Journal of Financial Risk Management, 2023, 12, 262-274
https://www.scirp.org/journal/jfrm
ISSN Online: 2167-9541
ISSN Print: 2167-9533

Unlocking Causal Relationships in Commercial Banking Risk Management: An Examination of Explainable AI Integration with Multi-Factor Risk Models

Bing Hu[1*], Yi Wu[2*#]

[1]Financial Intelligence Program, International Technological University, San Jose, USA
[2]UC Berkeley, Berkeley, USA
Email: #yi-wu@berkeley.edu

How to cite this paper: Hu, B., & Wu, Y. (2023). Unlocking Causal Relationships in Commercial Banking Risk Management: An Examination of Explainable AI Integration with Multi-Factor Risk Models. *Journal of Financial Risk Management, 12,* 262-274.
https://doi.org/10.4236/jfrm.2023.123014

Received: August 29, 2023
Accepted: September 25, 2023
Published: September 28, 2023

Abstract

The 21st century has ushered in transformative digital technologies, notably Artificial Intelligence (AI), which has the potential to redefine commercial banking risk management, especially in the current complicated geopolitical context. This paper examines the integration of Explainable AI into traditional multi-factor models used in commercial banking. Traditional models, while foundational, often struggle to decipher intricate causal relationships between various risk factors, especially with limited data. With the advent of AI, especially machine learning techniques like Bayesian networks and random forests, there is an opportunity to enhance these models by capturing intricate risk interdependencies and predicting future risks more precisely. We delve deep into the nuances of XAI, emphasizing its potential in making AI's decision-making transparent and interpretable, addressing the "black box" challenge. Furthermore, we explore the application of Explainable AI in detecting causal relationships within restricted datasets, underscoring the importance of techniques like cross-validation, regularization, and bootstrapping. The paper concludes by highlighting the need for a synergistic approach, combining Explainable AI's capabilities with the robustness of traditional models, setting the stage for future research in this promising nexus of technology and finance.

Keywords

Artificial Intelligence, Commercial Banking, Causal Relationships,

*These authors shared first authorships.
#Corresponding author.

Explainable Artificial Intelligence, Banking Risk Management, Transparency and Interpretability

1. Introduction

The rapid proliferation of digital technologies in the 21st century has dramatically transformed various aspects of human life and industry operations. Artificial Intelligence (AI) stands out among these innovations, influencing numerous sectors and revolutionizing problem-solving and decision-making methodologies.

Modern commercial banking, with its interconnected financial systems and evolving regulatory landscapes, demands robust risk management strategies (Jobst & Gray, 2013). Traditional multi-factor models, while instrumental, often fall short in capturing the intricate causal relationships among various risk factors. Additionally, the commercial banking sector's data explosion, combined with a pressing need for transparent decision-making, signals the urgency for more advanced risk management approaches. The failure of Silicon Valley Bank underscores the ramifications of lapses in regulatory compliance, emphasizing the importance of adaptability and foresight in banking operations (Hu & Wu, 2023).

A promising solution to these challenges lies in the integration of AI into commercial banking risk management. AI's prowess in handling vast data, discerning patterns, and accurate forecasting enhances traditional risk management models. Furthermore, the emergence of Explainable AI, as illustrated in **Figure 1**, addresses the "black box" dilemma of AI, striving to make AI's decision-making processes transparent and comprehensible.

This paper seeks to explore XAI's potential in refining traditional multi-factor models for commercial banking risk management. We will navigate the intricacies of XAI, its role in demystifying complex risk structures, and the hurdles faced when merging XAI with conventional models. By delving deep into XAI's theory and practical applications, this paper aims to further the dialogue on XAI's role in commercial banking and pave the way for continued research in this dynamic intersection of technology and finance.

2. Literature Review

Machine Learning (ML) and Artificial Intelligence (AI) have profoundly impacted various sectors, with financial sector being a key beneficiary, which has been rapidly increasing the use of Artificial Intelligence/Machine Learning systems (Boukherouaa et al., 2021). As delineated in **Table 1**, various capabilities of artificial intelligence correspond to distinct roles within the realm of risk management in commercial banking.

Current banking risk management models encompass various factors that potentially influence the risk profile of a bank. However, traditional models struggle

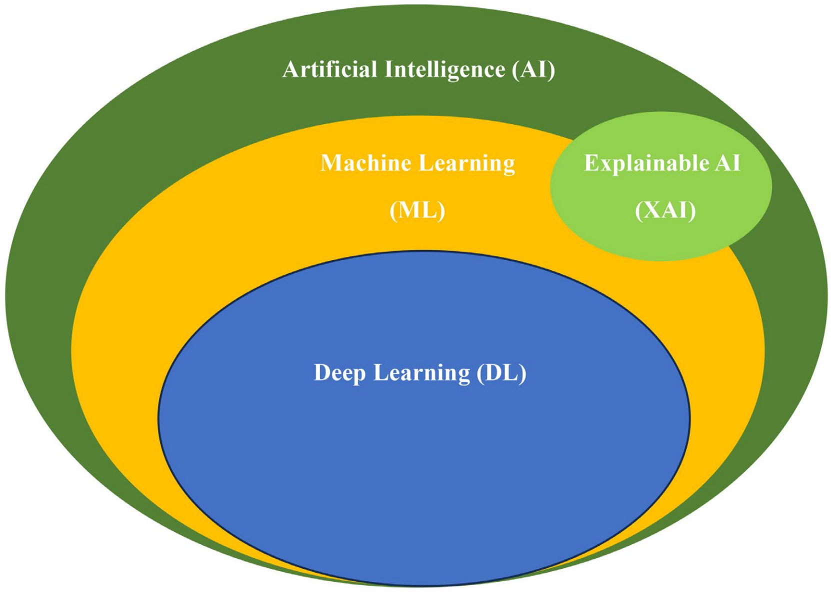

Figure 1. Relationship of explainable AI with AI.

Table 1. Artificial intelligence and machine learning capabilities (Chandola et al., 2009).

Capability	Related with Commercial Banking Risk Management
Forecasting	Credit risk scoring, economic and financial variables forecasting, risk management, etc.
Natural language processing	Chat bots, contract reviewing, and report generation.
Image recognition	Assist with carrying out certain anti-money laundering/combating the financing of terrorism (AML/CFT) requirement.
Anomaly detection	Insider trading, credit card and insurance fraud detection, and AML/CFT are some of the applications that leverage this capability.

with determining the causal relationships among these factors, making it challenging to predict and manage potential risks accurately. The integration of machine learning techniques into commercial banking risk management presents a promising solution. Machine learning's predictive capabilities outperform traditional statistical methods in many instances. For instance, support vector machines (SVMs) and neural networks can identify patterns and trends within large datasets more effectively, helping predict future outcomes with greater accuracy (Sarker, 2021).

In the context of multi-factor risk models, machine learning may be particularly beneficial in identifying and understanding causal relationships. Machine learning algorithms can process vast amounts of data to discern patterns that might remain hidden with traditional models. These patterns can offer valuable insights into the causal relationships between different risk factors. A significant example of machine learning's application in understanding causal relationships in multi-factor risk models is demonstrated in the work of OECD (2021). They employed machine learning techniques to a dataset from a large multinational bank and uncovered several new causal relationships between risk factors previously overlooked. These findings have the potential to revolutionize risk management strategies.

Despite the promising advancements, there remain challenges in adopting machine learning in commercial banking risk management. Firstly, some scholars point out these advanced algorithms require large volumes of high-quality data to function effectively. This requirement might pose a problem for banks with limited or incomplete datasets (OECD, 2021). Additionally, the "black box" nature of some machine learning models may pose transparency issues, making it difficult for regulators to understand how these models arrive at their decisions (Faggella, 2020). To summarize, integrating machine learning with multi-factor risk models in banking risk management may facilitate the understanding of causal relationships, which can, in turn, improve risk management. However, challenges such as data requirements and model transparency need to be addressed to fully leverage machine learning's capabilities.

European Central Bank raised a satellite panel model to translate a given Macro Economic Scenario into Risk Parameters at the firm level (ECB, 2017), while the concept of satellite model was first mentioned by IMF in 2013, as **Figure 2** shows. The concept of distinct satellite risk models emerged in the context of post-crisis large scale stress testing activities organized by banking regulators worldwide. A key design of these stress testing exercises is that a set of common scenarios are specified as applying to all firms and those are then translated into specific firm models using proprietary firm specific data.

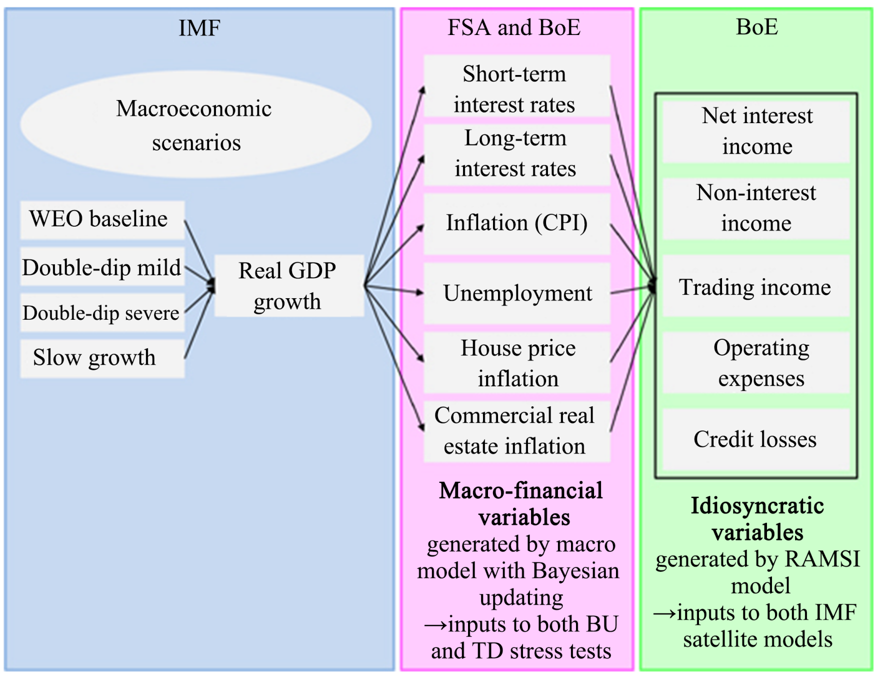

Figure 2. Example of satellite model estimations for bank solvency stress testing: U.K. FSAP. Source: Jobst et al. (2013).

To explore its theoretical basis, the current state-of-the art satellite model for PD translation is Bayesian model averaging (BMA) (Raftery, 1995). It has a long track record as being a reliable tool for generating scenario-conditional projections for credit risk and is being adopted by more and more central banks and institutions. However, with the easier access to regression models and the advent of new predictive models in the field of machine learning, the question arises if there are other models that could deliver better results. In 2011, Turing Award laureate and the father of Bayesian networks, Judea Pearl, predicted that artificial intelligence was entering a bottleneck phase in its current development and advocated for a greater focus on causal inference in AI. In his book, *The New Science of Cause and Effect in Data Science and Artificial Intelligence*, he proposed that data science is shifting from a data-centric paradigm to a science-centric paradigm, leading to a sweeping "causal revolution" across various research domains. Yoshua Bengio and Yann LeCun have publicly asserted that causal reasoning constitutes a crucial approach for enhancing the generalization capabilities of machine learning and deep learning (ML/DL) models.

3. Traditional Multi-Factor Models in Commercial Banking Risk Management

In modern banking operations, commercial bank risk management is an indispensable facet, as depicted by the intricate system shown in Figure 3. This necessity stems from the fact that banks operate within a dynamic financial ecosystem, one that's replete with a diverse array of inherent risks. The objective

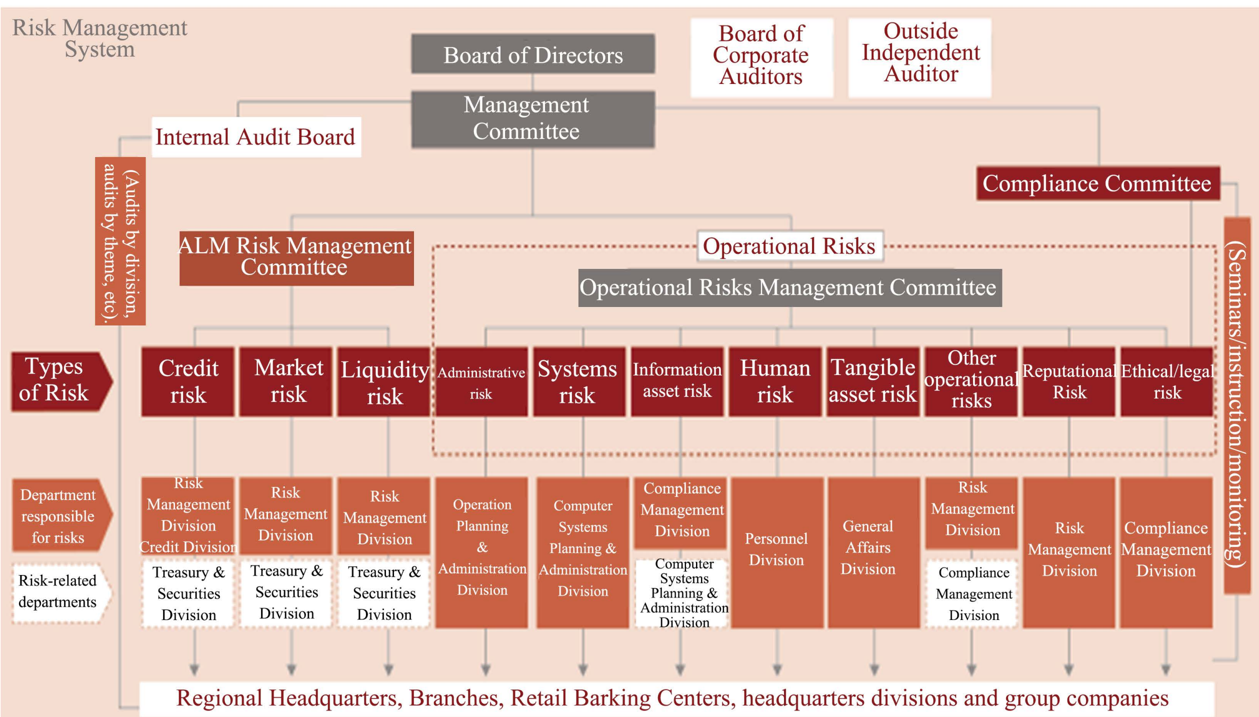

Figure 3. Commercial bank risk management system. Source: Hyakugo bank group.

is to foresee these risks and devise appropriate strategies to alleviate their potential negative impacts. An instrumental approach to this has been the adoption of multi-factor models, which have held a cornerstone role in traditional bank risk management. At the core of these models are several key risk types that banks must navigate. These risks typically include credit risk, market risk, operational risk, liquidity risk, and business risk. In the traditional multi-factor model, each risk type contributes to a bank's total risk portfolio in a specific, quantifiable manner.

Credit risk pertains to the possibility of a debtor defaulting on their obligations, causing financial loss for the bank. Traditional multi-factor models would examine multiple credit-related factors like the debtor's credit history, income stability, and the economic climate, to gauge the potential risk of lending to a particular customer. Market risk, another pivotal aspect, involves losses that may result from changes in market prices, like interest rates or foreign exchange rates. Multi-factor models would consider numerous factors, such as volatility in the exchange rate or the interest rate sensitivity of the bank's assets and liabilities, to assess market risk. Operational risk revolves around potential losses from inadequate or failed internal processes, people, systems, or external events. Banks have traditionally assessed operational risk by examining historical data on internal losses, external loss data, business environment, and internal control factors. Liquidity risk, the possibility that a bank will not be able to meet its obligations as they come due without incurring unacceptable losses, is also factored into these models. Banks traditionally evaluate liquidity risk by observing factors like cash flow projections, funding diversification, and liquidity gap analysis. Lastly, business risk is assessed by observing factors like the competitive environment, changes in customer behavior, and changes in regulation that might affect a bank's profits.

Despite their utility, these traditional multi-factor models are not without limitations. The fundamental constraint lies in their ability to accurately capture the complex interrelations among different risk types. For instance, during an economic downturn, credit risk and market risk may become more interlinked as falling asset prices could trigger defaults. Unfortunately, traditional models may not be sophisticated enough to fully capture these dynamic interdependencies, which could lead to an underestimation of the total risk. Additionally, traditional multi-factor models tend to rely heavily on historical data. This approach assumes that the future will largely reflect the past, an assumption that can be unreliable during periods of rapid change or unprecedented events. The 2008 financial crisis served as a stark reminder of this limitation, as many banks found that their traditional risk models were ill-equipped to predict the magnitude of the crisis.

4. Integrating Explainable AI into Multi-Factor Models in Commercial Banking Risk Management

The advent of Artificial Intelligence (AI), as shown in **Figure 4**, has offered new

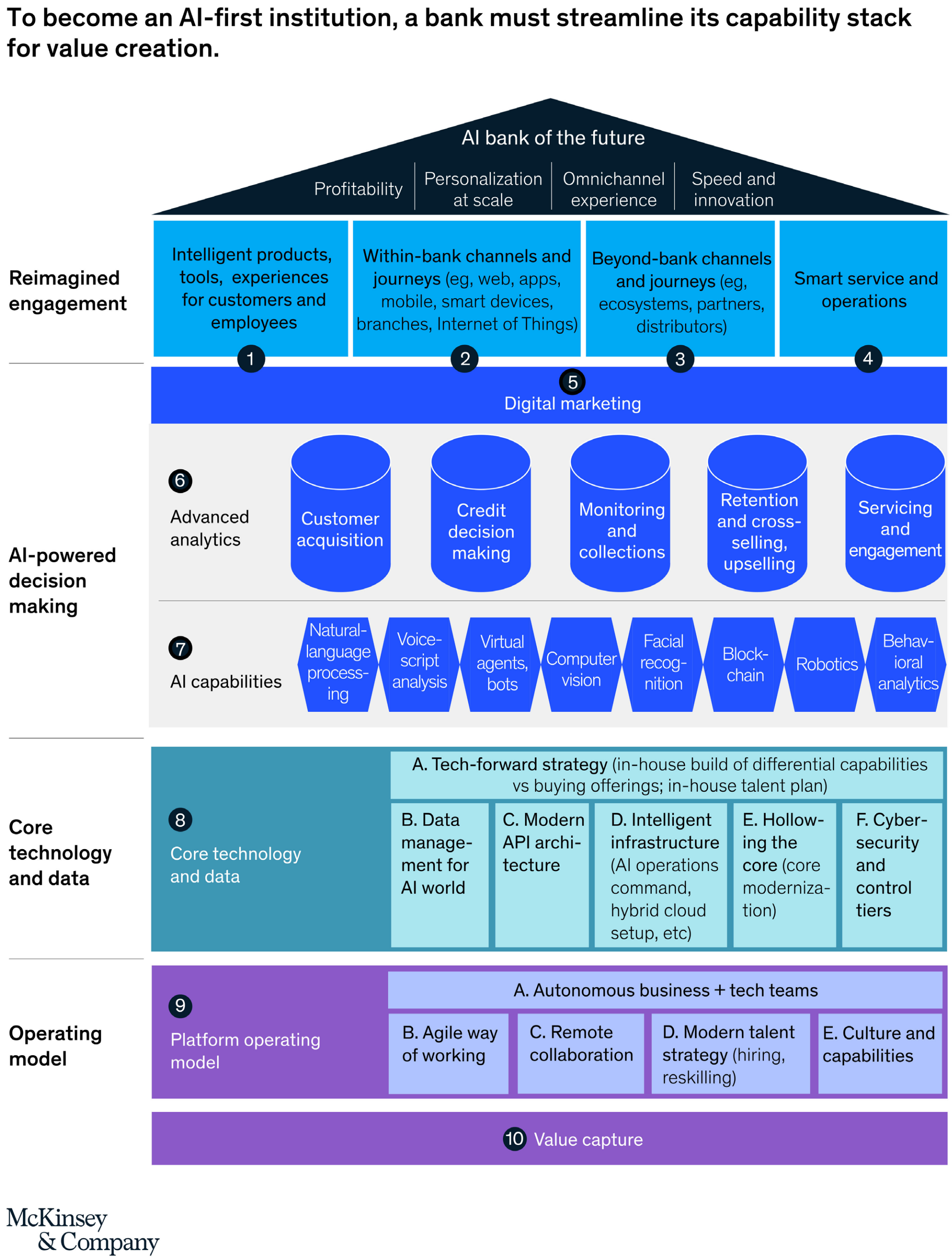

Figure 4. The changes of commercial bank after integrating AI. Source: McKinsey & company.

perspectives and potential solutions for long-standing challenges across various sectors, including commercial banking. In the field of commercial bank risk management, Explainable AI has emerged as a revolutionary tool that can significantly enhance traditional multi-factor models. The enhanced models can unravel the complex interdependencies among different risk factors, predict future risks more accurately, and improve overall risk management.

The unique feature of Explainable AI that makes it invaluable for risk management is its ability to learn from data and make predictions. Explainable AI can process vast amounts of data, identify complex patterns, and make accurate predictions based on those patterns. This predictive capability can significantly improve risk management by enabling banks to anticipate potential risks and take preventive measures.

One area where Explainable AI can significantly enhance multi-factor risk models is in the prediction of credit risk. Traditional credit risk models consider various factors such as the borrower's credit history, income stability, and macroeconomic conditions. However, these models often fail to capture the complex interactions between these factors. Explainable AI, with its advanced algorithms, can analyze vast amounts of data, understand these interactions, and predict credit risk with greater accuracy (Bussmann et al., 2021). Moreover, AI can also enhance the prediction of market risk. Market risk involves potential losses resulting from changes in market prices such as interest rates or foreign exchange rates. AI algorithms, such as machine learning, can analyze historical and real-time data, understand market trends, and predict changes in market prices. This predictive capability can enable banks to take proactive measures to mitigate potential losses resulting from market fluctuations (Jobst & Gray, 2013).

Operational risk, which stems from inadequate or failed internal processes, people, and systems, can also be effectively managed using Explainable AI. Explainable AI can analyze large volumes of historical and real-time data to predict potential operational failures. For instance, machine learning algorithms can analyze historical data on system failures to predict potential system breakdowns, enabling banks to take preventive measures (Boukherouaa et al., 2021). In terms of liquidity risk, Explainable AI can help banks predict cash flow patterns more accurately. Through AI's deep learning algorithms, banks can analyze vast amounts of data, understand complex cash flow patterns, and predict future cash flows. This predictive capability can enable banks to better manage their liquidity and avoid potential liquidity crises. Explainable AI's application in managing business risk is also notable. Business risks arise from changes in the competitive environment, customer behavior, or regulations. Explainable AI can help banks understand these changes by analyzing vast amounts of data and predicting future trends. For example, Explainable AI have the potential to analyze data on customer behavior to predict future customer needs and preferences, enabling banks to adapt their strategies accordingly.

5. Explainable AI and the Detection of Causal Relationships: Working with Limited Data

Artificial Intelligence (AI) has been instrumental in uncovering hidden patterns and extracting valuable insights from complex data sets. A particular area where Explainable AI proves its worth is in the detection of causal relationships, especially when working with limited data.

Traditional statistical methods of determining causality often rely on the assumption of large sample sizes. However, the real-world seldom provides us with perfect, extensive datasets. In many scenarios, the available data might be limited due to several factors such as privacy constraints, resource limitations, or the novelty of the field under investigation. It is in these contexts that Explainable AI, and more specifically machine learning techniques, can play a pivotal role. Machine learning, a subset of AI, enables the identification of causal relationships with comparatively small datasets. It does so by using the available information to learn the underlying patterns and structures that can imply causality. This is significantly different from traditional correlation-based methods that often confuse correlation with causation.

A specific machine learning approach suited to discovering causal relationships from small datasets is the use of Bayesian networks. Bayesian networks are probabilistic graphical models that represent the set of variables and their conditional dependencies via a directed acyclic graph. They provide a succinct representation of the joint probability distribution and enable the modeling of complex stochastic processes. By their design, they inherently capture the causal structure among variables.

Bayesian networks are a family of probability distributions that admit a compact parametrization that can be naturally described using a directed graph. The general idea behind this parametrization is surprisingly simple. Recall that by the chain rule, we can write any probability p as (Zhang, 2021) (Equation (1)):

$$p(x_1, x_2, \cdots, x_n) = p(x_1)\, p(x_2 \mid x_1) \cdots p(x_n \mid x_{n-1}, \cdots, x_2, x_1). \tag{1}$$

A compact Bayesian network is a distribution in which each factor on the right-hand side depends only on a small number of *ancestor variables* x_{A_i} (Equation (2)):

$$p(x_i \mid x_{i-1}, \cdots, x_1) = p(x_i \mid x_{A_i}). \tag{2}$$

For example, in a model with five variables, we may choose to approximate the factor $p(x_5 \mid x_4, x_3, x_2, x_1)$ with $p(x_5 \mid x_4, x_3)$. In this case we write $x_{A_5} = \{x_4, x_3\}$.

When the variables are discrete (which will often be the case in the problems we will consider), we may think of the factors $p(x_i \mid x_{A_i})$ as probability tables, in which row correspond to assignments to x_{A_i} and columns correspond to values of x_i; the entries contain the actual probabilities $p(x_i \mid x_{A_i})$. If each variable takes d values and has at most k ancestors, then the entire table will contain at most $O(nd^{k+1})$ entries. Since we have one table per variable, the entire probability distribution can be compactly described with only $O(nd^{k+1})$ parameters (compared to $O(d^n)$ with a naive approach). With a Bayesian network, we can manage limited data by using the principle of "Occam's Razor," stating that simpler models should be preferred over more complex ones when they fit the data similarly well. This principle guides the search for the best model that explains the data and helps avoid overfitting when working with small sample sizes (Bargagli Stoffi et al., 2022).

Random forests, another AI method, are also beneficial for identifying causal relationships in limited data scenarios. The operational algorithm is illustrated in Figure 5, wherein this technique employs the generation of multiple decision trees, utilizing the mode of their outcomes for the ultimate prediction. It enables identifying non-linear dependencies and interactions between variables, which are crucial for detecting causal relationships. Causal discovery algorithms, such as the PC algorithm (shown as Figure 6), have also shown promise in revealing causal structures from limited data. The PC algorithm is a constraint-based method that combines statistical tests of independence with graph-theoretical concepts to uncover causal relationships.

However, finding causal relationships with limited data through AI does not come without challenges. Overfitting, where the model captures noise instead of the underlying pattern, is a key concern. Furthermore, even with AI, small sample sizes can lead to less reliable and less generalizable results. To overcome these challenges, techniques such as cross-validation, regularization, and bootstrapping can be applied. Cross-validation helps assess how well the model will generalize to unseen data. Regularization techniques prevent overfitting by adding a penalty term to the loss function. Bootstrapping, or resampling, can increase the apparent size of the data set, providing more robustness in the models.

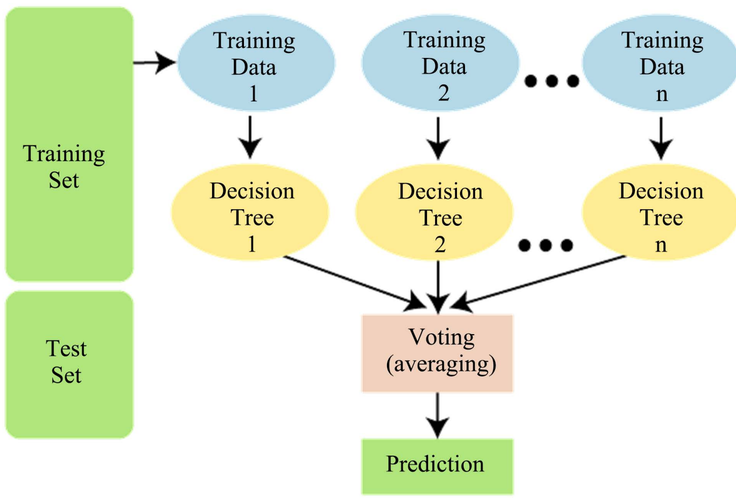

Figure 5. Working of the random forest algorithm.

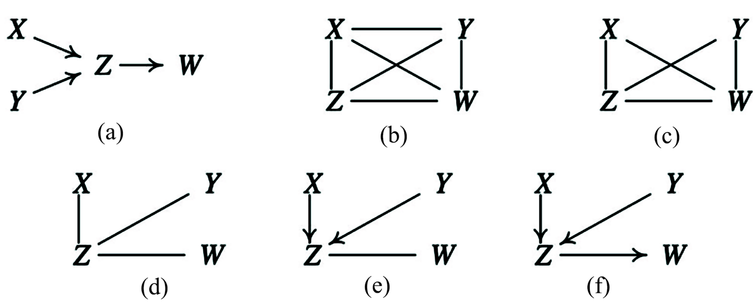

Figure 6. Working of the PC algorithm. (a) Original true causal graph. (b) PC starts with a fully-connected undirected graph. (c) The X − Y edge is removed because X ⊥ ⊥ Y. (d) The X − W and Y − W edges are removed because X ⊥ ⊥ W | Z and Y ⊥ ⊥ W | Z. (e) After finding v-structures. (f) After orientation propagation.

6. Discussion and Further Study

In the current state of some machine learning models, often described as "garbage in, garbage out", the opacity of algorithms can result in erroneous outputs if the inputs are flawed. This lack of transparency can pose severe challenges for financial institutions that must constantly justify and explain their risk assessments and operational decisions. For example, in the consideration of implementing ChatGPT within the banking system, several key concerns arise that make it an unsuitable choice. The core of these concerns revolves around control, openness, and interrelated key factors essential for the banking ecosystem. To foster trust and compliance, there needs to be a concerted effort towards developing AI systems that are not only powerful but also transparent and accountable. Integrating these aspects into Explainable AI could be likened to inserting an "eye" into the black box, providing a clear line of sight into the "why" and "how" behind every decision. For the commercial banking, where accountability and traceability are not just ethical practices but legal necessities, this transparent approach to AI may well be the path towards a more secure and responsible future.

The role of Explainable AI in detecting causal relationships within limited datasets, particularly in the context of commercial banking risk management, offers profound potential for future research and practice. This exploration into integrating Explainable AI with traditional multi-factor models has revealed a paradigm shift in understanding risk, modeling causal relationships, and utilizing limited data for decision-making. Explainable AI's capacity to process complex data and uncover hidden patterns can significantly enhance multi-factor risk models' predictive accuracy. Bayesian networks, random forests, and causal discovery algorithms, for instance, facilitate the extraction of meaningful information from small datasets. They achieve this by capturing causal structures, identifying non-linear dependencies, and performing statistical tests of independence. However, the application of these methods must be carefully managed to avoid overfitting and to ensure the results' reliability and generalizability.

The discussion on Explainable AI and limited data has highlighted the necessity of cross-validation, regularization, and bootstrapping as safeguards against the inherent challenges in dealing with small datasets. While these methods show promise, they are not silver bullets and need to be considered as part of a holistic approach to model design and validation. As we have navigated through the complexity of integrating Explainable AI with multi-factor risk models, we acknowledge that the implementation of these advancements may face obstacles. Notably, the need for specialized expertise and ethical considerations is around data use. The transformation in commercial banking risk management driven by AI necessitates comprehensive frameworks that address data privacy, security, and governance. Moreover, the human factor in AI-driven decision-making must not be overshadowed by the technological advancements. The interpretability of Explainable AI models, a significant concern in AI ethics, becomes even more critical when dealing with limited data. Transparent, interpretable models will not only build trust among decision-makers but also promote a culture of accountability in Explainable AI deployment.

For future research, an area of interest could be the development of hybrid models that combine traditional statistical techniques with Explainable AI methods to leverage the strengths of both. The integration of Explainable AI in multi-factor models opens avenues for further research into AI's role in other areas of banking, such as fraud detection, customer segmentation, and credit scoring. Another direction for future research is the application of Explainable AI in the detection of causal relationships in other industries and domains. The lessons from the banking sector could provide a starting point for these explorations. Further, expanding the scope beyond causal relationships, Explainable AI's potential in predicting future trends using limited data, or under conditions of uncertainty, presents a fascinating area of study.

In conclusion, the journey towards integrating Explainable AI into commercial banking risk management has only just begun. As the field continues to evolve, so will our understanding of the interactions between Explainable AI, multi-factor risk models, and limited data. It is through this continuous exploration and learning that we can fully unlock Explainable AI's potential in commercial banking risk management.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Bargagli Stoffi, F. J., Cevolani, G., & Gnecco, G. (2022). Simple Models in Complex Worlds: Occam's Razor and Statistical Learning Theory. *Minds & Machines, 32,* 13-42. https://doi.org/10.1007/s11023-022-09592-z

Boukherouaa, E. B., AlAjmi, K., Deodoro, J., Farias, A., & Ravikumar, R. (2021). *Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance* (Vol. 2021). International Monetary Fund. https://doi.org/10.5089/9781589063952.087

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics, 57,* 203-216. https://doi.org/10.1007/s10614-020-10042-0

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Computing Surveys, 41,* Article No. 15. https://doi.org/10.1145/1541880.1541882

European Central Bank (ECB) (2017). *Occasional Paper Series. Macroprudential Stress Test of the Euro Area Banking System.* https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op226~5e126a8e37.en.pdf

Faggella, D. (2020). *AI Transparency in Finance—Understanding the Black Box.* https://emerj.com/partner-content/ai-transparency-in-finance/

Hu, B., & Wu, Y. (2023). AI-Based Compliance Automation in Commercial Bank: How the Silicon Valley Bank Provided a Cautionary Tale for Future Integration. *International Research in Economics and Finance, 7,* 13-15. https://doi.org/10.20849/iref.v7i1.1356

Jobst, A., & Gray, D. F. (2013). *Systemic Contingent Claims Analysis: Estimating Market-Implied Systemic Risk* (Vol. 2013). International Monetary Fund. https://doi.org/10.5089/9781475572780.001

Jobst, A., Ong, L. L., & Schmieder, C. (2013) *A Framework for Macroprudential Bank Solvency Stress Testing: Application to S-25 and Other G-20 Country FSAPs.* https://doi.org/10.5089/9781616355074.001

OECD (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers.* https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology, 25,* 111-163. https://doi.org/10.2307/271063

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science, 2,* Article No. 160. https://doi.org/10.1007/s42979-021-00592-x

Zhang, R. C. (2021). *Bayesian Networks.* https://zhangruochi.com/Bayesian-networks/2020/03/23/