RISK MANAGEMENT MAGAZINE

Vol. 16, Issue 1
January – April 2021

AIFIRM

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 16, Issue 1, January – April 2021

Direttore Responsabile (Editor in Chief)
Maurizio Vallino (AIFIRM General Director, Banca Carige, Italy)

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

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

Magazine printed on 30th April 2021

Peer review process on papers presented for publication

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

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

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

Editorial regulation

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

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

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

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

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

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

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

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

Processo di referaggio degli articoli proposti per la pubblicazione

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

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

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

Regolamento redazionale

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

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

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

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

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

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

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

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

Referees Year 2020

How to maximize profitability and minimize risk with dynamic stress testing

Ioannis Akkizidis (Wolters Kluwer)

Abstract

This paper offers a novel approach for optimizing banks' accounts and portfolios by using both static and dynamic simulation analysis to perform stress tests using several strategies and scenarios driven by exogenous shocks, such as the Covid-19 pandemic. The results of this analysis are explored and discussed using real cases in which dynamic analysis in stress scenarios has been applied to specific banking portfolios that may be impacted by Covid-19.

Keywords: static and dynamic analysis, Covid-19, stress testing, portfolio analysis, portfolio management strategy

DOI 10.47473/2020rmm0079

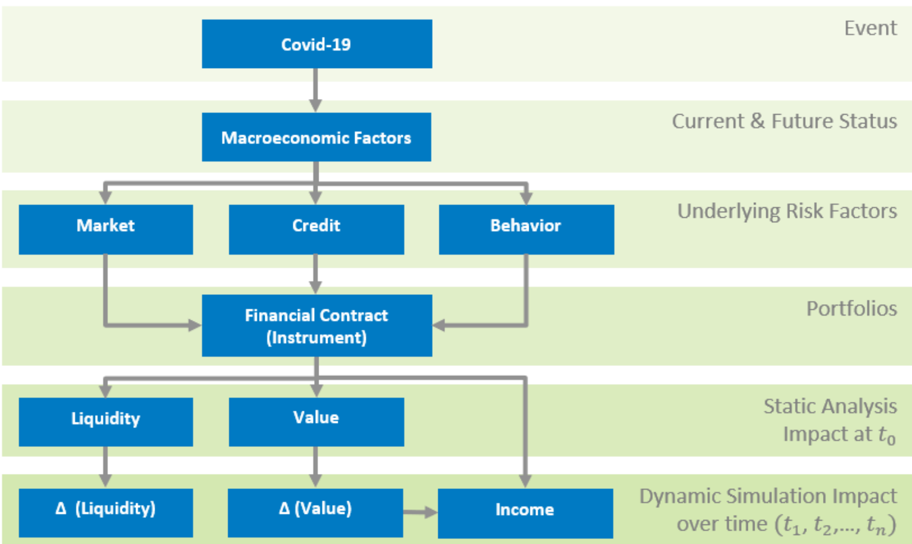
1. Introduction

The arrow of cause and effect seldom flies in one direction in a complex structure like the economy or financial system. It goes back and forth, and sometimes it heads off on the most unpredictable tangents.
That reality underpins much of modern banking regulation, particularly the ever more stringent stress testing procedures that banks must follow. Stress testing used to be a simple exercise in simulating potentially dangerous conditions: If X happens, what is the impact on Y, where X might be a sudden rise in interest rates or the default of one of a bank's major counterparties, and Y might be some aspect of the bank's capital position?
A sudden stress in the economy can develop from threats large and small, foreseen and unforeseen. Covid-19 is an obvious example of the large and unforeseen variety. Since the outbreak of the pandemic, authorities have been under pressure to adjust stress testing approaches to better assess the vulnerability of the banking sector and, if needed, adjust capital or liquidity positions at an individual institution. Supervisors have also drawn attention to the adjustments of assets, liabilities and off-balance-sheet positions over time.
The pandemic has persuaded credit institutions in Italy and other European Union countries, on their own initiative, to validate and adjust their models to reflect changes to underlying risk factors and enhance strategies related to their credit and investment portfolios. This is crucial for ensuring portfolio stability over multiple timeframes. Securing portfolio stability under stress scenarios helps evaluate the conditions under which a bank can continue providing credit and investing in the best mix of assets. The strategies it adopts must be robust so that it can optimize profitability under different forward-looking stress scenarios and react quickly when a certain scenario applies.

1.1 Methodology: a blueprint for building a stress test

Before considering the shock provided by Covid-19, a stress testing scenario must include certain elements to provide an institution with useful information for analysis and forecasting. *Figure 1* below depicts the broad outline of the design of a stress analysis. It should be a useful starting point in considering the impact of Covid-19.
Covid-19 has been influencing macroeconomic conditions that affect the underlying market, credit and behavior risk factors. These factors are the inputs in financial analysis and the design of the stress scenarios. They are used to calculate the elementary outputs in financial analysis, including expected cash flow, and the value and income of the financial contracts making up a bank's accounts and portfolios. Stressing the input factors as defined at the analysis date t_0, changes in the outputs indicate the risk of those factors to the liquidity and value of the existing portfolio. When the input factors are simulated and stressed over future time buckets, $t_1, t_2, \dots t_n$, banks evaluate the impact in terms of changes (Δ) to the liquidity, value and income of portfolios.

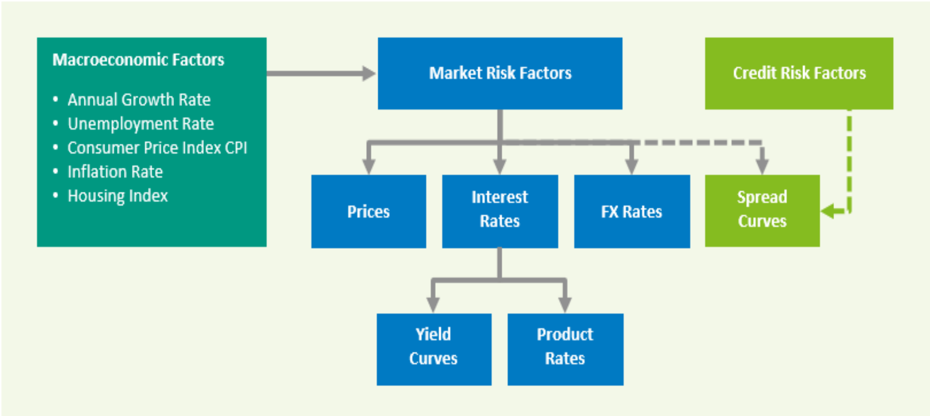
Figure 1: Framework in stress scenarios



2.Stressing market risk factors

Although macroeconomic conditions are considered underlying factors in stress scenarios, when calculating values and cash flow, only stress due to standard market risk factors is considered. As illustrated in *Figure 2* below, the main risk factors are prices, interest rates expressed as term structures of yield, as well as product rates and spread curves. The basis of credit spread curves is credit and counterparty risk factors driven by the markets. Stressing credit spreads will directly impact the values of expected income and market liquidity; the high probability of defaults and the resulting low credit ratings indicate cancellation of the contractual cash flows, credit losses and loss in value.

Figure 2: Market risk factors applied in stress scenarios

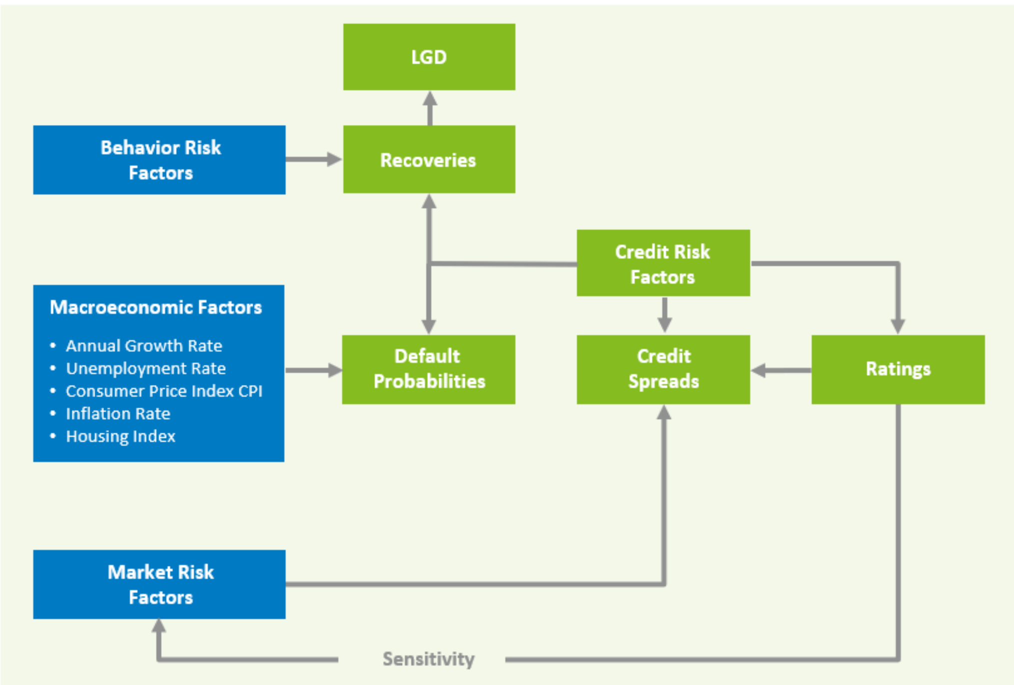


2.1 Stressing credit and counterparty risk factors

Credit exposures may become volatile due to market and counterparty behavior risks, resulting in credit losses. For instance, the value of a loan may change due to underlying market risk factors, such as credit spreads and the yield curve used to discount cash flows. Moreover, the availability of a credit line may change the exposure's value size. Yet most credit exposures are collateralized, fully or partially, and so banks must stress the collateral's value and the guarantor's rating status. Finally, stressing credit exposures can lead to systemic risk.

Stress scenarios also must model a counterparty's default probability. One way to measure the likelihood of default is to observe macroeconomic factors, such as the unemployment rate, together with counterparty-specific indicators, such as income. But the industry has chosen other paths, as this observation may become too complex and volatile, demanding an additional data layer that hardly can be up to date. One method that is often used is to estimate credit ratings and their probability of change over a certain time, for instance by using migration matrices. *Figure 3* below illustrates credit risk factors, together with their interdependencies with other factors considered in the design of credit stress scenarios.

Figure 3: Interdependencies among factors applied in stress scenarios



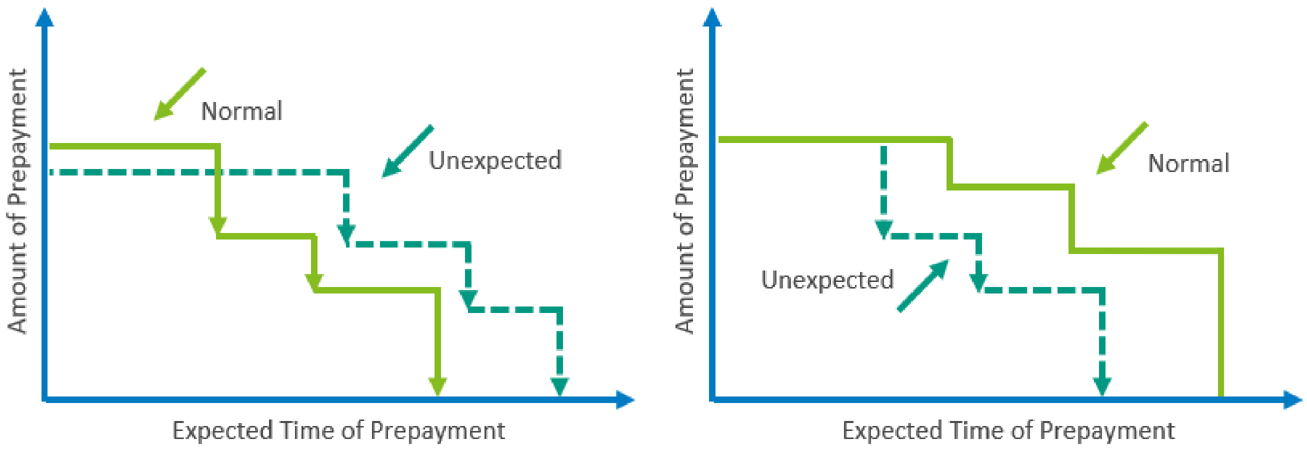
2.2 Stressing behavior risk

Behavioral elements are the most challenging factors to incorporate into stress tests because they rely on assumptions and historical data. Since behavior could have extreme domino effects in the financial industry, stress scenarios must contain behavioral elements, such as customer activity.

There are market-related behavior scenarios, such as the exercise of savings account withdrawals, replication, prepayment and sales. There is also credit risk-related behavior, referring mainly to expected recovery, use of credit lines and used at default. Behavior does not always follow market and credit stress conditions, however. This is often the case with non-maturity contracts, such as savings accounts.

Behavior also can depend on the structure of a financial contract, combined with stressed market and credit conditions. In the design of behavioral stress scenarios, banks stress two dimensions: time and amount. *Figure 4* below illustrates the behavior cases of withdrawals and remaining principal. Stress behavior directly impacts the liquidity and value of credit portfolios.

Figure 4: Prepayment behavior based on market conditions under normal and stress financial risk conditions



3. Strategies on the evolution of credit portfolios under stress conditions

Systemic shocks like Covid-19 may increase the threat to a financial institution's solvency. Institutions, therefore, must measure the strength of their assets, liabilities on and off the balance sheet, and all credit and investment portfolios, under stress scenarios considering such a shock. The results should be used to forecast the impact on profit and loss over multiple periods. The P&L analysis must factor in the strategies that an institution has applied. The best strategies encompass rollovers of current positions and maximize their evolution to assume a constant balance-sheet composition, while including new market conditions.

Financial instruments may be rolled over in a portfolio after they mature and, together with new positions, reflect the growth of the existing portfolios and the introduction of new accounts and portfolios in a new balance-sheet composition. One has to define the volume, type and pricing assumptions of these contracts, and consider the evolution of the underlying risk factors under expected and stress conditions.

Under stress conditions, institutions must consider two prominent cases:

i. Roll-down or runoff scenarios: Underlying risk factors as defined at the date of analysis are not only deterministically shocked but evolve along one or more specified market scenarios. The paths can be defined by the bank – dynamically – and could describe the condition in which the portfolio will change when a certain event happens.

ii. Going-concern scenarios: These encompass all risk factors related to market conditions and behavior change for existing accounts and the generation of new business. They can change in an interdependent way, as new business evolution depends on prevailing risk conditions. In other words, the assumption is that the company is keeping a similar investment alive, but against newer risk factor conditions.

The liquidity, value and income of existing and new financial contracts under stress scenarios must be analyzed dynamically, taking into account the interdependencies of multiple risk factors, to optimize future portfolios. The ideal liquidity situation may not give rise to the ideal credit risk situation, and the current conditions could either directly or indirectly influence one another, where the ideal portfolio will need to change against the current risk appetite of the bank across different risk types.

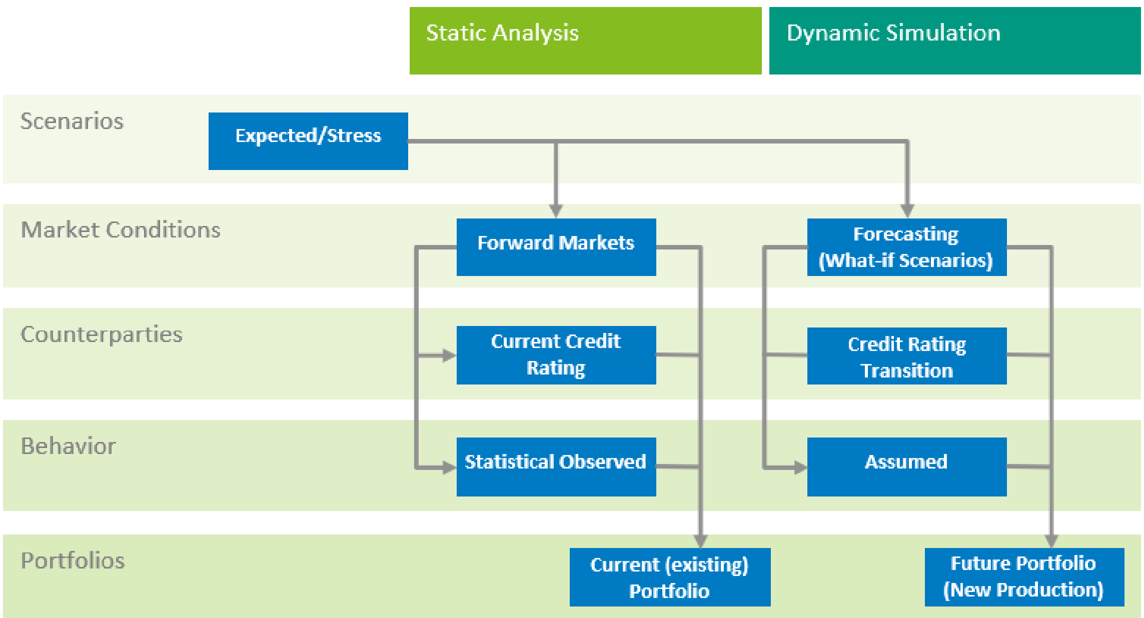
4. Dynamic simulation

Static analysis aims to evaluate the past and adjust present actions accordingly. Apart from cash, a portfolio's present value depends on income to be derived in the future. This brings us to dynamic analysis.

In dynamic analysis, everything – market and credit conditions, counterparty characteristics, behavior assumptions, new business – is seen to be in flux because the future remains undetermined and forever changing. Thus, in dynamic simulation:

- Market conditions, for example the risk factor of yield curves, are no longer derived from prices observed at $t_{0;}$ instead they are forecast using a simulation process.
- The counterparties' status, for example default probability and credit rating, may change as time passes. The rating transition can follow probabilities driven by the evolution of market conditions, their correlations with one another, and the counterparties' idiosyncratic characteristics.
- The generation of new business depends on current and future market and credit conditions. Portfolios of assets and liabilities are being rolled over, but, given the plans for new business, assets and liabilities are also being generated, leading to a growing balance sheet. Growth can follow predefined strategies and is included in the over-time analysis of the portfolios. These strategies depend very much on an institution's type, and the underlying risk factors to which current and future portfolios are exposed. Given the stress scenarios on risk factors, the bank identifies its risk appetite for the portfolios. A retail bank's strategy may focus on how loans or savings accounts are added to the balance sheet, while an investment or private bank may be most concerned with developing investment accounts and portfolios.
- The simulated contracts are generated by defining the characteristics of planned future financial instruments. This is most efficiently done with a well defined set of financial contract types. For instance, new bonds generated by using a principal of maturity contract type and then defining the contract characteristics, such as the targeted principal amount, maturity date, cycles of interest payments, and the counterparty's rating class.
- As new positions are generated, given the evolution of the markets and new counterparties assigned to these positions, assumptions about their behavior also must be considered. If the bank's strategy is to structure a new portfolio to provide facilities, say, the credit lines' possible exercising should be defined. Given the evolution of market conditions, scenarios of expected and stress behavior on exercising the facilities also should be applied.
- Given that parameters may change in a discrete or interdependent way, new business generation depends on those changes and how they develop. Stress on those parameters also can be applied. Liquidity, value and various risk measures can be analyzed dynamically, along with income and funds transfer pricing. Within this category fall earnings at risk, dynamic stress and value at risk (VaR), as well as dynamic liquidity and liquidity at risk (LaR). During the dynamic analysis, the potential impact - positive and negative - on a bank's income from applying stress scenarios to existing and future portfolios shows how robust the bank is to risk factors. *Figure 5* below illustrates the elements considered in the flow of static analysis and dynamic simulation.

Figure 5: Elements in static analysis and dynamic simulation



4.1 Stress testing under Covid-19

We have seen the impact of the pandemic on the Italian market, with variations across cities and regions. The impact on credit risk and on counterparties is significant and plays into the recognition of expected and incurred credit losses. In addition, alongside the EU, the national government has taken steps to limit the spread of Covid-19 and put a series of measures in place to support local businesses and people who are temporarily unemployed.

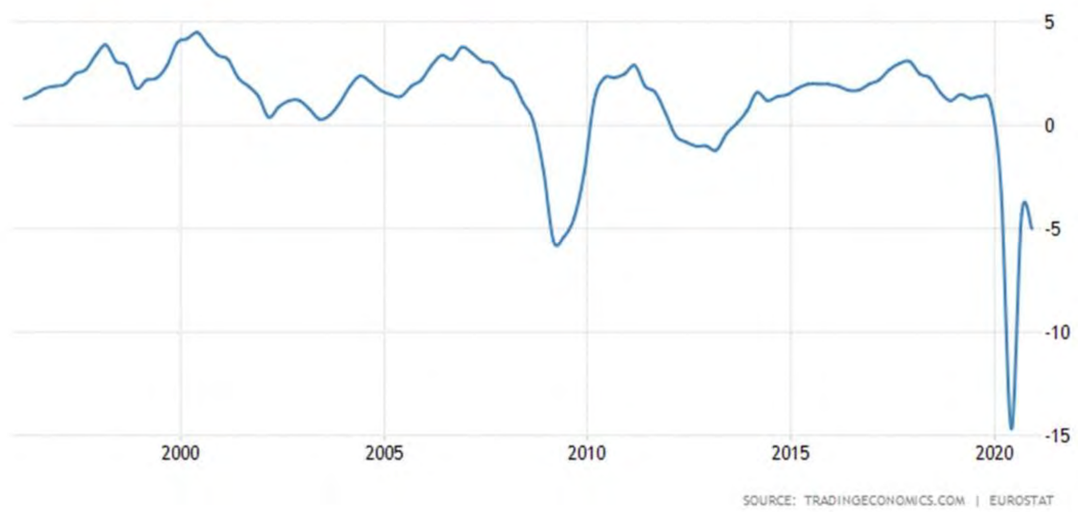
This could have a positive and a negative effect. The positive effect is to give counterparties that would have become insolvent due to the pandemic, but were sound businesses under normal operating conditions, a break until the pandemic is over, and so limit the amount of credit loss. The negative effect is to keep on life support businesses that otherwise were headed into default, which is not necessarily money well spent.

Looking at the world economy, lockdowns in response to the pandemic created unprecedented recessionary conditions. But governments' provision of liquidity limited the damage and helped maintain financial stability, so after the record deterioration of economic output, a record recovery is underway.

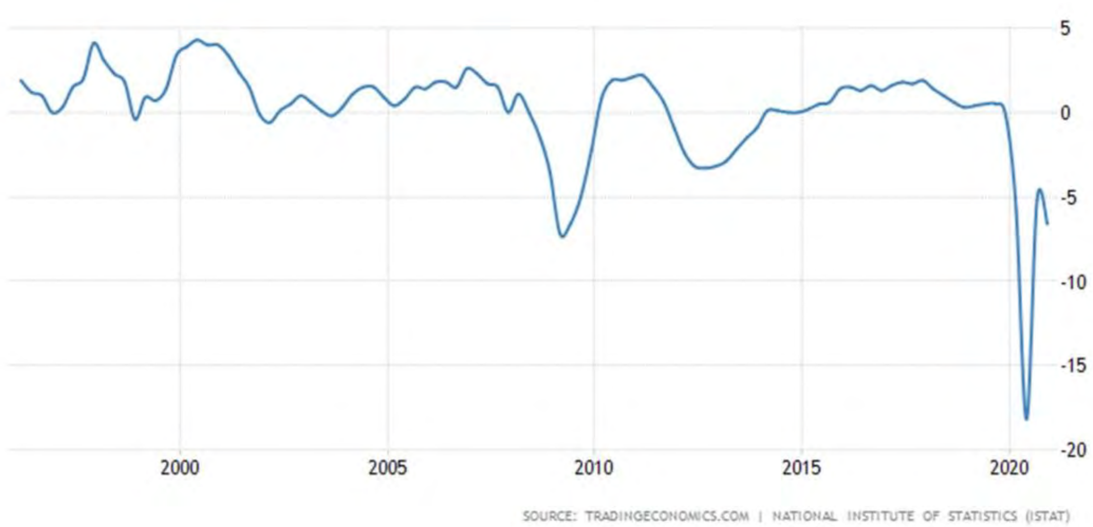
The pandemic continues to affect the economy, so banks must define and execute stress testing scenarios for current and future portfolios that take the potential effects of the pandemic into account.

Below are the GDP annual growth rates for the euro zone and Italy, which demonstrate the sharp economic contraction and the sharp rebound.

Euro Area GDP Annual Growth Rate



Italy GDP Annual Growth Rate



4.2 Examining different stress testing strategies under Covid-19

In the following exercise, we will consider stress testing and strategies for new business applied to typical banking portfolios containing loans, bonds, stocks and deposits to gauge how institutions are likely to – or should – alter their models to take into account the long-term impact of the pandemic and optimize the performance of their businesses.

The table below presents exposures to three broad types of risk: market, credit and behavioral. Market risk can have an impact through changes to yield curves (YC), foreign-exchange rates (FX) and credit spreads (CS).

Credit risk can be reflected in changes in credit ratings and probability of default (PD). The contributors to behavioral risk are the exercise of prepayments or withdrawals.

Portfolios and accounts considered in bank stress testing referring to our case on EU/Italian credit institutions

Portfolio Account	Risk Factors		
	Market	Credit	Behavior
Stocks & indexes	Prices Increases by 80% Decreases by 20%		
Corporate loans denominated in developing country currencies	YC Increases 2.5% p.a. Decreases 0% p.a.	Ratings Downgrading three notches	
	FX 50% drop		
	CS based on rating changes		
Government bonds & Corporate bonds denominated in euros	YC Increases 1.5% p.a. Decreases 0.1% p.a.	Ratings (unchanged)	
	CS (unchanged)		
Retail loans (mortgage and long-term loans)	YC Increases 2% p.a. Decreases 0.1% p.a.	Ratings PDs (unchanged)	Prepayments reduced by 50%
	CS (unchanged)		
	CS		
Current & Saving Accounts			Withdrawals (stable)

4.3 Scenarios based on interest rates and counterparty default probabilities

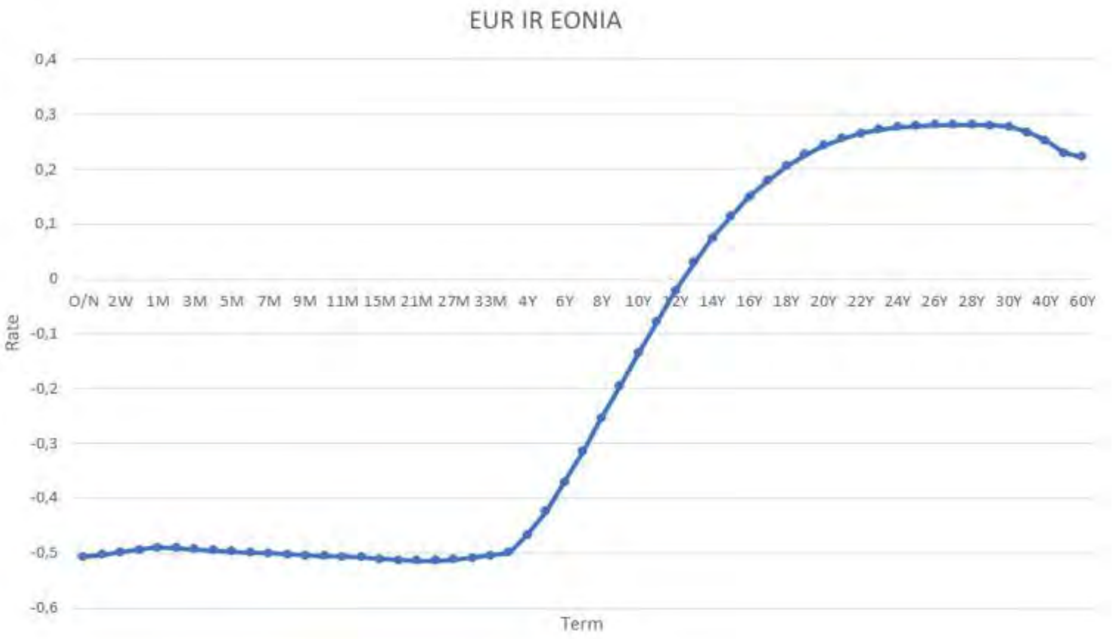
Policymakers worldwide have used massive fiscal and monetary stimulus programs to manage the pandemic's impact. As a result, public debt has soared in developed countries.

To manage the debt and reduce it in the long run, governments will need to keep interest rates lower than inflation and economic growth rates by maintaining their policies of financial repression. Very low or negative rates on all major interest rate curves and developed countries' government bonds are almost certain to continue in the long run (see *Figure 6*), despite the unprecedented amounts of fiscal stimulus. But if rapid growth should send rates higher in developed markets, shock on yield curves should be applied to evaluate the credit portfolios' impact on value and liquidity.

Moreover, under ordinary market conditions, the probability of defaults is not expected to be high. Any stress applied for steepening yield curves will directly impact default probabilities and credit ratings.

A strategy is to roll over the mortgage and long-term loan portfolios aligned with the scenarios mentioned above.

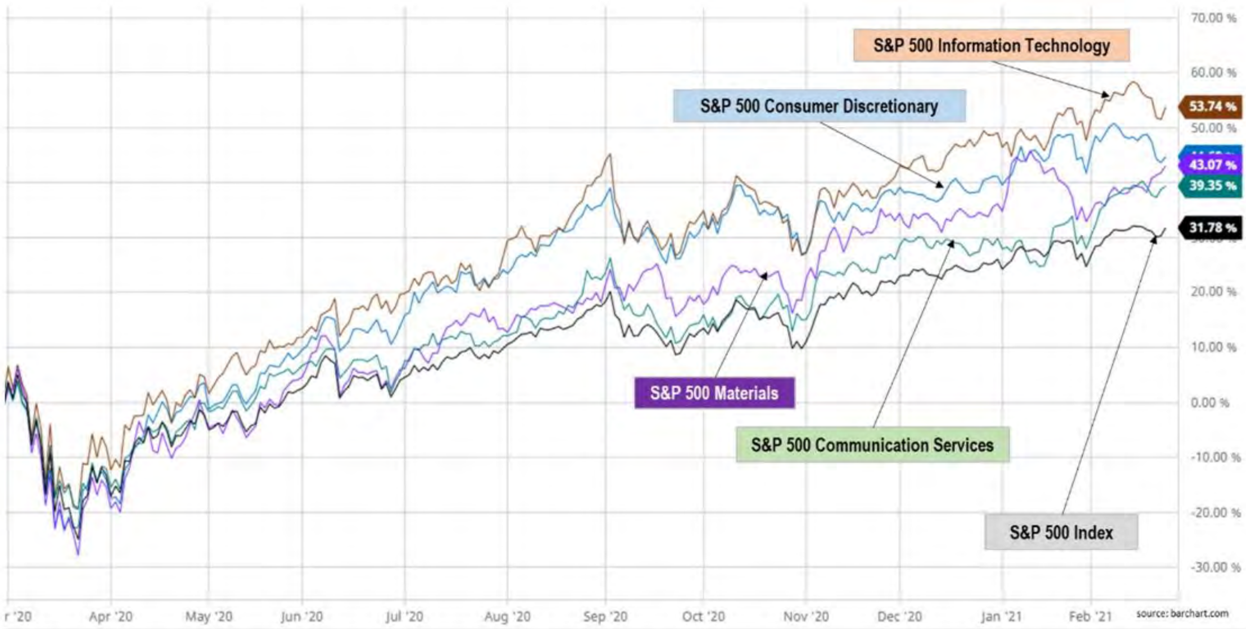
Figure 6: Long-term projection of EONIA



4.4 Scenarios based on stocks

Stock prices plunged in March 2020 as the severity of the pandemic became apparent, then quickly recovered to varying degrees. Growth sectors like technology outperformed, while economically sensitive sectors such as energy and basic materials lagged. *Figure 7* below displays how the Covid-19 crisis has impacted S&P 500 performance over the last year.

Figure 7: Performance of the S&P 500 during the Covid-19 crisis



4.5 Scenarios based on FX rates and counterparty ratings

With short-term interest rates in mature economies expected to remain clustered around zero percent for the foreseeable future, the main factor driving foreign-exchange movements is likely to be differences in economic growth and inflation. No significant revaluations among the main currencies are expected, therefore, as developed countries' economic recovery policies have many common features. Thus, FX rates need not feature in stress scenarios used by institutions in these economies.
Emerging economies often try to devalue their currencies against those of large, developed countries to remain competitive. The public debt of developing countries is not large, but private and corporate debt has increased significantly, mainly debt denominated in foreign currencies. A devaluation of the local currency makes debt repayment more challenging, therefore, especially for companies that depend on domestic sales.

The devaluation of developing countries' currencies may continue, albeit to a lesser extent for countries with vigorous export activity, and to an even lesser extent for countries that produce minerals, metals, energy or other commodities. Thus, stress on FX rates must be considered in the scenarios applied in portfolios exposed to these economies.

4.6 Scenarios based on counterparty behavior

During the pandemic, deposits into savings accounts may be reduced. After the pandemic, current and savings accounts will continue to earn negligible or negative rates, but deposits are still expected to grow. Thus, no stress in withdrawals is applied.
Low interest rates are likely to result in declining loan prepayments, especially on mortgage portfolios, and an increase in new mortgages.

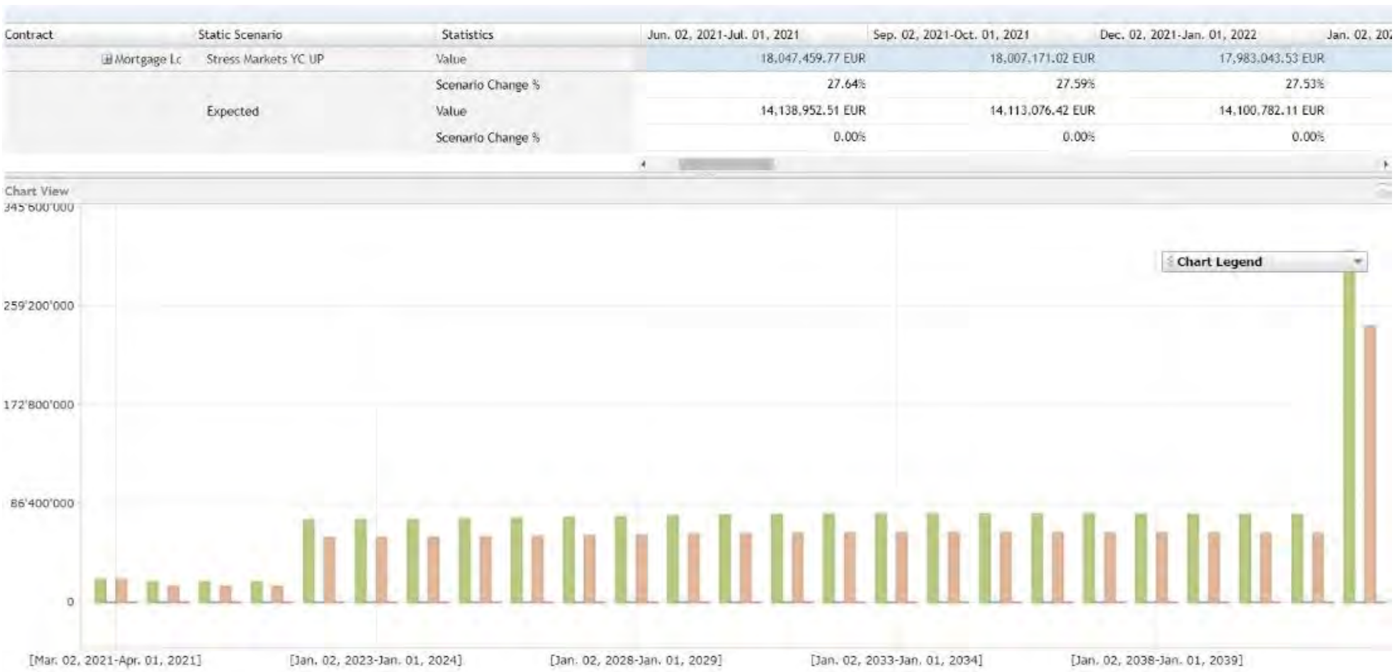
5. Results and discussion

For this exercise, a corporate portfolio has been structured with domestic euro-denominated loans and foreign loans denominated in Turkish lira. The domestic portfolio component is still profitable, with steady cash flows and a low default probability for the corporate counterparties, including small and medium-sized enterprises. As a result, expected credit losses for the domestic portfolio are low. The foreign loans may incur significant losses, however, for reasons that soon will be made clear.

In designing stress tests to reflect the impact of Covid-19, we focused on portfolios containing the following asset classes: stocks and stock index funds, government bonds, mortgage loans, and corporate loans denominated in local and foreign currencies. The counterparties are in the euro zone. Ratings are BBB for Italian government bonds and AAA for German government bonds. The euro zone corporates are rated from A to BB, and foreign instruments are rated B. The retail corporates carry high credit ratings and a low probability of default. The market conditions (YCs) are following the standard euro zone term structures, for example the Euro Overnight Index Average (EONIA). The long-term (20 to 35 years) mortgages are floating-rate loans. The corporate portfolio comprises floating-rate loans maturing in five to seven years. Stocks are concentrated in the technology sector.

In our stress scenario, we applied an increase of 200 basis points across the yield curve. The change in net present value (NPV) under those conditions was only 0.49%, whereas expected cash flow would rise. Under the same scenario, however, the increase in interest rates would cause losses in an investment portfolio holding rate-sensitive assets. Moreover, even though government bonds denominated in strong currencies, such as euros, dollars and yen, are highly rated, the negative yields that some of them carry could make the issues illiquid, as the prospect of earning a negative yield limits investor demand for such bonds. In the portfolio under study, fair value dropped by 18%. See *Figure 8* below to observe the impact of negative interest rates on expected cash flows of mortgage loans.

Figure 8: The impact of negative interest rates on expected cash flows under current and stress scenarios on long-term mortgage loan portfolios



Credit ratings for the foreign counterparties were between BB and B before the pandemic. The stress scenario called for each holding to be downgraded three notches. That resulted in a 20% decline in NPV for the annuity type of contracts and a 40% decline for the regular amortized loans.

Given that the latter instruments are more common among corporate loans, the impact of the downgrades is significant. For corporate portfolios exposed to foreign counterparties, and denominated in foreign currency, two types of stress are applied, covering counterparty risk due to the three-notch downgrades, and foreign-exchange risk, in this case a 50% drop in the lira. The fair value of such portfolios was reduced by almost 60%, creating high market and credit losses.

Government bonds play a significant role in collateral management for banking portfolios, as shown in the stress testing scenario (see *Figure 9* below). If more debt is issued with negative interest rates, the risk will grow that these assets will become illiquid. Finally, a reduction by 50% on prepayments should be applied to the mortgage and long-term retail loans. Still, in the last five years, the prepayment rate has been very low, so any stress does not impact the credit portfolios' expected cash flows.

Figure 9: The impact of negative interest rates on expected cash flows on a bond portfolio under current and stress scenarios

6. Conclusion

The pandemic continues, but there are signs that it will be brought under control before long. A return to normal social, business and economic conditions is coming.
Under current and short- to medium-term market conditions, credit institutions must perform stress testing in portfolios affected by macroeconomic and financial risk factors exacerbated by the pandemic.
Furthermore, after the pandemic abates, banks must ensure that rolling over existing portfolios and generating new ones will result in positive income. That requires banks to define strategies for the new portfolios on a going-concern basis by simulating the evolution of financial risk factors under expected and stress scenarios.
This paper discussed how banks might apply stress as a consequence of the Covid-19 crisis and, therefore, how observing both the input factors and results of the stressed portfolios on values and liquidity can guide future portfolios' strategies.
For instance, as observed in the analysis and results, given the likelihood that interest rates will remain low, so will banks' finance costs. Thus, mortgage and corporate credit portfolios are expected to increase in volume and be profitable as long as the counterparties keep default probability low and steady.
Notably, exposures to counterparties in emerging markets contain high credit risk, so banks may want to take extra care in rolling over portfolios with risky counterparties. Banks holding highly rated government bonds may face some challenges regarding negative interest income and liquidity.
Stock portfolios may increase in volume, so price volatility must be included in the stress scenarios. Since crises seem to appear frequently in highly integrated and interconnected markets, banks must be diligent in applying stress scenarios to existing and future portfolios.

Ioannis Akkizidis

Bibliography

- BIS, Financial Stability Institute, Oct 2020, Issue No 11, *Stress-testing banks during the Covid-19 pandemic* (https://www.bis.org/fsi/fsibriefs11.pdf)
- BIS Working Papers No 848, March 2020, *Implications of negative interest rates for the net interest margin and lending of euro area banks* (https://www.bis.org/publ/work848.pdf)
- European Banking Authority, Dec 2020, *Risk Assessment of the European Banking System*
- European Central Bank, March 2021, Macroeconomic projections, https://www.ecb.europa.eu/pub/projections/html/index.en.html
- Willi Brammertz, Ioannis Akkizidis, Wolfgang Breymann, Rami Entin. Nov. 2011, *Unified Financial Analysis: The Missing Links of Finance*

Moving from IBORs to Alternative Risk Free Rates

Marco Bianchetti (Intesa San Paolo), Umberto Cherubini (Università di Bologna) [1] and Veronica Falco (KPMG)

Abstract

In this short note we briefly review the state of the art of the ongoing transition from interbank rates (IBORs) to alternative risk free rates, with a focus on LIBOR and EUR benchmark rates. This note is a reduced version of a position paper published by AIFIRM in December 2019 [1], reporting more details regarding the impacts of the transition on Bank's internal processes, updated to December 2020.

JEL classifications: E43, G15, G18.

Keywords: IOSCO, FSB, ECB, EMMI, IBA, BMR, LIBOR, EURIBOR, EONIA, €STR, benchmark rate, interest rate, risk-free rate, overnight rate, discounting.

DOI 10.47473/2020rmm0080

1. IBORs Transition Overview

The Inter-Bank Offered Rates (IBORs) have been widely used by the market players as benchmarks for an enormous number of market transactions and a broad range of financial products since their invention by M. Zombanakis in 1969 [2] and their successive standardization by the British Bankers' Association in 1986 [3].

Currently, IBORs are the predominant interest rate benchmark for USD, GBP, CHF, EUR and JPY derivatives contracts [4]. EURIBOR is the most widely used interest rate benchmark for EUR contracts [5]. They are calculated through contributions from panel banks, and they reflect the offered rates for interbank unsecured wholesale deposits. IBORs indexed OTC derivatives and ETDs represent approximately 80% of IBOR-linked contracts by outstanding notional value, and thus derivatives represent the focus for global transition and reform initiatives. Going forward, this focus will include other products, such as securities, loans and mortgages.

After the LIBOR manipulation scandals [6][15], in 2013 IOSCO issued a set of principles that administrators of financial benchmarks should comply with, stating that interest rates must be reliable, robust and reflect real transactions [7].

By that time, the G20 had also mandated the Financial Stability Board (FSB) with conducting a global review of the main benchmarks and plans for their reform, in order to ensure that these were coherent and coordinated to the extent possible. In its 2014 report "Reforming major interest rate benchmarks" [8] the FSB recommended:

- strengthening existing reference rates by underpinning them, to the greatest extent possible, with transaction data;
- developing alternative, nearly risk-free reference rates.

In the euro area, the reform efforts were accelerated by the adoption of the EU Benchmarks Regulation (BMR) on 8 June 2016 [9], which codifies the IOSCO Principles into EU law and defines critical benchmarks that need a robust framework: EONIA, EURIBOR, LIBOR, STIBOR, WIBOR. Among other requirements, since 1 January 2018 BMR requires to include fallback clauses in specific type of contracts and permit the usage of critical benchmarks not compliant to the BMR until 31 December 2021.

Following these new requirements, in particular, EONIA, EURIBOR and LIBOR, were the subject of a deep reform, accelerated in the case of LIBOR from the statement of the Financial Conduct Authority (FCA) that confirmed it will no longer compel banks to submit LIBOR post December 2021.

In order to lead the market through the reform and with the will to be the link between market participants and regulators, each jurisdiction established a Working Group (WG) to define the Alternative Risk Free Rate (Alt-RFR) for the different currencies with which IBORs are contributed.

2. Features of the Alternative Risk Free Rates

Starting from the IOSCO principles and the following Authorities' guidelines, the Alt-RFRs are:

- transaction based, including non-bank counterparties deals;
- secured or unsecured
- reflecting the borrowing costs from wholesale market including non-bank counterparties.

[1] **Acknowledgements**: the authors acknowledge fruitful discussions with members of the AIFIRM Commission on Market Risk.
Disclaimer: the views expressed here are those of the authors and do not represent the opinions of their employers or affiliates. They are not responsible for any use that may be made of these contents.

Table 1 below reports the main features of the reference risk free rates identified by the WGs: they will side next to the IBORs and eventually substitute them.

Jurisdiction	EU	US	UK	Switzerland	Japan
Rate	Euro Short Term Rate (€STR)	Secured Overnight Financing Rate (SOFR)	Reformed SONIA	Reformed SARON	Reformed TONAR
Working Group	Working Group on Euro RFR for the Euro area	Alternative Reference Rates Committee	Working Group on Sterling Risk-Free Reference Rates	National Working Group on Swiss Franc Ref. Rate	Study Group on Risk-Free Reference Rates
Rate administrator	European Central Bank	Federal Reserve Bank of NY	Bank of England	SIX Swiss Exchange	Bank of Japan
Live	Pre-€STR: 15/3/2017 €STR: Oct. 2019	3 April 2018	Yes	Yes	Yes
Transition plan	No	Yes	No	No	No
Tenor	Overnight	Overnight	Overnight	Overnight	Overnight
Secured	No	Yes	No	Yes	No
Publication time	T+1 (8:00 CET)	T+1 (8:30 ET)	T+1 (9:00 GMT)	T (12:00, 16:00, 18:00 CET)	T (17:15 JST)
Market	EUR borrowing cost in the wholesale sector from MMSR banks	USD loans collateralized by Treasury Securities	GBP borrowing wholesale bilateral or broker transactions	CHF interbank repo transactions	JPY uncollateralized overnight call rate market
Data	Data from MMSR, size > 1 mln€	Data from BNYM and DTCC	Data from Sterling Money Market daily collection	Data from the order book of SIX Repo Ltd electronic trading platform	Data provided by money market brokers
Formula	Volume-weighted trimmed average (25%)	Volume-weighted trimmed average	Volume-weighted trimmed average (25%)	Volume-weighted average	Volume-weighted average

Table 1: main features of the reference risk free rates identified by the WGs.

3. Focus on the EURO area

3.1 From EONIA to €STR

In February 2018, the European Central Bank (ECB), the Financial Services and Markets Authority (FSMA), the European Securities and Markets Authority (ESMA) and the European Commission established the working group on euro risk-free rates (the ECB WG, [10]). The working group was tasked with (i) identifying risk-free rates which could serve as the basis for an alternative to the current benchmarks used in a variety of financial instruments and contracts in the euro area, (ii) identifying best practices for contractual robustness, and (ii) developing adoption plans – and, if necessary – a transition plan for legacy contracts which reference existing benchmarks.

The ECB WG works to provide guidelines and recommendations to market participants, in order to facilitate a smooth transition: its recommendations apply to different areas of impact (legal, accounting, risk management, etc.). In particular, the Working Group recommended €STR as euro risk-free rate on 13th September 2018 [11] [12]. Some of the key properties of €STR are:

- significant and steady volumes, markedly above EONIA volumes. On average, around 30 banks report data each day out of a pool of 52 MMSR reporting banks, which ensures that there is sufficient underlying data to calculate a reliable rate;
- very stable with an average daily volatility of just 0.4 basis point. Comparing the performance of the so called pre-€STR with that of EONIA over the period from March 2017 to July 2018, pre-€STR was very stable and was trading at a spread of around 9 basis points below EONIA.

Since 1 October 2019 €STR is published and EONIA is computed as EONIA = €STR + 8.5 bps, a one-off spread provided by the ECB, calculated as the arithmetic average of the daily spread between EONIA and pre-€STR (data from 17/04/2018 until 16/04/2019), after removing the 15% of observations from the top and the bottom of the sorted series.

Also the timing changed: while EONIA was published at 19.00 CET on each business day (T), €STR is published at 8:00 CET on the next business day (T+1). In case of errors in the €STR calculation that affect the rate value by more than 2 bps, €STR is revised and re-published on the same day at 09:00 CET. As a consequence of the recalibrated methodology, also EONIA is published on the next business day (T+1) at 9:15 CET.

EONIA will be published until 3 January 2022, when it is discontinued. Before its discontinuation, market participants have to perform a series of activities to be ready. The ECB WG issued a lot of recommendations to address a smooth transition and the milestones are:

1 creation of a new market based on €STR-linked derivatives: at the beginning, the €STR OIS curve was EONIA OIS curve – 8,5bps but, with the passage of time, the €STR OIS curve is being built;

2 PAI and discounting regime switch performed by CCPs on 27 July 2020: LCH, EUREX and CME switched from EONIA to €STR all their EUR OTC derivatives in clearing;
3 PAI and discounting regime switch to be performed by counterparties with respect to their derivatives positions under bilateral CSAs: several banks are dealing each other to agree how and when perform the switch;
4 Decision by market makers and brokers on how to quote non-linear/volatility/correlation derivatives. Currently Cap/Floor are still quoted versus EONIA, while Swaption are quoted versus €STR: the way is still long but it is traced;
5 Decision by market participants to revise risk-free net present values and xVAs pricing models or to perform new valuation adjustments.

The transition from EONIA to €STR has a number of consequences on the valuation of derivatives, as outlined e.g. in [13].

3.2 From EURIBOR to Hybrid EURIBOR

EURIBOR is the commonly used term rate for euro denominated financial contracts. EURIBOR reflects the rate at which wholesale funds in euro can be obtained by credit institutions in EU and EFTA countries in the unsecured money market, and seeks to measure banks' costs of borrowing in unsecured money markets [5].

In 2016, EURIBOR was declared a critical benchmark by the European Commission, so its administrator, the Euro Money Markets Institute (EMMI), has conducted in-depth reforms in recent years in order to meet the BMR requirements, by strengthening its governance framework and developing a hybrid methodology in order to ground the calculation of EURIBOR, to the extent possible, in euro money market transactions.

In July 2019, the supervisor of EURIBOR, the FSMA, granted authorisation to EMMI for hybrid-EURIBOR under the BMR. This authorisation provides confirmation that EMMI and the EURIBOR hybrid methodology meet the requirements laid down in the BMR and that EURIBOR may continue to be used in new and legacy contracts.

Starting from the end of 2019 all panel banks contribute their data following the "hybrid" determination methodology developed by EMMI, based on a 3 levels hierarchy, as illustrated in Figure 1 below.

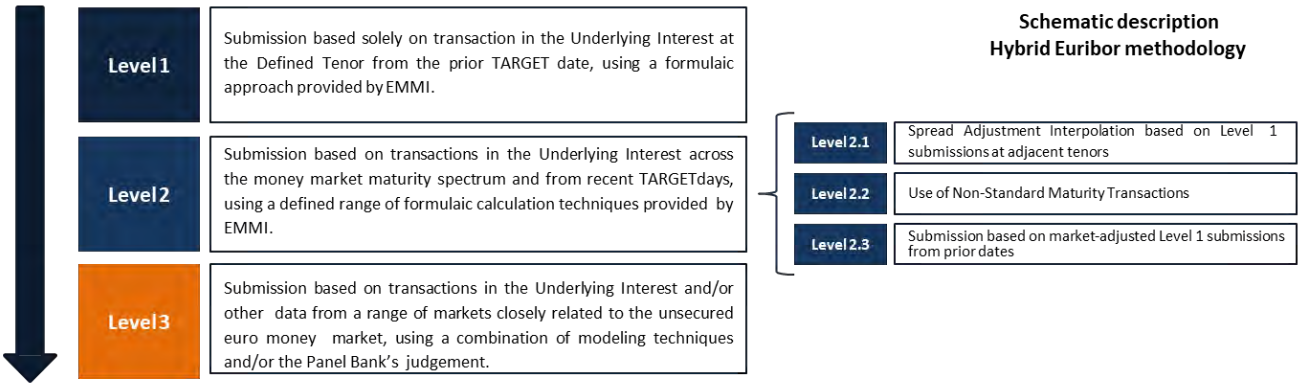


Figure 1: scheme of the hybrid-EURIBOR methodology adopted by EMMI in 2019 [5].

The ECB WG recommended to use the €STR term structure as a fallback to EURIBOR [14]. On 23 November 2020 the ECB WG issued two consultantions on fallback trigger event and €STR-based fallback whose results will be shared during the ECB WG meeting in February 2021 [15]. Starting from the industry's feedbacks, the ECB WG will issue recommendations on fallback rates to be applied to different products. It is worth to be highlighted that €STR will be the EURIBOR fallback rate, but the calculation methodology will depend on the products to be applied.

4. Focus on LIBOR

The Financial Stability Board and Financial Stability Oversight Council have both publicly recognized that the decline in wholesale unsecured term money market funding by banks poses structural risks for unsecured benchmarks, including LIBOR. Although significant progress has been made by the LIBOR Administrator (ICE Benchmark Administration – IBA) in strengthening the governance and processes underlying LIBOR, the scarcity of underlying transactions poses a continuing risk of a permanent cessation of its production after the end of 2021.

Andrew Bailey, then the Chief Executive of the United Kingdom's Financial Conduct Authority (FCA) highlighted this end-2021 timeline in a speech in 2017 [16] and the FCA recently reemphasized [17] that the central assumption that firms cannot rely on LIBOR being published after the end of 2021 has not changed and that this should remain the target date for all firms.

On December 4, 2020, IBA published its consultation, with deadline January 25, 2021, on its intention to cease the publication of the LIBORs (CHF, GBP, JPY and EUR) on December 31, 2021, considering some postponements only for USD LIBOR until June 30, 2023 [4].

Different WGs are providing recommendations to lead a smooth transition from USD, GBP, CHF, JPY and EUR LIBOR to, respectivey, SOFR, SONIA, SARON, TONA and €STR: with 13 months left until LIBOR could become unusable, it is important that market participants accelerate their transition efforts, having in mind that:

- new LIBOR cash products should include fallback language as soon as possible;
- third-party technology and operations vendors relevant to the transition should complete all necessary enhancements to support Alt-RFR by the end of this year;
- New use of LIBOR should stop, with timing depending on specific circumstances in each cash product market.
- For contracts specifying that a party will select a replacement rate at their discretion following a LIBOR transition event, the determining party should disclose their planned selection to relevant parties some months prior to the date that a replacement rate would become effective.

Considering the cleared USD OTC derivatives, the CCPs switched the PAI and discounting regime in October 2020 from Fed Fund Rate to SOFR through a complex mechanism. Since there is no fix spread between EFFR and SOFR, the switch resulted in cash compensation, to manage the valuation change, and swap compensation, to manage risk profile change.

The first next milestone that USD market participants have to reach before the USD Libor discontinuation is the PAI and discounting regime switch for derivatives under bilateral CSAs.

The second next milestone for the LIBOR WGs is to lead the market participants in the construction of a term rate structure or address the impacts that the only use of overnight rate compounding could cause (e.g. some derivatives cannot be priced with compounded rates).

5. Focus on ISDA work on Derivatives

In 2016, the Official Sector Steering Group (OSSG) formally launched a major initiative to improve contract robustness and address the risks of widely-used interest rate benchmarks being discontinued. The OSSG invited ISDA to lead this work as it pertained to derivative contracts – the largest source of activity for the IBORs.

ISDA [18] conducts its work through different WGs: ISDA Americas and Europe Benchmark WG, ISDA APAC Benchmark WG, ISDA JPY Benchmark WG, ISDA EU Benchmark Regulation Advocacy Group and the ISDA IBOR Fallback Implementation Subgroup.

To address the risk that one or more IBORs are discontinued while market participants continue to have exposure to that rate, counterparties are encouraged to agree to contractual fallback provisions that would provide for adjusted versions of the RFRs as replacement rates.

ISDA developed fallbacks that would apply upon the permanent discontinuation of certain IBORs and upon a 'non-representative' determination for LIBOR. ISDA will amend the 2006 ISDA Definitions by publishing a 'Supplement' to the 2006 ISDA Definitions on January 25, 2021: transactions incorporating it, that are entered into on or after the date of the Supplementwill include the amended floating rate option (i.e., the floating rate option with the fallback). Transactions entered into prior to the date of the Supplement (so called "legacy derivative contracts") will continue to be based on the 2006 ISDA Definitions as they existed before they were amended pursuant to the Supplement, and therefore will not include the amended floating rate option with the fallback.

ISDA has published a protocol [19] to facilitate multilateral amendments to include the amended floating rate options, and therefore the fallbacks, in legacy derivative contracts. By adhering to the protocol, market participants would agree that their legacy derivative contracts with other adherents will include the amended floating rate option for the relevant IBOR and will therefore include the fallback. The protocol is completely voluntary and will amend contracts only between two adhering parties (i.e., it will not amend contracts between an adhering party and a non-adhering party or between two non-adhering parties). The fallbacks included in legacy derivative contracts by adherence to the protocol will be exactly the same as the fallbacks included in new transactions that incorporate the 2006 ISDA Definitions and that are entered into on or after January 25, 2021.

Marco Bianchetti, Umberto Cherubini and Veronica Falco

Bibliography

[1] AIFIRM Position Paper n. 16, “From IBORs to RFRs: Impacts on Banks’ Processes and Procedures”, December 2019, http://www.aifirm.it/wp-content/uploads/2020/01/2019-Position-Paper-16-From-IBORs-TO-RFRs.pdf

[2] Finch, Gavin and Vaugham, Liam, “The Man Who Invented the World’s Most Important Number”, 29 November 2016, https://www.bloomberg.com/news/features/2016-11-29/the-man-who-invented-libor-iw3fpmed

[3] Michaud, François-Louis and Upper, Christian, “What drives interbank rates? Evidence from the Libor Panel”, BIS quarterly review, 3 March 2008, https://www.bis.org/publ/qtrpdf/r_qt0803f.htm.

[4] ICE LIBOR, https://www.theice.com/iba/libor

[5] Euro Interbank Offered Rate (EURIBOR) rate. https://www.emmi-benchmarks.eu/euribor-org/about-euribor.html.

[6] See e.g. Wikipedia and references therein https://en.wikipedia.org/wiki/Libor_scandal

[7] OICV-IOSCO, “Principles for Financial Benchmarks” Final Report, July 2013, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD415.pdf.

[8] Financial Stability Board, “Reforming Major Interest Rate Benchmarks”, 22nd July 2014. https://www.fsb.org/2014/07/r_140722/.

[9] Regulation (EU) 2016/1011 of the European Parliament and of the Council of 8 June 2016 on indices used as benchmarks in financial instruments and financial contracts or to measure the performance of investment funds and amending Directives 2008/48/EC and 2014/17/EU and Regulation (EU) No 596/2014. June 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R1011.

[10] Working group on euro risk-free rates, https://www.ecb.europa.eu/paym/interest_rate_benchmarks/WG_euro_risk-free_rates

[11] European Central Bank. Overview of the euro short-term rate (€STR). https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_short-term_rate/html/eurostr_overview.en.html.

[12] European Central Bank. The euro short-term rate €STR methodology and policies. June 2018. https://www.ecb.europa.eu/paym/initiatives/interest_rate_benchmarks/shared/pdf/ecb.ESTER_methodology_and_policies.en.pdf.

[13] Scaringi, Marco and Bianchetti, Marco, “No Fear of Discounting - How to Manage the Transition from EONIA to €STR” (September 2, 2020). Available at SSRN: https://ssrn.com/abstract=3674249.

[14] Working group on euro risk-free rates, https://www.ecb.europa.eu/press/pr/date/2018/html/ecb.pr180913.en.html; https://www.ecb.europa.eu/paym/interest_rate_benchmarks/WG_euro_risk-free_rates/shared/pdf/20180913/Item_3_High_level_implementation_plan.pdf

[15] Working group on euro risk-free rates, https://www.ecb.europa.eu/pub/pdf/other/ecb.pubcon_EURIBORfallbacktriggerevents.202011~e3e84e2b02.en.pdf; https://www.ecb.europa.eu/pub/pdf/other/ecb.pubcon_ESTRbasedEURIBORfallbackrates.202011~d7b62f129e.en.pdf

[16] Bailey, Andrew, “The future of LIBOR”, Bloomberg, 27 July 2017, https://www.fca.org.uk/news/speeches/the-future-of-libor

[17] Financial Conduct Authority, Transition from LIBOR | FCA

[18] ISDA, Benchmark Reform and Transition from LIBOR https://www.isda.org/2020/05/11/benchmark-reform-and-transition-from-libor/

[19] ISDA, ISDA Launches IBOR Fallbacks Supplement and Protocol ISDA-Launches-IBOR-Fallbacks-Supplement-and-Protocol.pdf

Artificial Intelligence: the Application of Machine Learning and Predictive Analytics in Credit Risk

Stefano Bonini (Università di Bologna - Università degli Studi di Milano), Giuliana Caivano (Accenture Strategy & Consulting)[1]

Article submitted to double-blind peer review, received on 4th September 2019 and accepted on 20th January 2021

Abstract

In the last years Machine Learning (and the Artificial Intelligence), is experiencing a new rush thanks to the growth of volume and kind of data, the presence of tools / software with higher computational power and cheaper data storage size (e.g. *cloud*). In Credit Risk Management, the PD (Probability of Default) estimation has attracted lots of research interests in the past literature and recent studies have shown that advanced Artificial Intelligence (AI) methods achieved better performance than traditional statistical methods tied to simplified Machine Learning techniques. The study empirically investigates the results of applying different advanced machine learning techniques in estimation and calibration of Probability of Default. The study has been done on big data sample with more than 800,000 Retail customers of a panel European Banks under ECB Supervision, with 10 years of historical information (2006 - 2016) and 300 variables to be analyzed for each customer. The study shows that *neural network* produces a higher population riskiness ranking accuracy, with 71% of Accuracy Ratio. However, the authors' idea is that *classification tree* is more interpretable from an economic and credit point of view. In terms of model calibration, *cluster analysis* produces rating classes more stable and with a predicted risk probability aligned with the observed default rate.

***

Negli ultimi anni il Machine Learning e, più in generale, il mondo dell'Intelligenza Artificiale, sta acquisendo un nuovo slancio grazie alla crescita del volume e della varietà dei dati, a processi di elaborazione / strumenti con elevata potenza computazionale oltre agli spazi per l'archiviazione dei dati sempre più a buon mercato (es. *cloud*). Nell'ambito del Credit Risk Management, la modellizzazione della PD (Probabilità di Default) ha attirato l'interesse accademico nella letteratura passata e recenti studi analizzati dagli autori hanno mostrato che l'applicazione di tecniche di Intelligenza artificiale (IA) avanzate permette di ottenere performance migliori rispetto alla statistica tradizionale legata a tecniche di Machine Learning più semplificate. In questo paper si analizzano empiricamente i risultati derivanti dall'applicazione di diverse tecniche avanzate di machine learning nella stima e calibrazione del parametro di Probabilità di Default. Lo studio è stato condotto su un campione contenente oltre 800.000 clienti Retail di un panel di banche europee sotto supervisione della BCE, con 10 anni di informazioni storiche (2006 - 2016) e 300 variabili da analizzare per ciascun cliente. I risultati mostrano una maggiore accuratezza del ranking della popolazione (in termini di rischiosità) ottenuto attraverso l'applicazione di *reti neurali*, con un valore di Accuracy Ratio (AR) del 71%. È idea degli autori, tuttavia, che al di là delle prestazioni ottenute, l'*albero di classificazione* risulti essere maggiormente interpretabile da un punto di vista economico e creditizio. In termini di calibrazione del modello, l'applicazione della *cluster analysis* genera classi di rating stabili e con una rischiosità stimata allineata alla rischiosità empirica osservata.

Key Words:
Risk Management, Credit Risk, Machine Learning, Big Data, Data Analysis, Advanced Predictive Analytics

DOI 10.47473/2020rmm0081

1 Introduzione

Il tema dell'Intelligenza Artificiale è la buzz-word del momento e il mercato bancario ha iniziato a scoprirla solo negli ultimi anni, ma affonda le sue radici nel passato: basti pensare a tutti quei ricercatori che iniziarono a indagare se fosse possibile l'apprendimento dei computer a partire dai dati. È questo l'assunto alla base del Machine Learning (o apprendimento automatico), ossia che i computer possano imparare ad eseguire dei task semplicemente osservando le relazioni esistenti tra i dati, imparando dai dati con una efficacia tanto maggiore quanto maggiore è la disponibilità di informazioni.

Gli ultimi anni sono stati caratterizzati da una rivoluzione tecnologica e digitale che offre nuove opportunità per il miglioramento e l'efficientamento delle prassi operative e l'adozione di approcci metodologici più avanzati in diversi campi di ricerca. In un contesto sempre più competitivo con riduzione dei margini di profitto il Machine Learning e, più in generale, il mondo dell'Intelligenza Artificiale, sta acquisendo un nuovo slancio grazie alla crescita del volume e della varietà dei dati, a processi di elaborazione / strumenti con elevata potenza computazionale oltre agli spazi per l'archiviazione dei dati sempre più a buon mercato (es. *cloud*). Il Machine Learning, in particolare, può ricoprire un ruolo chiave sia in ambito tecnologico sia di business, consentendo alle istituzioni finanziarie di gestire al meglio grandi moli di dati e facilitare l'adattamento e la ricalibrazione dei modelli.

Negli ultimi anni sono state molte le tecniche di Machine Learning pensate per la stima di variabili binarie, in molti campi della scienza. Nell'ambito del Credit Risk Management, in particolare, la modellizzazione della PD (Probabilità di Default) ha attirato l'interesse accademico nella letteratura passata e recenti studi analizzati dagli autori hanno mostrato che l'applicazione

[1] Le opinioni espresse rappresentano esclusivamente il punto di vista degli autori e non riflettono necessariamente quello dell'Istituto/Azienda d'appartenenza.

di tecniche di Intelligenza artificiale permette di ottenere performance migliori rispetto alla statistica tradizionale sia nell'applicazione ai problemi di credit scoring ([16], [27]) sia nella stima della Probabilità di Default (cfr. [3], [4], [7], [8], [20]).

Obiettivo principale di questo studio è evidenziare la rilevanza della scelta degli algoritmi, dei parametri, la selezione delle variabili (caratteristiche) rilevanti, il ruolo dei criteri di valutazione e l'importanza del contributo esperto nella definizione della Probabilità di Default di un portafoglio di prestiti. Lo studio presentato nel paper, inoltre, cerca di rispondere anche ad una serie di quesiti tipici che emergono quando si ha a che fare con l'applicazione di algoritmi e tecniche statistiche avanzate: come risolvere o raggiungere un determinato obiettivo?

Dalla letteratura analizzata emergono diversi paper che applicano tecniche di Machine Learning a un campione di dati con elevato numero di osservazioni e relativo ad un solo intermediario finanziario. Il nostro paper si differenzia dalla letteratura esistente per una serie di motivazioni: mentre la letteratura accademica analizzata si focalizza sulla creazione di una misura cardinale dell'evento default (cfr. [25]) utilizzando algoritmi di classificazione generalizzati e tecniche alberi di classificazione, il nostro studio è focalizzato, da una parte, sulla capacità di ordinamento / ranking, in particolare su come diversi algoritmi di machine learning e deep learning prevedono l'evento default. In aggiunta, il paper si focalizza anche sulle variabili selezionate da ciascun algoritmo come drivers di rischio e, infine, analizza il potere di calibrazione delle stime ottenute tramite l'utilizzo di tecniche di tipo non supervisionato per la definizione delle classi di rating.

Il nostro studio si basa sull'applicazione di algoritmi già utilizzati in letteratura, ad es. in [10] gli autori utilizzano alberi decisionali, regressione logistica e random forest per l'analisi del livello di delinquency dei consumatori utilizzando dati relativi a sei differenti banche. In [23] gli autori applicano alberi decisionali ad un portafoglio mutui per prevedere l'evento default e confrontano i risultati con tecniche k-nearest neighbors (KNN), reti neurali artificiali (ANN) e modelli probit.

L'utilizzo di modelli classici di regressione (logistica e lineare, cfr. [12]) è ben noto nel mondo bancario, pertanto nel nostro studio abbiamo utilizzato la regressione logistica come benchmark e comparato la sua capacità di fitting (in termini di *Accuracy Ratio* e *Tasso di corretta classificazione*) con quella di altri modelli non parametrici annoverati tra le tecniche di machine learning e deep learning nella letteratura più recente. Abbiamo utilizzato tre approcci: alberi di classificazione, random forest e deep learning (rete neurale) applicandoli ad un campione contenente oltre 800.000 clienti Retail di un panel di banche europee sotto supervisione della BCE, con 10 anni di informazioni storiche (2006 - 2016) per valutare non solo la capacità di fitting di ciascun modello rispetto alla regressione logistica ma anche la combinazione di variabili selezionate da ciascun modello.

Il paper è cosi strutturato: nella Sezione 2 è presentata una descrizione delle principali logiche metodologiche sottostanti agli algoritmi utilizzati nell'ottica di illustrarne il funzionamento rispetto all'obiettivo / variabile target da modellizzare. La Sezione 3 illustra i principali criteri utilizzati per la classificazione e il confronto dei risultati mentre la Sezione 4 descrive il dataset utilizzato per lo studio empirico e i principali risultati ottenuti. Infine, la Sezione 5 fornisce le conclusioni dello studio e una vista delle possibili evoluzioni della ricerca.

2 Aspetti teorici delle metodologie più diffuse in letteratura[2]

In questa sezione sono descritti gli algoritmi di machine learning utilizzati per:

a) costruire un modello di ranking / scoring della popolazione utilizzata (*apprendimento supervisionato*);

b) calibrare lo score e definire quale probabilità di default associare a ciascuna classe di rating (*apprendimento non supervisionato*).

2.1 Apprendimento supervisionato per la definizione dello scoring

Il primo obiettivo – costruire un modello di ranking della popolazione – si presta ad essere formulato come un problema di apprendimento supervisionato che rappresenta una delle tecniche di machine learning più utilizzate in letteratura.

Nel framework del *supervised learning*, un *"learner"* si presenta con coppie di input / output dai dati storici in cui i dati di input rappresentano attributi pre-identificati per essere utilizzati nel definire il valore dell'output. I dati di input sono comunemente rappresentati come vettori e, in funzione dell'algoritmo di apprendimento scelto, possono consistere in valori continui e/o discreti con o senza dati mancanti. L'apprendimento supervisionato risolve un problema di tipo *regressivo* quando l'output è una variabile continua, di *classificazione* quando l'output è una variabile discreta.

Una volta presentati i dati di input / output, il compito del *learner* è trovare una funzione che mappi correttamente i vettori di input verso i valori di output, ad esempio memorizzando tutte le precedenti coppie di valori input / output. Anche se questo metodo mappa correttamente le coppie di valori di input / output nel campione di training, è improbabile che il modello funzioni nel prevedere i valori di output se i valori di input sono diversi da quelli contenuti nel dataset di training o se il dataset di training contiene *"noise"*. In questo contesto, la sfida dell'apprendimento supervisionato è trovare una funzione che generalizzi oltre il dataset di training, in modo che la stessa sia in grado di mappare accuratamente input verso output *out-of-sample*.

[2] Parte dei dettagli metodologici qui illustrati è estratta dal position paper AIFIRM #14 "*Intelligenza Artificiale: l'applicazione di Machine Learning e Predictive Analytics nel Risk Management*"

Ad esempio, nel caso specifico del nostro studio, l'output del modello è una variabile continua tra 0 e 1 che può essere interpretata (sotto certe condizioni) come una stima della probabilità di un cliente di andare in default entro i 12 mesi successivi date certe caratteristiche del cliente stesso e/o del prodotto in oggetto.

Nel nostro studio abbiamo costruito diversi modelli di ranking / scoring della popolazione utilizzando e confrontando tra loro, in particolare, quattro approcci di apprendimento supervisionato:

1. Regressione logistica;
2. Albero decisionale (CART);
3. Random Forest;
4. Rete neurale (*deep learning*).

La regressione logistica è un'estensione del modello di regressione lineare in cui la relazione lineare alla base di quest'ultimo modello è aggiustata attraverso una trasformazione esponenziale, chiamata trasformazione logistica. In particolare, la regressione logistica analizza la relazione tra multiple variabili indipendenti e una singola variabile dipendente dicotomica - nel caso dello studio in oggetto la variabile "*good*" /"*bad*" – tramite la stima di un punteggio di probabilità e con l'obiettivo di discriminare al massimo i due gruppi individuati dalla variabile dicotomica.

$$y_i = f(w_i) = \frac{1}{1+e^{-w_i}} \quad (1)$$

Dove la variabile indipendente w_i è data dalla funzione lineare degli indicatori selezionati:

$$w_i = \alpha + \sum_{j=1}^{m} \beta_j x_{i,j} \quad (2)$$

Combinando le equazioni definite e aggiungendo il termine di errore, si ottiene il modello logit come:

$$y_i = \frac{1}{1+e^{-\alpha-\sum_j \beta_j x_{i,j}}} + \varepsilon_i \quad (3)$$

Il campo di valori generati dalla funzione logistica (il " codominio " della funzione) è ora limitato all'intervallo (0,1). Ciò garantisce che la variabile dipendente y_i sia sempre compresa tra 0 e 100% e può pertanto essere correttamente interpretata come una probabilità di default.
Modelli lineari come la regressione logistica hanno l'indubbio vantaggio di produrre buoni risultati laddove l'aspettativa sia di avere funzioni lineari e di essere comprensibile e spiegabile. Di contro, tale modello non gestisce in maniera efficiente le variabili categoriche e la presenza di elevata correlazione tra le variabili può generare problemi; inoltre le performance possono risentire in caso di variabili non lineari e spesso si osserva una propensione all'*underfitting*.

Logiche diverse sono invece alla base dei modelli *CART (Classification and Regression Trees)*, ossia un insieme di tecniche di stima molto utilizzate nell'ambito del Machine Learning, applicabili sia a problemi di classificazione sia di regressione (cfr. [31]) e in cui una variabile dipendente o output (discreta o continua) è legata ad un insieme di variabili indipendenti (o di input) attraverso una sequenza ricorsiva di semplici relazioni binarie (da qui il riferimento ad "albero"). L'insieme delle relazioni ricorsive divide lo spazio multidimensionale delle variabili indipendenti in distinte "aree" in cui la variabile dipendente è tipicamente assunta, nel caso di un albero regressivo come nello studio in oggetto, come legata linearmente alle variabili indipendenti con parametri univoci per ciascuna "area".

In *Figura 1* è rappresentato un modello CART con due variabili indipendenti non negative (x_1, x_2) anche note come "*feature vector*" e una variabile dipendente discreta che assume due valori, "*good*" e "*bad*". La sequenza di relazioni ricorsive binarie rappresentate nell'albero in *Figura 1* suddivide il dominio di (x_1, x_2) in cinque aree distinte determinate dai parametri L_1 ….L_4.

In particolare, questo modello implica che tutti i valori di (x_1, x_2) con $x_1 < L_1$ e $x_2 < L_2$ sono associati a un outcome "bad" e tutti i valori di (x_1, x_2) con $x_1 < L_1$ e $x_2 >= L_2$ sono associati a un outcome "good".

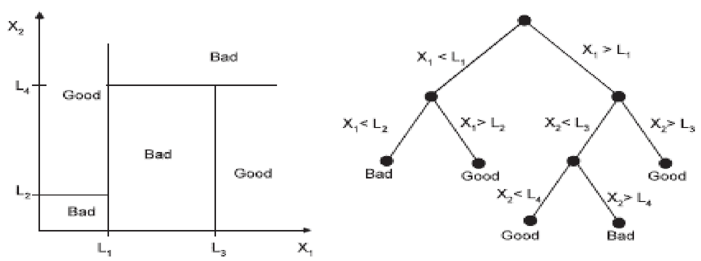


Figura 1 – Esempio di albero regressivo con variabile target binaria

I parametri (Lj) sono scelti per minimizzare, in ciascun passaggio, la distanza tra la variabile dipendente e i valori fittati all'interno di ciascuna categoria e massimizzare invece quella tra le diverse categorie. Questi due vincoli vengono incorporati nella formula della funzione obiettivo:

$$D = \sum_{i=1}^{K} \left\{ \sum_{j \in S_i} \left(y_i - \widehat{\beta}_i \right)^2 \right\} = \sum_i D_i, \tag{4}$$

dove y_i e $\widehat{\beta}_i$ sono rispettivamente i valori della variabile target (nel caso dello studio in oggetto 0 e 1) e il parametro ad essi associato presenti all'interno di uno dei K sottospazi dei dati S_i. In ciascun passaggio il processo si ripete, cercando tra i dati dei sottoinsiemi il valore di soglia per la variabile in grado di ottenere la divisione ottimale. Questo processo viene iterato fino a quando non si verificano determinate condizioni che ne determinano l'arresto. Una di queste cause può essere, ad esempio, la creazione di un sottospazio di dati aventi la stessa categoria della variabile target (nel *Classification Tree*) o che possiedono gli stessi valori (nel *Regression Tree*).

Quando l'algoritmo genera un albero particolarmente fitto e complesso, composto da molti rami e foglie, il risultato può risultare scarsamente interpretabile, per l'elevato numero di tagli e per la tendenza al sovradattamento dei dati (*overfitting*). È pertanto necessario ridurre l'albero tramite una procedura automatica chiamata "potatura": una tecnica che, partendo dal modello completamente sviluppato, elimina sequenzialmente i rami non utili ai fini della stima o con la minore carica informativa. Definendo la funzione di perdita come segue:

$$C_\alpha(K) = \sum_{i=1}^{K} D_i + \alpha K \tag{5}$$

dove K è la dimensione dell'albero considerato in ogni singolo passo e α il parametro associato al costo computazionale del modello, verrà ad ogni passo rimossa la foglia la cui eliminazione comporta il minore aumento della funzione obiettivo $\sum_{i=1}^{K} D_i$. La procedura continua fino a quando il valore di $C_\alpha(K)$ sarà stabilizzato.

Quello degli alberi decisionali è un algoritmo molto utilizzato nella pratica e spesso citato nella letteratura poiché presenta grandi vantaggi rispetto alle altre tecniche di Machine Learning. Risulta infatti uno dei modelli più informativi, grazie alla sua alta semplicità logica che permette di comunicare facilmente le regole alla base della sua struttura, mettendo in evidenza quali sono i principali driver implicati nella stima. Correlato a questo fatto, gli alberi risultano essere un ottimo metodo automatico di riduzione della dimensionalità dei dati, selezionando soltanto le variabili più importanti ai fini dell'approssimazione dei dati. Un altro vantaggio di questo modello è la sua ridotta complessità computazionale anche quando la mole di osservazioni e il numero di variabili è molto alto: proprio per questo, i CART sono spesso utilizzati come strumento base di varie tecniche di combinazione di stimatori. In generale, gli alberi decisionali sono facili da interpretare e da costruire ma uno degli svantaggi principali è la loro tendenza all'overfitting e sono fortemente dipendenti dalle caratteristiche del dataset di training.

Una diretta evoluzione dei modelli CART sono le tecniche di *Random Forest*. Una Random Forest è uno speciale classificatore formato da un insieme di classificatori semplici (Alberi Decisionali), rappresentati come vettori random indipendenti e identicamente distribuiti, dove ognuno di essi contribuisce per la classe più popolare in input (cfr. [34]. Ciascun albero all'interno di una Random Forest è costruito e addestrato a partire da un sottoinsieme casuale dei dati presenti nel training set: gli alberi pertanto non utilizzano quindi il set completo, e ad ogni nodo non viene più selezionato l'attributo migliore in assoluto, ma viene scelto l'attributo migliore tra un set di attributi selezionati casualmente. La casualità è un fattore che entra quindi a far parte della costruzione dei classificatori e ha lo scopo di accrescere la loro diversità e diminuirne così la correlazione. Il risultato finale restituito dalla Random Forest è che la media dei risultati di ciascun albero (nel caso di utilizzo per il *forecasting*), o la classe restituita dal maggior numero di alberi nel caso la Random Forest sia utilizzata a fini di *clustering*.

In letteratura le Random Forest ottengono risultati estremamente consistenti nelle stime probabilistiche (cfr. [9], [27], [28], [32]) e sono spesso state oggetto di confronto con i metodi parametrici classici [cfr. [21] testandoli su diversi tipi di dati. Rispetto al singolo albero decisionale, tuttavia, risulta meno intuitivo e facile da spiegare e può risultare complicata la calibrazione dei parametri nel tempo.

A rappresentare, infine, un ampio insieme di tecniche *machine learning* sono le reti neurali (*neural network*). Il termine *neural network* nasce come modellizzazione matematica di quello che in passato si riteneva essere il meccanismo di funzionamento del cervello animale (cfr. [23]). Una rete neurale è sostanzialmente uno schema di regressione non lineare (cfr. [25]) a due o più stadi ([1], [47]) costituito da strati di neuroni che, collegati tra loro da ideali bottoni sinaptici, mettono in relazione le variabili di input con quelle di output. Il neurone, in sostanza, è interpretabile come una funzione matematica (definita funzione primitiva) delle variabili esplicative ([10],[12], [18]). Il processo di apprendimento della rete neurale consiste nell'identificare i coefficienti delle funzioni di rete – sigmoidi - ([22], [17],[14]) che legano tra loro i neuroni (ed esprimono pertanto le relazioni che intercorrono tra le variabili di input a quelle di output) attraverso la minimizzazione di una funzione obiettivo ([13], [11]) espressa come scarto quadratico medio tra il valore reale dell'output ed il valore calcolato ([33], [6]).

Pur riuscendo a catturare le relazioni non-lineari e non-monotòne che intercorrono tra la PD e le variabili esplicative, tali modelli presentano numerosi inconvenienti: arbitrarietà nella scelta di molti parametri e soprattutto difficoltà di interpretazione dei risultati (spesso vengono indicati come black box).

2.2 Apprendimento non supervisionato

A differenza dell'apprendimento supervisionato, quello non supervisionato prevede l'utilizzo di dati non strutturati o senza etichetta. Le tecniche di apprendimento non supervisionato, in particolare, consentono di osservare la struttura dei dati e di estrapolare informazioni cariche di significato. In queste tecniche non esiste però una variabile o una funzione obiettivo note a priori, a differenza di quanto accade invece nell'apprendimento supervisionato. Nel nostro studio abbiamo utilizzato logiche di clusterizzazione basate su un algoritmo *K-means*. Tale algoritmo (cfr. [30]) è una metodologia di clustering che permette di suddividere N osservazioni in K cluster, nei quali ciascuna osservazione appartiene al cluster avente il punto medio a questa più prossimo: tale obiettivo viene raggiunto dalla metodologia minimizzando la varianza totale intra-cluster. Esprimendo il concetto in termini formali: dato un insieme di osservazioni $(x_1, x_2, \dots, x_N)$, dove ciascun elemento può essere rappresentato da un vettore reale a d dimensioni, il K-means clustering ha lo scopo di partizionare le N osservazioni in K $(\leq N)$ insiemi $S = \{S_1, S_2, \dots, S_K\}$ in modo da minimizzare la varianza espressa dalla WCSS (Within-Cluster Sum of Squares).

In termini matematici, l'obiettivo è il seguente:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{K} \sum_{x \in S_i} \|x - \mu_i\|^2 = \underset{S}{\operatorname{argmin}} \sum_{i=1}^{K} |S_i| \operatorname{Var} S_i \tag{6}$$

Dove μ_i è la media dei punti in S_i.

L'algoritmo standard impiega una tecnica iterativa di aggiustamento: dato un insieme iniziale di K medie $m_1^{(1)}, \dots, m_k^{(1)}$, la procedura evolve alternando le due fasi seguenti:

I. *Fase di Assegnazione (Assignment step):* viene assegnata ciascuna osservazione al cluster la cui media è caratterizzata dalla distanza euclidea minima. Matematicamente significa partizionare le osservazioni impiegando un diagramma di Voronoi (Voronoi diagram) generato dalle medie.

$$S_i^{(t)} = \left\{x_p : \left\|x_p - m_i^{(t)}\right\|^2 \leq \left\|x_p - m_j^{(t)}\right\|^2\right\} \ \forall j, i \tag{7}$$

Dove ciascun x_p è assegnato ad uno ed un solo $S^{(t)}$

II. *Fase di aggiornamento (Update step):* sono calcolate le nuove medie che costituiranno i centroidi delle osservazioni nel nuovo cluster:

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j \tag{8}$$

L'algoritmo converge quando non avvengono più cambiamenti significativi alla configurazione trovata. Rispetto alla formulazione base, sono state proposte numerose varianti atte ad incrementare le performance di ricerca del metodo. Tra le più popolari si ricorda il *K-medians* clustering, il *K-means ++* e la soft *k-means* (detta anche *Fuzzy C-means*).

3 Applicazione del Machine Learning al Rischio di Credito: stima della Probabilità di Default

L'idea alla base del nostro studio è stata quella di utilizzare le tecniche multivariate di machine learning supervisionato sopra citate per arrivare alla quantificazione del merito creditizio (Probabilità di default – PD) della clientela a fini di erogazione / concessione di nuovo credito, sfruttando poi l'efficacia della cluster analysis per giungere ad una rappresentazione discreta (scala di rating) del merito creditizio di ciascun cliente.

3.1 Campione di dati utilizzato

Lo studio in oggetto è stato condotto su un campione contenente oltre 800.000 clienti Retail di un panel di banche europee vigilate BCE, con 10 anni di informazioni storiche (da gennaio 2006 a dicembre 2016). A livello di cliente si è studiato il potere informativo e predittivo di un set di dati riconducibile alle seguenti fonti informative:

- *Credit Bureau:* sono state analizzate informazioni afferenti alle seguenti categorie di prodotti:
  - *Carte:* Importo residuo utilizzo carta, numero di contratti attivi, numero di carte in possesso;
  - *Prodotti non rateali:* numero di contratti attivi, importo accordato, importo utilizzato, importo sconfinato;

- *Prodotti rateali:* importo rata mensile, importo rate residue, importo rate scadute e non pagate, numero totale di contratti attivi;
- *Dati complessivi (banca e sistema):* numero di banche affidatarie a sistema, numero contratti attivi presso l'istituto, numero di contratti attivi a sistema, totale importi scaduti non pagati, Score Credit Bureau, presenza sofferenze a sistema, presenza di protesti a sistema, presenza di pregiudizievoli a sistema, presenza di note negative.

- *Prodotto*: a livello di prodotto sono state utilizzate informazioni relative a importo rata mensile, rapporto rata / reddito, importo richiesto, valore dell'immobile, grado dell'ipoteca e tipologia di immobile (in caso di mutuo), durata e scopo del finanziamento;
- *Informazioni socio-demografiche:* Nazionalità, Area geografica di residenza, Anni di residenza presso l'indirizzo attuale, Anzianità bancaria, Anzianità lavorativa, Età, Tipo contratto di lavoro del richiedente (tempo determinato, indeterminato, etc.), Tipo controparte (persona fisica, cointestazione, garante affidato, etc.), Professione, SAE, Situazione abitativa (es. proprietà, affitto…), Stato civile, Possesso carta di credito (aggiuntiva), reddito netto annuo da lavoro, reddito netto annuo (comprensivo di altri redditi), possesso di immobili.

3.2 Costruzione vettori di input

Per ciascuna mese di riferimento del campione (nel periodo compreso tra gennaio 2006 e dicembre 2016) è stata analizzata la dinamica dei passaggi a default (past-due a 90 giorni, inadempienze probabili e sofferenza) delle pratiche erogate in ciascun mese nei 12 mesi successivi, costruendo in tal modo la variabile target.

Dato l'obiettivo principale del nostro studio, ossia di trovare – attraverso metodologie di Machine Learning diverse – le migliori combinazioni tra le informazioni sopra citate nel prevedere l'evento default, si riportano di seguito alcune analisi grafiche finalizzate a mostrare la relazione esistente tra l'andamento delle singole variabili e il tasso di default sull'intero campione utilizzato. Tali analisi sono riportate – a titolo esemplificativo - solo per le variabili ritenute più esplicative per ciascuna area informativa considerata.

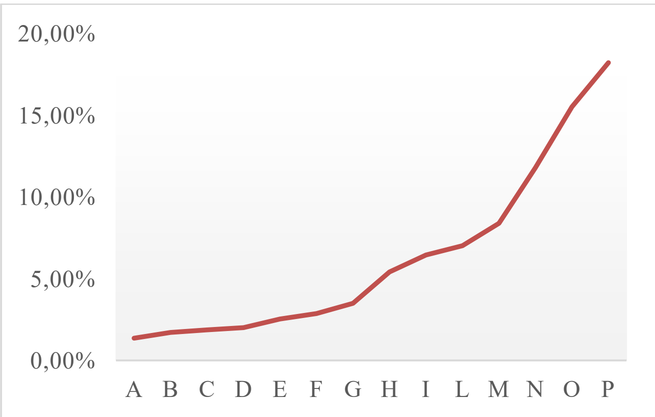


Figura 2 – Distribuzione CBScore vs Tasso di default

Come evidenziato dall'analisi grafica, lo score Credit Bureau mostra – in linea con le aspettative, un trend positivo rispetto al tasso di default: all'aumentare della classe di score / rating, aumenta la rischiosità osservata.

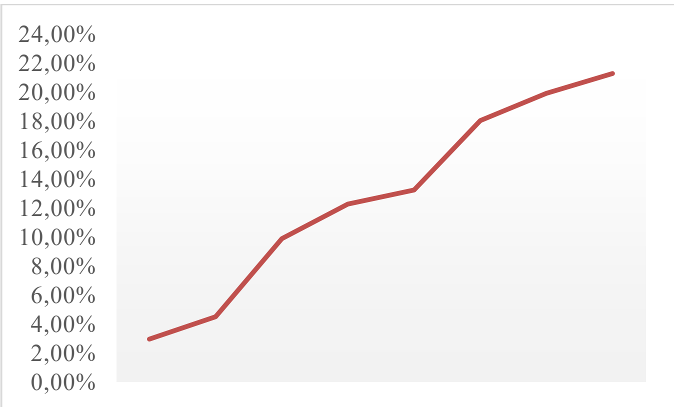


Figura 3 – Distribuzione importi scaduti vs Tasso di default

Anche nel caso dell'importo totale dei rapporti scaduti.

Figura 4 – Distribuzione rapporto rata / reddito vs Tasso di default

Per selezionare il set di variabili da sottoporre alle tecniche multivariate di machine learning supervisionato sopra menzionate, le variabili iniziali a disposizione sono state sottoposte ai "classici" trattamenti di normalizzazione legati alla gestione dei valori mancanti e al trattamento di valori anomali. Nello specifico tutte le variabili con percentuale di valori mancanti superiore al 15% sono state escluse a priori dal processo.

La selezione delle singole variabili è stata poi effettuata combinando l'analisi grafica sopra riportata con una regressione logistica e l'imposizione dei seguenti vincoli predefiniti per ciascuna variabile analizzata:

- Coerenza del segno del coefficiente con il senso economico atteso tra la variabile e il tasso di default;
- Significatività statistica del coefficiente (*p-value* inferiore al 5%);
- Capacità predittiva di ciascuna variabile misurata attraverso un Accuracy Ratio superiore al 10%.

Le variabili cosi selezionate sono state poi sottoposte ad un'analisi di correlazione, eliminando pertanto quelle con correlazione superiore a |0.5|.

La tabella seguente riporta le variabili che sono state utilizzate ai fini della costruzione del modello multivariato.

Tabella 1 – Input finali utilizzati per la selezione dei modelli multivariati di Machine Learning

Input del modello	
Sociodemografici	***Credit Bureau***
Nazionalità	**Carte**
Area geografica di residenza	Importo residuo utilizzo carta
Anni di residenza presso l'indirizzo attuale	Numero contratti attivi
Anzianità bancaria	Numero di carte in possesso
Anzianità lavorativa	**Prodotti non rateali**
Età	Numero contratti attivi
Numero garanti collegati al rapporto principale	Importo accordato
Numero componenti della pratica (co-obbligati e garanti)	Importo sconfinato
Tipo contratto di lavoro del richiedente (tempo determinato, indeterminato, etc.)	Importo utilizzato
Tipo NDG (persona fisica, cointestazione, garante affidato, etc.)	**Prodotti rateali**
Professione	Importo rate mensilizzate
SAE	Importo rate residue
Situazione abitativa (es. proprietà, affitto…)	Importo rate scadute e non pagate
Stato civile	Numero totale contratti attivi
Carta di credito - carta aggiuntiva	**Dati di sistema**
Reddito netto annuo da lavoro	Numero di banche affidatarie a sistema

Reddito netto annuo (comprensivo di altri redditi)	Numero contratti attivi presso l'istituto
Possesso immobili	Numero contratti attivi a sistema
Informazioni di prodotto	Totale importi scaduti non pagati
Importo rata mensile	Score Credit Bureau
Rapporto rata / reddito	Presenza Sofferenza a sistema
Importo richiesto	Presenza protesti a sistema
Valore dell'immobile	Presenza pregiudizievoli a sistema
Durata del finanziamento	Presenza di note negative
Grado dell'ipoteca	
Scopo del finanziamento	
Tipologia immobile	

3.3 Modelli multivariati di Machine Learning: principali evidenze

In questa sezione descriviamo le evidenze derivanti dall'applicazione degli algoritmi di machine learning utilizzate per costruire i modelli di previsione del default sul nostro campione di pratiche erogate tra gennaio 2006 e dicembre 2016 su un portafoglio di controparti Retail derivanti da un panel di banche vigilate ECB.
Come già illustrato nella sezione 2.1, la costruzione di modelli di previsione probabilità di default è un tipico problema di apprendimento supervisionato, che rappresenta una delle tecniche di machine learning più utilizzate. Nel framework di apprendimento supervisionato, un "*learner*" è rappresentato da coppie di valori di input / output sui dati storici dove i dati di input rappresentano gli attributi predefiniti da utilizzare per determinare il valore di output. I dati di input sono comunemente rappresentati come un vettore e, in funzione dell'algoritmo di apprendimento, possono consistere in valori continui e/o discreti con o senza dati mancanti. Il problema dell'apprendimento supervisionato è un tipico problema "regressivo" quando l'output è continuo, di "classificazione" quando l'output invece è di natura discreta. Obiettivo del "*learner*" è trovare una funzione che mappi correttamente i vettori di input rispetto ai valori di output. Un possibile approccio per questo mapping è memorizzare tutti i precedenti valori di coppie di input / output. Anche se questo approccio mappa correttamente le coppie di valori input / output nel dataset di training, è poco probabile che funzioni nella previsione dei valori di output se i valori di input sono diversi da quelli presenti nel dataset di training o quando quest'ultimo contiene "*noise*". Pertanto, l'obiettivo dell'apprendimento supervisionato è trovare una funzione che generalizzi al di là del dataset di training, cosi che la funzione trovata possa mappare accuratamente coppie di input / output anche su campioni out-of sample.
L'output del nostro modello è una variabile continua con valori tra 0 e 1 che può essere interpretata (sotto certe condizioni) come una stima della probabilità di andare in default nei successivi 12 mesi di vita di un contratto, date specifiche variabili di input.

Definizione del modello multivariato

Per la costruzione del modello previsivo abbiamo costruito e confrontato tra loro tre algoritmi Machine Learning:

- *Rete neurale* a tre strati, basata sull'algoritmo di *backpropagation*, completamente connessa e feed-forward;
- *Modello CART* che utilizza nella "fase di potatura" l'indice di Gini come funzione obiettivo per la riduzione dell'albero;
- *Modello Random Forest*.

Tali metodologie sono state scelte in quanto maggiormente diffuse in letteratura e i risultati a cui si è pervenuti, in termini di performance e capacità predittive sono stati confrontati con quanto invece ottenuto con il tradizionale approccio di regressione logistica (anch'esso, ricordiamo, annoverabile tra le tecniche di Machine Learning di tipo supervisionato).
Si riportano di seguito le performance ottenute:

Tabella 2 – Performance metodologie di Machine Learning

Metodologia	Accuracy Ratio	CCR
Rete Neurale	71%	86%
Random Forest	68%	81%
Albero di classificazione	66%	79%
Regressione Logistica	66%	77%

Dal confronto tra i modelli, eseguito in termini di Accuracy Ratio (AR) e Correct Classification Rate (CCR), il ranking prodotto dalle reti neurali rappresenta il modello migliore, con un valore di AR pari al 71% e di CCR pari all' 86%.

Le Random Forest hanno prodotto performance leggermente inferiori e paragonabili a quelle degli alberi di classificazione, con valori di AR rispettivamente pari a 68% e 66% e di CCR del 81% e 79%.

È nostra idea tuttavia che le reti neurali abbiano dato un risultato così poco superiore perché siamo partiti da indicatori standard. Un algoritmo di rete riuscirebbe a fare la differenza nel momento in cui andassimo ad aggiungere nuove informazioni con dati anche meno strutturati.

Molto importante è comunque sottolineare, al di là delle prestazioni ottenute, che l'albero di classificazione risulta essere, tra i tre approcci di machine learning "avanzati", l'approccio maggiormente interpretabile da un punto di vista economico e creditizio, mentre gli altri due non permettono una buona e diretta comprensione dei risultati e dei legami tra le variabili di input e quella di output.

È questo il motivo per cui si è privilegiato il ranking ottenuto con gli alberi di classificazione ai fini della calibrazione delle PD.

3.4 Calibrazione del modello

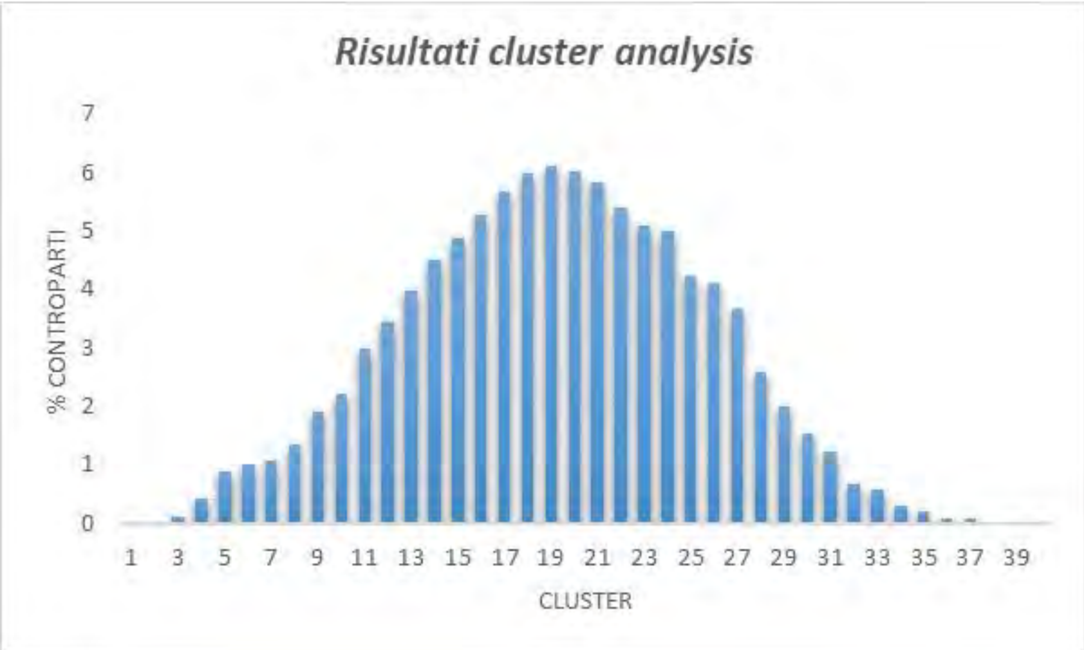
L'ultimo step di un modello di stima di PD è rappresentato dalla calibrazione degli score ai fini della loro trasformazione in PD passando attraverso la creazione di classi di rating.

Abbiamo pertanto sottoposto gli scores derivanti dal modello multivariato identificato dall'albero di classificazione a una calibrazione di tipo bayesiano, ancorando gli scores alla Central Tendency di lungo periodo e infine identificando la rating *scale* più adeguata e compliant con i requisiti normativi.

La creazione delle scale di rating è stata fatta ricorrendo ad un approccio di machine learning di tipo *unsupervised*, in particolare la **clusterizzazione *k-means*** con l'applicazione dei seguenti parametri:

- *Set iniziale dei parametri:* identificazione di 40 cluster iniziali, scala finale con massimo 11 classi di rating;
- *Split dei cluster:* concentrazione superiore al 15%.

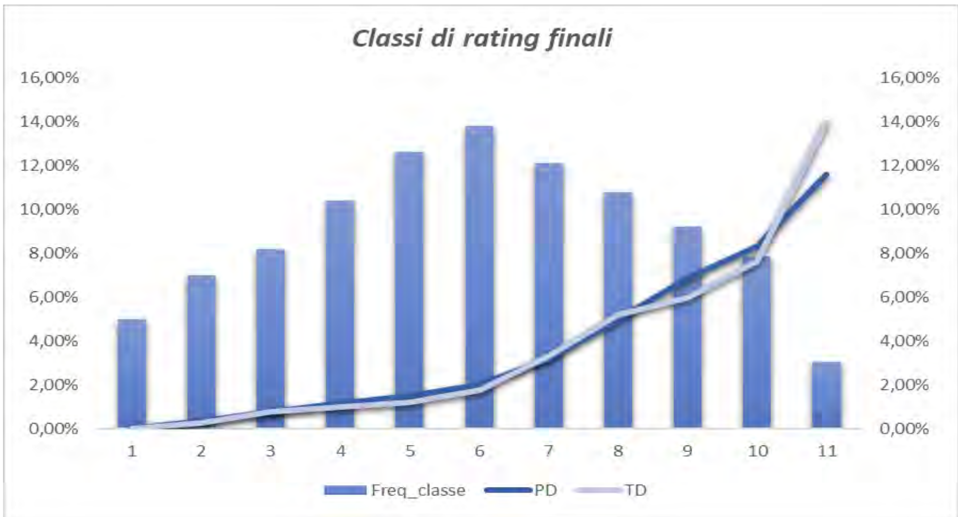
Figura 5 – Distribuzione popolazione per cluster



I cluster sopra identificati sono stati sottoposti ad un algoritmo combinatorio iterativo finalizzato a valutare:

- La forma corretta della scala di rating (simmetria e forma della campana);
- La presenza di concentrazione in ciascuna classe di concentrazione non elevata;
- Risultato test di calibrazione (Binomiale e *chi-square*);
- Monotonicità del trend di PD / Tasso di default.

Figura 6 – Classi di rating finale con applicazione Machine Learning



4 Conclusioni

Il paper ha mostrato in che modo diverse metodologie di Machine Learning possono essere applicate all'interno del framework complessivo di stima della Probabilità di default. Abbiamo in particolare comparato le tecniche più comunemente utilizzate per la modellizzazione della PD con tecniche più avanzate (Rete Neurale, Albero di classificazione e Random Forest). Per far questo abbiamo utilizzato un campione di dati molto ampio (2006-2016) basato su dati panel di diverse banche europee sotto supervisione della BCE, composto da oltre 800.000 clienti Retail e un rilevante numero di indicatori da analizzare per ciascun cliente. L'albero di classificazione, pur mostrando capacità predittiva leggermente inferiore a Random Forest e Reti Neurali, è stato considerato molto più interpretabile e pertanto utilizzato per l'ultimo step dell'applicazione: la creazione di classi di rating attraverso un algoritmo di *k-means*. È nostra idea che le reti neurali abbiano dato un risultato così poco superiore perché siamo partiti da indicatori standard. Un algoritmo di rete riuscirebbe a fare la differenza nel momento in cui andassimo ad aggiungere nuove informazioni con dati anche meno strutturati.
Un ulteriore sviluppo di questa ricerca è rappresentato dall'applicazione di ulteriori tecniche di machine learning al campione, eventualmente estendendo l'analisi anche a un portafoglio Corporate.

Stefano Bonini and Giuliana Caivano

Bibliography

[1] ALTMAN, E., BARBOZA, F., KIMURA, H. (2017). *Machine learning models and bankruptcy prediction*. Expert Systems with Applications 83: 405-417

[2] ADDO P., GUEGAN D., HASSANI B. (2018). *Credit Risk Analysis using Machine and Deep Learning*. Documents de travail du Centre d' Economie de la Sorbonne

[3] ANTUNES F., RIBEIROA B., PEREIRA F. (2017). *Probabilistic modeling and visualization for bankruptcy prediction.* Applied Soft Computing 60: 831–843.

[4] ARMINGER G., KRUPPA J., SCHWARZ A., ZIEGLER, A. (2013). *Consumer credit risk: Individual probability estimates using machine learning.* Expert Systems with Applications.

[5] AUGUSTIN L., BOULESTEIX A., T STROBL C. (2007). *Unbiased split selection for classification trees based on the Gini index.* Computational Statistics & Data Analysis.

[6] BHAVSAR H., PANCHAL M. H. (2012*). A review on support vector machine for data classification.* International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(10), pp-185.

[7] BONINI S., CAIVANO G. (2013). *Survival analysis approach in Basel2 Credit Risk Management: modelling Danger Rates in Loss Given Default parameter.* Journal of Credit Risk, 9 (1).

[8] BONINI S., CAIVANO G. (2016). *Estimating loss-given default through advanced credibility theory.* The European Journal of Finance 22 (13).

[9] BREIMAN L. (2001). *Random forests.* Machine Learning 45 (1): 5-32.

[10] BUTARU F., QUINGQUING C., BRIAN C., SANMAY D., ANDREW W. L., and AKTARE S. (2016). *Risk and risk management in the credit card industry.* Journal of Banking and Finance 72: 218–39.

[11] CHAUDHURI A., DE K. (2011). *Fuzzy support vector machine for bankruptcy prediction.* Applied Soft Computing 11: 2472–2486.

[12] CHEN C., SCHWENDER H., KEITH J., NUNKESSER R., MENGERSEN K., MACROSSAN P. (2011). *Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression.* IEEE/ACM Trans Computer Biol Bioinform 8: 1580–1591.

[13] CHEN S., HARDLE W., MORO R.A. (2006). *Estimation of Default Probabilities with Support Vector Machines.* SFB 649, Economic Risk, Berlin, SFB 649 discussion paper, No. 2006-077.
[14] CHEN M.L., TSAI C.F. (2010). *Credit rating by hybrid machine learning techniques.* Applied Soft Computing
[15] DE GROOT J. (2016). *Credit risk modeling using a weighted support vector machine.* Mathematical Institute, UTRECHT University, Master Thesis
[16] DWYER D.W., STEIN R.M. (2006). *Inferring the default rate in a population by comparing two incomplete default databases.* Journal of Banking & Finance 30: 797–810.
[17] FONSECA P., LOPES H. (2017). *Calibration of Machine Learning Classifiers for Probability of Default Modelling.* James Finance, Crowd Process Inc.
[18] GHODSELAHI A. (2011). *A hybrid support vector machine ensemble model for credit scoring.* International Journal of Computer Applications, 17(5), 1-5.
[19] HARDLE W., MORO R.A., SCHAFER D. (2005). *Predicting Bankruptcy with Support Vector Machines.* SFB 649, Economic Risk, Berlin, SFB 649 discussion paper, No. 2005-009.
[20] HARDLE W., MORO R.A., HOFFMANN L. (2010). *Learning Machines Supporting Bankruptcy Prediction*. SFB 649, Economic Risk, Berlin, SFB 649 discussion paper, No. 2010-032
[21] HERNÁNDEZ-LOBATO D., SHARMANSKA V., KERSTING K., LAMPERT C.H., QUADRIANTO N. (2014). *Mind the nuisance: Gaussian process classification using privileged noise.* Advances in Neural Information Processing Systems 27: 837–845.
[22] HOSMER D. W., LEMESHOW S. (2000). *Applied logistic regression.* 2nd ed. New York: Wiley.
[23] HUANG Z., CHEN H., HSU C.J., CHEN W.B., WU S. (2004). *Credit rating analysis with support vector machines and neural networks: a market comparative study.* Decision Support Systems 37: 543-558.
[24] KABIR M.J., KANG B.H., LIU Y., WASINGER R., ZHAO Z., XU S. (2015). *Investigation and improvement of multi-layer perception neural networks for credit scoring.* Expert Systems with Applications.
[25] KHANDANI A.E., KIM J., LO A.W. (2010). *Consumer credit-risk models via machine-learning algorithms.* Journal of Banking & Finance.
[26] KHASHMAN A. (2010). *Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes.* Expert Systems with Applications.
[27] LESSMANN S., BAESENS B., SEOW H. V., THOMAS L.C. (2015). *Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research.* European Journal of Operational Research 247 (1): 124–136.
[28] LIAW A., WIENER M. (2002). *Classification and regression by random forest.* R News 2 (3): 18-22.
[29] LOH W.Y. (2011). *Classification and regression trees.* WIREs Data Mining Knowledge Discovering 1: 14–23.
[30] MIN J. H., LEE Y. (2005). *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters.* Expert Systems with Applications, Vol. 28, Issue 4, pp.603-614.
[31] STEINBACH M., TAN P.N. (2009). kNN: *k-nearest neighbors.* In: Wu, X., Kumar, V. (eds.). The top ten algorithms in data mining. Chapman & Hall/CRC: 151–162.
[32] STEINBERG D. (2009). *CART: classification and regression trees*. In: Wu, X., Kumar, V. (eds.). The top ten algorithms in data mining. Chapman & Hall/CRC: 180–201.
[33] STROBL C., BOULESTEIX A., AUGUSTIN L. T. (2007). *Unbiased split selection for classification trees based on the Gini index.* Computational Statistics & Data Analysis 52 (1): 483-501.
[34] TSAI C. F., CHEN M. L. (2010). *Credit rating by hybrid machine learning techniques.* Applied soft computing, 10(2), 374-380.
[35] WRIGHT M. N., ZIEGLER A. (2017). *Ranger: A fast implementation of random forests for high dimensional data in C++ and R.* Journal of Statistical Software 77:1-17

A possible holistic framework to manage ICT third-party risk in the age of cyber risk

Andrea Giacchero and Jacopo Moretti (Cassa Depositi e Prestiti)[1]

Article submitted to double-blind peer review, received on 30th March 2020 and accepted on 16th February 2021

Abstract

Third-party risk for external ICT services, which concerns both the outsourced services and the third-party products, is a crucial issue for a financial institution, because a cyber attack on a vendor can be a threat for the data of its customers.

For this reason, financial institutions should adopt a holistic risk management framework to stress the effectiveness of the mitigating actions even when they engage a third-party provider.

Risk analysis of external ICT services is necessary to prepare proper mitigation plans that provide enough resources allocation. This paper proposes a possible management framework whose aim is providing indications on security measures and controls to implement against the possible sources of ICT third-party risk, and defining a proper internal process that a financial institution should adopt. In this context, the framework also embodies a model to pick the best vendor among those that a financial institution could choose for an ICT service, which is based on a risk assessment technique focused on the three information security dimensions (confidentiality, integrity, and availability) and on the Borda method.

*****

La gestione dei rischi connessi a servizi ICT esterni (sia servizi ICT in outsourcing sia quelli forniti da terze parti) è un tema cruciale per gli istituti finanziari, dal momento che un attacco cyber nei confronti di un fornitore rappresenta una minaccia per i dati dei suoi clienti.

Gli istituti finanziari dovrebbero implementare un framework di risk management olistico al fine di stressare l'efficacia delle azioni di mitigazione anche nel caso in cui decidano di affidarsi ad un fornitore terzo.

L'analisi dei rischi connessi ai servizi ICT esterni è propedeutica alla definizione di piani di mitigazione che prevedano un'adeguata allocazione di risorse. Il presente paper propone un possibile framework di gestione che mira a fornire indicazioni sulle misure di sicurezza e sui controlli da implementare in merito alle possibili fonti di rischio e a definire un robusto processo interno di gestione. Il framework prevede un modello per scegliere il miglior fornitore - tra una lista di possibili fornitori - per un servizio ICT: tale modello è basato su un risk assessment incentrato sulle tre dimensioni di sicurezza delle informazioni (riservatezza, integrità e disponibilità) e sul metodo Borda.

Keywords: third-party risk; operational risk management; ICT risk assessment.

DOI 10.47473/2020rmm0082

1. Introduction

New technologies entail an evolution of the business models that leads financial institution to implement external Information and Communication Technologies (ICT) services to quickly put in place solutions not available within the organization.

Since 2017, ICT outsourcing risk (together with legal and reputational risks) is one of the main concerns for the Top Management of the European banks according to the yearly survey on European banking system published by European Banking Authority (EBA). Furthermore, outsourcing risk is one of the Top 10 operational risks for 2019 according to a survey of Risk.net[2]. Also, the Operational Riskdata eXchange Association (ORX) highlights that third-party risk is one of the Top 5 emerging operational risks[3]. Thus, there is a wide agreement between authorities and practitioners in considering ICT third-party risk as a priority for a financial institution. The adoption of financial technology (fintech) of third-party service providers poses operational risks that need to be carefully managed to maintain an effective oversight of the emerging risks related to new technologies, which may require specialist competencies (Basel Committee on Banking Supervision, 2018).

In this context, a financial institution should increase its operational resilience, which refers to the ability of absorbing and adapting to shocks due to the impact of any disruption to the outsourced function or failure of the service provided (European Banking Authority, 2019A). External ICT services, indeed, embodies new risk sources in the concept of outsourcing risk, which is the risk that <<*engaging a third party, or another Group entity (intra-group outsourcing), to provide ICT systems or related services adversely impacts the institution's performance and risk management*>> (European Banking Authority, 2017).

The purpose of this paper is to define a third-party risk management framework for external ICT services (which concerns both the outsourced services and the third-party products) that a financial institution could implement.

[1] The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper

[2] https://www.risk.net/risk-management/6470126/top-10-op-risks-2019#cxrecs_s.

[3] https://managingrisktogether.orx.org/research/operational-risk-horizon-2019

The proposed framework, inspired by the EBA Guidelines (European Banking Authority, 2019A) and by the Circular 285/2013 of Banca d'Italia (Banca d'Italia, 2013), describes the steps that a financial institution should put in place to handle the resort to an external ICT service, providing the implementation of organizational information security measures and evaluate the adequacy of such measures. Indeed, a financial institution must identify, assess, monitor, manage, and mitigate all the risks associated with an external ICT service. More precisely, the framework provides a model that contains the guidelines to choose the best vendor for a given ICT service using a risk assessment methodology based on the three information security dimensions, namely confidentiality, integrity, and availability. However, the framework we propose can also be useful for companies of other industries that use external ICT services, indeed, the phases of the management process described in Section 3 represent general rules to follow.

When a financial institution needs to implement a new ICT service always faces a so-called 'make-or-buy' decision, namely choosing which activities <<should be provided 'in-house', and which should be bought in>> (Ford & Farmer, 1986). Walker & Weber studied the effects of cost and uncertainty (Walker & Weber, 1984), as well as the interaction between the market competition and uncertainty (Walker & Weber, 1987) on 'make-or-buy' decisions. According to the authors, the comparative analysis of production costs is the strongest criterion that influences the decision, even though both the volume uncertainty and the supplier market competition present significant effects. However, in recent years, regulators are increasingly drawing the attention also to risk management aspects.

A financial institution that decides to resort to an external ICT service must clearly define which part of the service remains under the domain of the organization and which part is handled by the vendor (Boardman & Sauser, 2008). Understanding the boundary between internal and external domain is crucial to define rôles and responsibilities in managing an external ICT service (Power, Desouza, & Bonifazi, 2006).

All the innovative projects implemented by a financial institution together with a third-party provider in the ICT field could increase the ICT risk, especially because information asymmetry could weaken the effectiveness of the oversight on vendor's information security measures (Banca d'Italia, 2019).

An external ICT service requires an agreement of whatever sort between a financial institution and a supplier. This agreement must prescribe how the supplier execute an internal process, a service, or an activity on behalf of the financial institution. The definition of the agreement with a third-party must specify whether the ICT service is an outsourced service or a third-party product. In case of an outsourced ICT service, the financial institution should assess whether would realistically be able to implement the service (even if the company has never implemented the service). Note that, as a general rule, outsourcing is a viable solution to carry out the business services of a company for several reasons, indeed, it permits e.g. to focus on core business functions, to compensate a lack of capabilities, and to acquire quickly current technologies. For larger-sized companies, ICT outsourcing is mainly motivated by strategy, while firms with smaller ICT staff and fewer resources resort to ICT vendor because of economic reasons (González, Gascó, & Llopis, 2016). In this context, ICT third-party providers bring extensive world-class resources, such as access to new technology, tools and techniques that a financial institution may not have, together with and a competitive advantage through expanded skills (Ghodeswar & Vaidyanathan, 2008). ICT vendors strengthen the value chain of their customers making available the industry best practices to use assets more effectively and efficiently (Farrell, 2004).

For many organizations, outsourcing and third-party products are 'silver bullets' to solve operational drawbacks (Power, Bonifazi, & Desouza, 2004). In particular, purchasing external ICT services represents one of the most important and successful strategies to reduce organization's ICT cost and to focus the efforts on their core business rather to ICT operational activities. However, ICT outsourcing can be the source of serious information security risks. Therefore, risk analysis of external ICT services is necessary to prepare proper mitigation plans that provide enough resources allocation (Khidzir, Mohamed, & Arshad, 2013).

As a general rule, a financial institution resorts to an outsourcing agreement when it decides to commit to a third-party the management and the responsibility of an ICT service that supports its core business, due to the lack of adequate internal resources and expertise. On the other hand, we refer to a third-party product when the external ICT service supports activities different from the core business. We refer to a third-party product even when an ICT service supports core business activities, but the government of the system remains under the responsibility of the financial institution.

A financial institution should implement a unique internal process to assess and manage the ICT third-party risk, both for the outsourced ICT services and the third-party products. However, the outsourced ICT services require a more complex approach to execute an ex-ante assessment on the potential outsourcing that involves several functions of the financial institution (e.g., Risk Management, Compliance, ICT). During the assessment, the financial institution should - inter alia - provide a cost-benefit analysis and stress both its ability to re-internalize the activities and the reliability of the potential vendors. The risk assessment framework should embody tacit and explicit knowledge together, even if the former factor is more difficult to formalise or standardise in the renewal of a third-party provider contract with respect to the latter one (Currie, 2003). In other words, as it creates a structure of formal and informal relationships, mitigation of ICT outsourcing risk also depends on relationship management between vendor and client (Levina & Ross, 2003). For the particular case of application development vendors, a company should also evaluate the size and typology of the project, and also future business (Gopal, Sivaramakrishnan, Krishnan, & Mukhopadhyay, 2003). However, today the recourse to outsourced ICT services and third-party products is a common practice among financial institutions (regardless of their dimensions), which concerns several kind of ICT services (e.g., applications, software development, software maintenance, network and server management). Thus, as ICT third-party risk management is a main concern for all the financial institutions (Haller & Wallen, 2016), the Board of Directors and senior management of a financial institution must ensure that outsourced activities are conducted in a safe-and-sound manner (Board of Governors of the Federal Reserve System, 2013).The paper is organized as follows: Section 2 provides a list of potential third-party risk sub-categories. In Section 3 and in its sub-sections is elicited a possible framework to manage ICT third-party risk while the conclusions are presented in Section 4.

2. Third-party risks

In the context of a globalized economy, the complexity of the outsourcing projects is continuing to increase also due to the number of parties involved (Power, Desouza, & Bonifazi, 2006).

Consequently, a financial institution - instead of focusing on the short term costs - should decide to outsource an activity both considering its long-term needs in terms of know-how and quality, and the risks that the outsourcing could imply (Prasad & Prasad, 2007). According to Gandhi et al. (Gandhi, Gorod, & Sauser, 2012), the identification of the third-party risk typologies and their prioritization becomes a key factor to the success of an outsourcing project and to guarantee a competitive advantage.

For these reasons, the decision concerning the outsourcing of an ICT service is subject to a periodic review (Carter, Maltz, Yan, & Maltz, 2008). Layton et al. state (Layton, Zechnich, & al., 2008) that the risk identification phase should be preliminary with respect to the decision of picking an outsourcer and that this phase should provide a risk management holistic approach for the entire outsourcing life cycle. In the first place, according to (Bott & Milkau, 2015), before resorting to an outsourcer a financial institution should manage a strategic risk which concerns a 'make or buy' decision to optimize the available resources (e.g., cash flows, workforce).

ICT third-party risk is an endogenous risk for a financial institution, because it concerns a choice between outsource or doing an activity, and between several possible vendors for a given activity (Aubert, Benoit, Patry, & Suzanne, 2005).
Nakatsu and Iacovou (Nakatsu & Iacovou, 2009) use a Delphi survey[4] to list the main risks related to ICT outsourcing. In a decreasing order of importance, the authors cite the lack of communication, poor change controls, absence of top management support, failure to manage end-user expectations, lack of customer's project management know-how, and inadequate vendor's staffing by vendor.
A complete list of potential risks that could be connected to a third-party relationship is difficult. However, below we try to synthesize the main ICT third-party risks[5]:

a) choosing an inappropriate vendor that negatively influences the execution of a project. A common error is choosing a vendor with a procurement process based on a 'race to the bottom' economic offer (the so-called 'winner's curse' phenomenon), which could push the vendor to stress its managers to achieve the expected profits even in adverse circumstances for the customer (Feeny, Lacity, & Willcocks, 2005). Thus, a financial institution should choose a vendor on the basis of several attributes like reputation, experience, and price instead of just picking the one with the lowest cost (Ruzaini, Aris, Arshad, & Mohamed, 2008);
b) insufficient expertise of the vendor which, once concluded the agreement, sends its most talented employees in search of new customers (Sullivan & Ngwenyama, 2005);
c) overdependence on a single vendor, which would be difficult to substitute in case of evident inefficiencies without an exit strategy (Bahli & Rivard, Validating Measures of Information Technology Outsourcing Risks Factors, 2005). Since many ICT projects are not separable, the management by different providers may turn out to be difficult (Fan, Suo, & Feng, 2012). However, re-internalizing an outsourced function could be a quite difficult and long process, especially in case of a full outsourcing (Harland, Knight, Lamming, & Walker, 2005), mainly because of the high costs involved in recreating the ICT department and hiring the staff for it (Earl, 1996);
d) economic aid to a vendor under financial stress (so-called 'step-in risk') to protect the company from potential reputational damages due to the vendor;
e) violation of the contractual clauses with respect to the execution of the activities from the vendor (Carter, Maltz, Yan, & Maltz, 2008);
f) loss of know-how, due both to an inadequate training of the internal staff and to an insufficient transfer of knowledge from the vendor (Verwaal, Verdu, & Recter, 2008);
g) disruption of the business continuity, which could cause huge financial losses;
h) security threats, both for possible personal data leakage and cyber-attacks. The risks linked to the storage and transfer of data stems from the faculty of accessing to systems and to customer data attributed to the vendor. Indeed, the cyber risk management in financial institutions should consider several items, including third-party risk management (e.g., especially in case of cloud computing providers who are not subject to the regulation of the financial sector authorities) (Financial Stability Board, 2018). The typology of contractual agreement directly influences potential information security threats (Alner, 2001).

Information security risk is one of the most critical risk sub-category of ICT third-party risk (Davison, 2003). Typical examples of information security risks are theft of personal data, information leakage, extraction and unauthorized manipulation of intellectual properties (Hinson, 2007). These risks, caused by lack of control on threats and vulnerabilities, refer to natural or man-made events that could have an adverse impact on organizational assets (Kaplan R. , 2004). Regardless

---

[4] <<Technique using a group of people who are either involved or interested in the research topic to generate and select a more specific research idea>> (Saunders, Lewis, & Thornhill, 2009).
[5] See also (de Sá-Soares, Soares, & Arnaud, 2014) and (Shroff & Bandi, 2018) for a comprehensive catalog information system outsourcing risks and for a list of possible misconduct risk, respectively.

of the benefits of outsourcing services, a company must consider all the activities, internal processes, and controls to protect information, data, and their underlying infrastructures. In this context, confidentiality, integrity, and availability represent the core values of information security (Vorster & Labuschagne, 2005) that a financial institution should always monitor. Confidentiality refers to the restrictions on the use of different kinds of information, while integrity is the assurance that information has not been adulterated and availability is the guarantee that only authorized users have access to information and connected assets when required (Parker, 2002). The growing interconnectivity between a financial institution and their third-parties amplifies the risk that the attack vector is a third-party vendor (Mallinder & Drabwell, 2013).

Other critical risks in ICT outsourcing projects regards complexity management, due to the use of new technologies, and team risks, both concerning communication problems and conflicts between customer and seller due to divergent work styles (Abdullah & Verner, 2012). In addition to the risks described above, a financial institution should also consider other kind of risks, such as liquidity, interest rate, pricing, legal, and foreign currency translation risk (Federal Deposit Insurance Corporation, 2019).

Third-party risk can be also related to frequent staff and senior management changes at the service provider (McCahery & de Roode, 2018). Furthermore, a company could face a strong opposition if its staff consider outsourcing as a threat to their jobs (Brooks, 2006), consequently, the situation of uncertainty caused by outsourcing could lead to low productivity, loss of motivation, and anxiety (Walden & Hoffman, 2007).

Financial institutions should make a risk analysis to weigh the dependencies from third parties in their supply chain, aiming at identifying the best solutions to improve cyber resilience (especially in case of third parties not subject to banking supervision) and at improving ICT risk management (Basel Committee on Banking Supervision, 2018). To guarantee a strong oversight on cyber risk, financial institutions should implement an ex ante risk assessment activity and an ongoing monitoring based on the same framework established for ICT risk management. An integrated approach is necessary to manage the risk assessment process across the entire supply chains (Kleindorfer & Saad, 2005). To strengthen their resilience and to deal with unexpected events, financial institutions should increase their flexibility (Waters, 2011). In particular, developing digital resilience is necessary to design business processes and technology architectures together with the cybersecurity defenses needed to protect critical information assets (Kaplan, Bailey, O'Halloran, Marcus, & Rezek, 2015). The Tiber EU (European Framework for Threat Intelligence-based Ethical Red Teaming) issued by the European Central Bank provides a set of guidelines to financial institutions to execute the penetration testing activity (on a voluntary or on a mandatory basis) aiming at verifying the cyber-resilience of the institutions. If the infrastructure outsourced and a third party is included in the scope of the test, it is necessary to include all the information about that third party (European Central Bank, 2018).

Concerning both the security of data (focusing, in particular, on personal and confidential data) and of the applications, financial institutions should choose vendors which present high information security standards levels, constantly monitoring the respect of these standards. In view of the weakest link principle, information security is a broader theme that interests all the player of the markets, instead of being a worry just for the single participant and for the suppliers of critical ICT services (European Banking Authority, 2019B). Also the G7 Cyber Expert Group - whose mission is addressing the increase in sophistication, frequency, and persistence of cyber threats in the financial sector - stressed the importance of establishing an adequate strategy and a strong framework for information security in terms of nature, dimension, complexity, risk profile, and culture of the single financial institution, especially for ICT services provided by third parties (European Commission - G7 Cyber Expert Group, 2016).

ICT third-party risks are currently exacerbated due to tighter regulatory requirements in the field of data processing and of data security (Vőneki, 2018). More precisely, for the personal data that are stored (or that are going to be stored), information security measures should be applied both to the automatic and to the manual data processing, so that the protection of any physical person is technologically neutral and does not depend on the techniques employed (European Parliament, 2016).

3. A possible ICT third-party risk management framework

The ex-ante analysis of the ICT third-party risk requires the implementation of a risk assessment based on the three drivers (confidentiality, integrity, and availability) provided by the EBA (European Banking Authority, 2019B), and on specific further information on the vendors, which could be necessary due to the riskiness of the external ICT services. The main aim of the risk assessment is to verify the adequacy of the security measures to mitigate ICT third-party risk. In case of an outsourced core function, a financial institution could provide a more detailed due diligence about the future trustworthiness of the vendor to express a risk opinion on the decision to resort to a third-party.

Risk exposure, once made explicit, converts the unexpected into an option selected intentionally, being aware of what it is selected and what it is discarded. Thus, the risk assessment essentially concerns an analysis of the vendor's financial statement and strategy to determine its creditworthiness. Furthermore, the risk assessment provides a risk concentration analysis based on the vendor portfolio for ICT services. In this context, a financial institution should continuously monitor the adequacy of the procedures and of the security measures adopted by the vendors of its external ICT services. This monitoring constitutes an ex-post analysis aiming at identifying possible changes that could undermine the stability and the performance of the same vendors.

The risk opinion on the decision to entrust to a third-party an ICT service is generally preliminary with respect to the settlement of the contractual agreements, which define the responsibilities about the management of the ICT services mentioned in the contract. In line with a risk-based approach, the higher the riskiness of an ICT service, the higher the number of the protection clauses. This opinion should also assess the risk of supplier's ecological or social misconduct (Khidzir, Mohamed, & Arshad, 2013).

Taking inspiration from EBA Guidelines (European Banking Authority, 2019A), the ICT third-party risk management internal process can be divided into three main phases:

a) **ex-ante analysis of initiative approval and vendor selection**, in which the financial institution should assess the potential impact of the agreement with a third-party in terms of operational, compliance, and reputational risk, as to verify whether the agreement increases the risk exposure. Specifically, the risk assessment should consider both the data classification and the relevance of the reference operation to evaluate – inter alia – potential risks as losing the direct control of the critical components of the external ICT service, data leakage, and unauthorized use of company's tools subsequent to a cyber-attack;
b) **contractualization**, to establish proper protection clauses, in coherence with the service configuration. For instance, the protection clauses should prescribe that the vendor respects the information security policies of the customer[6] and should provide one or more exit strategies;
c) **monitoring of the external ICT services**, which should provide (i) a monitoring procedure on vendor's activities, according to a method and a frequency compliant with the riskiness of the internal processes supported, (ii) an internal Contact Person (CP) for the external ICT service dedicated to the implementation of the security measures against potential threats and vulnerabilities, and (iii) periodic reports to the internal control functions.

Thus, a financial institution that resorts to an external ICT service must adopt a control framework for all these phases of the procurement process.

The following sub-sections provide a more detailed description of the security measures related to the three main phases of the procurement process.

3.1 Ex-ante analysis of initiative approval and vendor selection

Before the initiative approval, the financial institution should nominate an internal CP for the ICT service[7], which has the appropriate skill to classify the external ICT service, to put in place the preliminary risk assessment, and to control the activities carried out by the vendor. The CP is responsible for the management of the contract and, in particular, both for executing controls on external ICT services and for sending periodic reports to the internal control functions.

The set of external ICT service would constitute an inventory, which should contain several information (e.g., data typology, internal processes supported by the ICT service, list of users) that may contribute to make a more precise risk assessment of the ICT services. To be effective, the CPs should continuously update the inventory for any substantial change (e.g., purchase of a new external ICT service, new and different kind of data inserted in the service).

A CP, using the information of the inventory, classifies an external ICT service (before its purchase), distinguish between the outsourced services and the third-party products. According to the classification of the external ICT service classification, the CP defines a proper set of security measures. As a matter of principle, a financial institution should adopt the same set of security measures for all the external ICT services, regardless of their classification.

However, the staff that a company dedicates to these controls is generally limited, thus, it is convenient to divide external ICT services into homogeneous clusters and rank these services in terms of their riskiness. Note that, as a general rule, outsourced ICT services require a higher number of security measures, while, among the third-party products, the company should identify (at least) two different clusters to distinguish the services that require stronger measures from those with a non-material riskiness (e.g., hardware assistance).

As already mentioned, for the outsourced ICT services, the financial institution should assess its ability to re-internalize the activities if necessary. Thus, the company should maintain the technical and managerial skills to re-internalize the outsourced activities, complying with the current regulatory requirements.

Before the conclusion of an agreement with the vendor, a financial institution should perform a preliminary risk assessment. Indeed, after the classification of an ICT service within a cluster, to comply with the regulator, the financial institution should assess the inherent riskiness of that service. As a general rule, one should estimate inherent riskiness for all the contracts related to a vendor of an ICT service. The aim of the risk assessment is ranking - in a decreasing order of riskiness - the potential vendors for an ICT service, starting from those with the highest rating class.

The preliminary risk assessment should, however, estimate a residual riskiness to verify that the vendor respects the standard of the financial institution in terms of information security measures. A financial institution should perform this assessment in case of new ICT services, updates of an existing one, major incidents that significantly modify the risk profile of the vendor, and/or for significant changes in the internal processes that the ICT services support.

More precisely, the analysis provides the following steps:

- Inherent risk calculation
- Vendor's controls adequacy calculation
- Residual risk calculation

[6] Including a correct treatment of the data based on their classification (e.g. confidential, sensitive).
[7] It would be desirable to choose an employee of the ICT function.

Often, when a company decides to acquire a new ICT service, it has not historical loss data and (sometimes) process owners are inexperienced in managing the operations, thus the estimation of probability and of the potential impacts stemming from risk events is very hard. For this reason, the risk assessment approach described in this paper does not require expert opinions. Indeed, the methodology provides a list of multiple-choice questions centered around the three dimensions of information security, namely confidentiality, integrity, and availability (Khidzir, Mohamed, & Arshad, 2013).

The main aim of this risk assessment is comparing the riskiness of the potential vendors of a given ICT service for a company. This approach is coherent with a security-by-design concept in which companies use ICT services to develop their products in step with their cybersecurity. Indeed, a company should run the risk assessment during the vendor selection for an ICT service.
We propose a qualitative model that match the scores of the inherent risk and the vendor's control adequacy to proper rating classes (represented on a color scale). The model combines the inherent risk rating and the vendor's control adequacy rating by using a heat map[8] to calculate a residual risk rating.

Inherent risk calculation

A first set of multiple-choice questions concerns the estimation of the three inherent risk ratings, one for each information security dimension of the ICT service. The information needed to answer these questions should be known for a company so that the CP could estimate the inherent riskiness of the pair ICT service/vendor by its own. For illustrative purposes, Table 1 provides an example of questions and possible answers (together with a numerical score) for each of the three information security dimensions. Note that the methodology provides a 'yes or no' answer for all the questions (this is a convenient approach to make the output of the analysis easier to interpret).

Information security dimension	Question	Possible answers (score_IR)
Confidentiality	Does the ICT service handle confidential data?	Yes (1) No (0)
Integrity	Does the ICT service handle data to be included in the financial statement?	Yes (1) No (0)
Availability	Does the ICT service provide more than 100 users?	Yes (1) No (0)

Table 1 - Possible questions for the calculation of the inherent risk rating

For instance, a company could calculate this inherent risk rating assigning a score to each of the possible answers of every single question. Summing up the scores of all the questions, a company could estimate an inherent risk rating by choosing appropriate cutoffs for the range of possible scores. Table 2 lists the set of inherent risk rating classes (in a decreasing order of significance) based on an illustrative set of ranges for the quotient 'score_IR summation / max potential score_IR summation' (IR_SCORE).

Risk Rating class	IR_SCORE
Very High	100% ├ 75%
High	75% ├ 50%
Medium	50% ├ 25%
Low	25% ├ 0%

Table 2 - Inherent risk rating classes

In Table 2 we match the 4 score ranges defined to as many risk rating classes, where different levels of criticality are represented by different colors on an appropriate color scale. For instance, the red class stands for a 'Very High' inherent risk.

Vendor's controls adequacy calculation

Each vendor must provide the information to answer to a second set of multiple-choice questions that aims at verifying its internal control systems. These answers lead to a controls adequacy rating of a given vendor with respect to a given ICT service (to calculate this rating, a company should sum up the score of all the questions, as for the inherent risk rating). Table 3 provides some possible questions and answers (together with a numerical score) to determine the vendor's controls adequacy with respect to each of the three information security dimensions.

[8] A two-dimensional data representation in which colors represent the output.

Information security dimension	Question	Possible answers (score_CA)
Confidentiality	Does the vendor have a Security Operation Center?	Yes (1) No (0)
Integrity	Does the vendor shield all its devices with an anti-malware protection?	Yes (1) No (0)
Availability	Does the vendor back up the data?	Yes (1) No (0)

Table 3 - Possible questions for the estimation of the vendor's controls adequacy rating

Table 4 lists the set of vendor's controls adequacy classes based on an illustrative set of ranges for the quotient 'score_CA summation / max potential score_CA summation' (CA_SCORE).

Vendor's controls adequacy rating class	CA_SCORE
Adequate	100% ⊢ 75%
Partially adequate	75% ⊢ 50%
Partially inadequate	50% ⊢ 25%
Inadequate	25% ⊢ 0%

Table 4 - Vendor's internal controls adequacy rating classes

In Table 4 we match the 4 score ranges defined to as many vendor's internal controls adequacy rating classes, where different levels of criticality are represented by different colors on an appropriate color scale.

Residual risk calculation

Combining the inherent risk rating and the vendor's controls adequacy rating (both represented on the color scale), by using for example the illustrative heat map in Fig. 1, a company can estimate the residual riskiness for each of the information security dimensions. The colors of the heat map represent the residual risk areas: the red cells indicate 'Very High' risks, the orange cells indicate 'High' risks, the yellow cells indicate 'Medium' risks, and the 'Green' cells indicate 'Low' risks. The overall residual riskiness, which is a qualitative measure to compare the potential vendors of an ICT service, is equal to the highest residual riskiness among those of the information security dimensions due to a prudential approach.

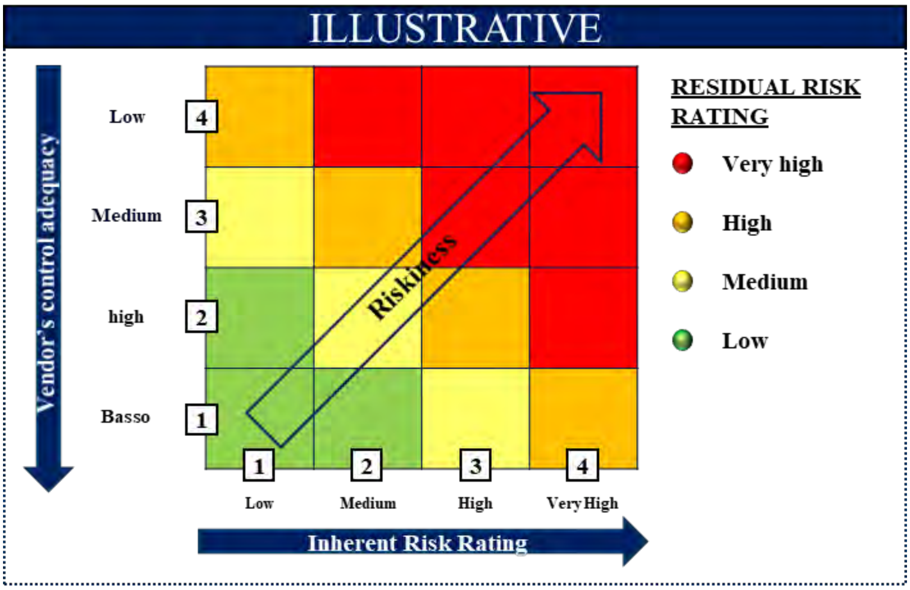


Figure 1 - Vendor's internal controls adequacy classes

The results of the preliminary risk assessment are useful but not binding for the conclusion of the agreement with a vendor. Indeed, these results provide a qualitative riskiness of a given ICT service and an indication on the maturity degree of both the technological and managerial measures that the vendor implemented. Thus, the residual riskiness is one of the criteria of the vendor selection activity. Furthermore, the company should always verify that the residual riskiness of the ICT service is compatible with its risk appetite: otherwise, the company could decide either not to implement the ICT service or to tolerate the risk. Indeed, deciding whether implementing a new external ICT service depends both on the risk appetite and on the strategic business choices.

The residual risk rating classes allow to rank the potential vendors of an ICT service in terms of their residual riskiness. However, two or more vendors could present the same residual risk for a given ICT service. In this case, the methodology provides the application of the Borda Method (Lansdowne & Woodward, 1996) to rank - in a decreasing order of riskiness - vendors with the same residual risk. More precisely, the approach provides that the vendors would be ranked by 3 criteria, namely:

1. Score of the questions to estimate vendor's controls adequacy of the confidentiality dimension (CON_CON)
2. Score of the questions to estimate vendor's controls adequacy of the integrity dimension (CON_INT)
3. Score of the questions to estimate vendor's controls adequacy of the availability dimension (CON_AVA)

The Borda Method provides to sort the scores of each criterion to achieve 6 rankings of the potential vendors of an ICT service. In case of a 'tie' between the scores of two or more vendors for a given criterion, Borda Method assigns a position equal to the average of the associated positions to the vendors in a 'tie' situation. From this hypothesis follows that, considering M_j vendors in a 'tie' situation, their ranking is equal to:

$$P_j = \frac{1}{2}\left(2C_j + 1 + M_j\right) \qquad \text{Eq. (1)}$$

Where

P_j position of the vendors in a 'tie' situation;

M_j number of vendors in a 'tie' situation in the j-th position;

$C_j = \sum_{s=1}^{j-1} M_s$ number of vendors having a better ranking with respect to those in a 'tie' position.

Then, Borda method brings the three rankings comparable using the so-called 'Borda Count' defined by the following equation:

$$b_j = \sum_k \left(N - r_{jk}\right),$$

Where

b_j Borda count of the j-th vendor;

N overall number of vendors;

r_{jk} ranking of the j-th vendor for the k-th criterion.

Then, Borda method requires to calculate the so-called Borda Rank sorting the Borda counts of each vendor for each criterion. Table 5 and Table 6 provide a numerical example concerning five vendors with 'very high' residual riskiness (the data in this table is artificial, however, the scores simulate practical cases). Note that the higher the score of a vendor's controls adequacy, the lower its riskiness and, consequently, the lower its ranking.

VENDOR	Overall residual riskiness	CON_CON	CON_INT	CON_AVA	Borda rank CON_CON	Borda rank CON_INT	Borda rank CON_AVA
Vendor A	Very high	50	30	30	5	2.5	2.5
Vendor B	Very high	45	50	30	3.5	5	2.5
Vendor C	Very high	20	30	25	1	2.5	1
Vendor D	Very high	30	25	45	2	1	5
Vendor E	Very high	45	40	35	3.5	4	4

Table 5 - Numerical example of Borda Method (part 1)

VENDOR	Borda count CON_CON	Borda count CON_INT	Borda count CON_AVA	Borda count overall	Borda rank overall
Vendor A	0	2.5	2.5	5	3
Vendor B	1.5	0	2.5	4	5
Vendor C	4	2.5	4	10.5	1
Vendor D	2	4	0	6	2
Vendor E	1.5	1	1	4.5	4

Table 6 - Numerical example of Borda method (part 2)

Table 7 contains the final ranking of the vendors and shows that the riskiest vendor is C while the less risky is B.

VENDOR	Borda rank overall
Vendor C	1
Vendor D	2
Vendor A	3
Vendor E	4
Vendor B	5

Table 7 - Final risk ranking of the numerical example

3.2 Contractualization

A proper contractual framework must consider the 'modularity' of the protection clauses regarding security measures which should both consider the reference cluster and the characteristics of the ICT service. More precisely, regardless of the cluster chosen for the ICT service, contractual agreements should at least provide the protection clauses defined by EBA's guidelines (European Banking Authority, 2019B). However, all the contractual agreements related to external ICT services must contain a clause that obliges the vendor to supply the documentation needed to the company to realize an adequate ex-post monitoring on the services. Furthermore, the agreements should also provide the possibility for the customer to execute audits on the ICT service and on-site inspections. In this context, a weak agreement could lead to difficulties in managing the contract (Smuts, Kotzé, van der Merwe, & Loock, 2015).
The outsourcing contract should provide, at the end of the outsourcing relationship, that the vendor guarantees maintenance and/or training to key personnel of the financial institution (Jothi Kandan & Idris, 2010).
The management of an ICT service outsourcing contract provides a cost for a financial institution, as well as the cost re-integrate the outsourced activity into its internal processes (Nordås, 2020), which is a faculty explicitly written in the agreement.

3.3 External ICT services Monitoring

The first line of defence of a financial institution (typically, the support and the business functions) has to put in place controls - on a continuous basis - on the external ICT services. To comply with EBA's guidelines, these controls must meet a risk-based approach, indeed, the number of controls is proportionate to the riskiness of the purchased service, to the supported activities, and/or to the cluster to which the service belongs.

The set of ex-post controls on external ICT services can be included in an inventory, in which the company specifies the characteristics of the controls (e.g., typology, frequency, reference documentation). In this context, company should send to the supplier a periodic questionnaire to identify any deterioration of the adequacy degree of the controls. In this questionnaire, the supplier must specify any cyber incidents, providing a detailed description of the event and of the root causes: the financial institution should analyze these incidents to decide whether to continue the supply relationship. More precisely, the financial institution should verify whether the supplier understood the causes and put in place effective countermeasures.

3.4 Reporting to the internal control functions

The CPs should transmit to the internal control functions of the financial institution a synthetic report with the main evidences of the control activities. This report should include, at least, a qualitative judgment on the results of the controls that highlights any criticality, the list of vendor's incidents and the security measures that the vendor adopted before and after every single incident, and indicators on the respect of contractual service levels in the reference period of the report (including a comparative analysis with subsequent periods and the indication of possible payment due to contractual penalties).

To guarantee full knowledge and governance of the risk factors that affect an external ICT service, the CP should promptly inform the internal control functions of the company over anomalies that could compromise the service or could give rise to material risks. In this manner, the CP and the internal control functions can decide whether making a communication to the governing bodies.

4. Conclusions

The recent growing resort to third parties for ICT services, which can contribute to increase the consequences of cyber risk, entails many challenges for financial institutions that have to guarantee a mitigation strategy against their ICT third-party risk exposure. In this context, the first challenge is probably that of defining a framework to increase efforts on vendor's control especially to manage cyber risk, which is still much less understood than many other risk types as credit risk or market risk (Cohen, Humphries, Veau, & Francis, 2019). This increase could even jeopardize the survival of some existing arrangements (Mourselas, 2019).

In this paper, we present a holistic ICT third-party risk management framework. The first step is an ex-ante analysis of initiative approval and vendor selection, we stress the need for a preliminary risk assessment to choose the best vendor for an ICT service. The risk assessment technique provides the three following steps:

- Inherent risk calculation, which aims at estimating a rating of the potential riskiness of the ICT service as a function of its characteristics (this activity is carried out by the CP);
- Vendor's controls adequacy calculation, for which the company requests to the vendor several information on the security measures in place;
- Residual risk calculation, obtained as a combination of the inherent risk rating and of the vendor's controls adequacy. This activity includes the application of the Borda Method when two or more vendors present the same residual risk.

Once chosen the best vendor, in the second step (contractualization), the financial institution should write down the agreement with the vendor including all the necessary protection clauses. Note that the protection clauses represent the indispensable prerequisite to guarantee an effective monitoring of the vendor on an ongoing basis, which is the main activity of the third (and last) step.

Outsourcing of ICT services (in particular, cloud services) is one of the best solutions for the financial institutions that will face the near future main trends concerning these services, namely the growing necessity of quickness, easiness, and accessibility. In this context, it is necessary to analyze the possible outcomes of this kind of outsourcing (Hanafizadeh & Zare Ravasan, 2018), because the third-party risk related to ICT services will become more and more crucial for the cyber security programs that aim at guaranteeing the integrity, confidentiality, and traceability of data. Indeed, according to Khan & Estay (Khan & Estay, 2015), a cyber-attack <<not always comes from the front door>>. Cyber security is a really relevant issue, since many silent attacks are carried out every day. Cyber risk is different with respect to the other security risks, due to the ease at which a hacker can identify vulnerabilities (Singer & Friedman, 2014). Cybercriminals aims at obtaining sensitive and personal information continuously analyzing vulnerabilities modifying their strategy due to the security measures of the targeted company (PWC, 2015). Furthermore, also a senior official at the European Central Bank affirmed that financial institutions that use external data storage and analogous digital technologies have a good chance of being hacked (Comfort, 2019). Again, according to FBI (Federal Bureau of Investigation) Director James Comey, there are big companies that have been hacked and those who don't know they have been hacked (Cook, 2014).

Financial institutions should strengthen security measures on vendors of ICT services asking adequate information security standards. Thus, financial institutions need to rethink to the ICT third-party risk mitigation[9]: on the one hand, companies should put in place a framework integrated at all the levels of the company to anticipate future trends of the threats. On the other hand, financial institutions should use more and more modern technologies (e.g., machine learning, deep learning, and analytics) and different data sources to guarantee an effective vendor selection.

Taking into consideration agency theory and transaction cost theory, we identify four main risk scenarios associated with external ICT services: lock-in, contractual amendments, unexpected transition and management costs and disputes and litigation (Bahli & Rivard, 2003). These scenarios are mainly due to the absence of a proper oversight on vendor's activity that could entails many drawbacks, in terms of (i) loss of control, (ii) weakening of both the development ability and creativity, (iii) decrease of the employee motivation (which could intend the outsourcing of an activity as a potential source for job losses), (iv) high economic costs for the transactions, and (v) threats for the data confidentiality. These drawbacks could be exacerbated in case of an ill-defined contract, if the monitoring of the service levels is insufficient or when the financial institutions did not define an exit strategy to prevent potential criticalities.
Theft of intellectual property is another kind of strategic risks related outsourcing (Aron, Clemons, & Reddi, 2005). Therefore, trust between the two parties is one of the key factors for success in the outsourcing arrangement, useful to avoid legal action and to establish a long-term relationship (Babin, Bates, & Sohal, 2017).

The management of a financial institution has to build key in-house capabilities and to learn how to manage outsourcing (Lacity, Khan, & Willcocks, 2009). This aspect is crucial because the client company's staff may see outsourcing as a threat to their jobs and, thus, may oppose to this solution (Brooks, 2006). Last but not least, fourth-parties (namely the supplier's suppliers) could be an additional risk source in the context of ICT third-party risk (Awasthi, Govindan, & Gold, 2018). A possible strengthening of the ICT third-party risk management framework is a theme left for future research. In conclusion, the need of implementing an effective control framework for the ICT third-party risk is increasingly debated among financial institutions because it represents a priority in which they must invest in the near future.

Andrea Giacchero and Jacopo Moretti

[9] See (de Sá-Soares, Soares, & Arnaud, 2014) for a comprehensive catalog of mitigation actions.

Bibliography

- Abdullah, L., & Verner, J. (2012). Analysis and application of an outsourcing risk framework. *The Journal of Systems and Software, 85*(8), 1930-1952.
- Alner, M. (2001). The effects of Outsourcing in Information Security. *Information Systems Security, 10*(2), 35-43.
- Aron, R., Clemons, E., & Reddi, S. (2005). Just right outsourcing: understanding and managing risk. *Journal of Management Information Systems, 22*(2), 37-55.
- Aubert, B. A., Dussault, S., Patry, M., & Rivard, S. (1999). Managing the risk of IT outsourcing. *In Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, pp. (pp. 10-pp). IEEE.
- Aubert, B. A., Patry, M., & Suzanne, R. (2005, Fall). A Framework for Information Technology Outsourcing Risk Management. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 36*(4), 9-28.
- Awasthi, A., Govindan, K., & Gold, S. (2018). Multi-tier sustainable global supplier selection using a fuzzy AHP-VIKOR based approach. *International Journal of Production Economics, 195*, 106-117.
- Babin, R., Bates, K., & Sohal, S. (2017). The Role of Trust in Outsourcing: More Important Than the Contract? *Journal of Strategic Contract and Negotiation, 3*(1), 38-46.
- Bahli, B., & Rivard, S. (2003). The Information Technology Outsourcing Risk: a Transaction Cost and Agency theory-based Perspective. *Journal of Information Technology, 18*(3), 211-221.
- Bahli, B., & Rivard, S. (2005). Validating Measures of Information Technology Outsourcing Risks Factors. *Omega, 33*(2), 175-187.
- Banca d'Italia. (2013, December). *Circolare n. 285 del 17 dicembre 2013 34° aggiornamento.* Retrieved from https://www.bancaditalia.it/compiti/vigilanza/normativa/archivio-norme/circolari/c285/aggiornamenti/Circ-285-Rist-int-34-agg.pdf
- Banca d'Italia. (2019, December). *Indagine Fintech nel sistema finanziario italiano.* Retrieved from https://www.bancaditalia.it/compiti/vigilanza/analisi-sistema/approfondimenti-banche-int/Allegato_2_Indagine_Fintech.pdf
- Basel Committee on Banking Supervision. (2018, December). Cyber-resilience: Range of practices. Retrieved from https://www.bis.org/bcbs/publ/d454.pdf
- Basel Committee on Banking Supervision. (2018, February). Sound Practices: Implications of FinTech Development for Banks and Bank Supervisors. Retrieved from https://www.bis.org/bcbs/publ/d431.pd
- Board of Governors of the Federal Reserve System. (2013). *Guidance on Managing Outsourcing Risk.* Retrieved from https://www.federalreserve.gov/supervisionreg/srletters/sr1319a1.pdf
- Boardman, J., & Sauser, B. (2008). Systems thinking: Coping with 21st century problems. *CRC Press*.
- Bott, J., & Milkau, U. (2015). Outsourcing Risk: A Separate Operational Risk Category? *Journal of Operational Risk, 10*(3).
- Brooks, N. (2006). Understanding IT Outsourcing and its Potential Effects on IT Workers and Their Environment. *The Journal of Computer Information Systems, 46*(4), 46-53.
- Carter, J., Maltz, A., Yan, T., & Maltz, E. (2008). How procurement managers view low cost countries and geographies. *International Journal of Physical Distribution & Logistics Management, 38*(3), 224-243.
- Cohen, R. D., Humphries, J., Veau, S., & Francis, R. (2019). An investigation of cyber loss data and its links to operational risk. *Journal of Operational Risk, 14*(3).
- Comfort, N. (2019, August 19). *ECB Says the Next European Bank Hack Is Just a Matter of Time*. Retrieved from Bloomberg: https://www.bloomberg.com/news/articles/2019-08-19/ecb-says-the-next-european-bank-hack-is-just-a-matter-of-time
- Cook, J. (2014, October 6). FBI Director: China Has Hacked Every Big US Company. *Business Insider*. Retrieved from http://www.businessinsider.com/fbi director-china-has-hacked-every-big-us-company-2014-10
- Currie, W. (2003). A knowledge-based risk assessment framework for evaluating web-enabled application outsourcing projects. *International Journal of Project Management, 21*(3), 207-217.
- Das, T., & Teng, B. (2000). Instabilities of strategic alliances: an internal tensions perspective. *Organization Science, 11*(1), 77-101.
- Davison, D. (2003, December). *Top 10 Risks of Offshore Outsourcing*. Retrieved from ZDNet: https://www.zdnet.com/article/top-10-risks-of-offshore-outsourcing/
- de Sá-Soares, F., Soares, D., & Arnaud, J. (2014). Towards a theory of information systems outsourcing risk. *Procedia Technology, 16*, 623-637.
- Earl, M. J. (1996). The risk of outsourcing IT. *Sloan Management Review, 37*(3), 26–32.
- European Banking Authority. (2017, September 11). Guidelines on ICT Risk Assessment under the Supervisory Review and Evaluation Process - SREP. *EBA/GL/2017/05*.
- European Banking Authority. (2019A, February). Final Report on EBA Guidelines on outsourcing arrangements. *EBA/GL/2019/02*.
- European Banking Authority. (2019B, November). EBA Guidelines on ICT and security risk management. *EBA/GL/2019/04*.

- European Central Bank. (2018, May). TIBER-EU FRAMEWORK (How to implement the European framework for Threat Intelligence-based Ethical Red Teaming). Retrieved from https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework.en.pdf
- European Commission - G7 Cyber Expert Group. (2016, October). *G7 fundamental elements of cybersecurity in the financial sector*. Retrieved from https://ec.europa.eu/info/system/files/cybersecurity-fundamental-elements-11102016_en.pdf
- European Parliament. (2016). General Data Protection Regulation. *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
- Fan, Z.-P., Suo, W.-L., & Feng, B. (2012). Identifying risk factors of IT outsourcing using interdependent information: An extended DEMATEL method. *Expert Systems with Applications, 39*(3), 3832–3840.
- Farrell, D. (2004). Beyond offshoring. *Harvard Business Review, 82*(12), 82-90.
- Federal Deposit Insurance Corporation. (2019, June). Compliance Examination Manual: Third Party Risks. Retrieved from https://www.fdic.gov/resources/supervision-and-examinations/consumer-compliance-examination-manual/documents/7/vii-4-1.pdf
- Feeny, D., Lacity, M., & Willcocks, L. (2005). Taking the measure of outsourcing providers. *MIT Sloan Management Review, 46*(3), 41-48.
- Financial Stability Board. (2018, November). *Cyber Lexicon.* Retrieved from https://www.fsb.org/wp-content/uploads/P121118-1.pdf
- Ford, D., & Farmer, D. (1986). Make or buy—a key strategic issue. *Long Range Planning, 19*(5), 54-62.
- Gandhi, J., Gorod, A., & Sauser, B. (2012). Prioritization of outsourcing risks from a systemic perspective. *Strategic Outsourcing: An International Journal, 5*(1), 39-71.
- Ghodeswar, B., & Vaidyanathan, J. (2008). Business process outsourcing: an approach to gainaccess to world-class capabilities. *Business Process Management Journal, 14*(1), 23-38.
- González, R., Gascó, J., & Llopis, J. (2016). Information Systems Outsourcing Reasons and Risks: Review and Evolution. *Journal of Global Information Technology Management, 19*(4), 223–249.
- Gopal, A., Sivaramakrishnan, K., Krishnan, M., & Mukhopadhyay, T. (2003). Contracts in Offshore Software Development: An Empirical Analysis. *Management Science, 49*(12), 1671-1683.
- Haller, J., & Wallen, C. (2016). *Managing third party risk in financial services organizations: a resilience-based approach.* Retrieved from https://resources.sei.cmu.edu/asset_files/whitepaper/2016_019_001_473742.pdf
- Hanafizadeh, P., & Zare Ravasan, A. (2018). An empirical analysis on outsourcing decision: the case of e-banking services. *Journal of Enterprise Information Management, 31*(1), 146-172.
- Harland, C., Knight, L., Lamming, R., & Walker, H. (2005). Outsourcing: assessing the risks and benefits for organisations. *International Journal of Operations & Production Management, 25*(9), 831-850.
- Hinson, G. (2007, December). Top Information Security Risk for 2008. CISSP Forum. Retrieved from http://www.naavi.org/cl_editorial_08/Top_information_security_risks_for_2008.pdf
- Jothi Kandan, M., & Idris, N. A. (2010). *Guidelines on Information Security in ICT Outsourcing.* CyberSecurity Malaysia. Retrieved from https://www.cybersafe.my/pdf/guidelines/Guidelines_on_Information_Security_in_ICT_Outsourcing.pdf
- Kaplan, J., Bailey, T., O'Halloran, D., Marcus, A., & Rezek, C. (2015). *Beyond Cybersecurity: Protecting Your Digital Business.* John Wiley & Sons.
- Kaplan, R. (2004). A matter of trust. In *Information Security Risk Management Handbook.* Auerbach Publications: Boca Raton.
- Khan, O., & Estay, D. A. (2015). Supply chain cyber-resilience: Creating an agenda for future research. *Technology Innovation Management Review, 5*(4), 6-12.
- Khidzir, N. Z., Mohamed, A., & Arshad, N. H. (2013, December). ICT Outsourcing Information Security Risk Factors: An Exploratory Analysis of Threat Risks Factor for Critical Project Characteristics. *Journal of Industrial and Intelligent Information, 1*(4), 218-222.
- Kleindorfer, P., & Saad, G. (2005). Managing disruption risks in supply chains. Production and operations management. *14*(1), 53-68.
- Lacity, M., Khan, S., & Willcocks, L. (2009). A Review of the IT Outsourcing Literature: Insights for practice. *The Journal of Strategic Information Systems, 18*(3), 130–146.
- Lansdowne, Z. F., & Woodward, B. S. (1996). Applying the borda ranking method. *Air Force Journal of Logistics, 20*(2), 27-29.
- Layton, M., Zechnich, D., & al., e. (2008). The risk intelligent approach to outsourcing and offshoring. *Risk Intelligence Series*(8). Retrieved from www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_risk_riskintelligent_
- Levina , N., & Ross, J. (2003). From the Vendor's Perspective: Exploring the Value Proposition in Information Technology Outsourcing. *MIS Quarterly, 27*(3), 331-364.
- Mallinder, J., & Drabwell, P. (2013). Cyber security: A critical examination of information sharing versus data sensitivity issues for organisations at risk of cyber attack. *Journal of business continuity & emergency planning, 7*(2), 103–111.
- McCahery, J., & de Roode, A. (2018). Governance of financial services outsourcing: Managing misconduct and third-party risks. *European Corporate Governance Institute (ECGI)-Law Working Paper No. 417/2018, Tilburg Law School Research Paper Forthcoming*.

- Mourselas, C. (2019, June 10). *Financial firms toil to meet new EU rules on outsourcing*. Retrieved from https://www.risk.net/risk-management/6714296/financial-firms-toil-to-meet-new-eu-rules-on-outsourcing
- Nakatsu, R., & Iacovou, C. (2009). A comparative study of important risks factors involved in offshore anddomestic outsourcing of software development projects: A two-panel Delphi study. *Information & Management, 46*(1), 57–68.
- Nordås, H. K. (2020). Make or buy: Offshoring of services functions in manufacturing. *57*(2), 351-378.
- Parker, D. (2002). Toward a New Framework for Information Security. In S. Bosworth, & M. Kabay, *Computer Security Handbook. 4th ed.* New York: John Wiley & Sons.
- Power, M., Bonifazi, C., & Desouza, K. (2004). The ten outsourcing traps to avoid. *Journal of Business strategy, 25*(2), 37-42.
- Power, M., Desouza, K., & Bonifazi, C. (2006). *The Outsourcing Handbook: How to Implement a Successful Outsourcing Process.* London: Kogan Page Publishers.
- Prasad, A., & Prasad, P. (2007). Moving out: toward understanding the complexity of outsourcing in the age of globalization. *Business Renaissance Quarterly, 2*(3), 67-91.
- PWC. (2015). Insurance 2020 & Beyond: Reaping The Dividends Of Cyber Resilience. Retrieved from http://www.pwc.com/gx/en/insurance/publications/assets/reaping-dividends-cyber-resilience.pdf
- Ruzaini, S., Aris, H. S., Arshad, N. H., & Mohamed, A. (2008). Conceptual framework of risk management in IT outsourcing project. *WSEAS Transactions on Information Science & Applications, 5*(4), 816-831.
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students.* Pearson education.
- Shroff, B. P., & Bandi, R. K. (2018). Impact of IT Multisourcing on vendor opportunistic behaviour-A research framework. *ACIS 2018 Proceedings.* Sidney.
- Singer, P., & Friedman, A. (2014). *Cybersecurity: What Everyone Needs To Know.* Oxford, UK: Oxford University Press.
- Smuts, H., Kotzé, P., van der Merwe, A., & Loock, M. (2015, October). Threats and opportunities for information systems outsourcing. *In 2015 International Conference on Enterprise Systems (ES)* (pp. 110-120). IEEE.
- Sullivan, W., & Ngwenyama, O. (2005). How are public sector organizations managing IS outsourcing risks? An Analysis of Outsourcing Guidelines from three Jurisdictions. *Journal of Computer Information Systems, 45*(3), 73-87.
- Verwaal, E., Verdu, A., & Recter, A. (2008). Transaction Cost and organizational learning in strategic outsourcing relationships. *International Journal of Technology Management, 41*(1-2), 38-54.
- Vőneki, Z. T. (2018, December). OPERATIONAL RISK AFTER THE CRISIS. *Economy and Fincance (GÉP), 5*(4). Retrieved from http://www.bankszovetseg.hu/magunkrol.cshtml?lang=eng
- Vorster, A., & Labuschagne, L. E. (2005). A framework comparing different information security risk analysis methodology. *In Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, (pp. 95-103).
- Walden, E. A., & Hoffman, J. J. (2007). Organizational form, incentives and the management of Information Technology. Opening the black box of Outsourcing. *Computers & Operations Research, 34*(12), 3575-3591.
- Walker, G., & Weber, D. (1984). A transaction cost approach to make-or-buy decisions. *Administrative science quarterly, 29*(3), 373-391.
- Walker, G., & Weber, D. (1987). Supplier competition, uncertainty, and make-or-buy decisions. *Academy of Management journal, 30*(3), 589-596.
- Waters, D. (2011). *Supply chain risk management: vulnerability and resilience in logistics.* London, UK: Kogan Page Publishers.

Design of an algorithm for an adaptive Value at Risk measurement through the implementation of robust methods in relation to asset cross-correlation

Marco Bagnato (SoftJam), Anna Bottasso (University of Genoa) and Pier Giuseppe Giribone (Banca CARIGE – University of Genoa)

Article submitted to double-blind peer review, received on 5th March 2021 and accepted on 8th April 2021

Abstract

This study proposes an algorithmic approach for selecting among different Value at Risk (VaR) estimation methods. The proposed metaheuristic, denominated as "Commitment Machine" (CM), has a strong focus on assets cross-correlation and allows to measure adaptively the VaR, dynamically evaluating which is the most performing method through the minimization of a loss function. The CM algorithm compares five different VaR estimation techniques: the traditional historical simulation method, the filtered historical simulation (FHS) method, the Monte Carlo method with correlated assets, the Monte Carlo method with correlated assets which uses a GARCH model to simulate asset volatility and a Bayesian Vector autoregressive model. The heterogeneity of the compared methodologies and the proposed dynamic selection criteria allow us to be confident in the goodness of the estimated risk measure. The CM approach is able to consider the correlations between portfolio assets and the non-stationarity of the analysed time-series in the different models. The paper describes the techniques adopted by the CM, the logic behind model selection and it provides a market application case of the proposed metaheuristic, by simulating an equally weighted multi-asset portfolio.

Key Words:

Value at Risk (VaR), Historical VaR, Filtered Historical Simulation (FHS), Monte Carlo VaR, GARCH volatility model, Bayesian Vector Autoregressive (BVAR), Commitment Machine (CM)

DOI 10.47473/2020rmm0083

1) Introduction

The adoption of VaR measures in financial markets dates back to the 90s; since then, the available models for VaR estimation have increased both in quantity and in complexity. However, there is no preferred VaR method that performs "better" in absolute terms; each methodology considers a particular aspect of the return distribution and its performance depends on the context and on the specifications used to implement each model.

An important risk factor for any investment portfolio is the cross correlation of assets, which tends to increase during periods of financial crisis [1]. Another important risk factor is the time-varying behavior of conditional volatility (see section 3). Therefore, the objective of this work is to create a Risk Management system based on models that take into account the cross-correlation between assets and the trend of variance over time, and to propose a possible solution to the problem of selecting among several VaR models.

In particular, the proposed solution consists in a flexible and adaptive selection criterion that allows to choose, day by day, the estimation of the VaR that reasonably represents financial market conditions. We adopted five VaR models, whose structural principles are described in section 3: traditional historical VaR, filtered historical simulations (FHS) VaR, Monte Carlo VaR with cross correlated assets, Monte Carlo GARCH VaR with cross correlated assets, Bayesian VaR. These different models are evaluated considering different aspects of the time series; three out of these five methods examine the correlation among portfolio assets, while the other two, based on the "historical" approach (historical VaR and filtered historical VaR), are less sophisticated but quite flexible and easy to implement.

Starting from the cited methods, a validation phase is carried out, based on the frequency of VaR threshold violation, with an approach based on backtesting and statistical tests to verify the hypothesis of unconditional coverage (binomial test and traffic light approach). Subsequently, a CM is proposed.

The CM evaluates the five approaches day by day by calculating every day, for each of them, the value assumed by a given loss function in the previous days. The model with the best performance is then chosen to estimate the VaR of the following day. The use of loss functions for evaluating a VaR model has many examples in related literature. A particular attention has been directed towards the use of a loss function that weights the negative differences between VaR and returns according to a quadratic function [17]. Section 4.1 presents a theoretical discussion of the properties of the loss function adopted in this study. Two different variants have been tested based on the number of observations used to calibrate the CM. For each of these two variants, the overall performance and selection capacity of the CM have been assessed, by analyzing the losses with respect to the VaR threshold and the frequency of VaR violations. The results clearly indicate that the CM is able to make an efficient selection among the various methods, by choosing VaR thresholds that are less likely to be violated and register the smallest losses in terms of negative differences between VaR and actual returns.

The proposed risk management approach is extremely customizable. Indeed, thanks to the flexibility of the code written in MATLAB, the CM approach can be used to select among a great variety of VaR methods.

2) Dataset

For the purposes of the analyses introduced in section 1, we built an equally weighted portfolio by using 4 historical time series retrieved from the info-provider Bloomberg®. These series track four different indices representing three of the main asset classes available to investors (equities, gold, bonds). The components of these indices are representative of the investment choices of most financial intermediaries and allow us to represent a balanced portfolio:

- European Stock Index (SXXP Index): the Stoxx 600 index tracks the trend of large, mid and small cap stocks in 17 different European countries. With its 600 components, it allows to simulate a highly diversified equity portfolio across the UK, Switzerland and the Eurozone.
- US Stock Index (RAY Index): The Ray Index includes 3000 listed companies which represent (in terms of market cap) 98% of the universe of US listed shares, allowing US stock markets to be incorporated into the portfolio.
- World Bond Index (Legatruh Index): the Bloomberg Barclays Global Aggregate Index collects investment grade debt listed on 24 markets, in both developed and emerging economies. The inclusion of this index allows to increase diversification by adding a second asset class distinct from the stock market and diversifying the geographic risk.
- Gold (XAU USD currency): this series tracks the historical exchange rate between gold and the US dollar. Gold has traditionally been considered a safe-haven asset and its inclusion can offer significant diversification potential.

Since the European stock index is denominated in euro, we have retrieved the historical Euro/Dollar exchange rates for the analysed period and used it to convert all the data into dollars. In order to achieve a reasonable sample size, we decided to analyze the data of the daily closing prices for the period from 1st June 2000 to 30st September 2020. This time span contains a total of 5305 market observations for each asset considered. The analysis requires the choice of a time window that allows to dynamically evaluate the evolution of the risk measures and of all other relevant variables (particularly the evolution of cross correlations in order to have a correct measurement of portfolio risk). This observation window ('rolling windows') must be large enough to be statistically significant, but at the same time it should not be too wide in order to concretely capture the effects of relatively short-term shocks (for example the collapse and subsequent recovery of the markets due to the Covid-19 pandemic in the spring of 2020).

In order to balance the above-mentioned trade-off, in accordance with the practice used for this type of analysis, we decided to use a rolling window of 260 observations, which is equal to one year and it is considered large enough for an overall analysis of market risk. Scientific literature [8] shows a remarkable consistence of VaR estimations calculated on data from a 250-day time window. Such consistence is measured by root mean squared relative bias, that is a measure of the relative distance between different risk measures calculated for a given interval. (See tab A2 of [8] for a comparison with other observation windows).

Unless otherwise specified, in the context of this study, all the measurements obtained will always refer to the 260-day rolling window. This means that, for example, the VaR calculated for the 270th daily observation will be calculated on the data ranging from the 10th to the 269th day. At each iteration, the calculation moves forward by a step equal to one day, so the VaR of the 271st day will be calculated on the data ranging from the 11th day to the 270th day, and so on; the same logic is maintained when we consider all the inputs that contribute to VaR estimation (mean, variance, correlation, etc.) that are always calculated on the same rolling window.

Figure 2.1 Historical trend of portfolio components

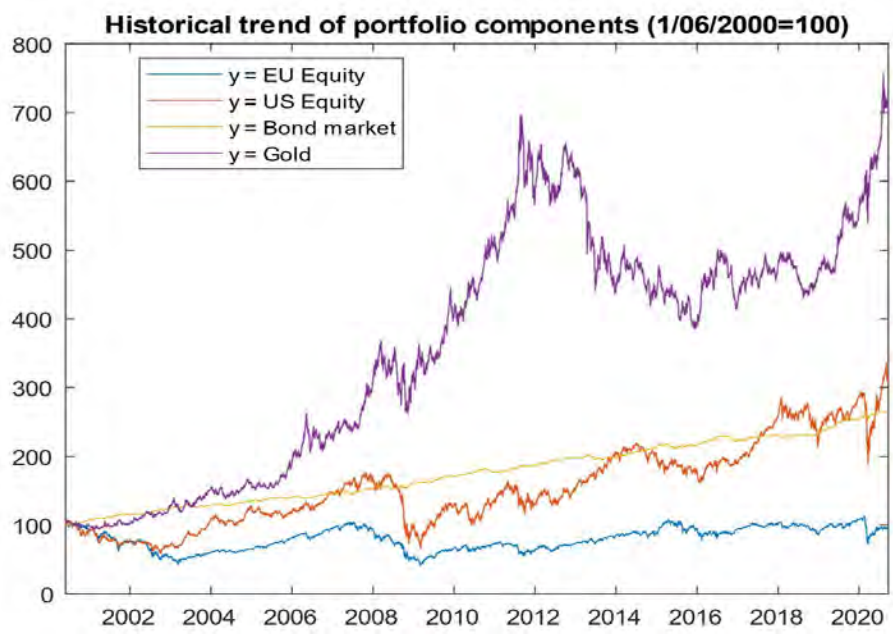


Figure 2.1 shows the trend of the 4 indices over time, after normalizing their price to 100 on the first observation day (1st June 2000).

In order to conduct a correct risk analysis, it is necessary to consider the characteristics of daily returns distribution both for the four indices considered and for the portfolio. Table 2.1 shows the main descriptive statistics (minimum, maximum, median and the first four moments) for each historical time series. The various data are calculated on the entire history available without the use of rolling windows.

Table 2.1 Analysis of portfolio and components time series

	Distributional and descriptive statistics of daily returns						
	Min	Max	Median	Average	St. Dev.	Kurtosis	Skewness
US equity	-11%	10%	0,026%	0,006%	1,21%	10,49	-0,22
EU equity	-12%	13%	0,061%	0,032%	1,44%	12,06	-0,12
Bond market	-2%	2%	0,024%	0,020%	0,16%	11,60	-0,41
Gold	-9%	11%	0,045%	0,042%	1,07%	9,19	-0,19
Portfolio	-7%	6%	0,04%	0,025%	0,65%	12,77	-0,27

All series show a high level of kurtosis and this seems to suggest a non-normal distribution for our data. In order to check for data distribution, we performed a Kolmogorov Smirnov test at 5% significance level that suggested us to reject the normality hypothesis for all the four time series (p value ~ 0 for all the series).

This allows us to reject the hypothesis of normality of our dataset and prevents the adoption of the variance-covariance method in the VaR calculation. Consequently, this approximated method of estimation has not been adopted in this study. The second consideration is about the most volatile among the indices i.e., the two equity indices. The matrix of daily correlations across the indices, reported in Table 2.2, shows that the greatest correlation is observed between the US stock index and the European stock index.

Table 2.2 Cross correlation matrix of portfolio components

Cross correlation matrix of portfolio components				
	EU Equity	US Equity	Bond market	Gold
EU Equity	1,00	0,51	-0,29	-0,01
US Equity	0,51	1,00	-0,18	0,002
Bond market	-0,29	-0,18	1,00	0,18
Gold	-0,01	0,002	0,18	1,00

The correlation between portfolio assets can be considered as a possible risk factor. The greater riskiness of the two equity indices has been confirmed by the Euler decomposition of portfolio risk, that suggests that more than 75% of the portfolio volatility is generated by the two most volatile equity indices. More specifically, the Euler decomposition attributes portfolio risk among the portfolio components as follows: 32,3% to the European stock index, 43% to the American stock index, 1,2% to the bond market index, 23,5% to the gold.

This calculation highlights how the correlation structure is by itself a risk factor, as it has the potential to amplify the losses due to the most volatile indices in the portfolio.

Another important feature of the cross-correlations between assets is the opportunity to use them to build less procyclical VaR models. Over the sample period, the two major negative events (the 2008 crisis and the Covid shock) came after a long period of positive equity market returns; in both cases, the value of the portfolio reached an all-time high just before the crisis. This is an obvious issue for risk management: the most common models of VaR are strongly backward looking, with obvious negative effects when suddenly indices shift from growth to collapse. However, cross-correlations between assets can help to solve this problem: in both cases mentioned above, the correlation between the two equity indices started to rise before the onset of the crisis. In the case of the 2008 crisis, in the previous two years both the variance of each of the two indices and their covariance increased, while before the 2020 crisis the two variances were stable.

3) VaR methodologies

After choosing the dataset we have built a risk management system that can serve as a basis for the analyses. In particular, we focused on the calculation of VaR and Expected Shortfall (ES). Although the analysis of this work focuses on the VaR threshold, the CM can also estimate an ES value from the corresponding VaR threshold for every time step. This makes the approach more versatile and offers - in perspective - greater possibilities for evaluating its performance and a better risk analysis with a coherent and subadditive risk measure (the ES). In fact, the CM selects a method for simulating price distribution that can be used both for VaR and for ES calculation.

As already mentioned, the value of the VaR threshold has been calculated on a daily basis, thus returning a total of 5043 VaR measurements for each considered method. All the methodologies have been estimated on a 95% confidence interval to calculate the losses occurring in the market day following the calculation date.

For the purpose of calculating these two quantities, 5 different methods have been implemented in order to take into account both sudden changes in variance (or, in other words, to distinguish between conditional and unconditional variance) and the effects of cross-correlations. These 5 methods can be classified into 3 families of models:

- Base Methodology: Historical VaR. This is a very simple method that is affected by rather simplifying assumptions on the future distribution of returns. However, given its computational simplicity and its sensitivity to negative tails of market returns, we decided to include it in the system and employ it together with other more sophisticated methods. A more advanced version has also been adopted (Historical VaR with filtered simulations, or FHS) which takes into account the short-term conditional volatility for each calculation date.

- Stochastic Differential Equation (SDE) Integration: Monte Carlo methods. The Monte Carlo method is a numerical simulation technique based on the integration of stochastic differential equations, which in this case is applied to the performance of the returns of the selected assets. This is an extremely flexible simulation method; in this work it has been implemented by simulating the effects of the correlation between assets on the dynamics of process innovations in two different versions (with and without GARCH volatility to consider the heteroskedasticity of the return time series).

- Econometric Bayesian methods: Bayesian Vector Autoregressive. This methodology is more complex than the Monte Carlo method, both theoretically and in terms of implementation. In particular, the Bayesian method allows us to simulate the joint trend of correlations and volatility in an extremely flexible way, including them in the stochastic component of the model.

We now proceed to expose the theory, the operational principles and the implementation of the methodologies discussed.

3.1) Basic Methodology: historical VaR

The first family of models included in the risk management system is the historical method. This technique is the simplest of the five models adopted in this paper: it models the VaR threshold for a certain day as the quantile of the returns of the n immediately preceding market days (in this case n = 260). This is a backward-looking method which assumes that the distributional characteristics of past returns are a good proxy for analysing future returns. It relies on the assumption that the past situation of the market reflects the future one. Such an assumption is impossible in the case of portfolios of new financial products for which we have no previous past prices realization [3]; in our case, the main issues are related to the non-stationarity of the historical time series considered. However, considering its popularity and its simplicity, we decided to combine this approach with more complex techniques whose calibration could however create potential model risks.

In order to further diversify the approach adopted, a historical Var method based on "filtered historical simulations" (FHS) has also been employed. The FHS method “filters” the various returns, rescaling them using short-term volatility [13]. In this paper, the short-term volatility is calculated taking into account 100 market days, while the long-term volatility is based on a sample of 260 market days, in line with the same time windows used in the other methods.

In this way, the distribution of returns also incorporates information on the most recent volatility. In more formal terms, by assuming to have a distribution of unfiltered returns x(t), it is possible to derive the distribution of filtered returns X(t) as:

$$X(t) = x(t) * \frac{\sigma(t)_{SHORT\ TERM}}{\sigma(t)_{LONG\ TERM}} \quad (1)$$

Thereby, if the ratio of conditional short-term volatility to long term volatility increases, the corresponding return increases in absolute value, and if it is negative it assumes more weight in the calculation of the VAR; the opposite holds if volatility decreases. This allows to consider the effects of changes in volatility on VaR: in periods of financial crisis when the market returns decline and volatility increases (and therefore $\sigma_{SHORT\ TERM} > \sigma_{LONG\ TERM}$), negative returns are rescaled and increase in absolute value, making VaR thresholds theoretically more robust.

3.2) Monte Carlo method with and without GARCH volatility

The second class of models that has been implemented is the Monte Carlo method, of which two possible variants are proposed. The Monte Carlo method in this context is interpreted as a numerical method that allows to simulate the possible trajectories of one or more assets that follow a Brownian geometric motion. A Brownian geometric motion is meant as a stochastic process defined by the SDE:

$dS_t = \mu S_t dt + \sigma S_t \mathrm{d}W_t$ (2)

Where μ is the mean of asset returns, σ>0 is standard deviation, S(t) is the price of the asset at time t and W(t) is a Wiener process, that is a stochastic process defined by independent increments over time with mean equal to 0 and variance equal to the time interval considered: W(T)-W(0) is normally distributed with mean 0 and variance T.

It can be shown that equation (2) can be rewritten in its final form as:

$S_T = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\varepsilon_T\sqrt{T}\right]$ (3)

Where ε_T is distributed as a standardized normal with zero mean and unit variance. By means of equation (3), using the mean and variance of returns, the price at the beginning of the period and the time interval of the simulation as input parameters, it is possible to simulate multiple paths of asset price evolution.

Each individual path produces a simulated value S_T which represents a possible final price value. The final price S_T of the single simulation is used to estimate the various possible returns; these returns are combined into a single sample from which the 5% quantile (VaR) and the Expected Shortfall are calculated.

However, the setting of this work also requires the evaluation of the effects of correlations between assets. Consequently, two different variants of the Monte Carlo method have been implemented which generalize the "basic" method described, including the correlations in the calculation of the error term. In order to incorporate the correlation matrix for the four assets in the Monte Carlo simulation, the Cholesky decomposition can be used.

Assuming you have a set of unrelated random numbers $\vec{\varepsilon} = \varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots\ldots \varepsilon_T$, Cholesky decomposition can be used to transform them into a set of correlated variables $\vec{a} = a_1, a_2, a_3, \ldots\ldots a_T$. If $\vec{a}$ and $\vec{\varepsilon}$ are column vectors with N rows, and **R** is the correlation matrix, it is possible to apply the following transformation:

$\vec{a} = \boldsymbol{M}\vec{\varepsilon}$ (4)

Where $\boldsymbol{M}$ is a matrix that must satisfy the condition $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{R}$. The matrix $\boldsymbol{M}$ can be obtained by applying the Cholesky decomposition to $\boldsymbol{R}$. Subsequently, the correlated shocks ($\vec{a}$) are substituted to the errors (ε) in equation (3). From this point, the various possible paths of the assets are simulated, thus obtaining a set of possible values of the returns from which to calculate VaR and ES with the quantile method.

With regards to the implementation of the model in the MATLAB environment, the Cholesky function has been used to transform the correlation matrix $\boldsymbol{R}$ into an upper triangular matrix $\boldsymbol{M}$ that would guarantee the respect of the condition $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{R}$. Subsequently, the Hadamard product has been used to multiply, for each simulation, the innovations and the $\boldsymbol{M}$ matrix.

The number of simulations adopted for each VaR and ES value has been set equal to 50000. Once the value of the single innovation has been obtained, the simulated value of the returns of each asset in T + 1 has been calculated for each simulation, applying equation (3).

In each simulation, the mean of the prospective returns of the individual assets has been calculated in order to jointly obtain the return values of the assets and the portfolio. From these replicated simulations, VaR and ES have been calculated.

In order to implement the stochastic differential equations for the Monte Carlo engine, we did not use 'built in' functions already coded in the MATLAB toolboxes. Such a choice guarantees maximum flexibility in the design phase and allows to customize, for example, the dynamics that describes the stochastic process.

For this reason, the code has been validated through the pricing of three different options: a call option written on one asset, a European arithmetic average spread option written on two assets and an exotic option written on three assets (option on Maximum of two spread options) [7]. In all cases the results have been compared with a different numerical method, obtaining aligned values up to the basis points.

However, the presented Monte Carlo method could have some flaws, as it does not consider the non-stationarity of the time series. In particular, sudden increases in variance are observed in periods of crisis, a clear sign of the possible presence of heteroskedasticity.

In order to test the presence of this phenomenon, an Engle test with a confidence interval of 95% has been conducted. For each analyzed return series, the test led to the rejection of the null hypothesis of heteroskedasticity. Table 3.1 reports the P value of the two tests performed in this section.

Another very common feature in historical financial time-series is the autocorrelation of returns. With regards to this aspect, the autocorrelation of the residuals has been tested using a Ljung-Box test at 95% confidence interval applied to the first lag of the data.

The results of the test lead to reject the null hypothesis of independence of the residuals for the US equity index and the European equity index, the two more volatile components of our portfolio as outlined in section 2 (Tab 2.1). P values of this test are shown in Table 3.1.

Table 3.1 P-values for the statistical tests

P value of Engle Test and Ljung Box test for the analyzed time series					
		US equity Index	European Equity index	Bond index	Gold index
P-values of statistical test	Engle Test	0.001	0.002	0.008	0.007
	L-B, 1st lag	0.026	0.015	0.2	0.05

Starting from these results we have decided to estimate a GARCH model (Generalized AutoRegressive Conditional Heteroskedasticity) [6]. In a GARCH process, the conditional variance depends on the long-term unconditional volatility, p most recent values of the variance and the square of the last q past returns, according to equation (5):

$$\sigma_t^2 = V_L\gamma + \sum_{i=1}^{p} \alpha_i * u_{t-i}^2 + \sum_{i=1}^{q} \beta_i * \sigma_{t-i}^2 \quad (5)$$

Where V_L is the unconditional (or long term) volatility, u_{t-1}^2 is the squared log-return observed in t-1, and σ_{t-1}^2 is the conditional volatility observed in t-1. γ_i, α_i and β_i are the three weights whose sum is equal to 1. In the case of a GARCH (1,1) and assuming ω=$V_L\gamma$, the equation (5) can be rewritten as (6):

$$\sigma_t^2 = \omega + \alpha_i * u_{t-1}^2 + \beta_i * \sigma_{t-1}^2 \quad (6)$$

By applying a Maximum Likelihood (ML) approach, it is possible to estimate the three parameters ω, α and β, obtaining then γ, where γ =1-α-β. Writing the estimated variance in T as $v_t = \sigma_t^2$ and assuming that the probability distribution of u conditional to the variance u_t^2 is normal, the ML equation that has to be maximized becomes:

$$L = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v_i}} \exp\left(\frac{u_1^2}{-2v_i}\right) \quad (7)$$

By applying the natural logarithm, eq. (7) can be rewritten as:

$$L = \sum_{i=1}^{m} \left[-\ln\left(v_i - \frac{u_1^2}{-v_i}\right)\right] = \sum_{i=1}^{m} L_i \quad (8)$$

The next step has a computational nature: by using a traditional numerical optimization approach such as the Nelder-Mead simplex, it is possible to obtain the value of the weights that maximize L. Once these weights have been estimated, Equation (6) has been substituted in Equation (3); in other terms, the GARCH volatility structure has been plugged in the Monte Carlo simulation that describes the dynamics of correlated assets with the Cholesky decomposition.

3.3) Econometric methodology: Bayesian VaR.

In order to further diversify the approach used to describe the dynamics of correlated assets, we decided to use an additional approach related to Econometric methods: a BVAR - Bayesian Vector Autoregressive model. This model allows to include in the stochastic component the uncertainty that results not only from the shocks, but also from the variation of the correlation between the indices over time. Taking as a reference an econometric model written in the form:

$$y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \Sigma) \quad (9)$$

The main parameters are the vector of the coefficients β and the variance-covariance matrix of the errors Σ, which in this case is distributed according to a multivariate Normal. The principle of Bayesian analysis consists in putting together the information that is available in advance on the distribution of these parameters (the so-called prior distribution) with the information that we can obtain from the data (i.e., the likelihood function). In this way it is possible to obtain a new probability function that takes into account both factors, the so-called posterior distribution. The essential step for putting together the prior distribution and the likelihood function is the Bayesian rule. For a vector of parameters θ and a dataset y, given the density function f (y | θ), the Bayesian rule can be expressed as:

$$\pi(\theta|y) = \frac{f(y|\theta)}{f(y)} \pi(\theta) \quad (10)$$

The formula states that, given y, the probability that the "true value" of the parameter vector is θ is equal to the likelihood function of the data multiplied by the a priori distribution of the vector of parameters π (θ) and divided by the density of the data f (y). The vector of the parameters mentioned above is made up of two different elements: the vector of the coefficients β and the variance-covariance matrix of the errors Σ. For each of these elements, it is necessary to specify a prior probability distribution that allows - together with the likelihood function - to implement the “Bayesian rule”. One of the most widely used prior distribution is the "Minnesota prior".

The Minnesota prior assumes that the variance-covariance matrix Σ is already known. Therefore, only the vector of the coefficients β remains to be estimated: for this purpose, it is necessary to identify the likelihood function of β, f (y| β), and a prior distribution π(β). The starting point is the likelihood function: equation (9) implies that y is distributed as a normal multivariate distribution with mean Xβ and variance-covariance matrix Σ. Various techniques can be employed in order to estimate the matrix Σ. With enough computational power it is possible to relax the hypothesis of the diagonality of the Σ matrix described by [5], which is then derived from the variance-covariance matrix of a similar VAR model estimated via Ordinary Least Squares – OLS. Consequently, the maximum likelihood function can be written as:

$$f(y|\beta,\Sigma) = (2\pi)^{-nT/2}|\Sigma|^{-\frac{1}{2}} \exp[-1/2(y - \bar{X}\beta)'\bar{\Sigma}^{-1}(y - \bar{X}\beta)] \qquad (11)$$

The notation can be simplified by collecting the terms that do not depend on β:

$$f(y|\beta,\Sigma) = \alpha \exp[-1/2(y - \bar{X}\beta)'\bar{\Sigma}^{-1}(y - \bar{X}\beta)] \qquad (12)$$

Looking at the distribution of β, it is supposed to follow a normal multivariate distribution with mean β_0 and variance Ω_0. In the original formulation [10] the expected value for each parameter (which contributes to the specification of the β_0 vector) is equal to 1 for the first lag and equal to 0 for the following lags since most of the time series are characterized by the presence of a unit root. The variance-covariance matrix Ω_0 is a diagonal matrix whose terms are defined by a set of parameters usually derived from the econometric theory. For a discussion of the estimation of these parameters, see the reference literature [5]

The chosen approach uses a slightly more complex variant of prior distribution with respect to the Minnesota prior: the normal-inverse-Wishart prior [9]. The main difference is related to the variance-covariance matrix of the errors Σ, which is also an unknown and is no longer known in advance (in line with the approach adopted in the Monte Carlo GARCH which also stochastically considers non-homoskedastic components). It is assumed that Σ follows an inverse Wishart distribution which has as input parameters the matrix Ω and the number of degrees of freedom v. In mathematical notation: $\Sigma \sim W^{-1}$ (Ω, ν), where the matrix Ω is equal to the amount (v – number of parameters – 1) multiplied by the diagonal matrix that contains in the main diagonal the variance of the errors of each single variable calculated with an equivalent number of AR models.

Furthermore, in this setting, the variance-covariance matrix of β is not symmetric: the hypothesis of independence between the variables is relaxed; the new variance-covariance matrix of β is obtained by multiplying Σ by φ_0, a matrix with each dimension equal to the number of variables. The new distribution of the coefficients is therefore:

$$\pi(\beta) \sim N(\beta_0, \Sigma \otimes \varphi_0) \ (13)$$

Where the elements of φ_0 depend on a series of hyperparameters set a priori according to the model. In this case, a notable aspect of this approach is the relationship between the prior distribution of the parameters β and the matrix Σ.

A rather common problem in the calculation of the correlation coefficient between assets is the effect of variance: in conditions of relatively low volatility on the market, an apparent increase in correlation between the assets can be noticed, as the declining volatility decreases the denominator of the correlation coefficient. Conversely, if volatility increases, the denominator increases and it is possible to observe an apparent decline in correlation between assets. The term "apparent" is used here because, in fact, there is not a real variation in the risk linked to the correlation between assets: simply, in proportion, the co-movements between the assets are larger or smaller with respect to the variability of each asset.

However, depending on what the stochastic component of the model is, the effects on VaR estimation can be quite significant. In the Monte Carlo VaR, discussed in 3.2, this risk is reduced by the fact that the correlated shocks of an asset are multiplied by the corresponding standard deviation of each asset, thus offsetting any "apparent" effects due to changes in volatility. However, if one wants to analyse a BVAR model where the only varying parameter is the cross correlation between two or more assets, then the difference between the "real" and the "apparent" variations would become much more marked. This can be a very important theoretical problem while trying to generalize the models.

By way of example, looking at the S&P500 index data from 1992 to 2010, it is possible to observe a series of recurring patterns in the correlation matrix between different sectors and, starting from these patterns, to identify 8 "states" of the market characterized by particular configurations of this correlation matrix [12]. More importantly, the authors of this study show that the market only moves along continuous states. This is an interesting starting point for any subsequent calibration of the a posteriori distribution of the coefficients of interest. In the case of the Bayesian VaR, which adopts the normal-inverse-Wishart prior, this problem is implicitly solved by creating a proportional relationship between the variance - covariance matrix of the coefficients and the Σ matrix.

More generally, the calculation of a VaR with the Bayesian approach starts from the a posteriori distribution: once a distribution for the parameters has been defined, n possible values of each parameter are calculated, thus obtaining n possible simulations of the returns of each asset. Starting from these values, with the techniques already described for the other methods based on quantiles, VaR and Expected Shortfall are computed.

In order to implement the BVAR model, the first step has been a model selection analysis in order to identify the most suitable number of lags for the model. First, a maximum number of 10 lags has been set for the model. Subsequently, a specific VAR model has been estimated in a structured form, for each of these lags.

Once a VAR model has been estimated for each lag, the value of two information criteria has been calculated for each model: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).
The best model has been identified for each of the two information criteria: if both suggested a model with a certain number of lags, that model has been adopted; in case of disagreements, the model with the best BIC has been chosen: it is the criterion that attributes the greatest penalty (log (n)) for the number of parameters, thus making the modelling more prudent. Note that this caution does not only derive from a modelling issue: a big problem encountered in building the Bayesian model is the computational time.

Models that consider more than one lag make it problematic to conduct a sufficiently large number of simulations. In about 99% of the more than 5000 iterations replicated for the entire dataset, this strategy chooses a VAR (1) as the best model for all four assets in the portfolio, which has been consequently adopted for the entire dataset. This criterion indicates rather low orders of representation, orienting us towards parsimonious models in terms of the number of parameters. Obtaining a simulation of the returns for each asset is a computationally complex challenge for the implementation of the Bayesian method. Estimating a Bayesian VAR for each iteration and conducting 10,000 simulations may require many hours, if the code is not parallelized, depending on the processors used.

4) The Commitment Machine

4.1) Design phase

As already highlighted, the need to have a precise VaR and ES estimation for T + 1 in T makes it necessary to selectively use the models described in section 3. To this end, it has been decided to opt for an algorithmic solution, that we refer to as Commitment Machine (CM), which, given a set of VaR methods considered statistically reliable, allows to select on each day T a method that has been more performing in the previous period and to use it in order to calculate VaR and ES thresholds for the T+1 observation.

The time interval in which the CM evaluates the performance of the models is defined as the 'observation window': so, for example, if we use data ranging from T-49 to T, the observation window is equal to 50.

Given these specifications, the CM algorithm has been defined starting from 3 elements:

1. A set of methods for calculating the risk measures used (VaR and ES), whose adequacy has been tested by the CM, together with the portfolio return data to be used for the backtesting and for the calculation of the loss function.
2. A loss function to be minimized that allows the meta-heuristic to select the different calculation methods of VaR and ES.
3. An observation window of n observations ranging from T- (n-1) to T which is used for estimating the loss function.

The first step to proceed with the construction of the CM is the selection of the components, i.e., the choice of which methods of VaR and ES calculation are considered valid for the purpose of inclusion in the CM. To this end, it has been considered appropriate to evaluate the single models on the basis of their overall forecast performance. These models have been calibrated to provide a minimum return threshold corresponding to a 95% confidence interval. This implies that we must expect to have lower returns than the corresponding VaR thresholds in 5% of the observations.

Consequently, the frequency of VaR overruns has been selected to be the performance measure. This frequency is expected to be slightly higher than 5%: it is likely that in the event of a market collapse, the implemented VaR methods will need more than one day to recalibrate the estimated VaR and ES values. However, this percentage must not be much higher than 5%: otherwise, the system of risk management methods could not be considered robust and reliable. Furthermore, a frequency of violation greater than the theoretical value (in this case, 5%) by a wide margin could lead to an inspective intervention in order to understand the low effectiveness of the risk measure adopted.

Table 4.1 displays the percentage of days in which portfolio returns are lower than the daily VaR threshold calculated by the model, for each method, during the whole time span of our dataset

Table 4.1 VaR violation frequency for the five daily VaR estimation models

VaR Threshold violation frequency				
Historic	FHS	Monte Carlo	Monte Carlo GARCH	BVAR
5.44%	5.46%	5.37%	5.31%	5.21%

The results show that all five methods have a violation frequency greater than 5%, but apparently not much higher. Consequently, we decided to select all of the five models as potential candidates for the CM VaR selection. The correctness of this selection method has also been verified in a more rigorous way using two different statistical tests:

1) Binomial Test: The binomial test compares the observed number of violations, x, with the expected number. The observed number of violations is assumed to follow a binomial distribution. By using the properties of the binomial distribution, it is possible to construct a test statistic p expressed as:

$$Z_{bin} = \frac{x-Np}{\sqrt{Np(1-p)}} \quad (14)$$

N is the number of observations, p = (1 - VaR level) is the probability of observing a violation if the model is correct, and x is the number of violations observed. This value is compared with the expected value $x - Np$, that is supposed to follow a normal distribution with zero mean. The null hypothesis states that it is possible to observe $x - Np$ violations, that is, Z_{bin} lies in the confidence interval of the theoretical normal distribution. In the case of the five VaR models mentioned above, the binomial test leads to accept the null hypothesis in all cases.

2) "Traffic Light test": the TL test, proposed by the Basel committee [2], measures the probability of observing a number of violations equal to or less than x (i.e., the cumulative probability from 0 to x). The probability of observing a certain number

of violations is supposed to follow a binomial distribution similar to the one seen for the binomial test. The test is called "traffic light test" because, depending on the realized performance (the number of overruns), the test classifies a VaR estimation model in one of the following three zones, denoted by a different colour:

- Red zone: the probability of observing a number of violations equal to or smaller than the number actually observed is 99.99%. Results are in the tail of the distribution: it is very unlikely that a VaR model with x violations on N observations is correct. The model is therefore discarded.
- Yellow zone: the probability is between 95% and 99%. The number of violations is very high but not excessively so the model is acceptable, but on the condition that the calculated VaR thresholds are strengthened.
- Green area: The probability is 95% or less. The model is accepted.

All selected VaR models fall within the green zone.

The second step needed in order to design the CM is the choice of the loss function to be minimized: it is necessary to specify a performance objective that can be used to evaluate different VaR estimation methods, taking into account the goal of obtaining a method that produces a percentage of overruns as close to 5% as possible (i.e., a theoretically solid model, able to reliably predict the actual value of the estimated quantile).

A first possible performance measure could be, again, the violation frequency. However, this is a measure with low sensitivity for limited time intervals: in certain periods, it is likely that the percentage of violations of the five methods is the same depending on the market phase considered. A second problem with the violation frequency is that it is a partial measure: the magnitude of violations is not factored in. There might be many small violations or few very large ones, where the return even falls below the Expected Shortfall thresholds.

The alternative approach is to focus on losses: in such case, the CM evaluates the VaR models based on the cumulated losses in the observation window. The use of a loss function to compare the performance of different VaR estimation methods has already been discussed in literature, starting with [11] that suggest three different possible loss functions (LF) that reflect specific concerns of risk management and tend to grow in the case of VaR estimation model failure. From this work various analyses of possible LF have been derived and the most common is the "regulatory loss function" proposed by [17], which incorporates one of the LFs proposed by [11].

The regulatory loss function has a functional form similar to the one described in equation (16) that is adopted in this work, but it weights the losses according to a quadratic function; it is supported by another LF which also takes into account the opportunity costs related to the capital absorbed by VaR in periods in which the VaR threshold is not exceeded. Compared to these LFs proposed by [17] and [11], the two LFs tested in this work (equations (15) and (16)) are easier and allow an immediate comparison between the different VaR techniques.

In order to adopt an approach that is as flexible as possible, we decided to test 2 different loss functions, one characterized by the objective of minimizing violations and one with the objective of minimizing the overall losses.

Defining R_T ad V_T respectively as the returns at date T and the corresponding VaR threshold, the loss function for a certain VaR model takes the functional form described in equation (15) if we want to minimize violations, while equation (16) is more suitable in case we want to focus on losses minimization.

$$f_1(x) = \begin{cases} 1, & R_T < V_T \\ 0, & R_T \geq V_T \end{cases} \qquad (15)$$

$$f_2(x) = \begin{cases} \mathrm{Vt} - \mathrm{Rt}, & R_T < V_T \\ 0, & R_T \geq V_T \end{cases} \qquad (16)$$

Minimizing the objective function, therefore, is equivalent to iteratively choosing the method of VaR estimation which leads to lower losses or overruns. It is assumed that a "weak" model in the previous period maintains this weakness also in T + 1: since the five models used are structurally very different, it is expected that they can capture the market trend with a different degree of precision, depending on the market conditions in the period considered.

The LF is not the only possible way to take into consideration temporal variations in the ability of single models to predict the actual value of the VaR: in literature, for example, methods of estimating the VaR are based on a weighted average of the VaR thresholds estimated by several models [4].

This approach has been discarded because it is more difficult to calibrate and returns less informative results than simply selecting the best model (meaning that it is more difficult to interpret, also in terms of performances).

Although the performance measure takes into account the VaR threshold, the CM is designed to calculate both VaR and ES: when a certain method is selected as "the best performing", not only the VaR value in T+1, but also the corresponding expected shortfall value on the same date is calculated by using that method.

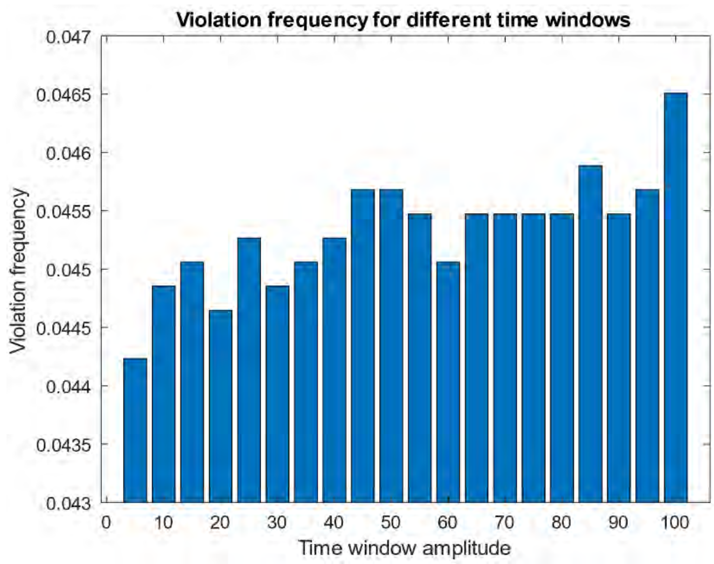
4.2) Calibration phase

The last step for the CM design is the determination of the time interval used for its calibration. This process highlights an important trade-off. While, as the observation window expands, there is a more consistent estimate of the relative effectiveness of the various models, on the other hand a further argument in favour of a short observation window is the rapidity of reaction; among the main issues for risk management are sudden moments of market collapse. Such events occur very quickly, making it advisable, under certain circumstances, to have a CM that adapts as quickly as possible.

In order to balance this trade-off, we decided to test the algorithm on 20 possible observation windows of various length between 5 and 100 days (5, 10, 15 .... 100 days). To evaluate the performance of the CM ex-post, we decided to calculate, for each simulation, both the frequency of VaR violations (given the regulatory and management importance of this indicator) and the overall average of the losses occurred (i.e., the sum of the VaR overruns divided by the number of observations) for the VaR thresholds calculated by each loss function. The results of this calibration show that the loss function based on losses below the VaR threshold performs better than the one based on violations. More specifically, by taking the mean of the results of all of the possible simulations (20, one for each simulated observation window) we find that the CM adopting a LF based on the overruns has a frequency of violations of 4.82%, while the CM based on losses has a frequency equal to 4.53%. For all observation windows, as can be easily imagined, the LF that tries to minimize losses below the VaR threshold has minor losses. The superiority of a loss-based function with respect to both performance measures considered (losses and overruns) is supported by the data. Therefore, we decided to adopt a LF that aims to minimize losses.

Figure 4.1 shows the various violation frequencies achieved by the CM for observation windows ranging from 5 to 100 days. On the horizontal axis we show the different amplitude values of the observation windows for which the percentage of violations has been calculated (vertical axis).

Figure 4.1 Violation frequency of the CM



The first important consideration that can be deduced from the figure is that the logic behind the CM appears to be quite reasonable; indeed, for any interval tested, the frequency of violations is less than 5%, and it is between 4.42%, measured with a 5-day interval, and 4.65%, measured with a 100-day interval. The CM is effective in reducing overruns, although this effectiveness decreases as the observation window is extended. The high sensitivity of the LF leads to very prudent results if applied over very short time intervals; however, the proposed goal is to have a robust method for estimating a 95% confidence interval and a violation frequency far below 5% means having an excessively cautious model compared to the initial specifications. Considering Figure 4.1, a violation frequency near to 4.5-4.6% would represent a reasonable result, given that it is better to have an excess of prudence than an excess of risk: in this sense, having a percentage of overruns of 4.6%, for example, is much better than having a percentage of 5.4%, although in both cases the absolute distance from the target is the same. To this end, two different time windows have been chosen to calibrate the CM: one at 50 days and one at 85 days. The reason for this double choice is that the violation frequency is relatively stable between 45 and 95 days and choosing values within this range seems to ensure that we adopt a reliable model at the same time.

Table 4.2 shows the main characteristics of the two different CMs, differentiated according to the number of observations considered each time (50 for the first variant, 85 for the second).

Table 4.2: Summary of the two selected CM variants

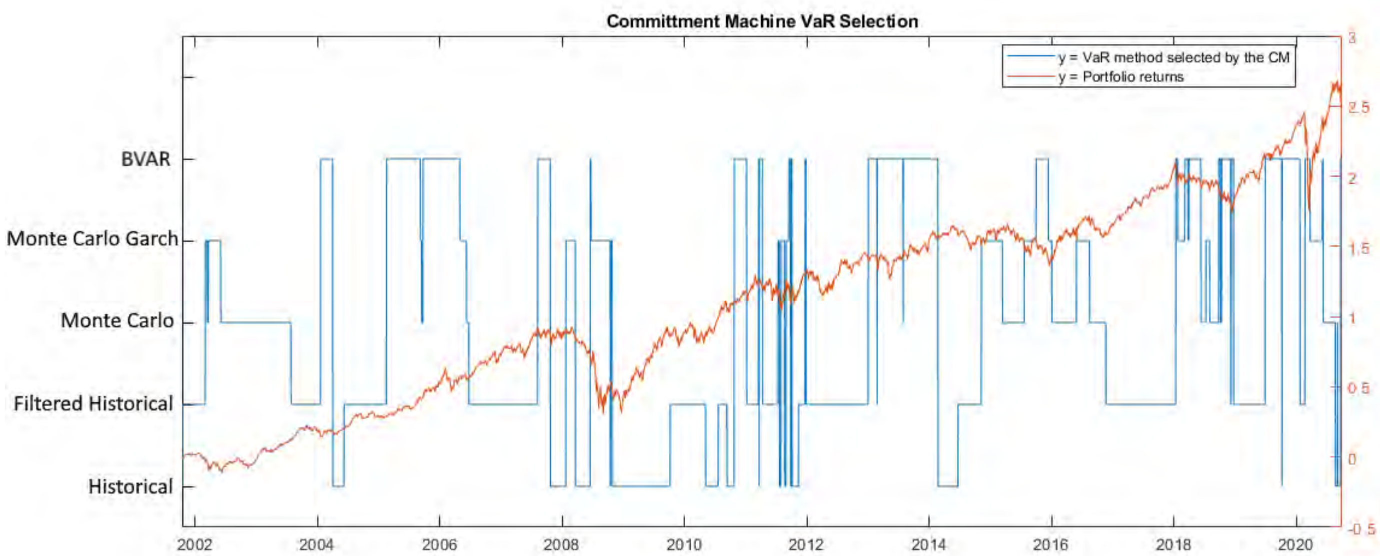
	Violation Frequency	**Frequency of VaR selection by CM**				
		Historical	Filtered Historical	Monte Carlo	Monte Carlo Garch	Bayesian
CM (50)	4,92%	13,4%	37,3%	14,3%	11,7%	23,3%
CM (85)	4,82%	13,0%	37,4%	12,5%	14,1%	23,0%

The two CM variants CM(50) and CM(85) are very similar in terms of how frequently the various VaR methods are chosen. The only difference is that the version based on the shorter observation window chooses more frequently the Monte Carlo method without GARCH volatility and less frequently the Monte Carlo method which includes the GARCH. It is important to notice that using two different CMs makes it possible to verify that the heuristic underlying the CM is efficient. In section 5, the results (in terms of VaR selection) of both of these CMs were evaluated. From the point of view of the robustness of the methodology, using multiple variants can confirm that the design flexibility of the CM does not translate into an excessive variability of the results. While in fact the adoption of a different observation window is optimal for the needs of different subjects and increases the flexibility of the algorithm during implementation, it is clear that the CM (with the same selected methods) should provide a consistent approach. Indeed, choosing a window observation of 50 or 85 days should not lead to too different selections, otherwise the algorithm of the CM would be excessively unstable and the possibility to choose between multiple observation windows would add a further layer of complexity to the model risk, instead of making the implementation more flexible.

5) Results

This section focuses on the analysis of the results of the two different CM variants discussed above. In particular, the analysis is divided into two steps: a disaggregated analysis of the performances of each CM on the days in which one of the five methods has been selected (historical VaR, filtered historical VaR, Monte Carlo, Monte Carlo GARCH, Bayesian) and a general analysis of the performances of the two different CMs over the entire dataset. In order to distinguish between the different CMs, the notation proposed in section 4 is maintained: CM(50) is the CM that uses an observation window - in accordance with the adopted heuristics - of 50 trading days, while CM(85) is the CM that uses a window of 85 days.
Starting from the analysis of the methods selected by the various CMs, we first analyze the CM(50), i.e., the CM calibrated over a time interval of 50 days. This variant of CM is particularly interesting because it shows how the optimization algorithm performs on relatively short time intervals, where the information set available is more volatile. It should be noticed that in 83.4% of the periods considered, the CM(50) and the CM(85) choose the same VaR method. While this data highlights the consistency of the algorithm adopted, on the other hand it allows us to take the VaR selection of one of the two as representative of both CMs, thus making this analysis much easier.
Figure 5.1 shows, for the CM(50), the type of VaR model selected by the algorithm (labelled on the y axis) and the performance of the portfolio (measured by cumulative returns) from the beginning of the dataset to the end.
By observing Figure 5.1, two significant aspects of the CM can be noticed:
1) Stability: for many long periods, the CM always chooses the same method for calculating the VaR. This is a very important factor, as it highlights how CM's choices are consistent over time. There is no random selection, but a logical choice. This stability is less evident when the Bayesian method is selected since Bayesian estimators tend to produce more volatile values [10].
2) In the majority of the phases characterized by positive portfolio returns, the CM chooses the historical methods, particularly in 2007, in the period 2009-end 2011, in 2014 and in 2017. The fact that in these periods mean returns are positive makes it more difficult for the Monte Carlo VaR to generate a prudential forecast, compared to the historical VaR. Conversely, the VaR estimated with the historical method remains very conservative, even in a context characterized by positive returns [15]. In times of market downturn, however, the CM switches to the Monte Carlo VaR, with or without GARCH (particularly in the subprime mortgage crisis and in the fall of 2016) and, during the past two years of declining trend, it also adopts the Bayesian VaR, whose estimates tend to be more conservative than the other methods employed [14]. In the case of the 2008 crisis, for example, the historical VaR is initially chosen and after some observations, as the market trend reverses, the CM switches to the Monte Carlo VaR, which is more prudent also due to the trend of cross-correlations.

Figure 5.1 CM model selection



A significant element for evaluating the performances of the two different variants of CM is their ability to select, for each day, a model whose ex-post performance is better than the performances of the other models. Before making this comparison, it is advisable to have a general overview of the performance of the single VaR methods over the days they have been selected, by using two metrics: the violation frequency and the average loss.
The average loss is simply the average, for each method, of the values assumed by the loss function introduced in section 4 (equation (16)) on the days in which that method is selected; the average loss of the historical VaR, for example, represents the total sum of the negative differences between returns and Historical VaR in the days in which the CM selects the historical VaR, divided by the total number of days in which the historical VaR is selected. Tables 5.1 and 5.2 show the values of these two indicators for each VaR method selected by each of the two CMs.

Table 5.1 Violation frequency of different VaR methods in the day in which they are selected by the CMs

	Average violation % when selected				
	Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	2,91%	3,02%	2,89%	9,75%	6,48%
CM(85)	2,53%	3,61%	2,82%	8,36%	6,01%

Table 5.2 Average VaR loss for different methods in the day in which they are selected by the CMs

	Average VaR loss when selected				
	Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	0,014%	0,015%	0,013%	0,074%	0,039%
CM(85)	0,010%	0,016%	0,011%	0,077%	0,039%

Tables 5.1 and 5.2 show that, on average, both the CMs choose the Bayesian method and the Monte Carlo GARCH method in the days in which VaR violations are more frequent (Tab 5.1) and there are greater VaR losses (Tab 5.2). This is consistent with the fact that these two methods are selected during periods of market crisis, as shown in Figure 5.1.
To better contextualize these results, it is important to evaluate the performances of other VaR methods observed in the same days. To this end, we proceed to further break down the analysis and we consider, for each CM, the days in which each VaR method is selected. The dataset has been divided into 5 sets, each containing the days in which one of the 5 methods is selected. Then, the average violation and VaR losses have been calculated for each method in each of the five sets, and these data have been compared with the performances of the other methods on the days of the same set.
Table 5.3 shows the violation percentage frequency. Here is an explanation of how to read Table 5.3: the first of the five rows of the CM(50) shows, for each VaR estimation method, the percentages of overrun of the VaR threshold calculated on the days in which the method selected is the unfiltered "base" historic VaR.
It is expected that, on the days in which the CM chooses the historical method, this percentage will be lower for the historical VaR, i.e., the CM is able to choose, in T+1, the method for which it has the smallest chance of seeing one exceeding the VaR threshold.
In order to simplify the analysis, the performances of each method on the days in which it is selected have been highlighted in bold. For each row, these performances are expected to be the best. The analysis of the results shows a significant selection ability of the CM; with the exception of the historical unfiltered VaR, in all considered cases, the CM is able to manage the choice of the method that guarantees the lowest violation frequency. The historical VaR is an exception, probably because in 2008 it has been chosen by the CM in the very first days of the outbreak of the financial crisis (see Figure 5.1).

Table 5.3 Analysis of violation frequency of VaR methods

		Average violation % when selected				
		Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	Historic	**2,91%**	2,75%	3,21%	3,52%	2,91%
	F. Historic	3,85%	**3,02%**	4,40%	4,34%	3,90%
	Montecarlo	4,04%	3,75%	**2,886%**	3,319%	4,473%
	Montecarlo Garch	10,99%	12,77%	10,284%	**9,752%**	10,461%
	BVAR	7,55%	8,35%	7,282%	6,927%	**6,483%**
CM(85)	Historic	**2,53%**	2,37%	3,165%	3,481%	2,532%
	F. Historic	4,44%	**3,61%**	4,874%	4,819%	4,436%
	Montecarlo	3,65%	3,48%	**2,819%**	3,317%	4,312%
	Montecarlo Garch	9,384%	11,144%	8,798%	**8,358%**	9,238%
	BVAR	7,271%	7,810%	6,732%	6,373%	**6,014%**

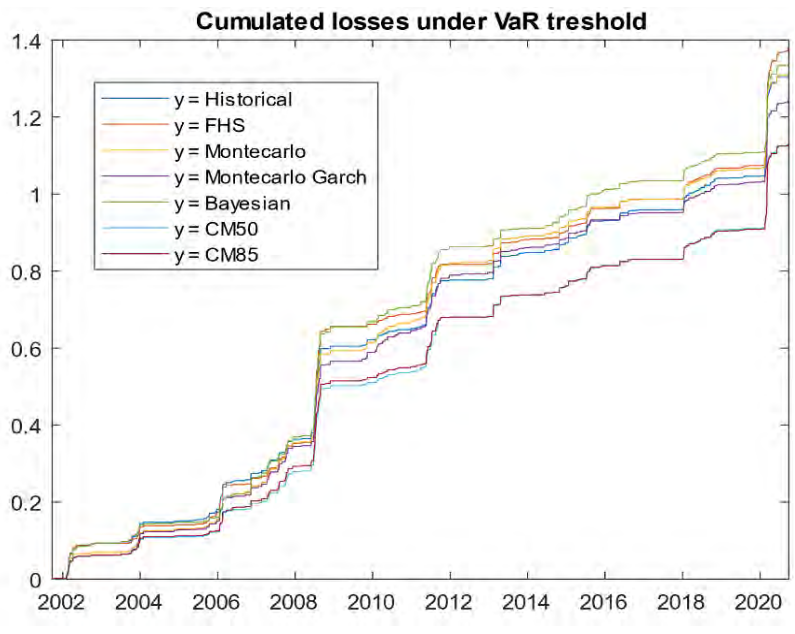
The Monte Carlo GARCH method is selected on the days in which, on average, the frequency of violations of the various methods is very high. As already indicated in the design phase, its ability to take into account both the correlation between returns and the non-stationarity issues makes it quite useful in times of crisis, when VaR violations are generally more frequent. The Bayesian VaR, when selected by the CM, is also the best and this represents a noteworthy result, considering the high sensibility of the Bayesian model to parameter estimations. However, it should be considered that the heuristic underlying the CM is different since the aim is to minimize not the frequency of violations, but the overall losses.
In this sense, a more convincing performance measure is the average of the losses realized in the days in which our two CMs select the various VaR methods. Resuming the objective function discussed in section 4, this measure is equal to the average value of such function on the analysed days.
In this way, it is possible to derive an ex-post value that indicates how the selected method really minimizes the value of the objective function compared to the other methods. Higher values indicate higher losses in the days considered and vice versa. Table 5.4 shows the values of the average losses, organized in a similar way to that seen in Table 5.3.

Table 5.4 Analysis of the average VaR loss methods

		Average VaR loss when selected				
		Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	Historic	**0,014%**	0,014%	0,017%	0,017%	0,016%
	F. Historic	0,015%	**0,013%**	0,018%	0,018%	0,017%
	Montecarlo	0,013%	0,012%	**0,011%**	0,011%	0,015%
	Montecarlo Garch	0,074%	0,090%	0,066%	**0,061%**	0,075%
	BVAR	0,039%	0,040%	0,038%	0,035%	**0,034%**
CM(85)	Historic	**0,0099%**	0,0098%	0,013%	0,013%	0,012%
	F. Historic	0,018%	**0,016%**	0,021%	0,020%	0,020%
	Montecarlo	0,012%	0,011%	**0,0094%**	0,0095%	0,014%
	Montecarlo Garch	0,063%	0,077%	0,056%	**0,052%**	0,064%
	BVAR	0,038%	0,039%	0,036%	0,034%	**0,032%**

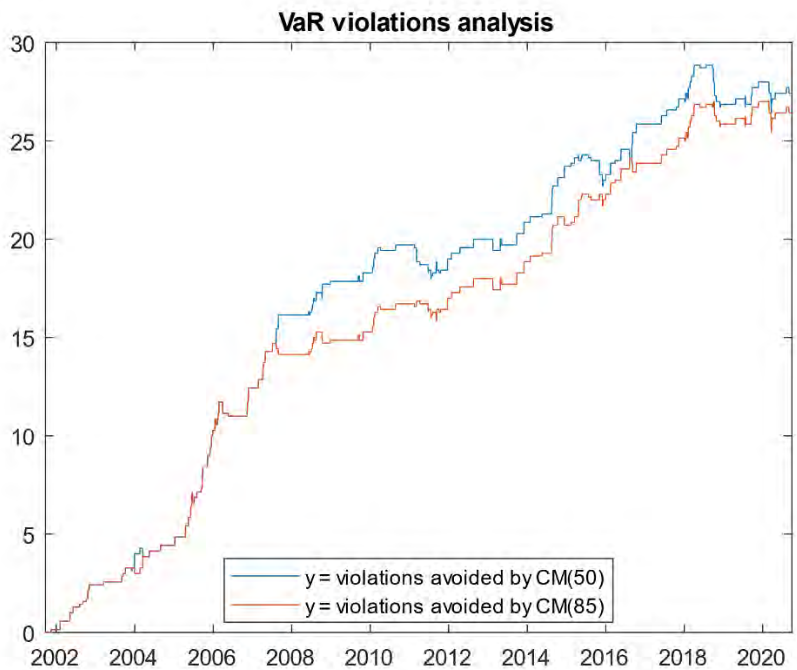
Results suggest that the CM effectively manages to accurately select the various methods. In each row, the minimum loss is associated with the method selected by the CM; for example, in the last row it can be seen that, on the days in which the CM(85) selects the Bayesian method, this method is the one with the lowest average losses (0.032% against 0.034-0.039% of the other methods). From Table 5.4 we further notice the difficulty of the two CMs in making the best use of the unfiltered historical VaR.
Starting from this data, it is appropriate to take a step back and aggregate the analysis once again in order to evaluate the overall performance of the five VaR methods (Historical, Filtered Historical, Monte Carlo, Monte Carlo GARCH, Bayesian) and of the two CMs over the entire time series.
Figure 5.2 shows, for the time interval considered, the cumulative losses with respect to the VaR threshold of the five methods considered by the CM and of the two variants of CM. The cumulative loss on a certain date t is equal to the sum of the historical values from 0 to t of the loss function introduced in section 4, equation (16) and it represents the total amount of losses below the VAR threshold up to that day.

Figure 5.2 Cumulative losses of VaR methods and CMs



The data show two periods in which the portfolio registers high losses, specifically the subprime mortgage crisis and the Covid shock in the spring of 2020. Between these two periods there is a further phase of losses – of smaller magnitude than the other two - corresponding to the crisis of the sovereign debts in the Eurozone. During these two negative market events, the cumulative losses increased considerably both for the five VaR methods used and for the CM built on these methods. The subprime mortgage crisis is a key moment for this analysis. Indeed, the two CMs suffer lower losses in such period, while the two methods with the worst performances are the Monte Carlo VaR (without GARCH) and the Bayesian Var.
This is a further endorsement of the validity of the heuristic approach adopted in the design phase, which over the course of the length of our time series allows the algorithm to avoid choosing, among the five methods, the ones that have the worst performance.
A further element of interest is the temporal trend of the cumulated losses: considering Figure 5.2, almost a third of the losses are realized in the two rare market events (the subprime mortgage crisis and the 2020 crisis). It is important to underline how, during the 2020 crisis, the VaR method (out of the three initially chosen) that accumulates the smallest losses is the Bayesian VaR. This data confirms the validity of a Bayesian approach in a scenario in which the market situation changes (more or less literally) overnight.
Another evidence of the goodness of our results can be seen by taking into account the amount of VaR violations that are avoided by the CMs. This measure can be estimated by the difference between the average number of VaR violations computed by the five methods and the VaR violation committed by the CMs. For example, a difference of 3 means that, on average, there are 3 days in which the VaR threshold calculated by the CM is not violated, while the other methods have VaR threshold violations. Figure 5.3 shows this difference: CM algorithm is successful at avoiding VaR violations and the smooth growth of the measure over time suggests a stability of the positive performances of the CM optimization algorithm.
It is worth to notice that the logic of CM approach allows to calculate the Expected shortfall (ES) for a given portfolio: once a VaR method is selected by the CM, the mean of returns under VaR threshold can be easily calculated.
This is a valuable feature considering that ES is a coherent risk measure and thus better suited as a risk measure for portfolio optimization [16].

Figure 5.3: Analysis of daily VaR threshold violations



6) Conclusions

In this paper we have designed an algorithm which performs an automatic choice among different VaR Methods, based on the minimization of a loss function that takes into account the negative returns below the VaR threshold.
The implemented Commitment Machine (CM) handles five different estimation techniques: traditional historical VaR, filtered historical simulations (FHS) VaR, Monte Carlo VaR with cross correlated assets, Monte Carlo GARCH VaR with cross correlated assets, and Bayesian VaR. These different models are able to take into account different econometric aspects of the time series, particularly non-stationarity and cross correlation.
The analysis of the CM performances, tested on the realized returns of an equally weighted portfolio made up of four market indices from 2001 to 2020, shows that the use of the algorithm provides interesting advantages compared to the implementation of a single VaR method.
Thanks to its adaptive logic selection, the CM records lower VaR violations and losses than the five methods individually implemented.
The flexibility of the code written in MATLAB environment guarantees the possibility of generalizing the analysis by including other VaR estimation methods, even different from the ones used in this work. It also offers further possibilities for alternative implementations, for example by modifying the loss function in order to consider the needs of the various entities involved in the risk assessment.

In conclusion, our applied analysis provides significant evidence in favor of the goodness of the design of the proposed CM, thus making it a useful tool for managing portfolio risk.

Marco Bagnato, Anna Bottasso and Pier Giuseppe Giribone

Bibliography

[1] Ang A., Chen J. (2002) “Asymmetric Correlations of Equity Portfolios”. Journal of Financial Economics, vol. 63, issue 3, pp. 443-494
[2] Basilee Committee on Banking Supervision (1996) “Supervisory Framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements”
[3] Bottasso A., Giribone P. G., Martorana M. (2019) “The emerging market of guarantees of origin: analysis and design of a quantitative measurements system for monitoring financial risks”, Risk Management Magazine Vol. 14 N. 2
[4] Casarin R., Chang C. L., Martin J. J., McAleer M., Amaral T. P. (2012) “Risk management of risk under the Basel Accord: A Bayesian approach to forecasting Value-at-Risk of VIX futures”, Mathematics and Computers in Simulation, Vol. 94
[5] Dieppe A., Legrand B., Roye R. (2018) “The Bayesian Estimation, Analysis and Regression (BEAR) toolbox”, Technical guide
[6] Engle R. F. (1982) “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”, Econometrica 50, pp. 987-1007.
[7] Haug E. G. (2007) “The complete guide to Option Pricing Formulas”, McGraw Hill, cap 8.1.4
[8] Hendricks D. (1996) “Evaluation of Value-at-Risk Models Using Historical Data”, Economic Policy Review, Vol. 2, No. 1
[9] Karlsson S. (2012) “Forecasting with Bayesian Vector Autoregressions”, Handbook of Economic Forecasting
Volume 2, Part B, 2013, pp. 791-897
[10] Litterman R. B. (1986) “Forecasting with Bayesian Vector Autoregressions – Five years of experience”, International Journal of Forecasting, Vol. 4, No. 1, pp. 25-38
[11] Lopez J. A. (1998) “Methods for evaluating Value-at-Risk estimates”, Federal Reserve Bank of New York, research paper No. 9082
[12] Münnix M. C., Shimada T., Schäfer R., Leyvraz F., Seligman T. H., Guhr T., Stanley H. E. (2012) “Identifying States of a Financial Market”, Scientific Reports, Vol. 2, No. 644
[13] Perez P., Murphy D. (2015) “Filtered historical simulation Value-at-Risk models and their competitors”, Bank of England Working Paper No. 525
[14] Pollard M. (2007) “Bayesian VaR and Capital Charge Puzzle” - papers.ssrn.com/sol3/papers.cfm?abstract_id=1116542
[15] Pritsker M. (2006) “The hidden danger of historical simulation”, Journal of Banking & Finance, Vol. 30, Issue 2, pp. 561-582
[16] Quaranta A. G. , Zaffaroni A. (2008) “Robust optimization of conditional value at risk and portfolio selection”, Journal of Banking & Finance, Vol. 32, Issue 10, pp. 2046–2056
[17] Sarma M., Thomas S., Shah A. (2002) Selection of Value-at-Risk models, journal of financial forecasting, Volume 22, Issue 4, pp. 277-358

Banks' Corporate Governance: lessons learnt from the Great Financial Crisis

Enrica Rimoldi (Indipendent analyst, Non-Executive Director, INSEAD Certificate in Corporate Governance)

Article submitted to double-blind peer review, received on 12th January 2021 and accepted on 21st April 2021

Abstract

After acknowledging that one of the key factors that contributed to the Great Financial Crisis were the failures by banks in their Corporate Governance, standard-setting bodies have reinforced banks governance standards in order to reduce the shortcomings observed during the crisis. In March 2020, the Bank for International Settlement issued the paper "Bank Boards – a review of post-crisis Regulatory approaches" that, taking stock of specific aspects of the post-crisis Regulatory approaches used in 19 jurisdictions to strengthen Board oversight at banks, reviews the "Fitness and Propriety" assessment that these jurisdictions use to ensure that bank Board members are suitably qualified.

Investigating the empirical evidence provided by scientific literature on the relationship among Corporate Governance and the profitability of the banks during the Great Financial Crisis, the results of this paper support some of the choices made by the Regulators to enhance the banks' Corporate Governance in order to mitigate similar risks that banks could face subsequently.

*****

Come noto, uno dei fattori che contribuì alla Grande Crisi Finanziaria è ascrivibile alle carenze nella Corporate Governance delle banche. Con tale consapevolezza, i Regolatori hanno successivamente rafforzato gli standard di Corporate Governance delle banche e, nel Marzo 2020, la Bank for International Settlement ha emesso il documento "Bank Boards – a review of post-crisis Regulatory approaches". Tale elaborato rivede i criteri dei "Fitness and Propriety" assessments utilizzati da 19 giurisdizioni successivamente alla Grande Crisi Finanziaria per assicurare che i Board members delle banche siano adeguatamente qualificati al ruolo da ricoprire.

I risultati del presente elaborato - che investiga le evidenze empiriche provenienti dalla letteratura scientifica sulle relazioni tra la Corporate Governance e la profittabilità delle banche durante la Grande Crisi Finanziaria - supportano alcune scelte fatte dai Regolatori al fine di mitigare i rischi che le lacune in Corporate Governance emerse durante la Grande Crisi Finanziaria non si ripresentino in crisi successive.

Keywords

Accountability, Bank, Board, Chair, Committee, Corporate Governance, Financial Crisis, Fit & Proper, Fitness & Propriety, Pandemic, Profitability.

DOI 10.47473/2020rmm0084

1. Introduction

This paper investigates if the failure of banks during the Great Financial Crisis - actually the time when effective Corporate Governance should have been more vital - was effectively related to their Corporate Governance and the relationship, if any, between their Corporate Governance and performance.

"*Failures in Corporate Governance were one of the key factors that contributed to the Great Financial Crisis"*[1] is one of the most important lessons learnt on the Governance of banks from the so-called Financial crisis that, after the crisis entailed by Covid-19 started in March 2020, has become more current than ever.

The Great Financial Crisis exposed shortcomings in banks' Corporate Governance practices. Investigations by national authorities and international organizations found that bank Boards were constrained by "groupthink", deferred excessively to senior management, allocated insufficient time to oversee activities and lacked experience and knowledge. Other weaknesses included ineffective Board structures and poorly designed compensation frameworks that led to excessive risk-taking.

Following the Great Financial Crisis, standard-setting bodies have reinforced bank governance standards in order to reduce the shortcomings observed during the crisis. In March 2020, the Bank for International Settlement issued the paper "Bank Boards – a review of post-crisis Regulatory approaches" that, taking stock of specific aspects of the post-crisis Regulatory approaches used in 19 jurisdictions to strengthen Board oversight at banks, reviews the "Fitness and Propriety" (F&P) assessments that these jurisdictions use to ensure that bank Board members are suitably qualified. The contents of this paper are summarized in paragraph 2. Bank Boards – a review of post-crisis Regulatory approaches. Outline.

With the aim to better understand and give sense to the choices made by Regulators after the Great Financial Crisis, this paper refers to the relevant scientific literature on how (i) the performance in one crisis has strong predictive power for following crises and how (ii) independence, time commitment, board size and diversity of the Board members affected the banks' profitability during the Great Financial Crisis. The acknowledgment of this evidence could be considered by both the Chair of the Board in leading the Board and by the Nomination Committees in determining the composition of the Boards, in order to mitigate the risk that the performance in one crisis predicts the performance during the following crisis.

This work unfolds as follows: Section 2. Bank Boards – a review of post-crisis Regulatory approaches. Outline, Section 3. Empirical results and Section 4. Conclusions.

[1] Financial Stability Institute - FSI Insights on policy implementation No 25 - Bank Boards – a review of post-crisis Regulatory approaches, 17 March 2020 https://www.bis.org/fsi/publ/insights25.htm

2. Bank Boards – a review of post-crisis Regulatory approaches. Outline

On March 17, 2020 (at the beginning of the crisis entailed by Covid-19), the Bank for International Settlement issued the paper "Bank Boards – a review of post-crisis Regulatory approaches" (BIS F&P review) about the shortcomings in banks' Corporate Governance practices revealed by the Financial Crisis occurred in years 2007 – 2008 (the so called Great Financial Crisis).

In particular, the study (i) reviews the Fitness & Propriety assessments used by 19 jurisdictions[2] to ensure that the banks' Board members are suitably qualified and (ii) includes a survey on the guidance of Board composition and structure issued in the 19 jurisdictions.

The review of F&P assessment approaches identifies useful practices for supervisory authorities that can also be useful for banks. In particular, in F&P assessment authorities might consider, where appropriate:

- the search of Regulatory powers to approve Board candidates,
- the identification of which aspects of the Fitness criterion can be enhanced to help support desired outcomes. These include clarifying the "expertise" requirements of Board candidates (particularly of the Board Chair and the Chair of Board subcommittees); assessing the time commitment of Board candidates, considering their external obligations; incorporating the "independence of mind" concept, which goes beyond determining whether candidates have a conflict of interest; and
- outlining the role of interviews in the assessment process.

In determining formal independence, supervisory assessments might be improved by defining more concrete attributes for an Independent Non-Executive Director (INED); establishing maximum INED tenure limits; and monitoring how often INEDs dissent from the majority opinion.

Based on this paper, three are the key elements of a sound and effective banks' Corporate Governance: (1) The Fitness and Propriety of the key decision-makers (Table 1), (2) the structure and composition of Boards and their Committees (Table 2) and (3) the accountability and remuneration process (Table 3). The most important topics highlighted by the BIS F&P review are summarized in the three tables below.

Table 1 The Fitness and Propriety of the key decision-makers
Selection process • The initial and ongoing assessment of a person's individual suitability is the bank's responsibility • Many Boards have established a Nomination Committee to oversee the selection process • Given the importance of ensuring the Board and senior management of a bank have the optimal mix of skills and experience, Nomination Committees are often tasked with the implementation of tools that (i) can help to identify candidates that meet the specific needs of the bank and (ii) support the assessment of those candidates against the criteria set out by the Regulatory authorities (skills matrices) • The need of an open search processes for director roles is becoming more relevant • Supervisory authority should evaluate the processes and criteria used by banks in the selection of Board members and senior management but the approaches vary across the surveyed jurisdictions. Most jurisdictions incorporate the formal approval of directors as part of the Regulatory framework • While in ECB jurisdiction the Prior approval of an initial appointment is required only in some cases, for 13 out of 19 jurisdictions assessed, the prior approval is required for all banks. In case of renewal, for 7 out of 19 jurisdictions assessed, the approval is required (in ECB, the approval is required only in some cases) • Almost all jurisdictions have the power to remove or disqualify existing Board directors prior to the end of their mandate
Fitness and Propriety criteria • The fit and proper assessment rests on a combination of principle-based and prescriptive guidance across surveyed jurisdictions • Proportionality is only applied with respect to the Fitness component of the F&P which typically covers aspects such as expected skills and experience. Proportionality does not apply to reputation and integrity features

Fitness criteria
• Of the two criteria, the "Fitness" element is where Regulatory requirements are less prescriptive
• The Fitness subcomponent of the F&P assessment includes expertise, practical experience, conflicts of interest, time commitment and, in ECB jurisdiction, independence of mind
• **Expertise** includes candidate's education and theoretical knowledge
• **Practical experience** focuses on the candidate's current and previous business positions
• **Conflicts of interest**:
  o banks are responsible for identifying any current or potential conflict: the assessment is based on criteria similar to those considered to determine whether a person is formally independent
  o must be (i) adequately disclosed, (ii) managed via a person who is not party to relevant discussions and (iii) avoided where significant. Importantly, these aspects apply to all directors at all times, and not only in relation to assessing formal independence
• **Time commitment:** the comprehensive assessment considers not only the number of directorships held, but also the size,

[2] The 19 jurisdictions are: Nigeria, South Africa, Brazil, Chile, United States, Australia, China, Hong Kong SAR, India, Malaysia, Philippines, Thailand, Belgium, ECB, Germany, Netherlands, Russia, United Kingdom, Bahrain.

nature, scale and complexity of the institutions where those directorships are held. The existence of any other professional or personal commitments is considered given the growing expectations regarding the amount of time that directors must commit to

- **Independence of mind**: In ECB jurisdiction, this feature is a key component of the F&P assessment process. By considering:
  - the candidate's character, supervisory authorities assess their ability to constructively engage with directors and senior management and, where necessary, mitigate the risk of "groupthink"
  - the candidate's behavioral attributes and the existence of relationships between the candidate and the bank, any factors which may impede them from taking an impartial perspective in discharging their responsibilities is taken into consideration. A link to the related party transactions strengthens this aspect of the framework
- **Interviews:** 10 out of 19 jurisdictions include interviews in the F&P assessment. Among others, interviews are an opportunity to sound out the candidate's understanding of the supervisors' expectations

Propriety criteria

- Of the two criteria, the "Propriety" element - that includes reputation and integrity features - is where Regulatory requirements are more prescriptive
- The definition of Propriety is consistent across the surveyed jurisdictions and rests on a consideration of whether the person has: been convicted of any crime relating to dishonesty and/or integrity; been the subject of an adverse finding in a civil action by any court; been adjudged bankrupt; been disqualified by a court or other competent bodies as a director or manager of a corporation; been a director of a company which has been wound up by a court on the application of creditors; failed to satisfy a judgment debt under a court order resulting from a business relationship; and a record of non-compliance with statutory codes as well as a record of disciplinary or other supervisory actions

Table 2
Structure and composition of Boards and their Committees

Board Diversity

- Both the Board's composition and diversity are critical to its effectiveness: Effective Corporate Governance requires a Board of directors to listen, contribute, challenge and, when necessary, push back against senior management. A Board should comprise a mix of executive directors (EDs), non-executive directors (NEDs) and independent non-executive directors (INEDs), so that it can draw on a depth and breadth of insights, perspective and experience. A broad range of skills, competencies, philosophies, life experience and diversity encompass more than gender, age, and ethnicity. Only in such a way, diversity will bolster the Board's strategic and risk-decision making abilities
- The Board and its Nomination Committee are responsible for ensuring that the Board collectively has the necessary skills, experience and expertise
- Most jurisdictions specify a minimum of three to five directors for a bank, with few specifying a maximum number and also require Boards to formulate a succession plan
- Some jurisdictions also set specific limits on the maximum tenure of an NED, to ensure that Boards benefit from fresh perspectives
- Several jurisdictions have Regulatory requirements or supervisory expectations for gender diversity: these are contained primarily within the respective Corporate Governance codes, most of which operate under a "comply or explain" rule, with disclosure and explanation required where an institution is not fully compliant with the defined principles
- Diversity helps Boards to answer the increasing need to consider the interests of a wider range of stakeholders
- Several jurisdictions have Regulatory requirements or supervisory expectations for gender diversity
- The Board's composition, including the minimum number of Independent Non-Executive Directors as well as the types of Committees required, varies across jurisdictions, according to a bank's size and the nature of its business operations. Some jurisdictions set specific limits on the maximum tenure of an INED
- There is no guidance on what constitutes the optimal size for a Board of directors and this will necessarily depend on the nature and scope of its business operations
- In jurisdictions where a dual Board structure has been adopted (including China, Germany and the Netherlands) requirements for employee representation equal to 33% of the supervisory Board members exist (or higher depending on the bank's size)

The Chair

- The Board Chair is arguably the most important position in a Board
- Even if in some jurisdiction the separation of the Chair position from that of the Chief Executive Officer (CEO) continues to be actively debated, ECB jurisdiction the Chair of the Board in its supervisory function must not be simultaneously the Chief Executive Officer
- Most authorities have issued guidance on their expectations for Board Chairs given the nature of the role and its influence on the effectiveness of Corporate Governance
- Expectations on the requisite of expertise, practical experience and time commitments required for the position of Chair are higher than the expectations set for directors. Moreover, ECB requires more experience of prospective Board Chairs in relation to other Board nominees

Independent Directors

- The definition of independence focuses on the nature of the relationships between a director and a bank and is mostly based on when the director is not independent (negative criteria):
  - **Not to have professional relationships:** most definitions prescribe a period of at least two years within which a person must not have been employed in an executive role or been a material shareholder, professional adviser, consultant, supplier or a client of the bank
  - **Not to have personal relationships**: a person who is a family member or is otherwise related to a material shareholder or to the senior management of a bank is not considered independent
  - **Limited tenure on Board**: It ranges from 6 to 12 years, with an average of 9 years. The expiration of the tenure does

- not necessarily impede a director from remaining on the Board. However, the person is then no longer considered as independent
  - **Not to be major shareholder or associate**: most definitions prescribe a period of at least 2 years within which the person must not be a material shareholder
- Supervisory requirements concerning independence will have the greatest impact where these incorporate the need to have a certain percentage of independent directors
- In a number of jurisdictions, the Chair of the Board and/or its Committees are required to be independent. Tenure also restricts this designation

Board Committees
- Specialized Committees are an accepted practice to increase efficiency and focus on specific areas
- The composition requirements are related to (i) the mix of INEDs and NEDs and (ii) who should be nominated Chair. It is worth noting that in ECB jurisdiction, the establishment of an Audit Committee, a Risk Committee and a compensation Committee is specifically stated: the mandate of each should be established in a formal charter and each Chair should be an INED
- The BCBS Guidelines[3] specifically prescribe the establishment of an Audit Committee, a Risk Committee and a Compensation (or Remuneration) Committee, noting that the mandate of each should be established in a formal charter and that the Chairs should each be an INED. These guidelines also recommend that a Nomination Committee and an Ethics Committee be established
- Currently, two jurisdictions require banks to establish an Ethics or Culture Committee (Hong Kong SAR and South Africa)
- In jurisdictions where it is specified that the Chair of the Committees must be an INED, it is also commonly stated that the Chair of the Board cannot be the Chair of any Committees but can only be a Committee member
- Most jurisdictions require the Chair of each Committee to be an INED and, in addition, most jurisdictions also require the majority of the Committee members to be INEDs
- Specific Regulatory requirements for the composition of the Audit Committees are prescribed in most jurisdictions while less specific guidance is provided on the competencies expected from directors who are members of the Risk, Compensation and Nomination Committees
- Communication process: The BCBS Guidelines specifically note the need for (i) the Risk Committee and the Audit Committee to establish protocols to facilitate the exchange of information; and for (ii) the Compensation Committee to work closely with the Risk Committee in evaluating the incentives created by the remuneration system
- In ECB jurisdiction, Board Committees should not comprise the same directors, while the cross-participation of the Chairs at the meetings of other Committees is considered a constructive way to ensure that salient matters are referred to, and discussed across, Committees
- Lesson learnt: weaknesses in communication between Board Committees and between Senior Management and the Board of the Commonwealth Bank of Australia contributed to inefficient Corporate Governance. Proper and timely Board reporting by executive management in combination with informal Board meetings may help to better face this challenge

Table 3
Accountability and remuneration process

Accountability
- Accountability for any failure in governance resides with the Board of directors as a collective body

Remuneration policy
- In ECB jurisdiction, the remuneration expectations of the members of the Board should be consistent with their powers, tasks, expertise and responsibilities. Fixed remuneration should be permanent, predetermined, non-discretionary and irrevocable, while variable remuneration should be based on performance
- The variable component of the remuneration, for each individual:
  - Shall not exceed 100% of the fixed component. Only with special approval may it be increased to 200% of the fixed component
  - At least 50% comprise a balance of shares, equivalent ownership rights, share-linked or equivalent non-cash instruments, in the case of non-listed institutions
  - At least 40% is subject to deferral arrangements

3. Scientific literature

Taking into consideration the evidence of the document "Bank Boards - a review of post-crisis Regulatory approaches" briefly summarized above, the aim of this section is twofold: to underpin the main topics considered by Bank for International Settlement with the scientific literature available on bank governance and to highlight its correlation (if any) to the performance and value, which refer to before and during the Great Financial Crisis.

Without accessing to restricted scientific data bases, the research was done using the free scientific articles published online. At the beginning of the research, the sample was made up of 35 articles scouted by two principal keywords that must always be present (financial crisis and bank) and other five ancillary keywords (Corporate Governance, loan, director, Board, performance, women). As a second step, I selected articles published after the year 2009. Furthermore, in the third step, I considered the articles in which the sample selection or the evidence were related to European or American or worldwide

[3] Guidelines on the corporate governance principles for banks, updated in 2015 by Basel Committee on Banking Supervision (BCBS) incorporating the key lessons from the Great Financial Crisis.

commercial banks (listed and non-listed). As the fourth and last step, I read the Abstract, the Introduction and the Conclusion of each of the 13 articles selected. Being aware that this sample is limited and not fully representative of all the available scientific literature on banks Corporate Governance, I believed that it was worth continuing with the analysis and share with the readers the results reached as reported in section 4. Conclusion.

Two general topics supported by the scientific literature on banks and the crisis that are worth reporting are related to (i) the learning curve of banks during the crisis and (ii) the relationship among Corporate Governance and regulations, market returns and the evaluation of banks during the crisis.

As for the **learning curve of banks during the crisis**, the study carried out by Fahllenbrach, Prilmeier and Stulz (2011) investigated whether banks' performance during the 1998 crisis - the crisis triggered by the default of Russia which set off a dramatic chain reaction within the entire global economic system - could be considered a prediction of the performance witnessed during the financial crisis of 2007 and 2008. They observed 347 banks and made 3 hypotheses: the first - the "learning hypotheses" - was based on the fact that organisations and executives that perform poorly in a crisis learn to do things differently and consequently cope better with the next crisis (negative correlation). The second - the "business model hypothesis" - was based on the fact that banks, after facing a crisis, did not change their business model, either because it would not have been profitable or for other reasons (positive correlation). The last - the "null hypotheses" - was based on the assumption that the returns during the two crises were unrelated. The authors found that banks that had been negatively affected in the 1998 crisis, subsequently, neither amended their business model nor became more risk adverse and that for each percentage point loss in the value of the equity in 1998, banks lost an annualised 66 bp during the financial crisis. Consequently, they concluded that the performance in one crisis has strong predictive power for the succeeding crisis.

Concerning the **relationship between Corporate Governance and performance** during the crisis, Cornett, McNutt and Tehranian in 2010 demonstrated that Corporate Governance was significantly correlated to 2008 market returns for larger banks but less for smaller banks. Conversely, they found that the decline in stock performance brought about by the weakness of Corporate Governance controls prior to and during the financial crisis was less significant for smaller banks.

Mixed empirical results were found by Peni and Vähämaa in 2011: while they found that the American listed commercial banks with stronger Corporate Governance had higher profitability in 2008, the results of the analysis also indicated that strong Corporate Governance practises did not create shareholder value in the banking industry during the market meltdown due to the negative effect on their stock market evaluation. Moreover, the authors demonstrated that banks with strong Corporate Governance - providing higher stock returns in the immediate aftermath of the crisis - mitigated its adverse effect on the credibility among stock market participants from March 2009 onwards.

Beltratti and Stulz (2011) found that there was no systematic evidence that a stronger regulation entailed better performance of Banks during the crisis. They found evidence that Banks from countries that imposed more restrictions in 2006 fared better during the crisis. Since there was no evidence that these banks had had fewer risks before the crisis, the authors hypothesized that banks with more restrictions on their activities before the crisis had higher returns. This being due to the fact that they did not have the opportunity to diversify their activities that were likely to perform poorly during the crisis.

Returning to the document "Bank Boards - a review of post-crisis Regulatory approaches", I highlight here below the conclusions of the scientific literature on Independence, Time commitment, Board size and Diversity during the financial crisis.

Independence

The definitions of independence provided by surveyed jurisdictions is set out from a negative standpoint (that is, the focus is on when a director is not considered independent) and is based on the nature of the relationships between a director and a bank (see Table 2). All surveyed jurisdictions provide some guidance on independence. Many of them prescribe at least a two-year time limit within which a person must not have engaged in such relationships. In addition, most countries have made the independence assessment time-bound by restricting the period a director can remain on the Board and still be considered independent.

Moreover, some countries include in the Fitness criteria the concept of "independence of mind" which goes beyond determining whether candidates have a conflict of interest and is related to the ability to challenge directors and senior management and avoid (or at least mitigate) the risk of groupthink (see Table 1).

Based on the scientific literature, during the financial crisis (i) the level of independence decreased, (ii) the bank with shareholder-friendly Boards generally fared worse as a consequence of the higher risk assumed before the crisis.

The results of the study carried out by Cornett, McNutt and Tehranian in 2010 was related to the change in Corporate Governance measures put in place by banks during the Great Financial Crisis and to how this shift negatively affected their performance. In general, they found that during the crisis, banks reported a decrease in performance and those with weak Corporate Governance controls didn't perform as well as the others. In particular, they observed the decrease of the Board's independence and that Boards failed to meet more frequently than before the financial crisis, and also that CEOs continued to serve also as Board Chair. These changes were more evident for the largest banks.

Analysing the performance of a sample of 164 banks in 32 countries, Beltratti and Stulz (2011) found that not necessarily the so called "good governance" is put in place in the interest of shareholders as they found no particular relationship between the banks' performance during the crisis and the standard values of Corporate Governance. In particular, they found that banks with more shareholder-friendly Boards[4] generally fared worse during the crisis. This being due to the fact that they, in order to

[4] Shareholders-friendly Boards could be defined as those who act in the best interest of bank shareholders.

create more value for shareholders before the crisis, left the bank more exposed to risks that adversely arose during the crisis and led to the realization of not-forecasted poor outcomes. Regardless of the independence of the Board members, indeed, Beltratti and Stulz found consistency between the risks taken by the bank before the financial crisis and the huge and unexpected significant losses accounted for during the crisis.

Consistent results were reached by Aebi, Sabato and Schmid. Indeed, in 2011, they demonstrated that, before the crisis, banks were pushed by the Board toward both the maximization of the shareholder wealth and the risks that were assumed to create wealth. During the credit crisis, these choices turned out poorly. Moreover, Erkens, Hung and Matos in 2012 - using a data set made up of 296 financial firms from 30 countries that were at the centre of the crisis - found that banks with both more independent Boards and higher institutional ownership had worse stock returns during the crisis. This was because banks that took more risk prior to the crisis resulted in larger shareholder losses during the crisis and because banks with more independent Boards raised more equity capital during the crisis which then resulted in the transfer of wealth from existing shareholders to debtholders.

Conversely, Grove, Patelli, Victoravich and Pisun (2011) did not find a consistent association between non-independent directors and performance.

Time commitment

When considering time commitment, included in the Fitness section of the F&P assessment, factors such as the number of directorships held, the size, nature, scale and complexity of the institutions where those directorships are held and the existence of any other professional or personal commitments and circumstances, form part of the Regulatory assessment process (see Table 1).

Time commitment is not a feature measurable per se: indeed, the scientific literature analyzed include concepts such as busy directors and Board meeting frequency that can be assumed as expression of the Time commitment.

Grove, Patelli, Victoravich and Pisun (2011) did not find the concave relationship between busy directors (both insider or outsider) and supposed performance: this means that when directors become too busy, or when there are too many busy directors sitting on the Board, the Board's ability to monitor effectively and efficiently is not significantly reduced.

As for the relationship between the number of Board meetings and the performance of the banks, in 2011 Grove, Patelli, Victoravich and Pisun demonstrated that the Board meeting frequency is positively associated with financial performance. In contrast, Bussoli in 2013 - analyzing to what extent the size of the Board affected the performance of Italian listed banks between 2006 and 2009 - evidenced inefficiencies in Corporate Governance due to the significant and inverse relationship between the profitability of the banks and the average attendance at Board of directors' meetings and at Committees' meetings. However, Bussoli found a direct relationship between the intensity of the activity of the Committees and the performance of the bank that can be read as a result of the attitude and the ability of Committees to limit the inefficiency of the Board.

Board size

The structure and the composition of Boards and their Committees, along with the F&P of key decision-makers, are some of the common Regulatory elements which provide a sound basis for the effective Corporate Governance of banks.

There is no definite Regulatory guidance on what constitutes the optimal size for a Board as it depends on the nature and scope of the business of the bank: while larger Boards can draw on a broader range of skills, capabilities and perspectives, smaller Boards may find decision-making more efficient, thereby providing more time for strategic discussion. Moreover, most jurisdictions specify a minimum of three to five directors for a bank, with few specifying a maximum number (see Table 1 and 2).

Grove, Patelli, Victoravich and Pisun (2011) demonstrated that Board size has a concave relationship with bank performance and loan quality as an increase in Board size is associated with financial performance. However, when the Board becomes too large, the increase can impair performance. Moreover, they showed that large Boards do not effectively monitor the lending activities of the bank that, in turn, results in lower asset quality.

Based on the Regulatory guidance, the role of the Chair remains pivotal in both large and small Boards: moreover, a number of jurisdictions retain the power to approve the dual-hatting of the Chair and CEO positions even if they restrict this discretion to limited circumstances assessed on a case-by-case basis. Consistent results were highlighted by Grove, Patelli, Victoravich and Pisun (2011) which demonstrated that CEO duality is negatively associated with financial performance.

Diversity

BCBS Guidelines state that a Board should comprise individuals with a balance of skills, diversity and expertise that should be correlated with the size, complexity and risk profile of the bank. The G20/OECD Principles also recognize the importance of bringing diversity of thought to Board discussions, stating that "*countries may wish to consider measures such as voluntary targets, disclosure requirements, Boardroom quotas and private initiatives that enhance gender diversity on Boards and in senior management*". Regulators expect that Board diversity help to mitigate against 'group think' and to expand the Board's focus to a broader range of stakeholders (see Table 2).

Consistent results were found in scientific literature: Bussoli in 2013 demonstrated the existence of a significant direct relationship between the percentage of women on the Boards and the performance of the banks.

Three Italian women researchers, Schwizer, Saoana and Cucinelli in 2013 contributed to the research on diversity considering not only gender but also nationality: they investigated the relationship between Board diversity and the performance and the cost of equity of Italian listed companies between 2007 – 2009. They concluded that the presence of women determined an increase in the frequency of the meetings of both the Board and the Audit Committee. On the other hand, they found that, due to logistic reasons, the presence of foreign members had a negative impact on the internal organization. Lastly, the research did not highlight any relationship with diversity (in terms of gender and nationality) and the cost of equity.

Enlarging the sample, similar results were reached by Andries, Seyed and Stoica in 2017 using a dataset of 156 banks from Central and Eastern Europe during 2005-2012. They assessed that banks with more women directors or chairwomen assumed lower risk and reached higher performance. The analysis first showed that banks with a chairwoman and a higher proportion of women among Board members recorded a higher level of profitability and tended to have a lower level of credit losses. Additionally, the results suggested that the higher the number of women among Board members, the greater the stability of the bank during the Great Financial Crisis. The results also revealed that the local regulatory framework affected the relationship between board gender diversity and bank performance and risk.

Islam and Md Nurul in 2020, with a sample made up of 102 U.S. listed commercial banks, found that NPLs are negatively related to board independence, CEO duality and the number of Committee meetings. Moreover, they found that, during the Great Financial Crisis, a large board size and the presence of women directors may also help lower NPLs.

4. Conclusion

The above-mentioned scientific literature - that, as said above, does not fully represent the empirical analysis available - supports the strengthening of the Corporate Governance of Banks introduced by the Regulators after the Great Financial Crisis as a measure to mitigate the risk that the performance in one crisis predicts the performance for the succeeding crisis.

In addition, this sample of scientific literature shows mixed results on the nexus between Corporate Governance and bank performance. Firstly, it supports the actions put in place by Regulators in order to strengthen the banks' Corporate Governance that, during the Great Financial Crisis, was significantly correlated to the banks' market returns in 2008. Moreover, banks with strong Corporate Governance - even if they did not create shareholder value in the banking industry during the market meltdown due to the negative effect on their stock market evaluation - mitigated the adverse effect of the crisis on their credibility among stock market participants from March 2009 onwards. Based on this scientific literature, it is also true that banks from countries that imposed more restrictions in 2006 fared better during the Great Financial Crisis because, thanks to such restrictions, they did not have the opportunity to diversify their activities that were likely to perform poorly during the crisis.

The enhancement of the requirements on independence introduced by the Regulator after the Great Financial Crisis mitigated the risk that, during the next crises, the level of independence of Boards decreased and the Board failed to meet more frequently. Contrasting results were found in this scientific literature regarding the relationship between shareholder-friendly Boards and performance during the crisis. Any consistent association between non-independent directors and performance was demonstrated.

As supported by the above-mentioned scientific literature, during the Great Financial Crisis the Board's ability to monitor effectively and efficiently was not significantly reduced even in the presence of busy directors. Nevertheless, the significant and inverse relationship between the profitability of the banks and the average attendance at Board of directors' meetings and at Committee meetings, evidenced inefficiencies in Corporate Governance. The request for Committees by the Regulators increased the performance of banks during the crisis due to the direct relationship between the intensity of activities carried out by the Committees and the performance of the bank.

During the Great Financial Crisis, Board size had a concave relationship with bank performance and loan quality: an increase in Board size was associated with financial performance however when the Board became too large the increase impaired performance. Without prejudice to the proportionality principle, the prescription of a high number of minimum seats on the Board could represent a weakness in Corporate Governance. CEO duality is also negatively associated with financial performance.

As far as the gender diversity concerns, the scientific literature analyzed reported that the higher the women on Boards, the lower is the risk assumed and the greater the profitability and the stability of banks are.

Enrica Rimoldi

Bibliography

1. Abdul Gafoor C.P., Mariappan V., Thyagarajan S., (2018), "Board characteristics and bank performance in India", *IIMB Management Review*
2. Adams R.B., Mehran H., (2008), "Corporate Performance, Board Structure, and Their Determinants in the Banking Industry", *Federal Reserve Bank of New York Staff Reports n. 330*
3. Aebi V., Sabato G., Schmid M., (2011)," Risk management, corporate governance, and bank performance in the financial crisis", *Journal of Banking & Finance 36*
4. Andries, A. M., Seyed M., Stoica O., (2017), "The Impact of Board Gender on Bank Performance and Risk in Emerging Markets", *https://ssrn.com/abstract=3041484 or http://dx.doi.org/10.2139/ssrn.3041484*
5. Beltratti A., Stulz R. M., (2011), "The Credit Crisis Around the Globe: Why Did Some Banks Perform Better?" *Fisher College of Business Working Paper Series*
6. Bekiaris M., Efthymiou T., Koutoupis A. G., (2013), "Economic crisis impact on corporate governance & Internal audit: the case of Greece", *Corporate Ownership & Control / Volume 11, Issue 1*
7. Blinco S., Galbarz M., Hohl S., Zamil R., (2020). " Bank boards – a review of post-crisis regulatory approaches", *Financial Stability Institute Insights on policy implementation No 25*
8. Boon-Hui Chai B., Tan P.S., Goh T.S., (2015), "Banking Services that Influence the Bank Performance", *6th International Research Symposium in Service Management, IRSSM-6*
9. Bubbico R., Giorgino M., Monda B., (2012), "The impact of Corporate Governance on the market value of financial institutions: empirical evidences from Italy", *MPRA Paper 45419, University Library of Munich, Germany*
10. Bussoli C, (2013), " Corporate governance and bank performance in Italy", *Financial System in troubled water, Routledge international studies in money and banking, pages 40-61*
11. Carney R.W., Child T.B., Li X., (2020), "Board connections and crisis performance: Family, state, and political networks", *Journal of Corporate Finance 64*
12. Cornett, Marcia & Mcnutt, Jamie & Tehranian, Hassan, (2010), "The financial crisis, internal corporate governance, and the performance of publicly-traded U.S. bank holding companies", *SSRN Electronic Journal. 10.2139/ssrn.1476969*
13. Dedu V., Chitan G., (2013)," The influence of internal corporate governance on bank performance - an empirical analysis for Romania", *Procedia - Social and Behavioral Sciences 99*
14. Eluyela D.F., Akintimehin O.O., Okere W., Ozordi E., Osuma G.O., Ilogho S.O., Oladipo O.A., (2018), "Board meeting frequency and Firm performance: examining the nexus in Nigerian deposit money banks", *Helyon 4 (2018) e00850*
15. Eluyela D.F., Akintimehin O.O., Okere W., Ozordi E., Osuma G. O., Ilogho S.O., Oladipo O.A., (2018), "Datasets for board meeting frequency and financial performance of Nigerian deposit money banks", *Data in Brief 19 1852 -1855*
16. Erkens D.H., Hung M., Matos P., (2012),"Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide", *Journal of Corporate Finance 18*
17. Esubalewa A.A., Raghurama A., (2020), "The mediating effect of entrepreneurs' competency on the relationship between Bank finance and performance of micro, small, and medium enterprises (MSMEs), *European Research on Management and Business Economics 26*
18. Fahlenbrach R., Prilmeier R.,. Stulz R.M., (2011)," This time is the same: using bank performance in 1998 to explain bank performance during the recent financial crisis", *National Bureau of Economic research NBER Working Paper n. 17038*
19. Garcìa-Alcober M.P., Prior D., Tortosa-Ausina E., Illueca M., (2019), " Risk-taking behavior, earnings quality, and bank performance: A profit frontier approach", *BRQ Business Research Quarterly*
20. Grove H., Patelli L., Victoravich L.M., and Xu T., (2011), "Corporate Governance and Performance in the Wake of the Financial Crisis: Evidence from US Commercial Banks", *Corporate Governance: An International Review*, 19(5): 418–436
21. Islam, Md Nurul, (2020), "The Impact of Board Composition and Activity on Non-Performing Loans", *Awards for Excellence in Student Research and Creative Activity. 7.* https://thekeep.eiu.edu/lib_awards_2020_docs/7
22. Karas A., (2020), "Russian bank data: Reasons of bank closure", *Data in brief 29 105343*
23. Lafuente E., Vaillant Y., Vendrell-Herrero F., (2019),"Conformance and performance roles of bank boards: The connection between non-performing loans and non-performing directorships", *European Management Journal*
24. Mkhaiber A., Werner R.A.,(2020), "The relationship between bank size and the propensity to lend to small firms: New empirical evidence from a large sample", *Journal of International Money and Finance 110*
25. Naser N, (2019), "A Comprehensive Analysis of European Banking Soundness Theoretical Study", *Financial Markets, Institutions and Risks, Volume 3, Issue 2*
26. Onakoya A.B.O., Ofoegbu D.I., Fasanya I.O., (2012), "Corporate governance and bank performance: a pooled study of selected banks in Nigeria", *European Scientific Journal,* December edition vol. 8, No. 28, ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431
27. Peni E., Vähämaa S., (2012)," Did Good Corporate Governance Improve Bank Performance during the Financial Crisis?", *J Financ Serv Res, 41:19-35*
28. Schwizer P., Soana M.G., Cucinelli D, (2013), " The advantages of board diversity. An empirical analysis of the Italian market.", *Financial System in troubled water, Routledge international studies in money and banking, pages 79-96*
29. Staikouras P.K., Staikouras C.K., Agoraki M.E.K., (2007), "The effect of board size and composition on European bank performance ", *Eur J Law Econ 23:1-27*
30. Takacs A., Szucs T., Kehl D., Fodor A., (2020), "The effect of fair valuation on banks' earnings quality: empirical evidence from developed and emerging European countries", *Heliyon 6*
31. Tu DQ. Le, Ngo T., (2020), "The determinants of bank profitability: A cross-country analysis", *Central Bank Review 20 65-73*
32. Tunay K.B., Tunay N., Akhisar I. (2015),"Interaction Between Internet Banking and Bank Performance: The Case of Europe", *Procedia - Social and Behavioral Sciences 195*
33. Villanueva-Villar M., Rivo-López E., Lago-Pe˜nas S., (2016)," On the relationship between corporate governance and value creation in an economic crisis: Empirical evidence for the Spanish case", *BRQ Business Research Quarterly, 19, 233-245*
34. Zaharia D.L., Lazar A., Tilea D.M., (2013), "The Role of Internal Audit regarding the Corporate Governance and the Current Crisis", *5th World Conference on Educational Sciences – WCES 2013*
35. Zephaniah C.O., Ogbab I.E., Izogo E.E., (2020), "Examining the effect of customers' perception of bank marketing communication on customer loyalty", *Scientific African 8*

Risk Management Magazine
Volume 16, Issue 1, January – April 2021

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


VACCINO
ANTI
DEBITO
PUBBLICO
SALUTI
DALLA B.C.E.
Silva