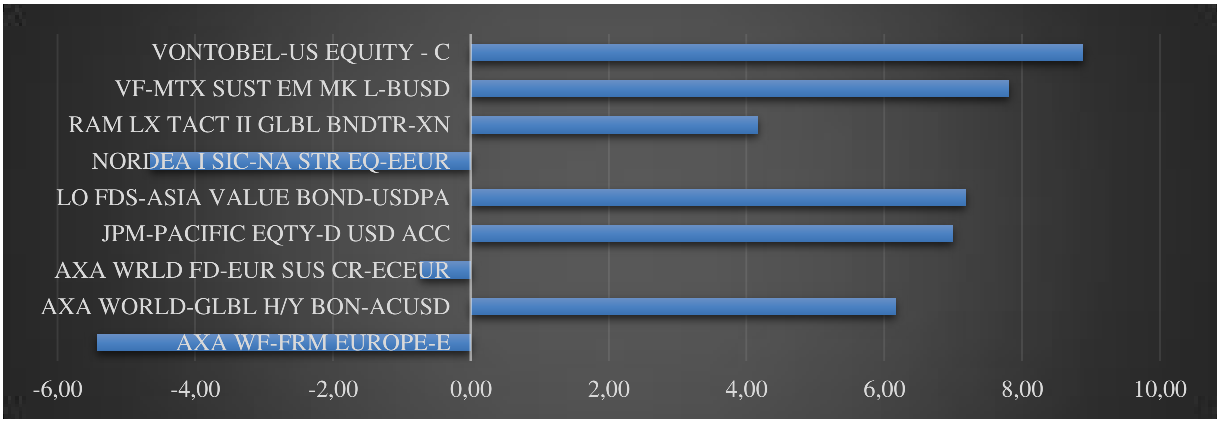


Fig. 12B, Moderate Risk Portfolio: Impatto Scenario Inflation Risk + Geo Risk

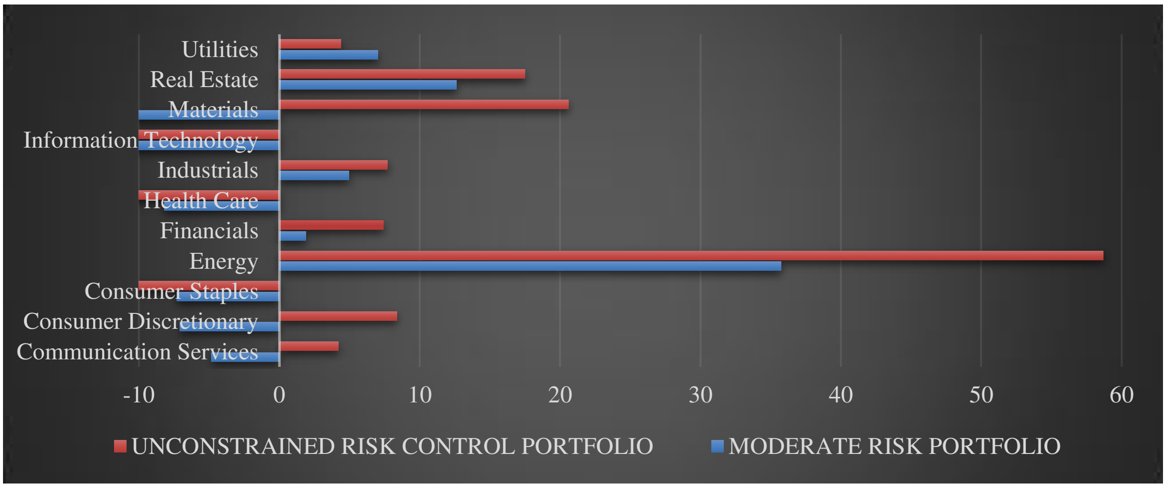


Fig.13, Equity: Impatto Scenario "Inflation Risk + Geo Risk"

Il confronto tra le misure di VaR ed Expected Shortfall pre e post scenario ci permette di effettuare una serie di considerazioni.
Nel caso dell'Unconstrained Risk Control Portfolio, la performance positiva (+4,93%) non mostra un impatto apprezzabile su VaR ed Expected Shortfall, anche in presenza di un leggero aumento della volatilità (8,05%). Confermata la proprietà della omogeneità delle misure di rischio adottate.
Situazione diversa per il Moderate Risk Portfolio. Anche in presenza di una performance leggermente positiva (+0,81%) si assiste ad una consistente riduzione del VaR (5,72 da 7,21) e dell'Expected Shortfall (da 7,84 da 9,88) spiegata in parte dalla riduzione della volatilità del portafoglio (7,24%) ed in parte determinata da una variazione della struttura di correlazione fattoriale a livello di singole strategie, assieme a quelle dove rendimenti negativi si associano ad aumenti del VaR ed Expected Shortfall dovuti quasi totalmente ad una maggiore volatilità, se ne osservano altre che mostrano comportamenti apparentemente contrastanti, ossia dove a fronte di rendimenti positivi si associano forti riduzioni del rischio.

Fig. 14, VaR ed Expected Shortfall post-scenario "Inflation Risk + Geo Risk"

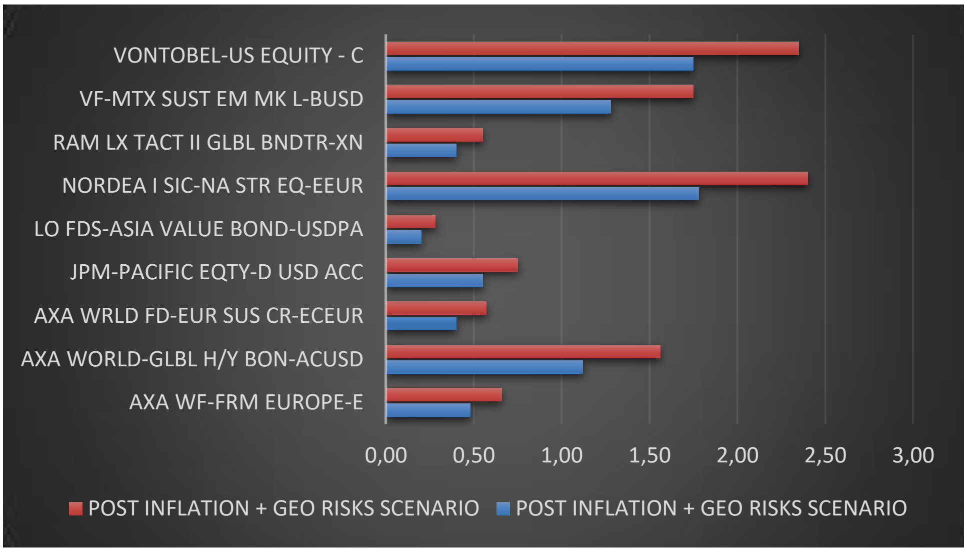


Fig.15, VaR & Expected Shortfall post scenario Moderate Risk Portfolio

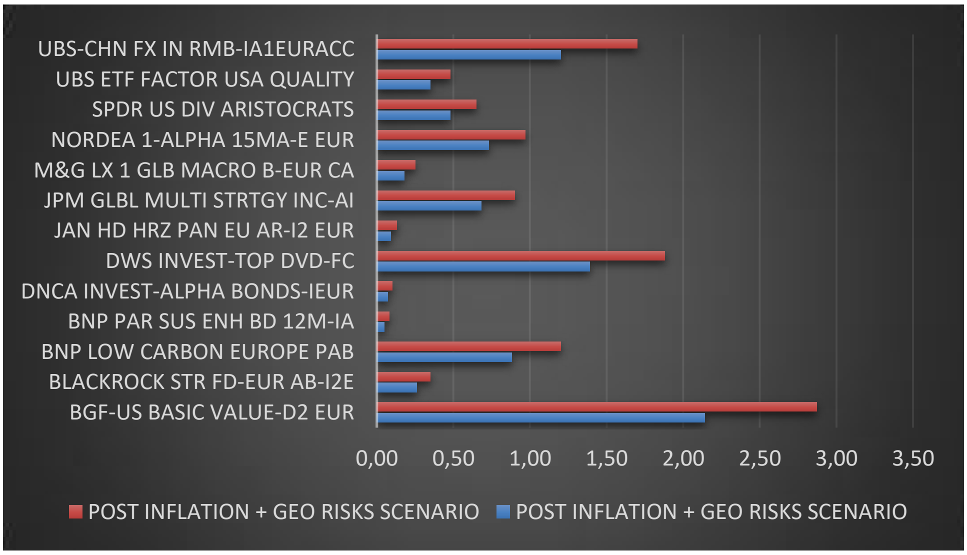


Fig.16, VaR & Expected Shortfall post scenario Unconstrained Risk Control Portfolio

8. Conclusioni

I risultati ottenuti nel presente articolo mostrano l'importanza di affiancare alle misure di rischio "classiche" come VaR e Expected Shortfall, uno stress testing/ scenario analysis framework che permetta all'investitore di poter comprendere anticipatamente quali possano essere gli impatti non solo a livello di rendimento ma di tenuta del profilo di rischio.
L'importanza di tale metodologia, che prevede un doppio step operativo per valutarne l'impatto a livello di portafoglio, è sottolineata anche dalla variazione in termini di VaR ed Expected Shortfall in virtù di una variazione nella struttura di correlazione a livello fattoriale.
Da ultimo, è importante osservare come l'utilizzo di adeguate strategie alternative permettano una maggiore potenziale resilienza rispetto ad uno scenario avverso.

Bibliografia

- C. Altavilla, L. Brugnolini, R. S Gurkaynak, R. Motto e G. Ragusa "Measuring euro area monetary policy" Journal of Monetary Economics, 108:162–179, 2019.
- E. Angelini, N. Bokan, K. Christoffel, M. Ciccarelli, S. Zimic "Introducing ECB-BASE: The blueprint of the new ECB semi-structural model for the euro area" Working Paper Series n°2315, Settembre 2019
- F. Brayton, T. Laubach e D. Reifschneider "The FRB/US Model: A Tool for Macroeconomic Policy Analysis" Board of Governon of the Federal Reserve System Aprile 2014
- S. Domit, F. Monti e A. Sokol "A Bayesian VAR benchmark for COMPASS" Staff Working Paper n°583 Bank of England, 22 Gennaio 2016
- Jordi Galı and Mark Gertler "Inflation dynamics: A structural econometric analysis" Journal of Monetary Economics, 44(2):195–222, 1999.
- K. Kim, T. Laubach, M. Wie "Macroeconomic effects of large-scale asset purchases: new evidence. 2020.
- J. Kim and C. C. Finger, "A Stress Test to Incorporate Correlation Breakdown," Journal of Risk 2, no. 3 (Spring 2000): 5–19.
- P. Kupiec, "Stress Testing in a Value at Risk Framework," Journal of Derivatives 6 (1999): 7–24.
- Philip R Lane "The yield curve and monetary policy" public lecture for the centre for finance and the department of economics at university college london, 2019.
- G. Macchia "Economic recovery and inflation risk: what is the "price" to manage debt?" Risk Management Magazine – Volume 16, Issue 2, pag. 21 – 34.
- Jose Menchero, Lei JI "Multi-Asset Class Risk Model -MAC2" Bloomberg Professional Service Offering (2016)
- K. Rogoff "The purchasing power parity puzzle"Journal of Economic Literature, 34(2):647–668, 1996.
- J. Rudd e K. Whelan "Modeling inflation dynamics: A critical review of recent research" Journal of Money, Credit and Banking 39, 155-170 (2007)
- J. Rush, D. Hanson, M. Cousin, B. Van Roye "Bloomberg Economics Forecast Models for the U.S, U.K and Euro Area" Bloomberg Professional Service (2020)
- Hull, John C. "Risk Management and Financial Institution – 5th Edition" Wiley (2018)
- M. Weale e T. Wieladek "What are the macroeconomic effects of asset purchases? Journal of monetary Economics, 79:81–93, 2016.

Estimation of Flood Risk on a residential mortgages portfolio

Luca Bartolucci (Prometeia)[1], Guido Luciano Genero (Intesa Sanpaolo), Maurizio Pierigè (Prometeia)[1], Fabio Verachi (Intesa Sanpaolo),

Article submitted to double-blind peer review, received on 11th July 2022 and accepted on 26th July 2022

Executive Summary

In the context of the rapid changes that have occurred in recent years, characterized by veritable 'black swans' such as the COVID-19 pandemic and extreme weather events that are occurring with increasing frequency, the issue of climate change has come into the focus of banking regulators and supervisors.
Therefore banking institutions, if they are subject to the Single Supervisory Mechanism, have been called upon to develop (and, subsequently, to integrate into their business practices) methodologies for the identification, quantification and management of such risks, mainly under the profiles of:

- Transition Risk, associated with policies undertaken by governments to foster climate change mitigation and adaptation;
- Physical Risk, associated with the occurrence of extreme climatic events and its impact on the bank's assets.

This paper analyzes one of the most significant hazards within the Physical Risk domain, which is Flood Risk. The measurement is focused on the prospective evolution of the flood events on a portfolio of mortgages secured by residential properties. The impact of this risk driver is subsequently reflected through the movement of appropriate transmission mechanisms on the LGD and PD parameters relating to the exposures in the scope.
Finally, the effect on loan adjustments is provided, by recalculating the expected losses that result from the stressed projections. The flood risk projection is executed on a long-term timeframe, developing over 3 climate scenarios up to 2050.
The choice of this hazard is due to its relevance in terms of frequency of events and harmfulness, a relevance that is confirmed by its inclusion in both the top-down climate stress testing exercises carried out by the ECB and in the bottom-up climate stress testing exercise promoted by the ECB itself in 2022 and carried out by the SSM Banks.

A comprehensive simulation framework, structured as follows, is then presented:

- a macro-climate scenario simulation engine;
- the downscaling of these scenarios to obtain localized climate effects on individual properties;
- the transmission of these effects into a depreciation formula for the individual property;
- the LGD stress associated with the devaluation of the collateral property, and the PD stress that goes along with it, obtained by correlation.

Key Words:

Climate risk, Physical Risk, Flood Risk, macro-climate scenario, LGD

Literature review

The research about the effects of flood events on mortgage portfolios is quite new, and mostly focused on American properties. Kousky et al. (2020) provide a comprehensive literature review on the topic, highlighting the low rate of (US) insured properties in areas prone to flood risks, and investigating potential behavioural reasons for this low take-up level of insurance.

In the work of Ratnadiwakara and Venugopal (2020), the relationship between flood risk exposure and lower-income borrowers is proven, as well as the higher likelihood of mortgage default for houses bought after a major flood event. Calabrese et al. (2021) proposed an additive Cox proportional hazard model with time-varying covariates to estimate the impact of the exposure to flood risk (in interaction with heavy rainfall) and tropical cyclones on mortgage default in Florida.

They also performed a scenario analysis in 2050 under RCP 4.5 scenario. The main evidence from scenario analysis is an increase in the risk of default for properties located in areas with a projected increasing exposure to flood risk. Caloia and Jansen (2021) instead proposed to carry out a flood risk stress test for the Netherlands, by estimating (with a *damage function* approach) damages to properties deriving by flooding for several return periods, which are then feeded into PD and LGD satellite models to estimate their impact on the bank's balance sheets.

This stream of research is integrated, in this paper, with a sophisticated Integrated Assessment Model (IAM) used for the long-term projection of economic and climatic variables.
The research on IAMs traces back to the seminal works of Nordhaus, who was awarded the Nobel Prize in Economics in 2018 with the following motivation *"[..] integrating climate change into long-run macroeconomic analysis"*. Indeed, Nordhaus (1991) was the first work to evaluate the feedback effects between economic activity and the climate through the GHGs emissions channel, and his prolific scientific production has widened the frontier of climate change economics (see Nordhaus (2018)).
As it is known, the feedback loop works as follows: GHG emissions are a byproduct of economic activity that affect the climate which in turns impacts economic activity through negative effects on productivity and damage to the capital stock. Further developments of this research stream (see Barrage 2020) allow for the modeling of climate transition policies, such as carbon taxes, and their distortionary effect on economic activity.

[1] We would also like to thank our team colleagues with special reference to: Giacomo Novelli (PhD, Prometeia) and Michele Catalano (PhD, IIASA - International Institute for Applied System Analysis), who oversaw the identification of exposures and the measurement of hazard, as well as Marco Brandolini, Andrea Lugli, Cristiana Moriconi, and Luca Zanin who oversaw the vulnerability analysis and the transmission of impacts on credit risk parameters.

This type of models allows to properly estimate the welfare costs of taxing emissions against the economic costs associated with not taxing emissions, allowing GHGs to rise.
One final tool for the evaluation of the welfare effects of climate transition policies on a global scale is the use of an OLG (Overlapping Generations) model, which is able to properly discount the welfare of future generations against the costs needed for the current transition policies (Kotlikoff et al. 2019).

General framework

Following the entry into force of the Paris Agreements and the European Union's adoption of the commitment to achieve climate neutrality by 2050 ('Fit for 55'), the European financial sector is called upon to play a leading role in the ecological transition process.

From a regulatory point of view, this role is to be outlined as follows:

- on one hand, as a role in monitoring the environmental sustainability of investments, according to the technical criteria outlined in the EU Taxonomy, with the aim of directing capital flows towards 'green' activities;
- on the other hand, through the monitoring and management of environmental, social and governance (ESG) risks.

Climate risks can be divided between risks related to transition policies (such as carbon taxes, ETS price increases, energy performance constraints) and risks related to the impact of weather events, i.e. physical risks (which in turn are divided between risks related to extreme weather events, i.e. *acute* risks, and risks related to permanent changes in weather phenomena, i.e. *chronic* risks).

This paper analyses one of the most significant hazards within this last category, measuring the prospective evolution of the **flood risk on a portfolio of mortgages** secured by residential properties.

The impact of this risk driver is then reflected through the movement of appropriate transmission mechanisms on the LGD and PD parameters related to the exposures in the perimeter. Finally, the effect on impairments flows is estimated, through the recalculation of the expected loss.

The choice of this hazard is due to its relevance in terms of frequency of events and harmfulness, a relevance that is confirmed by its inclusion in both the top-down climate stress testing exercises carried out by the ECB and in the bottom-up climate stress testing exercise promoted by the ECB itself in 2022 and carried out by the SSM Banks, as well as by the availability of data (which makes possible a more punctual and reliable mapping phase, compared to other less consolidated and/or more complexly measured hazard events).

The methodology for flood risk evaluation can be broken down along the following axes:

1. Identification of **exposure**: the geographical location of properties, especially in terms of proximity to waterways
2. **Hazard measurement**: it consists of the forecast of the flood phenomenon, expressed in terms of flood depth. The forecast is calculated from the application of the IAM macroclimatic model, which estimates, among other output variables, temperatures, winds and precipitation. The forecast horizon is 2022-2050 and is based on three distinct NGFS scenarios: Orderly Transition, Disorderly Transition and Hot-House World, which were also used for the recent ECB-sponsored regulatory exercise 'Climate Risk Stress Test 2022';
3. **Vulnerability**: the vulnerability of individual exposures is derived from the characteristics of the individual property (e.g., in terms of construction material, the height of the house and the overall interior and exterior quality, etc.).

The combined analysis of these factors results produces an estimate of the expected impairment of the property under the different scenarios considered, which for the purposes of this analysis is the main stress factor for the exposure's credit risk parameters.
The final step is to recalculate the portfolio impairments with the stressed PDs and LGDs. The simulation time horizon is, consistently with the applied NGFS scenarios, 2050.
The assumption about the evolution of loan volumes is that of a static balance sheet: maturing exposures are replaced by equal volumes of new loans with the same financial characteristics. The starting portfolio consists of performing exposures exclusively.

Methodology: Exposure

The portfolio considered for the purposes of this use case consists of 8,835 residential mortgage exposures, secured by properties entirely located in Italy. The source database, extracted on data updated to 2021, provides information about:

- location (address), intended use, value of the property, "asset specific" information - if available;
- guaranteed amount and LGD of the exposure;
- PD of the entrusted counterparty.

Each exposure in the portfolio in scope is guaranteed by a single residential property; accordingly, in the remainder of this document, reference is made indiscriminately to properties and exposure lines.

As the following table shows, the portfolio is well diversified geographically across the country:

Table 1 – Geographical breakdown of properties

Region	Amount	#	Share (amount)
ABRUZZO	23.808.601	162	1,5%
BASILICATA	13.821.452	97	0,9%
CALABRIA	41.962.692	333	2,6%
CAMPANIA	167.516.597	885	10,5%
EMILIA ROMAGNA	126.059.866	719	7,9%
FRIULI VENEZIA GIULIA	38.786.340	265	2,4%
LAZIO	264.533.445	1.121	16,7%
LIGURIA	54.648.183	312	3,4%
LOMBARDIA	274.785.356	1.423	17,3%
MARCHE	26.915.106	175	1,7%
MOLISE	3.637.918	28	0,2%
PIEMONTE	64.304.739	418	4,0%
PUGLIA	131.427.410	873	8,3%
SARDEGNA	24.793.483	157	1,6%
SICILIA	22.377.485	143	1,4%
TOSCANA	148.797.334	784	9,4%
TRENTINO-ALTO ADIGE	4.509.421	18	0,3%
UMBRIA	22.976.036	169	1,4%
VALLE D'AOSTA	2.490.446	11	0,2%
VENETO	130.383.763	742	8,2%
TOTAL	1.588.535.672	8.835	100,0%

In order to provide a starting (static) representation of flood risk exposure, the aggregated portfolio is represented below according to risk clusters on a provincial basis (NUTS2), proposed by the ECB and used for the 2022 Climate Risk Stress Test.

Table 2 – Breakdown of properties by ECB risk bands

Flood Risk	Amount	#	Share (amount)
MINOR	683.217.576	3.547	43,0%
LOW	420.567.848	2.646	26,5%
MEDIUM	476.561.005	2.599	30,0%
HIGH	8.189.243	43	0,5%
TOTAL	1.588.535.672	8.835	100,0%

According to the flood risk classification established by the Regulator, four provinces[2], out of the entire Italian territory, are classified as exposed to a "High" risk.

In order to follow up on the exercise by means of the proposed methodological approach, which envisages a more granular association of the property with the territory in order to evaluate more precisely the environmental characteristics of the actual location (e.g., the individual properties are linked to the geographic cells for which the models simulate the manifestation of flood phenomena), the assets were geo-localised in order to obtain, starting from the address, the latitude and longitude.

The risk parameters (PD and LGD), to make the results more comparable by isolating only the effects of Flood Risk, have been made homogeneous for all exposures in scope, consequently the differences in PD and LGD levels obtained derive entirely from Flood Risk shocks.

Methodology: Hazard

The flood risk estimation as a physical phenomenon can be broken down into two distinct steps:

- rainfall forecasting;
- flood depth[3] estimation.

The estimation of rainfall and the consequent Flood Depth is carried out at the level of the individual property, traced back to the relevant geographical cell. Furthermore, above the whole procedure there is a top-down econometric model, i.e., an IAM model developed by Prometeia, which projects a set of macroeconomic and climatic variables over long-term time horizons (in this

[2] Provinces of Imperia, Sondrio, Verbano-Cusio-Ossola and Aosta.
[3] The exercise is focused on flood risk events originating from river floods; sea floods events have been excluded.

specific case a projection with a time horizon up to the end of the century is considered) at a high level of geographical aggregation (national). The equations are calibrated to NGFS[4] scenario values (June 2021 release).

Following the calibration of the IAM model on the NGFS scenarios, the methodology envisages further specification of the forecasts in order to produce climatologies - i.e., geo-referenced grids of climatic data (in this case the climatic data of interest are the precipitation for the hazard event under analysis). Given the significant variability of this specific climatic phenomenon (floods), it was decided to produce the climatologies with a high granularity, defining cells of $1km^2$, to adequately capture the riskiness to flood events of individual building units.

This specification is achieved by relating the national forecasts to the SSP/RCP scenarios developed within the World Climate Research Programme[5]. At this stage an initial downscaling between the Prometeia-NGFS macro scenarios and the SSP/RCP scenarios is performed. This calculation makes it possible to use the climatologies of the SSP/RCP scenarios but calibrated to the paths of the NGFS scenarios in terms of climate variables, and to exploit their vast analytical potential in terms of geolocation of atmospheric phenomena. Downscaling can thus be defined as the process of inferring more granular climate data than the source data, covering a wider area.

To be able to use climatologies at the desired level of granularity (which, for the purpose of this paper, consists of the single property), starting from the climatologies identified in the previous point, a further step of statistical downscaling of climate metrics on the geographical cells of interest is carried out on the basis of a high granularity historical dataset[6].

This calculation process then allows a probability distribution of the flood depth to be estimated for each forecast instant, for each cell and for each scenario. The probability distribution of Flood Depth is calibrated and validated on the main global hydrogeological[7] models and is forecasted for both river flooding and urban.

From the probability distributions estimated in the previous point, for the purpose of this exercise, 5 return periods[8] were isolated: 10, 50, 100, 200 and 500 years.

Methodology: Vulnerability

Estimation of the structural depreciation of residential property

The process for estimating the theoretical depreciation of the property due to flood risk starts by translating the flood depth into structural damage of the property. For the estimation of structural impairment, the damage function library (estimated by the JRC[9] and available for Europe) is used. The function allows the theoretical structural damage resulting from flooding to be quantified based on the location of the property and certain of its characteristics. These include:

- reconstruction costs;
- vulnerable surface area (squared meters) of the property to the flood phenomenon;
- maintenance status;
- property type;
- number of floors (cadastral) of the building.

On these quantities, to allow the function to be applied even in the presence of missing information, ad hoc models were created to estimate and complete the minimum information set. Finally, to correctly assess the value of the damage over the entire time span under analysis, the results of the theoretical damage function at time t0 are projected through the use of a property price forecasting model, integrated with the IAM model.

Once the value of the damage has been simulated, this is then transformed into a percentage of devaluation through the application of a mathematical model which, combining the extent of the damage itself and the territorial specificity of the reference property market (purchase prices), calculates as a final result the monetary loss (expressed in current euros) connected to the flood event for each asset, year, scenario and return period.

Calculation of expected depreciation

The next step is to calculate the expected write-down (or loss), known in the catastrophe risk literature as the Average Annual Loss: each percentage write-down is multiplied by the respective 1-year marginal probability of occurrence.

Equation 1 – Calculation of expected Average Annual Loss

$$AAL_t\% = q_{sv}^{RP_{500}} p_{RP_{500}} + q_{sv}^{RP_{200}} p_{RP_{200}} + q_{sv}^{RP_{100}} p_{RP_{100}} + q_{sv}^{RP_{50}} p_{RP_{50}} + q_{sv}^{RP_{10}} p_{RP_{10}}$$

[4] Specifically, of the six overall scenarios provided by NGFS in the release, those actually used are as follows: Orderly Transition – Net Zero 2050, Disorderly Transition – Delayed Transition, Hot-house world – Current Policies.
[5] Coupled Model Intercomparison Project Phase 6: https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6 .
[6] Database WorldClim.
[7] HadGEM2 Model: https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-hydrology-variables-derived-projections?tab=doc .
[8] For example, the flood depth of a flood event associated with a 500-year RP can be interpreted as the flood depth of an event occurring on average once every 500 years (in a specific location, at a specific date, in a specific scenario).
[9] Huizinga, J., De Moel, H. and Szewczyk, W., (2017). Global flood depth-damage functions: Methodology and the database with guidelines, EUR 28552 EN, Publications Office of the European Union, Luxembourg.

The result of the AAL%, calculated on the portfolio and multiplied by the appraised value of the properties in scope, allows the calculation of the average pure risk premium across the whole simulation (i.e., the "fair" insurance premium that one would have to pay to protect against flood risks in the absence of a mark-up by the insurance company). The results of this metric, calculated as the Euros required to insure 100,000€ of collateral value, are as follows:

Table 3 – Portfolio Pure Risk Premium

Scenario	Premium
Orderly Transition	4,3€
Disorderly Transition	5,2€
Hot-House World	6,0€

Following this calculation, from the AAL forecasts for each property, stressed appraisal values that include the expected effect of Flood Risk are estimated for each year and scenario. Once these appraisal values have been estimated for each simulation year, for LGD calculation purposes the expected loss calculated over an 8-year period from the date of default is estimated at each future date. This observation horizon represents a conservative estimate of the time required [10] for the workout of (secured) non-performing loans in Italy.

Methodology: Impact on risk parameters and impairments

LGD Stress

Having produced as output from the previous step a forecast of the depreciation of the property, the resulting LGD stress is calculated for each individual position, based on the cumulative depreciation of the property relative to the starting value. The formula is applied as follows:

Equation 2 – LGD stress based on collateral depreciation

$$LGD_t = LGD_0 \cdot \left(1 + \left(\frac{-(1 - LGD_0)}{LGD_0} \cdot V\%_{value_t}\right)\right)$$

where:

- $V\%_{value_t}$ is the cumulative percentage (de)growth of the appraisal value.

PD stress

As described in the previous sections, the elaboration is carried out on the basis of three NGFS-derived long-run scenarios. It should be noted that for PD stress purposes the impacts of macroeconomic variables are sterilized: the only impact considered is the impact from flood events. Therefore, the impact from flood events is calculated using a correlation formula between LGD and PD. Although the literature on this subject is still scarce, empirical evidence [11] has been found of a significant deterioration in the credit quality of real estate loan exposures following heavy flood events.
To measure the correlation between PD and LGD, we chose to apply the formula proposed by Frye-Jacobs [12], which establishes the following correlation structure between conditional LGD (denoted LGD1) and stressed PD (denoted PD^{ST}) [13]:

Equation 3 – Frye-Jacobs model

$$PD_{ST} = \frac{\Phi\left[\Phi^{-1}[PD_{ST}] - \frac{\Phi^{-1}[PD_{TTC}] - \Phi^{-1}[PD_{TTC} \cdot \mathrm{LGD}_0]}{\sqrt{1-\rho}}\right]}{LGD_1}$$

The basic assumption of this model is that PD and LGD are distributed according to the Vasicek [14] portfolio model and the PD_{ST} is assumed to be equal to the conditional Default Rate [15].

---

[10] https://www.infodata.ilsole24ore.com/2016/04/27/npl-la-media-italiana-per-il-recupero-dei-crediti-e-7-8-anni/ .
[11] Carolyn Kousky, Mark Palim & Ying Pan (2020) Flood Damage and Mortgage Credit Risk: A Case Study of Hurricane Harvey, Journal of Housing Research.
[12] J. Frye, M. Jacobs Jr., Credit loss and systematic loss given default, The Journal of Credit Risk Vol 8, 1-32, Spring 2012.
[13] For the purposes of this exercise, no assumptions were made on the valuation of the correlation parameter ρ, which is set at zero. This choice is equivalent to removing the systematic (macroeconomic) risk factor common to all exposures in the portfolio, focusing exclusively on the idiosyncratic risk factors of individual exposures, namely Flood Depth.
[14] Vasicek, O. (2002), 'The distribution of loan portfolio value', Risk 15(12).
[15] See note 2 related to Equation (7) in António dos Santos "The relation between PD and LGD: an application to a corporate loan portfolio" - https://www.bportugal.pt/sites/default/files/anexos/papers/re202009_en.pdf

Results

To facilitate the reading of the results, the portfolio is segmented into four risk clusters. The segmentation is defined as follows:

- by sorting all exposures in the sample according to expected losses;
- by assigning to cluster "1" all exposures for which expected losses are zero;
- dividing the remaining exposures into three dimensionally homogeneous clusters.

FD_OT, FD_DT and FD_HHW denote the average Flood Depths (in meters) realized in the respective scenarios (Orderly, Disorderly and Hot-House). The outcome of the process is as follows (the values reported below refer to a Return Period of 500 years, projected to 2050):

Table 4 – Flood depth by risk clusters and scenarios

Cluster	Amount (€)	#	FD_OT (m)	FD_DT (m)	FD_HHW (m)
1	1.415.823.202	7.504	0,70	0,72	0,77
2	60.243.080	439	0,83	0,86	0,93
3	57.120.019	438	0,86	0,89	0,96
4	55.349.371	454	1,29	1,34	1,45
	1.588.535.672	8.835	0,74	0,77	0,82

Overall, we can see how the portfolio is predominantly distributed over geographic units of negligible risk, which is why they are placed in cluster 1. We also show how the risk classes established above are distributed with respect to the NUTS2 zones (provinces) identified by flood risk in the Climate Stress Test:

Table 5 – Flood depth by ECB risk classes

	# (€)	1(m)	2 (m)	3 (m)	4 (m)
MINOR	3.547	0,76	0,84	0,80	0,99
LOW	2.599	0,77	0,99	1,01	1,55
MEDIUM	2.646	0,79	0,95	1,07	1,70
HIGH	43	0,85	1,38	1,77	-

It should be noted that the results, in terms of Flood Depth forecasts, are substantially consistent, on average, with the mapping provided by the ECB: higher risk classes correspond to higher Flood Depths. Below is the numerical evidence of the exercise, in terms of stress on risk parameters and overall forecasts and adjustments.

LGD

Table 6 – Stressed LGD by scenario

ORDERLY	2021	2030	2040	2050
1	13,5%	13,5%	13,5%	13,5%
2	13,5%	13,5%	13,5%	13,5%
3	13,5%	13,7%	13,7%	13,8%
4	13,5%	14,7%	15,7%	16,8%

DISORDERLY	2021	2030	2040	2050
1	13,5%	13,5%	13,5%	13,5%
2	13,5%	13,5%	13,5%	13,6%
3	13,5%	13,7%	13,8%	14,0%
4	13,5%	15,1%	16,1%	17,4%

HOT-HOUSE	2021	2030	2040	2050
1	13,5%	13,5%	13,5%	13,5%
2	13,5%	13,5%	13,5%	13,7%
3	13,5%	13,7%	13,9%	14,1%
4	13,5%	15,1%	16,4%	17,9%

LGD stress intensity is significantly concentrated on risk cluster 4; there is a high similarity of the Disorderly and Hot-House World scenarios.

PD

Table 7 – Stressed PD by scenario

ORDERLY	2021	2030	2040	2050
1	0,67%	0,67%	0,67%	0,67%
2	0,67%	0,67%	0,67%	0,67%
3	0,67%	0,72%	0,72%	0,75%
4	0,67%	1,00%	1,33%	1,79%

DISORDERLY	2021	2030	2040	2050
1	0,67%	0,67%	0,67%	0,67%
2	0,67%	0,67%	0,67%	0,70%
3	0,67%	0,74%	0,75%	0,79%
4	0,67%	1,13%	1,49%	2,06%

HOT-HOUSE	2021	2030	2040	2050
1	0,67%	0,67%	0,67%	0,67%
2	0,67%	0,67%	0,69%	0,73%
3	0,67%	0,74%	0,77%	0,84%
4	0,67%	1,13%	1,61%	2,34%

Applying the Frye-Jacobs formula results in a high shift for PD, significantly higher than that of LGD, relative to the initial level of the respective starting points.

Balance-Sheet and Impairments

Table 8 – Balance-sheet projections and impairments, by scenario

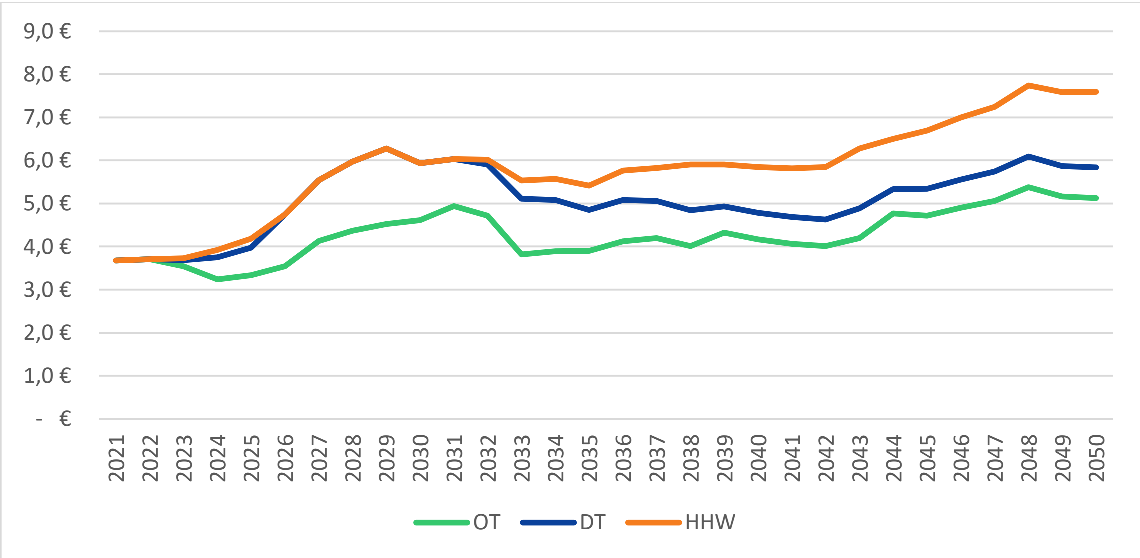
		ORD	ORD	ORD	HHW	HHW	HHW	DIS	DIS	DIS
	2021	2030	2040	2050	2030	2040	2050	2030	2040	2050
Gross Book Value	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**	**1.589**
Performing	1.589	1.537	1.485	1.432	1.536	1.483	1.427	1.536	1.483	1.429
1	*1.416*	*1.370*	*1.324*	*1.277*	*1.369*	*1.321*	*1.272*	*1.369*	*1.322*	*1.274*
2	*60*	*58*	*56*	*54*	*58*	*56*	*54*	*58*	*56*	*54*
3	*57*	*55*	*53*	*52*	*55*	*53*	*51*	*55*	*53*	*51*
4	*55*	*54*	*52*	*50*	*54*	*52*	*50*	*54*	*52*	*50*
NPL	0	52	104	156	52	106	162	52	105	159
Impairments [Stock]	**4**	**28**	**58**	**87**	**29**	**59**	**91**	**29**	**59**	**90**
Performing	4	4	4	4	4	4	4	4	4	4
1	*3,4*	*3,3*	*3,1*	*3,0*	*3,3*	*3,1*	*3,0*	*3,3*	*3,1*	*3,0*
2	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*
3	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,2*	*0,2*	*0,1*	*0,1*	*0,1*
4	*0,1*	*0,2*	*0,3*	*0,4*	*0,2*	*0,4*	*0,5*	*0,2*	*0,3*	*0,5*
NPL	0	25	54	84	25	55	87	25	55	86
Coverage Performing	0,2%	0,2%	0,2%	0,3%	0,2%	0,3%	0,3%	0,2%	0,3%	0,3%
1	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*
2	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,3%*	*0,2%*	*0,2%*	*0,3%*
3	*0,2%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*	*0,3%*
4	*0,2%*	*0,4%*	*0,5%*	*0,8%*	*0,4%*	*0,7%*	*1,1%*	*0,4%*	*0,6%*	*0,9%*
LGD Performing	13,5%	13,5%	13,6%	13,6%	13,5%	13,6%	13,7%	13,5%	13,6%	13,6%
Impairments [Flow]	*1,62*	*2,92*	*2,94*	*3,01*	*2,99*	*3,08*	*3,26*	*2,99*	*3,01*	*3,13*

Conclusions

The main evidence from this exercise is that a large majority of the portfolio falls into the negligible risk cluster. The most exposed part of the portfolio (risk class 4) consists of 3.5% of the overall exposure: for these positions, the coverage ratio in 2050 reaches between 3 and 5 times the initial level, depending on the scenario. It also emerges that even for risk classes 2 and 3, the effect is modest overall, compared to the baseline scenario: compared to the initial level, the coverage ratio at 2050 grows between 1% and 6% for class 2 exposures, while it grows between 15% and 31% for class 3 exposures. This outcome is consistent with the Flood Depth of the respective clusters. Geographical factors and characteristics of the different real estate zones contribute to this outcome: the analytical mapping of locations and its association with climatologies allows for a more accurate grasp of the actual vulnerability to flood events.

The evidence shown in Figure 1 shows a generalized upward trend in the risk premium from around 2040 onwards, resulting in an increase in this premium in the Hot-House world scenario of around 33% compared to the Orderly Transition scenario.

Figure 1 – Evolution over time and scenarios of the pure risk premium



From a technical point of view, potential areas for refining this type of analysis include:

- to elaborate further on the correlation between portfolio defaults (and the consequent parametrization of the ρ parameter of the Frye-Jacobs model, which, if valued, would lead to more severe results in terms of increase in expected loss);
- to evolve the calibration of the depreciation formula, based on the height of the property: a flat on the fifth floor, for example, would not be directly affected by a 'standard' flood event, but its value would be affected by damage to garages, common areas located on the ground floor, and so on. To this end, it is important to enrich the source property database as much as possible to reduce the use of statistical proxies;
- to integrate real estate price forecasts according to the NGFS/ECB reference scenarios into the methodological framework, capturing their market revaluation effect;
- to explore further the transmission channels on PDs, the relevance of which, emerging from the analysis, requires further contributions / in-depth analysis;
  to integrate the management of insurance coverage on the various positions into the methodological framework and, subject to data availability, refine the PD stress methodology to distinguish the riskiness of insured positions from that of uninsured positions.

Finally, future developments of this kind of analysis are going to be determined by regulatory (Stress Test exercises, ESG Pillar III, EU Taxonomy, etc.) and managerial needs, as the Banks will progressively integrate Physical Risks measurement into the strategic and business processes. The first step, on the regulatory side, is an adequate **mapping** of the bank exposure to Physical Risk – on immovable properties and NFCs as well (e.g ESG disclosure). A further development requires banks to equip themselves with a damage estimation methodology (e.g. ICAAP), such as the one presented in this paper for real estate assets. However, to fully integrate physical risk considerations into strategic choices, it is necessary to integrate **forward-looking climatic metrics**, in strategy and business processes. A not exhaustive list of use cases follows:

1. credit allocation should take into account the trade-off between different transition scenarios: economic sectors which may are less impacted by transition impacts (say because lower energy intensive) may be adversely affected by extreme climatic events, which are going to be more frequent and damaging on a longer timeframe;

2. due to the long maturity (and consequent riskiness) of mortgage exposures a forward-looking physical risk metric should be adopted and integrated into loan underwriting process;

3. because of the need to consider the uncertainty inherent into physical risk:
    a. loans pricing should be enhanced taking into account risk exposure;
    b. opportunities linked to insurance products, which may be provided to the customers or directly purchased by the Bank, can be progressively adopted.

Bibliography

- Barrage L. (2020) "Optimal Dynamic Carbon Taxes in a Climate–Economy Model with Distortionary Fiscal Policy," Review of Economic Studies, Oxford University Press, vol. 87(1), pages 1-39.
- Calabrese R., Dombrowski T., Mandel, A., Pace R.K. and Zanin L. (2021) Impacts of Extreme Weather Events on Mortgage Risks and Their Evolution under Climate Change: A Case Study on Florida
- Caloia F. and Jansen D.J. (2021), Flood Risk and Financial Stability: Evidence from a Stress Test for the Netherlands. De Nederlandsche Bank Working Paper No. 730
- Dos Santos A. R. (2020) "The relation between PD and LGD: an application to a corporate loan portfolio", Banco de Portugal paper available on https://www.bportugal.pt/sites/default/files/anexos/papers/re202009_en.pdf.
- Frye J., Jacobs Jr. M. (2012), Credit loss and systematic loss given default, The Journal of Credit Risk Vol 8, 1-32
- Huizinga, J., De Moel, H. and Szewczyk, W., (2017). Global flood depth-damage functions: Methodology and the database with guidelines, EUR 28552 EN, Publications Office of the European Union, Luxembourg.
- Kotlikoff L.J., Kubler F., Polbin A., Sachs J.D., and Scheidegger S. (2019) Making carbon taxation a generational win win. NBER Working Paper, (25760)
- Kousky C., Palim M. & Pan Ying (2020) Flood Damage and Mortgage Credit Risk: A Case Study of Hurricane Harvey, Journal of Housing Research.
- Kousky C., Kunreuther H., LaCour-Little M. & Wachter S. (2020) Flood Risk and the U.S. Housing Market, Journal of Housing Research
- Nordhaus, W. D. (1991). To Slow or Not to Slow: The Economics of The Greenhouse Effect. The Economic Journal, 101(407), 920–937.
- Nordhaus W.D. (2018) Projections and uncertainties about climate change in an era of minimal climate policies. American Economic Journal: Economic Policy, 10(3):333–60, 2018.
- Ratnadiwakara D. & Venugopal B. (2020) Do Areas Affected by Flood Disasters Attract Lower-Income and Less Creditworthy Homeowners?, Journal of Housing Research

Current and prospective estimate of counterparty risk through dynamic neural networks

Alessio Agnese (Duferco Energia), Pier Giuseppe Giribone (Banca Carige - Gruppo BPER, University of Genoa), Francesca Querci (University of Genoa)

Article submitted to double-blind peer review, received on 27th April 2022 and accepted on 12th July 2022

Abstract

The estimate of the probability of default plays a central role for any financial entity that wants to have an overview of the risks of insolvency it may incur by having economic relations with counterparties. This study aims to analyze the calculation of such measure in the context of counterparty risk from a current and prospective standpoint, by using dynamic neural networks. The forecasting aspect in the calculation of such risk measure is becoming more and more important over time as current regulation is increasingly based on a "Through the Cycle" and not a "Point in Time" assessment, consequently giving fundamental importance to such estimate. To this end, three different models aimed at calculating the Probability of Default have been investigated: the CDS method, the Z-Spread method, and the KMV method (Kealhofer, Merton and Vasicek). First, the different techniques have been applied to one of the main suppliers of gas and energy in Italy as a reference company. Then, they have been applied to calculate the same risk measure on the 50 companies included in one of the most important European indices, the Euro Stoxx 50.

Key Words:
Default Probability, Counterparty Risk, Credit Default Swap, Corporate Bond, KMV model, Nonlinear Auto-Regressive (NAR) network, forecasting

1) Dynamic neural networks for forecasting a time series

Artificial Neural Networks (ANNs) are a field of Machine Learning and they represent the cornerstone of Deep Learning algorithms. They owe their name and structure to the human brain, as they emulate the way in which biological neurons send signals to each other by acquiring knowledge of the external environment (Arbib, 2003). There are different types of neural networks, and they are classified according to the different purposes for which they are used. The perceptron is the oldest and simplest form of artificial neural net created by Frank Rosenblatt in 1958. It has an input layer, a single hidden layer and an output layer (Rosenblatt, 1958). Feedforward neural networks, also called multi-layer perceptrons (MLPs), are formed by an input layer, two or more hidden layers and an output layer (Haykin, 1994). Convolutional neural networks (CNNs) are similar to feed-forward networks, but their architecture is much more sophisticated (Cun et al., 1990). They are typically used for image recognition, pattern recognition and computer vision (Krizhevsky et al., 2012). This article focuses on recurring neural networks, or RNNs, which are characterized by feedback loops used for modeling time series to make predictions about future results (Kolen and Kremer, 2001). They are widely used in finance, for example for making predictions (Decherchi and Giribone, 2020). In order to make time series predictions using dynamic neural networks, it is necessary to use a sequence of values as input, in the first place, and, subsequently, to set the corresponding network in such a way so that it can use its previous values to best interpret the non-linear relations present in it (Tsay, 2010). We focus on recurring networks of the Non-linear Auto-Regressive network (NAR) type, which exclusively use the endogenous variable to perform this task (Beale et al., 2014). When a neural network is designed with these purposes, the hypothesis behind the reasoning is that the value in t of the time series can be a function of its past values. Furthermore, the prediction can be carried out at different depth levels: for example, if the goal is to find the value of a share on the following day, a step ahead prediction will be implemented. Likewise, if an analyst is interested in the trend that a certain variable may have over time, he will perform a multiple-step ahead prediction (Giribone et al., 2018). The quantity and quality of available data play a crucial role in the training of the network. Indeed, the available dataset needs to be as deep and complete as possible in order to obtain a statistically reliable forecast (Géron, 2019). The autoregression of the network can be mathematically represented by the following relationship: $y(t) = f(y(t-1), y(t-2), \ldots, y(t-n))$ where $y(t)$ is the time series to be modeled and $y(t-1)$, $y(t-2)$, $y(t-n)$ are the past values of the time series itself up to $t-n$ time lags (Agosto and Giribone, 2019). Similarly to the static version, NARs must also be trained using a gradient method. Through this procedure, the statistical model calibrates its own parameters in order to best interpret the input data. In the design of the architecture, particular attention was paid to the potential problem of over-fitting by dividing the sample of available data into a training-validation-test set (Giribone, 2020). In addition, different network configurations have been examined by putting as a parameter the number of hidden layers, the number of neurons and the lags with which the forecasting problem is to be econometrically faced. The selection measures to reach the best predictive architectures were designed according to the absence of self-correlation of the error, of the R^2 and of the RMSE (Root Mean Squared Error) calculated on the out-of-sample test (Bonini *et al.,* 2019). From an architectural point of view, ANNs are composed of layers of nodes that contain an input layer, one or more hidden layers and an output layer. In a dense artificial neural network, the neurons which belong to the next layer are fully connected with the neurons which are in the current layer. Every connection is characterized by a weight associated to the arch which links the two neurons, and every neuron is itself characterized by an activation threshold and a further parameter called bias. If the output of any single node of the architecture is above a threshold value, this node is activated, sending signals to the next layer of the network. Conversely, if the output is below the threshold value, no information is passed to the next layer (Rojas, 1996). Neural networks rely on training data for their training phase and for improving their accuracy over time. Once optimized, these learning algorithms are very powerful tools and allow to classify and organize data at a very high speed (Fonseca and Lopez, 2017). To understand the working principles of a neural network at an elementary level, each single node can be considered as a linear regression model composed of many inputs, many weights, a threshold and an output (Freeman and Scapura, 1992). Once an input level has been determined, the weights are assigned through a numerical optimization routine: the greater the weight associated to these hyperparameters, the greater the importance of the signal processed by the network. Subsequently, the inputs are multiplied by their respective weights and then added together. Once the process is finished, this first output obtained is passed through an activation function which determines a further output; if the output exceeds a certain significance threshold of the transfer function, then the node is activated, passing the data to the next level of the network: substantially, the output of one node becomes the input of the next node. This process of passing data from one level to another defines this type of neural network as a feedforward network (Caligaris *et al.*, 2015). During the training of the model, one of the needs is to evaluate its accuracy, for example through the Mean Squared Error (Chollet, 2018). The goal is therefore to reach a convergence point or a local minimum, through a progressive adjustment of the weights included in the algorithm (Principe, 2000). Neural networks are for the most part feedforward, that is, the signal flows in only one direction from the input to the output. However, the model adopts the back-propagation technique during the training, that is, information provided by the minimization of the cost function moves in the opposite direction: from output to input. The backpropagation allows to calculate and attribute the error associated to each neuron, allowing to appropriately adjust and calibrate the parameters (Rumelhart *et al.*, 1986). It is beyond the scope of this article to provide further technical details on the operating principles of a NAR. Interested readers can refer to the works mentioned above.

2) Current and prospective estimation of the Probability of Default using CDS premiums

Estimating the Probability of Default is of fundamental importance for financial institutions as it provides a summary information on the creditworthiness associated with the counterparty. The objective of this section is to analytically explain how to estimate the probability of default through CDS premiums and to provide two operational examples of calculation: the first is implemented by estimating the default probability at the present valuation time, so-called time t_0, and the second is implemented by calculating the same measure, but in a prospective way, that is after a forecasting of the CDS premiums made through dynamic neural networks. Given that the CDS is an insurance that hedges the holder from the loss in case of default of the issuer, it can directly measure the counterparty risk. As a result, this methodology is often preferred by the market players in comparison to others. In other words, if the counterparty in question has listed CDSs, it is preferable to use them for calculating the probability of default as such derivatives are expressly aimed at hedging credit risk (Bottasso *et al.,* 2019) through the following formula (Hull, 2015):

$$PD(T) = 1 - \exp\left(-\bar{\lambda}(T)T\right) \quad (1)$$

Where:

$PD(T)$ is the probability of default before T;

$\bar{\lambda}(T)$ is the hazard rate

T is the time to maturity expressed in years;

The above stated formula can be derived through the theory of probability by proceeding with the following steps. The hazard rate at time t is defined in such a way that $\lambda(t)\Delta t$ is the probability of default between time t and time $t + \Delta t$ conditional on the fact that there be no default before. In this sense, $\mathrm{SV}(t)$ is the cumulative probability that the company survives at time t.

Therefore, the conditional default probability between time t and time $t + \Delta t$ is:

$$PD(t, t + \Delta t) = \frac{\mathrm{SV}(t) - \mathrm{SV}(t + \Delta t)}{\mathrm{SV}(t)} \quad (2)$$

Since expression (2) is equal to $\lambda(t)\Delta t$, then we have:

$$\frac{\mathrm{SV}(t + \Delta t) - \mathrm{SV}(t)}{\mathrm{SV}(t)} = -\lambda(t)\Delta t \quad (3)$$

$$\frac{\mathrm{SV}(t + \Delta t) - \mathrm{SV}(t)}{\Delta t} = -\lambda(t)\mathrm{SV}(t) \quad (4)$$

By setting $\Delta t \to 0$ and assuming $\mathrm{SV}(t)$ differentiable (i.e., $\mathrm{SV} \in C^1$)

$$\frac{d\mathrm{SV}(t)}{dt} = -\lambda(t)\mathrm{SV}(t) \quad (5)$$

And by solving the Ordinary Differential Equation for $\mathrm{SV}(t)$, the following general solution is obtained:

$$\mathrm{SV}(t) = \exp\left[-\int_0^t \lambda(\tau)d\tau\right] \quad (6)$$

Defining $PD(t)$ as the issuer's probability of default at time t such that $PD(t) = 1 - \mathrm{SV}(t)$ we therefore have:

$$PD(t) = 1 - \exp\left[-\int_0^t \lambda(\tau)d\tau\right] = 1 - \exp\left[-\bar{\lambda}(t)t\right] \quad (7)$$

Where $\bar{\lambda}(t)$ is the average hazard rate or, equivalently, the default intensity, between time 0 and time *t*.

In the case of the CDS market, the premium can be seen as a direct compensation received by the insurer for the potential event of default by the issuer until maturity $t = T$. This means that the average loss rate between time 0 and time T can be annually approximated by the CDS premium $s(T)$, typically expressed in basis points.

Assuming now that the average hazard rate during such period is $\bar{\lambda}(T)$ and taking into consideration that the Recovery Rate (RR) in a standard Credit Default Swap is 40% (Giribone *et al.,* 2014), the average loss rate can be expressed by the quantity: $\bar{\lambda}(T)(1 - RR)$. This means that the following relationship is reasonably valid:

$$\bar{\lambda}(T)(1 - RR) = s(T) \to \bar{\lambda}(T) = \frac{s(T)}{(1 - RR)} \quad (8)$$

2.1) Pre-processing data for the default probability estimation at the current time

Before calculating the issuer's probability of default, a preliminary analysis of the dataset was necessary in order to verify, and potentially correct, the anomalies present in it. As a first step, after retrieving the CDS premiums at different maturities from Bloomberg®, Matlab was used to check the data and their consistency. The first check consists in analyzing the homogeneity of the dataset by verifying that there are no missing data, or that the CDS premiums are characterized by a daily frequency in the trading days considered in the analysis. In the case of anomalies of this kind, the missing values were linearly interpolated. After having interpolated any missing data, the various time series were grouped monthly to have a single monthly value for each different CDS maturity. To do this, the arithmetic mean of the different CDS Premiums was calculated for each month in order to obtain a dataset that is more aligned with the prospective time span of interest. The last check that was implemented, before the actual calculation of the probability of default, was the one relating to the presence of any potential outlier in the dataset. To search for such potentially odd values, the procedure is as follows: the historical series of the CDS Premiums were grouped two by two in ascending order (i.e., first grouping: 6M and 1Y, second grouping: 1Y and 2Y, etc.), then the difference between the different historical series was calculated and, on such difference, the mean and the standard deviation were calculated. The final step consists in constructing the interval within which the values are expected to fall. If the differences exceed the confidence interval, the premiums of the corresponding CDS are interpolated. The procedure can be exemplified with the following logic of preparation of the time series for CDS with tenors of six months and one year:

$$CDS6M = [CDS6M_1, CDS6M_2, \dots, CDS6M_i \dots, CDS6M_n]$$

$$CDS1Y = [CDS1Y_1, CDS1Y_2 \dots, , CDS1Y_i, \dots, CDS1Y_n]$$

With $i = 1, \dots, n$ where n is the length of the time series. This procedure has been repeated for each CDS tenor (6M-1Y-2Y-3Y-4Y-5Y-7Y-10Y).

$$D_{1Y-6M} = [D_1 = CDS1Y_1 - CDS6M_1, \dots, D_i = CDS1Y_i - CDS6M_i, \dots, D_n = CDS1Y_n - CDS6M_n]$$

$$D_{2Y-1Y} = [D_2 = CDS2Y_1 - CDS1Y_1, \dots, D_i = CDS2Y_i - CDS1Y_i, \dots, D_n = CDS2Y_n - CDS1Y_n]$$

This procedure has been repeated for each couple of CDS tenors.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} D_i \quad (9)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(D_i-\mu)^2}{n}} \quad (10)$$

$$\inf = \mu - 3\sigma$$

$$\sup = \mu + 3\sigma$$

$$\text{if}: \begin{cases} \inf \le D_i \le \sup \rightarrow \text{market } CDS\ premium \\ D_i < \inf \quad \text{or} \quad D_i > sup \rightarrow \text{interpolated CDS premium} \end{cases}$$

Where:

D is the vector containing the differences of the values contained in the CDS vectors;

μ is the mean of the differences of the values vector;

σ is the standard deviation of the vector;

inf is the lower extreme of the interval;

sup is the upper extreme of the interval;

As a result, the linear interpolation between two contiguous values (i-1 e i+1) of the time series has been carried out only when the CDS premiums provided by the market are considered outliers, i.e. they exceed the threshold of their mean ± 3 standard deviations. Once the control phase on the dataset was completed, the default probability was estimated for the various historical series of the CDS according to formula (1).

2.2) Estimation of the forecasted default probability

For the calculation of the prospective default probability, it was first necessary to forecast the CDS Premiums for each selected maturity. For forecasting purposes, NAR artificial neural networks were used. For each CDS maturity, 27,750 networks (for a total of 222,000) were tested in order to select the 8 best networks (one for each CDS reference tenor) for forecasting the data (Table 1). The selection of the most performing networks has been conducted only among the models characterized by the absence of auto correlation error (Econometric Test) and a good $R^2 = 0.95$ on validation set (Quality Accuracy). Among these architectures the best networks are chosen in accordance with the minimum out-of-sample gap. Table 1 also shows the In-Sample error and the average between the In-sample and Out-of-Sample error. This last quantity has been taken into account in the choice of the best network when the machine learning model with the lowest out-of-sample errors produced unfeasible forecasts (for example prospective negative premiums).

Non Linear Auto Regressive Best Networks					Econometric Test (Boolean)		Quality Accuracy
					TRUE		≥0.95
CDS	Number of Delays	Neurons (First Layer)	Neurons (Second Layer)	Neurons (Third Layer)	In Sample RMSE	Out Sample RMSE	MEAN IN SAMPLE AND OUT SAMPLE RMSE
6M	23	6	5	5	3.618442834	3.71288727	3.66566506
1Y	23	8	10	7	2.042700524	2.76245372	2.59477949
2Y	23	5	9	10	5.448432725	2.17786316	3.81314794
3Y	25	7	10	3	7.080359036	3.85242165	5.46639035
4Y	24	10	10	3	7.665926988	3.82910442	5.74751570
5Y	25	5	4	2	11.01484139	11.19604652	11.10544396
7Y	23	7	10	6	7.294506369	6.95375541	7.12413089
10Y	24	3	8	10	7.411680334	6.04787764	6.72977898

Table 1 Architecture of the Best networks: CDS method

After having trained and selected the various best networks, we proceeded with the forecasting of the various CDS Premiums and with the calculation of the default probability.

The figures $\tau = \{0.5,1,2,3,4,5,7,10\}$ have been estimated, as the year fractions t vary and for each tenor provided by the market, the corresponding listed spread $s(t, t+\tau)$ and the corresponding $PD(t, t+\tau)$.

In particular:

$t = 0$ corresponds to the first sampling date, which is 18th June 2008.

$t = 11.84$ corresponds to the date on which the network training set ends, that is 29th February 2020. The training set is characterized by the black line in Figure 1 both for the spreads, $s(t, t+\tau)$ and for the derived probability, $PD(t, t+\tau)$.

$t = 12.85$ corresponds to the date on which the network test set ends, i.e., 28th February 2021. The test set is characterized by the red line in Figure 1 both for the spreads, $s(t, t+\tau)$ and for the derived probability, $PD(t, t+\tau)$.

For $t > 12.85$ the forecasting is implemented. Both the predicted values of $s(t, t+\tau)$ and the value of $PD(t, t+\tau)$; they are characterized by the blue line in Figure 1.

By way of example, the graph shows the premiums of the various CDS with tenor equal to 7 years and the corresponding default probability estimated using the CDS method.

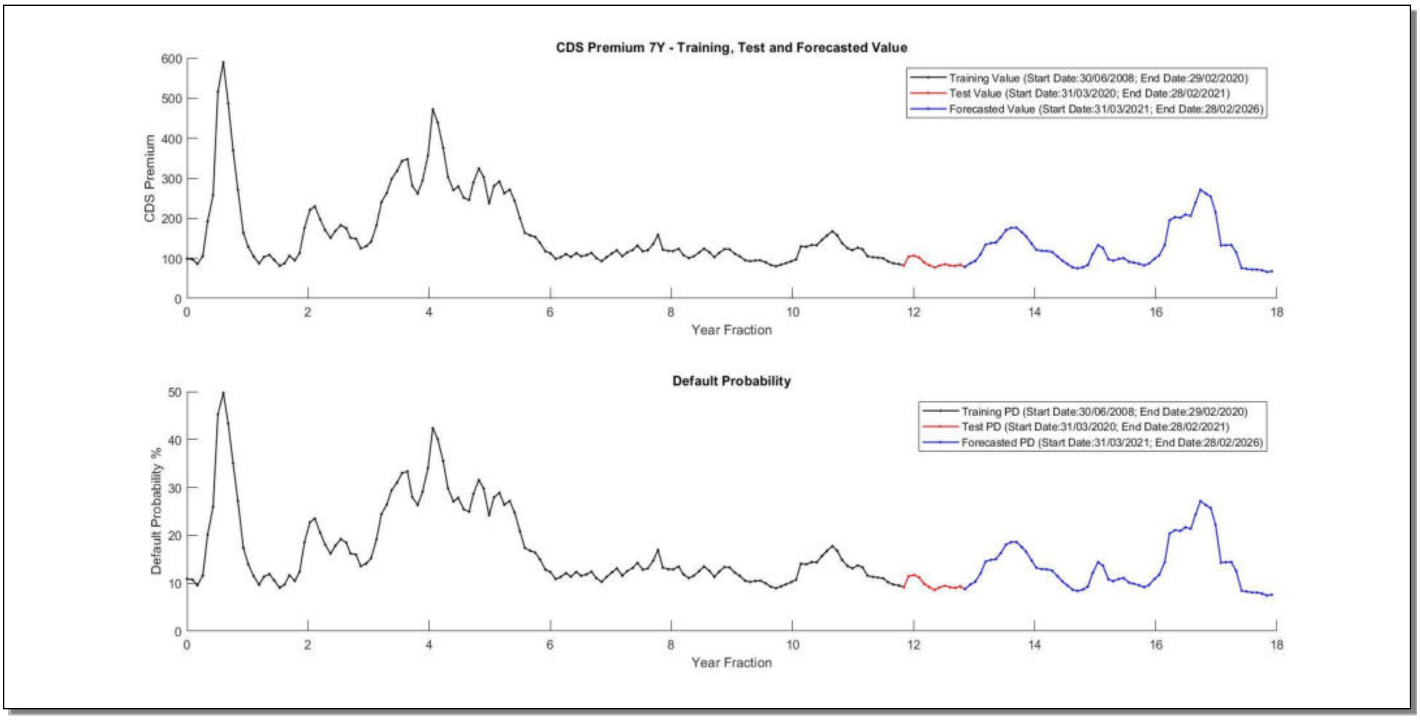


Figure 1 CDS premium (tenor: 7 years) and Default Probability

The further graphs below represent the surfaces of the CDS Premiums and of the Default Probability: the fractions of year are represented on the x-axis, the different tenors are represented on the y-axis and the different premiums are on the z-axis (Figure 2) and the corresponding implied probabilities (Figure 3).

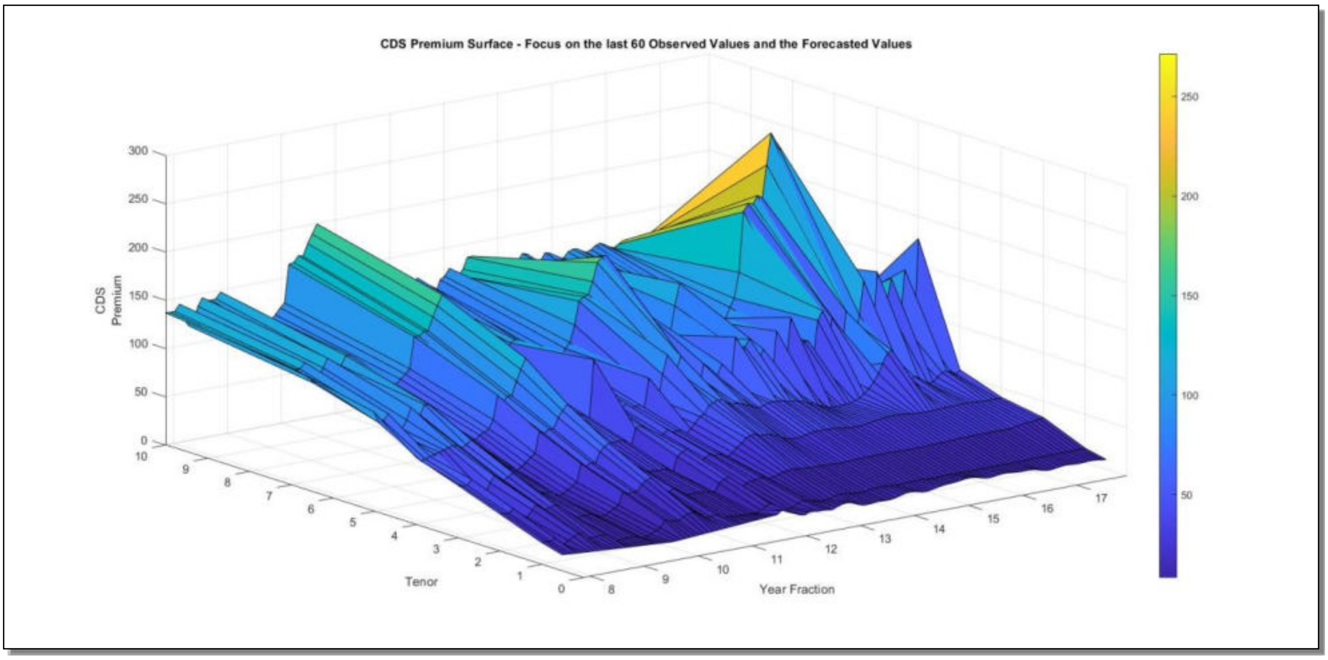


Figure 2 Observed and Forecasted CDS premium surface

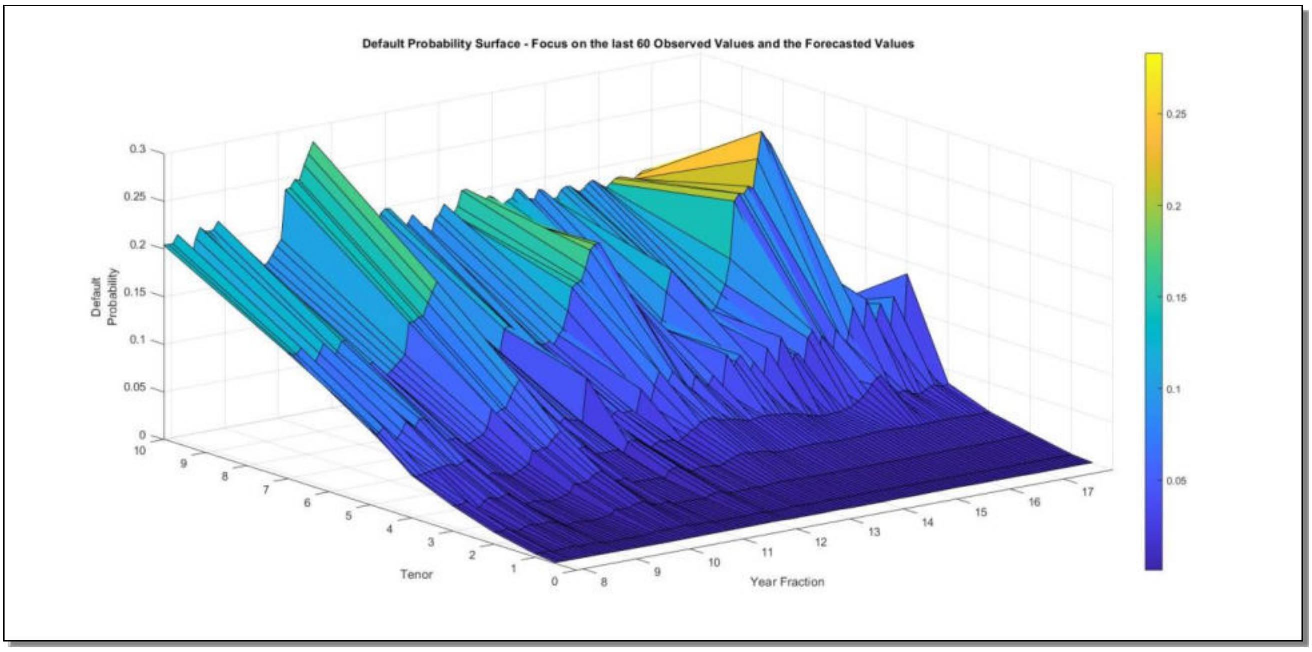


Figure 3 Default Probability surface obtained with the CDS method

3) Current and prospective estimation of the Probability of Default using listed bonds

The second approach for calculating the Default Probability consists in estimating the Z-Spread and using such figure as a proxy for a synthetic CDS premium. The Z-Spread indicates the excess return that a bond must give with respect to the risk-free rate and, usually, this yield differential is the compensation for a potential default of the bond issuer (Hull, 2015).

A clearer explanation of this risk transfer can be provided by considering a portfolio composed as follows:

A corporate bond with a 5-year maturity and a yield of 5%;

A long position on a 5-year Credit Default Swap, which has a cost of 250 basis points per year.

This portfolio is approximately equivalent to a long position in a risk-free instrument offering a rate of return of 2.5% per annum. The effect that the CDS has on the composition of the above-mentioned portfolio is to "transform" a risky instrument, such as a corporate bond, into a risk-free instrument. This is easily understood from the fact that, if the issuer of the bond does not incur a default, the return for the holder of the above-mentioned portfolio is equal to 2.5%, or 5% profit deriving from the corporate bond minus 2.5% which is the Credit Default Swap premium.

On the other hand, if the issuer were to go bankrupt, the investor would have a profit of 2.5% until the event of default and the full notional capital would be repaid thanks to the CDS. Furthermore, this capital can be reinvested at the risk-free rate for the time between the default and the maturity of the security.

In theoretical terms, the spread of a T-year CDS (s), should be equal to the difference, in terms of yield, between a corporate bond with a maturity of T-years and a risk-free security with the same maturity. In mathematical terms, this can be expressed as follows:

$$s_T = y_T - r_T \quad (11)$$

Where:

s is the excess spread;

y is the yield of the corporate bond;

r is the return on the risk-free security.

T is the maturity.

If the above stated would not happen, arbitrage opportunities would arise.

3.1) Procedural example

In order to illustrate the calculation of the Z-Spread, a corporate bond issued on 18th May 2020 at a price equal to 99.308 was taken into consideration. The bond expires 6 years from the date of issue (18th May 2026) and pays an annual coupon of 1.25%. Based on the Moody's rating scale, this bond is rated Baa, which means that it is considered a security with a moderate credit risk. On the day our analysis was carried out, i.e., 26th February 2021, the bond's market price was 105.595. The goal is to price the bond with the characteristics listed above. The feature to be evaluated is the spread composed of the part related to liquidity risk plus the credit risk, however the bonds suitable for this purpose should be selected among the most liquid securities, which means that the liquidity risk can be considered negligible and, therefore, the spread is almost entirely associated with credit risk. In order to understand the impact of credit risk, a theoretical valuation model was constructed and, therefore, the bond was first priced at the risk-free rate. The zero rates were derived from the term structure of the 6-month Euribor which, typically, in the context of pricing a fixed-income instrument, is used as the best proxy for the risk-free rate. The formulation of cash flows is shown in Table 2

Payment Date	Year Fraction	Cash Flows	Zero Rate	Adjusted Zero Rate	Adjusted Discounted Factor	Net Present Value
18/05/2021	0.222	1.25	-0.516%	-0.516%	1.0011487	1.2514359
18/05/2022	1.222	1.25	-0.460%	-0.460%	1.0056497	1.2570621
18/05/2023	2.222	1.25	-0.454%	-0.454%	1.0101677	1.2627096
18/05/2024	3.222	1.25	-0.407%	-0.407%	1.0132148	1.2665185
18/05/2025	4.222	1.25	-0.348%	-0.348%	1.0148243	1.2685304
18/05/2026	5.222	101.25	-0.283%	-0.283%	1.0148966	102.7582784
Z-Spread	0	Dirty Price	109.0645349			
Value Date	26/02/2021	Accrued	0.9726027			
Issue Date	18/05/2020	Clean Price	108.0919322			

Table 2 Risk-free pricing of the corporate bond examined as of 26th February 2021

Where:

the Payment Dates are the dates on which the coupon of the bond is paid and, at maturity, the repayment of the notional;

the Year Fractions (YF) are the fractions of the year corresponding to the Payment Dates;

the Cash Flows (CF) are the cash flows generated by the bond, i.e., the coupons until 18th May 2025 and the Face Amount plus the coupon on 18th May 2026;

regarding the values of the Zero Rate (ZR) they have been interpolated following the corresponding year fractions of the Euribor term structure;

the Adjusted Zero Rate (AZR) is the risk-free rate adjusted according to the Z-Spread (Z);

the Adjusted Discounted Factor (ADF) is the adjusted discount factor which, in this case, is equal to the discount factor as there is no additional spread;

the Net Present Values (NPV) correspond to the present value of the cash flows;

the Dirty Price (DP) is the sum of the NPVs including the accrued interest (Accrued Interest);

the Clean Price (CP) is the sum of the NPVs, net of accrual;

the Valuation Date corresponds to the day on which the pricing of the corporate bond was made;

the Issue Date corresponds to the day on which the corporate bond was issued.

With the bond valued in these terms, i.e. at the risk-free rate (which is why the Z-Spread is equal to 0), a value of 108.092 is obtained.

Since the considered corporate bond is a liquid bond, the market provides a contribution which, in this case, is equal to 105.595, lower than the result of the theoretical model built above.

To take creditworthiness into account, the bond should be priced considering a spread. In order to identify the exact value of the Z-Spread that returns the market value of the corporate bond, the problem should be solved by means of a Goal Seeking method programmed in Matlab (Byrd *et al*., 1994).

Payment Date	Year Fraction	Cash Flows	Zero Rate	Adjusted Zero Rate	Adjusted Discounted Factor	Net Present Value
18/05/2021	0.222	1.25	-0.516%	-0.057%	1.0001276	1.2501595
18/05/2022	1.222	1.25	-0.460%	-0.001%	1.0000180	1.2500225
18/05/2023	2.222	1.25	-0.454%	0.004%	0.9999053	1.2498817
18/05/2024	3.222	1.25	-0.407%	0.052%	0.9983277	1.2479096
18/05/2025	4.222	1.25	-0.348%	0.111%	0.9953429	1.2441786
18/05/2026	5.222	101.25	-0.283%	0.176%	0.9908704	100.3256296

Z-Spread	45.852	Dirty Price	106.5677815
Value Date	26/02/2021	Accrued	0.9726027
Issue Date	18/05/2020	Clean Price	105.5951788

Table 3 Risk-Adjusted pricing of the examined corporate bond as of 26th February 2021

In order to estimate a synthetic term structure of CDS we need to consider several liquid bonds of the same issuer and to repeat the calculation for all days in which the market provides a price. Starting from the summary of this term structure of the CDS premiums, the Probabilities of Default are obtained with the technique described in the previous section. The 16 bonds selected for the considered issuer are shown in Table 4.

ID	Coupon	Maturity	Payment Frequency	Day Basis
A	2.625	22/11/2021	Annual	ACT/ACT
B	0.750	17/05/2022	Annual	ACT/ACT
C	3.250	10/07/2023	Annual	ACT/ACT
D	1.750	18/01/2024	Annual	ACT/ACT
E	0.625	19/09/2024	Annual	ACT/ACT
F	1.000	14/03/2025	Annual	ACT/ACT
G	3.750	12/09/2025	Annual	ACT/ACT
H	1.500	02/02/2026	Annual	ACT/ACT
I	1.250	18/05/2026	Annual	ACT/ACT
J	1.500	17/01/2027	Annual	ACT/ACT
K	1.625	17/05/2028	Annual	ACT/ACT
L	1.125	19/09/2028	Annual	ACT/ACT
M	3.625	29/01/2029	Annual	ACT/ACT
N	0.625	23/01/2030	Annual	ACT/ACT
O	2.000	18/05/2031	Annual	ACT/ACT
P	1.000	11/10/2034	Annual	ACT/ACT

Table 4 Corporate bonds used for the analysis

In this example as well, the probability of default was calculated both at time zero and prospectively. This required the use of NAR neural networks, to implement the various forecasting of the values (de Simon-Martin *et al.,* 2020).

Therefore, 27,750 different network architectures were tested for each single bond, for a total of 440,000, in order to select the 16 best networks used for forecasting the different Z-Spreads, whose architecture is shown in Table 5.

The criteria used for the selection of the most performing networks remain the same that we have previously described for Table 1.

Non Linear Auto Regressive Best Networks					Econometric Test (Boolean)		Quality Accuracy
					TRUE		≥0.95
Bond	Number of Delays	Neurons (First Layer)	Neurons (Second Layer)	Neurons (Third Layer)	In Sample RMSE	Out Sample RMSE	MEAN IN SAMPLE AND OUT SAMPLE RMSE
A	17	9	9	5	1.525955908	0.65624776	1.09110183
B	17	8	7	1	1.387777305	0.63832195	1.01304963
C	11	1	4	2	1.665444521	0.88516771	1.27530612
D	16	8	8	4	1.720082629	0.59825059	1.15916661
E	18	9	7	5	1.587440954	0.63539117	1.11141606
F	18	9	5	6	1.68540112	0.63272568	1.15906340
G	16	8	7	8	1.756757209	0.71813277	1.24244499
H	12	8	5	8	2.071197384	0.59747441	1.33433590
I	19	10	10	8	0.76682641	0.56315685	0.66499163
J	12	6	10	6	1.029262298	0.87700847	0.95313538
K	3	8	1	1	1.359849175	0.93094883	1.14540900
L	19	10	3	9	1.722871502	0.72863780	1.22575465
M	9	8	9	8	2.243363579	0.78521178	1.51428768
N	17	10	8	10	1.426791275	0.61658974	1.02169051
O	20	9	6	9	0.754451999	0.47453179	0.61449190
P	18	8	2	2	1.099202997	0.58227227	0.84073763

Table 5 Architecture of the Best networks: Z-spread method

By way of example, the graph of bond O is shown in Figure 4, in which, in the upper part, the different time series used and the respective forecasting of the synthetic CDS premiums are shown and, in the lower part, a detail of the forecasting and the last 44 observed values.

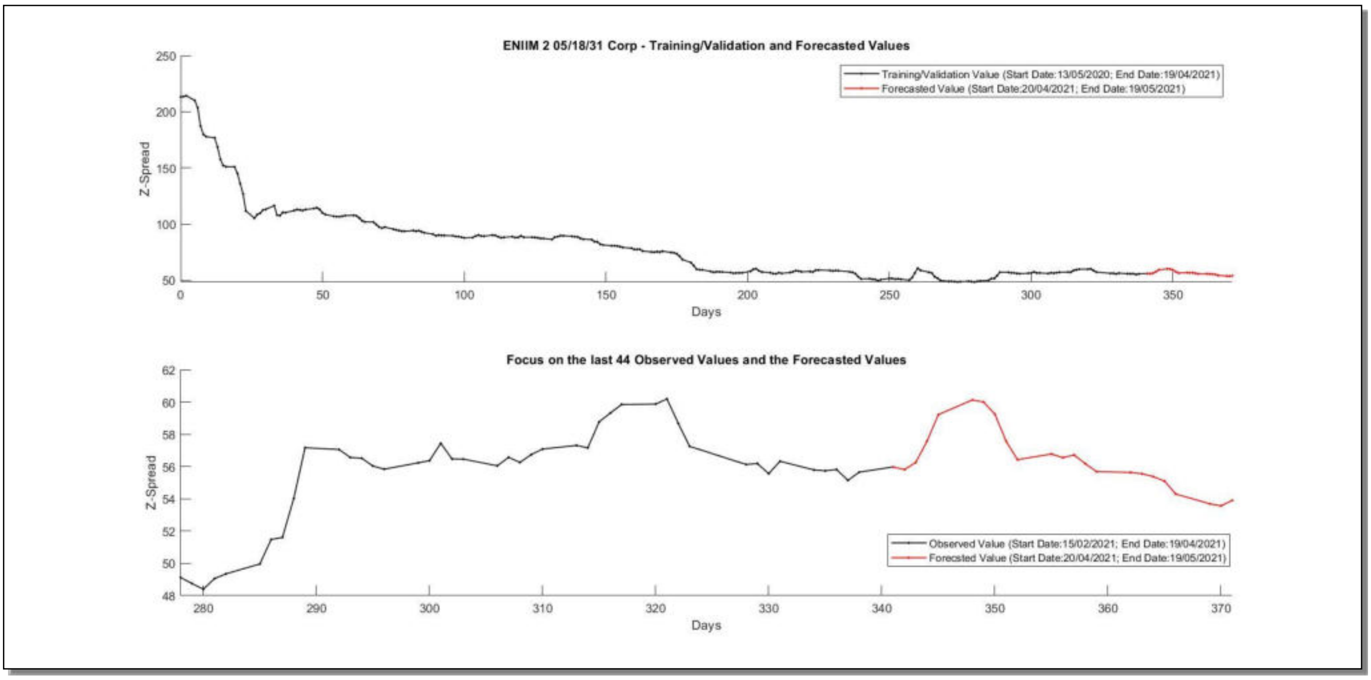


Figure 4 Training / Validation and Forecasting for Bond with ID: O

Once all the necessary forecasting for the prospective estimate of the default probability had been implemented, a further step was required which is not present in the operational example of the Credit Default Swaps. In fact, each single bond is characterized by its own time to maturity (as shown in Table 4).

It was thus necessary to regularize the Z-Spreads for the standard market tenors, that is: 6 months, 1 year, 2 years, 3 years and so on up to 12 years. Therefore, for each single day, the synthetic spread corresponding to each standard tenor of the term structure was interpolated across the bonds.

At this point, following this standardization procedure, the problem to be solved becomes like the case of the CDS.

The surfaces shown in Figures 5 and 6 show a summary of the results obtained.

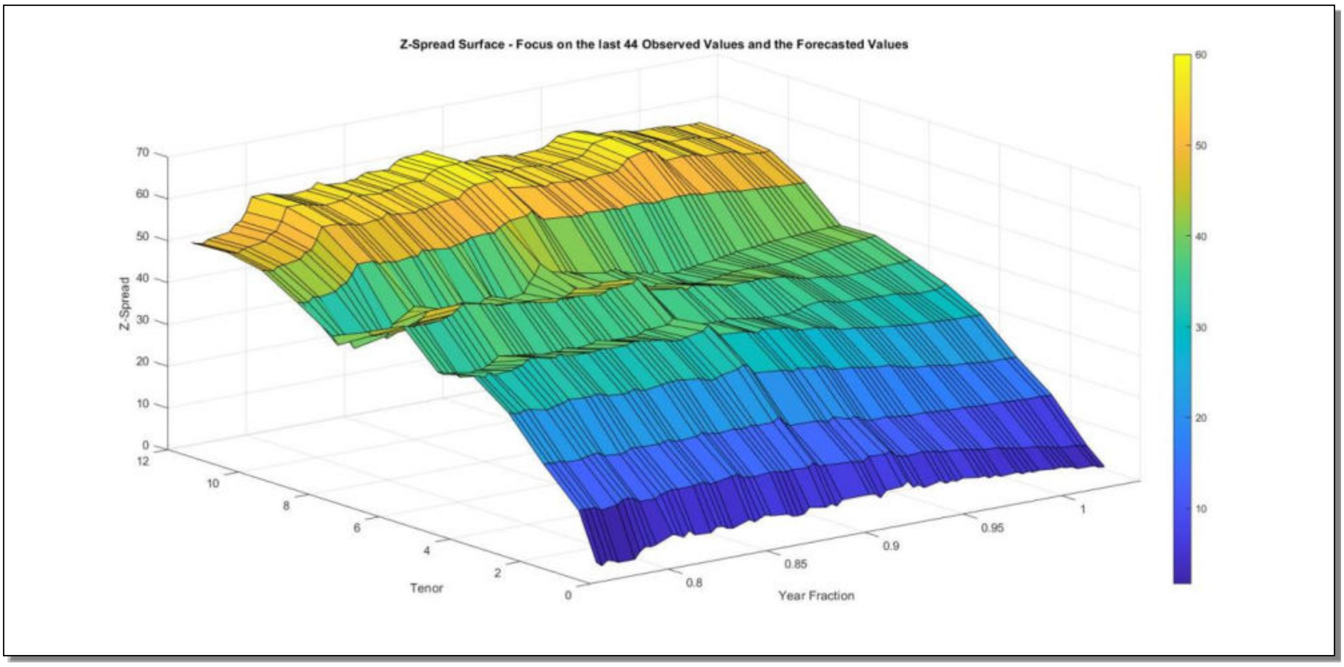


Figure 5 Surface of the implicit spreads on the analyzed bonds

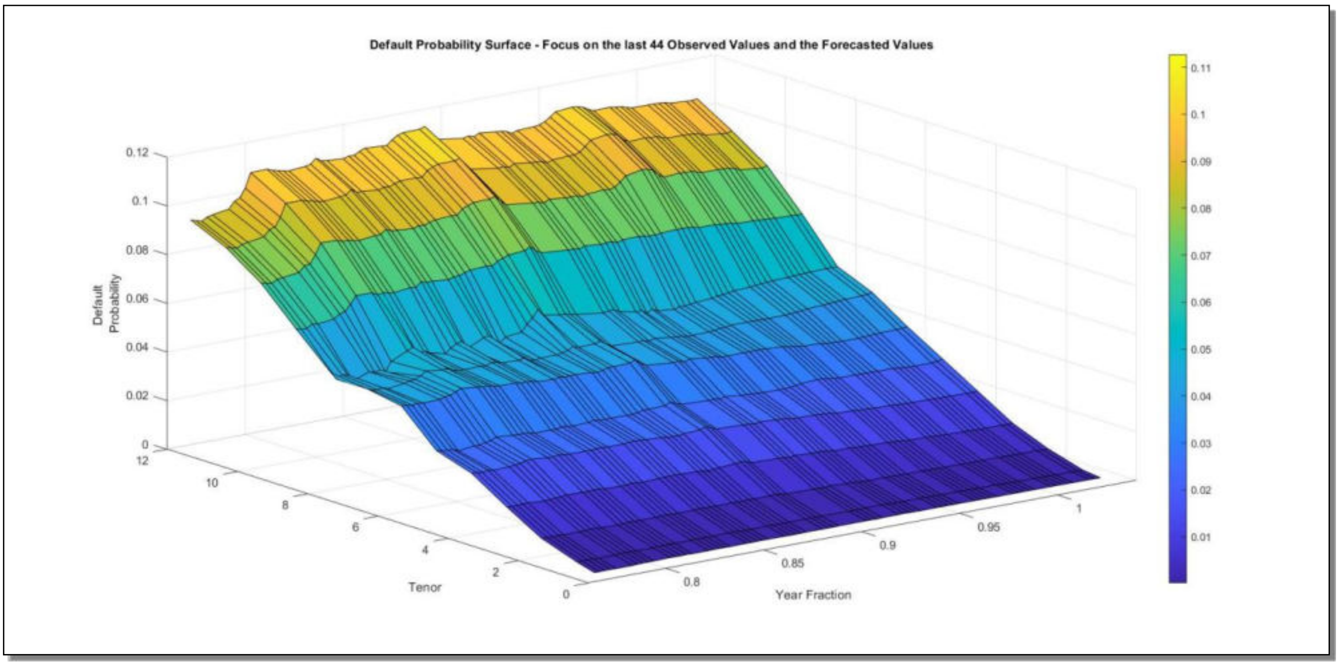


Figure 6 Default Probability surface obtained with the Z-spread method

4) Current and prospective estimation of the Probability of Default using the KMV method

If the company in question does not have listed Credit Default Swaps or listed corporate bonds on the markets, the KMV model (Kealhofer, Merton and Vasicek) can be used to estimate the probability of default (Bharath and Shumway, 2004).

KMV is a structural model usually implemented in credit risk management and many different versions of this model were successfully implemented (Agosto & Moretto, 2012). Its application is suggested in the estimation of the counterparty risk when market data are not available and the main source of information for the corporate analysis is the balance sheet. As a result, this approach is not suitable for a short-term forecasting ($T < 1$).

The KMV method is based on the equity price and on the balance sheet of the company in question. The assumptions underlying the Merton model can be divided into the following 4 sections:

The Debt is homogeneous and has a time to maturity equal to T;

The capital structure of the company in question is given by debt and equity. Consequently, it is true that: $V_A(t) = D(t) + V_E(t)$, where: $V_A(t)$ is the value of the assets at time *t*, $D(t)$ is the debt to be repaid and $V_E(t)$ is the value of the company's equity at time *t*;

The model assumes that value of the firm assets follows a Brownian geometric motion of the following type: $dV_A = \mu_A V_A dt + \sigma_A V_A dW_t$, where: μ_A is the expected instantaneous rate of return, σ_A is the volatility and dW_t is a Wiener process;

The hypothesis of perfect markets applies: there are no taxes; there are no restrictions on short selling; the market is completely liquid, i.e. investors can buy and sell any asset at market price; sellers and buyers have the same as risk free rate and such interest rate is constant over the reference time span.

Based on the previous assumptions, Robert Cox Merton proposed, in 1974, a model where a company's equity is an option on the assets of the company. Based on the structural model proposed (Merton, 1974), if $V_A(T) < D$ it is likely, at least in theory, that the company surely defaults at time T; in this case the equity at time T is equal to 0 ($V_E(T) = 0$). Likewise, if $V_A(T) > D$, then the company in question should repay its debt at time T and, in this case, the value of the equity at that same time will be equal to: $V_E(T) = V_A(T) - D$. In strictly mathematical terms, the amount of the company's equity at time T is given by the following pay-off:

$$V_E(T) = \max(V_A(T) - D, 0) \quad (12)$$

Given the previous statement, the analogy with a European-type option is clear: the value of the equity is similar to the payoff of a call option on the value of the assets with a strike price equal to the payment requested on the debt.

Under the assumptions of this model, the traditional Black-Scholes-Merton (BSM) model for option pricing (Black and Scholes, 1973) can be applied and, consequently, the following equations are valid:

$$V_E(t) = V_A(t)\phi(d_1) - \exp\big(-r(T-t)\big)D\phi(d_2) \quad (13)$$

$$d_1 = \frac{\ln\left(\frac{V_A(t)}{D}\right) + \left(r - \frac{1}{2}\sigma_A^2\right)(T-t)}{\sigma_A\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma_A\sqrt{T-t}$$

Where:

T is the time when the valuation is made;

σ_A is the asset volatility;

r is the risk-free rate at time T;

$V_A(0) - V_E(0)$ is the value of the debt at time 0;

$\phi(-d_2)$ is the risk neutral probability that the company will default on its debt at maturity T.

However, in order to calculate the default probability $\phi(-d_2)$, a further relationship is necessary since there are two unknown parameters in the above equations, namely the value and the volatility of the assets (respectively: V_A and σ_A). The other variables of the model are, on the other hand, directly observable or calculable: the risk-free rate (r) can be selected by referring to the Euribor term structure; the value of the debt (D) is directly observable from the financial statements of the company in question; the equity value (V_E) can be calculated by multiplying the number of shares of the company by their unitary market value; the volatility of the equity (σ_E) can be calculated using the implied volatility or through a traditional backward looking econometric approach such as the GARCH. Applying Ito's lemma, the stochastic dynamics that regulate the behavior of V_E:

$$dV_E = \mu_E V_E dt + \sigma_E V_E dW_t \quad (14)$$

can be re-written as:

$$dV_E = \left(\frac{1}{2}\sigma_A^2 V_A^2 \frac{\partial^2 V_E}{\partial V_A^2} + \mu_A V_A \frac{\partial V_E}{\partial V_A} + \frac{\partial V_E}{\partial t}\right)dt + \sigma_A V_A \frac{\partial V_E}{\partial V_A} dW_t \quad (15)$$

By comparing the terms of equations (14) and (15) we can derive the following relationship:

$$\sigma_E V_E = \sigma_A V_A \frac{\partial V_E}{\partial V_A} \quad (16)$$

Following the theory of the Black-Scholes-Merton model, the term $\frac{\partial V_E}{\partial V_A}$ corresponds to one of the so-called Greeks of a European option, in particular it corresponds to the sensitivity measure defined as Delta (Δ^E), and it is equal to: $\phi(d_1)$.

At this point, to obtain the unobserved values of V_A and σ_A, the following non-linear system of two equations and two unknowns has to be solved:

$$\begin{cases} f_1(V_E, \sigma_E) = V_A \phi(d_1) - \exp\big(-r(T-t)\big)\, D\phi(d_2) - V_E = 0 \\ f_2(V_E, \sigma_E) = \frac{V_A}{V_E}\phi(d_1)\sigma_A - \sigma_E = 0 \end{cases} \quad (17)$$

With $\frac{\partial f_1}{\partial V_A} = \phi(d_1) > 0$

This is because, just like the Delta in the BSM model framework, f_1 is an increasing function of V_A and this implies that $f_1(V_A)$ only has one solution. For the same reason, f_2 also has a unique solution.

Once the previous non-linear system has been solved numerically with Matlab, all the necessary data to calculate the default probability using $\phi(-d_2)$ are available:

$$DD = d_2 = d_1 - \sigma_A\sqrt{T} = \frac{\ln\left(\frac{V_A}{D}\right) + \left(r - \frac{1}{2}\sigma_A^2\right)(T)}{\sigma_A\sqrt{T}} \quad (18)$$

The term d_2 within this model is often indicated with the term Distance to Default (DD).

Considering now that the assets follow a Brownian geometric motion and, consequently, $V_A(t)$ is distributed as a log-normal with an expected value at time t equal to:

$$V_A(t) = V_A \exp\left\{\left(r - \frac{1}{2}\sigma_A^2\right)(T-t) + \sigma_A W_{T-t}\right\} \quad (19)$$

We can state that the default probability, $PD(T-t)$, for $t = 0$ can be calculated as follows (Löffler and Posch, 2011):

$$PD(T) = \Pr[V_A(T) < D] = \Pr\left[V_A \exp\left\{\left(r - \frac{1}{2}\sigma_A^2\right)T + \sigma_A W_T\right\} < D\right] = \quad (20)$$

$$= \Pr\left[W_T < \frac{\ln\left(\frac{D}{V_A}\right) - \left(r - \frac{\sigma_A^2}{2}T\right)}{\sigma_A}\right] = \Pr\left[Z < \frac{\ln\left(\frac{D}{V_A}\right) - \left(r - \frac{\sigma_A^2}{2}T\right)}{\sigma_A\sqrt{T}}\right] =$$

$$= \Pr\left[Z < -\frac{\ln\left(\frac{V_A}{D}\right) + \left(r - \frac{\sigma_A^2}{2}T\right)}{\sigma_A\sqrt{T}}\right] = \Pr[Z < -DD] = \phi(-DD)$$

4.1) Procedural example

Two operational examples of default probability calculation will be conducted using the KMV method. The first consists in calculating the measure of the default probability at time 0, that is taking as reference data those of the latest available financial statements of one of the main Italian energy companies, which is 31st December 2020, at the date of preparation of this paper.

Once all necessary data have been found from Bloomberg®, the issuer's default probability is calculated following the procedure illustrated in the previous section.

The second operational example consists in the calculation of the prospective default probability using different methodologies for forecasting data and for modeling the other parameters required for the calculation of the relevant figures.

The equity value was calculated by multiplying the stock price as of 31st December 2020 by the number of company shares on the same date, obtaining the data from Bloomberg®. The share price was 8.548, the number of shares in the company was 3,572,550,000 and, consequently, the equity value was equal to: V_E = 8.548 × 3,572,550,000 = 30,538,157,400 euros.

As regards the amount of debt, it is always obtained from the balance sheet and here it is equal to 31,704,000,000.

The term structure of the risk-free rate with tenor equal to 6 months (Euribor 6M) was used as the value for the risk-free rate and the value of the zero rate at one year is equal to -0.533%.

The last parameter required in order to set up the system of equations (17) is the volatility of the equity. This was estimated at 53.58% using the close-to-close method based on the daily returns of the share recorded in the past year and annualized with a factor of 252, i.e., the number of trading days within the considered year. (Haug, 2010).

Table 6 displays a summary of the results obtained

Parameters	VE	D	sigma E	r (RISK FREE RATE)	STOCK PRICE	NUM. OUTSTANDING	FIRM VALUE (VA)	ASSET SIGMA (SIGMA A)
Values in Millions of Euro	30,538.16	31,704.00	53.58%	-0.533%	8.548	3,572.55	62,385.93	26.32%
Default Probability	0.00775627%							

Table 6 Estimation of the current default probability using the KMV method

As a second step, the objective is to estimate the prospective default probability. In order to obtain this figure, we need to investigate which is the most reasonable technique to use to determine the behavior of the input data of the one-year model.

Regarding the value of the equity, the number of shares was assumed to be constant during the following year and the prediction of the share price was obtained through NAR. In order to project the value of the debt into the relevant time span, we assumed that it follows a Log-Normal distribution. According to this assumption, the percentage change in the value of the debt over a very short time span is normally distributed. Defining as:

μ_D the expected value of returns over a year

σ_D the volatility of the share price over a year.

The mean and the standard deviation over a time period of Δt are approximately $\mu_D \Delta t$ and $\sigma_D\sqrt{\Delta t}$. Thus, we can state that:

$$\frac{\Delta D}{D} \sim \phi(\mu_D \Delta t, \sigma_D^2 \Delta t) \quad (21)$$

Where: ΔD is the change in the value of debt over a period of time Δt, and $\phi(m, v)$ denotes a normal distribution with mean equal to m and variance equal to v.

The BSM model implies that:

$$\ln D_T - \ln D_0 \sim \phi\left[\left(\mu_D - \frac{\sigma_D^2}{2}\right)T, \sigma_D^2 T\right]$$

$$\ln \frac{D_T}{D_0} \sim \phi\left[\left(\mu_D - \frac{\sigma_D^2}{2}\right)T, \sigma_D^2 T\right]$$

$$\ln D_T \sim \phi\left[\ln D_0 + \left(\mu_D - \frac{\sigma_D^2}{2}\right)T, \sigma_D^2 T\right] \quad (22)$$

Where D_T is the value of the debt at a future time T and D_0 is the value of the debt at time 0. From this, we can conclude that the variable $\ln D_T$ is normally distributed and consequently, D_T is log-normally distributed. The mean of $\ln D_T$ is: $\ln D_0 + \left(\mu_D - \frac{\sigma_D^2}{2}\right)T$ and the standard deviation is: $\sigma_D\sqrt{T}$.

Three different scenarios were constructed: the best scenario with a confidence level of 50%, the average scenario with a confidence level of 75% and the worst scenario with a confidence level of 99%. Table 7 shows different scenarios constructed for the forecasted estimation of the debt.

CONFIDENCE LEVEL	Z	DEBT (LOWER BOUND)	DEBT (UPPER BOUND)
50.00%	0.0100	32629.41	32695.94
60.00%	0.2500	31841.50	33504.99
75.00%	0.6745	30494.18	34985.33
95.00%	1.9600	26752.03	39879.18
99.00%	2.5760	25125.22	42461.28

Table 7 Debt simulation scenarios

To obtain a prospective estimate of the capital volatility (σ_E), we proceeded using a GARCH (1,1) to model such value. However, before proceeding with the use of this econometric model, a statistical test was preliminarily used, the Ljung-Box Q-Test, to verify the presence of heteroskedasticity in the relevant historical series, i.e. the time series of the stock prices. (Tsay, 2010).

The Ljung-Box Q-Test (Ljung and Box, 1978) is defined as follows

H_0: the data are independently distributed (i.e., the correlations, ρ in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process). In mathematical terms $\rho_1 = \rho_2 = \cdots = \rho_n = 0$

H_a: the data are not independently distributed; they exhibit serial correlation.

The test statistics is:

$$Q = N(N+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{N-k} \quad (23)$$

Where N is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k and h is the number of lags being tested. Under H_0 the statistic Q asymptotically follows a $\chi^2_{(h)}$

The econometric test applied on the residuals of the time series indicates that the evidence is insufficient to reject the null hypothesis of no residual autocorrelation through 20 lags, consequently it is reasonable to use a GARCH type model.

The GARCH (Generalized AutoRegressive Conditional Heteroschedasticity) is a model of generalized auto regressive conditioned heteroskedasticity presented by (Bollerslev, 1986). The more general expression of the GARCH(p,q) model however evaluates the σ_n^2 starting from the p observations of u^2 and from the most current estimates of the variance rate q. The GARCH (p, q) formula can be generalized as follows:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{q} \alpha_i u_{n-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{n-i}^2 \quad (24)$$

It should be emphasized that, by setting p = 0, the result obtained is the expression of the ARCH (q). By setting $\omega = V_L\gamma$ in the GARCH equation, the estimation model can be rewritten as follows:

$$\sigma_n^2 = \omega + \sum_{i=1}^{q} \alpha_i u_{n-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{n-i}^2 \quad (25)$$

The above expression is generally used in order to estimate the parameters since, once ω, α and β are known, we can calculate γ as a difference: $\gamma = 1 - \alpha - \beta$. The Long-Term Variance, on the other hand, is equal to the ratio: $\frac{\omega}{\gamma}$.

The equation representing the GARCH (1,1) is the following:

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (26)$$

Where V_L is the long-term variance.

A fundamental condition is given by the fact that the sum of the weights must be equal to one: $\alpha + \beta + \gamma = 1$. The parameters (1,1) of GARCH (1,1) indicate that the calculation of σ_n^2 is focused on the most recent observation of the u^2 and on the closest estimate of the variance rate. It is required that $\alpha + \beta < 1$ since, otherwise, the weight assigned to the Long-Term Variance would become negative. The model described here recognizes that, over time, the variance tends to converge towards a long-term average level (V_L) with an associated weight equal to $\gamma = 1 - \alpha - \beta$. The final stage, for the realization of the GARCH (1,1) model, is the estimation of the required parameters starting from the historical price series. The most common method for estimating them is that of the maximum likelihood estimation, MLE (Maximum Likelihood Estimation). The first step for the estimation of the model parameters consists of defining the estimated variance for day i as: $v_i = \sigma^2$ and assume that u_i follows a normal conditional probability distribution. The maximum likelihood function (L) to be maximized with respect to the model parameters is given by (Francq and Zakoian, 2010):

$$L = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v_i}} \exp\left(-\frac{u_i^2}{2v_i}\right)$$

Applying the logarithm to the previous equation, the maximum points of the function do not change, but the calculations are simplified:

$$L = \sum_{i=1}^{m} \left[-\ln(v_i) - \frac{u_i^2}{v_i}\right] = \sum_{i=1}^{m} L_i \quad (27)$$

The parameters that allow the maximization of L were found using a gradient descent algorithm (Byrd *et al.*, 1994) programmed in Matlab around the optimum zone (Figure 7).

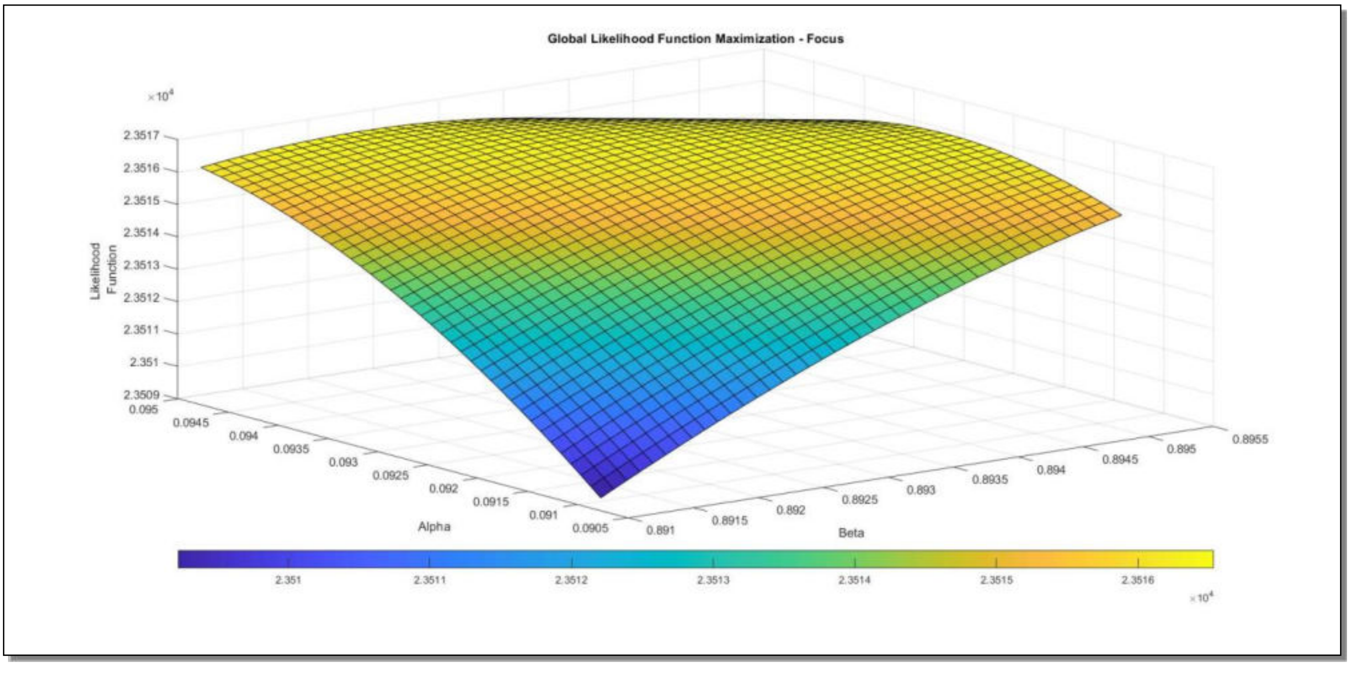


Figure 7 The detail of the surface of the Log-Likelihood Function around the optimum zone

Figure 8 shows the term structure of the volatility calculated from these, in accordance with the equation (Hull, 2015):

$$\sigma(T) = \sqrt{252\left(V_L + \frac{1-\exp(-aT)}{aT}[V(0) - V_L]\right)} \quad (28)$$

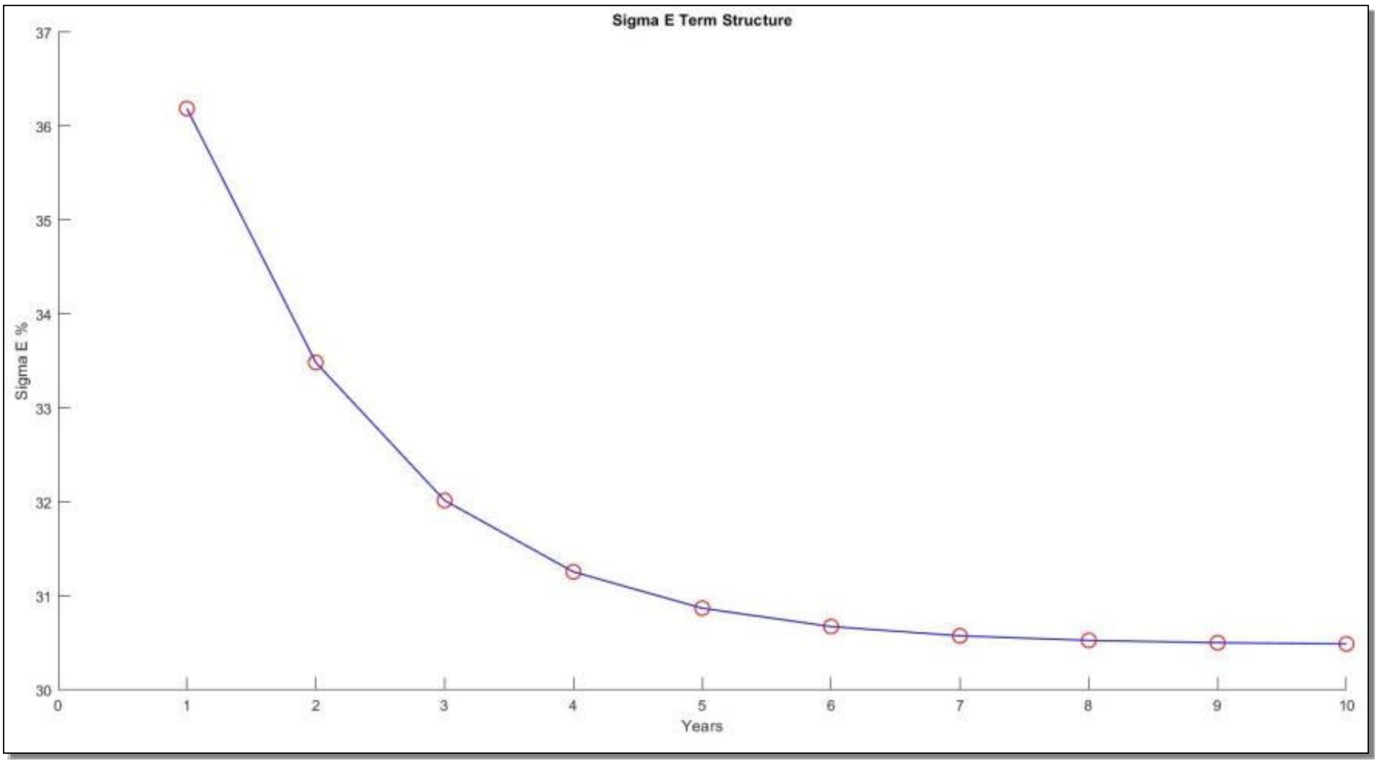


Figure 8 The term structure of the volatility estimated by GARCH (1,1)

As regards the rates used for the calculation of the prospective default probability, they were chosen by calculating the forward rate between 1 year and 2 years ($F_{1Y,2Y}$) and the zero rates ($r_{0Y,1Y}$ and $r_{0Y,2Y}$) implied by the term structure of the 6-month Euribor at the date of analysis:

$$F_{1Y,2Y} = \frac{(1+r_{0Y,2Y})^2}{(1+r_{0Y,1Y})} - 1 = \frac{(1-0.00524)^2}{(1-0.00533)} - 1 = -0.515\% \quad (29)$$

Once all the future estimates of the parameters required for the application of the KMV model were obtained, they were used for the estimate of the prospective default probability following the same procedure.

Table 8 shows the results obtained in calculating the relevant figure for the three different debt scenarios previously described.

TIME 1 VALUES (31/12/2022) - Values in Millions of Euro			
Parameters	**Best Scenario**	**Average Scenario**	**Worst Scenario**
VE	30,023.71		
D	32,695.94	34,985.33	42,461.28
sigma E	36.19%		
r (RISK FREE RATE)	-0.515%		
STOCK PRICE	8.404		
NUM. OUTSTANDING	3,572.55		
FIRM VALUE (VA)	**62,888.24**	**65,189.41**	**72,707.77**
ASSET SIGMA (SIGMA A)	**17.28%**	**16.67%**	**14.95%**
Default Probability	**0.01215340%**	**0.01476090%**	**0.02425420%**

Table 8 Forecasted default probability using the KMV method

5) Applications of the previous methodologies on the companies included in the EuroStoxx 50

The goal of this section is to apply the analyzed methodologies to the companies making up one of the most famous European indices: the Euro Stoxx 50 (SX5E Index) in order to calculate, for each of them, the default probability at time 0 and the same figure as a forecasted value.

For each of the companies included in the reference index, the more appropriate forecasting technique has been selected.

For the calculation of the Probability of Default, the hierarchical principle previously discussed in this report was followed, namely: in the presence of Credit Default Swaps listed on the market, the method used will be that of the CDS; if there are no listed CDSs, but there are corporate bonds listed on the markets, the methodology used will be that of the Z-Spread and finally, in the absence of listed CDS or corporate bonds, the methodology will be that of the KMV.

Tables 9 and 10 show all 50 companies in the index with a progressive ID aimed at identifying them easily later and, in the last column, the forecasting technique that was used to obtain the Default Probability.

In particular, number 1 indicates that the methodology used will be that of the CDS, number 2 indicates that the methodology will be that of the Z-Spread and, finally, number 3 indicates that the methodology will be that of the KMV.

SX5E INDEX				
FORECASTING TECHNIQUE: 1 (CDS METHOD) / 2 (Z-SPREAD METHOD) / 3 (KMV METHOD)				
ID	Ticker (Bloomberg)	Name	Stock Price (30/06/2021)	Forecasting Technique
01	OR FP Equity	L'Oreal SA	375.80	**3**
02	DG FP Equity	Vinci SA	89.99	**1**
03	ASML NA Equity	ASML Holding NV	579.40	**2**
04	SAN SQ Equity	Banco Santander SA	3.22	**1**
05	PHIA NA Equity	Koninklijke Philips NV	41.79	**1**
06	TTE FP Equity	TotalEnergies SE	38.16	**1**
07	AI FP Equity	Air Liquide SA	147.66	**1**
08	CS FP Equity	AXA SA	21.39	**1**
09	BNP FP Equity	BNP Paribas SA	52.87	**1**
10	BN FP Equity	Danone SA	59.37	**1**
11	EL FP Equity	EssilorLuxottica SA	155.64	**3**
12	VIV FP Equity	Vivendi SE	28.33	**1**
13	MC FP Equity	LVMH Moet Hennessy Louis Vuitton SE	661.30	**1**
14	KER FP Equity	Kering SA	737.00	**1**
15	AMS SQ Equity	Amadeus IT Group SA	59.32	**3**
16	SAF FP Equity	Safran SA	116.92	**3**
17	AD NA Equity	Koninklijke Ahold Delhaize NV	25.07	**1**
18	IBE SQ Equity	Iberdrola SA	10.28	**1**
19	INGA NA Equity	ING Groep NV	11.14	**1**
20	LIN GY Equity	Linde PLC	243.35	**2**
21	PRX NA Equity	Prosus NV	82.47	**3**
22	ITX SQ Equity	Industria de Diseno Textil SA	29.71	**3**
23	KNEBV FH Equity	Kone Oyj	68.80	**3**
24	FLTR ID Equity	Flutter Entertainment PLC	152.70	**3**
25	ISP IM Equity	Intesa Sanpaolo SpA	2.33	**1**

Table 9 Companies belonging to the Euro Stoxx 50 index at the analysis date - first part

SX5E INDEX				
FORECASTING TECHNIQUE: 1 (CDS METHOD) / 2 (Z-SPREAD METHOD) / 3 (KMV METHOD)				
ID	Ticker (Bloomberg)	Name	Stock Price (30/06/2021)	Forecasting Technique
26	ENI IM Equity	Eni SpA	10.27	**1**
27	ENGI FP Equity	Engie SA	11.55	**1**
28	ABI BB Equity	Anheuser-Busch InBev SA/NV	60.81	**1**
29	ADYEN NA Equity	Adyen NV	2060.50	**3**
30	SAN FP Equity	Sanofi	88.36	**1**
31	ENEL IM Equity	Enel SpA	7.83	**1**
32	IFX GY Equity	Infineon Technologies AG	33.82	**2**
33	SU FP Equity	Schneider Electric SE	132.68	**3**
34	ALV GY Equity	Allianz SE	210.30	**1**
35	AIR FP Equity	Airbus SE	108.44	**1**
36	BAYN GY Equity	Bayer AG	51.21	**1**
37	BMW GY Equity	Bayerische Motoren Werke AG	89.31	**1**
38	CRH ID Equity	CRH PLC	42.50	**2**
39	BAS GY Equity	BASF SE	66.44	**1**
40	SIE GY Equity	Siemens AG	133.62	**1**
41	VOW3 GY Equity	Volkswagen AG	211.20	**1**
42	MUV2 GY Equity	Munich Re	230.95	**1**
43	SAP GY Equity	SAP SE	118.84	**2**
44	RI FP Equity	Pernod Ricard SA	187.20	**1**
45	ADS GY Equity	Adidas AG	313.90	**3**
46	DTE GY Equity	Deutsche Telekom AG	17.81	**1**
47	DPW GY Equity	Deutsche Post AG	57.36	**1**
48	DAI GY Equity	Daimler AG	75.30	**1**
49	DB1 GY Equity	Deutsche Boerse AG	147.20	**2**
50	VNA GY Equity	Vonovia SE	54.52	**2**

Table 10 Companies belonging to the Euro Stoxx 50 index at the analysis date - second part

In the following three sub-sections, the relevant figure, i.e. the Default Probability, will be calculated for each of the companies included in the index as of 30th June 2021.

5.1) Use of the CDS method

For each of the individual companies the Default Probability was calculated at time 0 and prospectively. To calculate the relevant forecasted figure, it was necessary, as seen before, to use artificial neural networks to forecast the CDS premiums.

NARs and the network training methodology already described were used. The only difference is the fact that the range in which to look for the best network architecture has been reduced. In the search for the best network, the number of Lags was varied from 1 to 15, the number of neurons in the first layer from 1 to 20, the number of neurons in the second layer from 0 to 20 and the third layer was not considered, thus leaving the number of neurons constant at 0.

The reason for this change in the search for the best network is the aim at reducing the computational time taken by the machine to select the best network.

Table 11 displays the result of applying the CDS method to the SX5E Index companies.

In particular, the following figures are shown: the Default Probability at time 0, the Forecasted Default Probability, and the architecture of the best network used for forecasting.

SX5E INDEX					
METHOD 1 - CDS METHOD					
ID	Ticker (Bloomberg)	Name	Time 0 Default Probability	Forecasted Default Probability	Best Network [Lag; 1st Layer; 2nd Layer; 3rd Layer]
02	DG FP Equity	Vinci SA	0.00141566	0.00181783	[12; 8; 11; 0]
04	SAN SQ Equity	Banco Santander SA	0.00167316	0.00071439	[15; 17; 6; 0]
05	PHIA NA Equity	Koninklijke Philips NV	0.00099950	0.00105345	[6; 13; 6; 0]
06	TTE FP Equity	TotalEnergies SE	0.00141566	0.00131872	[2; 9; 8; 0]
07	AI FP Equity	Air Liquide SA	0.00108275	0.00121097	[11; 8; 6; 0]
08	CS FP Equity	AXA SA	0.00167491	0.00161944	[2; 17; 17; 0]
09	BNP FP Equity	BNP Paribas SA	0.00230348	0.00281530	[4; 9; 0; 0]
10	BN FP Equity	Danone SA	0.00154048	0.00179698	[14; 12; 5; 0]
12	VIV FP Equity	Vivendi SE	0.00309137	0.00407886	[12; 3; 18; 0]
13	MC FP Equity	LVMH Moet Hennessy Louis Vuitton SE	0.00149472	0.00126877	[9; 9; 13; 0]
14	KER FP Equity	Kering SA	0.00140343	0.00400380	[14; 13; 4; 0]
17	AD NA Equity	Koninklijke Ahold Delhaize NV	0.00156544	0.00698189	[14; 8; 6; 0]
18	IBE SQ Equity	Iberdrola SA	0.00204525	0.00185069	[5; 18; 9; 0]
19	INGA NA Equity	ING Groep NV	0.00238968	0.00488713	[10; 20; 19; 0]
25	ISP IM Equity	Intesa Sanpaolo SpA	0.00456957	0.00452180	[5; 6; 12; 0]
26	ENI IM Equity	Eni SpA	0.00248554	0.00190261	[9; 12; 10; 0]
27	ENGI FP Equity	Engie SA	0.00221837	0.00325333	[7; 12; 7; 0]
28	ABI BB Equity	Anheuser-Busch InBev SA/NV	0.00388259	0.00240518	[10; 7; 19; 0]
30	SAN FP Equity	Sanofi	0.00106194	0.00277703	[12; 15; 12; 0]
31	ENEL IM Equity	Enel SpA	0.00218133	0.00243555	[13; 6; 4; 0]
34	ALV GY Equity	Allianz SE	0.00121505	0.00167887	[4; 8; 15; 0]
35	AIR FP Equity	Airbus SE	0.00288598	0.00137127	[3; 13; 17; 0]
36	BAYN GY Equity	Bayer AG	0.00262911	0.00249248	[9; 20; 16; 0]
37	BMW GY Equity	Bayerische Motoren Werke AG	0.00178762	0.00315165	[14; 14; 11; 0]
39	BAS GY Equity	BASF SE	0.00135703	0.00215369	[11; 2; 20; 0]
40	SIE GY Equity	Siemens AG	0.00166906	0.00146033	[6; 13; 5; 0]
41	VOW3 GY Equity	Volkswagen AG	0.00317865	0.00237291	[10; 20; 8; 0]
42	MUV2 GY Equity	Munich Re	0.00132675	0.00200233	[14; 13; 13; 0]
44	RI FP Equity	Pernod Ricard SA	0.00151779	0.00068621	[14; 19; 0; 0]
46	DTE GY Equity	Deutsche Telekom AG	0.00205281	0.00224608	[14; 17; 5; 0]
47	DPW GY Equity	Deutsche Post AG	0.00091625	0.00248906	[13; 9; 18; 0]
48	DAI GY Equity	Daimler AG	0.00237785	0.00044437	[8; 15; 18; 0]

Table 11 Application of the current and prospective version of the CDS method on the SX5E Index

5.2) Use of the Z-Spread method

The companies included in the index, for which Credit Default Swaps are not available, but listed corporate bonds are available, are considered here. In this case, as before, the Default Probability was also both calculated at time 0 and forecasted. In order to calculate the forecasted measure, in this operational example it was also necessary to carry out a forecasting of the Z-spreads using the NARs. The procedure followed to search for the best networks is completely similar to the one we already discussed. Table 12 shows the result of the application of the Z-Spread method on the index companies which do not have listed CDSs on the markets but do have listed corporate bonds both at time 0 and prospectively, as well as the architecture of the best network used to the different forecasts.

5.3) Use of the KMV method

Lastly, we considered the companies in the index which do not have either listed Credit Default Swaps or listed corporate bonds, therefore the available balance sheet data has been used for calculating the probability of default using the KMV method has been used. To calculate the forecasted figure, it was necessary to use the different methods described in section 4.1 to obtain the requested forecasted values. In particular, the neural networks were used for the forecasting of the different stock prices in order to calculate the forecasted equity value (V_E), the Log-Normal distribution was used to estimate the different forecasted debt scenarios and finally a GARCH (1,1) was used to obtain the forecasted value of the capital volatility (σ_E). Table 13 shows the result of applying the KMV method. In this case, the various Default Probabilities have not been included as the values obtained are extremely low, in line with the checks carried out with the DRSK risk assessment module available on Bloomberg®. The values highlighted in the table are instead the number of shares of the company, the value of the share (at time 0 and forecasted), the value of the equity (at time 0 and forecasted), the value of the debt at time 0 and the 3 different prospective scenarios, the value of the capital volatility (σ_E) at time zero and the same forecasted value obtained with the GARCH (1,1) (also highlighting the parameters α, β and ω of the GARCH), the value of the assets (V_A) at time 0 and in the three envisaged scenarios and, finally, the value of the asset volatility (σ_A) at time 0 and in the 3 different scenarios.

SX5E INDEX					
METHOD 2 - Z-SPREAD METHOD					
ID	Ticker (Bloomberg)	Name	Time 0 Default Probability	Forecasted Default Probability	Best Network [Lag; 1st Layer; 2nd Layer; 3rd Layer]
03	ASML NA Equity	ASML Holding NV	0.00177343	0.00592898	[15; 14; 6; 0]
20	LIN GY Equity	Linde PLC	0.00171860	0.00030275	[14; 9; 12; 0]
32	IFX GY Equity	Infineon Technologies AG	0.00363643	0.00187741	[13; 16; 9; 0]
38	CRH ID Equity	CRH PLC	0.00417472	0.00561251	[10; 12; 16; 0]
43	SAP GY Equity	SAP SE	0.00171348	0.00016373	[13; 7; 19; 0]
49	DB1 GY Equity	Deutsche Boerse AG	0.00157778	0.00053382	[15; 20; 3; 0]
50	VNA GY Equity	Vonovia SE	0.00358036	0.00822622	[11; 10; 3; 0]

Table 12 Application of the current and prospective version of the Z-spreads method on the SX5E Index

SX5E INDEX							
METHOD 3 - KMV METHOD (risk-free rate: - 0.432%)							
ID	Shares Outstanding	(Stock Price) [Forecasted S. P.]	(VE_0) [VE_1]	(D_0) [D_1] [Best; Medium; Worst]	(σ_{E0}) [σ_{E1}] {α; β; ω}	(VA_0) [VA_1] [Best; Medium; Worst]	(σ_{A0}) [σ_{A1}] [Best; Medium; Worst]
01	557.67	(375.80) [303.70]	(209,572.386) [169,363.13]	(2,451.60) [2,247.95; 4,148.29; 23,948.26]	(0.1804) [0.2428] {0.0745; 0.9105; 0.0000038}	(212,034.51) [171,620.81; 173,529.38; 192,415.07]	(0.1783) [0.2396; 0.2370; 0.2126]
11	441.75	(155.64) [142.21]	(68,753.97) [62,823.22]	(11,413.00) [14,917.95; 20,053.19; 46,756.78]	(0.1905) [0.2454] {0.0501; 0.9384; 0.0000029}	(80,216.38) [184,345.66; 189,503.14; 216,322.34]	(0.1633) [0.2231; 0.2170; 0.1901]
15	450.50	(59.32) [71.27]	(26,723.66) [32,105.78]	(5,783.00) [6,398.08; 7,333.09; 10,834.22]	(0.2557) [0.2986] {0.8558; 0.8949; 0.0000060}	(32,534.60) [38,531.62; 39,470.68; 42,987.00]	(0.2100) [0.2488; 0.2429; 0.2230]
16	426.26	(116.92) [112.67]	(49,838.32) [48,024.76]	(6,860.00) [7,486.4; 8,297.83; 11,139.12]	(0.3486) [0.3795] {0.0800; 0.9100; 0.0000059}	(56,731.47) [55,543.65; 56,538.59; 59,212.21]	(0.3062) [0.3281; 0.3234; 0.3078]
21	1616.29	(82.47) [87.36]	(133,295.44) [141,197.12]	(6,971.91) [6,579.74; 7,756.06; 12,417.90]	(0.2014) [0.1638] {0.1105; 0.7907; 0.0000096}	(140,297.53) [147,805.35; 148,986.76; 153668.78]	(0.1913) [0.1565; 0.1552; 0.1505]
22	3114.86	(29.71) [28.63]	(92,542.49) [89,186.65]	(6,089.00) [2,037.47; 7,643.29; 336,008.76]	(0.2322) [0.2859] {0.0670: 0.9140; 0.0000061}	(98,657.85) [91,232.94; 96,863.03; 426,649.88]	(0.2178) [0.2795; 0.2632; 0.0598]
23	518.71	(68.80) [72.39]	(35,687.25) [37,550.80]	(559.20) [584.61; 725.66; 1,346.93]	(0.2122) [0.2626] {0.0630; 0.9141; 0.0000066}	(36,548.87) [38,137.94; 38,279.60; 38,903.56]	(0.2089) [0.2586; 0.2576; 0.2535]
24	174.63	(153.09) [121.78]	(26,734.11) [21,266.17]	(3,834.16) [3,295.26; 7,339.21; 72,573.75]	(0.3014) [0.3128] [0.1004; 0.8407; 0.0000023]	(30,586.8) [24,575.73; 28,637.22; 94,154.56]	(0.2634) [0.2707; 0.2323; 0.0707]
29	30.45	(2,060.50) [1,213.42]	(62,742.23) [36,948.74]	(128.36) [174.93; 243.55; 627.91]	(0.3928) [0.6365] {0.4246; 0.4919; 0.0001371}	(62,871.15) [37,124.43; 37,193.34; 37579.37]	(0.3920) [0.6335; 0.6323; 0.6258]
33	567.07	(132.68) [132.37]	(75,238.85) [75,065.79]	(11,399.00) [10,713.88; 11,507.08; 14,116.47]	(0.2538) [0.3038] {0.0753; 0.9135; 0.0000047}	(86,583.04) [85,826.15; 86,622.79; 89,243.50]	(0.2205) [0.2657; 0.2633; 0.2555]
45	195.07	(313.65) [309.42]	(61,183.71) [60,357.93]	(6,279.00) [6,384.44; 8,540.15; 19,633.66]	(0.3039) [0.2998] {0.0449; 0.9281; 0.0000096}	(67,491.81) [66,770.07; 68,935.13; 80,076.77]	(0.2460) [0.2710; 0.2625; 0.2260]
VALUE OF: SHARES OUTSTANDING - VE - D - VA IN MILLIONS OF EUROS							

Table 13 Application of the KMV method in current and prospective version on the SX5E Index

6) Conclusions

The present paper illustrates the application of three different methods for estimating the Default Probability as a measure of the counterparty risk: the CDS method, the Z-Spread method and the KMV method (Kealhofer, Merton and Vasicek).

All the above-mentioned methodologies have been validated both by following the reference scientific literature and by means of the calculation modules made available by one of the main information providers used by professionals in the sector, including the YAS, DRSK and CDSW modules of Bloomberg®.

As already extensively discussed, this figure was estimated both at the current level, using the market data observed at the time of valuation (t_0), and as a forecasted measure, using simulations on the model's inputs to this end.

For the perspective values, we have conducted econometric tests (such as absence of autocorrelation errors), checked the goodness of fit on validation sets measured by R^2, conducted out-of-sample tests on different architectures created by varying the number of the main parameters of the NAR networks (neurons, layers, and lags).

In particular, the paper wants to focus on this specific aspect, providing a set of methodologies if the relevant figure is presented as a forecasted value and not in a current standpoint, in line with the growing attention paid by regulators on the so-called "Through the cycle" measurement.

The use of artificial neural networks plays therefore a primary role in the whole research: the forecasting obtained with the aid of dynamic ANNs proved to be robust in terms of econometric measures and allowed to obtain a robust estimate of the data required for the forecasting.

The above has proved valid, both for the practical examples carried out on one of the main gas and energy suppliers in our country, and for those conducted on the 50 companies included in the Euro Stoxx 50 (SX5E Index).

There is no lack of further insights on this matter. For example, the set of techniques illustrated could be applied to calculate the Default Probability of an issuer not belonging to the European area, but to the US area.

Another interesting potential application could be the construction of a committee machine which, by comparing the dynamic artificial neural networks with the traditional econometric methods aimed at forecasting, could choose, time after time, the most performing solution in terms of goodness of the estimate, thus improving the heterogeneity of forecasting methods (Bagnato *et al.,* 2021).

The authors want to thank the anonymous Referees whose suggestions helped improving this article.

Bibliography

[1] Agosto A., Giribone P. G. (2019), Artificial Intelligence in Risk Management: Machine Learning methods in application to market risk, AIFIRM (Italian Association of Financial Industry Risk Management) Conference, Milan, 11th April 2019

[2] Agosto A., Moretto E. (2012), "Exploiting default probabilities in a structural model with nonconstant barrier". *Applied Financial Economics*, 22(8), 667-679

[3] Arbib A.M. (2003), "The Handbook of brain theory and neural networks". The MIT press

[4] Bagnato M., Bottasso A., Giribone P. G. (2021). "Implementation of a Commitment Machine for an Adaptive and Robust Expected Shortfall Estimation". Frontiers in Artificial Intelligence: A.I. in Finance (https://doi.org/10.3389/frai.2021.732805)

[5] Beale, M. H., M. T. Hagan, and H. B. Demuth (2014). Neural Network Toolbox, User's Guide. The Mathworks Inc.

[6] Bharath, S.T. and Shumway, T (2004). "Forecasting Default with the KMV-Merton Model". Working Paper, The University of Michigan. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=637342)

[7] Black F. and Scholes M. (1973). "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy,* 81, 637-654

[8] Bollerslev T. (1986). "Generalized autoregressive conditional heteroscedasticity". *Journal of Econometrics,* 31(3), 307-327

[9] Bonini B., Caivano G., Cerchiello P., Giribone P. G. (2019) – "Intelligenza Artificiale: l'applicazione di Machine Learning e Predictive Analytics nel Risk Management" – XIV Position Paper AIFIRM (Associazione Italiana Financial Industry Risk Managers)

[10] Bottasso A., Giribone P. G., Martorana M. (2019). "Analisi e progettazione di un sistema di misure quantitative per il monitoraggio dei rischi finanziari delle garanzie di origine". *Risk Management Magazine*, 14(2), 27-45

[11] Byrd R., Nocedal, J.; Schnabel, R. B. (1994). Representations of Quasi-Newton Matrices and their use in Limited Memory Methods". *Mathematical Programming,* 63(4), 129–156

[12] Caligaris, O., P. G. Giribone, and S. Ligato (2015). "Applicazioni delle reti neurali feed-forward per la ricostruzione di superfici di volatilità". *AIFIRM Magazine* (Italian Association of Financial Industry Risk Managers), 10(1), 4-19

[13] Chollet F. (2018), "Deep Learning with Python". Manning Publications

[14] Cun, Y. L. et al. (1990). "Handwritten digit recognition with a back-propagation network". In: Advances in Neural Information Processing Systems, 396–404

[15] Decherchi, C. and P. G. Giribone (2020). "Stima prospettica delle misure finanziarie di rischio mediante reti neurali dinamiche: un'applicazione al mercato statunitense". *Risk Management Magazine,* 15(1), 50–69

[16] de Simon-Martin M., Bracco S., Rosales-Asensio E., Piazza G., Delfino F., Giribone P. G. (2020). "Electricity Spot Prices Forecasting for MIBEL by using Deep Learning: a comparison between NAR, NARX and LSTM networks" – International Conference on Environment and Electrical Engineering - EEEIC 2020, Technical Area: Regulation and Electricity Markets

[17] Fonseca P. G. and Lopez H. D. (2017), "Calibration of Machine Learning Classifiers for Probability of Default Modelling". James Finance, Crowd Process Inc. (https://doi.org/10.48550/arXiv.1710.08901)

[18] Francq C. and Zakoian J. M. (2010). "GARCH Models: Structure, Statistical Inference and Financial Applications". Wiley

[19] Freeman J. A. and D. M. Skapura (1992), *"Neural Networks Algorithms, Applications, and Programming Techniques".* Addison-Wesley, Massachusetts

[20] Géron, A (2019). “Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems”. O’Reilly
[21] Giribone P. G. (2020). “Seasonality Modeling through LSTM Network in Inflation-Indexed Swaps” – Data Analytics: The Ninth International Conference on Data Analytics - Special Session: FinTech Risk Management
[22] Giribone, P. G., Ligato S., and Penone F. (2018). “Combining robust Dynamic Neural Networks with traditional indicators for generating mechanic trading signals”, *International Journal of Financial Engineering,* 5(4). (https://www.worldscientific.com/doi/abs/10.1142/S2424786318500378)
[23] Giribone P. G., Ligato S., Raviola P. (2014). “Studio ed implementazione della metodologia Credit Value Adjustment in un framework di pricing automatico” – ASSIOM Forex Newsletter, 25-36
[24] Haug E. G. (2010). “The complete guide to Option Pricing Formulas”. McGraw-Hill
[25] Haykin, S. (1994). “Neural networks: a comprehensive foundation”. Prentice Hall PTR
[26] Hull J. C. (2015). Options, Futures and other Derivatives. Wiley
[27] Kolen J. F. and C. S. Kremer (2001). “A field guide to dynamical recurrent networks”. Wiley-IEEE Press
[28] Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). “ImageNet Classification with Deep Convolutional Neural Networks”. In: Proceedings of the 25th International Conference on Neural Information Processing Systems
[29] Löffler G., Posch P. N. (2011). “Credit risk modeling using Excel and VBA”, Wiley Finance
[30] Ljung G. M. and Box G. E. P. (1978). "On a Measure of a Lack of Fit in Time Series Models". *Biometrika*. 65(2), 297-303
[31] Merton R. C. (1974). “On the pricing of corporate debt: The risk structure of interest rates”. *The Journal of Finance,* 29(2), 449-470
[32] Principe, J. C. (2000). Artificial Neural Networks – The Electrical Engineering Handbook. CRC press
[33] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408
[34] Rojas R. (1996), "Neural Networks – a systematic introduction", Springer Verlag
[35] Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). “Learning representations by back-propagating errors”. In: Nature, 533–536
[36] Tsay R. S. (2010). “Analysis of Financial Time Series”. Wiley

Risk Management Magazine
Volume 17, Issue 2, May – August 2022

Direttore Responsabile (Editor in Chief)
Maurizio Vallino

Condirettore (Managing Editor)
Corrado Meglio

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. +39 389 6946315

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


RICORDI DI GIOVINEZZA
SMOG DI MILANO
2050
RUMORE DI CITTA'
IL MOTORE A SCOPPIO
EDITRICE BOOM
ACQUA CON MICRO PLASTICHE
OLIO DI PALMA
SILVA