BANCA D'ITALIA
EUROSISTEMA

# Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

## The expert assessment within Banca d'Italia's in-house credit assessment system

by Lorenzo Esposito, Massimo Guglielmi, Francesco Monterisi, Simone Narizzano and Marco Orlandi

March 2026

Number 79

BANCA D'ITALIA
EUROSISTEMA

# Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

## The expert assessment within Banca d'Italia's in-house credit assessment system

by Lorenzo Esposito, Massimo Guglielmi, Francesco Monterisi, Simone Narizzano and Marco Orlandi

Number 79 – March 2026

*The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of the Bank of Italy in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.*

*The views expressed in the papers are those of the authors and do not necessarily reflect those of Banca d'Italia.*

*The series is available online at www.bancaditalia.it.*

*Printed copies can be requested from the Paolo Baffi Library: richieste.pubblicazioni@bancaditalia.it.*



ISSN 2724-6418 (online)
ISSN 2724-640X (print)

Banca d'Italia
Via Nazionale, 91 - 00184 Rome - Italy
+39 06 47921

*Designed and printing by the Printing and Publishing Division of Banca d'Italia*

# THE EXPERT ASSESSMENT WITHIN BANCA D'ITALIA'S IN-HOUSE CREDIT ASSESSMENT SYSTEM

**by Lorenzo Esposito*, Massimo Guglielmi †, Francesco Monterisi †, Simone Narizzano †, Marco Orlandi †**


### Abstract

This study investigates the role of the expert assessment – conducted by two analysts following the production of a rating based solely on the statistical model – within Banca d'Italia's in-house credit assessment system of Italian non-financial firms (ICAS). We have two aims: to document recent methodological enhancements, including the integration of climate-related risks and of sector analysis, and to provide an estimate of the contribution of the expert assessment to the rating process. The study leverages over 25,000 assessments produced by analysts between 2016 and 2022, including corporate default events. The recent methodological innovations have enhanced the transparency and consistency of the expert assessment, facilitating the integration of new risk sources. Our empirical results show that the expert assessment significantly improves both the predictive power and the discriminatory power of the full ratings obtained through ICAS, compared with the ratings based solely on the statistical model, with an increase of the AUROC of around 2 percentage points. Furthermore, the expert assessment protects the performance of ICAS, particularly during periods of macroeconomic stress.



**JEL Classification:** D82, G21, G24, G32, E52.

**Keywords:** Model-based ratings; Credit risk; Collateral; Central bank refinancing; Credit rating.



### Sintesi

Il lavoro analizza il ruolo della valutazione esperta – condotta da due analisti successivamente alla produzione del rating basato sul solo modello statistico – nel sistema interno di valutazione del merito creditizio delle imprese non finanziarie (ICAS) della Banca d'Italia. L'obiettivo è duplice: documentare i recenti miglioramenti metodologici apportati alla valutazione esperta dell'ICAS, tra cui l'integrazione dei rischi climatici e dell'analisi settoriale, e fornire una stima del contributo della valutazione esperta al processo di rating. Lo studio si basa su oltre 25.000 valutazioni formulate dagli esperti tra il 2016 e il 2022. Le recenti innovazioni metodologiche hanno migliorato la trasparenza e la coerenza delle valutazioni degli esperti, favorendo l'integrazione di nuove fonti di rischio. I risultati empirici mostrano che la valutazione esperta migliora significativamente sia il potere predittivo sia il potere discriminante dei rating completi prodotti dall'ICAS rispetto ai rating prodotti dal solo modello statistico, con un incremento dell'AUROC di circa 2 punti percentuali. Inoltre, la valutazione esperta preserva la performance dell'ICAS specialmente durante periodi di stress macroeconomico.



* Banca d'Italia, Milan Branch.

† Banca d'Italia, Financial Risk Management Directorate.

# CONTENTS

## 1. Introduction[1]

Since 2013 the Banca d'Italia has been operating the in-house credit assessment system (ICAS). The system plays an important role in monetary policy implementation by ensuring the quality of a vast amount of credit claims pledged as collateral in refinancing operations.[2] ICAS consists of two components: (i) the statistical model, covering a broad sample of Italian firms, and (ii) the expert assessment, applied to the subsample of the most significant firms for collateral purposes. ICAS ratings are aligned with the Eurosystem high credit quality standards and contribute to the effective transmission of monetary policy.

The structure of the Banca d'Italia's ICAS reflects a common approach in the banking and finance industry. Statistical models, valued for their objectivity, scalability, and reliance on historical data, offer speed and consistency in credit assessment. However, they may fall short of capturing forward-looking variables or incorporating qualitative factors. The expert assessment introduces a forward-looking perspective and qualitative judgement, helping to address the limitations of purely statistical approaches, albeit at the cost of increased subjectivity.

At the Banca d'Italia, the expert assessment is carried out by two credit analysts who may either confirm the rating obtained from the statistical model (the 'statistical rating') or revise it upward or downward. Starting from the statistical rating, the analysts independently review the main characteristics of the firm (size, business sector, geographical location, etc.) along with the variables underlying the statistical assessment; they then assess a set of predefined risk profiles (also dubbed as 'assessment modules') and determine a score for each of them. Each score reflects whether the analyst believes that an individual module improves, confirms or worsens the risk assessment of the statistical model. The module scores are subsequently weighted and aggregated to determine the final rating, also dubbed as the 'full rating'. If the two analysts broadly agree and their assessment does not lead to a major improvement vis-à-vis the statistical rating, the process stops. Otherwise, the rating proposal is submitted for further review to the Rating committee, composed of senior staff members, that takes the final decision.

This study has two aims. First, it illustrates the expert assessment framework and its recent methodological developments, including the integration of climate-related risks and major changes in the sector analysis. Second, it provides evidence on the value added by the expert assessment to the whole credit evaluation process. The current methodology benefits from improvements carried out over time to increase the

[1] We are grateful to Antonio Scalia for his useful comments. We also thank Alessio Martone and Stefano Di Virgilio for their fundamental contribution in the development of the Bayesian analysis that is at the basis of the overall assessment. A special thanks goes to Chara Prassa and Apostolos Katsafados of Bank of Greece, for some very useful insights about validation.

[2] The ICAS rating process of the Banca d'Italia is managed in accordance with the principles of the Eurosystem Credit Assessment Framework (ECAF). Similar systems are also employed by other National Central Banks (NCBs) of the Eurosystem.

homogeneity, reliability, and transparency of the qualitative credit assessment conducted by the analysts. In particular, a comprehensive review was carried out in 2023.

An internal validation exercise performed before the review of the methodology showed that the expert assessment had significantly enhanced the discriminatory power, the predictive power and the stability of the statistical model.[3] Specifically, over the review period the AUROC[4] had improved by nearly 2 percentage points, from 85.3 to 87 percent. Furthermore, the expert assessment had introduced some degree of conservatism, thus strengthening the robustness of the system during periods of heightened uncertainty, such as financial crises or abrupt macroeconomic shifts.

Our validation results, as well as the substantial enhancement under the current version, suggest that the gains of the expert assessment have significantly increased compared with the estimates presented in this study, which may be viewed as a lower bound. We will compute an empirical estimation of the value added by the current methodology once an adequate statistical sample will be available.

The remainder of the paper is organized as follows: Section 2 reviews the best practices in the credit risk industry; Section 3 outlines the Eurosystem framework within which the Banca d'Italia's ICAS operates; Section 4 describes the ICAS expert assessment model in its current version; Section 5 describes the main findings from the internal validation exercise as applied to the previous version of the expert assessment, which was employed until 2023; Section 6 discusses how analysts can deal with emerging risk categories, like ESG risks and geopolitical risks; Section 7 concludes.

## 2. The expert assessment of credit risk

Counterparty credit risk is the primary source of risk in banking operations, as noted e.g. by the Basel Committee and national banking regulators (BCBS, 2024). The development of a robust credit rating system is thus of paramount importance for each financial institution as well as for the financial system. To this end, credit risk service providers and lenders managing credit risk systems have established

[3] The discriminatory power concerns the ability of a model to distinguish firms on the basis of their future status (default or non-default) over a predefined time horizon. So, it should reflect the following properties: i) specificity, that is the ability to correctly classify the units for which the default does not occur; ii) sensitivity, that is the ability to correctly classify the units for which the default occurs. The predictive power (or calibration quality) refers to the model's ability to identify the 'real' probability of default for an individual debtor or a class of homogeneous debtors. Predictive power tests compare the number of defaults that actually occurred in a given rating class with the number of defaults predicted by the model. The stability of a rating system is the ability to distinguish between real causes/effects and purely random relationships; unstable rating systems have a disappointing performance (i.e. they lose discriminatory power or predictive power) if applied to databases other than that for which they were developed.

[4] The discriminatory power of a model is typically assessed with the Area under the receiver operating characteristic (AUROC). AUROC is a measure of the ability of the model to assign higher PDs to firms that will default compared with financially sound firms. By construction, the AUROC statistic ranges from 0 ('the model is completely wrong') to 1 ('the model discriminates perfectly'), whereas 0.5 indicates a purely random model. In practice, a rating system with AUROC $\geq 0.7$ is usually considered as adequate.

comprehensive databases, tools, and procedures that collectively define credit risk management activities. Corporate credit analysis is centered around the company’s structural ability to generate income and service its debt.

The practice of corporate credit analysis generally relies on common components as concerns input data and models. Ratings are usually updated annually, in accordance with the fiscal year, as a market and regulatory standard (BCBS, 2015). Major credit risk agencies (CRAs) manage a range of models, because ratings cannot be applied uniformly to borrowers of different size or from different sectors (Garcia Alcubilla and Ruiz del Pozo, 2012). Pro-cyclicality and reputational risks may affect ratings (Amato and Furfine, 2003). Markets and financial players have long valued the comparability and the general clarity of CRA ratings. The use of these ratings for regulatory purposes increased in the 1990s, in the run-up to the great financial crisis (EU, 2009).

The CRAs’ rating methodology is based on two components: a quantitative model and a qualitative assessment. The output of the quantitative model results from the integration of a few predictive variables of the default event and serves as the starting point for the rating process. The qualitative assessment incorporates expert and market views, based on general guidelines. The qualitative assessment is even more important for large corporates and in the current international economic and financial context marked by exceptional uncertainty (Ballis *et al.*, 2024). CRAs prioritize sector risk as it helps understand the outlook for the corporation and its competitive environment. Governance and group structure are also examined, owing to their effect on corporate strategy, business model, transparency, and integrity. Rating agencies carry out due diligence and meetings with the firm’s top management as part of the qualitative assessment of the company. The rating is accompanied by an ‘outlook’, that shows the likely direction of the rating over the medium term, i.e. in twelve months time.[5]

Internal ratings-based (IRB) models managed by banks are subject to certain requirements to be eligible for regulatory purposes, at the outset and on an ongoing basis.[6] Like CRAs’ models, IRB methodologies: (i) vary significantly with firm size;[7] (ii) include a quantitative component and a qualitative assessment; (iii) share the same input data. However, IRB models have a comparative advantage over CRAs because banks own a richer information set concerning payment flows, which is particularly important in the case of SMEs.

[2] As for banks, our second source of ratings, in the 1990s, several major financial institutions developed advanced credit risk models such as CreditMetrics (J.P. Morgan) and CreditRisk+ (Crédit Suisse), that have since become market standards, and, with Basel II, also regulatory standards (Gallo, 2021).
[6] See BCBS, 2001 and EBA, 2017.
[7] For instance, transactional behaviour detected from current account flows is less important for large corporates than small and medium enterprises (SMEs).

CRAs faced significant criticism after the Enron and WorldCom scandals, and even more so following the 2008 financial crisis, with accusations of lax rating standards and potential conflicts of interest (Alp, 2013). The credibility of CRAs rating models was questioned (deHaan, 2017) due also to increased competition (Becker and Milbourn, 2011). Standard setters observed that supervisors should be cautious against over-reliance on internal models for credit risk management (BIS, 2015). More recently, geopolitical factors may have contributed to cyclicality in credit ratings, with data indicating optimistic ratings during risky periods, linked to conflict-of-interest issues (Singhal *et al.*, 2024).

## 3. The Eurosystem credit assessment framework: the role of ICASes

The collateral framework is one of the pillars for the implementation of monetary policy. All Eurosystem liquidity providing operations require counterparties to provide adequate collateral. The credit risk framework is designed to safeguard the Eurosystem's balance sheet against losses arising from the default of a monetary policy counterparty. The Eurosystem Credit Assessment Framework (ECAF)[8] lays down rules and procedures that ensure that all assets eligible for monetary policy operations meet high credit quality standards. With a view to accepting a broad range of marketable and non-marketable assets as collateral, the Eurosystem relies on three credit assessment sources (Narizzano *et al.*, 2024): 1) CRAs accepted as ECAIs; 2) ICASes of National Central Banks (NCBs); 3) counterparties' IRB systems that are accepted for setting banks' regulatory capital requirements.

ICASes play an important role by enabling counterparties to pledge credit claims granted to non-financial counterparties (NFCs) as collateral even in the absence of an IRB system.[9] Notably, as of June 2025 the Eurosystem applies the prioritization of ICAS ratings, where available, over those from other credit assessment sources to establish the eligibility of credit claims and the applicable valuation haircuts.[10]

The quality and reliability of ICAS ratings are ensured by a set of Eurosystem rules. Although ICAS ratings are not subject to the CRA Regulation of the European Union, they must comply with certain standards that reflect the industry best practices in terms of organization, resources, and governance.[11] ICASes must ensure that the development of methodologies, validation and performance monitoring, and the rating of entities, are allocated to different units run by separate management lines. The resources devoted to credit assessment should be proportional to the number of rated entities. ICASes estimate credit risk as the probability of default (PD) over a one-year horizon, based on a common default definition, as

[8] For more information, see https://www.ecb.europa.eu/paym/coll/risk/ecaf/html/index.en.html.
[9] ICASes may assess NFCs of any industry, size and/or legal form; ICAS NCBs should inform the Eurosystem about the criteria used to select the entities to be assessed by the system, for example in terms of sector and size. Monetary policy counterparties can request a rating for a specific NFC upon submission of assets to be potentially mobilised as collateral.
[10] See harmonised-rules-for-Eurosystem-collateral-management.
[11] Regulation (EC) No 1060/2009 and its subsequent amendments.

per Article 178 of the Regulation (EU) No 575/2013 – CRR. In the Italian case, the PD is mapped to the Banca d'Italia's credit risk classes and to the credit quality steps (CQS) on the Eurosystem harmonized rating scale (see Appendix 1 for details). The Eurosystem's guidelines for ICASes emphasize accuracy, consistency, and comparability in rating activities while allowing for some flexibility in implementation (Auria *et al.*, 2021).

Under normal circumstances, ICASes ensure adequate collateral availability for a wide range of counterparties with different business models, thereby facilitating the smooth implementation of monetary policy. The acceptance of bank loans as collateral mitigates the need for counterparties to hold marketable assets solely for collateralizing monetary policy operations. Besides, bank loans incur in relatively low opportunity costs compared to marketable assets, which are predominantly used as collateral in private market repo transactions.

During periods of market stress, ICASes may enhance the monetary policy transmission mechanism by allowing banks to increase the proportion of non-marketable collateral, especially if marketable assets become scarce or their value declines. This role was exemplified by the Eurosystem's measures implemented in April 2020, amid the pandemic-induced financial and economic crisis, which eased the use of bank loans to corporates and households as collateral.[12]

ICASes thus contribute to the diversification of risks on the Eurosystem balance sheet and reduce the reliance on external credit rating agencies. Besides, NCBs may use ICAS ratings for other purposes along with monetary policy operations, such as emergency liquidity assistance (ELA) operations, financial stability analysis, economic research, and supervisory tasks.

## 4. Methodology

### 4.1 General principles and new features

The ICAS rating methodology follows a two-stage approach that combines a statistical model with the expert assessment (Giovannelli *et al.*, 2020).

The statistical model employs quantitative variables and yields the statistical rating. It comprises a system of logit models with two independent components, each providing a separate credit score (Narizzano *et al.*, 2024):

[12] This increase was to be achieved by expanding the use of credit claims as collateral, particularly the so-called Additional Credit Claims Framework (ACC) that allow NCBs to enlarge the scope of eligible credit claims for counterparties in their jurisdictions. Credit claims, and in particular ACCs, are typically mobilised by relying on ICAS or IRB assessments. The effectiveness of these measures was demonstrated by the significant increase in such collateral that was observed in several jurisdictions in 2020.

- financial component: it employs a logit regression on annual financial statement data, such as the debt sustainability ratio, financial structure, and liquidity ratios. This component consists of eleven sub-models that account for different sectors and types of financial statement;
- credit behaviour component: it employs a logit regression on individual data from the National Credit Register (NCR). Three sub-models are designed for different firm size (micro, small, and medium-large).[13]

In line with banking industry practices, the two components are estimated separately and then combined into the final model through an additional logistic regression that yields the final statistical score. The integration of the two components is carried out with four models based on size (micro, small, medium, and large firms), since the relative importance of financial and credit behaviour information varies with firm size.

The expert assessment involves the financial analysts' assessment of credit risk with a forward-looking perspective, employing a wide range of information sources. The analysis can either confirm the rating obtained from the statistical model or revise it upward or downward. In particular, two credit analysts review the statistical rating employing quantitative data and qualitative information. Each analyst independently evaluates predefined assessment modules that capture firm-specific features (financial statement, financial flexibility, business sector, governance, etc.). They assign a score to each module based on their evaluation of the firm's creditworthiness, in comparison with the statistical rating. These scores are then weighted and aggregated to obtain the full rating.

The Banca d'Italia has identified several principles and desirable properties for the expert assessment methodology:

- objectivity and credibility: credit ratings must be perceived as reliable and trustworthy by all users;
- quality, completeness, and integrity: the methodology should encompass all key factors affecting default risk. The data used in developing credit ratings must be of high quality and sourced from reliable providers;
- transparency and disclosure: the methodology should guide analysts through a clear and traceable decision-making process;[14]
- homogeneity: the rating process should ensure a uniform treatment of information to maintain consistency across analysts and over time; [15]

[13] We follow the European Commission definition (Recommendation 2003/361/EC). For more information, see https://ec.europa.eu/growth/smes/sme-definition_en.
[14] Each phase of the credit assessment should be consistently and transparently reflected in the full rating. Any deviation from the general methodology should be explained and justified using a 'comply or explain' approach.
[15] Similar business cases should be evaluated in a homogeneous manner to support comparability and methodological coherence.

- robustness: the weighting of the qualitative and quantitative stages should be based on robust analysis, ideally grounded in observed data and/or aligned with industry best practices;
- flexibility: the rating process should allow analysts to address all potential business idiosyncrasies without being constrained by overly rigid rules.[16]

Following the internal validation exercise conducted in 2023, several improvements have been introduced in 2024 to improve the quality and consistency of expert assessments. Structured questionnaires (or checklists) have been developed to guide analysts through the evaluation of each risk profile. This ensures a more comprehensive and consistent assessment of all relevant risk factors, reducing the possibility of omissions or subjective biases. A Bayesian aggregation methodology now combines the partial scores attributed to each risk area. This provides analysts with a coherent statistical framework that supports final decisions (to confirm, upgrade, or downgrade the initial model-based rating) based on the overall assessment derived from the single profiles modules.

Analysts may preliminarily adjust the statistical PD by changing some predictive variables or adding new ones, where justified by firm-specific accounting practices or data peculiarities (e.g. finance leasing not included in financial debt according to Italian GAAP financial statements). Sector and group analyses have been strengthened to better capture a firm's competitive position within its economic sector and its financial interlinkages within the group.[17] The coverage of risk profiles has been expanded, with the inclusion of climate-related risks in line with evolving standards for integrating sustainability considerations into credit assessments.

### 4.2 Procedure

The expert assessment procedure is organized into the following steps, which analysts are required to perform sequentially through the dedicated modules that capture firm-specific characteristics (Figure 1):

- initial statistical rating adjustment;
- financial statement (module 1);
- financial flexibility (module 2);
- governance (module 3);
- sector analysis (module 4);
- third-party opinions (module 5);
- climate change (module 6);
- overall assessment and stand-alone full rating;

[16] The value added of qualitative assessment often lies in the ability to integrate subjective or 'soft' information that is typically not included in statistical models. However, the exercise of such flexibility should be properly documented and consistently monitored throughout the rating process.
[17] If the company is affiliated to a group, we estimate the intensity of the interlinkages within the group and establish the impact of the creditworthiness of the parent company at a consolidated level.

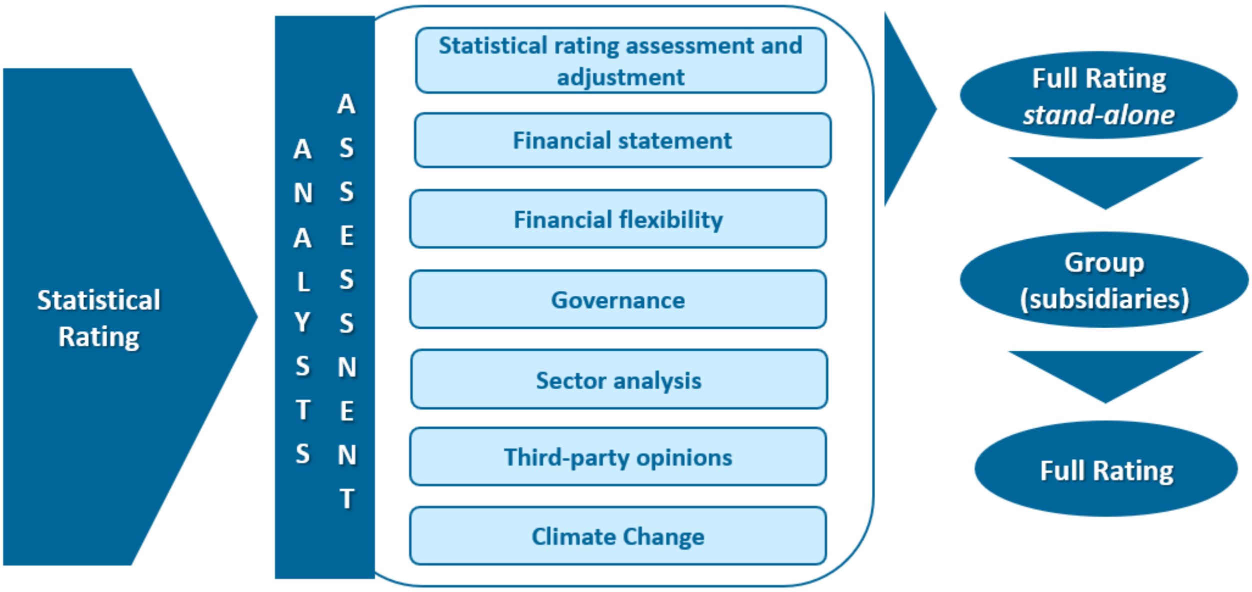
- for subsidiaries only: group analysis (module 7) and full rating.

For each module the analysts assign a score (the 'module score') on a five-point scale: very positive, positive, neutral, negative, and very negative, corresponding to the cardinal values +2, +1, 0, -1, and -2, respectively. These scores result from the aggregation of the answers given by the analysts to the checklist for each module. The answers are standardized on the same five-point scale as above.

The stand-alone rating is obtained by means of the Bayesian integration of the scores assigned to modules 1-6 (see Section 4.2.8). In the case of subsidiaries, in module 7 an adjustment to the stand-alone rating may be made based on the strength of the interlinkages with the group and the creditworthiness of the parent company.

Automated aggregation criteria ensure consistency across different analyses. However, analysts may override both the scores and the stand-alone rating, provided that such deviations are properly explained.

**Figure 1 – The expert assessment procedure**




The expert assessment is carried out in sequential steps by two different analysts; in specific cases the Rating committee reviews the assessment and makes the final decision on the full rating. The first analyst assesses the modules. Every score is accompanied by an explanation; this should be sufficiently detailed to allow the second analyst and, possibly, the Rating committee[18] to check the motivation. The second

[18] The Rating committee is composed of senior managers within the Financial Risk Management Directorate. The analysts involved in the assessment are also required to attend the committee meeting. The Committee is responsible for the review of the ratings that involve significant financial and/or reputational risks. It also acts as the decision-making body in case of discrepancies between the ratings put forward by the two analysts. The Committee takes the final decision, which can involve a rating upgrade of more than 1 notch.

analyst reviews and double-checks the coherence of the analysis performed by the first analyst, under a four-eyes approach.

### 4.2.1 Initial statistical rating adjustment

The expert assessment starts by reviewing the quality and consistency of the input data for the statistical rating. This step serves also to integrate any additional information obtained by the analysts through their own research on the firm (financial reports, NCR and Analytical Credit[19] data) that may lead to a revision of the statistical rating. The assessment of the financial statement component concerns: financial leases, cash and cash equivalents, non-recurring items in the income statement, dividends, interest expenses, hybrid shareholders' loans. The assessment of the credit behaviour component involves average utilization rates and financial distress variables (see Appendix 2 for details). The statistical rating resulting from the initial adjustment made by the analysts is the starting point for the expert assessment.

### 4.2.2 Financial statement

Financial statement analysis involves the examination of a broader set of indicators than those considered by the statistical model. This module includes peer group analysis and any other new information (e.g., infra-annual reporting) with a forward-looking perspective. The empirical literature confirms the role of financial ratios towards the prediction of firm default.[20]

Financial statement analysis covers four aspects: (i) profitability; (ii) capital adequacy and financial structure; (iii) debt sustainability; (iv) liquidity and cash generation.[21]

The corresponding sections include a broad set of indicators available over the last four financial years, accompanied by a quintile-based classification in the peer group of companies (Figure 2).

[19] Analytical Credit (AnaCredit) is a dataset containing detailed information on individual bank loans in the euro area, harmonized across all Member States. The AnaCredit dataset project was launched in 2011 and data collection started in September 2018. On 18 May 2016 the ECB adopted Regulation (EU) 2016/867 on the collection of granular credit and credit risk data (known as 'AnaCredit Regulation), following the principles approved by the Governing Council in 2015.
[20] Modigliani and Miller (1958), Altman (1968).
[21] The web-based platform automatically provides analysts with detailed information on financial reporting data (balance sheet, income, cash-flow statement, etc.) to facilitate their assessment.

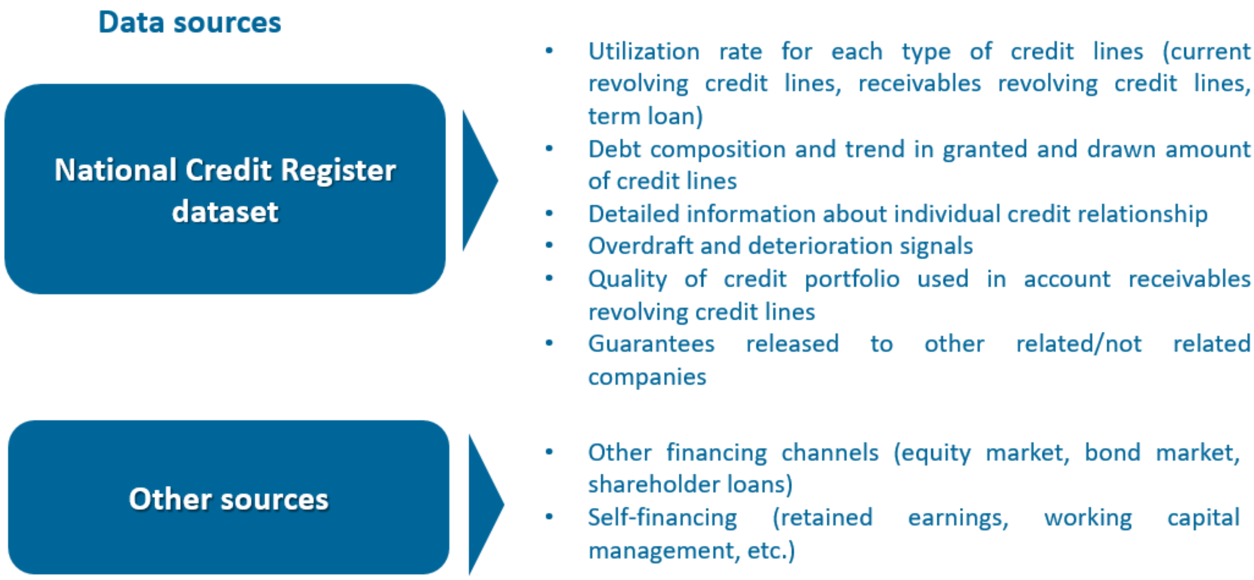
**Figure 2 - Financial statement**


The first analyst notes the elements obtained from financial reporting and any other useful information. The first and second analyst fill out independently a checklist. Each answer is weighted, contributing to the module score. This can be reviewed by both analysts. Once the assessment is completed, the procedure weights equally the profile scores to yield an overall score of the financial statement module. The final score can be reviewed by both analysts, enabling them to reach independent conclusions.

#### 4.2.3 Financial flexibility

As part of the financial flexibility module, the analysts evaluate the company's ability to mitigate the negative impact on financial resources of external shocks relating to the economic cycle or of internal shocks linked to its own activity. The assessment involves an in-depth analysis of the NCR dataset and of additional information (Figure 3).

**Figure 3 - Financial flexibility**




The NCR dataset provides extensive information on different credit lines held by each firm. The dataset distinguishes three classes of loans: term loans, current account revolving credit lines, and accounts receivable revolving credit lines for both short term and medium to long term credit lines. For each credit line, the dataset records drawn and undrawn amounts, excess and unauthorized overdrafts, and default classifications.

Several studies confirm the role of credit behaviour in discriminating between good and bad firms.[22] Credit line usage and excess overdraft are significant explanatory variables of default, especially for small business firms (Norden and Weber, 2010; Giannozzi *et al.*, 2013). Additionally, past delinquencies are positively correlated with default (Gallucci *et al.*, 2022). Firms with significant credit line usage may face liquidity issues, and high leverage or poor financial performance may hinder their ability to acquire additional funding (Zhao *et al.*, 2014). The utilization rate for current account revolving credit lines and excess overdrafts are good in discriminating between non-defaulted and defaulted Italian SMEs.

Analysts have access to the credit-related flows primarily from the NCR. This module automatically provides data covering the last 13 months. This information helps assess the relationship between the firm and the banking system, which is particularly important in Italy where firms, especially small ones, heavily rely on bank financing. Changes in the usage patterns of granted credit lines or the frequency of overdrafts are critical for identifying potential tensions in financial resources.

The analysis is structured along three sections, leading to the score on financial flexibility.

The first section contains NCR data on: i) the amount drawn and granted for short-term credit lines according to different technical forms, such as current account revolving credit lines and accounts

[22] Agarwal *et al.*, 2006; Jiménez *et al.*, 2009; Sufi, 2009.

receivable revolving credit lines; ii) the distribution of credit lines by maturity; iii) detailed data on term loans and lease financing. The analysts assess the amount of undrawn short-term financing lines to check for sufficient financial flexibility in the event of a liquidity shortage. While the statistical model incorporates the used-to-granted-credit ratios, it does not account for the absolute amount of undrawn credit lines. Analysts therefore complement the statistical model's output by evaluating whether the residual availability on short-term credit lines is sufficient to absorb any temporary liquidity shortfall. They also assess the adequacy of the firm's debt structure in relation to its net working capital and the evolution of financial leverage.

In the second section the analysts examine the quality of credit receivables and check for potential risks related to off-balance sheet guarantees issued by the firm and negative fair value on derivatives. Besides, the analysts assess whether the firm has actual or potential access to additional financing sources, such as private equity capital, debt and capital markets, the international banking system. The number of active banking relationships, the nature of the intermediaries (local vs big banks, factoring and leasing intermediaries, etc.), and their changes over time can provide additional insight.

The last section contains information on overdrafts and non-performing loans over the past 13 months. The analysts check whether a fractional default[23] has occurred using both NCR and AnaCredit sources.

### 4.2.4 Governance

The academic literature has long highlighted the importance of corporate governance in the context of corporate stability and efficiency.[24] Firms with better management practices tend to perform better: they are larger in scale, more profitable, have higher labour productivity, and are more likely to export (Bloom and Van Reenen, 2007). Strong corporate governance can also mitigate conflicts of interest between shareholders and company management, thus reducing risks for creditors. It also contributes to greater transparency and better disclosure of information, which are crucial elements for assessing credit risk (Diamond, 1984). Finally, the composition of the board of directors, particularly the presence of independent directors, strong internal control mechanisms and well-structured risk management correlate with a robust corporate governance (Anderson *et al.*, 2004).

The analysis of this risk factor is conducted through a questionnaire (or checklist), to evaluate the key aspects of corporate governance. that are known to influence a firm's risk profile and creditworthiness.

[23] The default definition aggregates the whole default information into a single default indicator. Two thresholds are applied:

- a materiality threshold of 2.5 per cent of the defaulted amount over the total exposure of the debtor;
- persistence for three consecutive months.

Over a given monitoring period, fractional default is equal to the maximum proportion in default which fulfils the two above conditions. In the case of bankruptcy, insolvency, judicial administration or similar measures (legal default) the materiality is considered equal to 100 per cent and no persistency is required.

[24] Jensen and Meckling, 1976; Shleifer and Vishny, 1997.

Factors such as stock exchange listing and adherence to governance codes of conduct signal transparency and external accountability. The structure and functioning of the board - especially the presence of independent directors, managerial competence, and internal committees - are viewed as indicators of effective oversight. Robust internal controls, including rules on conflicts of interest and related-party transactions, help reduce operational and reputational risks. The quality and transparency of financial reporting, along with the auditor's findings, are also considered as essential for the evaluation of the reliability of disclosed information. Adverse legal rulings, public controversies, and exposure to social risks are monitored as potential red flags. Overall, strong governance practices are viewed as a positive signal, enhancing the credibility of the firm's rating in addition to the results of quantitative models.

Based on the analysts' responses, the procedure computes the module risk score. The analysts may override the score, explaining the reason for the change in a specific comment line.

#### 4.2.5 Sector analysis

This module has three sections: (i) strategic positioning, (ii) country risk, and (iii) environmental risk. Based on the analysts' responses, a score is produced for each section.[25]

With reference to strategic positioning, the analysts consider the conditions and potential evolution of the economic sector in which the company primarily operates, in view of the macroeconomic outlook. To support the analysis of the economic environment and industrial sectors, industry forecasts and other data are obtained from an external provider. The sector outlook is evaluated from both a structural and a forward-looking perspective. The analysts examine the company's positioning within the sector, considering factors such as market size/share, product differentiation, brand recognizability, innovation, and diversification of offerings. Additionally, the analysis encompasses broader sectoral dynamics, including the sensitivity of sector demand to macroeconomic trends, the sector's stage in its lifecycle, the intensity of competition, capital requirements, and the economies of scale.

The country risk section is applicable if net revenues produced abroad exceed 30 per cent of total revenues. In this case, the score attribution is guided by a checklist and quantitative indicators obtained from external providers. These indicators measure the country's attractiveness in terms of export and investment opportunities, political and legal risks, and social situation.

The section relating to environmental risk assesses the possibility that the company business may cause (or be negatively affected by) environmental issues. The exposure to environmental factors is strictly related to the sector in which the firm operates. Examples include air, water, and soil pollution, loss of

[25] Environmental risks and opportunities are included in this area of the expert assessment because the exposure to environmental factors (i.e. Energy and Fuel Management, Air and water pollution, Waste of raw materials, Water usage, Biodiversity and ecological impact) is related to the sector in which the firm operates.

biodiversity, soil deterioration and habitat loss, and exposure to hazardous chemicals. The environmental risk questions address the company's exposure to environmental risks, its reliance on natural resources, the impact of its production processes, and its ability to operate sustainably. The analysts also consider reputational and legal risks arising from lack of corporate attention to environmental sustainability and their negative externalities. The analysts rely on sector-level opinions and scores produced by an external provider.

Finally, the overall risk score for this module is reviewed by the analysts. Similarly to the other modules, the analysts may override the overall score, providing the reason for the adjustment in a dedicated comment.

### 4.2.6 Third-party opinions

This module requires that analysts consider three data sources, if available:

- the firm analysis carried out by other financial institutions;
- the ratings assigned to the company by rating agencies;
- the company score produced by an independent credit assessment source.

When assessing third-party opinions, analysts are required to map each third-party rating or score into the equivalent ICAS rating. To reconcile possible differences across external ratings, the analysts are guided through a list of questions for each external rating source. Based on the answers, a risk score is computed.[26] The analysts can override the automatic score, explaining the reason for the deviation in a dedicated comment.

### 4.2.7 Climate change risks

The assessment of climate change risks covers transition risk and physical risk.[27] The analysts assess both categories and integrate them into a single score. Given the unique nature of these risks, which are not captured by the statistical model, in this module analysts assign a score that reflects the firm's exposure to climate risks using a best-in-class/best-in-universe approach.

[26] This choice is related to the fact that IRBs and ECAIs are eligible rating sources for the ECAF.

[27] Climate change and related policies generate risks through two main channels:
i) physical risk, stemming from the disruption of economic activity caused by increasingly frequent and severe extreme weather events, which may impair borrowers' ability to meet their financial obligations;
ii) transition risk, arising from the socioeconomic responses to climate change and the potential adverse effects of climate policies aimed at reducing greenhouse gas (GHG) emissions—particularly on carbon-intensive industries. The transition may also generate opportunities for firms, potentially leading to a positive impact on their credit risk profile.

The analysis of transition risk has two stages, in line with the ECB's minimum requirements.[28] First, we estimate the impact of a carbon tax on the firm's creditworthiness (see Di Virgilio *et al.*, 2024). The 12-months' PD of the firm is computed under different NGFS scenarios. The resulting transition risk score is compared with those of external data providers. The qualitative approach for transition risk involves a checklist including items such as best-in-class/universe firm behaviour, governance sensitivity, transition plans. The analysts answer qualitative questions (e.g., on internal governance, commitment of corporate bodies, insurance coverage, investments, and damages already suffered from physical risks) using information collected from NFR and financial statements.

The quantitative methodology for the evaluation of physical risk employs a benchmark physical risk score provided by an external provider. Hazards such as landslides and floods are captured through a synthetic indicator on a five-point scale. The resulting physical risk score considers the location of the company headquarters and operating units, but it does not account for mitigating actions (such as insurance or physical devices) undertaken by the company to reduce potential risks. In the subsequent qualitative assessment, the analysts incorporate possible mitigation measures if the information is available, for instance, from the non-financial statement. The assessment is supported by a questionnaire about recent disaster events and insurance coverage.[29]

An internal model is later used to integrate the transition and the physical risk scores. The final score contributes to the stand-alone rating.

### 4.2.8 Overall assessment and stand-alone full rating

The integration of the expert assessment into the statistical rating is performed using Bayesian statistical techniques. We update the statistical rating with new information, given by the scores provided by the various modules. The reformulation of the PD estimate can be seen as an extended version of the standard logistic model. Formally, the logistic model for the PD is defined by the following sigmoidal function:

$$PD = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots)}}$$

The extended model incorporates the original logistic model and the module scores which capture the judgment of the analysts. The extended model implements a Bayesian approach using numerical

[28] See *Common minimum standards for incorporating climate change risks into in-house credit assessment systems in the Eurosystem*, European Central Bank, Economic Bulletin Issue 6, 2022, box 6.

[29] Additionally, analysts can leverage individual information from a survey (launched in spring 2024 and planned to be repeated annually) involving a sample of Italian non-financial firms. The objective of this survey is to gather climate change risk information at an individual firm level from entities not subject to non-financial reporting requirements.

techniques such as Markov Chain Monte Carlo (MCMC), which is standard practice for this type of application. In our extended model, we use the formula:

$$PD = \frac{1}{1 + e^{-(Score + \alpha_1 z_1 + \alpha_2 z_2 + \cdots)}}$$

where *Score* is the result of the original logistic model, and $z_1$, $z_2$, ... are the module scores.

We note that the above is a simplified representation of the methodology, for illustrative purposes. While the traditional logistic model follows a frequentist approach, the extended model adopts a Bayesian approach, whereby the PD is not a point estimate, but it is obtained as the median of a probability distribution. Unlike the statistical PD, the Bayesian model strictly requires numerical methods such as MCMC.

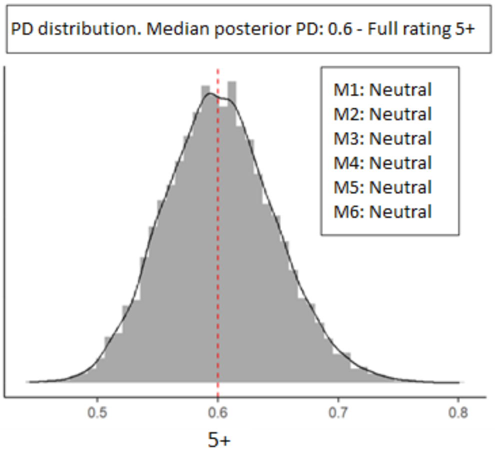
In frequentist statistics, probability is interpreted as the long-term frequency of an event in repeated samples. In contrast, the Bayesian view sees probability as a measure of belief or credibility which can be updated as new information becomes available using Bayes' theorem. This approach allows for greater flexibility and the inclusion of a priori knowledge in the model. The Bayesian approach is particularly suitable when working with small samples and allows for the integration of analysts' assessments in a flexible and yet rigorous manner. This enhances both the accuracy and the value of expert assessment. While the model requires somewhat complex computational techniques, its output is rather intuitive and easy to interpret. It also provides an objective benchmark for full ratings, against which analysts can compare their own assessment. The internal rules foresee that the analysts retain the option to override the resulting rating under certain conditions. They may lower the rating without restrictions, but they can raise it by one notch only. If they intend to raise it by more than one notch, they must submit the proposal for review to the Rating committee.

In the following example, we illustrate how module scores affect the posterior distribution of PD estimates to yield the full rating. The process begins with the statistical rating. Let's assume that our representative firm has a statistical rating equal to 5+ on the Banca d'Italia's scale. A neutral expert assessment would confirm this estimate, while the integration of positive or negative module scores could produce a different full rating (Figure 4).

**Figure 4 - Posterior of PD distribution according to module scores - Examples**

*(statistical PD: 0.6 percent; statistical rating: 5+)*

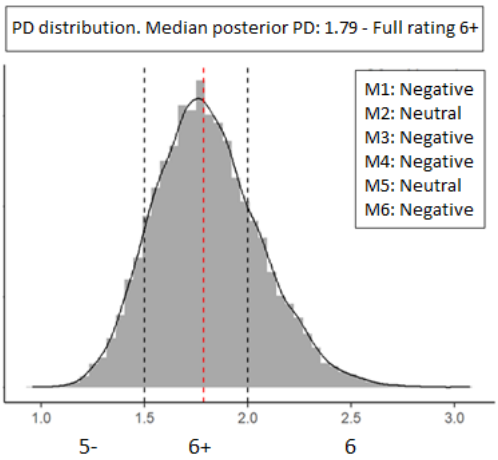
**Panel A - All module scores are neutral**




The median of the posterior PD distribution coincides with the statistical PD statistical rating; besides posterior PDs are concentrated in the interval [0.5–0.7] corresponding to the full rating 5+

**The model yields the full rating 5+**

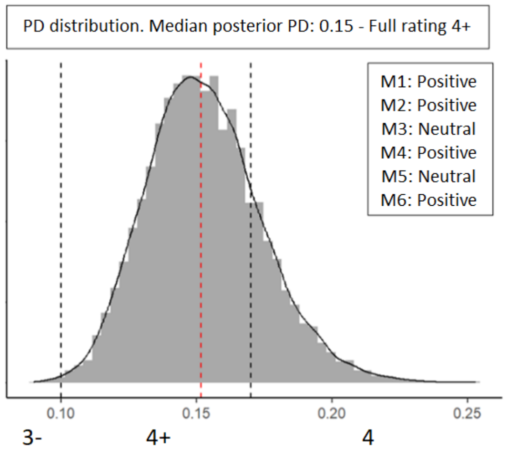
**Panel B - Prevalence of negative scores**




The median of the posterior PD distribution falls in the interval of PDs corresponding to the full rating 6+, worse than the statistical rating. Most posterior PDs fall in the interval corresponding to 6+ [1.5–2.0] but a non-negligible percentage of posterior PDs falls in the interval corresponding to the full rating 6 [2.0-3.0]

**The model yields the full rating 6+, but also 6 is plausible**

**Panel C - Prevalence of positive scores**




The median of the posterior PD distribution falls in the interval of PDs corresponding to the full rating 4+, better than the statistical rating (+3 notches). Most posterior PDs fall in the interval corresponding to 4+ [0.1–0.17], but a non-negligible percentage of posterior PDs falls in the interval corresponding to the full rating 4 [0.17-0.3]

**The model yields the full rating 4+, but also 4 is plausible**

### 4.2.9 Group analysis

For subsidiaries, group analysis is conducted once the stand-alone full rating is assigned. The methodology foresees that the higher is the degree of integration between the company and its group, the greater is the weight assigned to the parent company in the full rating of the subsidiary. To support this assessment, analysts are provided with a set of indicators and information on group entities. They are expected to identify the appropriate parent company, assess the strength of the interlinkages, and evaluate the parent's creditworthiness, drawing on the ratings available within ICAS or from external sources. Group analysis is divided into three steps. In the first one, companies are sorted. The first analyst is required to identify the parent company.[30] If the company under assessment does not belong to a group or is a parent company at the highest level of consolidation, the final assessment coincides with the stand-alone assessment of the company and the process stops.

In the second step, the first and second analyst independently fill out a checklist on the interlinkages between the subsidiary and its parent company. Responses are classified into three levels of strength: weak, medium, and strong. The questions pertain to three dimensions: i) governance and corporate integration; ii) economic and operational integration; and iii) financial integration. Each dimension is equally weighted to obtain the overall intensity of the interlinkages. The analysts may apply a motivated override to the resulting level of intensity; for instance, if the subsidiary is expected to be divested in the near term, a lower degree of integration is typically assumed. The subsequent step consists in the credit risk assessment of the parent company. The first analyst verifies whether a full rating for the parent company is available, based on financial data from the same year as that of the individual entity. If the parent company is a foreign entity, the analyst checks for an ICAS rating issued by other National Central

[30] The second analyst must explicitly validate the choice made by the first analyst.

Banks. In the absence of such ratings, the analysts provide a qualitative assessment of the parent's financial position relative to the individual entity, on a five-point scale (from 'very positive' to 'very negative'), drawing on all available information, including third-party assessments (e.g., external ratings, IRB PDs) and the statistical ICAS rating, if available.

In the third step, based on previous information, the system generates a rating proposal. If a full rating for the parent company is available, a weighted average is computed between the individual and parent company ratings, where the weights reflect the interlinkage intensity. In the absence of a rating for the parent company, the system applies an adjustment (notching upward or downward) using a transition matrix that incorporates both the parent's creditworthiness and the interlinkage factor. In all cases, the proposed rating can be overridden by either analyst. Generally, when a strong link is identified between an entity and its group, the subsidiary's rating cannot exceed that of the parent company. If the subsidiary's rating exceeds that of the parent company, the former rating must be submitted to the Rating committee for approval.

## 5. Validation

### 5.1 General principles

The validation of the expert assessment is based on the dataset covering the period from 2016 to 2022, with over 25,000 records.[31] The validation exercise draws on interviews with staff from the Banca d'Italia's division responsible for managing the ICAS. The dataset comprises the expert assessment output grouped by analyst, firm and date. For each observation, the statistical rating, the full rating, the module scores assigned by the analysts and any default event in the 12 months following the rating assignment are recorded. The exercise addresses the following questions:

- does the expert assessment provide incremental value relative to the statistical model? If so, is this value statistically significant?
- what is the relationship between each module of the analysis and the full rating? Specifically, how does each module contribute to the explanatory power and performance of the full rating?
- are there significant differences across analysts in deriving the full rating? If that is the case, do these differences produce a different performance of the full rating?

We perform backtesting on both the statistical rating and the full rating, evaluating their discriminatory power, predictive power and performance stability. We adopt the following guiding principles:

[31] The default state for each assessment is binomial; we chose this type of default to have a larger dataset (evidence of fractional defaults is only available from January 1, 2019).

i) discriminatory power: we favour rating systems that exhibit higher values of the AUROC;[32]

ii) predictive power: we prefer conservative default forecasts; hence we favour rating systems whose forecasted default rate closely aligns with the observed default frequency;

iii) stability: we value time consistency in both discriminatory and predictive power; we therefore prefer rating systems that exhibit stationary behaviour over time.

Besides, the validation process seeks to ensure broad accessibility and interpretability of its results**.** Therefore, we follow the testing procedures recommended by the ECAF.

We evaluate whether any changes in discriminatory power, predictive power and stability arising from the transition from the statistical rating to the full rating are statistically significant. Concerning the predictive power, we apply a rule of the ECAF framework: two rating systems are considered as significantly different if the difference is due to one or more rating classes.

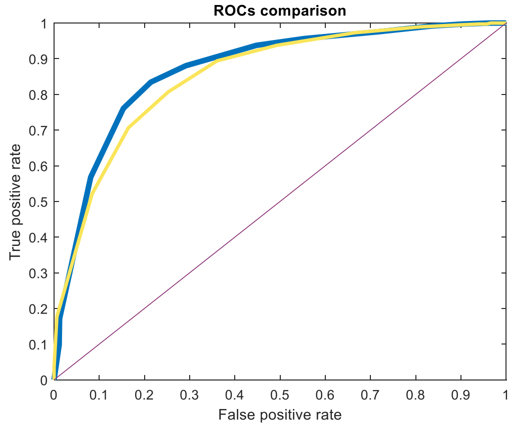
**5.2 Performance**

We present the results concerning the overall performance of the expert assessment, without disaggregating the specific contributions of individual components to the full rating (see the following sections for a detailed analysis). We examine the alignment between full ratings and observed defaults, evaluating the discriminatory power and the predictive power of ratings and the stability of their performance. We estimate the incremental value added by the expert assessment over the statistical ratings and assess its significance.

The results indicate that the expert assessment improves the discriminatory power of ICAS. The AUROC increases from 85.3 percent for the statistical rating to 87 percent for the full rating (Figure 5).

[32] The Receiver Operating Characteristic (ROC) curve is a plot that visualizes the performance of a binary classification model, showing the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) at different decision thresholds. It shows how the model performs when the level of 'caution' in predicting the defaults varies. The area under this plot is the AUROC (Engelmann and Rauhmeier, 2011).

**Figure 5 – ROC curves for statistical and full rating**
*(2016-2022)*




The yellow line refers to the statistical rating, the blue line to the full rating.

The ROC curve for the full rating lies strictly above that for the statistical rating across the entire range of thresholds. This implies that, in terms of discriminatory power, the full rating strictly dominates the statistical rating. To evaluate the significance of this improvement, we apply a bootstrapping procedure to estimate the distribution of AUROC differences simulated between the two types of rating for each firm. The observed difference falls in the tail of the empirical distribution, outside of the 95 percent confidence interval, indicating statistical significance. These results are robust to dataset stratification, since similar conclusions hold when the analysis is repeated by firm size and year (Tables 1-2).

**Table 1 – AUROC, by firm size**
*(2016-2022, percentage values)*

| | Statistical rating | Full rating |
|---|---|---|
| **Micro** | 71.7 | 79 |
| **Small** | 77.1 | 84.3 |
| **Average** | 86.4 | 87 |
| **Large** | 87.3 | 89.3 |

**Table 2 – AUROC, by year**
*(percentage values)*

| | Statistical rating | Full rating |
|---|---|---|
| **2016** | 82.1 | 86.9 |
| **2017** | 82.4 | 83 |
| **2018** | 89.7 | 89 |
| **2019** | 85.6 | 87.9 |
| **2020** | 87.2 | 91.2 |
| **2021** | 86.6 | 91.2 |
| **2022** | 84.2 | 85.9 |

A bootstrapping exercise reveals that some of the differences between the statistical rating and the full rating reported in the above tables are statistically significant, while others are not.[33] We conclude that the full rating consistently achieves higher AUROC values than the statistical rating across all firm size categories in nearly all years in the sample. These findings suggest that the expert assessment improves the discriminatory power and is stable over time.

Secondly, the expert assessment enhances the predictive power of ICAS. The full rating tends to be more conservative than the statistical rating, often resulting in lower ratings (Table 3). While this may reduce precision in matching actual defaults, it increases prudence. In practice, the analysts frequently assign lower ratings than those implied by the statistical model, reflecting a cautious stance. This behaviour is corroborated by additional statistical tests, including the Redelmeier test.[34]

Although the full rating is less precise in a narrow predictive sense, it shows greater robustness to year-to-year fluctuations in default rates. The statistical rating does not always forecast deteriorating credit conditions, whereas the full rating tends to overestimate defaults. This conservative bias enables the expert assessment to act as a safeguard against unexpected defaults, thereby contributing to a more stable and resilient predictive performance over time.

**Table 3 - Multiple test for statistical rating and full rating**
*(2016 - 2022)*

| **Statistical rating** | | | | | |
|---|---|---|---|---|---|
| **CQS** | **Members number** | **Lower bound** | **Expected defaults** | **Upper bound** | **Defaults** |
| CQS1&2 | 1489 | 0 | 1 | 5 | 1 |
| CQS3 | 6933 | 7 | 19 | 34 | 8 |
| CQS4 | 7673 | 34 | 56 | 81 | 29 |
| CQS5 | 2734 | 18 | 34 | 54 | 32 |
| CQS6 | 4325 | 63 | 91 | 122 | 105 |
| CQS7 | 2090 | 57 | 84 | 113 | 127 |
| CQS8 | 221 | 24 | 41 | 59 | 64 |
| **Full rating** | | | | | |
| **CQS** | **Members number** | **Lower bound** | **Expected defaults** | **Upper bound** | **Defaults** |
| CQS1&2 | 937 | 0 | 1 | 5 | 0 |
| CQS3 | 6145 | 5 | 16 | 30 | 9 |
| CQS4 | 6809 | 29 | 49 | 72 | 13 |
| CQS5 | 3904 | 28 | 49 | 72 | 22 |
| CQS6 | 3561 | 49 | 74 | 102 | 47 |
| CQS7 | 1868 | 50 | 75 | 103 | 71 |
| CQS8 | 2241 | 454 | 516 | 580 | 204 |

Note: the table above refers to the assignment to individual debtors of the full ratings produced by ICAS, according to which the debtors are allocated to the different CQS. For each CQS we apply a binomial test with 2 tails centered on the average rating. The aim of the test is to verify if the expected defaults are different from the observed defaults. The test confidence level is 0.95. The numbers in red signal that the number of the observed defaults is greater than the number of the expected ones.[35]

[33] The subsample analysis by firm size and year is based on relatively small samples; as such the corresponding results should be interpreted with caution due to limited statistical robustness.
[34] For technical details, see Appendix 3.
[35] For further details, see Narizzano *et al.* (2024), pag. 44 - 46.

### 5.3 Behaviour of the analysts[36]

In this section we examine how the community of analysts derives the full rating from the module scores and how this rating performs in relation to observed defaults. We thus view the performance of the analysts without distinction across the individuals. We proceed as follows:

- we estimate a regression model to identify how the analysts combine the module scores, taking note of systematic deviations from the general guidelines;
- we assess the effectiveness of the score aggregation in predicting defaults, in terms of discriminatory power and predictive power.

On average we find that:

- the analysts combine the module scores in a rather discretionary fashion;
- certain modules significantly contribute to the performance of ICAS, others do not.

#### 5.3.1 Deviations from the guidelines

We run the following regression, for each firm and year (the suffixes are suppressed for simplicity):

$$ln\,\left(\frac{PD_{full}}{PD_{statistical}}\right) = \beta_0 + \beta_{FS} * AST_{FS} + \beta_{FF} * AST_{FF} + \beta_{GV} * AST_{GV} + \beta_{SA} * AST_{SA} + \beta_{GR} * AST_{GR} + \beta_{3P} * AST_{3P} + \eta\ ,$$

where:

- the dependent variable is the natural logarithm of the ratio between the full rating, assigned by the analysts for each firm and year, and the statistical rating;
- the independent variables are the module scores assigned by the analyst.

The regression results[37] reveal that the analysts' weighting of the modules partially diverges from the guidelines. Specifically:

- the financial statement score receives the highest weight in the transition to the full rating;
- the financial flexibility score and the third-party opinions scores are roughly equivalent;
- the remaining modules appear to have a limited impact on the full rating.

#### 5.3.2 Performance of the analysts

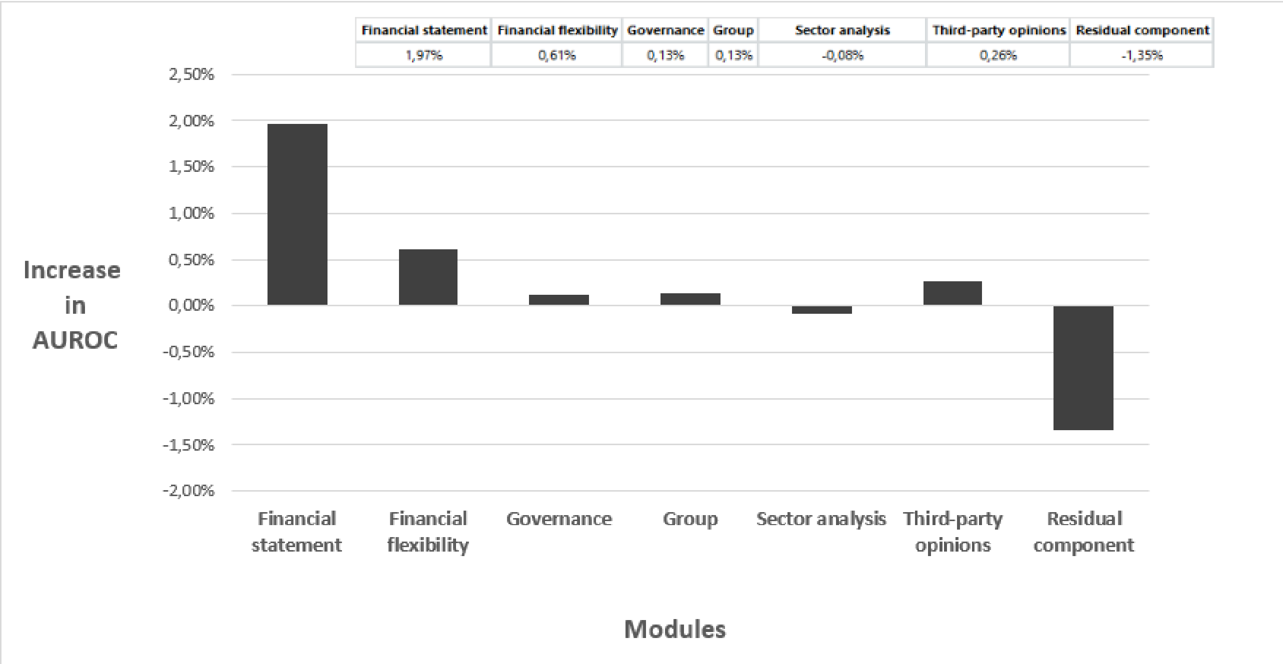
Starting from the regression results, we reconstruct the full ratings by sequentially adding the contributions of the constant term and each module score to the baseline statistical rating.

[36] In this section we examine the first analyst's behaviour.
[37] For technical details, see Appendix 4.

At each step, we obtain a distinct set of marginal contributions of each module to the full rating. We then compare the contributions against observed defaults, which enables us to measure the corresponding discriminatory power and predictive power.

Figure 6 illustrates the breakdown of the total discriminatory power attributable to the full ratings:

**Figure 6 - Contribution to discriminatory power**




Note: the x-axis reports the names of the modules, the y-axis shows the increase in discriminatory power.

We test whether each module provides a statistically significant increase in discriminatory power, and whether the module contributions are significantly different from each other.[38]

Regarding the discriminatory power, we find that:

- the financial statement module provides the largest marginal contribution, followed by the financial flexibility and the third-party opinions modules;
- these three modules are the only ones that make a positive and significant contribution;
- the sector analysis module, while not statistically significant, exhibits a negative marginal contribution;

[38] These significance analyses were conducted using appropriate bootstrap procedures, identifying the threshold quantiles at a 95 percent confidence level.

- a residual component, attributable to other independent variables[39] or to individual choices made by the analysts, shows a negative contribution that turns out to be statistically significant.

Building on these findings, we decompose the overall predictive power of the full ratings, isolating the marginal contribution of each module to the system's forecasting accuracy.[40] The main results are as follows:

- the transition from the statistical rating to the full rating results in a more conservative assessment of credit risk;
- none of the six modules, considered in isolation, significantly enhances the predictive power of the rating system;
- the inclusion of the constant term ($\beta_0$) in the regression introduces a shift in the rating level, changing the assessment from a relatively lenient to a conservative stance. In other words, even in the absence of specific signals from the modules, the analysts tend to worsen the statistical rating, showing a precautionary bias.

### 5.4 Heterogeneity in the analysts' behaviour

We extend[41] the previous analysis to investigate the following questions:

- whether the individual analyst behaviour significantly deviates from that of the group;
- whether any such deviations result in a statistically significant improvement in rating performance.

We analyze some partitions of the original dataset, each corresponding to a different analyst. For each sub-dataset we estimate a regression model similar to the one described in the previous section. These results reveal substantial heterogeneity across analysts. Specifically, we find that:

- the analysts often differ in the selection of modules and their weights;
- some analysts do not employ all the modules in their assessment.

These findings reveal that the analysts employ in practice a broad range of rating strategies. Each strategy features a distinct combination of modules and weighting schemes. We then set out to assess the performance of these strategies in terms of their discriminatory power and predictive power. Concerning the discriminatory power, we find that some strategies yield performance improvements relative to the average analyst, while others lead to performance deterioration. In terms of predictive power, no individual strategy appears to produce significant differences compared to the average. None of the

[39] This is the case for instance of the climate change risk considerations, formerly not considered, and recently introduced in the expert assessment.

[40] For each contribution we run the multiple binomial test (Dohler, 2011; Dudoit *et al.*, 2003) as reported in Table 3. The predictive power of two contributions is compared by examining the results of their respective tests.

[41] For methodological details and results, see Appendix 5.

strategies significantly outperforms the analysts as a group in either discriminatory power or predictive power. We conclude that the heterogeneity in analysts' behaviour does not translate into performance gains.

### 6. Developments: geopolitics and ESG

As shown above,[42] the validation highlights, among other things, that on average analysts used other elements beyond the variables included in the modules. This phenomenon suggests to extend the expert assessment to additional factors with a view to improving its contribution to the performance of the rating system.

Geopolitical considerations are playing an increasing role in credit ratings, as country risk may be affected by rising global tensions. The pandemic and the war in Ukraine have accelerated the reconfiguration of global supply chains, a trend that began in the aftermath of the 2008 financial crisis. The future of a liberalized, rule-based global trade system is uncertain, with the US exemplifying this shift (Mercurio, 2024). International trade with geopolitically aligned partners and trade tariffs are increasing for economic and political reasons. Global supply chains have shown their fragility during the pandemic and afterwards (Shih, 2022). From a political standpoint, the number of new global trade restrictions has increased from approximately 650 in 2017 to over 3,000 in 2023 (MGI, 2024), while 'friendshoring', the alignment of supply chains with the political orientation of the country, is spreading (Yellen, 2022). These developments have led to a closer alignment between states and firms, influencing credit ratings, and triggering rapid demand shifts across sectors, with a move away from export-led strategies (Cherif and Hasanov, 2024). The approaches of CRAs and IRBs must adapt to this volatile environment by identifying which industries are likely to benefit or be adversely affected by the global changes.

The importance of environmental, social and governance (ESG) concerns has grown among the public, governments, and in corporate strategies. The number of companies acknowledging climate change risks in their financial reporting has significantly increased (KPMG, 2020). ESG-related investments have surged in recent years, prompting CRAs to explicitly integrate ESG factors into their rating methodologies (EC, 2022). However, ESG scores are heterogeneous across rating providers.[43] This heterogeneity has fueled the investors' skepticism about the reliability of ESG ratings (Del Giudice *et al.*, 2024). Nonetheless, the importance of integrating ESG factors into corporate credit assessment is increasing, as shown by studies linking carbon neutrality to credit risk (Umar *et al.*, 2021) and green innovation to default risk (Meles *et al.*, 2023). ESG issues may affect corporate risk both directly, through the business

[42] See, in particular, paragraph 5.3.2.
[43] This variability in rating criteria has been empirically shown to result in divergent opinions on the same firms (Billio *et al.*, 2021), with differences in weighting and social and governance indicators being the primary drivers of these divergences (Capizzi *et al.*, 2021).

model, and indirectly, via reputational channels. Numerous studies have shown that sustainable companies tend to exhibit a stronger financial performance.[44] Financial market data corroborate these findings, with ESG funds outperforming their peers (Morgan Stanley, 2024). The rationale is that sustainability becomes profitable when it is embedded in the firm's core business model (Kim and Yoon, 2023).[45]

Due to market demand and the growing body of research linking ESG factors to investment outcomes and financial performance, CRAs nowadays incorporate ESG considerations into their ratings, albeit in different ways. ESG factors can affect credit risk either directly, by influencing a specific rating component such as the governance quality, or indirectly, by reducing profitability or liquidity through additional costs. It is also essential to establish whether the impact stems from one-off events, such as a controversy, or it reflects a structural, sector-wide shift that may alter long-term risk.

CRAs assess whether climate-related physical and transition risks are significant for the firm. The most frequent approach is to integrate ESG factors into the standard rating procedure, treating them as a component of the overall assessment, despite the often-limited quality of ESG data. Given the heterogeneity of the indicators, CRAs typically develop E, S and G scores separately. The IRB approach to ESG is similar to CRAs'. Since IRB models are used for regulatory purposes related to capital requirements, it seems likely that, should ESG-related requirements materialize, banks would align their own ESG ratings more closely (EBA, 2022).

In line with this trend, ICAS has recently been tasked with the integration of ESG factors into its credit assessment process.

## 7. Conclusions

This study shows the contribution of the expert assessment in the Banca d'Italia's ICAS system towards the predictive power and discriminatory power of credit ratings. By integrating qualitative considerations within a robust statistical framework, the expert assessment improves the PD estimates and serves as a safeguard against the inherent limitations of statistical models, particularly during periods of rising economic uncertainty.

[44] Notably, a comprehensive study by Whelan *et al.* (2021) reviewed over 1,000 research papers from 2015 to 2020, finding a positive relationship between ESG and financial performance in 58 per cent of the corporate studies, with only 8 per cent showing a negative relationship.

[45] Companies that merely declare sustainable intentions, such as signing the Principles for Responsible Investment, do not show distinguishable performance improvements, despite widely advertising their affiliation (Kim and Yoon, 2023).

The validation results confirm that the expert assessment enhances the accuracy and the conservatism of credit ratings. Notably, the AUROC improves significantly. Furthermore, the adjustments introduced by the analysts ensure that the ICAS framework responds to evolving economic conditions.

Our findings underscore the importance of an adequate structure for the expert assessment process. The recent introduction of detailed checklists and Bayesian aggregation techniques has significantly enhanced transparency, homogeneity, and accountability. These advancements mitigate any subjective bias and allow the ICAS system to address complex, forward-looking factors such as climate risks and sector dynamics.

Our work suggests several avenues for future development. The expert assessment framework lends itself to further enhancements, notably in the areas of ESG integration and geopolitical risk, and to the use of new data sources, to ensure that the system remains at the forefront of the credit risk assessment practice.

In conclusion, our study reaffirms the importance of the interplay between statistical rigour and expert insight in corporate credit analysis. By bridging these dimensions, the ICAS system exemplifies a comprehensive and resilient approach, providing valuable insights for risk management, financial institutions, and researchers. It is important to keep the rigorous standards of the system as well as its ability to incorporate new developments in credit risk analysis, to integrate emerging risk factors, such as climate risks and geopolitical risk, and new estimation techniques, in particular artificial intelligence tools.

## References


Agarwal S., Ambrose B.W., Liu C., 2006. Credit Lines and Credit Utilization, *Journal of Money, Credit and Banking*, 38, 1-22.

Alp A., 2013. Structural Shifts in Credit Rating Standards, *Journal of Finance*, 68(6), 2435-2470.

Altman E.I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, 23, 589-609.

Amato J.D., Furfine C.H., 2003. Are Credit Ratings Procyclical. Bank for International Settlements Working Paper, 129.

Anderson R.C., Mansi S.A., Reeb D.M., 2004. Board characteristics, accounting report integrity, and the cost of debt. *Journal of Accounting and Economics*, 37, 315-342.

Auria L., Bingmer M., Mateo C., Graciano C., Charavel C., Gavilá S., Iannamorelli A., Levy A., Maldonado A., Resch F., Rossi A.M., Sauer S., 2021. Overview of Central Banks' In-house Credit Assessment Systems in the Euro area, *Occasional Paper Series*, European Central Bank, 284, Frankfurt.

Ballis A., Ioannidis C., Sifodaskalikis E., 2024. Structural Shifts in Bank Credit Ratings, *Journal of Financial Stability*, 73.

BCBS. 2001. The Internal Ratings-Based Approach, Bank for International Settlements, Basel. BCBS

BCBS. 2015. Developments in Credit Risk Management Across Sectors: Current Practices and Recommendations, Bank for International Settlements, Basel. BCBS

BCBS. 2024. Guidelines for Counterparty Credit Risk Management, Bank for International Settlements, Basel. BCBS

Becker B., Milbourn T., 2011. How did increase competition affect credit ratings?, *Journal of Financial Economics*, 108, 493-514.

Billio M., Costola M., Hristova I., Latino C., Pelizzon L., 2021. Inside the ESG Ratings: (Dis)agreement and Performance, *Corporate Social Responsibility and Environmental Management*, 28, 1426–1445. Billio *et al.*

BIS. 2015. Developments in credit risk management across sectors: current practices and recommendations. BIS

Bloom N., Van Reenen J., 2007. Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, 122, 1351–1408.

Capizzi, V., Gioia, E., Giudici, G., Tenca, F., 2021. The Divergence of Esg Ratings: An Analysis of Italian Listed Companies. *Journal of Financial Management, Markets and Institutions*, 9, 2150006. Capizzi *et al.*

Cherif R., Hasanov F., 2024. The Pitfalls of Protectionism: Import Substitution vs. Export- Oriented Industrial Policy, *IMF Working Paper,* 24/86, International Monetary Fund, Washington.

deHaan D., 2017. The Financial Crisis and Corporate Credit Ratings, *The Accounting Review*, 4, 161-189.

Del Giudice, A., Gallucci, C., Santulli, R., 2024. I rating ESG: un confronto internazionale, Università Cattolica del Sacro Cuore, Milano.

Diamond D.W., 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3), 393–414.

Di Virgilio S., Faiella I., Mistretta A., Narizzano S., 2024. Assessing Credit Risk Sensitivity to Climate and Energy Shocks: Towards a Common Minimum Standards in Line with the ECB Climate Agenda, *Journal of Policy Modeling*, 46(3): 552-568.

Dohler, S., 2011. Validation of Credit Default Probabilities Using Multiple-Testing Procedures, *The Journal of Risk Model Validation*, 4(4): 59-92.

Dudoit S., Ge Y., Speed T. P., 2003. Resampling-based multiple testing for microarray data analysis, *Sociedad de Estadistica e Investigacion Operativa*, *Test,* 12(1): 1–77.

Engelmann, B., Rauhmeier, R., 2011. The Basel III Risk Parameters, Estimation, Validation, Stress Testing - with Applications to Loan Risk Management, *Springer*, Berlin.

EBA. 2017. Guidelines on PD Estimation, LGD Estimation and Defaulted Exposures, European Banking Authority, Paris. EBA

EBA. 2022. The Role of Environmental Risks in the Prudential Framework, European Banking Authority *Discussion Paper* no. 2022/02, Paris.

EC. 2022. Targeted Consultation on the Functioning of the ESG Ratings Market in the European Union and on the Consideration of ESG Factors in Credit Ratings, *Summary Report*, European Commission, Brussels.

Gallo, R., 2021. The IRB Approach and Bank Lending to Firms, *Temi di discussione,* 1347, Banca d'Italia, Rome.

Gallucci C., Santulli R., Modina M., Formisano F., 2022. Financial Ratios, Corporate Governance and Bank-firm Information: a Bayesian Approach to Predict SMEs' Default, *Journal of Management and Governance*. Gallucci *et al.*

Garcia Alcubilla, R., Ruiz del Pozo, J., 2012. Credit Rating Agencies on the Watch List: Analysis of European Regulation, *Oxford University Press*, Oxford.

Giannozzi A., Altman E.I., Roggi O., Sabato G., 2013. Building SME Rating: is it Necessary for Lenders to Monitor Financial Statements of the Borrowers?, *Bancaria Editrice*, 10, 54-71.

Giovannelli F., Iannamorelli A., Levy A., Orlandi M., 2020. The In-House Credit Assessment System of Banca d'Italia, Occasional Papers, 586, Banca d'Italia, Rome.

Jensen, M. C., Meckling, W. H., 1976. Theory of the firm: Managerial behaviour, agency costs and ownership structure. Journal of Financial Economics, 3, 305–360.

Jiménez G., Lopez J., Saurina J., 2009. Empirical Analysis of Corporate Credit Lines, *The Review of Financial Studies*, 22, 5069-5098.

Kim S., Yoon A.S., 2023. Analyzing Active Fund Managers' Commitment to ESG: Evidence from the United Nations Principles for Responsible Investment", *Management Science*, INFORMS, 69, 741-758.

KPMG. 2020. The Time has Come: The KPMG Survey of Sustainability Reporting, Amsterdam. KPMG

Meles A., Salerno D., Sampagnaro G., Verdoliva V., Zhang J., 2023. The Influence of Green Innovation on Default Risk: Evidence from Europe, *International Review of Economic Finance*, 84, 692–710. *Meles et al.*

Mercurio, B., 2024. The Demise of Globalization and Rise of Industrial Policy: Caveat Emptor, *World Trade Review*, 23, 242–250.

MGI. 2024. Geopolitics and the Geometry of Global Trade, McKinsey Global Institute, New York.

Modigliani F., Miller M., 1958. The Cost of Capital, Corporation Finance and the Theory of Investment, *American Economic Review*, 48, 261-297.

Morgan Stanley. 2024. Sustainable Funds Outperformed Peers in 2023, *Institute for Sustainable Investment*, February 29. Morgan Stanley

Narizzano S., Orlandi M., Scalia A., 2024. The Banca d'Italia's statistical model for the credit assessment of non-financial firms, *Collana MISP,* 53, Banca d'Italia, Rome.

Norden L., Weber M., 2010. Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers, *The Review of Financial Studies*, 23, 3665-3699.

Redelmeier, D. A., Bloch, D. A., Hickman D. H., 1991. Assessing Predictive Accuracy: How to Compare Brier Score, *Journal of Clinical Epidemiology*, 44, 1141–1146.

Singhal H., Verma A., Chakraborty M., 2024. The Quality of credit Ratings amid Geopolitical Risk, *Economic Letters*, 234.

EU. 2009. REGULATION (EC) No 1060/2009 of the European Parliament and of the Council on Credit Rating Agencies, European Union, Bruxelles.

Shih W.C., 2022. Are the Risks of Global Supply Chains Starting to Outweigh the Rewards?, Harvard Business Review, March 21. Shih

Shleifer A., Vishny R.W., 1997. A survey of corporate governance. *Journal of Finance*, 52, 737–783.

Sufi A., 2009. Bank Lines of Credit in Corporate Finance: An Empirical Analysis, *The Review of Financial Studies*, 22, 1057-1088

Umar M., Ji X., Mirza N., Naqvi B., 2021. Carbon Neutrality, Bank Lending, and Credit Risk: Evidence from the Eurozone. *Journal of Environmental Management*, 296.

Verbeek, M., 2010. Econometria, Zanichelli, Bologna.

Whelan T., Atz U., Clark C., 2021. ESG and Financial Performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 – 2020, *NYU STERN and Rockfeller Asset Management*.

Yellen, J., 2022. Remarks by Secretary of the Treasury Janet L. Yellen on Way Forward for the Global Economy, Atlantic Council, April 13. Yellen

Zhao Y.J., Dwyer D.W., Zhang J., 2014. Usage and Exposures at Default of Corporate Credit Lines: an Empirical Study, *Journal of Credit Risk*, 10.

**Appendix 1. The rating scale adopted by the Banca d'Italia's ICAS**

**Table A1 – Rating scale**
*(percentage values)*

| Banca d'Italia Full ratings | Minimum PD | Maximum PD | Eurosystem Credit Quality Step |
|---|---|---|---|
| **1** | 0.000 | 0.001 | CQS 1 & 2 |
| **2+** | 0.001 | 0.01 | |
| **2** | 0.01 | 0.03 | |
| **2-** | 0.03 | 0.05 | |
| **3+** | 0.05 | 0.07 | |
| **3** | 0.07 | 0.09 | |
| **3-** | 0.09 | 0.10 | |
| **4+** | 0.10 | 0.17 | CQS 3 |
| **4** | 0.17 | 0.30 | |
| **4-** | 0.30 | 0.40 | |
| **5+** | 0.40 | 0.80 | CQS 4 |
| **5** | 0.80 | 1 | |
| **5-** | 1 | 1.50 | CQS 5 |
| **6+** | 1.5 | 2 | CQS 6 |
| **6** | 2 | 3 | |
| **6-** | 3 | 5 | CQS 7 |
| **7** | 5 | 25 | CQS8 |
| **8** | 25 | 100 | |
| **9** | 100 | 100 | Default |

## Appendix 2. Initial adjustment of the statistical rating

The analysts may adjust the statistical rating obtained from the financial component for the following considerations.

- Finance Leases: According to Italian GAAP, finance leases are usually recognized among off-balance sheet items. Italian accounting standards emphasize form, leading to the recognition of lease payments as operating costs over the term of the lease agreement. Italian GAAP stipulates that in a finance lease, involving the transfer to the lessee of most risks and rewards of asset ownership, the company should report in the financial statement notes a table detailing the implicit debt of the lease contract, the total amount of assets leased, yearly amortization, and interest expenses. Analysts must extract information from these notes and adjust for leasing according to IFRS methodology (known as the 'financial method). Under this method, companies recognize finance leases as assets and liabilities on the balance sheet, valued at their fair value and implicit financial debt, respectively. Annual lease payments are divided between interest expense and asset amortization on the income statement.
- Cash and Cash Equivalents: For conservatism, the statistical model for calculating net financial debt considers only cash, cheques, and bank and postal deposits as liquidity. Other financial instruments are classified as 'current financial assets'. Analysts must review the notes to the financial statement to include highly liquid instruments that are readily convertible to cash and subject to low risk of value change (e.g., sovereign bonds or time deposits) within liquidity. Conversely, companies may have cash or equivalents unavailable for general purposes (e.g., restricted cash). To ensure the statistical model accounts for the reduced amount of available cash to repay debt, analysts must exclude restricted cash from 'cash and cash equivalents'.
- Non-recurring Items in the Income Statement: The income statement information used by the statistical model sometimes includes positive or negative transactions that could be considered as 'non-recurring.' These non-recurring items should be reclassified among extraordinary items and excluded from operating income, as they are neither part of ongoing core operations nor an accurate reflection of future performance.
- Dividends: Dividends are proposed by the company's Board of Directors and approved at the general shareholders' meeting. The statistical PD typically reflects information as of the reporting date and does not incorporate declared dividends. Analysts are required to adjust for declared dividends by reducing equity and increasing short-term liabilities accordingly.
- Interest Expenses: Interest expenses are crucial variables for the statistical model, as they are used in predictive variables (e.g., interest expenses to EBITDA or return on debt). Sometimes, companies include items not related to financial debt within interest expenses (e.g., unwinding of the discount of certain assets or liabilities, foreign exchange gains and losses). This reporting

distorts the statistical ratios based on the relationship between interest expenses and financial debt. To address this issue, analysts must reclassify the portion of interest expenses not coherent with financial debt as operating expenses, preserving interest coverage ratios and interest rate risk ratios.

- Hybrid Shareholders' Loans: Some instruments possess characteristics of both equity and debt. For subordinated interest-free shareholders' loans without a specified maturity date, analysts must follow different reclassifications based on the type of company. For predominantly mutual cooperative corporations,[46] these loans must be reclassified as equity if they are considered financial debt by the statistical model. For other companies, these instruments should not be considered financial debt for calculating the ratios used by the statistical model (e.g., equity to net financial debt ratio or return on debt).

The analysts may adjust the statistical rating obtained from the credit behaviour component for the following considerations.

- Average Utilization Rates: These variables are defined as the ratio between the drawn amount and granted amount for each type of credit line (e.g., current account revolving credit lines, accounts receivable revolving credit lines, total short-term credit lines). When these ratios are not available due to the absence of short-term loans in the NCR, the model sets the ratio to 0 (best possible value) and introduces a specific dummy variable that adjusts the credit score for the average riskiness of the subset of firms within the estimation sample without this type of loan reported in the NCR. In specific cases, when analysts observe that one or more ratios are computed incorrectly due to a specific form of credit line (e.g. credit card), they must rectify the ratios used by the statistical model.
- Financial Distress: The statistical model employs variables such as the number of months, or absolute and relative amounts of overdrafts, for each type of credit line at various frequencies, as well as past delinquencies and default status. In this context, analysts must check for erroneous reporting or technical overdrafts and rectify them in the statistical model.

[46] According to Article 2511 of the Italian Civil Code, 'cooperatives' are companies with variable capital with a mutual purpose registered in the register of cooperative companies. Cooperative societies differ from other corporations because they are aimed at providing goods or services directly to the members, on more profitable terms than those offered by the market. They may be predominantly or not predominantly mutual, depending on whether there are conditions in the company's statute limiting the distribution of profits to members, and the activity is carried out primarily for the benefit of the latter.

**Appendix 3. Redelmeier test**

The Redelmeier test (Redelmeier *et al.*, 1991) verifies which one, between two rating systems, has the best predictive power. The test involves a comparison of the mean squared errors (MSE) computed by the two rating systems using the same data sample. The test checks whether the deviation of the realized MSE from its expected value, computed for the first rating system, is significantly different from the deviation of the realized MSE from its expected value, computed for the second rating system. The rating system with the lowest MSE is the best.

The test statistic is:

$$\mathrm{Z_R} = \frac{[\sum_{i=1}^{N}(\hat{\pi}_{\mathrm{i,m1}}^2 - \hat{\pi}_{\mathrm{i,m2}}^2) - 2(\hat{\pi}_{\mathrm{i,m1}}^2 - \hat{\pi}_{\mathrm{i,m2}}^2) * y_i]}{[\sum_{i=1}^{N}[(\hat{\pi}_{\mathrm{i,m1}}^2 - \hat{\pi}_{\mathrm{i,m2}}^2)^2 * (\hat{\pi}_{i,m1} + \hat{\pi}_{i,m2}) * (2 - \hat{\pi}_{i,m1} - \hat{\pi}_{i,m2})]]^{0,5}}$$

(where:

$\hat{\pi}_{i,mn}$ = expected default frequency from the model/rating system of rating *mn* (*n*=1,2) for the firm *i*;

$y_i$= true *default* state for the firm I (*y* = 0 – healthy firm –, 1 – defaulted firm –)

and follows a standard normal distribution under these hypotheses:

$$H_0: E[(E(MSE_{m1}) - MSE_{m1}) - (E(MSE_{m2}) - MSE_{m2})] = 0\ ,$$

$$H_1: E[(E(MSE_{m1}) - MSE_{m1}) - (E(MSE_{m2}) - MSE_{m2})] \neq 0$$

## Appendix 4. Contribution of the modules

We run the following regression:

$$ln\left(\frac{PD_{full}}{PD_{statistical}}\right) = \beta_0 + \beta_{FS} * AST_{FS} + \beta_{FF} * AST_{FF} + \beta_{GV} * AST_{GV} + \beta_{SA} * AST_{SA} + \beta_{GR} * AST_{GR} + \beta_{3P} * AST_{3P} + \eta \ ,$$

where:

- the dependent variable is the natural logarithm of the ratio between the full rating and the statistical rating;
- the independent variables are the scores assigned for each company by the analysts (denoted by AST for assessment) to the modules, identified by the acronyms:
- FS: financial statement
- FF: financial flexibility
- GV: governance
- SA: sector analysis
- GR: group
- 3P: third-party opinions
- $\beta_i$ is the coefficient of the $i_{th}$ module in the regression, including the intercept
- $\eta$ denotes the error term.

By construction, the dependent variable is defined on the set of real numbers; the independent variables are defined on the set of integers and are rescaled to allow a comparison between the modules. We can apply the OLS method.

The expert assessment practice requires the analysts to issue scores for each module. This provides a significant incentive for the analysts to incorporate any prior considerations on the modules into their evaluation. Therefore, it is assumed that a dynamic specification for the modules in the regression is unnecessary.

Running this regression, the Breusch-Pagan test and the Breusch-Godfrey test, with p-values below 1 percent, confirm the presence of heteroscedasticity and autocorrelation of the residuals. This suggests that the regression omits some variables.

Nevertheless, the primary interest here is to obtain the best indication regarding the behaviour of the analysts with respect to the modules. To this end, we continue using the coefficient estimates obtained in the first run of the regression, without including other explanatory variables and employing the standard errors provided by the Newey-West procedure (Verbeek, 2010) in the tests. The coefficients of all modules are significant in explaining the full rating (Table A.2).

**Table A.2 - Significance of the coefficients**

| Regressor | Coefficient | | |
|---|---|---|---|
| Intercept | 0.2115 | *** | (0.0050) |
| FS | -0.3790 | *** | (0.0061) |
| FF | -0.1436 | *** | (0.0052) |
| GV | -0.1224 | *** | (0.0060) |
| SA | -0.1248 | *** | (0.0047) |
| GR | -0.1076 | *** | (0.0054) |
| 3P | -0.1557 | *** | (0.0051) |

Note: in the 'coefficient' column the coefficients of the OLS regression described above are reported, the Newey-West standard errors are reported in parentheses. The asterisks (***,**,*) indicate a statistical significance of 1 per thousand, 1 percent, 5 percent respectively.

Additionally, these coefficients can be ranked (following a specific F-test) as shown in Table A.3.

**Table A.3 - Significance of the differences between the coefficients**

| Test hypothesis | F | Pr(F) | Significance |
|---|---|---|---|
| FS - FF = 0 | 688.53 | < 2.2e-16 | *** |
| FS - GV = 0 | 911.51 | < 2.2e-16 | *** |
| FS - SA = 0 | 899.73 | < 2.2e-16 | *** |
| FS - GR = 0 | 1,006.1 | < 2.2e-16 | *** |
| FS - 3P = 0 | 723.76 | < 2.2e-16 | *** |
| FF - GV = 0 | 6.52 | 0.01 | * |
| FF - SA = 0 | 6.79 | 0.01 | ** |
| FF - GR = 0 | 21.60 | 3.37e-06 | *** |
| FF - 3P = 0 | 2.70 | 0.10 | |
| GV - SA = 0 | 0.10 | 0.76 | |
| GV - GR = 0 | 3.22 | | . |
| GV - 3P = 0 | 20.38 | 6.375e-06 | *** |
| SA - GR = 0 | 5.52 | 0.02 | * |
| SA - 3P = 0 | 21.47 | 3.616e-06 | *** |
| GR - 3P = 0 | 43.26 | 4.896e-11 | *** |

Note: in the previous table, the 'F' column shows the values assumed by the F statistic used to test the hypotheses in the 'Hypothesis under test' column. The value of the p-value relating to column F is reported in the 'Pr(F)' Column. The related outcomes in terms of statistical significance are reported in the 'Significance' Column. The asterisks (***,**,*) indicate a statistical significance of 1 per thousand, 1 percent, 5 percent respectively. The dot (.) indicates a statistical significance of 10 percent. The absence of an asterisk or dot indicates a significance greater than 10 percent.

From the previous table, we note that in absolute terms:

- the largest coefficient pertains to the financial statement module;
- the coefficients for financial flexibility and for third-party opinions follow in magnitude;
- the coefficients for the remaining modules, though not significantly different from each other, exhibit the smallest magnitude.

## Appendix 5. Differences between the analysts

We study portions of the original dataset, each one referring to an analyst. To derive the behavioural characteristics of each analyst, we consider only those who performed at least 100 assessments over sufficiently differentiated firms. This approach results in the selection of over 50 analysts, with individual datasets containing between 100 and 700 assessments.

While the size of datasets can lead to a credible estimate of the analysts' patterns, the main feature of default events is their rarity. Consequently, the results regarding the performance of the analysts' patterns generally have only a heuristic value.

We conduct a regression analysis like the one shown in the previous section, for each analyst. The results are summarized in the following table, which, for simplicity, is only an excerpt from a larger table relating to all the selected analysts.

**Table A.4 - Regressions by analyst (excerpt)**

| | const | FS | FF | GV | SA | GR | 3P | F-test | Adjusted R square |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.2802 | -0.3792 | 0.0000 | 0.0000 | -0.1789 | -0.2133 | 0.0000 | 0.0000 | 0.5281 |
| 2 | 0.0963 | -0.2406 | -0.1448 | -0.1222 | -0.0768 | -0.0877 | -0.1050 | 0.0000 | 0.5317 |
| 3 | 0.1148 | -0.2502 | -0.0739 | -0.0816 | -0.0691 | -0.1254 | -0.1311 | 0.0000 | 0.5561 |
| 4 | 0.1704 | -0.4793 | 0.0000 | 0.0000 | -0.0972 | -0.1110 | -0.1627 | 0.0000 | 0.5995 |
| 5 | 0.0000 | -0.1599 | -0.2079 | -0.1986 | 0.0000 | 0.0000 | -0.1418 | 0.0000 | 0.5222 |
| 6 | ... | ... | ... | ... | ... | ... | ... | ... | ... |

The regression results show a fairly good explanatory power. Specifically:

- the null hypothesis of the F-test is rejected in all cases;
- the R-squared value is medium-high;
- the sign of the constant term is positive and aligns with expectations;
- the sign of the module coefficients is negative and aligns with expectations.

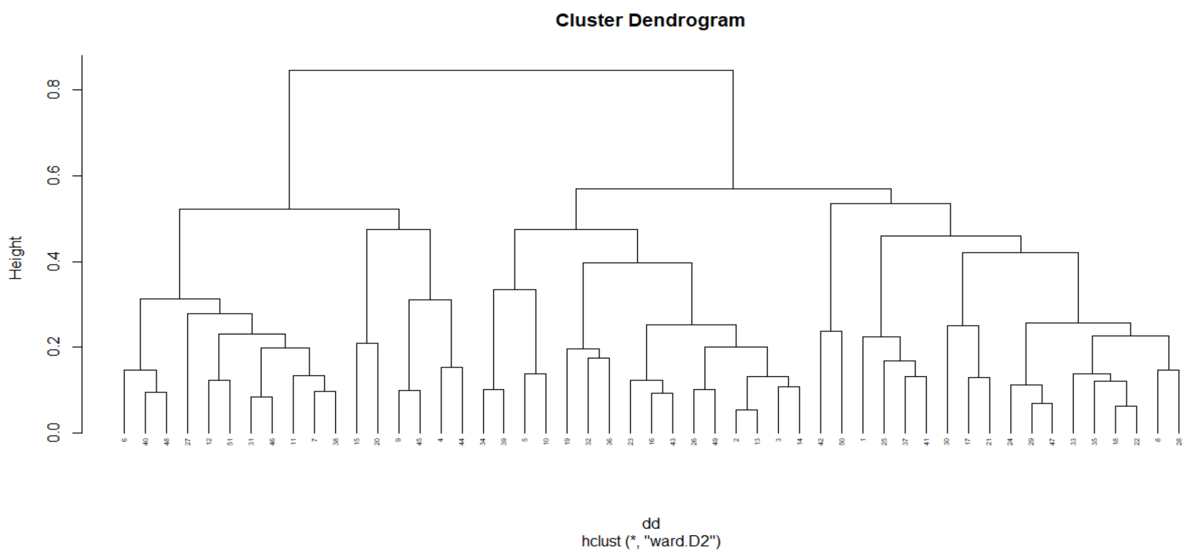
These regressions also indicate that the assessment methods may differ among analysts.

Approximately 80 per cent of the analysts differ from the average analyst as concerns the adoption of modules for the full rating. A significant part of this is attributable to the omission of one or two modules.

The analysts also differ in their choice of the modules that are not employed and of the weights that are attributed to the modules. These differences are often statistically significant. This enables the identification of a set of assessment strategies.

The dendrogram below provides a detailed view of the set of the assessments; we construct the dendrogram by computing proximity in terms of Ward's distance between individual analysts or groups of analysts. We compute the distance between the coefficients of the different regressions reported in Table A.4.

**Figure A.5 - Proximity between analysts**




Note: on the abscissa, we report the identification number for each analyst; on the ordinate, we report the Ward distance between analysts or groups of analysts.

The dendrogram helps in studying the strategies adopted by the analysts. For instance, we study strategies that appear as distant as possible in the dendrogram to see if they differ significantly in terms of discriminatory power and predictive power (and if they differ from that of the average analyst).

We conclude that the heterogeneity in the analysts' behaviour is not rewarding in terms of performance, neither is it in terms of discriminatory power or predictive power.

## Recently published papers in the 'Markets, Infrastructures, Payment Systems' series

n. 42 Report on the payment attitudes of consumers in Italy: results from the ECB SPACE 2022 survey, *by Gabriele Coletti, Alberto Di Iorio, Emanuele Pimpini and Giorgia Rocco*

n. 43 A service architecture for an enhanced Cyber Threat Intelligence capability and its value for the cyber resilience of Financial Market Infrastructures, *by Giuseppe Amato, Simone Ciccarone, Pasquale Digregorio and Giuseppe Natalucci*

n. 44 Fine-tuning large language models for financial markets via ontological reasoning, *by Teodoro Baldazzi, Luigi Bellomarini, Stefano Ceri, Andrea Colombo, Andrea Gentili and Emanuel Sallinger*

n. 45 Sustainability at shareholder meetings in France, Germany and Italy, *by Tiziana De Stefano, Giuseppe Buscemi and Marco Fanari* (in Italian)

n. 46 Money market rate stabilization systems over the last 20 years: the role of the minimum reserve requirement, *by Patrizia Ceccacci, Barbara Mazzetta, Stefano Nobili, Filippo Perazzoli and Mattia Persico*

n. 47 Technology providers in the payment sector: market and regulatory developments, *by Emanuela Cerrato, Enrica Detto, Daniele Natalizi, Federico Semorile and Fabio Zuffranieri*

n. 48 The fundamental role of the repo market and central clearing, *by Cristina Di Luigi, Antonio Perrella and Alessio Ruggieri*

n. 49 From Public to Internal Capital Markets: The Effects of Affiliated IPOs on Group Firms, *by Luana Zaccaria, Simone Narizzano, Francesco Savino and Antonio Scalia*

n. 50 Byzantine Fault Tolerant consensus with confidential quorum certificate for a Central Bank DLT, *by Marco Benedetti, Francesco De Sclavis, Marco Favorito, Giuseppe Galano, Sara Giammusso, Antonio Muci and Matteo Nardelli*

n. 51 Environmental data and scores: lost in translation, *by Enrico Bernardini, Marco Fanari, Enrico Foscolo and Francesco Ruggiero*

n. 52 How important are ESG factors for banks' cost of debt? An empirical investigation, *by Stefano Nobili, Mattia Persico and Rosario Romeo*

n. 53 The Bank of Italy's statistical model for the credit assessment of non-financial firms, *by Simone Narizzano, Marco Orlandi and Antonio Scalia*

n. 54 The revision of PSD2 and the interplay with MiCAR in the rules governing payment services: evolution or revolution?, *by Mattia Suardi*

n. 55 Rating the Raters. A Central Bank Perspective, *by Francesco Columba, Federica Orsini and Stefano Tranquillo*

n. 56 A general framework to assess the smooth implementation of monetary policy: an application to the introduction of the digital euro, *by Annalisa De Nicola and Michelina Lo Russo*

n. 57 The German and Italian Government Bond Markets: The Role of Banks versus Non-Banks. A joint study by Banca d'Italia and Bundesbank, *by Puriya Abbassi, Michele Leonardo Bianchi, Daniela Della Gatta, Raffaele Gallo, Hanna Gohlke, Daniel Krause, Arianna Miglietta, Luca Moller, Jens Orben, Onofrio Panzarino, Dario Ruzzi, Willy Scherrieble and Michael Schmidt*

n. 58 Chat Bankman-Fried? An Exploration of LLM Alignment in Finance, *by Claudia Biancotti, Carolina Camassa, Andrea Coletta, Oliver Giudice and Aldo Glielmo*

n. 59 Modelling transition risk-adjusted probability of default, *by Manuel Cugliari, Alessandra Iannamorelli and Federica Vassalli*

n. 60 The use of Banca d'Italia's credit assessment system for Italian non-financial firms within the Eurosystem's collateral framework, *by Stefano Di Virgilio, Alessandra Iannamorelli, Francesco Monterisi and Simone Narizzano*

n. 61 Fintech Classification Methodology, *by Alessandro Lentini, Daniela Elena Munteanu and Fabrizio Zennaro*

n. 62 The Rise of Climate Risks: Evidence from Expected Default Frequencies for Firms, *by Matilde Faralli and Francesco Ruggiero*

n. 63 Exploratory survey of the Italian market for cybersecurity testing services, *by Anna Barcheri, Luca Bastianelli, Tommaso Curcio, Luca De Angelis, Paolo De Joannon, Gianluca Ralli and Diego Ruggeri*

n. 64 A practical implementation of a quantum-safe PKI in a payment systems environment, *by Luca Buccella and Stefano Massi*

n. 65 Stewardship Policies. A Survey of the Main Issues, *by Marco Fanari, Enrico Bernardini, Elisabetta Cecchet, Francesco Columba, Johnny Di Giampaolo, Gabriele Fraboni, Donatella La Licata, Simone Letta, Gianluca Mango and Roberta Occhilupo*

n. 66 Is there an equity greenium in the euro area?, *by Marco Fanari, Marianna Caccavaio, Davide Di Zio, Simone Letta and Ciriaco Milano*

n. 67 Open Banking in Italy: A Comprehensive Report, *by Carlo Cafarotti and Ravenio Parrini*

n. 68 Report on the payment attitudes of consumers in Italy: results from ECB SPACE 2024 survey, *by Gabriele Coletti, Marialucia Longo, Laura Painelli, Emanuele Pimpini and Giorgia Rocco*

n. 69 A solution for cross-border and cross-currency interoperability of instant payment systems, *by Domenico Di Giulio, Vitangelo Lasorella, Pietro Tiberi*

n. 70 Do firms care about climate change risks? Survey evidence from Italy, *by Francesca Colletti, Francesco Columba, Manuel Cugliari, Alessandra Iannamorelli, Paolo Parlamento and Laura Tozzi*

n. 71 Demand and supply of Italian government bonds during the exit from expansionary monetary policy, *by Fabio Capasso, Francesco Musto, Michele Pagano, Onofrio Panzarino, Alfonso Puorro and Vittorio Siracusa*

n. 72 Statistics on tokenized financial instruments: A challenge for central banks, *by Riccardo Colantonio, Massimo Coletta, Riccardo Renzi*

n. 73 Credit Risk Assessment with Stacked Machine Learning, *by Francesco Columba, Manuel Cugliari, Stefano Di Virgilio*

n. 74 What if Ether Goes to Zero? How Market Risk Becomes Infrastructure Risk in Crypto, *by Claudia Biancotti*

n. 75 The Cyber Risk of Non-Financial Firms, *by Francesco Columba, Manuel Cugliari, Marco Orlandi, Federica Vassalli*

n. 76 Sustainability and financial innovation: The emerging role of Fintech for Good (F4G), *by Alessandro Lentini and Daniela Elena Munteanu*

n. 77 Hydrogeological and credit risk: the Italian firms' physical risk-adjusted probability of default, *by Manuel Cugliari, Simone Narizzano and Federica Vassalli*

n. 78 Liquidity Optimization in Gross Settlement Systems with Quantum Reordering: Application to TARGET2, *by Valerio Astuti, Adriano Baldeschi, Luca Bastianelli, Giuseppe Bruno, Ajit Desai, Danica Marsden and Riccardo Russo*