Unleashing the potential of Open Banking: three use cases for risk assessment

Rui Ying Goh

Doctor of Philosophy

University of Edinburgh

2024

Declaration

I declare that this thesis has been composed entirely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work contained herein is my own.

Acknowledgement

Time really flies - I am now at the end of my four-year PhD. Apart from academic development, my first living abroad life has been very memorable, thanks to everyone that have supported me throughout the journey.

I would like to express my greatest gratitude to my supervisory team. To Galina Andreeva, my primary supervisor, thank you for all the constructive discussions in shaping the project. Being a newbie in this domain, I greatly appreciate your expertise in guiding the project, bringing interesting inputs and ensuring the research moves in the right path. Thank you also for being very supportive and actively involved in all the meetings with the industrial partner. To Yi Cao, my secondary supervisor, thank you for your quick adaption when joining the team in my second year. I am very grateful for your active input that has brought new perspectives during the research development. Also, to Jonathan Crook, many thanks for having you in my initial PhD journey, I thank you for your bright ideas in all the brainstorming sessions at the early stage of the research to build up the foundation of this project. Additionally, this endeavour would not have been possible without the funders, EIT Digital and DirectID, to finance the research as well as Edinburgh Future Institute to set up this collaboration with the Business School of Edinburgh University.

Thanks to friendly colleagues in DirectID, especially Andy and Frank, thanks for always being open to questions and making me feel comfortable to ask for answers. I am also grateful to Johannes de Smedt, for being my host during my research visit to KU Leuven. Thank you for being the collaborator to work on one of the chapters in my PhD, I appreciate your guidance and advice when developing the methodology. I am also very lucky to have warm and lively PhD colleagues to share our research journey and motivate each other all the time.

Finally, and most importantly, to my parents, thanks for all the love and encouragement, who have inspired me to pursue my dream.

Lay summary

The implementation of Open Banking (OB) has brought revolutionary changes to the financial ecosystem. Based on the Second Payment Service Directive legislation foundation, OB requires banks to provide third-party service providers access to bank account data upon customers' authorisation. The UK OB standard enforces a secured bank data-sharing protocol via open Application Programming Interfaces for smooth and safe between-customer interactions. Such permission-based OB data access practice provides a new reliable data source to exploit insights on consumers' financial behaviour. Having access to the voluminous bank account transactions not only opens new opportunities for practitioners to provide data-driven personalised solutions to consumers but also brings challenges on scoping 'what' and 'how' to draw relevant insights from the data. This thesis focuses on leveraging OB transactions for decision analytics to support robust financial lives.

Cash flow transactions contain inherent exposure risks to financial threats where money inflow and outflow directly reflect consumers' response when dealing with turbulent economic conditions, money management practices or peculiar activities for illicit account use. Being able to model the underlying financial risks from OB transactions can effectively achieve a win-win for practitioners and consumers – financial firms can objectively assess day-to-day financial patterns to better understand consumer behaviour and effectively set portfolio management strategies tailored to the respective circumstances for risk mitigation. This thesis contributes to the methodological developments of three alternative personal financial risk assessment frameworks in the OB literature. Each framework represents a model designation (how) with a specified target indicator (what) for risk evaluation.

Chapter 2 examines the way to forecast future account-level volatility from previous transactional behaviour. We define a novel composite index of account financial volatility which simultaneously incorporates fluctuations in inflow, outflow, and balance. We also quantify multiple vulnerable characteristics from the transactional patterns and combine them to form a single vulnerability index. Using past volatility and vulnerability as predictors, we employ standard statistical and state-of-the-art machine learning regression models to forecast the target variable i.e., the financial volatility score. We emphasise interpretations of the non-linear relationship between vulnerability and volatility to reveal harmful volatile behaviour. The proposed personal volatility prediction has two practical implications. First, the predicted financial volatility score evaluates the degree of uncertainties in cash flow patterns to signal financial instability. Second, the model explainability provides transparent reasoning on which specific vulnerable characteristics drive risky (low/high) volatility to warn potential financial difficulty.

Chapter 3 tracks the evolution of distress risk over time. We consider the dynamic account balance to indicate the adequacy of financial buffer, to imply the risk level to face distress. We pioneer the use of a multivariate latent Markov model to exploit the underlying distress risk states from three observed target variables representing different aspects of observed balance i.e., duration (number of days in negative), materiality of the magnitude of negative amount (ratio of minimum to median balance), and fluctuation of negative balance (coefficient of variation of minimum balance). We include vulnerable-related transactional behaviour as the covariates to estimate and predict whether an account transit from one (latent) distress risk state to another, to capture financial improvement or deterioration. This statistical framework to assess dynamic distress risk is an essential tool to commensurate the implicit distress risk levels from account balance in a data-driven manner as well as monitor financial transitions alongside tracing the respective vulnerability drivers for timely and appropriate interventions to mitigate risks.

Chapter 4 detects potential suspicious accounts from cash flow transactions in a three-stage unsupervised learning approach. First, we propose a novel RFMP (Recency, Frequency, Monetary, Persistence) framework, an extension of the standard RFM marketing tool with the persistence dimension, to extract cash flow attributes from bank transactions. The persistence dimension is a vital component to capture regularity from the intervals between transaction occurrences and pinpoint anomalous patterns such as subtle consistency despite low F or bursty transaction spikes with large M, which would be neglected with only the RFM dimensions. Second, we build clusters of RFMP profiles as a baseline to distinguish typical vs atypical account behaviour. These profiles are crucial to not only enhance understanding of customer needs at the point of onboarding but also act as the guideline to ensure minority cluster is not confused with rare dubious behaviour. Third, we derive outlier scores as the proxy to measure the exposure risk to suspicious activities followed by interpreting the characteristics of outlier accounts to spot peculiar account behaviour, to be flagged for further investigations. The outlier detection scheme acts as a key account screening tool to efficiently identify accounts with high anomaly scores and justify the rationale to escalate an account for further due diligence.

This thesis investigates three different applications of OB data for personal financial risk assessment. We showcase the development of the financial risk models for OB analytics to work in practice. OB-driven decision analytics contribute to better risk management and improve financial lives in several ways. Early warning signals can nudge potential financial struggles for timely engagement and interventions, as well as enable responsible lending to avoid loan granting to potentially unaffordable customers. Interpretations of the transactional behaviour associated with financial risk strengthen the understanding of customer-specific problems to provide tailored preemptive or remedial solutions. Risk evaluation with OB data embraces financial inclusion to enable thin credit file consumers to have equal opportunity to undergo the onboarding assessment process to access financial products in the market. As the OB adoption rate continues to grow, the development of OB-based risk models in this thesis can spark innovative ideas to deploy OB analytics for practical applications as open data becomes the new paradigm in the financial landscape.

Abstract

This thesis includes three novel frameworks for modelling financial risks based on current account transactions of individuals within the area of Open Banking (OB). The implementation of OB allows third-party service providers to access customer bank account data, creating opportunities for personalised, data-driven solutions while posing challenges in extracting relevant insights from the extensive transactional data. The thesis contributes to methodological development in the OB literature, by considering three different applications for the transactional data with corresponding measures (what) and sets of models (how) for risk evaluation. The first application focuses on forecasting account-level volatility and linking it to previous vulnerable transactional behaviour. A novel composite index of account financial volatility, incorporating fluctuations in inflow, outflow, and balance, is proposed. Past volatility and vulnerability serve as predictors, and statistical and state-of-the-art machine learning regression models are employed to forecast the financial volatility score. The results provide the ranking of the degree of uncertainties in cash flow patterns that signal financial instability. In addition, the interpretation of the linear and non-linear relationship between vulnerability and volatility provides insights into situations when volatile behaviour signals financial problems. The analysis reveals diverse linear and non-linear relations between transactional patterns and risky volatile behaviour.

The second application introduces data-driven measures of financial distress from OB transactions and tracks the evolution of distress risk over time, by using of a multivariate latent Markov model. This model exploits underlying distress risk states from three different aspects of observed balance i.e., duration (number of days in negative balance), materiality of the magnitude of negative amount (ratio of minimum to median balance), and fluctuation of negative balance (coefficient of variation of minimum balance). The statistical framework assesses dynamic distress risk in a data-driven manner, incorporating vulnerability-related transactional behaviour as covariates to estimate and predict accounts' transition between latent distress risk states. This allows for effective ranking of distress risk levels and monitoring of vulnerability drivers of financial transitions, aiding appropriate interventions.

In the final use case, a novel RFMP (Recency, Frequency, Monetary, Persistence) representation is proposed for identifying potentially suspicious accounts based on cash flow transactions through

unsupervised learning. This data representation framework extends the well-known RFM model by incorporating a persistence dimension, captures regularity in cash flows and detects anomalous patterns often overlooked by traditional RFM dimensions. By clustering RFMP attributes, typical account behaviours emerge, such that they are commonly used for receiving salaries and facilitating payments for maintaining a living. These account patterns establish an understanding of normal behaviours that are used as a baseline. An outlier detection scheme derives an anomaly score as a proxy to measure exposure risk to potentially suspicious activities, followed by interpreting the outlier characteristics to justify the escalation of potentially suspicious accounts. This unsupervised framework can function as an efficient automated account screening tool to support the due diligence process. This thesis carries significant implications for retail banking practitioners. The proposed OB analytics offers early warning signals for timely engagement and interventions, providing a nuanced understanding of three specific problems for the customised implementation of pre-emptive or remedial solutions.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background: Open Banking

Open Banking (OB) is a permission-driven secured data-sharing practice to open bank data to third-party service providers (TPPs) e.g., fintech[1]. The Second Payment Service Directive (PSD2) is the legislation foundation for OB which requires banks to grant TPPs access to payment accounts upon explicit consumer consent. PSD2 regulates two types of service providers: Payment Initiation Service Providers (PISP) to directly initiate online third-party transfer payments, and Account Information Service Providers (AISP) to enable collection of account information (Brodsky & Oakes, 2017).

OB transforms the financial landscape by providing consumers with increased flexibility to manage their finances. It offers a unified platform for clear visibility into multiple accounts and improved payment accessibility (Figure 1.1). Traditional banking entails a one-to-one relationship between banks and customers, where the bank serves as the sole hub for managing account data. In contrast, OB fosters ecosystem-enabled relationships, allowing open access to bank data and services through third-party channels, upon consumers' consent. This approach enhances transparency and encourages nonbanks to actively participate in the ecosystem.

In the UK, the Competition Markets Authority (CMA) mandates nine major banks to implement OB through the Open Banking Implementation Entity (OBIE). The UK OB standard ensures that bank data is accessible through open Application Programming Interfaces (APIs), establishing a secure data-sharing environment for seamless interactions among customers, AISPs, and PISPs (Doyle et al., 2023). Open APIs have replaced screen-scraping, providing a secure protocol for third-party data access without requiring login credentials for bank data extraction[2]. The emergence of OB is thus a result from the convergence of regulations, market forces, and technological advancements (Gozman et al., 2018).

APIs serve as essential gateways, facilitating interactions between diverse applications (Figure 1.2). In this shared environment, consumers retain full control over their data, with the right to

[1] https://www.openbanking.org.uk/

[2] https://www.ukfinance.org.uk/system/files/Frequently-Asked-Questions-on-PSD2-and-Open-Banking.pdf

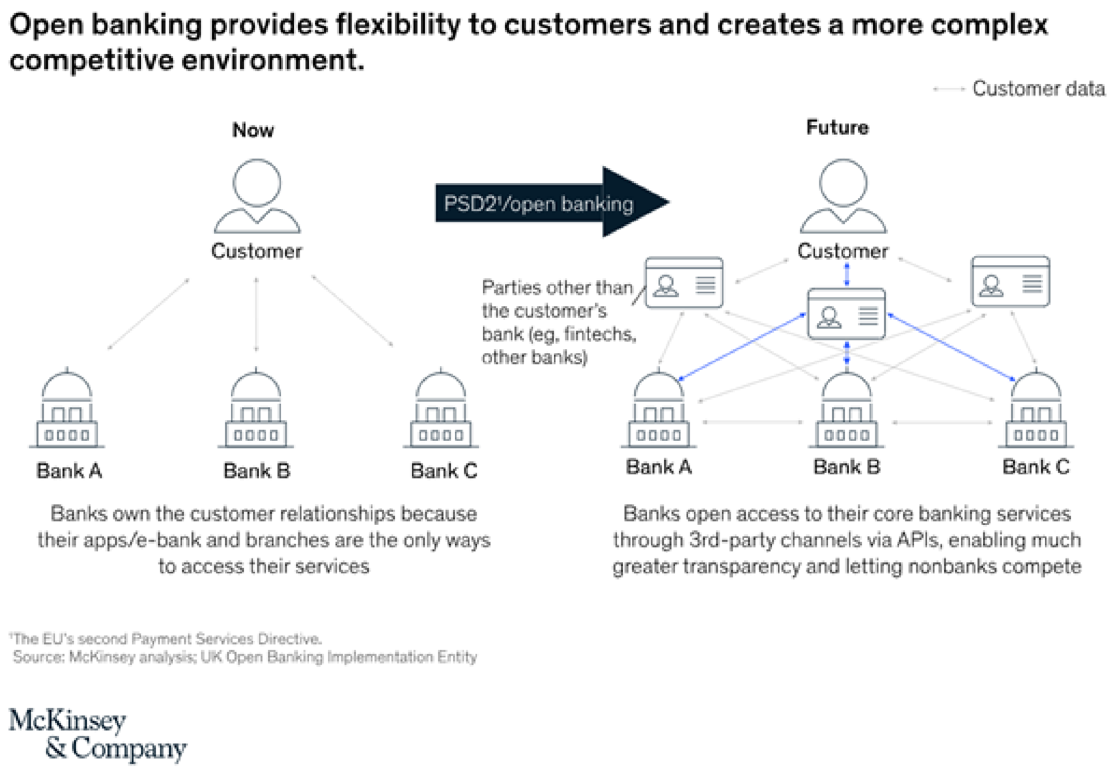


Figure 1.1: How Open Banking changes the traditional banking ecosystem

Source: https://www.mckinsey.com/industries/financial-services/our-insights/financial-services-unchained-the-ongoing-rise-of-open-financial-data

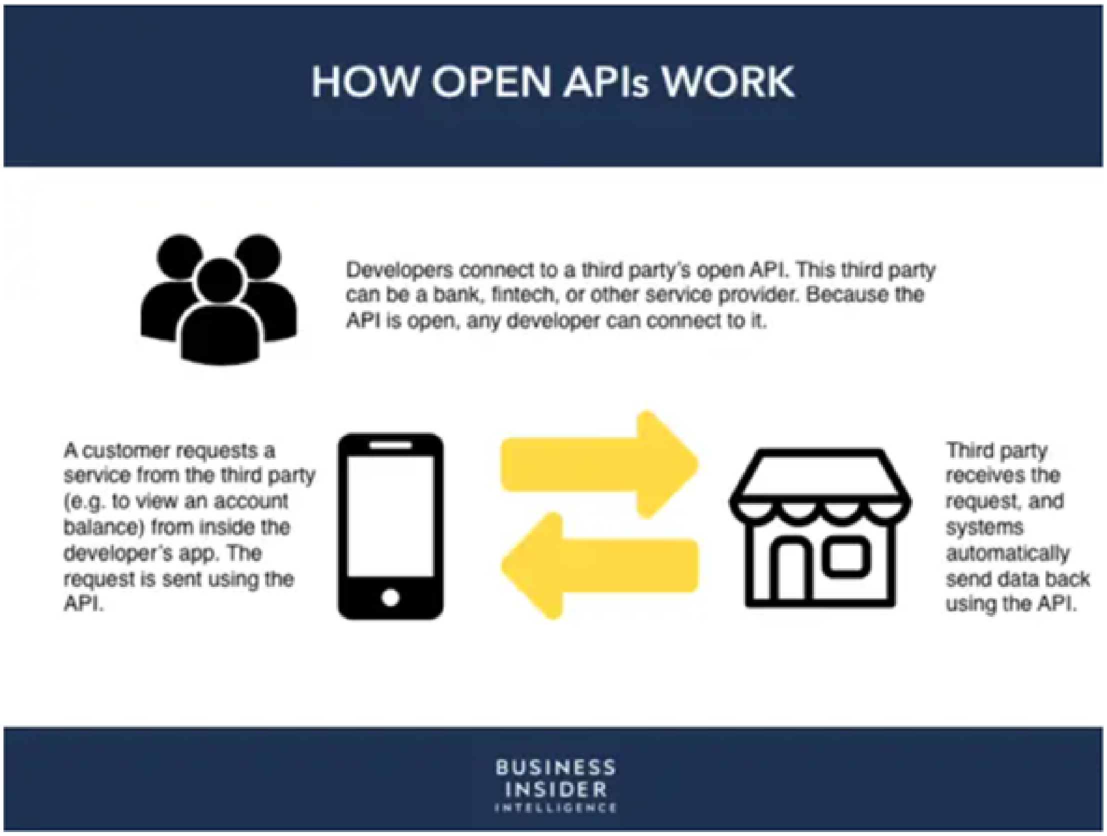


Figure 1.2: How Open Banking API works

Source: https://www.insiderintelligence.com/insights/open-banking-api-trends-explained

opt in or out of the ecosystem. Connecting third-party financial apps, consumers grant consent for the API gateway to initiate the exchange of data between TPPs and banks. APIs also serve as the critical connection for developers to innovate and deliver new solutions.

OB introduces a revolutionary concept of 'data democratisation', requiring banking entities to grant access to proprietary data under the owners' permission (Kassab & Laplante, 2022). This development empowers customers to leverage their data for personal benefit and fuels fintech innovations in the ecosystem. PISPs specialise in streamlining online transactions between entities, offering solutions tailored for scenarios such as substantial purchases involving multiple accounts or facilitating smooth money transfers across credit, checking, and savings accounts. Their primary focus is on enhancing the efficiency and ease of large-scale financial transactions. Conversely, AISPs concentrate on two aspects to enhance financial decision-making. Firstly, they provide a comprehensive dashboard summarising information from multiple accounts within a unified platform, enabling a holistic view of financial status for the analysis of income and expenditures to optimise cash flow[3].

The UK Open Banking (OB) has established a robust data sharing framework ensuring that consumers (data subjects) provide informed consent before sharing their financial data with data controllers (TPPs). This approach enables consumers to consent freely, without unnecessary behavioral manipulation. Under this framework, consent granted by consumers remains valid for 90 days, after which it must be renewed. Consumers have the option to revoke or modify their consent as needed. This system aligns with PSD2 regulations, which mandate strong customer authentication and secure API standards. Data controllers are required to facilitate easy viewing and revocation of consents through consumer-accessible mechanisms. The status of consents granted to different TPPs is displayed on dashboards, and consumers receive confirmation emails upon providing consent to TPPs. The UK consent model specifies key details for informed consent in data sharing: (i) identity of the data controller receiving rights, (ii) purpose of data use (e.g., payment/account details), (iii) duration of consent (typically 90 days), and (iv) procedures for consent expiration and revocation. The consent mechanism is implemented in three phases.

Phase 1 (**Consent**): Users are informed about the specific data requested and the purposes for which it will be used. Clear and detailed information about the time-bound nature of the permission is provided, and users retain the right to opt out. Initially, consumers are presented with consent details in the TPP interface, where they decide whether to grant consent. Subsequently, in the banking interface, consumers must authorize the sharing of consent details.
Phase 2 (**Authentication**): Users are informed about the consent process and the bank initiates authentication procedures to ensure the security of consumers' data.
Phase 3 (**Authorisation**): Users are presented with detailed consent requirements through the bank-user interface. They are prompted to approve or deny the TPP's request for access to their data. Responses from users are recorded and securely stored.

The AISPs extend their services to support borrowing decisions through an affordability assessment. This involves evaluating financial health indicators derived from categorized bank transactions (Equifax, 2021). Additionally, AISPs contribute to refining credit data by incorporating

[3]https://www.saltedge.com/use_cases/personal_finances

income and expenditure patterns for in-depth credit affordability checks (Experian, 2020b,a). They further play a crucial role in identity verifications during swift customer onboarding screening, leveraging bank accounts for seamless processes[4]. The implementation of Open Banking (OB) has not only fostered a competitive landscape within the financial service industry but has also paved the way for Open Finance, establishing an open data system across diverse economic sectors, including rent payments, insurance, wealth management, and more (Open Banking Implementation Entity, 2023a). This paradigm shift signifies a transformative evolution in the financial ecosystem, where data accessibility and innovation intersect to redefine how services are offered and consumed.

Despite the innovative potential, OB implementations pose certain threats. Empowering consumers with control over their data implies they have the discretion to refrain from revealing underperforming account details, potentially exploiting this discretion as a loophole to secure favourable deals (Podder, 2022). Fairness concerns arise as financial behaviour inferred from seemingly neutral OB data may reveal omitted personal information or protected characteristics (Kim et al., 2023). Addressing digital literacy gaps is crucial to prevent exclusion from digital banking services (Stefanelli & Manta, 2022). Additionally, traditional banks face challenges such as third-party disintermediation, security governance in open APIs, and technical adjustments to provide API functionalities to third parties (Gozman et al., 2018).

However, practitioners view OB more as an opportunity than a threat. Exploiting customer data for insights-based decision-making is crucial for success in the OB ecosystem. Tailoring propositions is essential for incumbent banks to maintain a competitive edge by offering personalised products and services to customers (Doyle et al., 2023).

1.2 Open Banking around the world

The evolution and adoption of OB practices have been significant on a global scale. Table 1.1 lists the regulatory frameworks and initiatives of OB across various regions[5]. These approaches can be categorised as either regulatory-driven or market-driven.

In the EU, OB practices have been predominantly shaped by regulatory frameworks[6]. The

[4]https://truelayer.com/blog/open-banking/open-banking-use-cases-guide/

[5]http://www.openbankingmap.com/

[6]https://noda.live/articles/open-banking-around-the-world

Table 1.1: Global Open Banking adaptation

Country	Regulation and Initiative
EU	Payment Service Directive 2 & 3 (PSD 2 & 3)
UK	Open Banking Implementation Entity (OBIE)
US	Initiatives by the Consumer Financial Protection Bureau (CFPB)
Australia	Consumer Data Right Act (CDR) & New Payments Platform (NPP)
Japan	2017 Banking Act Amendments
Singapore	API Exchange (APIX)
Mexico	The Fintech Law
Brazil	Open Banking system by Brazilian Central Bank
India	United Payments Interface (UPI)
Hong Kong	Open API framework

PSD2 framework stands out, aiming to streamline OB utilisation by mandating banks to grant TPPs access to consumers' bank account data, upon customer consent. PSD2 also lays the groundwork for enhancing digital payment security through Strong Customer Authentication (SCA) and facilitating data sharing between banks and TPPs via secure open APIs. Looking ahead, the proposed PSD3 in June 2023 aims to bolster the existing PSD2 framework. PSD3 seeks to expand OB regulations by introducing additional data sharing requirements, standardising APIs across a broader range of financial services, enhancing transparency for consumers regarding payment service terms and fees, and implementing more robust security measures to combat cyber threats. This revised edition of PSD promises to enhance OB implementation in the payments landscape, emphasising efficiency, security, and consumer protection. Mirroring the EU's approach, the UK translates PSD2 into domestic law, maintaining a regulatory stance on OB adoption[6]. The introduction of OBIE in the UK aims to establish standardized OB protocols, facilitating access to bank data among the major banks.

Similar regulatory-driven OB initiatives are underway in several Asia Pacific and Latin American countries[7]. In Australia, the proposed Consumer Data Right Act (CDR) shares similarities with PSD2 but extends data sharing beyond financial services to energy, telecommunications, and other sectors. Hong Kong's Open API framework outlines a phased approach to data sharing, starting with product information and gradually expanding to encompass transactional data and payment initiation services. In Brazil, a four-phase OB legislation mandates data sharing between financial

[7]https://thepaypers.com/expert-opinion/the-state-of-open-banking-in-apac-today

institutions, prioritising transparency and customer consent. The implementation progresses from sharing product information to personal and transactional data, ultimately encompassing currency exchange and investment information. Mexico's Fintech Law, though relatively recent, addresses OB by facilitating product information sharing and transactional data access, with further regulatory enhancements anticipated.

In the United States, OB adoption is primarily driven by market dynamics and industry practices[8]. The Financial Data Exchange (FDX) consortium recognises the commercial potential of leveraging bank data for informed decision-making, aligning with data sharing standards to promote OB adoption nationwide. The proposed Personal Financial Data Rights rules by the Consumer Financial Protection Bureau (CFPB) aim to foster market competition and decentralisation in banking, with a focus on safeguarding personal financial data against misuse and abuse.

Across Asia, market-driven OB initiatives are gaining momentum, despite the absence of formal regulations[2]. Singapore's Monetary Authority (MAS) introduces SGFInDex, a platform facilitating financial data consolidation for planning purposes. Japan's Banking Act revision mandates banks to develop open APIs, fostering collaboration with fintech firms to drive financial innovation. In India, the Unified Payments Interface (UPI) serves as a pivotal component of the India Stack, facilitating interoperable payments and laying the groundwork for OB advancement.

1.3 From Open Banking to Open Finance

The adoption of OB has not only fostered a competitive landscape within banking services but has also paved the way for the emergence of Open Finance (OF), representing an extension of OB principles[9]. Leveraging API technology for data sharing, OF creates a more interconnected and transparent financial ecosystem that transcends traditional banking boundaries. OF data encompasses a wide array of financial information, including bank transaction histories, investments, insurance policies, savings, mortgages, and pension details. This expansive OF ecosystem promises consumers enhanced access to their financial data, thereby catalysing further competition and innovation across the financial sector.

[8] https://www.consumerfinance.gov/about-us/newsroom/prepared-remarks-of-cfpb-director-rohit-chopra-on-the-proposed-personal-financial-data-rights-rule/

[9] https://www.openbanking.org.uk/open-finance/

OF offers numerous benefits to both consumers and businesses[10]. The interconnected nature of OF data empowers customers' decision-making processes by providing a comprehensive overview of their financial situation, enabling personalised insights and recommendations tailored to their unique goals and needs. OF also facilitates the consolidation of multiple investment accounts from various providers into a single platform, offering a holistic view of investment portfolios and enhancing management of profit and loss in trading activities. Additionally, the OF ecosystem enables easy comparison of financial products, such as personalised insurance options based on individual financial data and coverage requirements. Furthermore, OF streamlines pension management by facilitating transfers through APIs, bypassing cumbersome administrative procedures. Cross-border payments are also simplified within the OF framework, as it allows for the linking of accounts across different currencies and providers, thereby minimising currency conversion fees. Moreover, OF fosters innovation by enabling collaboration among financial service industries to co-create customer-centric solutions.

OB serves as the foundational pillar for the transition to OF. With the proliferation of OB connections in the UK and active development initiatives worldwide, OB has initiated a significant transformation in the banking sector, fueling innovations aimed at enhancing financial well-being through real-time insights into cash flow data. These OB advancements play a pivotal role in shaping the financial landscape within the OF ecosystem, ultimately contributing to the establishment of comprehensive financial footprints for consumers and businesses.

1.4 Motivation

Despite the significant progresses made towards establishing an open financial ecosystem, analytical exploration of Open Banking (OB) data, especially concerning consumer financial risk evaluation, remains an understudied area in the personal finance literature.

The analysis of financial status predominantly relies on subjective, self-reported information, which is susceptible to misinterpretation of the actual financial condition. This subjective data is often gathered through surveys or interviews to illustrate long-term (quarterly/yearly) income and spending behaviours (Financial Conduct Authority, 2021a). Additionally, it includes personal

[10] https://www.insurely.com/open-finance

perspectives on financial well-being, reflecting individuals' opinions on the burden of debt and money management (Consumer Financial Protection Bureau, 2017). While subjective data may offer insights into individuals' financial circumstances, it introduces biases when individuals intend to conceal their struggles during interactions with bank staff. Moreover, it may not accurately capture the timing of financial difficulties, leading to delayed intervention and support (Collard, 2011). In contrast, bank account transactions offer numerous advantages compared to self-reported information (Huang et al., 2018). The inherently objective nature of money flows minimizes the risk of dishonest exaggeration in both income and expenditure, providing a reliable foundation for evaluating financial activities. This objectivity complements subjective data, ensuring a more comprehensive understanding of individuals' financial behaviour while guarding against oversimplified portrayals. Consequently, bank account transactions serve as a primary source of verifiable financial records, underscoring the importance of developing methodologies to accurately quantify financial behaviour for robust risk management strategies.

Among the various account types linked to OB, including credit card accounts, mortgage accounts, and savings accounts, current account transactions offer distinctive properties for assessing consumer risk. Firstly, these transactions provide a detailed record of day-to-day financial activities, offering direct insights into consumer financial health. Unlike other account types that serve specific purposes such as loan repayments or savings, current account transactions encompass a wide range of activities, reflecting income and expenditure patterns in everyday personal finance. Secondly, current account transactions exhibit high granularity, on a daily basis, thereby capturing immediate changes in financial behaviour. Thirdly, connected current accounts provide a comprehensive view of the most recent twelve-month transactions, offering a real-time snapshot of financial status based on daily cash inflows and outflows. Given the unique characteristics of cash flow data, this thesis focuses specifically on current accounts connected through the OB API. A formalised framework for extracting multidimensional insights from cash flow transactions to assess financial risk remains an unexplored challenge in the OB literature.

OB-connected current account transactions have garnered attention in consumer risk assessment, with credit scoring standing as the predominant framework for evaluating credit default risk in the existing literature. Traditional credit scoring models have relied on applicant information or credit card repayment behaviours to forecast the likelihood of loan default (Thomas et al., 2017).

However, the advent of OB has prompted credit bureaus to explore the incorporation of current account cash flow data into credit scoring models. This integration allows for a more comprehensive representation of consumers' multidimensional financial behaviours and facilitates the utilisation of real-time cash flow data to reflect the most current financial circumstances. For instance, sudden changes such as job loss or taking out a new loan, which might not be immediately reflected in bureau records updated on a monthly basis, can now be captured more accurately (Equifax, 2021). Moreover, leveraging cash flow transactions in consumer risk scoring reduces reliance on traditional credit histories, thereby allowing thin file clients an opportunity for credit evaluation.

Notably, the inclusion of income and spending covariates in risk models has shown promising results in enhancing the predictive performance for delinquencies and defaults (Finlay, 2006; Khandani et al., 2010; Huang et al., 2018; Hjelkrem et al., 2022). Additionally, cash flow data exhibits a statistically significant correlation with financial well-being and distress, as shown by Gathergood & Guttman-Kenney (2016). However, while the integration of cash flow data holds immense potential in consumer risk modelling, its application remains focused within the credit domain, primarily aimed at predicting default (Khandani et al., 2010; Huang et al., 2018; Hjelkrem et al., 2022), assessing affordability (Finlay, 2006; Experian, 2020b,a), and evaluating over-indebtedness-related distress (Gathergood & Guttman-Kenney, 2016). Despite the prioritisation of risk assessment in OB, existing models primarily support credit risk decisions, neglecting exploration into other OB-specific use cases crucial for enhancing decision-making across various domains [11].

Establishing an OB-specific risk model presents a unique challenge: defining the target variable for assessment, an issue currently lacking clarity in OB literature. Having a specific risk variable tailored to OB is crucial for determining precisely what aspect of risk is under evaluation. While 'default' serves as a well-defined indicator for assessing consumer risk, its definition relies on credit histories from banks and bureaus, making it costly for third-party firms to acquire this information for developing their in-house risk assessment frameworks. In the context of personal finance, existing academic literature and industry practices lack a singular measure to capture the multi-faceted nature of financial challenges using bank account transactions. Distress or well-being is typically assessed in terms of overindebted scenarios, which encompass missed credit repayments

[11] The power of transactional data: https://experianacademy.com/blog/2023/04/14/the-power-of-transactional-data-from-open-banking/?fbclid=IwAR3yfADO8IAtSHDbU6Nzmly-fFJq1yxEWzV18mZgKMzHzAjqZPriGsYyOqk

and self-perceived stress related to debt clearance (Consumer Financial Protection Bureau, 2017; Gathergood & Guttman-Kenney, 2016; Gladstone et al., 2020). Kim et al. (2023) stand as the sole academic study proposing six binary financial vulnerability indicators using OB transactions; however, these dimensions are evaluated separately, lacking a comprehensive measure to assess risk levels holistically. Money management applications primarily focus on descriptive analyses of money inflows and outflows aggregated across multiple accounts, without providing a quantified score to represent the risks associated with financial patterns. The financial health index proposed by Equifax (2021) is computed from income and expenditure categories, overlooking the utilisation of account balances' adequacy in assessing risk. Financial institutions possess the flexibility to develop problem-specific risk frameworks using OB cash flow data, yet they encounter challenges in defining the target risk indicator to accommodate specific use cases.

The principal motivation behind this thesis is the need for new OB-based decision analytics that can bolster robust financial lives. OB emerges as a valuable resource for understanding consumers' financial circumstances, particularly during challenging economic conditions such as unprecedented events like the Covid-19 pandemic or escalating living costs. Instead of leveraging OB cash flow transactions for credit risk assessment, this thesis addresses three priority OB use cases [12], specifically in improving personal finance management, spotting consumer vulnerability, and detecting financial crime. The following paragraphs detail the unaddressed limitations for the use cases.

Use case 1: Improve personal finance management: Identifying financial difficulties is crucial for effective personal finance management, enhancing financial planning and overall well-being. Volatility in cash flow transactions, serves as a key risk indicator for assessing uncertainties in personal finance. Measuring and modelling this volatility is vital, as unstable behaviour poses challenges in accurately forecasting future earnings and affordability. Despite existing frameworks for depicting market uncertainty based on historical market closing prices (Xiao & Aydemir, 2007; Apergis et al., 2017; Barunik et al., 2016; Escobar-Anel et al., 2021), a dedicated volatility forecast model tailored to personal finance is yet to be developed. Developing an OB-specific cash flow volatility model requires exploring its connection with vulnerability. Vulnerability, indicating precursors of distress and diminished financial health compared to the average consumer, raises

[12] https://experianacademy.com/blog/2023/04/14/the-power-of-transactional-data-from-open-banking/?fbclid=IwAR3yfADO8IAtSHDbU6Nzmly-fFJq1yxEWzV18mZgKMzHzAjqZPriGsYyOqk

concerns regarding potential financial difficulties when driving volatility (Financial Conduct Authority, 2023a). Previous studies have mainly focused on investigating economic drivers of volatility through descriptive analyses, lacking predictive modelling (Mitchell, 2017; Trusts, 2017; Farrell & Greig, 2015, 2016). Additionally, while vulnerability-related studies have explored various indicators, the link between volatility and vulnerability remains unexplored (Noerhidajati et al., 2021; Šubová et al., 2021; Lin et al., 2021; Albacete & Lindner, 2013; Financial Conduct Authority, 2023a). Thus, a novel cash flow volatility forecast model leveraging OB current account transactions to explicitly reveal the relationship between vulnerability and volatility is yet to be developed.

Use case 2: Spotting consumer vulnerability: Identifying consumer vulnerability stands as another crucial use case within OB for practitioners. Regulatory emphasis on protecting vulnerable consumers ensures equitable access to financial products and support for underprivileged groups susceptible to varying degrees of distress. OB current account data captures dynamic cash flow transactions with daily granularity, facilitating timely alerts of financial vulnerability (FV) to mitigate the risk of financial deterioration. However, the temporal dimension of cash flow data remains overlooked in the personal finance context. Current practices typically assess poor financial status at specific time points (Gathergood & Guttman-Kenney, 2016; Kim et al., 2023; Harrison & Andreeva, 2021), neglecting to track changes in financial patterns in response to evolving financial struggles.

Moreover, the temporal effect of FV on potential distress remains an unstructured problem. Existing literature lacks a clear definition of objective risk measurement to quantify varying levels of financial troubles stemming from FV. While regulators have outlined key dimensions of FV, they do not specify methods for quantifying risks (Financial Conduct Authority, 2023a), and FV indicators derived from OB current account transactions exist as separate measures with no interconnectedness to gauge exposure risk to distress in a single metric (Kim et al., 2023). FV relates to the likelihood of encountering hardships (O'Connor et al., 2019), with lower (higher) FV attributes associated with lower (greater) levels of distress as financial circumstances evolve. Hence, modelling evolving FV to assess transitions across a spectrum of distress risk magnitudes is crucial for monitoring changes in FV and adopting a dynamic and proactive risk management approach. However, conceptualising temporal FV for risk assessment using OB cash flow transactions remains an unaddressed challenge in personal financial risk modelling.

Use case 3: Detecting financial crime: Know Your Customer (KYC) is a process by banks to obtain information about the customer to ensure bank services are not misused for illicit purposes. Utilising OB data for KYC risk assessment has received significant attention among practitioners as a crucial tool for preventing and controlling financial crimes. While OB connections serve as a means for identity verification, Podder (2022) has discussed the potential of OB to detect suspicious activities. However, this remains a conjecture without clear guidance on developing a KYC framework. In contrast to existing KYC tools tailored to specific financial crime contexts such as money laundering or credit card fraud (Ngai et al., 2011; Hilal et al., 2022), current accounts are vulnerable to various forms of crimes due to the multifunctional nature of cash flow transactions in managing daily finances. Thus, integrating OB cash flow data for data-driven KYC screening presents new challenges. Firstly, adapting transactional attribute representation to account for recurring transactions, like salary deposits and fixed merchant payments, is crucial to distinguish real suspicious patterns from normal behaviour effectively.

Secondly, in the absence of expert knowledge to label suspicious activities and without a specific crime context, searching for typical behaviour segments is important to define 'normality' for KYC profiling as well as guide the search for anomalous behaviour by scoring the deviation away from normal instances. Thirdly, scoring accounts with high anomaly risk necessitates complementary explanations to validate potential suspicious patterns. Existing KYC models often rely on subjective expert judgments to identify peculiar patterns (Dreżewski et al., 2012; Nian et al., 2016), rather than utilising model explainability to extract unique characteristics and validate flagged abnormalities in a data-driven manner. A KYC mechanism specifically tailored to OB cash flow transactions for automated bank account screening has yet to be formally investigated.

1.5 Objectives

The key question addressed in this thesis is how to leverage insights from OB data for informed financial decision-making. Our objective is to formulate three OB-based risk modelling frameworks, each dedicated to evaluating distinct aspects of financial risks. Each risk model entails the computation of transactional patterns as input attributes and the establishment of a specific measurement to define the target risk indicator. With a focus on three dimensions of financial risks, this thesis has the following specific objectives.

Financial Instability

- Define an account-level volatility index to quantify the degree of uncertainty in cash flow transactions.
- Predict account financial volatility based on past transactional patterns associated with vulnerability.
- Compare the predictive performance and interpretability of statistical and machine learning models in assessing account volatility.
- Alert potential distress by interpreting the relationship between vulnerability-related transactions and volatility.

Financial Vulnerability

- Define distress risk states as representations of varying degrees of financial difficulties in cash flow transactions.
- Predict transitions between distress risk states using vulnerable transactional behavior.
- Monitor the evolution of distress risk states over time.
- Examine the vulnerability drivers of financial transitions.

Financial Crime

- Identify distinctive profiles of bank accounts as baselines for understanding typical behaviour.
- Derive outlier scores as proxies for accounts' exposure risk to financial crime by measuring deviations from the norm.
- Escalate potential suspicious accounts by examining the presence of anomalous characteristics associated with dubious patterns.

1.6 Contribution

The primary contribution of this thesis lies in the methodological development for deploying statistical and machine learning (ML) techniques in three innovative OB use cases. Each chapter addresses a specific financial risk derived from OB cash flow transactions, providing empirical findings that carry practical implications for bank practitioners engaged in data-driven analytical decision-making.

In Chapter 2, three key contributions are made. Firstly, the chapter introduces a novel approach to predict account-level volatility using past transactional patterns. This pioneering work customises a volatility forecast model within the context of personal finance, leveraging bank transactions. The incorporation of volatility, alongside an exploration of underlying vulnerability, emerges as a key factor in detecting financial struggles. A composite index of account volatility is defined to represent uncertainties within cash flow transactions. Multiple vulnerability indicators are formulated and combined into a single measure, facilitating analytical account risk ranking in the OB domain. Secondly, the benchmark experiment comparing statistical and machine learning techniques provides a foundational reference for future OB-based risk models. Thirdly, the interpretation of non-linear transactional patterns delves into varying degrees of risky volatile behaviour. Beyond a simple linear relationship, the analysis highlights complex non-linear impacts of vulnerability, outlining different degrees of risk associated with financial hardships. This nuanced connection between volatility and vulnerability categorises accounts into distinct target risk groups, enhancing effective analysis for tailored interventions.

In Chapter 3, three notable contributions are presented. Firstly, the chapter pioneers the use of a latent Markov model to establish a dynamic financial vulnerability risk assessment framework for OB. This model's parameterisation enables a data-driven quantification of the varying stages of distress risk states, allowing practitioners to rank accounts' implicit risk of facing distress over time. Secondly, predicting the likelihood of transitioning between distress states from vulnerable-related transactional patterns facilitates the timely detection of financial improvements or deteriorations. The inclusion of time-varying vulnerable-related transactional patterns in the model estimation step enables the projection of future state transitions based on previous vulnerabilities. Thirdly, the interpretation of the relationship between financial vulnerability and state transitions reveals account-specific factors influencing shifts in financial status. This analysis aids in monitoring the tenacity of financial vulnerability, providing early warning signals of detrimental deterioration. The characteristics associated with dynamic financial states offer insights into distinguishing between transient and permanent vulnerabilities, contributing to consumer protection strategies.

In Chapter 4, four significant contributions are outlined. Firstly, the chapter designs a novel mechanism to evaluate accounts' exposure risk to financial crime, flagging suspicious accounts based on cash flow behaviour. This assessment framework introduces an algorithmic-driven approach to screen OB bank accounts during the customer onboarding due diligence process. Secondly, the proposal to represent cash flow behaviour with RFMP attributes (Recency, Frequency, Monetary, Persistence) serves as a fundamental component for assessing anomalous accounts. The extended persistence dimension effectively captures subtle regularities, such as persistent but low-frequency occurrences, enabling the recognition of habitual practices or implicit illicit transactions despite their infrequent appearance. Additionally, it identifies bursty activities, characterised by irregularly large counts or amounts, serving as an alert for spikes in scattered illegal patterns. Without this extended persistence dimension, these anomalies might have been misconstrued when relying solely on the RFM dimensions. Thirdly, in the absence of a 'suspicious' label, the identification of profiles of typical account behaviour serves as a baseline to distinguish anomalies. These account profiles enhance the understanding of consumer patterns, enabling financial firms to offer better product recommendations for new clients. Fourthly, an outlier detection scheme is derived to score accounts' anomaly risks as a proxy for exposure to financial crimes. The escalation of high anomaly risk accounts as 'suspicious' is justified by their association with RFMP attributes depicting peculiar activities. The proposed scheme improves efficiency for due diligence checks by focusing directly on high-risk (outlier) accounts to flag potentially suspicious accounts.

1.7 Thesis structure

This thesis is organised in the following structure. After the Introduction, Chapters 2, 3 and 4 will showcase the novel OB frameworks to address three different financial risks. Each chapter represents an independent work detailing the framework design with a self-contained literature review, methodology notations, and empirical analysis.

In Chapter 2, we forecast a vulnerability-driven volatility score to indicate potential financial difficulties. Section 2.2 reviews existing models for objective risk assessment in the financial domain, setting the stage for the motivation behind forecasting volatility in personal finance. Sections 2.3.1 through 2.3.5 describe the experiment setup, the definition of the volatility index as the target variable, the formulation of vulnerability covariates from transactional patterns, the techniques used to compute the composite index, and the statistical and machine learning models employed in the experiment. Section 2.4 analyses the reported results, and Section 2.5 concludes the findings in this chapter.

Chapter 3 introduces a dynamic distress risk model to assess the evolution of financial transitions. Section 3.2 reviews previous works on dynamic risk modelling approaches used in consumer risk management. Sections 3.3.1 to 3.3.4 detail the model estimation procedure to outline the implicit distress risk estimation from observed account balances, the formulation of vulnerability covariates measured from transactional records, and the estimation of the latent distribution capturing financial risk evolution. Section 3.4 discusses the model outputs with the interpretations of distress risk states, movements across financial states, and strategies to monitor transient and permanent vulnerabilities. Section 3.5 summarises the key findings in the chapter.

Chapter 4 identifies potential suspicious accounts by detecting outliers from cash flow attributes and provides justifications based on peculiar transactional behaviour. Section 4.2 introduces the background and previous literature on modelling financial crimes. Sections 4.3.1 to 4.3.3 outline the representation of cash flow attributes, the clustering setup, the word analysis procedure to profile clusters, and the outlier detection scheme to assess accounts' anomaly risk and examine dubious patterns. Section 4.4 reports findings on the profiles of typical account behaviours and the identified potential suspicious account activities followed by Section 4.5 to conclude the chapter.

Finally, Chapter 5 concludes the findings from each chapter, summarising the contributions in practical applications.

Note on the dataset used

All chapters within this thesis utilise OB current account transaction data retrieved from the OB API, as outlined in Table 1, extracted at different time points. The sharing of bank account data to third-party providers occurs only upon consumers' explicit consent. Within the scope of this thesis, our focus is on bank transactions connected through the UK AISPs. Each connection to a single account provides access to transaction data spanning up to 12 months. Consent is subject to renewal every 90 days, granting consumers the authority to either continue connecting recent transactions or opt out of further data sharing. The transactional records contain vital details, including date, brief description, credit/debit indicator, transaction amount, income/expenditure category label, and the account balance post-transaction. A comprehensive overview of the raw transactional records is summarised in Table 1.2.

Table 1.2: Description of raw transactional records

Variable	Description
AccountId	Account identifier
TransactionId	Transaction identifier
Transaction Date	Date when the transaction occurs
Description	Details of the transaction's counterparty
Transaction Type	Type of transaction (Credit: inflow/ Debit: outflow)
Amount	Transaction amount
Category	Income or expenditure categories*
End balance	Remaining balance after the transaction**

*assigned by the categorisation engine of the data provider

**calculated based on the balance amount in the latest day of the transactional record

During each extraction, both currently connected accounts and those with expired connections but enabled consent for data sharing are retrieved. Consequently, the timeframe of the extracted transactions spans various periods, rather than being limited to the most recent 12 months at the time of data extraction. Accounts featured in Chapters 2 and 4 are extracted in June 2021, while those in Chapter 3 are extracted in October 2022.

The OB bank account data is accessed securely through an online platform provided by DirectID[13]. This process involves three stages of data processing: retrieval, cleaning, and enrichment. In the retrieval stage, raw transaction data is extracted from the SQL database and stored in cloud storage before being imported into analytics platforms such as Azure ML Studio and Azure Databricks, which are utilised by DirectID. To clean the data, the raw transactions are imported into a data frame format resembling a table structure with rows representing transactions and columns containing details, as illustrated in Table 1.2. The descriptive statistics of the raw transactions are reported in Table 1.3.

We notice that missing values in the data frame arise from empty descriptions. These lines of transactions are kept in the experiment since we rely on other information (e.g., date, amount, transaction type, category), to obtain the data features. Then, duplicated accounts with identical transaction sets but different IDs, are removed. These duplicated accounts are created for testing purposes (by the engineering team) and do not indicate a different account. To refine the selection of accounts for the research, we have established two additional criteria. First, only accounts that have been connected within the last two years are considered. This is because the majority of accounts are connected within this recent period. Accounts connected between 2017 and 2019 are excluded as their relevance has diminished, and their limited sample size is deemed insufficient for representing contemporary financial behaviours in the OB context. Second, we include only accounts with at least twelve months of activity to ensure a sufficient number of data points for analysing financial behaviours and developing the prediction model. As a result, the final dataset comprises 11,150 accounts from raw data set A and 19,447 accounts from raw data set B.

The raw transactions data is then used for data enrichment i.e., to formulate features for the representation of financial behaviour. The enriched features depict income, expenditure and balance patterns, and they are designed according to the specific research objectives. Details on the data features are described in the coming Chapters 2, 3 and 4.

Due to the voluntary consent nature of the OB framework, self-selection bias could occur. Considering the primary interest of this thesis in exploring the potential insights from participants

[13] rebranded to Atto in 2024

Table 1.3: Descriptive statistics of the raw transactions field

Variable	Raw data A (Extracted in June 2021)	Raw data B (Extracted in October 2022)
AccountId	14170 accounts	29329 accounts
TransactionId	9960823 transactions	19853049 transactions
Transaction Date	Earliest: Jan 2018	Earliest: Aug 2017
	Latest: June 2021	Latest: October 2022
	number of accounts per year	number of accounts per year
	year 2018: 2	year 2017: 1
	year 2019: 124	year 2018: 2
	year 2020: 13621	year 2019: 6
	year 2021: 13702	year 2020: 33
		year 2021: 27174
		year 2022: 28297
Description	133 blank descriptions	56 blank descriptions
Transaction Type	Debit transactions: 779849	Debit transactions: 15801470
	Credit transactions: 9882874	Credit transactions: 4051579
Amount	Minimum: 0	Minimum: 0
	Mean: 4.046e+06	Mean: 724.14
	Standard deviation: 6.337e+09	Standard deviation: 4.565e+04
	First quartile: 5.95	First quartile: 6.10
	Median: 15.50	Median: 18.99
	Third quartile: 50	Third quartile: 74.16
	Maximum: 1.00e+13	Maximum: 1.23e+08
Category*	71 categories	65 categories
End balance**	Minimum: -1.00e+13	Minimum: -2.546e+07
	Mean: -2.389e+09	Mean: 1.249+04
	Standard deviation: 1.529e+11	Standard deviation: 1.944e+05
	First quartile: 5.81	First quartile: 8.487
	Median: 226.40	Median: 366.21
	Third quartile: 1555.70	Third quartile: 269.30
	Maximum: 2.000e+08	Maximum: 7.670e+06

*Refer to Appendix A for the detailed descriptive statistics of the categories

**Descriptive statistics is the summary of accounts' available balance i.e., end balance at the last day of transaction within the OB framework, selection bias is expected to have minimal implications in this research. However, when applying OB insights to risk assessment tasks (such as credit portfolio management or insurance premium pricing etc.), selection bias becomes a significant concern. This is because portfolios may consist of both OB and non-OB participants, potentially introducing biases in the conclusions drawn. Specifically, self-selection bias could occur where younger or more tech-savvy consumers are more likely to participate in OB due to their affinity for innovative solutions. As a result, conclusions drawn from OB data alone may not be representative of the broader population, impacting portfolio risk management strategies.

2 Volatility prediction via interpretable models through the lens of vulnerability: an application to personal account transactions

A part of this chapter has been published as a conference paper: Goh, R. Y., Andreeva, G., & Cao, Y. (2022, May). Predicting financial volatility from personal transactional data. In *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)* (pp. 1-8). IEEE.

2.1 Introduction

The EU Second Payment Services Directive (PSD2), also known as Open Banking (OB) in the UK, has been a game changer in the financial services sector. The Directive has enabled data sharing to regulated third-party financial service providers (TPP) based on consumers' consent. This initiative provided a unique opportunity to get an objective view of consumers' financial situation based on their account transactions and to get nuanced insights into their behaviour. These insights have improved financial decisions, helped individuals to practise better money management and expanded access to financial services (Open Banking Implementation Entity, 2023b). The Directive has also brought new challenges stemming from this new source of voluminous and unstructured data. One of such challenges is the need of new metrics, summarising various aspects of financial behaviour, that can be used in financial risk management. Finance literature on securities market views volatility as an important measure of risk and uncertainty from the fluctuations in historical daily market closing prices of financial instruments to mitigate investment losses (Xiao & Aydemir, 2007; Apergis et al., 2017; Barunik et al., 2016; Escobar-Anel et al., 2021). However, in the context of personal bank account transactions the concept of volatility is much less developed, and this chapter addresses this gap. Measuring and modelling volatility is important since it directly affects the ability of financial institutions to make accurate predictions of customers' behaviour, thus impacting the quality of financial planning and decisions.

This chapter examines the development of a volatility measure for bank account cash flow transactions, drawing on two key analogies from finance literature. First, akin to the concept of asset price volatility, where fluctuations in asset prices over time can impact financial stability, volatility in cash flow transactions can similarly reflect financial instability (International Monetary Fund, 2003). Asset price volatility, while inherent and not necessarily negative, depends on changes in underlying conditions (Analogy 1). Likewise, volatility in cash flows, driven by shifts in financial behaviour in response to challenging situations, may indicate financial instability.

In finance, asset price volatility is traditionally quantified by the standard deviation of the logarithmic changes in asset prices, assessed across various time frequencies—daily, monthly, quarterly, or annually - to evaluate short- or long-term investment risks (Granger & Poon, 2001) (Analogy 2). This chapter applies a comparable approach by using the coefficient of variation (CV) to calculate quarterly account volatility. This measure serves two key purposes: (i) it allows for the comparison of time series variation across different accounts, even when their means differ significantly, by expressing the standard deviation relative to the mean, and (ii) it aids in evaluating mid-term risk by highlighting significant fluctuations in cash flow over a quarter, which may indicate persistent financial uncertainty and potential financial struggles.

Yet, there is an additional concept of vulnerability that we explore in this chapter and its connection to volatility. Previous studies showed that volatility of personal finance can arise because of poor financial health and wellbeing. Volatility can be connected with poor budget planning, failures to make ends meet and difficulties to save money (Farrell & Greig, 2015, 2016). Under insufficient financial buffer situations, a mismatch between income and expenditure volatility, such that unstable income fails to cover irregular spending is associated with deteriorating financial wellbeing (Gladstone et al., 2020). Therefore, volatility is seen by many scholars as being driven by implicit risks leading to distress.

Developing a volatility forecast model for effective personal risk assessment has several challenges. First, unlike stock price volatility which is measured from the fluctuations in daily market closing prices (Xiao & Aydemir, 2007; Poon & Granger, 2003), there is no defined measure to quantify personal finance volatility. Second, the well-established framework for equity markets volatility forecast is not directly applicable to personal finance. The asset price time series has stylised facts i.e., common statistical properties across different financial markets that enables the models to learn the underlying volatility process from these characteristics but personal financial behaviour has completely different properties which are subject-dependent instead of driven by the securities market (Engle & Patton, 2001). Third, while higher stock price volatility directly indicates high market instability risk of equity, risk evaluation with personal volatility is not equally straightforward. High personal volatility could be a low distress risk when it is driven by non-harmful transactions such as income bonuses, tax refunds, or seasonal holiday spending (Farrell & Greig, 2015, 2016). Such additional complexity of personal volatility forecast then leads to the next challenge which requires emphasis to interpret the harmful vulnerability-related drivers of volatility apart from sole relation to past volatilities as in equity forecast.

Volatility under vulnerable circumstances should raise concerns because vulnerability is a precarious state of being more susceptible to distress. In the household context, high vulnerability and high level of household debt are the precursors of financial crisis (Noerhidajati et al., 2021; Šubová et al., 2021). Vulnerable consumers are more prone to distress because they are more likely to be excluded from financial services, and eventually suffer from low financial health as compared to an average consumer (Financial Conduct Authority, 2015). Thus, investigating the intertwined effect of volatility and vulnerability and assessing whether the cash flow uncertainty is related to a vulnerability-driven financial instability is essential for effective risk management.

To address the abovementioned problems, in this chapter, we aim to develop a personal account-level volatility model to forecast future volatility from past transactional behaviour. We define bank account volatility as the degree of fluctuations in cash flow activities over a quarter. We first propose a personal account volatility index which combines fluctuations in inflow, outflow, and balance to evaluate the variations in OB cash flow transactions. Then, we use the defined volatility index as the target variable and we include as predictors not only the past volatilities, but also past transactional behaviour that relate to various vulnerable conditions. We further construct a personal vulnerability index that summarises vulnerable characteristics in a single measure as the additional predictor to rank accounts' risk and identify target groups. Since OB is a relatively new field, we set up a benchmark experiment to compare the most popular statistical and state-of-the-art machine learning (ML) models in terms of predictive accuracy and explainability.

This chapter contributes to the personal finance and risk modelling literature in the following ways. First, we define a financial volatility composite index to measure the variations from all aspects available in transactional data i.e., inflow, outflow, and balance. Second, we examine the way to forecast the future volatility of existing or new personal accounts, based on previous transactional patterns. We quantify a wide range of vulnerability indicators and combine them into a single measure. This provides innovative analytical support to account risk management in the area of OB. Third, we benchmark several algorithms and show that spline regression is competitive with ML models in account volatility predictions because it captures non-linearities while preserving the intrinsic interpretability. The interpretations of non-linear transactional patterns provide insights into varying levels of harmful volatile behaviour. We also demonstrate how ML can be interpreted, and compare ML interpretations with those of the spline regressions. Lastly, we illustrate how target risk groups can be identified from the connection between volatility and vulnerability to help financial firms in setting up effective strategies for financial planning and consumer protection.

This chapter has the following structure. The next section 2.2 reviews related works. Section 2.3 outlines the methodology with data description, definitions of the target variable, and covariates, the models included in the benchmark experiment, and the model explainability approaches. Section 2.4 reports and discusses the results. We analyse the components in financial volatility, compare the predictive performance, interpret the non-monotonic patterns from spline regression's marginal effects to highlight varying degrees of risky volatile behaviours and identify target risk groups, as well as compare the explanability between spline regression and the best performing machine learning model. Finally, section 2.5 concludes the study.

2.2 Literature Review

Open Banking (OB) stands as a groundbreaking financial innovation, facilitating the secure sharing of users' bank account data with authorised third-party providers (TPPs). The increasing adoption of OB among digitally empowered consumers has led to a significant surge in connected bank transactions. This wealth of OB data, a direct reflection of evolving financial behaviors amidst dynamic economic conditions, has become an invaluable resource. However, leveraging this unstructured cash flow data for meaningful insights poses a new challenge. The current paradigm primarily relies on computing various metrics to assess financial risk. Fintech companies, for instance, aggregate diverse bank account balances, income details, and spending analyses, presenting them in dashboard summaries aimed at guiding financial planning (Open Banking Implementation Entity, 2023b). While this approach provides valuable information, its reliance on separate metrics for disposable income, income shocks, stability, and expenditure composition hinders a comprehensive assessment of risk.

To address this limitation, credit bureaus have taken steps to consolidate multiple metrics into a singular measure that encapsulates the overall health of OB transactions. For instance, Equifax (2021) introduce a durability index, derived from intersecting multiple capacity measures covering earnings, spending power, and credit utilisation. Additionally, Equifax (2022) devise a financial health index through ranking analysis of categorised inflows and outflows. Despite the efficiency of these single metrics in providing a straightforward assessment, they often overlook the inherent instability arising from the fluctuation of cash flows over time. The need for a novel metric that captures the progression of cash flows for effective financial risk management remains unmet. Developing a comprehensive measure that considers not only income and expenditure patterns but also addresses the inherent volatility in cash flows over time stands as a pivotal, yet unaddressed, challenge in the OB literature.

2.2.1 Risk metric for cash flow

Volatility, characterised by changes and unpredictability, is an essential aspect in the contemporary financial landscape. Factors such as zero-hour contracts, the gig economy, COVID-19 economic disruptions, and escalating inflations continue to amplify volatility, making it a crucial risk indicator for assessing the inherent fragility in financial progress. The essence of volatility lies in its role as a core measure of uncertainty, serving to identify implicit risks embedded in personal financial behaviour, given its close correlation to vulnerability.

Numerous studies have illuminated the adverse effects of income volatility on financial outcomes. A survey focusing on non-elderly households (Bania & Leete (2009)) discover that lower-income households experience the highest monthly income volatility, exposing them to episodic economic deprivation and an increased likelihood of slipping into poverty. Long-term tracking of year-to-year earnings in American households (Trusts (2017)) reveals income volatility as a prevalent phenomenon across demographics, with heightened significance for Hispanic and high school diploma populations. In the face of dramatic income fluctuations, stable households are better equipped to establish emergency funds, ensuring resilience against unexpected needs. However, nearly half of American households experience destabilised budgets due to volatile income (Aspen Financial Security Program (2016)). This inconsistency, stemming from irregular work schedules, manifests broader adverse effects such as stress, dissatisfaction, challenges in finding childcare, and a decline in the labour market.

Monthly income fluctuations, particularly large swings, signify vulnerable circumstances where consumers struggle to save, face difficulties in bill payments, encounter limited credit access, and grapple with unexpected expenses (Chen et al. (2015)). Examining income changes in the US household during the COVID-19 pandemic, Lin et al. (2021) observe improvements in rainy-day fund preparation but noted financial stress indicators, such as late mortgage payments or overdrawing, during income dips caused by furloughs. While existing studies often focus on the correlational relationship between income volatility and poor financial conditions, Peetz et al. (2021) identify a causal impact, revealing that participants with unstable income exhibit lower internal locus of economic control, leading to poorer financial planning decisions.

Beyond income volatility, there is a growing recognition of the need to consider volatility in expenses. Irregular income's adverse effects cascade into delayed and disrupted consumption, elevating the risk of poverty (Aspen Financial Security Program (2016)). The misalignment of income and expense volatility introduces anxiety, hindering individuals as they navigate financial ups and downs (Tescher & Schneider (2015)). Monthly volatility in low and moderate-income households (Hannagan & Morduch (2015)) underscores that closely tracking spending to income reflects poor financial conditions, whereas a greater misalignment provides more flexibility in spending options.

To capture more accurate cash flow timing, practitioners are increasingly relying on tracking monthly volatility through bank account transactions rather than relying on self-reported survey information (Farrell & Greig (2015)). While most studies emphasise the negative effects of volatility, Gladstone et al. (2020) contradict expectations, revealing that only balance volatility exhibits a significant negative relationship with financial well-being, while income volatility is non-significant, and expenditure volatility is positively correlated. This nuanced exploration, considering all three volatilities concurrently, sheds light on two distinct explanations. Firstly, when spending volatility closely mirrors income variations, individuals exhibit a high marginal propensity to consume income, consequently leading to a low account volatility and, in turn, indicating high wellbeing. Conversely, in scenarios where the timing of spending and income diverges significantly, high balance volatility becomes indicative of low wellbeing. These studies underscore the importance of jointly considering all three components—income, expenditure, and financial buffer—for a comprehensive understanding of personal financial risk.

Despite these insightful studies, a unified composite index encapsulating the variation in various aspects of financial behaviour for effective financial risk management remains an unmet challenge in the current landscape. Addressing this gap is crucial for developing a holistic understanding of personal financial risk.

2.2.2 Volatility forecast model

Measuring and modelling volatility holds paramount importance in assessing uncertainties stemming from fluctuating cash flows, enabling accurate predictions of consumer behaviour and facilitating effective financial planning. While volatility forecasting is a well-established tool in finance literature for gauging uncertainty in market securities and minimising investment loss, this concept has not garnered sufficient attention in the field of personal bank account transactions.

Previous studies on volatility (as reviewed in Section 2.2.1) have predominantly centered on descriptive analyses to elucidate the economic drivers of volatility, rather than focusing on predicting volatility for robust risk assessment. The primary drivers identified include irregular work schedules with sporadic hours (Mitchell, 2017), changes in household composition or illness (Trusts, 2017), payment schedules with five Fridays or seasonal shopping patterns (Farrell & Greig, 2015), and job transitions or alternative income from gig work i.e., irregular income-earning activities due to flexible non-standard working hours in the labour market heavily relying on temporary positions filled by freelancer, online platform worker or independent contractors (Farrell & Greig, 2016). Notably, Gladstone et al. (2020) employ a multiple regression model, incorporating inflow, outflow, and balance volatilities as predictors of financial well-being. However, their work is confined to exploring the link between bank account fluctuations and financial health, rather than forecasting volatility for risk assessment.

Developing a volatility forecast model presents a challenge in personal finance literature due to several limitations. Firstly, the established volatility forecast models in finance, designed for financial assets, lack a direct application to personal finance, as personal finance volatility lacks defined measures akin to those in financial assets. While financial assets' volatility is typically measured using standard deviation or conditional variance estimated with maximum likelihood procedures

(Xiao & Aydemir, 2007), personal financial behaviour exhibits distinct, subject-dependent properties. Secondly, personal financial behaviour does not adhere to stylised facts commonly found in asset price time series, such as fat-tail distribution, clustering of large and small moves, mean reversion in volatility, or asymmetric impact from positive and negative innovations (Poon & Granger, 2003; Engle & Patton, 2001). Thirdly, while higher stock price volatility directly signals high market instability risk, interpreting personal volatility is nuanced. Personal volatility can indicate low distress risk when driven by non-harmful transactions, such as income bonuses, tax refunds, or seasonal holiday spending (Farrell & Greig, 2015, 2016). Therefore, forecasting personal finance volatility requires scrutiny of underlying financial circumstances for effective risk evaluation.

While emphasising volatility under vulnerability is crucial for signalling potential distress, the literature on their interconnected relationship remains scarce. Furthermore, there is a lack of a universally accepted vulnerability measure in existing research. Some studies propose a household vulnerability index based on various distress conditions. For instance, Lin et al. (2021) conceptualise vulnerability as a life cycle measure, capturing the decline in household living standards by examining variations in income levels and the consumption-to-income ratio following the death of a spouse. Meanwhile, Albacete & Lindner (2013) calculate debt burden ratios and an income-to-expenditure ratio as vulnerability indicators to assess whether potentially vulnerable households pose a threat to the Austrian financial market. In another approach, Šubová et al. (2021) measures household vulnerability using total monthly debt payments and household gross monthly income, subsequently investigating the determinants of vulnerability. Additionally, Noerhidajati et al. (2021) formulate a vulnerability indicator by integrating objective information on debt, arrears, budgeting skills, resilience, and social exclusion, along with subjective perceptions related to the ability to sustain living standards, manage debt, handle shocks, and adapt to changes in income conditions.

Despite these quantitative measures of vulnerability within the household context, they are mainly based on self-reported survey information, which may introduce biases and inaccuracies. Offering guidance, Financial Conduct Authority (2023a) outlines the four key characteristics of vulnerable consumers—low resilience, low capability, poor health, and life events. However, there is no explicit suggestion on how to combine these characteristics. Notably, Kim et al. (2023) is the only study quantifying consumer vulnerability from OB transactions, although the binary indicators employed in this study do not gauge vulnerability levels on a numerical scale. This limited representation underscores the need for a more nuanced and comprehensive approach to quantifying and measuring vulnerability, particularly in the context of financial transactions and behaviours.

Moreover, the connection between vulnerability variables (representing multiple transactional patterns) and volatility remains an unexplored area in risk assessment modelling within personal finance literature. A comprehensive vulnerability index, amalgamating various transactional behaviours, is yet to be developed, posing a significant gap in current research. Addressing this gap would not only enhance our understanding of personal financial risk but also pave the way for more effective risk assessment and mitigation strategies.

2.2.3 Considerations of risk assessment modelling

Estimating risks derived from past behavior is crucial for mitigating risk and minimising losses (Apergis et al. (2017); Thomas et al. (2017); Xiao & Aydemir (2007)). Surprisingly, the personal finance literature lacks a developed risk model that incorporates account volatility predicted from vulnerability, presenting a noteworthy gap in the existing research. In the relatively nascent field of OB, the absence of baseline experiments has made the choice of modelling techniques challenging. To address this, we turn to credit scoring references for initial insights into setting up an OB-based risk model. Credit scorecards, with a long history of practical success in financial risk assessment (Thomas et al., 2017), provide a foundation for structuring the volatility model in personal financial risk evaluation.

Developing a volatility predictive model for risk assessment involves two key considerations. Firstly, the model should be capable of capturing complex non-linearities, as the relationship between volatility and vulnerability is intricate. Spline regression emerges as a promising model due to its demonstrated flexibility in handling irregular shapes. Notably, Giordani et al. (2014) augment logit with quadratic natural splines, specifically the truncated power basis, revealing non-linear and non-monotonic relationships between financial ratios and firm bankruptcy. This approach unveils credit rationing scenarios and highlighted cash flow risks associated with firms boasting high-earning ratios. Other studies, such as Luo et al. (2016), incorporate the cubic b-spline basis into discrete time survival models, enhancing predictive accuracy for credit card default hazard. Djeundje & Crook (2019a) further demonstrate the ability of spline basis functions to identify granular changes in impact magnitude compared to parametric survival models, revealing hidden patterns as credit card loan accounts age. In Djeundje & Crook (2019b), a generalised additive model with penalised spline is utilised to capture inherent patterns through a combination of smooth functions representing the risk factors for retail loan default. This approach allows for the representation of non-linear shapes in the data, revealing that identical applicants might face denial or acceptance when the linearity assumption is rigidly enforced. Spline regression, therefore, stands out as a viable alternative for constructing a personal volatility risk assessment model, providing insights into complex risk relationships and improving predictive performance for cost savings.

Beyond the statistical spline regression approach, machine learning, particularly tree ensemble and deep learning techniques, has gained prominence in modelling complex credit behaviour. Although there is no consensus on the best credit scoring model, benchmark studies suggest tree ensembles and deep neural networks as the new state-of-the-art due to their effectiveness in capturing non-linearities. For instance, Lessmann et al. (2015) recommend random forest, emphasising its balance between model simplicity and accuracy in feature importance extraction. Munkhdalai et al. (2019) find that both deep neural networks with multiple layers and extreme gradient boosting outperform traditional bureau credit scores, significantly reducing credit loss in predicting household debt delinquency. Tree-based algorithms and neural networks, as highlighted by studies like Ampountolas et al. (2021) and Moscato et al. (2021), consistently demonstrate high accuracy, especially in the absence of credit history or low default rate scenarios. Gunnarsson et al. (2021) further support the preference for extreme gradient boosting tree classifiers over other techniques, including logit and deep learning, in a large-scale benchmark study. As the interest in tree-based ensembles and deep neural networks grows for building consumer risk models, they emerge as the baseline for volatility risk modelling due to their superior predictive ability.

The second key concern in constructing a financial risk assessment model is model explainability. The ethical imperative of understanding and interpreting machine learning outputs, especially from black-box methods, is critical in credit risk management (Grennepois & Robin, 2019). Post-hoc interpretations of black-box machine learning models have gained attention, with Grennepois & Robin (2019) proposing a hybrid approach that combines outputs from logit and machine learning. This approach involves identifying well-classified populations by machine learning that are misclassified by logit, followed by extracting business rules using popular explainability tools such as feature importance, local interpretable model-agnostic explanations, or Shapley values. Bracke et al. (2019) introduce the quantitative input influence technique to measure the relative importance of inputs and their direction, uncovering risk drivers of mortgage default. By controlling for correlations via randomised interventions on input, the feature influence is computed, and Shapley values are applied to capture interactions, providing insights into varying non-linear behavior across different loan groups. In the context of peer-to-peer commercial lending, Bussmann et al. (2021) explore the correlation network of Shapley values for financial characteristics, facilitating the identification of variable contributions and effective grouping of risky and non-risky borrowers. These interpretability tools play a crucial role in bridging the gap between model complexity and user understanding, ensuring transparency and ethical usage of machine learning models in financial risk assessment.

While Shapley values have gained popularity for revealing black box properties, their use may lead to unrealistic observations that could mislead interpretations, especially when regressors are correlated. Pioneering an alternative approach, Bastos & Matos (2022) introduce Accumulated Local Effects (ALE) to explain how loss given default varies with firms' behaviour while keeping other covariates constant. They emphasise the benefits of ALE as a visual representation to scrutinize positive and negative, linear and non-linear, as well as convex and concave patterns in determining credit losses. Given the potential correlation among vulnerability transactional patterns, where poor behaviour is intertwined with other financial struggles, ALE emerges as the appropriate explainability tool to draw the vulnerability-volatility relationship. The successful applications of machine learning in financial risk modelling, combined with post-hoc explainability through tools like ALE, present a promising alternative for the development of an account-level volatility prediction model.

2.3 Methodology

2.3.1 Data setup

The objective of this chapter is to predict the future volatility of both existing and new accounts based on their historical transactional patterns. To achieve this, we adopt a standard setup commonly employed in consumer credit risk assessment, akin to credit scoring practices. Traditionally, application credit scoring assesses the risk of new borrowers, while behavioural credit scoring is applied to existing customers, guiding potential actions like adjusting credit limits. In this chapter, we align with this approach, with further details on credit scoring available in Thomas et al. (2017).

The dataset[14] is partitioned into a training set, an out-of-time in-sample test set (for existing accounts), and an out-of-time out-of-sample test set (for new accounts), as depicted in Figure 2.1. The training set comprises 70% randomly selected accounts, incorporating transactional data from the initial two quarters and volatility data from the third quarter. Consequently, the models are trained to predict the personal account's volatility one quarter ahead, utilising predictors/covariates from the preceding two quarters. The out-of-time in-sample test set applies the trained models to the same 70% randomly selected accounts (7805 accounts), predicting Q4 volatility based on transactions from Q2 and Q3. Meanwhile, the out-of-time out-of-sample test set employs the trained models on the remaining 30% of accounts (3345 accounts).

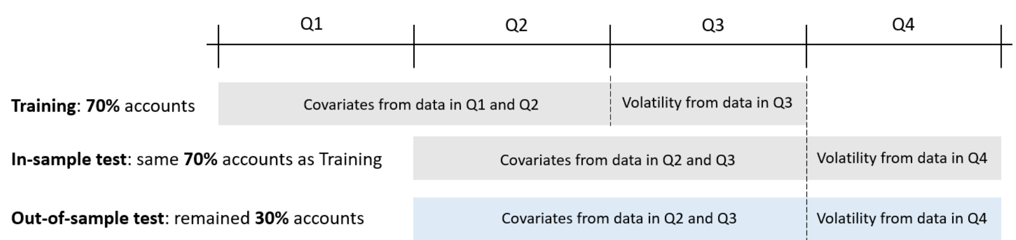


Figure 2.1: Training and test data split

2.3.2 Formulation of index

2.3.2.1 Definition of financial volatility index

This chapter introduces a novel volatility index that integrates various facets of volatility related to an account's available balance, income, and expenditure. The creation of a unified measure of financial volatility enables the ranking of customers based on the level of uncertainty they pose for modelling purposes. The volatility index is derived from three months of transactions, Q3 in the training set and Q4 in the test sets.

[14]described in Table 1.2

We define the account volatility index as a measure of the level of variation in cash flow activities observed over a quarter. We measure volatility on a quarterly basis because this time frame facilitates mid-term risk assessment. Fluctuations in cash flow over a quarter can indicate prolonged financial uncertainty (exceeding two months) and signal potential financial difficulties.To formulate the volatility index, three key variables are employed: daily balance, monthly inflow, and monthly outflow. The daily balance for each account constitutes a time series representing the end-of-day balance spanning three months, with a few exceptions for accounts with transactions ending in the middle of the month. Monthly inflow and outflow series are obtained by aggregating the daily credit (positive or incoming) and debit (negative or outgoing) amounts. The coefficient of variation (CV), a widely utilised measure of volatility in Finance, is employed to quantify the volatility of each account. The CV expresses the standard deviation of each data series relative to the corresponding mean, facilitating the comparison of variation across different time series, even when means vary significantly. The quarterly CV for each variable (inflow, outflow, balance) is expressed as

$$\mathrm{CV} = \frac{\sigma_t}{\mu_t}, \tag{1}$$

where σ_t is the standard deviation and μ_t is the mean of the variable over the total number of time points (t) in a quarter. These CVs serve as inputs for Principal Component Analysis (PCA) (detailed in Section 2.3.4) to determine optimal weights for the three dimensions of data, thereby constructing the volatility index.

2.3.2.2 Vulnerability index

We define a vulnerability index to measure the exposure risk to face financial struggles based on a combination of negative financial circumstances. The regulator (Financial Conduct Authority, 2021b) has outlined four key drivers of vulnerability: (1) poor health, impacting normal daily life; (2) experiencing unexpected life events such as unemployment or bereavement; (3) possessing low resilience to financial or emotional shocks; and (4) having low capability and insufficient knowledge in money management. While these drivers cannot be directly observed, identifiable transactions or patterns logically linked to the mentioned factors are considered. Consequently, a vulnerability index is formulated using a list of numeric attributes. These attributes represent different vulnerability aspects outlined by Financial Conduct Authority (2021b), which combined, give a generic overview of vulnerability risk. Table 2.1 explains the attributes.

Table 2.1: Attributes to form vulnerability index

Attribute	Vulnerability aspect	Remark
Total healthcare expenditure	Poor health	Commitment to medical treatment
Total inflow loss	Life event	Job loss
Total benefits received	Life event	Having disadvantage
Total debt amount	Low resilience	Insufficient money-in-hand
Debt-to-inflow ratio	Low resilience	Ability to pay debt
Buffer-to-outflow ratio*	Low resilience	Insufficient savings capacity
Outflow-to-inflow ratio	Low capacity	Overspending
Expenses of social activities or hobbies	Low capacity	Tight budget

*Buffer refers to the sum of savings and available balance

The PCA technique (explained in Section 2.3.4) is applied to assign weights and aggregate the attributes, facilitating the construction of the composite vulnerability index. Note that the attributes listed in Table 2.1 are not an exhaustive set for depicting vulnerability. This chapter is focused on establishing a baseline approach for developing a vulnerability index based on OB transactions for volatility prediction. Thus, creating a comprehensive vulnerability indicator is beyond the scope of this chapter. The vulnerability index presented here is a preliminary version that includes only a selection of representative attributes. Despite its basic nature, this approach offers valuable insights into key aspects contributing to vulnerability, and serves as a foundation for future development of a more robust indicator.

2.3.3 Covariate creation

The construction of covariates involves four main aspects that characterise financial behaviours: balance, inflow, outflow, and spending on specific categories. Beyond general summary measures depicting individuals' prevailing financial conditions, an exploration of downward trends is undertaken to identify potential signals of deteriorating financial circumstances. Two temporal perspectives are considered: a short-term trend, examining whether the current month has a lower value relative to the previous month, and a longer-term trend assessed over six months by scrutinising declining patterns over three-month rolling windows (given six months of transactional records, yielding four three-month rolling windows).

The balance serves as the foundation for extracting patterns on both a daily and monthly basis, giving rise to distinct groups of covariates detailed in Appendix B. The daily balance measures

(X1-X5) capture fluctuations throughout the day, while the monthly balance covariates (X6-X8) provide an assessment of the financial position at the end of each month. The identification of potential distress involves scrutiny of four overdraft use scenarios (X9-X12): (i) occurrence of going into overdraft at least once in a month, (ii) having a negative month-end balance, (iii) experiencing a larger negative month-end balance than a negative month open balance, and (iv) sustaining a negative month-end balance for more than half of the month. The latter two scenarios specifically highlight severe overdraft usage. Additionally, attention is directed towards analysing the dynamics of the lowest monthly balance (X13-X16), representing a worst-case scenario where a declining trend signifies deterioration in the worst financial position.

The analysis of income and expenditure behaviours is condensed from both incoming and outgoing transactions, scrutinizing variables such as amount, frequency, and the number of days certain transactions occur (X17-X34). Delving into spending behaviours within specific categories (X35-X92), the focus is directed towards discerning payment patterns in fixed costs, living expenses, and debts. Furthermore, the examination extends to leveraging expense information that may hint at potentially unhealthy financial conditions. For example, regular gambling may imply addictive behaviour, while occasional healthcare and medical expenses could signal unexpected life events.

Additionally, a composite measure of vulnerability (X93) is incorporated, encapsulating its four key drivers outlined by the regulator (Financial Conduct Authority, 2021b): poor health, experiencing unexpected life events, possessing low resilience and having low capability. The construction of the vulnerability index is detailed in Section 2.3.2.2.

Table 2.2 provides a concise overview of the covariates categorised into four main aspects—balance, inflow, outflow, and spending on specific categories. Three types of model setups are outlined to address different objectives. In Type 1, which focuses solely on prediction, the model employs the same variable from the past to predict the target variable of interest. Type 2 explores the influence of vulnerability on future volatility, recognising the potential dependence of future volatility on financial circumstances, especially financial difficulties. Finally, Type 3 investigates whether vulnerability risk contributes additional predictive power beyond past volatility. Note that all covariates under vulnerable circumstances are lagged one quarter to project volatility risk in the next quarter. For a more detailed description of each covariate, refer to Appendix B, where the comprehensive list, including the vulnerability index, gives a total of 93 covariates.

Table 2.2: Summary of covariates and types of model setup

		Included: ✓		
Aspect	Covariate	Type 1	Type 2	Type 3
(1) Volatility		✓		✓
volatility lag 1	Y_lag1			
volatility lag 2	Y_lag2			
(2) Vulnerable circumstances*			✓	✓
(A) Past transactional patterns				
(i) Balance				
daily	X1-X5			
monthly	X6-X8			
distress	X9-X12			
lowest balance	X13-X16			
(ii) Inflow	X17-X25			
(iii) Outflow	X26-X34			
(iv) Spending on specific categories	X35-X92			
(B) Vulnerability index	X93			

*vulnerable circumstances from the previous quarters (refer Figure 2.1)

2.3.4 Principal Component Analysis

Given a set of correlated variables, Principal Component Analysis (PCA) operates under the assumption that there exists an underlying latent structure capable of representing the original variables as new, uncorrelated latent variables referred to as principal components (PC). These PCs are linear functions of the input data and are estimated to maximise the variance, effectively capturing the information contained in the original variables. The optimal number of PCs is chosen to reduce the dimensionality of the data, providing the most effective weighted combinations of input attributes to construct a single composite index.

The correlation matrix of a standardised attributes is represented as $\mathbf{M}_{a\times a}$. Conducting PCA on $\mathbf{M}$ to extract eigenvectors and eigenvalues is equivalent to a linear transformation of the original variables. The eigenvector-eigenvalue pairs follow a sequential pattern where the first pair corresponds to the largest eigenvalue, capturing the most significant variation (maximum information retained) in the direction of the paired eigenvector. The subsequent pairs represent the next largest eigenvalues in orthogonal directions, each accounting for additional variation. For $a = 1, \ldots, A$ attributes, PCA yields $l = 1, ..., L$ principal components denoted as PC_l, each comprising the l-th eigenvalue λ_l and the l-th eigenvectors $\psi_{l,a}$ for each attribute a. Each PC_l is a weighted linear combination of the a attributes, with the weights defined by $\psi_{l,a}$. The percentage of explained variance for each PC is given by $\frac{\lambda_l}{\sum_{l=1}^{L} \lambda_l}$.

The composite index (CI) is derived from an optimal number of PCs (κ). To determine this, the percentage of variance explained by each PC is examined, ensuring that the selected PCs effectively capture information from the input attributes while discarding unnecessary noise. For each of the $i = 1, \ldots, n$ accounts, considering the input instance $D_{i \times a}$ with a dimensions of attributes, the CI for each account can be expressed as a weighted combination of the PCs,

$$\text{CI}_i = \frac{\sum_{l=1}^{\kappa} \sum_{a=1}^{A} \lambda_l (\psi_{l,a} D_{i,a})}{\sum_{l=1}^{\kappa} \lambda_l}, \tag{2}$$

such that the weights, w_a, signify the relative importance of each attribute a in the final CI, and it can be denoted as,

$$w_a = \frac{\sum_{l=1}^{\kappa} \lambda_l \psi_{l,a}}{\sum_{l=1}^{\kappa} \lambda_l}. \tag{3}$$

2.3.5 Models to be compared

The development of the volatility model draws inspiration from credit scoring models for two main reasons. First, initial trials using a conventional asset pricing model, such as GARCH, to model the volatility of daily available balance time series yielded unsatisfactory results. The model showed large residuals and non-significant coefficients, suggesting that the GARCH model's statistical assumptions are not well-suited to capture current accounts' cash flow variations. While asset price returns often follow specific statistical patterns, the GARCH model's poor fit for balance data could be due to the unique characteristics of cash flow transactions. Second, the concept of using historical transactional patterns to forecast future volatility risk mirrors the approach in credit risk modelling, where past repayment behaviour (behavioural scoring) or applicant information (application scoring) is used to predict future default risk. Therefore, the credit risk model framework can be similarly applied to volatility prediction in this chapter, aiming to assess consumer risk based on financial behaviour observed in account transactions.

The task of predicting a continuous numerical financial volatility (target variable) from transactional behaviour (covariates) resembles a regression problem. Consider y_i as the vector of target variables for $i = 1, 2, ..., n$ accounts and X_i as an $n \times k$ matrix representing k-dimensional covariates for n accounts. The generic form of regression is expressed as

$$y_i = f(X_i) + \epsilon_i, \tag{4}$$

where the function $f(\cdot)$ signifies the relationship between y_i and X_i through a set of estimated parameters obtained by optimising a specific loss function. The choice of regression model determines the optimisation of the loss function and the distribution setting for the residuals ϵ_i.

We evaluate the performance of both standard statistical models and several state-of-the-art machine learning techniques from the tree ensemble and neural network families. The choice of algorithms is influenced by previous studies on predictive accuracy in credit scoring (Lessmann et al., 2015; Gunnarsson et al., 2021; West, 2000; Kriebel & Stitz, 2022; Ala'raj et al., 2022), where these algorithms have reported superior performance. Additionally, we include statistical techniques, such as quantile and spline regressions, to account for non-linear patterns. Linear regression, as a widely used and straightforward algorithm, is also included as a benchmark for comparison with more advanced techniques.

2.3.5.1 Linear regression

Linear regression (LIN) models the relationship between y_i and X_i as a linear combination:

$$y_i = X_i\beta + \epsilon_i \tag{5}$$

where β is a $k \times 1$ vector of parameters obtained through Least Squares estimation, minimising the sum of squared residuals $\sum_{i=1}^{n} |y_i - X_i\beta|^2$, and ϵ_i represents the residuals. The assumptions of the linear regression model include:

1. X_i and y_i exhibit a linear relationship, captured by the estimated β.

2. X_i is a full-rank matrix to avoid multicollinearity among the covariates.

3. X_i is exogenous, ensuring $E[\epsilon_i|X_i] = 0$ to prevent the covariates from carrying useful information for predicting ϵ_i.

4. Residual term ϵ_i is assumed to have an independent and identical normal distribution, with $\epsilon_i|X_i \sim N[0, \sigma^2]$, ensuring non-correlated and homoscedastic errors with the same finite variance σ^2 across observations.

Following assumption (3) where $E[\epsilon_i|X_i] = 0$, the prediction $\hat{y}_i$ is the conditional mean on X_i,

$$\begin{aligned} \hat{y}_i &= E[y_i|X_i] \\ &= E[X_i\beta] + E[\epsilon_i|X_i] \\ &= X_i\beta. \end{aligned} \tag{6}$$

The least square estimated β coefficients serve as the best linear predictor, describing the marginal effect of X_i on y_i. In this context, a unit increase in the covariate is associated with a unit increase in the target variable (Cameron & Trivedi, 2005). However, certain assumptions in linear regression (LR) may be inappropriate for modelling bank account transactional patterns. For example, financial volatility might exhibit higher variation for accounts with a higher degree of income loss (non-homoscedastic), and the complex transactional behaviour could have varying impact levels on financial volatility (non-linearity). The following subsections outline two statistical models—linear quantile regression and spline regression—with more relaxed assumptions.

2.3.5.2 Linear quantile regression

Considering the potential of having a right-tailed distribution of volatility and the heteroscedasticity challenge in modelling the skewed variation in target variable, we consider the quantile regression in estimating the effects of covariates across low to high quantiles. Quantiles represent the cutoff values to divide a probability distribution into equal-sized groups. Quantile regression (LIN-Q) (Koenker & Hallock, 2001) employs this statistical measure to predict the conditional quantile $Q_q[y_i|X_i]$ instead of the conditional mean $E[y_i|X_i]$. The q-th quantile is the probability where y_i is less than or equal to the conditional quantile $Q_q[y_i|X_i]$

$$q = P[y_i \leq Q_q[y_i|X_i]]. \tag{7}$$

The regression equation is expressed as

$$y_i = X_i\beta_q + \epsilon_i, \tag{8}$$

where β_q represents the parameters to be estimated for different q quantile values by minimising an asymmetric error loss function,

$$\sum_{i;y\geq x_i\beta}^{n} q|y_i - x_i\beta_q| + \sum_{i;y<x_i\beta}^{n} (1-q)|y_i - x_i\beta_q|, \tag{9}$$

and ϵ_i is the non-parametric error term which relaxes the homoscedasticity normal error distribution assumption in LR with no specific distribution and not restricted to constant variance. The prediction $\hat{y}_i$ is

$$\begin{aligned}\hat{y}_i &= Q_q[y_i|X_i] \\ &= X_i\beta_q.\end{aligned} \tag{10}$$

Similar to LIN, the estimated β_q coefficients also represent the marginal effects of the covariates on the target variable but provide the flexibility to explore non-constant effects across low to high quantiles.

2.3.5.3 Spline regression

In situation where X_i has non-linear relationship with y_i, the linearity assumption can be relaxed by transforming each covariate in X_i with a certain basis function $g(x)$ (Hastie et al., 2009). For the r-th transformation, each covariate x is expressed as

$$f(x) = \sum_{r}^{R} \beta_r g_r(x). \tag{11}$$

Instead of a constant increasing or decreasing slope, the basis function representation captures non-linear patterns across different values of x.

Spline is a function constructed from continuous piecewise polynomials (Racine, 2014). Knot points are located within the data range to join the segments of adjacent polynomials together. The spline function represents each covariate x as a smooth curve, where increasing the degree of polynomials (D) and number of knots (K) give a more flexible curve representation. This chapter focuses on the regression splines based on the basis spline (B-spline) to express each covariate as

$$f(x) = \sum_{r=1}^{K+M} \beta_r B_r(x), \tag{12}$$

where $B_r(x)$ is the B-spline basis of polynomial degree D and order $M = D+1$ with K knot points, and β_r is the associated spline coefficients to be estimated.

The set of knots is a sequence of non-decreasing real numbers ($s_k \leq s_{k+1}$) denoted as

$$s_{-(M-1)} \leq ... \leq s_0 \leq ... \leq s_K \leq s_{K+1} \leq ... \leq s_{K+M}, \tag{13}$$

with K inner knots and augmented lower boundary knots $s_{-(M-1)} \leq ... \leq s_0$ and upper boundary knots $s_{K+1} \leq ... \leq s_{K+M}$ where these boundary knots are appended to allow the definition of the

B-spline basis in a recursive manner. This augmented knot set gives a total of K+2M knots, s_k, indexed by $k = 0, ..., K + 2M - 1$.

For each knot s_k ($k = 0, ..., K + 2M - 1$), a set of real-valued spline functions $B_{k,j}$ of j-th order at the k-th knot are recursively defined as

$$B_{k,j}(x) = \frac{x - s_k}{s_{k+j} - s_k} B_{k,j}(x) + \frac{s_{k+j} - x}{s_{k+j} - s_k} B_{k+1,j}(x) \tag{14}$$

where

$$B_{k,0}(x) = \begin{cases} 1 \quad \text{if} \quad s_k \le x \le s_{k+1}, \\ 0 \quad \text{otherwise}, \end{cases} \tag{15}$$

and $B_{k,0} = 0$ if $s_k = s_{k+1}$ to avoid division by zero (Prautzsch et al., 2002). Equation 14 shows that each component of $B_{k,j}(x)$ is the product of two local functions i.e., a constant function and a D-degree function. Thus, the B-spline basis functions have local supports and are numerically stable and efficient.

The spline regression (SPLINE) with B-spline basis is then expressed as

$$\begin{aligned} y_i &= \sum_{p=1}^{P} \sum_{r=1}^{K+M} \beta_r B_r(x_{i,p}) + \epsilon_i, \quad x \in [s_0, s_{k+1}] \\ &= \sum_{p=1}^{P} f_p(x_{i,p}) + \epsilon_i \\ &= F(X_i) + \epsilon_i, \end{aligned} \tag{16}$$

where the residual term has independent and identical normal distribution $\epsilon_i \sim N[0, \sigma^2]$, $F(X_i)$ is the simplified representation of the transformed X_i with the B-spline basis, and the unknown β coefficients can now be estimated with the least square method (James et al., 2013). The prediction $\hat{y}_i$ is then

$$\hat{y}_i = F(X_i). \tag{17}$$

Taking into consideration the flexibility of the B-spline functions in representing the covariates, the marginal effects will have different levels of changes when x increases, instead of only a single straight line slope.

The specification of SPLINE requires cautious selection of the spline order as well as the number and position of the knots. Higher polynomials order and number of knots provide more flexibility to fit non-linear pattern but pose the risk of overfitting. We focus only on quadratic and cubic splines since going beyond cubic specification is usually unnecessary (Hastie et al., 2009). To select the appropriate number of knots, the root mean squared error measure from ten-fold cross validations across $k = [1, 20]$ is manually examined, and the final k which gives the lowest error measure is picked. The knots are located based on the quantile values. Quadratic B-spline is the final specification used because it has lower cross validation errors than cubic B-spline.

The coming subsections describe the non-parametric tree-based and network based machine learning techniques, which are flexible to handle data patterns without specific statistical distribution.

2.3.5.4 Random forest

Random forest regression (RF) is a collection of independent unpruned decision tree regression via the bootstrap aggregation (bagging) technique (Breiman, 2001). Let $T = \{T_1, ..., T_B\}$ denote the regression trees, each tree T_b is a binary classification and regression tree (CART) that continuously divides the full region in each covariate x into subregions with a recursive binary splitting algorithm (Breiman, 2017). Every binary split creates two children nodes (N_L and N_R) such that each node represents a subregion,

$$N_L(p, c) = \{x|x_p < c\} \quad \text{and} \quad N_R(p, c) = \{x|x_p \geq c\}, \tag{18}$$

where the splitting point c for the p-th predictor is to minimise the residuals sum of squares,

$$\sum_{i=x_i \in N_L(p,c)} (y_i - \hat{y}_{N_L})^2 + \sum_{i=x_i \in N_R(p,c)} (y_i - \hat{y}_{N_R})^2. \tag{19}$$

The recursive splitting continues until the tree grows to its maximum i.e., all observations in the node has the same value, or there is only one observation in the leave node.

Following the bagging algorithm, each tree will be learned on a bootstrap sample from the training data (Cutler et al., 2012). Since every tree generated from bagging is identically distributed, the bias of the bagged trees equals to the individual tree. When the covariates are identically distributed but not independent, averaging the bagged trees remain at low bias but will have increased variance due to the correlated covariate pairs. To ensure low variance for the ensemble trees, the RF algorithm induces an additional step to randomly select $m \leq p$ input covariates for splitting when growing each tree. This strategy reduces the correlation between the trees and, consequently, reduces the variance. The RF regression predictor is then the average prediction from the independent B trees grown on a bootstrapped training sample,

$$\begin{aligned} \hat{y_i} &= \hat{f_{rf}^B}(X_i) \\ &= \frac{1}{B}\sum_{b=1}^{B} T_b(X_i). \end{aligned} \tag{20}$$

The key hyperparameters for RF are the total number of tree (B) and the maximum number of features (m) to be randomly selected for splitting in each tree.

2.3.5.5 Quantile gradient boosting tree regression

Gradient boosting tree (GB) is a strong tree ensemble model constructed by sequentially adding weak CART regression trees (Friedman, 2001). In every iteration, the added weak learner successively improves the predictions from the residuals of the previous tree. The objective function to be solved at each iteration t is represented as

$$\hat{\Theta}_t = \text{arg min} \sum_{i=1}^{N} L(y_i, f_{t-1}(X_i) + z_t(X_i, \theta_t)), \tag{21}$$

where $z_t(X_i, \theta_t)$ is the new tree to be added with parameters θ and $L(\cdot)$ is flexible to take any differentiable loss function. For the purpose of having a machine learning counterpart for the linear quantile regression, the GB quantile regression (GBR-Q) thus works on minimising the asymmetric loss function in Equation 9.

The GB algorithm (Hastie et al., 2009) starts with model initialisation $f_0(X_i)$ with a constant γ that minimises the loss function $L(\cdot)$,

$$f_0(x_i) = \text{arg min}_\gamma \sum_{i=1}^{N} L(y_i, \gamma). \tag{22}$$

Then, for each iteration $t = 1, ..., T$, the pseudo residuals r_{it} are computed as the negative gradient value i.e., partial differentiation of the loss function with respect to the previous prediction,

$$r_{it} = -\left[\frac{\partial L(y_i, f(X_i))}{\partial f(X_i)}\right]_{f(X_i)=f_{t-1}(X_i)}, \quad \text{for} \quad i = 1, ..., N. \tag{23}$$

A new regression tree is then trained with the covariates X_i against the pseudo residuals r_{it},

$$z_t(X_i) = \{(X_i, r_{it})\}_{i=1}^{N}. \tag{24}$$

The new learner is included with a multiplier γ_t that minimises the loss function,

$$\gamma_t = \arg\ \min_{\gamma} \sum_{i=1}^{N} L(y_i, f_{t-1}(x_i) + \gamma_t z_t(X_i)). \tag{25}$$

Finally, the model is updated via

$$f_t(X_i) = f_{t-1}(X_i) + \nu \gamma_t z_t(X_i), \tag{26}$$

where ν is a learning rate between $[0, 1]$ to adjust the step size to move towards the minimum of the loss function. A small ν value prevents dramatic improvements of the gradient boosting to avoid overfitting.

2.3.5.6 Extreme gradient boosting regression

Extreme gradient boosting (XGB) is an advanced implementation of GB where the ensemble model is also trained in an additive manner with a specific implementation to improve efficiency (Chen et al., 2015). The objective function at each iteration t is expressed as

$$\Theta_t = \sum_{i=1}^{N} L(y_i, \hat{y}_{i(t-1)} + z_t(X_i)) + \Omega(z_t). \tag{27}$$

The first part of Θ_t resembles the GB function to add a new tree $z_t(X_i)$ that minimise the residuals from the previous iteration.

XGB differs from GB in two ways (Chen et al., 2015). First, XGB enables regularisation with the extra complexity term, $\Omega(z_t)$, represented as $\Omega(z_t) = \delta T + \frac{1}{2}\lambda ||\omega||^2$ where z_t is a tree with structure q and leaf weights ω with T number of leaves, and $||\omega||^2$ denote the L2-norm of the leaf weight. Both δ and λ are the regularisation terms to avoid overfitting. Second, the parameters of a new tree are estimated from the second order Taylor expansion of the loss function instead of the first order partial derivative which computes the negative gradient in GB. Using squared error as the loss function $L(y_i, \hat{y}_i) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, the objective function in Equation 27 can be approximated as

$$\begin{aligned} \Theta_t &= \sum_{i=1}^{N} \left[2(\hat{y}_{i(t-1)} - y_i) z_t(X_i) + z_t(X_i)^2 \right] + \Omega(z_t) + \text{constant} \\ &= \sum_{i=1}^{N} \left[L(y_i, \hat{y}_{i,(t-1)}) + g_i z_t(X_i) + \frac{1}{2} h_i z_t^2(X_i) \right] + \Omega(z_t) + \text{constant}, \end{aligned} \tag{28}$$

the second line is the representation with Taylor expansion of the mean squared loss function up to the second order, g_i and h_i are the first and second order gradient statistics defined as

$$
\begin{aligned}
g_i &= \partial_{\hat{y}_{i(t-1)}} L(y_i, \hat{y}_{i(t-1)}) \\
h_i &= \partial^2_{\hat{y}_{i(t-1)}} L(y_i, \hat{y}_{i(t-1)}).
\end{aligned}
\tag{29}
$$

The new tree $z_t(X_i)$ is denoted as $\omega_q(X_i)$ to indicate a new tree with structure q and weights $\omega_{q(X_i)}$ and define a function $I_j = \{i | q(X_i) = j\}$ as the instance set of leaf j. By dropping the constant term, replacing z_t with the $\omega_{q(X_i)}$ notations and expanding the complexity term $\Omega(z_t)$, the objective function is now rewritten as

$$
\begin{aligned}
\tilde{\Theta}_t &= \sum_{i=1}^{N} \left[g_i \omega_{q(X_i)} + \frac{1}{2} h_i \omega^2_{q(X_i)} \right] + \delta T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \\
&= \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \delta T \\
&= \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2} \left(H_j + \lambda \right) \omega_j^2 \right] + \delta T,
\end{aligned}
\tag{30}
$$

the last line is a simplified representation with $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$.

The optimal ω_j^* for a given structure $q(X_i)$ and the optimal solution of the XGB objective function are

$$
\omega_j^* = -\frac{G_j}{H_j + \lambda}, \tag{31}
$$

$$
\tilde{\Theta_{t(q)}} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \delta T. \tag{32}
$$

A better tree structure $q(X_i)$ will have a lower score in Equation 32. XGB finds the tree structure by iteratively adding branches to the tree. The splitting criterios at a tree node is based on the Gain from the split,

$$
\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \delta, \tag{33}
$$

which is the information gain after splitting into the left (L) and right (R) leaves compared to before the split occurs. The regularisation terms δ controls how much improvement is required in a split to add the branch and λ adjusts the change of the optimal weights. The tree stops to grow when the Gain is insufficient to form a split or the tree reaches the maximum depth. Each tree is then added to form the final ensemble where a learning rate ν is multiplied on the new tree during the addition to control the rate of model fit. Apart from the regularisation terms, sampling is another strategy in XGBoost to control overfitting. The sampling is done by either selecting a subsample of the rows of training data or a subsample of the columns of the features to reduce the correlation in between the boosted trees.

2.3.5.7 Multilayer perceptron neural network

The multilayer perceptron neural network (MLP) comprises of one input layer, a number of hidden layers and one output layer (Bishop, 1995). Each hidden layer contains neurons that act as the medium to connect the entire network by input processing and output generation. A neuron z_v in a proceeding layer v is a linear combination of the output neuron vector a_u from the previous layer $u = v - 1$, denoted as

$$z_v = w_{uv} a_u + b_v. \tag{34}$$

The parameter w_{uv} is the weights connecting neurons in layer u and v. The bias term b_v is the intercept to increase or decrease the weighted linear sum for each new neuron. Note that the neuron for the first input layer takes the covariates such that $a = x$. The neurons in the hidden layers operate by (i) processing the input values with the weighted linear sum using the w_{uv} and b_v parameters and (ii) generating output with a non-linear activation function $\Phi(\cdot)$.

The training of the network parameters is done with the forward and backward propagation (Svozil et al., 1997). The forward pass transmits the output from one hidden layer to be the input for the subsequent hidden layer and the backward pass distributes the errors back through the network and adjust the weights by minimising the errors. An epoch is a complete forward and backward pass and it is repeated to train the parameters until reaching the stopping criterion. In the forward propagation step, each neuron is excited with the rectified linear unit activation function,

$$\Phi(z) = max\{0, z\}, \tag{35}$$

to compute the neuron output in each layer until the final layer,

$$a = \Phi(z). \tag{36}$$

The task to obtain the optimal parameters is by minimisation of the squared error loss function in the back propagation step for each layer v over all instances $n = 1, ..., N$,

$$L = \frac{1}{2}\sum_{n=1}^{N}(y_{n,v} - a_{n,v})^2, \tag{37}$$

where $y_{n,v}$ is the observed output and $a_{n,v}$ is the output computed from the activation function. Back propagation follows the gradient descent algorithm to adjust the parameters by taking the partial derivative of the loss function with respect to the weights (and bias). According to the chain rule, the partial derivative can be expressed as,

$$\begin{aligned}\frac{\partial L}{\partial w_v} &= \frac{\partial L}{\partial a_v} \cdot \frac{\partial a_v}{\partial z_v} \cdot \frac{\partial z_v}{\partial w_v} \\ &= -(y - a_v)\Phi'(z_v)a_u.\end{aligned} \tag{38}$$

Each partial derivative in Equation 38 can be obtained from Equation 37, 35 and 34 respectively. The change in weights in a layer v between the forward and backward pass is

$$\begin{aligned}\Delta w_v &= \eta\frac{\partial L}{\partial w_v} \\ &= \eta\delta_v a_u, \quad \delta_v = -(y - a_v)\Phi'(z_v),\end{aligned} \tag{39}$$

where η is the learning rate to adjust the magnitude of change of the weights. Note that δ_v is only applicable to back propagate the error from the output layer to the hidden layer when v is the final output layer with observed y. To allow the error back propagation across the hidden layers, δ_v is multiplied with the weights connecting neuron in the current (v) and previous (u) layers. The general notation of δ_u at a hidden layer u is then denoted as

$$\delta_u = \Phi'(z_u)\delta_v w_{vu}. \tag{40}$$

The update rule of the weights is now expressed as

$$\Delta w_u = \alpha\Delta w_u + \eta\delta_u a_u, \tag{41}$$

with α as the momentum term to speed up the training process and avoid being trapped in the local minima.

2.3.5.8 Deep belief network

Deep belief network (DBN) is a class of deep neural network constructed from a stack of subnetworks, known as Restricted Boltzmann Machine (RBM) (Hinton, 2009). RBM is an undirected neural network such that the visible units are connected to the hidden units but there are no visible-to-visible or hidden-to-hidden connections. A standard RBM is a two-layer network that connects binary hidden ($\mathbf{h}$) and visible ($\mathbf{v}$) units with symmetrically weighted connections (Hinton, 2012). The joint configuration of the visible and hidden unit pair has an energy defined as

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i \in \mathbf{v}} a_i v_i - \sum_{j \in \mathbf{h}} b_j h_j - \sum_{i,j} v_i h_j w_{ij}, \tag{42}$$

where v_i and h_j represent the visible and hidden units respectively, a_i and b_j are the corresponding biases, and w_{ij} is the connecting weights.

Given a partition function $\mathbf{z} = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}$ that sums over all possible paris of $(\mathbf{v}, \mathbf{h})$, the network assigns the probability of each configuration pair with

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})}, \tag{43}$$

and the marginal probability of the visible vector $\mathbf{v}$ is the sum of all possible hidden vector $\mathbf{h}$,

$$P(\mathbf{v}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}. \tag{44}$$

RBM has no direct connections between the hidden units (Hinton, 2012). Thus, the hidden unit activations are mutually independent given the visible unit activations, and vice versa,

$$\begin{aligned} P(h_j = 1|\mathbf{v}) &= \pi(b_j + \sum_i v_i w_{ij}) \\ P(v_i = 1|\mathbf{h}) &= \pi(a_i + \sum_j j_j w_{ij}), \end{aligned} \tag{45}$$

such that $P(h_j = 1|\mathbf{v})$ denotes the probability of each hidden unit given a randomly selected visible vector and $P(v_j = 1|\mathbf{h})$ denotes the probability of each visible unit given a hidden vector, and $\pi(\cdot)$ is the logistic function $\pi(x) = \frac{1}{1+e^{-x}}$.

The weight update Δw_{ij} is conducted by taking the partial derivative of the log probability of the training vector $\mathbf{v}$ with respect to the weight,

$$\begin{aligned} \frac{\partial \log P(\mathbf{v})}{\partial w_{ij}} &=< v_i h_j >_{data} - < v_i h_j >_{model} \\ \Delta w_{ij} &= \eta_{rbm}(< v_i h_j >_{data} - < v_i h_j >_{model}), \end{aligned} \tag{46}$$

multiplied with the learning rate η_{rbm} to adjust the step size of the weight change. The expectation of the data distribution is denoted as $< v_i h_j >_{data}$ where the unbiased sample can be obtained by

$P(\mathbf{h}|\mathbf{v})P(\mathbf{v})$. The expectation of the model is denoted as $< v_i h_j >_{model}$ and is estimated with the Contrastive Divergence (CD) algorithm instead of maximum likelihood estimation to save computational effort.

In every step of CD, Gibb sampling is run alternately to sample the states of hidden units independently given the states of the visible units and vice versa (Upadhya & Sastry, 2019). Each iteration of the alternating Gibb sampling update the hidden units in parallel with $P(h_j = 1|\mathbf{v})$ in Equation 45 and the reconstructed visible units are obtained by setting each v_i to 1 with a probability computed from $P(v_i|\mathbf{h})$ in Equation 45. Then, the change in weight is then computed by

$$\Delta w_{ij} = \eta_{rbm}(< v_i h_j >_{data} - < v_i h_j >_{reconstructed}). \tag{47}$$

A CD_T with T steps of Gibbs sampling recover the maximum likelihood estimation learning when $T \to \infty$.

The visible units in this experiment are continuous because the covariates are numerical variables. Hence, a Gaussian Bernoulli RBM (Gaussian visible units and binary hidden units) is employed and the energy function is now represented as

$$E(\mathbf{v}, \mathbf{h}) = \sum_{i \in \mathbf{v}} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j \in \mathbf{h}} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j w_{ij}, \tag{48}$$

where σ_i is the standard deviation of the Gaussian noise for $\mathbf{v}$. Due to computational difficulty to learn σ_i with CD_T, the data is normalised to have zero mean and unit variance, and set $\sigma = 1$ to obtain the reconstructed visible units in the CD algorithm (Hinton, 2012).

Each RBM layer acts as feature detector to capture underlying complex relationships via the weighted connections of the visible and hidden layers that is learned to model the input data distribution as close as possible. The DBN training resembles the MLP training process with forward and backward propagation algorithms. The key difference is that each hidden layer is now an RBM network.

2.3.5.9 Hyperparameter tuning for machine learning models

The ML techniques have hyperparameters that need to be pre-specified by the users. We conduct grid search tuning on the five-fold cross-validation of the training data to select the optimal hyperparameters. The tuning grid for the hyperparameters is set up based on literature recommendations (Lessmann et al., 2015; Gunnarsson et al., 2021) and manual explorations. For tree ensemble models, the candidate hyperparameters are chosen based on large-scale credit scoring benchmark experiments from (Baesens et al., 2021; Lessmann et al., 2015; Gunnarsson et al., 2021) where these hyperparameters grid setup has shown to provide sufficient candidates for effective tuning across different credit data sets. For neural network models, increasing the number of hidden layers increases model complexity and computational time. Thus, we restrict it to only five layers for MLP-NN. Prior to setting up the grid for DBN, a manual investigation with one, two, and three layers of RBM in DBN while fixing the other hyperparameters at default shows that increasing the layers does not improve cross-validation results but only increases computational effort. Thus, we only tune the hyperparameters for a single-layer DBN.

Table 2.3: Hyperparameters tuning grid for machine learning models

Model	Hyperparameters	Candidate values
RF	No. of CART trees	100, 250, 500, 750, 1000
	Maximum features	$\sqrt{\text{m}[0.1, 0.25, 0.5, 1, 2, 4]}$
		m = log_2(total number of features + 1)
GB-Q	No. of CART trees	100, 250, 500, 750, 1000
	Maximum features	$\sqrt{\text{m}[0.1, 0.25, 0.5, 1, 2, 4]}$
		m = log_2(total number of features + 1)
XGB	No. of CART trees	50, 100, 150
	Learning rate	0.3, 0.4
	Maximum depth	1, 2, 3
	Fraction of features subsample	0.5, 0.75, 1
	Fraction of input rows subsample	0.6, 0.8
MLP-NN	No. of hidden layers	1, 3, 5
	No. of neurons	5, 10, 15, 20
	Learning rate	0.01, 0.0001, 0.0001
	Momentum	0.1, 0.01, 0.001, 0
DBN	No. of neurons	10, 30, 50
	Learning rate	0.01, 0.1

2.4 Results and discussions

2.4.1 Volatility index

Table 2.4 presents the eigenvectors, eigenvalues, and the percentage of explained variance for each principal component (PC) estimated from the three dimensions of financial behaviours. The volatility index is constructed using the first two principal components, as they account for 88% of the information variance. The third component is discarded, considering 12% of the explained variance as unusual noise in the data.

Table 2.4: Principal Component Analysis estimates for volatility index

Dimensions	PC1	PC2	PC3	Weights*
cv_balance	0.1147	0.9934	0.005298	0.3915
cv_inflow	0.7027	-0.07734	-0.7073	0.3619
cv_outflow	0.7022	-0.08483	0.7069	0.3592
eigenvalues	1.655	0.9914	0.3543	
% explained variance	0.5515	0.3304	0.1181	

*computed from PC1 and PC2

The eigenvectors in Table 2.4 signify the contribution of each financial behaviour within the latent structures. Notably, the high weights of the first PC for inflow and outflow indicate that these components are more crucial for the volatility index. The contribution of balance fluctuation is further captured in PC2, providing additional information on overall volatility. The evenly-distributed weights in the last column of Table 2.4 suggest that all three dimensions are equally important in measuring individuals' overall volatility.

2.4.2 Vulnerability index

Table 2.5 displays the loadings of the PCs derived from the attributes. Six PCs are used to construct the vulnerability index as they capture the majority of the information in the attributes, explaining 92% of the variance. The eigenvectors (shown in the columns) reflect the importance of each attribute within the latent PCs. Attributes with relatively higher values, indicated in **bold** (greater than 0.5), demonstrate their significance across different PCs.

Table 2.5: Principal Component Analysis estimates for vulnerability index

Attribute	Vulnerability aspect	PC1	PC2	PC3	PC4	PC5	PC6	Weights
Total healthcare expenditure	Poor health	0.2101	-0.0827	**0.5084**	-0.3343	-0.6450	0.3945	0.2676
Total inflow loss	Life event	0.3725	0.2456	0.1352	-0.2586	**0.6386**	0.4518	**1.947**
Total benefits received	Life event	-0.0371	-0.2396	0.3799	**0.8124**	0.1097	0.3326	**1.121**
Total debt amount	Low resilience	**0.6248**	0.2068	-0.1114	0.1423	-0.0241	0.0323	**1.591**
Debt-to-inflow ratio	Low resilience	**0.5274**	0.1184	-0.3276	0.3189	-0.3271	-0.1776	0.7502
Buffer-to-outflow ratio*	Low resilience	-0.2239	**0.6395**	0.1011	0.0725	-0.1199	-0.0712	0.5844
Outflow-to-inflow ratio	Low capacity	0.1655	**-0.6319**	-0.2534	-0.1759	0.0694	0.0150	**-1.053**
Expenses of social activities or hobbies	Low capacity	0.2641	-0.1017	**0.6203**	-0.0568	0.1930	**-0.6990**	0.6930
eigenvalues		2.008	1.574	1.158	0.9718	0.9009	0.7357	
% of explained variances		0.2509	0.1968	0.1448	0.1215	0.1126	0.0920	

The weights in the final column highlight the varying degrees of importance of each attribute to the vulnerability index. Three key findings can be highlighted based on the highest weight values (greater than 1). First, life events, such as income loss and benefits received, have the greatest potential to affect vulnerability. This is likely because life events can significantly disrupt financial stability. For instance, a sudden job loss directly results in a loss of income and may leave individuals without immediate alternative sources, leading to financial stress.

Second, high credit usage also plays a significant role in vulnerability. Individuals who heavily rely on credit often lack sufficient funds for daily expenses, placing them in a precarious financial situation. Third, the outflow-to-inflow ratio exhibits a negative relationship with the vulnerability index. While one might expect a higher ratio—indicating overspending or impulsive behavior despite low income—to correlate with increased vulnerability, the negative sign suggests otherwise. This indicates a potential non-linearity, where a low outflow-to-inflow ratio, characterised by minimal spending, may be associated with restricted spending capability and, consequently, greater vulnerability.

2.4.3 Predictive performance

Table 2.6 presents the root mean squared error (RMSE) and mean absolute error (MAE) for both in-sample and out-of-sample test sets, providing an evaluation of the predictive performance across the three types of model setup.

Table 2.6: Predictive performance

	(1) Volatility				(2) Vulnerability				(3) Volatility and vulnerability			
	in-sample		out-of-sample		in-sample		out-of-sample		in-sample		out-of-sample	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Statistical												
LIN	0.7938	**0.6204**	0.8055	0.6356	0.8117	0.6357	0.8079	0.6366	0.7858	0.6161	0.7909	0.6259
LIN-Q	0.8144	0.6295	0.8190	0.6395	*0.8281*	*0.6407*	*0.8157*	0.6339	0.8008	0.6232	0.7993	0.6295
SPLINE	0.7944	0.6223	<u>**0.8020**</u>	0.6365	<u>0.8060</u>	<u>0.6314</u>	<u>0.8051</u>	<u>0.6347</u>	0.7898	0.6207	0.7957	0.6304
Tree ensembles												
RF	*0.8508*	*0.6650*	*0.8776*	*0.6892*	**0.7957**	**0.6215**	**0.8026**	0.6339	**0.7787**	**0.6121**	**0.7903**	0.6265
GBR-Q	0.8113	0.6292	0.8130	0.6392	0.8198	0.6384	0.8121	**0.6325**	*0.8020*	*0.6249*	*0.8053*	*0.6331*
XGB	**0.7916**	0.6205	0.8021	0.6365	0.8042	0.6291	0.8053	0.6351	0.7898	0.6191	0.8004	0.6328
Neural network												
MLP	0.7940	0.6222	0.7996	**0.6338**	0.8096	0.6340	0.8059	0.6365	0.7834	0.6164	0.7918	0.6284
DBN	0.7940	0.6212	0.8090	0.6396	0.8105	0.6313	0.8080	*0.6397*	0.7854	0.6183	0.7892	**0.6251**

Type 1 serves as the baseline scenario where only past values are utilised for prediction. Most models in Type 2 exhibit a slightly higher error than Type 1. This suggests that volatility covariates generally hold stronger predictive power than vulnerability covariates. Conversely, in Type 3, most models perform better than Type 1, indicating that vulnerability adds predictive power on top of the volatility covariates.

All models show highly competitive performance, with only a marginal 1-8% difference between the best (in **bold**) and worst (in *italic*) model results. Across the three model types, tree ensemble techniques, specifically RF, demonstrate the best performance with the lowest errors in most cases. In contrast, quantile regression models, LIN-Q and GBR-Q, display the least predictive performance in most cases. This indicates that predictions with median quantile (both LIN-Q and GBR-Q focus on $q = 0.5$) struggle to generalise well to unseen test sets.

SPLINE stands out as the statistical model with the lowest error in several cases (underlined in Table 2.6), particularly in models using vulnerability as predictors. Notably, LIN performs well out-of-sample when considering combined volatility and vulnerability information. This implies that standard linear models can approximate complex non-linear patterns effectively. Both LIN and SPLINE demonstrate robust performance compared to ML techniques, with only a marginal 1% difference relative to the best-performing ML techniques. However, SPLINE offers the additional advantage of unveiling underlying non-linear structures, providing a more in-depth and nuanced risk analysis (for details, refer to Section 2.4.5).

2.4.4 Linear regression: risky volatile situations

This section explores the connection between transactional patterns and volatility risk, highlighting potential financial struggles that may indicate harmful circumstances. Table 2.7 presents a multicollinearity check using the Variance Inflation Factor (VIF) to identify and eliminate covariates with high multicollinearity (VIF ≥ 5). This process ensures that the regression coefficients are reliable for interpretation, as it prevents the inflation of standard errors due to intercorrelation with other variables.

The VIF(all) column displays the VIF values for all covariates derived from the forward selection procedure. Note that the predictive performances reported in the previous section are based on this complete set of covariates. Removing covariates with high multicollinearity consistently reduces the VIF of the remaining variables. For accurate interpretation, linear and spline regressions are re-run using only those covariates with VIF ≤ 5.

Table 2.8 presents the marginal effects of linear regression in the three types of model setup. The ensuing discussion highlights precarious transactional patterns with a significant statistical effect on (high/low) volatility that may be indicative of financial difficulty.

The positive coefficients of lagged volatility (Y_lag1, Y_lag2) in the Type 1 model signify the persistence of past volatility for future predictions. The larger positive effect of Y_lag1 than Y_lag2 indicates a diminishing impact of past values, wherein conditions at the most recent time point

Table 2.7: Variance Inflation Factor (VIF) of covariates

Covariate	Short name	VIF(all)*	VIF
Y_lag1	volatility index in the previous quarter	1.382	1.287
Y_lag2	volatility index in the previous two quarter	1.430	1.372
X2	consecutive days in same balance amount	1.881	1.541
X5	transition from positive to negative balance	3.656	1.499
X4	extreme high or low balance amount	1.994	1.728
X7	month end balance less than month inflow	8.973	
X8	month end balance less than month open balance	7.392	
X9	take at least one overdraft in a month	7.884	
X12	negative month end balance where more than half of the month is in negative	2.880	1.214
X17	inflow lower than usual	11.34	
X24	decrease in days with inflow	6.573	
X28	days with outflow transactions	10.75	
X21	decrease in number of outflow transactions	7.237	
X30	continuous decrease in outflow amount	5.334	3.183
X35	spending on rent payment	1.070	1.063
X36	spending on telephone and tv	1.407	1.385
X42	spending on supermarket	1.686	1.522
X45	spending on convenience store	1.084	1.078
X51	spending on loan	1.263	1.252
X55	spending on gambling and gaming	1.040	1.035
X63	decrease in electricity and gas payment	1.280	1.276
X69	decrease in telephone and tv payment	2.821	2.569
X93	vulnerability index	4.373	3.760

*VIF for all covariates obtained from forward selection procedure.

Table 2.8: Linear regression estimates

Covariate	Short name	(1) Volatility Coefficient	(2) Vulnerability Coefficient	(3) Volatility and vulnerability Coefficient
(intercept)		-0.0008	-0.2013***	-0.0858***
Y_lag1	volatility index in the previous quarter	0.2998***		0.2551***
Y_lag2	volatility index in the previous two quarter	0.2433***		0.1683***
X2	consecutive days in same balance amount		2.049***	1.106***
X5	transition from positive to negative balance		6.955***	3.632***
X4	extreme high or low balance amount		2.862***	-0.0052**
X12	negative month end balance where more than half of the month is in negative		-0.093*	-0.0366
X30	continuous decrease in outflow amount		0.2900**	0.2023***
X35	spending on rent payment		-0.4637***	-0.2575*
X36	spending on telephone and tv		-0.7054***	-0.4788*
X42	spending on supermarket		-1.205***	-0.6210***
X45	spending on convenience store		1.096***	1.153***
X51	spending on loan		-0.5330***	-0.4604***
X55	spending on gambling and gaming		0.4949*	0.4616*
X63	decrease in electricity and gas payment		-0.3473***	-0.2172**
X69	decrease in telephone and tv payment		-0.3244***	-0.1590***
X93	vulnerability index		-0.0597	-0.1540*

***significant at $\alpha = 0.001$ **significant at $\alpha = 0.01$*significant at $\alpha = 0.05$

exert the strongest influence on the near future.

In the Type 2 model, significant covariates reveal which aspects of vulnerability are pertinent for volatility prediction. Covariate indicating financial buffer shortage, such as shifting from a positive to a negative balance (X5), is positively associated with volatility. This suggests that low financial capacity often leads to unstable cash flows. High volatility is linked to sudden changes in the balance amount (X4), implying that financial behaviors fluctuate due to unexpected events. Low spending capability with frequent reductions in spending activities (X30) also relate to high volatility. Individuals prioritising fixed and living expenses less (X35, X36, X42, X51) and those participating more frequently in gambling (X55) tend to exhibit higher volatility. This suggests that reductions in financial capacity and unhealthy addictions may contribute to increased cash flow fluctuations.

Nevertheless, certain facets of financial vulnerability can also contribute to low volatility, exemplified by frequent inability to meet financial obligations (X63, X69) and extensive overdraft usage (X12). This suggests that individuals reaching the bottom of their financial capacity (e.g., hitting an overdraft limit) and having limited alternatives may exhibit lower volatility, choosing to restrict their spending behavior to cope with their financial situation. Although vulnerability risk (X93) does not emerge as a significant predictor for volatility, the negative association hints that individuals in vulnerable circumstances tend to curtail their cash flow activities to sustain their standard of living.

In the Type 3 model, several transactional behaviors from Type 2 become non-significant (X12) or vice versa (X93). This is likely attributed to their strong correlations with the volatility covariates or the correlation of X93 with the other covariates (the attributes forming the vulnerability index represent several representative vulnerable circumstances that overlap with the transactional patterns here), masking their influence on future volatility. To examine the true effect of the vulnerability index (X93) with volatility risk, the results of linear regression solely with X93 is reported in Table 2.9.

Table 2.9: Linear regression with only vulnerability index

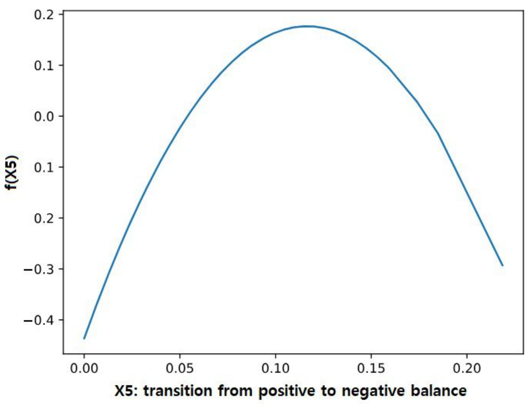
Covariate	Short name	Coefficient
(intercept)		0.099***
X93	vulnerability index	-0.4673***

Referring to the table, it is observed that using the vulnerability index as the sole covariate has a significant negative effect on volatility risk. The results suggest that potential inconsistencies in the statistical significance of the coefficients (e.g., X12 and X93) may stem from intercorrelation among covariates. Nonetheless, most vulnerability covariates demonstrate statistical significance in Type 3, suggesting that transactional patterns contain additional information to describe future volatility.

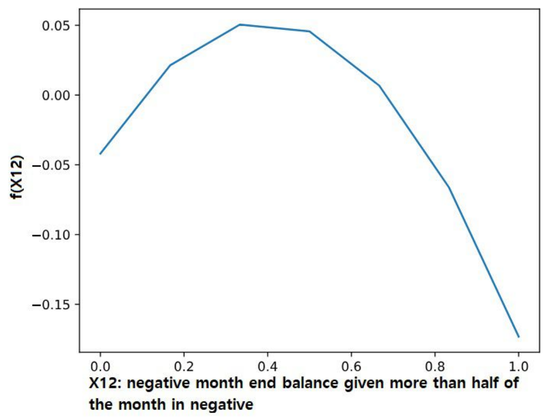
2.4.5 Spline regression: varying impact levels of transactional behaviours on volatility

This section delves into the interpretation of several covariates, exemplified from Type 2, to investigate how shifts in transactional behaviours influence financial volatility. We employ spline regression to analyse the marginal effects, aiming to discover relationships unfound in linear regression and highlight the nuanced impact of non-linear structures.

The marginal effects of overdraft use on volatility (Fig 2.2a, 2.2b) can be delineated through two scenarios. Occasional overdraft use, characterised by accounts sporadically transitioning from



(a) X5



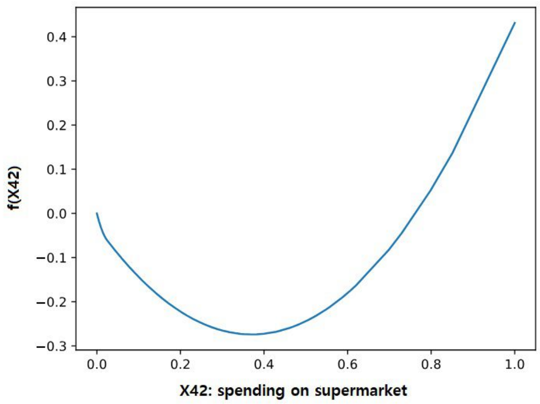
(b) X12

Figure 2.2: Marginal effects of X9 and X12

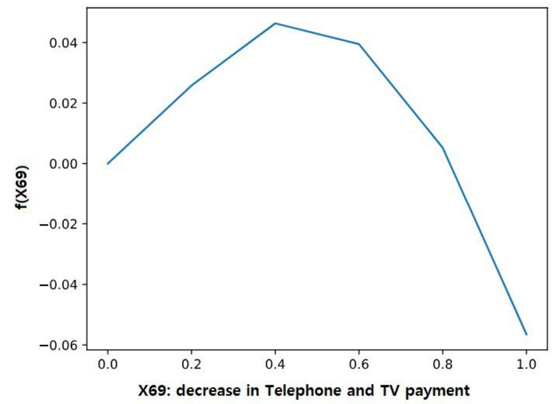
a positive to a negative balance, has a more pronounced impact on increasing volatility compared to regular overdraft usage (Fig 2.2a). However, the sharp decline in volatility beyond a certain threshold suggests that frequent extreme financial shortages, resulting in regular negative balances, reduces variations in cash flow. Consistent reliance on heavy overdrafts significantly diminishes volatility, while temporary use marginally heightens volatility (Fig 2.2b). The ascending patterns in both figures suggest a heightened risk of daily life disruption stemming from occasional overdrafts, possibly due to unforeseen events. Conversely, persistent reliance on heavy overdrafts due to limited financial capacity results in lower volatility.

In Figure 2.3a, the initial segment with a decreasing slope indicates that gradual increases in grocery spending consistently reduce volatility. However, beyond a certain point, further increases in spending lead to a sharp rise in volatility. Figure 2.3b shows a rising pattern, suggesting that volatility risk increases when an account experiences a greater number of months with reduced spending on telephone and TV subscriptions. This pattern implies that occasional missed payments contribute to higher cash flow fluctuations. Conversely, the decreasing trend on the right side of the figure suggests that lower spending capability, which may prompt a reduction in subscription costs, is associated with reduced volatility risk.

The decreasing pattern in Figure 2.4a shows that volatility drops steadily with the increasing number of months with lower spend on electricity and gas, which is logical - smaller values on



(a) X42



(b) X69

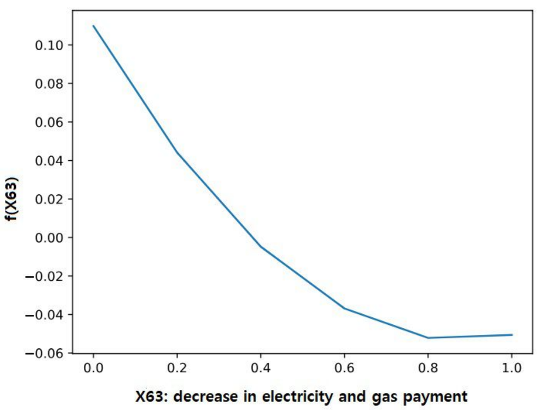
Figure 2.3: Marginal effects of X42 and X69

spending imply less variation. However, there is an almost flat section at the right-hand side, implying that volatility stabilises for very high values of number of months with lower spend on electricity and gas. After a certain level, cutting on utilities has very little effect on volatility.

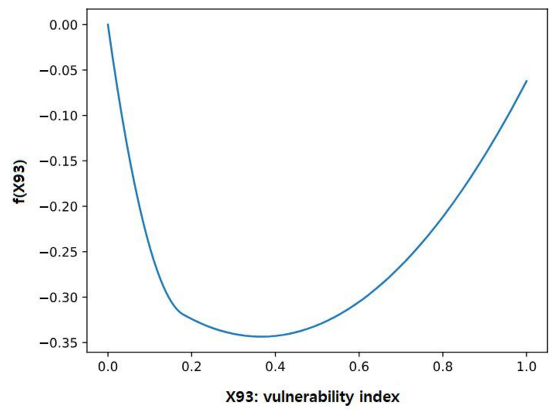
Figure 2.4b illustrates a sharp decline in volatility at lower values of vulnerability, followed by a steady increase with growing vulnerability risk. There is a slightly steeper rise at high vulnerability risk values. The figure encapsulates the intricate relationship between volatility and vulnerability—there is no straightforward linear pattern. The pattern varies depending on the level of vulnerability, and the next section will explore this in more detail.

2.4.6 Target risk groups

The non-linear relationship between volatility and vulnerability risk reveals distinct target groups that may require different approaches from both modelling and customer management perspectives. A potential segmentation strategy can be based on Figure 2.4b, with splits determined by the values of the vulnerability index. For instance, customers with a vulnerability index up to 0.4 may be classified as low-risk, with further sub-segmentation based on high or low volatility. Generally, this group may necessitate monitoring with minimal need for interventions. However, a subset with high volatility might require a different modeling approach for cash flow forecasting. For values ranging from 0.4 to 0.7, customers can be classified as medium-risk. Finally, the highest-



(a) X63



(b) X93

Figure 2.4: Marginal effects of X63 and X93

risk group consists of accounts with a vulnerability risk of 0.7 or above and high volatility. These accounts pose the risk of deteriorating financial circumstances, coupled with unstable behaviours, complicating predictions for future planning and interventions.

Special attention is warranted for the medium-risk group—accounts with moderate vulnerability risk and relatively high volatility. This group requires close monitoring for signs of further deterioration. Analysing transactional behaviours in both the high and medium-risk groups can help interpret the sources of risk and inform the development of effective mitigation strategies.

For accounts with low or no vulnerability risk, monitoring transactional patterns to identify drivers of (low/high) volatility serves as a preemptive measure to prevent vulnerability. Vulnerability is a complex personal financial risk, where individuals with currently low risk levels can face financial shocks and become vulnerable due to external factors. Hence, transparent explanations of the non-linear structures are crucial in identifying vulnerability-prone characteristics. Risky transactional patterns that exert a stronger impact on volatility, such as unusual use of overdraft or income shocks, can indicate an increased vulnerability to future uncertainties.

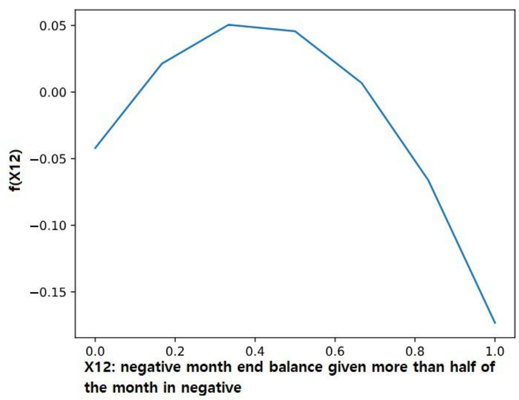
2.4.7 Marginal effects vs ALE

While SPLINE reports competitive predictive performance, machine learning algorithms hold a slight edge in Type 3 models. Therefore, it is crucial to consider the ways their outputs can be interpreted. We compare the marginal effects of spline regression with the post-hoc explanations obtained from ALE plots in XGB. Using several illustrative examples, we identify similarities and differences in non-linear patterns as reported in the two models. Subsequently, we emphasise the effectiveness and potential pitfalls of SPLINE and ALE in capturing underlying non-linearities.

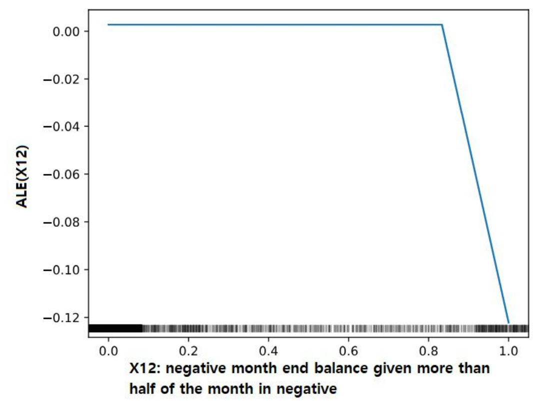
Both SPLINE and ALE show similar non-linear shapes in most covariates, as illustrated in Figures 2.5a-2.5f. Although the effects on captured covariates are not identical, both models demonstrate a high resemblance in terms of the overall trend (non-monotonic increase/decrease) and the presence of inflection points indicating changes in volatility risk levels. The marginal effects of SPLINE portray 'smoother' curves than ALE, attributed to the ALE approach's computation of local effects in grids. Given that SPLINE and XGB have completely different training procedures, a certain degree of variation in monotonicity portrayed here is expected.

On the other hand, certain covariates have disparate interpretations between SPLINE and ALE. Examining Figures 2.6a-2.6f, for instance, (i) ALE depicts a similar increasing trend with SPLINE at low values of X5 but continues to rise instead of declining after the turning point; (ii) SPLINE displays a curve in X with both rising and falling segments, while ALE shows a stepwise increase followed by a flat pattern; (iii) SPLINE illustrates a U-shaped relationship in X93, whereas ALE depicts several flat steps within its increasing and decreasing patterns. Several potential reasons may elucidate these distinct non-linear structures. Firstly, it could be attributed to the inherent capability of each model to capture patterns during training. Secondly, a closer look at the 'mini-bars' in the x-axis of ALE plots reveals that covariates showing differences have skewed distributions, resulting in the issue of insufficient data instances to adequately convey the concealed patterns.

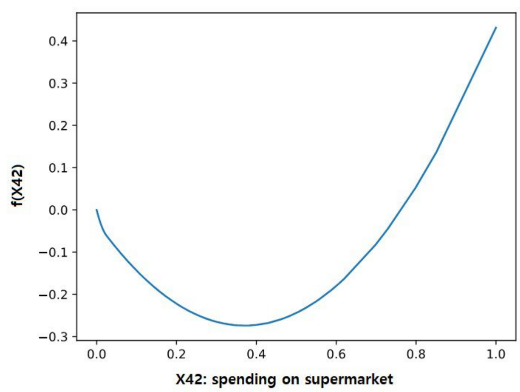
The high resemblance in non-linear patterns between SPLINE and ALE validates the effectiveness of both methods in uncovering hidden structures within the covariates. This implies that obtaining robust and interpretable predictions is feasible, regardless of one's inclination towards statistical parametric techniques or non-parametric machine learning. However, SPLINE's marginal effects hold an advantage by offering transparent explanations of how changes in transactional patterns are anticipated to impact changes in financial volatility. These effects are directly imbedded in model estimation and closely align with standard regression interpretation. In contrast, while ALE also provides appropriate interpretations, its post-hoc approach delineates how covariates influence



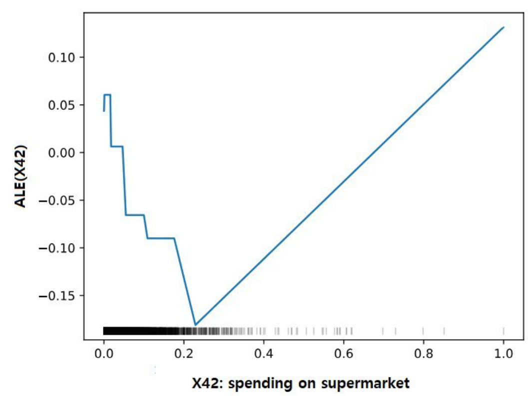
(a) SPLINE: X12



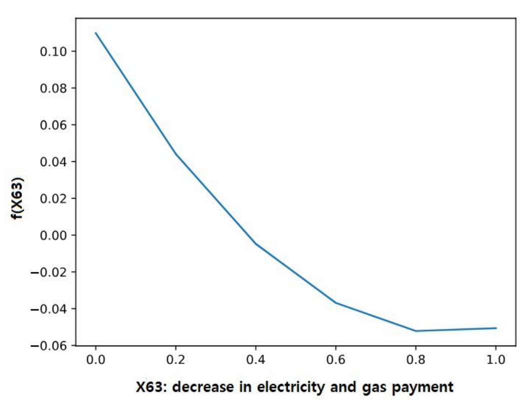
(b) ALE: X12



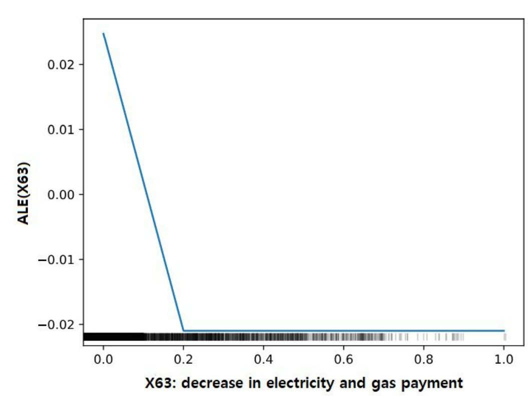
(c) SPLINE: X42



(d) ALE: X42

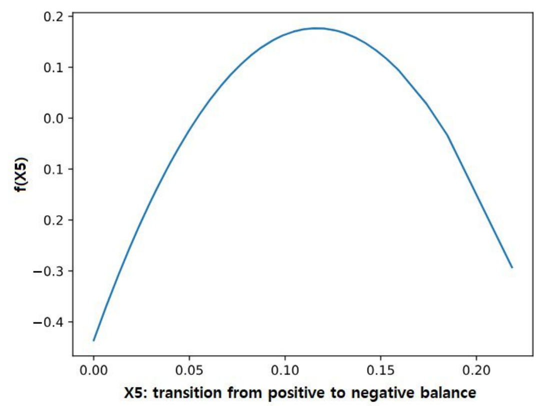


(e) SPLINE: X63

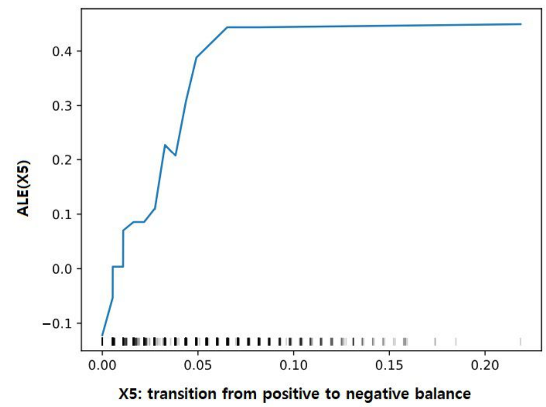


(f) ALE: X63

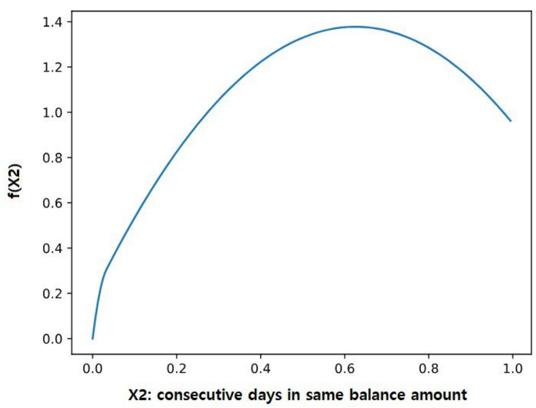
Figure 2.5: Similar patterns between marginal effects and ALE



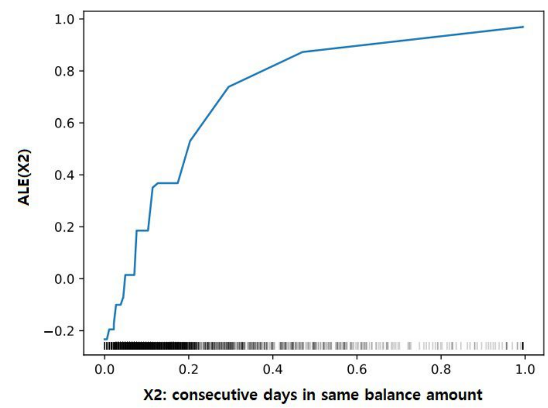
(a) SPLINE: X5



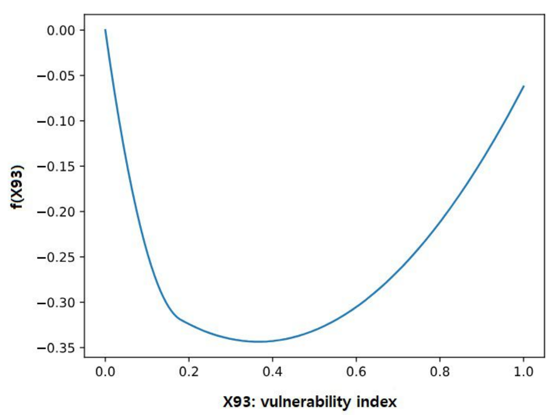
(b) ALE: X5



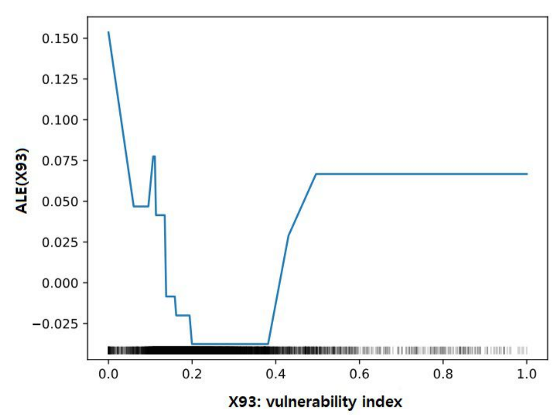
(c) SPLINE: X2



(d) ALE: X2



(e) SPLINE: X93



(f) ALE: X93

Figure 2.6: Different patterns between marginal effects and ALE

predictions on average over specified intervals. Nonetheless, it is crucial to acknowledge that the interpretation of the effect remains local and contingent on the intervals chosen.

The divergent patterns identified between SPLINE and ALE underscore potential pitfalls when employing marginal effects for white-box explanations. As evident from ALE, covariates with skewed distributions can distort their interpretation due to low frequencies in specific ranges of the distribution. Similarly, the marginal effects for these covariates may lack complete reliability owing to insufficient data points at the tail of the distribution. Therefore, interpretations of such covariates should be accompanied by a cautionary note, highlighting the potential for misleading shapes. Users are advised to proceed with further investigations if necessary.

2.5 Conclusion

This chapter explores an approach for predicting account volatility, assessing personal financial risk through (i) past volatility to gauge cash flow behavior stability, (ii) a diverse array of transactional patterns linked to vulnerability and driving risky volatility, and (iii) the vulnerability index.

Both past volatilities and vulnerability serve as crucial predictors of future volatility. Vulnerability enhances predictive power beyond past volatilities, shedding light on behavioural patterns that may lead to financial problems. While fluctuations in previous periods directly influence future behaviour, they are insufficient for distinguishing high or low financial risk. Despite a modest increase in predictive power, vulnerability predictors provide valuable insights into the specific circumstances behind unstable future cash flows. Several risky volatile situations are uncovered, linking detrimental high volatility to low financial capacity, sudden balance changes, financial instability from excessive debt, low spending capability, inconsistent financial practices, and unhealthy addictions. Conversely, harmful low volatility is associated with a reduced transactional level, which may also signal financial problems.

While the vulnerability index does not significantly affect future volatility, its non-linear impact on volatility forms the basis for segmenting accounts into different target risk groups. To ensure that past vulnerability does not persist to disrupt future cash flow stability, financial institutions should understand the source of risk and establish mitigation strategies. Monitoring the presence of risky volatile behaviours prone to vulnerability is crucial to prevent further deterioration of vulnerability in the low-risk group.

This chapter demonstrates that high predictive accuracy can be achieved with various models, offering users a choice of algorithms based on their preferences. Spline regression proves to be a robust model for predicting volatility, competitive with ML approaches. Spline regression excels in interpreting complex relationships between transactional patterns and volatility, unveiling implicit financial risks. The similar patterns illustrated by spline regression's marginal effects and accumulated local effects plots confirm its ability to handle non-linearities as effectively as machine learning models. Moreover, explanations from spline regression are integrated into the model training, in contrast to post-hoc estimation from predicted values in ALE.

Despite its insights into the complex relationship between volatility and vulnerability, this study has limitations. First, the OB data used provides transactions at the account level rather than the individual level. Future research could extend the volatility risk assessment to multiple accounts at the individual level. Second, the OB data provides the most recent 12-month transactional records, limiting the current experiment to short-term risk evaluation. Further investigations are needed to understand the long-term impact of transactional behaviours on volatility. Third, the risk assessment framework relies solely on OB data, potentially overlooking aspects of individuals' financial situations, such as credit and arrears status or ownership of financial assets. Future works could incorporate additional data or extend the current cross-sectional approach to a panel study to track dynamic financial behaviours.

The proposed vulnerability indicator has several limitations that should be addressed in future research. This composite index incorporates only a limited set of representative attributes, which may not fully capture all aspects of vulnerability. A more multidimensional approach is needed to ensure that these attributes effectively represent vulnerability. For example, summing healthcare expenses may not accurately reflect poor health, as these costs might include routine health maintenance (a high sum might be for good health) or be covered by insurance reimbursements (the medical costs indicating poor health shall look into insurance claims). While the current index provides an objective measure of vulnerability, future research could benefit from incorporating subjective factors, such as mental stress or insecurity, to offer a more comprehensive representation of financial vulnerability.

Another potential avenue for future research involves exploring the "stylised facts" pertaining to personal account volatility. Analogous to the volatility observed in asset market prices, which has been extensively studied in the literature and characterised by specific statistical properties, the realm of personal finance should similarly investigate the distinct characteristics of volatility derived from bank account transactions. Such an investigation serves as a strategic approach to enhancing modelling methodologies and deepening our understanding of consumer cash flow behaviour.

3 Assessment of distress risk transitions from vulnerability with a statistical latent Markov framework

The preceding chapter delves into an evaluation of cash flow volatility risk within a specific quarter, utilising a volatility metric to gauge instability risk at a fixed time point. However, recognising the dynamic nature of financial behaviours, this chapter extends its scope beyond a single quarter to examine how financial risk evolves across quarters.

3.1 Introduction

In recent years, consumer protection has become the primary focus for financial firms to ensure all customers receive fair treatment to access existing products to support their financial lives (CMA, 2019). Firms are expected to pay specific attention to the vulnerable group because they are underprivileged to engage effectively in the market with a high likelihood to obtain poor deals (Financial Conduct Authority, 2015). Moreover, vulnerability is the precursor of financial distress. Each vulnerable characteristic indicates a certain degree of financial difficulty which builds up the exposure risk to distress i.e., an extreme state of severe hardship. Early consumer engagement is particularly helpful to alleviate vulnerability and prevent distress (Financial Conduct Authority, 2021a). Adaptation to rising costs of living exacerbates the worsening of financial situations (Financial Conduct Authority, 2023b). Predicting the transition of financial conditions across varying distress risk levels from vulnerability is thus vital to detect potential harms at the very beginning for risk mitigation.

The existing practice to support vulnerable consumers has several limitations. Qualitative analysis to track the progression of financial positions over time with self-reported information could be problematic because individuals may not diagnose themselves as vulnerable (Financial Conduct Authority, 2015). It is also difficult to pinpoint the starting point of financial difficulties which eventually leads to delayed actions from firms (Collard, 2011). Hence, a quantitative approach to capture risk in customers' journeys over time with bank transactions is the recommended alternative strategy (Experian, 2019; Financial Conduct Authority, 2018). Open Banking (OB) creates a new opportunity for third-party service providers to access bank transaction data, facilitating detailed account-level analysis of changes in financial behaviour based on cash flow patterns. While Open Finance (OF) encompasses a broader range of financial data—including credit, insurance, and pensions—beyond just cash flow transactions, OB serves as the foundational framework for OF. The OB risk assessment framework is designed to provide customer-centric solutions based on direct consumer behaviour, enhancing consumer confidence in data sharing practices and supporting the advancement of OF.

Such a OB data sharing framework brings challenges on how to exploit vulnerability and distress from such new data as well as how to model the changing financial circumstances from the dynamic transactional patterns. The implementation of OB transactions to develop a risk assessment tool for customer support raise the following questions: (i) how to commensurate vulnerability from income and expenditure behaviour in bank transactions? (ii) How to define the varying stages of financial distress risk from cash flow transactions? (iii) How does an account move between the distress risk state i.e., improve or deteriorate over time? (iv) What are the vulnerable characteristics that drive the financial transitions?

An effective framework should account for the different degrees of risk as financial circumstances progress over time. Such a risk model has not been established in the OB field. Due to the well-developed dynamic risk assessment in the credit domain, these credit models serve as inspiration for analogous applications with OB transactions. There are two key approaches to represent changes in credit status throughout the book record duration i.e. multistate intensity model to estimate transition probabilities with the time-to-event survival analysis by capturing the occurrence of event-of-interest for the first time in the observed time window (Santos, 2018), as well as the cohort Markov model where the distribution of transition probabilities based on the Markov property such that the movement to a certain state depends on the current position (Ferretti et al., 2019). Several studies also compare both methods to obtain better credit risk transition estimations (Lando & Skødeberg, 2002; Schechtman, 2013; Schuermann et al., 2003). To understand the factors influencing the transitions, several studies employed the model specification to include covariates in the estimation of transition probabilities (Duffie et al., 2007; Kelly & O'Malley, 2016; Bocchio et al., 2022; Leow & Crook, 2014; Djeundje & Crook, 2018; Malik & Thomas, 2012; Frydman & Matuszyk, 2018; Frydman et al., 2019).

Modelling the fluid movements of financial circumstances in bank transactions have two differences compared to the credit risk migration. First, the existing dynamic credit models have explicitly-defined states such that default follows the formal 90 days past due criteria and the delinquency states are measured by arrears or behavioural scores. Yet, distress is not directly observable from cash flow because there are no specific criteria to quantify the varying degrees of exposure risk to distress. Second, the transition between states is estimated in a closed system with the presence of a permanent absorbing state where there is no path to exit i.e., default. In contrast to credit portfolios, information for a stagnant non-reverting distress is not available. For instance, accounts constantly in negative balances might not necessarily be completely unaffordable as long as consumers can handle the overdraft interests to keep up with living. Hence, it is ambiguous to claim consistent negatives as distress in absence of credit arrears or debt advice usage information.

We aim to assess the evolution of the underlying distress risk over time from vulnerabilities with OB transactions. The latent Markov model is a suitable solution to handle the above-mentioned contrasting points for OB application because (i) the implicit distress risks can be exploited via the estimation of the underlying latent states from the observed (account balance) data, and (ii) the estimation of the distress risk (latent states) evolution is not restricted to specify any absorbing state. Using the latent Markov model, we extract the implicit financial states from three facets of observed account balance information i.e., duration in negative balances, significance of the magnitude of negative amount (whether the negative value is large enough to deplete the available balance), and fluctuations of the balance amount. The discovered latent states serve as a strategy to quantify the varying stages of distress risk, which are then used as the target variable of interest, to track the quarterly changes over the year. Specifically, we focus on the extended latent Markov to also model the effects of covariates on the transitions. We use the vulnerable characteristics in the current quarter as the covariates to estimate and predict the transition of distress risks (improvement/deterioration) in the next quarter. The objective measures of vulnerabilities, covering the aspects of poor health, bad financial literacy, negative life events, and low financial capacity, are computed from the money inflow and outflow. The marginal effects of the significant covariates are analysed to reveal the vulnerability drivers of financial improvement and deterioration. We also predict the most likely sequence of the distress risk states over the quarters to assess how financial risks shift across time. By interpreting the evolving distress risks and the vulnerability drivers together, we differentiate transient vs permanent vulnerabilities and flag transitions that may lead to financial instability.

To the best of our knowledge, Medina-Olivares & Calabrese (2023) is the only study that has utilised changes in OB income and expenses to assess financial vulnerability transition risk using a latent Markov model. They examine the relationship between the number of payday loans taken out and the level of financial vulnerability, defining vulnerability as a state of being trapped in debt within their latent framework. Our proposed framework introduces several novelties to the OB literature. We define distress exposure risk based solely on account balance using OB data, without relying on credit usage information. We also present a novel approach by representing dynamic covariates of vulnerable circumstances derived from cash flow transactions, rather than using income and expenditure behaviours as covariates in risk predictions.

This research provides three key contributions to the OB literature. First, the pioneer application of the latent Markov model set up a data-driven framework to formalise the quantification of distress risk from account transactions. Such measurement scales the varying stages of distress at every time point to provide an objective assessment for enhanced diagnostic of undetected financial struggles due to misperceptions in subjective measures. Second, we estimate and predict the transition of distress risk state in the next quarter based on vulnerability in the current quarter. The quarterly predictions can monitor the migration path of financial risks to provide early warning signals of potential deterioration. Third, the relationship between vulnerability and risk transitions reveals the account-specific drivers of a change in financial status. Thus, practitioners can effectively promote consumer protection by designing customer-centric risk mitigation strategies to alleviate existing risks and guide financial improvement.

This chapter has the following structure. Section 3.2 reviews previous related works. Section 3.3 describes the data, outline the procedures to measure distress and quantify the covariates from transactional patterns as well as detail the estimation steps in multivariate latent Markov model to capture evolution of financial states over time. Then, Section 3.4 reports and discusses the obtained findings. Lastly, Section 3.5 concludes the research.

3.2 Literature review

3.2.1 Financial vulnerability and distress risk

Financial vulnerability (FV) is intrinsically linked to overall financial wellbeing. It represents the probability of encountering hardship in the future, with hardship manifesting as the actual inability to sustain one's standard of living (O'Connor et al., 2019). Recognising FV is imperative to identify early indicators of financial distress, allowing for proactive risk mitigation during the initial stages. Regulators underscore the significance of identifying FV, positioning it as a crucial element in consumer protection efforts. This ensures that financial institutions can proactively devise and implement interventions, guaranteeing equitable treatment for vulnerable consumers, as emphasised by the Financial Conduct Authority (FCA) (Financial Conduct Authority, 2015). In recent years, financial firms have actively embedded FV detection as a channel to explore the opportunities for timely intervention of financial hardships (Financial Conduct Authority, 2015; O'Connor et al., 2019).

Firms commonly assess consumer financial situations through subjective self-reported information. Individuals deemed potentially vulnerable often proactively reach out to these firms to communicate their circumstances and request financial assistance (Financial Conduct Authority, 2015). In the study by Collard (2011), it is noted that, beyond responding to consumer calls, firms take a proactive approach by reaching out to customers when specific financial difficulty risk indicators are triggered. These indicators include irregularities such as consistent use of unarranged overdrafts, changes in current account behaviour, or missed overdraft fee payments. The implementation of such early engagement strategies has proven successful in preventing the accrual of over-limit overdraft charges, aiding customers in reclaiming control over their financial management. Clients supported through this initiative have emphasised the importance of initiating contact at an earlier stage, well before overdraft charges accumulate or any fees are imposed. However, the reliance on self-perception to convey financial conditions introduces potential biases. Individuals may not recognise themselves as vulnerable, and staff interpreting these self-reports need to discern nuances in tone, keywords, and pitch to accurately understand the situation. Moreover, there is a delay in capturing financial difficulties, as some individuals may be overconfident and only reach out to firms well after the problem has initiated. Detecting FV from subjective data sources hinders proactive support due to the challenge of precisely pinpointing the onset of vulnerability. O'Connor et al. (2019) also highlight a disconnection between perceived financial conditions and actual financial states, emphasising the need to integrate both subjective and objective measures in assessing FV. The existing approach predominantly incorporates objective measures such as an individual's assets, credit, insurance, workforce, or education, without delving into daily cash flow patterns. Therefore, Open Banking (OB) emerges as a novel data source that facilitates the quantification of income and spending behaviours, enhancing the evaluation of FV.

Nonetheless, the literature lacks comprehensive exploration into constructing a FV measure solely from cash flow data. The FCA guidance, as outlined by Financial Conduct Authority (2023a), directs firms to identify vulnerability characteristics based on four primary drivers: the impact of poor health on daily tasks, occurrences of life events such as bereavement or job loss leading to mental stress and disrupting financial lives, low resilience without the ability to handle financial or emotional shocks, and low financial capability with inadequate money management skills for handling financial matters. While this guideline provides valuable insights into potential vulnerabilities, it does not establish specific rules for quantifying income and spending patterns to capture FV. A notable exception in the literature is the work by Kim et al. (2023), which extensively investigates FV dimensions using bank transaction data. Six indicators of FV, encompassing financial stability, capacity, and management ability, are derived from transaction records. This study highlights the efficacy of OB data in representing FV, even to the extent of inferring protected characteristics, underscoring the need for cautious use of OB data in risk assessment. However, these OB-specific indicators, while offering an objective representation of FV, are categorical and lack the ability to quantify the magnitude of FV in a numerical measure. This chapter aims to address this critical gap by formulating numerical attributes derived from cash flow transactions as the risk measures for FV.

The temporal dynamics of FV remain largely unexplored. Existing risk models have concentrated on projecting current behaviour to anticipate future risks, neglecting the intricate movements of financial circumstances over time. Kim et al. (2023) introduce financial behaviour attributes to forecast potential FV in upcoming periods, revealing that transactional patterns exhibit high predictability up to three months ahead, serving as effective early warning signals. However, this modelling framework forecasts FV at specific future time points rather than continuously tracking dynamic financial circumstances. This limits the ability to identify FV at its origin, thereby impeding timely support before FV transforms into actual hardships (Collard, 2011). Real-time FV identification through consistent tracking of financial stress throughout the entire consumer engagement lifecycle has been the key goal for firms aim to protect vulnerable consumers (Experian, 2019). OB holds the potential to provide valuable insights into consumers' transactional trends, offering a critical examination of emerging patterns. By monitoring changes in transactional records, firms can design interventions based on real behaviour and anticipate problematic situations (Financial Conduct Authority, 2023a). This chapter strategically focuses on capturing the dynamic nature of financial circumstances using OB data, aiming to continually track ongoing financial conditions and enhance the ability to identify and respond to evolving FV.

The connection between FV and the transitions in distress risk remains an unexplored domain. According to O'Connor et al. (2019), FV encapsulates the likelihood of experiencing hardship. Viewing distress risk as a spectrum spanning different degrees of hardships, a higher (lower) prevalence of FV characteristics correlates with an exacerbation (amelioration) of distress risk. Modelling the intrinsic relationship between FV and the evolution of distress encounters two main limitations. Firstly, existing literature has proposed numerical scales to assess financial conditions, yet none specifically highlight financial challenges related to cash flows. For instance, Consumer Financial Protection Bureau (2017) formulate a financial wellbeing scale by gauging individuals' perceptions of fully meeting financial obligations, feeling secure about their financial future, and having the capacity to make life-enjoying choices. Employing factor analysis and item response theory, they amalgamate a ten-item scale to derive the final wellbeing score, presenting a complementary measure to income, credit scores, and other financial condition metrics. Similarly, Gladstone et al. (2020) scale financial wellbeing using self-reported responses on financial control, ability to meet debt repayments, freedom to enjoy life without financial worries, and the burden of keeping up with commitments. Although they identify significant correlations between measures of bank account records and subjective wellbeing, a bank-transaction-based metric quantifying the extent of underlying financial difficulty is not developed. Gathergood & Guttman-Kenney (2016) define distress in the credit context by combining arrear payments records with self-perceptions of debt repayment burdens, running out of money, and managing bills. However, this definition overlooks cash flow patterns in measuring distress. Recognising the fluid state of financial circumstances, this chapter addresses these limitations by deriving metrics from account balances. This approach allows for the measurement of varying levels of financial struggles, providing a comprehensive representation of different distress risk states.

Secondly, existing methodologies for personal finance consumer risk lack modelling transitions into improved or deteriorated financial states. In a cross-sectional study, Gladstone et al. (2020) identify associations between bank account records—such as higher income, increased liquidity, and fewer days in overdraft—and higher financial wellbeing. Utilising both linear regression to examine significant correlations with wellbeing and non-parametric b-spline regression to explore complex non-linear relationships in objective bank records, the study offers alternative interpretations that guide attention toward consumers in need of additional support. However, the cross-sectional model only provides a snapshot of financial circumstances, failing to incorporate the dynamic and intricate processes of financial behaviour. On a similar note, Gathergood & Guttman-Kenney (2016) identify the debt-to-income ratio as a significant predictor of financial distress using a linear regression model. They extend their analysis with a fixed effect panel linear model to predict the occurrence of future distress one and two waves ahead. While this panel model assesses whether covariates at a specific time point can forecast future distress occurrences, it lacks the ability to track evolving patterns of financial states, either improving or deteriorating. Due to the survey data collection by waves, approximately spanning two years, the risk model in their study offers a long-term forecast, which may not be suitable for timely alerts regarding problematic conditions. Although the OB data contains transactional records that inform day-to-day cash flows, the personal finance literature lacks the development of a dynamic risk assessment model.

3.2.2 Modelling transitions of financial risks

Given the scarce literature on the OB consumer risk model, we draw inspiration from credit scoring methodologies to evaluate financial transition risks using OB data. Earlier studies have primarily focused on the comparison of various modelling techniques to establish a comprehensive framework for monitoring credit risk movements. The cohort Markov model, recognised as the standard approach for modelling risk transitions, operates under the assumption that credit transitions adhere to the Markov property. This property posits that the probability of each event is dependent solely on the state of the preceding event. Each research addresses distinct limitations associated with the cohort (discrete) Markov model and explores potential alternatives to enhance the precision of risk transition predictions.

Credit data can face censorship challenges when firms maintain incomplete records throughout the entire observation period due to early withdrawals or late entries. Consequently, the duration approach is considered an alternative to the cohort approach, aimed at incorporating all available data in the estimation process. The survival technique models the time-to-event process to predict transition intensities, thereby capturing dynamic risk movements. In the corporate credit landscape, Schuermann et al. (2003) delve into both the cohort (discrete) Markov and duration (survival) approaches to estimate firms' rating migration matrices. Their findings highlight significant statistical differences in the migration matrix between these two techniques, showcasing their effectiveness in capturing dynamic properties. Importantly, these statistical disparities directly translate into significant economic differences in estimating credit risk capital and credit derivative pricing. The study concludes that the survival approach offers a more efficient estimation compared to the cohort approach, mitigating the risk of over or underestimation. Similarly, in the consumer credit literature, Schechtman (2013) employ both cohort and duration techniques to estimate default matrices for consumer loan portfolios in four Brazilian banks. These matrices cover transitions not only into the past 90-day default but also recovery and multiple default severities. The authors introduce a mobility metric to assess the differences in delinquencies and their dynamics among the banks. Survival-estimated transition probabilities prove to be larger for typical default and recovery migrations, indicating higher efficiency gains compared to discrete estimation. The evaluation of mobility metrics contributes to detecting distinctions in the trajectories of states over time. Worsening metrics serve as early warning indicators for specific default definitions or can be used to radically reverse the behaviour of a migration path. Additionally, these metrics aid in identifying the most similar or dissimilar credit risk distances between banks to support supervisory purposes.

Apart from the survival model, Ferretti et al. (2019) conduct a comparative analysis involving the cohort Markov model and the mover-stayer model. They emphasise the crucial consideration of population heterogeneity in the credit rating of Italian small and medium enterprises (SMEs), categorising accounts into movers and stayers. The cohort Markov model tends to overestimate credit risk due to the inherent risk of assuming having an absorbing state in the pure Markov chain, wherein all accounts eventually default, thus resulting in excessive regulatory capital. Consequently, the mover-stayer model is proposed to estimate separate processes for the stayers and movers. The comparative assessment demonstrates that the mover-stayer model exhibits lower errors than the Markov chain for a more accurate analysis of the credit ratings evolution, facilitating an effective resource allocation process and enhancing the stability of the banking industry. Conversely, Santos (2018) exclusively concentrates on the continuous-time multi-state Markov model, without comparing with other techniques, to monitor the progression of consumer loans across recovery, restructuring, or default. The transition intensity matrices, in general, indicate a higher probability for loans to recover than to default. The estimated mean sojourn time in each state reveals that contracts tend to remain in a performing state for 17 to 20 months before transitioning to other, less favourable scenarios.

These studies have laid the groundwork for assessing credit risk transitions. However, the predicted sequence of risk states in these models is solely based on past status occurrences, overlooking the influence of external factors. This represents a significant limitation when attempting to apply these dynamic credit risk models to estimate OB-specific distress risk transitions. Understanding not only the predicted state sequences but also the factors driving these status shifts is crucial for practitioners to diagnose problems and implement appropriate remedies. To explore how FV patterns relate to shifts in distress status, incorporating covariates in the model estimation process is a key desired property for dynamic risk assessment. In the following subsection, we review these models, aiming to seek suitable applications in the context of OB.

3.2.3 Including covariate effects in the assessing the risk transitions

Credit status shifts often result from external factors. To incorporate individual-specific effects into the estimation of credit risk movements, the literature has expanded model specifications (reviewed in Section 3.2.2) by incorporating a regression model with two key approaches. The duration-based survival model is extended using the Cox proportional hazard specification, while the cohort Markov model is extended with the logit-based specification.

3.2.3.1 Survival-based model

To predict credit transition intensities, the Cox-based method has emerged as a common approach to simultaneously address the duration effect on changes in credit risks and additional characteristics that could impact the credit portfolio. In estimating transitions in corporate firms' ratings, Lando & Skødeberg (2002) incorporate previous downgrade events as covariates into the standard Cox survival method. Their findings reveal a significant downgrade momentum but no detectable effect on the upgrade intensity following a past upgrade. Investigating whether the duration spent in a specific rating influences the intensity of downgrade or upgrade, they show that a longer duration in a particular rating correlates with a lower probability of shifting to a new rating, irrespective of whether it involves an upgrade or downgrade. Instead of using previous rating upgrades or downgrades as input covariates, subsequent studies explore the influence of subject-specific or macroeconomic attributes as potential factors driving evolving credit risks using the survival-based technique.

Duffie et al. (2007) incorporate time-varying firm-specific and macro-covariates into the estimation of corporate default of US-listed industrial firms.. They propose a maximum likelihood estimator for multi-period survival probabilities, blending the Cox-based approach with time series analysis of covariates. This unique approach allows predictions for several additional periods into the future by considering the effects of mean reversion, volatilities, and correlations of the time series of covariates. Focusing on a two-state transition, their findings reveal a significant relationship between the default hazard rate and the economic state, particularly the distance-to-default, a volatility-adjusted leverage measure of the firm. Notably, their out-of-sample default predictions demonstrate a higher accuracy ratio compared to results published by Moody's, highlighting the effectiveness of including covariates in enhancing predictive power.

In mortgage lending, Kelly & O'Malley (2016) employ the standard Cox survival model to estimate default and cure probabilities. Their model includes loan characteristics and macroeconomic conditions, encompassing affordability and housing equity effects, to capture specific two-state transitions: default (moving from performing to default) and cure (transitioning from default to performing). The estimated coefficients highlight that an increase in affordability and housing equity effects is associated with a higher probability of default. Notably, labour market deterioration carries more significant weight than housing equity effects. When it comes to cures, the study reveals a scarring effect, indicating that the probability of a loan returning to a performing status decreases when it remains delinquent. While this in-depth analysis of covariates aligns with stress test requirements and provides valuable insights into understanding risk evolutions, it's important to note that focusing solely on two-state transitions, as observed in studies like Duffie et al. (2007) and Kelly & O'Malley (2016), may be insufficient for early warning. This limitation arises because the predictions directly pertain to the movement towards the severe default state rather than capturing milder risk states before a default occurs.

Several studies have undertaken modelling transitions across multiple risk states. Leow & Crook (2014) employ a Cox-based intensity model to predict the probability of delinquency and default for credit card loans, defining varying credit risk levels such as up-to-date, one month in arrears, two months in arrears, and default. The transition matrices are influenced by application variables and monthly behavioural patterns. Analysing factors affecting movements between states reveals unique insights compared to past literature. Notably, self-employed or unemployed debtors tend to stay out of default due to better account balancing abilities, while employed debtors with stable income face a higher probability of moving into default during unexpected income loss. The model performs well in predicting delinquent vs. non-delinquent states but exhibits limitations in predicting states between delinquency levels. Misclassifications tend to be on the conservative side, resulting in lower-cost consequences, where accounts are predicted to be in a lower state than they should be. Building upon the multistate delinquency framework in Leow & Crook (2014), Djeundje & Crook (2019a) extend the model by incorporating heterogeneity and macroeconomic variables, in addition to application and behavioural variables. A random effect term is introduced in the standard survival model to mitigate the impact of unobserved covariates and account for the dependence between transitions within the same account, aiming to obtain unbiased parameter estimates. Most application and behavioural variables exhibit expected signs with state movements, and macroeconomic covariates also demonstrate significance. Higher house prices, lower retail prices, and lower credit card interest rates correlate with a greater chance of recovery. However, compared to the standard Cox survival model, including random effects diminishes the significance of covariates and does not improve predictive accuracy.

In line with the approaches taken by Leow & Crook (2014) and Djeundje & Crook (2019a), Bocchio et al. (2022) introduce a survival-based multistate intensity model with covariate effects for mortgage loans. Their unique contribution lies in emphasising the consideration of recurrent events, as subjects may transit the same path more than once. This novel model specification addresses the complexity of handling multiple types of failures or events. The multistate model offers nuanced insights into the transitions between various levels of days past due (delinquency) and movement toward default. This enables lenders to understand the expected time an account remains current on repayments before missing any payments that could lead to default, thereby enhancing the estimation of expected credit losses. Several significant attributes associated with movement into delinquency or default include accounts with high current loan-to-value, higher outstanding balance, higher loan interest rates, elevated unemployment rates, and reductions in GDP growth.

3.2.3.2 Markov-based model

In addition to the survival approach, the Markov-based model stands out as another mainstream technique to include covariate effects for estimating credit risk movements. Malik & Thomas (2012) pioneers a Markov chain-based model with a cumulative logistic regression specification to capture the dynamics between behavioural score bands and the default state within a consumer loan portfolio. Economic variables, the age of the loan, and past score behaviour are incorporated into the estimation of the transition matrix, contributing to a deeper understanding of the drivers behind these transitions. Their results unveil that being new on the books and an increase in interest rates pose a greater risk of default or downgrading the behavioural score. Further investigation using a second-order Markov chain reveals the propensity of accounts to reverse and return in the direction they came from, indicating no momentum effect of deterioration on future credit risk. This finding contrasts with corporate ratings, as observed by Lando & Skødeberg (2002). The inclusion of additional drivers significantly enhances the accuracy of predicting default and high-risk states. Notably, for the second-order model, the accuracy improves primarily in predicting low-risk categories.

The mover-stayer (MS) model presents another Markov-based approach, particularly useful for addressing population heterogeneity within a credit portfolio, where stayers and movers follow distinct processes. Frydman & Matuszyk (2018) expand upon the conventional MS model by specifying stayers' probabilities as the logistic function of applicants' time-fixed attributes. This modification aims to predict the creditworthiness of instalment loan borrowers based on individual characteristics. The study finds that higher odds of being a stayer are associated with a positive credit score, higher education, a known residence, and supporting at least one family member. Leveraging these specific patterns, the study classifies accounts into stayers and movers, with stayers considered creditworthy due to their higher likelihood of repaying instalments promptly. Building on this research, Frydman et al. (2019) further extend the model to include covariate effects in the transition matrix. This enhancement provides a detailed understanding of the factors influencing movers' propensity to transition from the initial state to delayed payment or default. The study identifies lower odds of being a mover for borrowers taking a loan for a new car and those with a longer expected time to stay in a particular state.

While survival and Markov-based techniques with regression specifications are mainstream methods for investigating the specific drivers of credit risk transitions, deploying them to examine the influence of FV on the evolution of distress risks in the OB context faces two key restrictions. Firstly, these models monitor states of credit risk, defined by clear criteria like credit ratings, scores, or days past due. However, there are no clear definitions of distress states derived from cash flow transactions to serve as target variables in these models. The varying degrees of hardships are not directly observable from cash flow but are implicitly represented by the account balance situation. Secondly, state transitions in these credit models are estimated within a closed system, typically including a permanent absorbing state, often defined as 'default' in the credit domain. However, identifying such an irreversible state of distress is not straightforward. Continuous overdraft use (negative balance) does not necessarily indicate complete unaffordability, as observed in credit default scenarios, and may function as a strategy to manage daily expenses. Therefore, modelling distress risk movements from FV requires the ability to extract underlying states of financial difficulty from cash flow balances and enable the flexibility to move freely between all states without a non-reverting extreme state.

3.2.3.3 Latent Markov approach to assess dynamics of implicit risk states

The latent Markov model becomes crucial when the emphasis is on evaluating credit portfolio risk based on hidden states rather than observable credit states. In a study by Thomas et al. (2017), an estimated hidden Markov chain model leverages observed bond ratings and interest rates to uncover the underlying economic condition process, presumed to be mutually dependent on the two observables. This model identifies two latent states, categorising economic conditions as either good or bad. Similarly, Yu et al. (2019) delve into hidden economic states governed by a Markov chain of default sequences across different sectors. They compare three distinct reduced-form intensity-based models with a hidden Markov process to trace the movements of economic conditions over time. Examining the default sequences from the three sectors of consumer and service, energy and natural resources, as well as the leisure and media, they discern between good and bad economic states. The results align with real-world economic states, underscoring the model's efficacy in capturing hidden features and simulating credit default risks. However, it's important to note that applications by Thomas et al. (2017) and Yu et al. (2019) lack the incorporation of attribute effects, limiting the exploration of the drivers behind latent state movements.

Building upon the latent Markov framework, both Blümke (2022) and Medina-Olivares & Calabrese (2023) enhance the model by incorporating covariate effects into the specifications. In the case of Blümke (2022), utilising explicit corporate credit rating grades and the number of defaults from Standard & Poor's, they construct a hidden Markov model to objectively determine the number of implicit economic scenarios and predict scenario probabilities. The model identifies four latent states representing economic scenarios ranging from recession to stable economic growth. The model fit demonstrates a moderate improvement when the estimation of transition probabilities is conditional on macroeconomic covariates.

Medina-Olivares & Calabrese (2023) introduce a mixed Poisson hidden Markov approach to discern the varying levels of FV based on the number of payday loans a borrower obtains. To account for inter-borrower differences, they incorporate OB-derived transactional behaviours as explanatory variables to estimate the latent FV components. The model identifies two hidden states, namely FV and non-FV. An analysis of the covariates reveals that having more non-recurring income is associated with an increased need for payday loans when the borrower is in a state of FV. Conversely, when the borrower is financially stable (non-FV), more non-scheduled income lowers the reliance on payday loans. This nuanced insight provides a deeper understanding of the reasons behind payday lending. By monitoring FV states, the model identifies two dominant patterns: (i) 60% of payday loan borrowers persistently stay 12 or more consecutive weeks in FV, indicating consistent difficulties, and (ii) individuals in non-FV are most likely to persist in this state but exhibit a relatively higher probability of transitioning into FV once they do. This information assists practitioners in detecting high-risk customers who are likely to either not recover or default. While their framework includes OB transactional behavior as a covariate, it is grounded in the credit backdrop to define FV based on reliance on payday loans. To the best of our knowledge, the sole use of cash flow transactions to assess dynamic risks has not been developed in the personal finance context.

3.2.4 Limitation

The current credit risk transition models showcase their capabilities in (i) estimating and forecasting transition probabilities across two or multiple states of credit risk, (ii) incorporating covariate effects to interpret specific transition drivers, and (iii) tracking the sequence of evolving states. Although these models are standard tools in the credit domain, several considerations need to be taken into account for their adapted use in the OB context. The limitations to be addressed are summarised as follows.

Definition of target variable: Defining distress risk states requires extracting account balance situations to represent implicit financial hardships. Unlike most existing dynamic credit models that use credit statuses derived from credit ratings, credit scores, or arrear information as target variables, we aim to analogously extract distress states from cash flow data. While some studies have explored implicit economic scenarios from credit ratings (Thomas et al., 2017; Yu et al., 2019; Blümke, 2022), only Medina-Olivares & Calabrese (2023) has evaluated hidden consumer financial difficulties from payday loan usage. However, the two-state FV in Medina-Olivares & Calabrese (2023) doesn't comprehensively capture struggles in the personal finance context, and the binary representation is insufficient to scale the different stages of financial difficulties. Thus, developing a data-driven framework to extract implicit distress risk states from OB cash flow data remains an unmet challenge in the personal finance risk literature.

Covariates to be included: The integration of covariate effects into the estimation of credit risk transitions has enhanced the accuracy of credit loss estimations and provided insights into the specific drivers of recovery or deterioration. In the credit risk literature, there exists a well-established theory guiding the selection of covariates for investigating status transitions, including subject-specific characteristics (firms or borrowers), repayment behaviours, loan-specific characteristics (age of loan, credit limit, etc.), and macroeconomic variables. In contrast, distress risk transitions lack a quantitative framework to examine drivers comprehensively. While detecting FV holds potential for improving general financial well-being (Financial Conduct Authority, 2021a) and enhancing predictive capabilities for financial hardships (O'Connor et al., 2019), a quantitative framework to investigate how FV drives transitions between distress risk states has not been established for personal finance assessment.

Designing a monitoring tool: Practitioners have emphasised the significance of early consumer engagement to offer timely support, preventing those experiencing FV from slipping into actual distress (Collard, 2011; Financial Conduct Authority, 2023a). The credit risk transition framework relies on accurate forecasts of future transitions and understanding the drivers of specific transitions (upgrade or downgrade) to detect early signs of deterioration. Additionally, an in-depth analysis of the predicted sequence of states, tracking persistence to stay or frequency of movement, enables improved segmentation of low and high-risk portfolios. However, a data-driven monitoring tool that encompasses covariate interpretations and traces the progression of states is yet to be deployed in the OB context to signal initial signs of FV that could lead to further deterioration.

Model choice: We propose the adoption of the latent Markov approach to assess the evolving distress risk stemming from FV using cash flow data. While the credit risk literature typically favours survival and Markov-based approaches for estimating credit transitions, these methods are not suitable for dynamic assessment in the OB context due to the absence of explicit observable target variables, such as delinquency and default states. In contrast, studies employing the latent Markov approach have shown promise in defining hidden scenarios based on observed credit statuses in a data-driven manner. Furthermore, the latent state transitions in this approach are not confined to a closed system, eliminating the need for a permanent absorbing state (e.g., default). This flexibility aligns with the property of cash flow data, where account balances can fluctuate from positive to negative and vice versa. This may be indicative of a cash flow turnover strategy, like using overdrafts, rather than a permanent financial difficulty. However, tailoring the appropriate variant of the latent Markov model to OB risk assessment remains unexplored in the personal finance literature.

3.3 Methodology

3.3.1 Data setup

We work on 19447 personal accounts where their transactions span across 12 months[15]. We aim to detect indications of financial distress within cash flow activities. Distress typically refers to a deficiency in current cash flow to fulfill immediate financial commitments. Specifically, Keys et al.

[15]Details of the data can be found in Table 1.2

(2023) has defined personal financial strain as enduring difficulties in meeting credit obligations. These difficulties can be gauged through early warning signs such as credit card delinquency extending 30 days or more past the due date. Along the severity spectrum, the presence of debt in collections indicates a more critical level, where the credit account is declared as uncollectible and sold. At the far end of the spectrum, personal bankruptcy represents the most extreme form of financial distress. To depict financial challenges stemming from cash flow activities, we focus on analysing the cash flow balance position over an extended period spanning several months. A consistent state of negative balance or a reduced balance amount indicates a shortfall in financial resources. Unlike the conventional practice of monthly assessments in traditional financial reporting to evaluate firms' financial health, our approach centers on a quarterly timeframe. This focus ensures that our balance metrics, assessed over multiple months, effectively capture prolonged periods (beyond a month) of financial hardship evident in cash flow activities, thereby offering a more comprehensive indication of potential strain.

We aim to track how vulnerable transactional patterns impact the motion of distress risk states across the quarters. Specifically, as illustrated in Figure 3.1, we project the movement of distress risk state (detailed in Section 3.3.2) based on the vulnerability-related transactional behaviour (detailed in Section 3.3.3) in the current quarter. The predictive model is trained on 80% of the accounts and tested on the remaining 20% of the accounts. The covariates are lagged one quarter for one-month ahead prediction. Thus, the in-sample and out-of-sample predictions track the sequence of distress states over Q2→ Q3 → Q4 driven by previous transactions in Q1, Q2 and Q3 respectively.

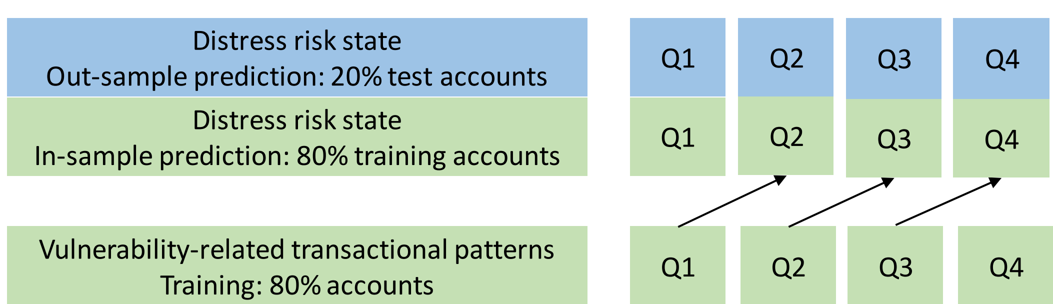


Figure 3.1: In-sample and out-of-sample prediction setup

3.3.2 Definition of distress

A distress metric aims to measure a range of financial challenges, spanning from minor to severe difficulties. In the area of OB, we propose to quantify distress by examining overdraft usage, as negative balances signify a shortfall, essentially representing borrowing from the bank. Given the absence of a definitive distress definition, we adhere to the methodology used in credit scoring, where credit default is assessed based on the number of days overdue and the overdue amount (Ernst & Yong, 2018).

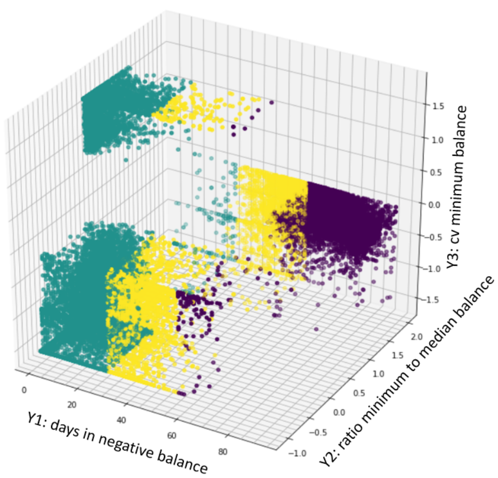
Measuring distress requires a relative evaluation of an individual's standard financial capacity. The same overdraft amount might represent a minor inconvenience for one person but a significant challenge for another. Additionally, it is crucial to consider fluctuations in financial circumstances from month to month. We utilise the minimum monthly balance as a proxy for the most challenging situation within the month. The coefficient of variation (CV) enables us to capture changes in this minimum balance over three months. A low CV value suggests that the account has consistently maintained a precarious state throughout the quarter. Contrarily, a high CV value indicates significant fluctuations in the minimum balance, reflecting either substantial increases or sharp declines over the quarter.

We integrate three elements of negative balances to measure distress: (i) the duration of negative balances ($y1$), (ii) the ratio of the (average) minimum balance to the (average) median balance ($y2$), and (iii) fluctuations in the minimum balance ($y3$), measured using the coefficient of variation (CV). The CV is calculated by dividing the standard deviation of the minimum balance by its mean, enabling a standardised comparison of variations between accounts, irrespective of significant differences in minimum balance amounts. Frequent overdraft usage is strongly associated with inferior credit performance, evidenced by a lower average credit score, a higher likelihood of possessing a subprime credit score, lack of available credit card balance, delinquent debt, and decreased financial capacity to meet bill obligations (Consumer Financial Protection Bureau, 2023). Therefore, identifying overdraft usage within cash flow balances serves as an effective method to discern ongoing financial difficulties and gauge the associated risks of distress. $y1$ indicates the duration of overdraft usage, while $y2$ reflects the magnitude of the overdraft amount, akin to the definition of credit default, which considers both the duration (90 days past due) and the significance of the overdue amount (Ernst & Yong, 2018). However, $y1$ and $y2$ alone may not adequately signify a consistent pattern of negative balances, hence the inclusion of $y3$ to account for this enduring challenging situation.

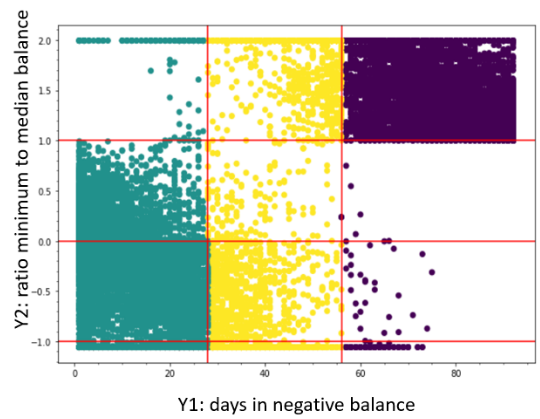
A two-stage threshold selection framework is proposed to divide each observed variable into categories depicting different levels of financial struggles. First, we input $y1$, $y2$ and $y3$ into the hierarchical clustering algorithm and interpret the dendogram to identify obvious groupings. We use the agglomerative hierarchical clustering technique with Ward linkage, implemented via the Python `scikit-learn` package. Prior to clustering, the observed variables are standardised to have a mean of zero and a standard deviation of one. Second, we visualise the groupings in a 3D space (Figure 3.2a) and determine cutoff points in the two-dimensional spaces(Figures 3.2b, 3.2c, 3.2d). During the threshold selection process, we exclude cases where $y1$ equals zero, denoting a positive balance. This exclusion is vital to prevent the creation of a sparse matrix and potential memory-related problems, as the majority of such instances fall into this category. Since $y1 = 0$ typically represents financially stable accounts, its omission does not undermine the identification of groups encountering financial difficulties. Following this exclusion, the threshold selection procedure proceeds with the remaining 8940 accounts. Figure 3.2a illustrates distinct thresholds for creating three separate groups.

Based on Figure 3.2b, the vertical lines at $y1 = 28$ and $y1 = 56$ are the thresholds to slice the data into three regions. In addition to the dropped $y1 = 0$, $y1$ eventually has four categories to classify the overdraft term length. For $y2$, we select the thresholds (horizontal lines) at $y2 = -1$, $y2 = 0$, and $y2 = 1$ because these cuts result in four categories with distinctive data distribution (as illustrated in Figure 3.2b) i.e. category 1 has even mixture of the three groups, category 3 has the yellow points being the minority, and categories 2 and 4 are dominated by the blue and purple points respectively. For $y3$, the thresholds are selected at $y3 = 0$ and $y3 = -0.80$. As observed from Figure 3.2c and 3.2d, the cut at $y3 = 0$ illustrates a clear cutoff to isolate the positive minimum balance fluctuations ($y3 > 0$) and $y3 = -0.8$ further splits the accounts into low ($y3 < -0.80$) and high ($-0.80 \leq y3 < 0$) volatile negative minimum amount where the purple points mainly gather in the high negative volatility category.

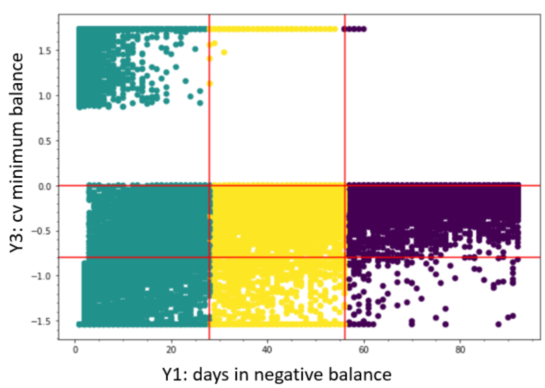
Apart from supporting the threshold choices, the proposed framework is also beneficial in dis-



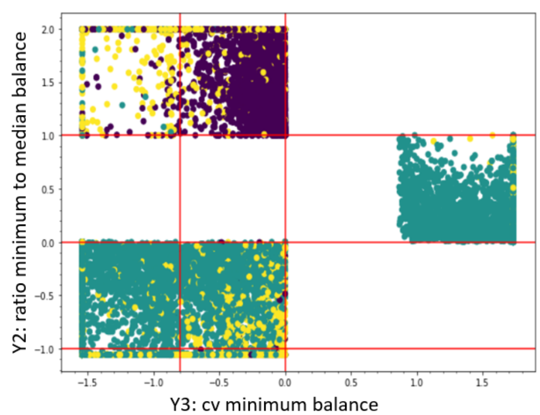
(a) 3D plot



(b) 2D plot: $y1$ vs $y2$



(c) 2D plot: $y1$ vs $y3$



(d) 2D plot: $y2$ vs $y3$

Figure 3.2: Groupings identified from hierarchical clustering examined from 3D and 2D plots

covering multiple underlying levels of distress (from the clusters) to hint the non-linear relationship between these three distress measurement variables. For instance, accounts could be in negative balance for many days yet the overdraft amount is immaterial with small amount, and vice versa. The accounts could also have many days in negative in the quarter but the monthly minimum balance fluctuates between positives and negatives. Thus, combining these variables to define a single distress measure is complex. The measure should account for the non-linearities when combining the three variables in quantifying the multiple levels of distress. A latent Markov model (detailed in Section 3.3.4) is employed to extract the underlying stages of distress from the three observed aspects of overdraft use.

3.3.3 Covariates

Consumers under vulnerable circumstances face great harm to fall into distress. We capture vulnerability from income and expenditure patterns as the covariates to predict movement away or towards potential distress. The computation of the covariates are inspired by the four key characteristics of vulnerability outlined by the Financial Conduct Authority (2021b): (i) low financial resilience describing the insufficient capacity situation to withstand financial shocks, (ii) negative life events relating to unforeseen occurrences of job loss, bereavement or relationship breakdown and etc., (iii) poor health due to illness that restricts the ability to conduct day-to-day tasks, and (iv) low capability representing lack of financial knowledge in money management. The covariates are lagged one quarter to project distress risk exposure in the next quarter. The full list of covariates are detailed in Table 3.1.

Each covariate is a numerical measure to summarise a specific vulnerability aspect. Financial resilience is covered by potential over-indebtedness approximated by negative balances situations (X1-X2) and financial capacity based on the available surplus to handle additional expenses (X3) and ability to spend with the given inflow (X4). Inflow fluctuations (X5) as well as changes in inflow (X6) and benefits amount (X7) attempt to capture income loss due to unemployment. Changes in healthcare cost (X8) imply expenses allocated to treat illness. Covariates X9-X16 serve as the proxy measures of financial literacy by evaluating whether the accounts have (i) the capability to afford debt payment with existing income (X9) such that credit acts as a medium to facilitate living rather than becoming a burden, (ii) proper money management skills to prioritise essential costs (X10), (iii) healthy spending habits under controlled budget with confidence to handle changes in

Table 3.1: List of covariates

Covariate	Vulnerability aspect	Abbreviation
negative opening balance	resilience	X1
negative median balance	resilience	X2
buffer to outflow ratio	resilience	X3
outflow to inflow ratio	resilience	X4
cv inflow	life event	X5
change in inflow	life event	X6
change in benefits	life event	X7
change in healthcare	poor health	X8
debt to inflow ratio	capability	X9
essential to discretionary ratio	capability	X10
change in fixed cost	capability	X11
change in living cost	capability	X12
change in discretionary cost	capability	X13
change in debt	capability	X14
change in transfer	capability	X15
change in gambling	capability	X16

key expenses (X11-X14), and (iv) harmful practices with overspending on miscellaneous cost (X15) or addictive gambling behaviour (X16).

3.3.4 Modelling the evolution of distress risk

Risk modelling in this chapter aims to assess the likelihood of transition between financial distress risk states i.e., improvement or deterioration, for consumer bank accounts. The Markov model provides a suitable framework to model the changes in distress risk due to the ability to account for the time-varying nature of financial behaviours. Cash flow patterns change over time to accommodate changing financial circumstances. Vulnerable characteristics developed in the current state are likely to impact financial transitions in the next time point. Thus, this aligns with the Markov assumption that the future distress exposure risk depends on the present state.

We apply the latent Markov model to estimate and predict the transition probabilities between the different states of distress risk. The evolution of the varying stages of distress risk is a non-observable latent process because the defined measures of distress (in Section 3.3.2) provide an implied financial difficulty from the observed balance situations. For each account, we examine how the past vulnerable transactional patterns drive the movement into a new distress level to predict a sequence of dynamic distress risks.

The Markov model represents the stochastic process of a system to move sequentially from one state to another following the Markov property i.e., the distribution of the future state is only dependent on the current state (Gagniuc, 2017). A latent variable model, in its generic form, is tailored to investigate the changes of an unobservable characteristic of interest which assume the system has a latent process following a Markov chain (Wiggins, 1973). We focus on an extended multivariate latent Markov (LM) which includes covariates to affect the latent distribution (Bartolucci et al., 2017). LM model has the strength to examine how the covariates affect the unobservable characteristic that is measured by the response variable. This is essential to distinguish the account-specific characteristics for initial and transition probabilities. The following subsections outline the multivariate LM model with the extension to include covariates, the parameters estimation and the sequence prediction procedures.

3.3.4.1 Multivariate latent Markov model with covariates

For a one-year transactional data, we set up a quarterly panel data with $T = 4$ time points for $t = 1, ..., T$. Let $\mathbf{Y}^{(t)}$ denotes the $J = 3$ observed response variables (in Section 3.3.2) of the balance aspects for $j = 1, ..., J$ categorical variables where each variable $Y_j^{(t)}$ in $\mathbf{Y}^{(t)}$ has c_j categories with $c_j = 0, ..., c_j - 1$. Then, $\mathbf{Y}$ represents the full response vector with $J \times T$ elements which are the union of all $J = 3$ observed balance variables stacked over all $T = 4$ quarters.

The model assumes the existence of a latent process $\mathbf{U} = (U^{(1)}, ..., U^{(t)})$ i.e., distress risk, that follows a first-order Markov chain with k number of latent states in a state space of $\{1, ..., k\}$. The latent process $\mathbf{U}$ affects the distribution of the response variables under the local independence assumption, such that the response vectors $\mathbf{Y}$ is conditionally independent given the latent process $\mathbf{U}$. The elements $Y_j^{(t)}$ of $\mathbf{Y}^{(t)}$ are also conditionally independent given $U^{(t)}$ at each occasion $t = 1, ..., T$. The model formulation can be represented as the conditional response probabilities

$$\phi_{jy|u}{}^{(t)} = P(Y_j^{(t)} = y|U^{(t)} = u), \qquad j = 1, ..., J, t = 1, ..., T, u = 1, ..., k, y = 0, ..., c_j - 1, \tag{49}$$

which are stored in a $c_j \times k$ matrices Φ_j to describe the distribution of each response Y_j given the k latent states (Bartolucci et al., 2017). The conditional distribution of $\mathbf{Y}$ given $\mathbf{U}$ is expressed as

$$P(\mathbf{Y} = \mathbf{y}|\mathbf{U} = \mathbf{u}) = \prod_{t=1}^{T} \phi^{(t)}_{y^{(t)}|u^{(t)}}, \tag{50}$$

where due to the local independence assumption,

$$\phi_{y|u}^{(t)} = \prod_{j=1}^{J} \phi_{jy_j|u}^{(t)}, \tag{51}$$

$\mathbf{y}$ is the realisation of $\mathbf{Y}$ with the subvectors $\mathbf{y}^{(t)} = (y_1^{(t)}, ..., y_J^{(t)})$ and $\mathbf{u} = (u^{(1)}, ..., u^{(t)})$.

Let $x^{(t)}$ denotes the vector of covariates (vulnerable characteristics in Table 3.1) at the t-th time point and let $\mathbf{X}$ represents the full covariates stacked over all $T = 4$ quarters. To account for the effect of covariates on the latent distribution (Bartolucci et al., 2014), the parameters of the latent process are conditional on $\mathbf{X}$, where the initial probabilities are

$$\pi_{u|x} = P(U^{(1)} = u|x^{(1)}), \qquad u = 1, ..., k, \tag{52}$$

and the transition probabilities are

$$\pi_{v|ux}^{(t)} = P(U^{(t)} = v|U^{(t-1)} = u, x^{(t)}), \qquad t = 2, ..., T, u, v = 1, ..., k. \tag{53}$$

The distribution of $\mathbf{U}$ given $\mathbf{X}$ is thus denoted as

$$P(\mathbf{U} = \mathbf{u}|\mathbf{X} = \mathbf{x}) = \pi_{u^{(1)}|x^{(1)}} \prod_{t=2}^{T} \pi_{u^{(t)}|u^{(t-1)}x^{(t)}}, \tag{54}$$

such that $\mathbf{x}$ is the realisation of the full vectors of covariates $\mathbf{X}$. The manifest distribution of the response variables is the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$,

$$P(\mathbf{y}|\mathbf{x}) = \sum_{\mathbf{u}} \pi_{u^{(1)}|x^{(1)}} \pi_{u^{(2)}|u^{(1)}x^{(2)}} ... \pi_{u^{(T)}|u^{(T-1)}x^{(T)}} \times \phi_{y^{(1)}|u^{(1)}x^{(1)}} ... \phi_{y^{(T)}|u^{(T)}x^{(T)}}. \tag{55}$$

The Baum-Welch forward recursion technique (Baum et al., 1970) is employed for an efficient computation of $P(\mathbf{y}|\mathbf{x})$ that involves the summation across all possible k^T configurations of U.

The LM model follows a multinomial logit parameterisation (Bartolucci et al., 2012) to consider the effect of covariates on the initial (Equation 56) and transition (Equation 57) probabilities,

$$log \frac{P(U^{(1)} = u|x^{(1)})}{P(U^{(1)} = 1|x^{(1)})} = \beta_{0u} + (x^{(1)})' \beta_{1u}, \qquad u \geq 2, \tag{56}$$

$$log \frac{P(U^{(t)} = v|U^{(t-1)}, x^{(t)})}{P(U^{(t)} = u|U^{(t-1)} = u, x^{(t)})} = \gamma_{0uv} + (x^{(t)})' \gamma_{1uv}, \qquad t \geq 2, u \neq v, \tag{57}$$

where $\beta_u = (\beta_{0u}, \beta'_{1u})'$ and $\gamma_{uv} = (\gamma_{0uv}, \gamma'_{1uv})'$ are the parameters to be estimated, and to be stored in the matrices β and Γ respectively.

3.3.4.2 Parameters estimation

This subsection describes the parameters estimation procedure with the Expected-Maximisation (EM) algorithm (Dempster et al., 1977). The EM algorithm is an iterative optimisation method to estimate the maximum likelihood of parameters in statistical models involving unobserved latent variables. There are two phases in the EM algorithm. The Expectation (E) step estimates parameters of the latent variables conditional on the observed data and the Maximisation (M) step maximises the estimated log-likelihood obtained in the E step to update the estimated latent variables. The procedures of EM algorithm are described below.

- Initialisation: A set of initial values is considered for the parameters.
- E-step: The posterior probability of each latent variable conditional on the observed data and current parameter estimates is computed. The incomplete data values (latent variables) are estimated based on the current parameter estimates. The log-likelihood of the observed data based on the current parameter estimates and the estimated latent variables are computed.
- M-step: The expected complete data log-likelihood obtained from the E-step is maximised to update the model parameters.
- Convergence: If the changes in log-likehood and parameter values between iterations are below a predefined threshold, the algorithm is considered to have converged and the iterations will stop, otherwise repeat the E and M steps until convergence is reached.

Considering the increased complexity due to the inclusion of covariates, Bartolucci et al. (2015) has proposed a three-step EM algorithm to more efficiently estimate the parameters. The latent Markov model framework is this study is implemented with the `LMtest` package in R. The input data frame must be formatted as long panel data, with covariates and observed target variables for each account stacked into a single column across all time points.

The EM optimisation is based on a complete data log-likelihood i.e., when the latent states are known for each account at each time point. The complete data is expressed as the triples $(\mathbf{u}_i, \mathbf{x}_i, \mathbf{y}_i)$ for $i = 1, ..., n$ accounts. Note that $\mathbf{u}_i = (u_i^{(1)}, ..., u_i^{(T)})$ is the sequence of latent states for account i, $\mathbf{y}_i$ represents the subvectors $\mathbf{y}_i^{(t)}$ with the elements $y_{ij}^{(t)}$ for $j = 1, ..., J$ response variables for account i and $\mathbf{x}_i$ represents the subvectors of covariates $x_i^{(1)}, ..., x_i^{(T)}$ for account i. The complete data log-likelihood (Bartolucci et al., 2015) is expressed as

$$
\begin{aligned}
l(\theta) = \sum_{i=1}^{n}[\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{u=1}^{k}\sum_{y=0}^{c_j-1} a_{ijuy}^{(t)} log\phi_{jy|u}^{(t)}+ \\
\sum_{u=1}^{k} b_{iu}^{(1)} logP(U_i^{(1)} = u|x_i^{(1)})+ \\
\sum_{t=2}^{T}\sum_{v=1}^{k}\sum_{u=1}^{k} b_{iuv}^{(t)} logP(U_i^{(t)} = v|U_i^{(t-1)} = u, x_i^{(t)})]
\end{aligned}
\tag{58}
$$

where θ is a vector all parameters in the model i.e., the conditional response probabilities, initial and transition probabilities, $a_{ijuy}^{(t)}$ is an indicator variable for account i to belong to latent state u with response y at time t, b_{iu} is an indicator variable for account i being in latent state u and $b_{iuv}^{(t)}$ is an indicator variable for account i to transit from u at time $t-1$ to state v at time t.

The EM algorithm (Dempster et al., 1977) alternates between the E-step and M-step until convergence occurs. The E-step computes the posterior expected value of the indicators ($a_{ijuy}^{(t)}$, $b_{iu}^{(t)}$, $b_{iuv}^{(t)}$) with a forward-backward recursions algorithm. The M-step maximises the posterior expected values in the complete log likelihood (in Equation 58) where the indicator variables are replaced with the corresponding expected value obtained from the E-step. Considering the slow convergence of EM with the complex LM model formulation to include large number of covariates. The parameters estimation for LM model with covariates follows a three-step approach (Bartolucci et al., 2015), which is summarised as follows.

- Step 1: Obtain the estimates of the conditional response probabilities $\hat{\phi}_{jy|u}$ and the temporary estimates of the marginal probabilities of the latent states $\hat{P}(U_i^{(t)} = u)$ (based on fitting on a basic latent class model without covariates (Wiggins, 1973)).

- Step 2: Obtain the posterior expected values of $b_{iu}^{(t)}$ and $b_{iuv}^{(t)}$ based on the results from Step 1.

- Step 3: Estimate β and Γ by maximising the components of the complete data log-likelihood with the latent structure parameters i.e., fitting the two multinomial logits (Equation 56 and 57) based on the weighted log-likelihoods below (Bartolucci et al., 2015),

$$l(\beta) = \sum_{i}\sum_{u} \hat{b}_{iu}^{(1)} logP(U_i^{(1)} = u|\mathbf{x}_i^{(1)}), \tag{59}$$

$$log(\Gamma) = \sum_{i}\sum_{t\geq 2}\sum_{u}\sum_{v} \hat{b}_{iuv}^{(t)} logP(U_i^{(t)} = v|U_i^{(t-1)} = u, \mathbf{x}_i^{(t)}) \tag{60}$$

3.3.4.3 Sequence prediction

The prediction of the entire sequence of latent states consists of two stages. The first stage is local decoding to predict the latent state of each account at each time point t. This is done by maximising the estimated posterior probability of $U^{(t)}$ i.e., $P(U^{(t)} = u|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})$, obtained from the EM algorithm (Dempster et al., 1977). However, the local decoding does not account for the joint probability of the latent sequence and could produce inconsistent sequences (Bartolucci et al., 2012).

The global decoding in the second stage obtains the maximal a posteriori likely sequence of states by adapting the Viterbi algorithm (Viterbi, 1967). This algorithm performs a forward recursion to obtain the prediction at each time t for each account followed by a backward recursion to predict the path of the latent states with the steps below (Bartolucci et al., 2012; Bartolucci & Farcomeni, 2009).

1. For $i = 1, ..., n$, $u = 1, ..., k$, compute

$$\hat{P}_i^{(1)}(u_i^{(1)}, y_i^{(1)}|x_i^{(1)}) = \hat{\pi}^{(1)}_{u_i|x_i}\hat{\phi}_{y_i^{(1)}|u_i^{(1)}}. \tag{61}$$

2. For $i = 1, ..., n$, $t = 2, ..., T$, $v = 1, ..., k$, compute

$$\hat{P}_i^{(t)}(v_i^{(t)}, y_i^{(t)}|x_i^{(t)}) = \hat{\phi}_{y_i^{(t)}|v_i^{(t)}} max_{u=1,...,k}\left[\hat{P}^{(t-1)}(u_i^{(t-1)}, y_i^{(t-1)}|x_i^{(t-1)})\hat{\pi}^{(t)}_{v_i|u_i x_i}\right]. \tag{62}$$

3. For $i = 1, ..., n$, find the optimal latent state,

$$\hat{u}_i^{*(T)}(y_i^{(T)}|x_i^{(T)}) = argmax_{u=1,...,k}\hat{P}_i^{(T)}(u_i^{(T)}, y_i^{(T)}|x_i^{(T)}). \tag{63}$$

4. For $i = 1, ..., n$ and $t = T-1, ..., 1$, find the state sequence at each point t,

$$\hat{u}_i^{*(t)}(y_i^{(t)}|x_i^{(t)}) = argmax_{u=1,...,k}\hat{P}_i^{(t)}(u_i^{(t)}, y_i^{(t)}|x_i^{(t)})\hat{\pi}^{(t+1)}_{\hat{u}_i^* y_i|u_i}. \tag{64}$$

3.4 Results and discussions

Since the k number of latent states is not known a priori, we fit several models with $k = [2, 4]$ followed by the examination of AIC/BIC. We select $k = 3$ because this model corresponds to the least information loss (minimum AIC/BIC) with no convergence problem. $k = 4$ reports a relatively lower AIC/BIC but the model estimation process could not converge where the estimated parameters might be biased and not reliable for prediction purpose. $k = 5$ is not further attempted due to the

Table 3.2: Selecting number of latent states k

	$k = 2$	$k = 3$	$k = 4$
AIC	128849.28	100852.53	95024.40
BIC	129361.98	102076.89	97220.60
Convergence	Converged	Converged	Could not converge

convergence issue with $k = 4$ which implies additional parameters results in an over-complex model.

Hence, the latent Markov model discovers three underlying states of potential distress from the observed balanced scenarios. Based on the three latent states, the subsequent discussions will cover (i) the characteristics of the latent states, (ii) the distribution of the latent process to understand the movement between the states, (iii) the effect of covariates of the latent distribution, and (iv) the predicted sequence of states on new accounts.

3.4.1 States of financial distress risk

Table 3.3 reports the estimated emission probability matrix to reveal the chances of observing the balance scenarios conditioned on the latent states. These conditional response probabilities point out the key components in each latent state (in bold) to specify their respective properties. Based on the output in Table 3.3, the latent states can be defined as exhibiting varying degrees of i.e., low, medium, and high risk, to face financial distress.

Latent state 1 represents a medium-risk state. The accounts are most likely to be in a short-term negative balance for 1-27 days over a quarter (77.4%), have a positive buffer to handle a relatively low magnitude of negative balance (66.3%), and unexpectedly fall into negative minimum balances as denoted by the large negative fluctuations over the months in a quarter (61.4%). On the other hand, we also discover secondary characteristics in the medium-risk state that reflects a regular overdraft practice, potentially as a strategy to keep up with living. The accounts have relatively lower chances to have negative balances for a longer duration of 28-55 days (22.5%), have insufficient positive funds to cover a large negative situation (29.6%), and show consistent negative minimum balances with low negative fluctuations across the quarter (38.1%).

Latent state 2 depicts a low-risk state in a positive balance situation. The accounts are most

Table 3.3: Estimated emission probability

Observed target variables	Category	Latent states 1 :medium	 2 :low	 3 :high
$y1$: Days in negative balance				
0: $y1 = 0$	0	0	**0.9133**	0
1: $1 \leq y1 < 28$	1	**0.7741**	0.0841	0
2: $28 \leq y1 < 56$	2	0.2254	0.0023	0.2454
3: $y1 \geq 56$	3	0.0005	0.0004	**0.7546**
$y2$: Ratio of minimum to median balance				
0: $y2 < -1$	0	0.2957	0	0.0743
1: $-1 \leq y2 < 0$	1	**0.6625**	0	0.0033
2: $0 \leq y2 \leq 1$	2	0.0008	**1**	0.0208
3: $y2 > 1$	3	0.0410	0	**0.9017**
$y3$: Minimum balance CV				
0: $y3 < -0.80$	0	**0.6142**	0	0.0400
1: $-0.80 \leq y3 < 0$	1	0.3811	0	**0.9324**
2: $y3 > 0$	2	0.0047	**1**	0.0276

likely to be in positive balances throughout the whole quarter (91.3%), have adequate funds such that the median minimum balance has never been negative (100%), and the positive fluctuations indicate the level of changes in the amount does not bring the accounts to a lack of money position (100%). In contrast, latent state 3 portrays a high-risk state with persistent dependence on overdraft. The accounts have the highest probability to fall into negative balances for more than 56 days in a quarter (75.5%), be in a severe deficit position with simultaneous negative median and a large magnitude of negative minimum (90.2%), and constantly remain at the same level of negative with low negative fluctuations in the quarter (93.2%).

3.4.2 Dynamics of the distress risk states

Table 3.4 outlines the distribution of the distress risk states in two parts. First, the estimated initial probabilities show the proportion of accounts to be in each state at the first observation time point. Second, the estimated transition probabilities indicate how likely are the accounts to stay or shift to the other states. We can examine the dynamics of the distress risk state across the quarters from the reported latent process distribution (Table 3.4). Since the covariates are lagged one period to project the risk state in the next quarter, the observed time window covers from Q2 to Q4.

Table 3.4: Distribution of latent process

		Initial probabilities		
		Latent states		
		1:medium	2:low	3:high
Q2		0.1316	0.7227	0.1457

		Transition probability matrix		
		Latent states		
		1:medium	2:low	3:high
Q2 → Q3	medium	0.5370	0.3147	0.1429
	low	0.0918	0.8386	0.0696
	high	0.2405	0.3287	0.4308
Q3 → Q4	medium	0.5360	0.3151	0.1489
	low	0.0915	0.8376	0.0709
	high	0.2385	0.3283	0.4332

Maintaining a stable financial state at a positive balance level throughout the year is a standard account behaviour. Most of the accounts start at low risk (72.2%) and are unlikely to deteriorate where 84% of the accounts stay in low risk throughout Q2 to Q4, and merely less than 10% of them worsen into medium or high risks. On the contrary, a low proportion of accounts are exposed to medium (13.2%) and high (14.6%) risks at the beginning of the observation period. Medium-risk accounts have the most active movements with a probability of 31.5% to stay. These accounts tend to revert to low risk (54%) rather than step into high risk (15%). This implies that accounts facing transitory financial struggles utilise short-term overdrafts to facilitate an optimistic recovery pattern. Despite the high-risk situation, these accounts can still escape rather than being trapped in this state.

These findings reflect the flexibility of current accounts to move across all risk states. Thus, we can track the varying stages of financial difficulty to measure the exposure risk to distress, although the model cannot directly detect a severe non-reverting distress state.

3.4.3 Estimated regression results: insights into vulnerability driving potential distress

Apart from tracking the dynamic financial positions over the quarters, we also examine the estimated regression parameters affecting both the initial probability (β coefficients estimated from Eq (56), reported in Table 3.5) and transition probability distribution (Γ coefficients estimated from Eq (57), reported in Table 3.6 and 3.7) to explore the key vulnerable drivers of the latent process. The interpretation of the regression coefficients is such that the logit of the model outcome (latent state) relative to the referent group is expected to increase/decrease by a positive/negative unit change in the covariate when the others are held constant.

Table 3.5: Estimated parameters for initial probability distribution

Covariate	Abbreviation	Latent states 2:low	 3:high
intercept		-79.76**	33.00*
negative opening balance	X1	-2.420****	-0.3342*
negative median balance	X2	-1.818****	4.467****
buffer to outflow ratio	X3	178.9***	-0.3133
outflow to inflow ratio	X4	123.84***	53.74***
cv inflow	X5	1.423****	-0.3156
change in inflow	X6	1.525	-25.39
change in benefits	X7	1.505	-3.141
change in healthcare	X8	3.218	4.542
debt to inflow ratio	X9	35.56****	1.207****
essential to discretionary ratio	X10	-0.5615	-1.008
change in fixed cost	X11	-3.861	0.2308
change in living cost	X12	-5.727	-7.329
change in discretionary cost	X13	-5.873	-8.055
change in debt	X14	2.757	-4.164
change in transfer	X15	-20.08	-10.99
change in gambling	X16	1.353	-10.14

significant at α =0.10* 0.05** 0.01*** 0.001****

Table 3.5 analyses the effect of covariates on the initial distress risk states with the medium risk state as the referent group. The significant covariates highlight the different characteristics with respect to each distress state. The accounts are more likely to be at low risk when they are financially stable in the previous quarter with adequate funds to serve expenses (X3) and rarely get into negative (X1, X2). Low-risk state is also associated with high inflow fluctuations (X5) when the accounts receive extra income (X6, X7) to boost the available resources, as well as a wealthy status with the ability to service more debt (X9) and cater more expenses (X4).

The opposite signs in most of the estimated logits between the high and low-risk states indicate that they are driven by contrasting circumstances. Having insufficient funds in the past quarter is one of the drivers to have high distress risk in the future. These accounts usually start with a positive balance (X1) but often end up with a negative median balance (X2), indicating they lack money after paying off necessary expenses. Although the buffer-to-outflow ratio (X3) is not significant, the negative sign validates that high-risk accounts are short of money to cushion extra expenditures. Low earnings without additional inflows, denoted by the non-significant X5, X6, and X7, also increase the chances to be at high risk. Potential financial struggles with the reliance on higher debts at low income (X9) or unhealthy practices to have high outflow regardless of the low inflow situation (X4) also increase the probability to be at high risk.

In the subsequent time periods, the model predicts whether a transition occurs for each account. Tables 3.6 and 3.7 present the estimated parameters for two types of transitions i.e., escaping from (improve) or moving into (deteriorate) a relatively higher risk state in the future. Generally, the key properties driving financial improvements (Table 3.6) are being financially resilient with substantial funds to buffer unexpected expenses (X3) and rarely falling into negative balances (X5). In addition, spikes of additional income streams (X6, X7) provide a larger pool of money to enable movement toward a relatively lower risk state. Having lower transfer payments (X15), probably due to the cut down of miscellaneous costs or ceased borrowings from friends and families, also associates with improvements.

Each type of improvement is linked to a specific set of characteristics. The high-to-medium transition is more likely to occur when the accounts can better manage their finance by restricting spending (X8) to retain more funds and reducing debt (X9) dependence to escape from long-term negatives. The obligation to handle increasing fixed and living costs (non-significant X11, X12) could potentially explain the reason for the accounts to revert only to a medium risk instead of completely recovering to the positive low risk.

Table 3.6: Estimated parameters for transition probability distribution (improvement)

Covariate	Abbreviation	Latent states 1:medium → 2:low	3:high → 1:medium	3:high → 2:low
intercept		31.12**	-93.84****	33.90
negative opening balance	X1	-0.7902****	-1.237****	-0.6241****
negative median balance	X2	-1.438****	-1.684****	-2.577****
buffer to outflow ratio	X3	77.01****	162.2****	60.33****
outflow to inflow ratio	X4	11.13***	-16.59***	1.032***
cv inflow	X5	0.9402****	0.6947**	0.7376***
change in inflow	X6	44.97	114.62**	73.10
change in benefits	X7	4.003	8.765	1.992
change in healthcare	X8	6.592	5.516	7.841
debt to inflow ratio	X9	7.658****	-1.688****	4.741****
essential to discretionary ratio	X10	-1.610	-1.618	0.3850
change in fixed cost	X11	-2.563	13.21	6.021
change in living cost	X12	1.433	4.782	-0.4655
change in discretionary cost	X13	5.007	8.402	5.871
change in debt	X14	-2.226	-4.696	-3.578
change in transfer	X15	-154.1****	-91.93****	-153.3****
change in gambling	X16	-1.354	-13.64	-18.76

significant at α =0.10* 0.05** 0.01*** 0.001****

Getting back to low risk (either from high or medium risks) has a contrasting behaviour to the high-to-medium transition where the results report X4 and X9 to be significantly related in the opposite way. The positive X4 reflects that the accounts with higher capacity and confidence to spend without struggles are more likely to escape from the negative status. There are two possible interpretations for the positive X9. First, the probability to return to low risk increases when the accounts are capable to service more debt with their income. Second, a full escape from negatives i.e., a jump from high to low risk, is more likely to happen when the accounts take on more credit as a cashflow turnover strategy to maintain a positive balance. The non-significant negatives in X11 and X12 indicate that budgeting, probably to the extent of sacrificing certain fixed or living expenses, increases the likelihood to regain a healthy positive status.

Table 3.7: Estimated parameters for transition probability distribution (deterioration)

Covariate	Abbreviation	Latent states 1:medium → 3:high	2:low → 1:medium	2:low → 3:high
intercept		36.00	105.1****	21.45
negative opening balance	X1	0.0125	2.754****	-1.923
negative median balance	X2	2.666****	1.296	7.765****
buffer to outflow ratio	X3	17.79	-211.95****	-47.35
outflow to inflow ratio	X4	-4.220***	-41.16	-7.120
cv inflow	X5	1.277****	-0.6906****	0.1667
change in inflow	X6	-187.33***	-3.033	-3.168
change in benefits	X7	-4.210	1.190	-12.23
change in healthcare	X8	22.88***	-3.580	1.270
debt to inflow ratio	X9	4.223****	-14.42	-4.161
essential to discretionary ratio	X10	-7.439	1.732	-5.933
change in fixed cost	X11	0.2172	-3.630	-0.2707
change in living cost	X12	6.522	13.06**	2.609
change in discretionary cost	X13	6.529	-0.6918	-2.587
change in debt	X14	6.176	-2.102	2.7225
change in transfer	X15	63.67****	2.129	4.154
change in gambling	X16	-7.096	4.719	-0.120

significant at α =0.10* 0.05** 0.01*** 0.001****

In contrast to financial improvements, deterioration is generally driven by low resilience when the accounts lack money and frequently fall into negative (X1, X2). Among the three types of deterioration, the model could not capture other significant covariates in the low-to-high transition, potentially due to insufficient occurrences. For the remaining two, the results depict different transactional behaviour with respect to each transition, as reported in Table 3.7.

The accounts tend to move from low to medium risk when they have low inflow fluctuations (X5). This is because financially tight accounts without extra earnings spikes would fail to increase their available fund and hence the need to use overdraft. The significant positive X12 explains that higher living costs increase the chances of low-to-medium transition. From the non-significant negative signs, low-risk accounts are more likely to become medium-risk when they have to reduce their expenses in fixed (X11) and discretionary expenses (X13) to prioritise essential spending (X10), denoting a struggling behaviour to remain financially stable. Besides, further deterioration from medium to high risk is more likely to occur when the accounts face unforeseen hardships such as dramatic income loss (X5, X6) and increased healthcare costs due to abrupt life events (X8). The positive association between the medium-to-high transition and X9 also signals the necessity to rely on more credit facilities when the low income is insufficient to make ends meet.

To summarise, we highlight how the distress risk state is driven by different aspects of vulnerability. Financial resilience is a significant aspect affecting accounts' distress risk. Deficit balance not only leads to being at high risk at the initial observation point but also transitioning into a worse financial state. In contrast, accounts with surplus balances generally start with low risk and are able to alleviate potential distress circumstances (medium/high risk) to low risk. Life events have different effects on distress risks. Positive events (receiving extra salaries to elevate financial buffer) and negative incidents (job loss, illness, accidents, etc) posit the accounts in low and high risks respectively. While bad spending habits regardless of low inflow lead to high-risk state, interestingly, accounts portraying financial capability with proper money management do not necessarily bring them back to the low-risk state because budget planning by cutting down expenses signals a financially-constrained scenario.

3.4.4 Predictions on test set accounts

3.4.4.1 Monitoring risks from the predicted sequence of states

The predicted sequence of states across the quarters (Table 3.8) can be used as a risk management tool to monitor the evolvement of vulnerable circumstances and potential distress. Table 3.8 (part 1) presents the predicted number of test set accounts with respect to each distress risk state in each quarter. The slight increment in the low-risk state denotes the gradual movement of the accounts to a steady financial position. The high-risk accounts show a fluctuation trend which peaks in Q3 followed by a decrease in Q4. This implies that some accounts experience drastic short-term difficulty with quick recovery. Lastly, the consistent decreasing pattern in the medium-risk state indicates a high tendency for the accounts to leave such intermediary position, and step into either a non-struggling or struggling state.

Table 3.8: Predictions on test set accounts

Part 1: Predicted accounts in each state		medium	low	high
Q2		572	2764	554
Q3		558	2717	615
Q4		522	2781	587
Part 2: Predicted transition of accounts		medium	low	high
$Q2 \rightarrow Q3$	medium	331	149	*92*
	low	*171*	2555	*38*
	high	56	13	485
$Q3 \rightarrow Q4$	medium	307	189	*62*
	low	*140*	2558	*19*
	high	75	34	506

The second part in Table 3.8 summarises the between-states movements of the accounts at Q2-Q3 and Q3-Q4. Each cell in the transition matrix denotes a shift from the previous state (row) to the current state (column). The diagonal elements represent accounts that have stayed in the same state. The upper and lower triangular parts are the accounts that have moved to a different state in the next quarter which represent financial improvements or deterioration (*italic*). These results provide two streams of strategies to track the tenacity of vulnerability and examine financial stress.

First, accounts that consecutively remain at the same risk state in all quarters can be directly segmented into the least, mild, and severe stress levels. These static situations refer to long-term circumstances where the accounts are constantly (i) in minimal exposure to vulnerability (low risk), (ii) financially restricted with a tight buffer (medium risk) or (iii) trapped in critical money deficits (high risk), throughout the whole duration. Second, the state transitions provide additional complexity to assess financial stress. Since transitions can occur at any time point, this will result in various possible sequences with different interpretations. These transitions reflect short-term financial adjustments when the accounts recover from or enter transient vulnerabilities. In the coming paragraphs, we detail specifically how monitoring certain improvement/deterioration, alongside the vulnerability drivers (in Section 3.4.3), can be a strategy to mitigate distress.

Improvements back into low risk (from high/medium risks) mainly indicate the ability to step into a stable status. However, it is important to follow up on the upcoming states after such a transition to confirm a full departure from the temporary vulnerable scenarios. For instance, further inspection is required for the high-to-low transition to ensure the acquired money (from extra income, debt, or saved up from cost-cutting) supports a persistent budget for a long-lasting stay in the low-risk state. Tracing the drivers behind the improvements could induce new ideas for short-term financial aids i.e., flexible credit schemes or interim incentives.

Deterioration specifies the downgrade of financial standards due to the occurrence of disruptive vulnerabilities. The transition towards high risk is the key point of risk which requires further invigilation in the subsequent period. If the accounts manage to recover, either immediately to low risk or gradually pathing ways through medium and then low risk, this sequence refers to a resilient behaviour to overcome transitory shocks. Yet, if the accounts remain at high risk, they experience destructive permanent vulnerabilities that may eventually lead to real distress. Being able to detect the transient or permanent vulnerable characteristics can support better remedial actions to alleviate potential distress. Besides, since the medium risk is an intermediary state which could exacerbate to high risk, the low-to-medium transition shall raise alert to avoid the transient vulnerability becoming a permanent one. Thus, understanding the specific vulnerable characteristics driving this transition can guide the setup of pre-emptive strategies to direct the temporary medium risk back to a stable position.

3.4.4.2 Predictive performance

The model predicts unobserved latent states rather than the directly observed target variables. Consequently, the rows denote the labels of the latent states, which are determined by mapping the observed responses $(y1, y2, y3)$ in the test set to these latent states, based on the estimated emission probabilities listed in Table 3.3. The columns indicate the predicted latent states as determined by the Viterbi algorithm (in Section 3.3.4.3).

Table 3.9 reports the test set account predictions at each quarter. The accuracy measures (**bold**) are the proportion of accounts correctly predicted. The rows represent the labels of the latent states whereas the columns are the predictions. Since the model aims to predict the unobserved latent states rather than the observed target variables, the risk state labels in the test set are obtained based on mappings between the observed and the latent states (estimated emission probability in Table 3.3). Then, the sequence of the latent states for each test set account is predicted following the steps in Section 3.3.4.3 to obtain the state sequences over the quarter.

Table 3.9: Correct predictions in each quarter

Time point	Confusion matrix				Accuracy
		medium	low	high	
Q2	medium	479	0	0	
	low	4	2764	6	
	high	89	0	548	**0.9746**
Q3	medium	481	0	3	
	low	3	2717	9	
	high	74	0	603	**0.9779**
Q4	medium	470	0	1	
	low	5	2781	6	
	high	47	0	580	**0.9851**

The results show a high accuracy at around 97%. From the confusion matrices, the model has the weakness to mistakenly predict high risk as medium risk. The caveat for the high accuracy is that the computation of both the latent state target variables and the predictions involve the emission probability matrix and thus the predictions are highly matched.

3.5 Conclusion

We initiate a statistical framework with the latent Markov model for an objective assessment of personal financial condition. We predict an account's transition probabilities to a less or more severe stage of distress risks in the next quarter based on the vulnerable characteristics in the current quarter from cash flow transactions.

Three implicit states of distress are discovered from the observed negative account balance information. The financial states represent the scale of financial difficulty i.e., low, medium, and high exposure risk to distress. The low-risk state shows a stable position with a sufficient positive balance and almost no overdraft use. The medium-risk state indicates either a minimal or mild level of financial difficulty - the accounts rarely fall into a low negative balance but generally have a positive balance (minimal) or frequently take up low overdraft amounts as a strategy to keep up with living (mild). The high-risk state depicts a consistent overdraft reliance with low fluctuations of high negative amounts in the quarterly observation window. This provides a data-driven approach to objectively measure distress.

The relationship between vulnerability and financial transitions distinguishes the different characteristics between improvement vs deterioration. Movement from high to medium risk is associated with financial capability such as practising money management to restrict spending and debt dependence. Full recovery to low risk requires financial resiliency to have sufficient income to service more debt or be accessible to alternative credits for cashflow turnover. On the other hand, deterioration from low to medium risk is driven by low financial resilience when a financially constrained account has no extra income to boost the low buffer. Unexpected hardships such as sudden unemployment or commitment to illness treatments are the main vulnerability factors for the transition towards high risk. Understanding the vulnerability drivers to increase or decrease the likelihood of transitions provides valuable insight to effectively investigate the specific problem harming financial health.

The evolution of financial states alongside the vulnerability drivers reveals the tenacity of vulnerability to evaluate financial stress. Static financial states over the quarters depict long-term situations to be stable with minimal vulnerable harms (low risk), under financial pressure with limited buffer (medium risk), or critical money deficits (high risk). Temporary vulnerability relates to improvement when an account completely departs from high or medium risks to remain at low risk afterward whereas permanent vulnerability refers to a non-reverting deterioration to high risk. The worsening from low to medium risk is not only the initial sign of temporary vulnerability to slip into financial struggles but also the starting point to developing into permanent vulnerability. The joint analysis of the predicted sequence of states and the vulnerability drivers enables practitioners to set remedial or pre-emptive actions according to the specific short or long-term vulnerable characteristics to alleviate distress risk.

This research can be extended in several ways. Since out-of-sample prediction of financial transition on new accounts is the focus here, future work shall consider out-of-time prediction if longer time span data is available. The developed framework can also be applied to track long-term panel data e.g. five-year-long instead of only one-year transactional record, for a more accurate estimation of vulnerability-driven distress risk evolution. The account-based empirical analysis may lose the holistic financial position of an individual who could own multiple accounts. Thus, the assessment could be run on consolidated account information by the same individual. The proposed framework can also be integrated into the credit risk model to support stress testing with time-varying vulnerable characteristics or distress risks.

The Markov-based model discussed in this chapter is a "memoryless" model, meaning that state transitions are not influenced by the past or the duration an account has been in a distress state. However, the length of time an account remains in a particular state does affect the likelihood of future outcomes, such as improvement or deterioration. Future research shall explore alternative frameworks to better model scenarios involving recurring events. Furthermore, the class imbalance problem remains unaddressed in this chapter. Given that the majority of accounts fall into the low distress risk category, there may be an underestimation of the high distress risk states due to their lower occurrence frequency. Therefore, future research efforts should prioritise addressing the class imbalance problem within the latent states to ensure more accurate risk assessment and analysis.

4 Detecting suspicious bank accounts in absence of labels

Both volatile cash flow patterns (discussed in Chapter 2) and dynamics in financial vulnerability (explored in Chapter 3) indicate distinct behavioural characteristics among accounts. Such accounts may represent outliers exhibiting unusual patterns. For instance, accounts displaying exceptionally low or high volatility risk could signal significant financial challenges, while abrupt shifts in financial circumstances may hint at unforeseen events. The primary focus of this chapter is the identification of outlier accounts, aiming not only to deepen our understanding of account behaviours but also to detect potential suspicious activities stemming from anomalous patterns.

4.1 Introduction

Know Your Customer (KYC) is fundamental for financial crime prevention and control, vital for evaluating the risk associated with potential clients at the start of a new business connection (Financial Conduct Authority, 2023a). Developing KYC profiles involves Customer Due Diligence (CDD) measures, including two crucial components: (i) authenticating client identities, encompassing details like their name, location, sources of income, and more, and (ii) examining past suspicious activities from various sources. These KYC profiles are instrumental in determining the need for enhanced due diligence for new clients. However, obtaining comprehensive information during the initiation of a business relationship often provides a limited perspective of clients' behaviour, without a complete view of the nature and intentions of new clients.

The enactment of the Payment Service Directive 2 has established the legal framework for Open Banking (OB) in the UK, enabling authorised third-party financial services to access verified sources of bank account transactions from various financial institutions, contingent upon consumers' consent (Doyle et al., 2023). This milestone presents an opportunity to exploit extensive KYC risk profiles from OB cash flow transactions to derive client behaviour from their day-to-day activities. These data-driven KYC profiles offer a twofold advantage. First, they provide objective information to validate subjective judgements in flagging suspicious activities to streamline the KYC onboarding process. Second, they allow firms to tap into new client segments, those with limited financial records in other databases, with the enriched explored individuals' financial situations.

The identification of suspicious bank accounts through OB data is a new challenge in the KYC literature. Current accounts, in particular, present unique complexities due to their versatile nature in managing everyday finances, rather than being solely designated for purposes like mortgage payments or savings. Consequently, current accounts are exposed to various contexts of financial misconduct, necessitating a wide range of anomaly detection measures to identify suspicious activities. However, existing KYC tools possess distinct characteristics compared to current accounts, where they have been restricted to specific crime contexts such as money laundering, fraud in credit card purchases, or insurance claim (Ngai et al., 2011; Hilal et al., 2022). Developing an OB-specific KYC assessment for current accounts remains an area that has not been thoroughly explored.

The existing tools for detecting financial crime in the literature exhibit several limitations. Firstly, the Recency, Frequency, and Monetary (RFM) representation of transactional data effectively extracts spending patterns from card purchasing records, but it fails to fully quantify the complexities of cash flow transactions (Whitrow et al., 2009; Baesens et al., 2021). Current accounts reflect daily financial behaviour, characterised by recurring transactions e.g., salary deposits and fixed merchant payments, which is a unique property not found in money transfers or card purchases. Since the RFM framework overlooks the intervals between transactions, the regularity of repeated patterns is not taken into account. The persistence measure can effectively capture regular patterns and spot anomalies resulting from subtle persistence and sporadic activities (Belth et al., 2020), yet it has not been integrated to represent cash flow attributes for KYC screening. Secondly, prior works have predominantly relied on binary predictions with outlier detection techniques to identify suspicious activity without investigating the segments within the non-suspicious group (Luna et al., 2018; Stripling et al., 2018; Vandervorst et al., 2022; Carcillo et al., 2021; Nian et al., 2016). Due to the absence of expert-labeled data in the OB domain, grouping typical behaviours helps define ‘normality’ to guide the detection of anomalies.

Thirdly, the common way in the current KYC model to identify suspicious patterns is by analysing the predictions with subjective expert judgments (Ngai et al., 2011; Hilal et al., 2022; Bhattacharyya et al., 2011; Chen et al., 2015; Segovia-Vargas et al., 2022; Carneiro et al., 2017). The data-driven approach to extract unique anomalous characteristics to justify flagging abnormalities with model explainability is rarely studied. The goal in this chapter is to detect potential suspicious accounts via an unsupervised manner by addressing the following questions: (i) What distinct profiles of bank account behavior emerge? (ii) To what extent does each account deviate from the others? (iii) Do any outlier accounts present high anomaly risks? If so, what specific anomalous characteristics are associated with suspicious activities?

This chapter presents a novel three-phase KYC support system to identify potential suspicious accounts within the OB domain. First, we introduce an RFMP (Recency, Frequency, Monetary, Persistence) framework aimed at representing cash flow attributes derived from current account transactions. The extension of the persistence dimension seeks to address the neglected regularity and highlight subtle or bursty activities, which are not accounted for within the RFM framework. Second, we conduct cluster analysis to explore RFMP profiles, enabling the identification of multiple potential groups corresponding to normal account behaviour. Additionally, word analysis on transaction descriptions further describes the financial activities within each cluster. This segmentation approach focuses on uncovering various patterns of normal account behaviour, rather than using the binary approach that solely distinguishes typical from atypical behaviour. This method enriches KYC risk profiling and reduces the risk of misclassifying minority behaviours as dubious. Third, we compute outlier scores to differentiate outliers from non-outliers, followed by extracting the characteristics of potential anomalies from the RFMP attributes through an explainable outlier detection model. The data-driven explanations provide justifications for flagging suspicious accounts.

This proposed KYC strategy leverages OB data to improve customer onboarding processes. This RFMP data representation offers a customised framework to represent cash flow attributes specific to current accounts. By introducing the persistence measure to quantify regular recurring transactions, practitioners can detect subtle anomalies that might be overlooked when relying only on RFM. Based on the multiple KYC risk profiles, financial institutions can better understand the distinctive, normal financial behaviours, to provide appropriate product recommendations tailored to clients' behaviour. This KYC framework also enhances the efficiency to flag suspicious accounts for further due diligence through automated screening of current accounts into low and high anomaly risk groups, supported by explanations of peculiar activities. Overall, the proposed KYC mechanism fosters financial inclusion by extending the KYC assessment opportunity to clients with limited financial records, granting them equal access to financial products in the market.

The chapter is organised as follows. Section 4.2 review past related studies. Section 4.3 presents the methodology by describing the data representation framework, outlining the clustering setup to build risk profiles and detailing on the assessment process to rank the accounts' anomaly risk to flag suspicious accounts. Section 4.4 discusses the reported results and lastly Section 4.5 concludes.

4.2 Literature review

4.2.1 KYC-related studies

Know Your Customer (KYC) serves as the central process within financial crime prevention systems, dedicated to combating an array of illicit activities, encompassing money laundering, fraud (such as mortgage, insurance, and credit card fraud), bribery, and investment-related offences Financial Conduct Authority (2023a). Establishing a robust system and controls empowers firms to identify, prevent, and deter financial crime effectively. A pivotal component of the KYC framework is the Customer Due Diligence (CDD) check, enabling firms to gain a comprehensive understanding of the risks associated with individual relationships and to implement an appropriate level of due diligence to mitigate identified risks. Firms are required to gather pertinent information about the customer, and when applicable, beneficial owners, in order to ascertain the purpose and intended nature of the customer's relationship with the firm. This ensures the acquisition of sufficient data for a holistic assessment of the risk associated with the business relationship, thereby providing a solid foundation for subsequent monitoring Financial Conduct Authority (2023a).

Identity verification is the standard practice of Customer Due Diligence (CDD), ensuring the authenticity of new clients and safeguarding against malicious intentions Financial Conduct Authority (2023a). In an effort to streamline the data collection process from diverse, trusted sources and mitigate delays, the integration of blockchain technology has emerged as a compelling and secure solution within the KYC landscape (Yadav & Chandak, 2019; Al Mamun et al., 2020; Kapsoulis et al., 2020). Yadav & Chandak (2019) propose an algorithm aimed at conducting KYC verification only once for each customer, regardless of the number of institutions the customer seeks to engage with. Leveraging distributed ledger technology, their innovative solution operates as a smart contract, enabling the secure sharing of verification results based on predefined conditions encoded as a computer program. Al Mamun et al. (2020) develop a document-sharing platform using blockchain networks to simplify the uploading and sharing of KYC documents among financial entities. Kapsoulis et al. (2020) introduce a privacy-centric decentralized architecture for KYC process implementation. This architecture employs a permissioned blockchain network to automate the protection of sensitive user data. While these studies have significantly enhanced the secure sharing of user data through blockchain technology, identity verification remains a cumbersome task within KYC due to the necessity of sourcing data from multiple parties.

Despite its vital role in combating financial crime, the rigorous demands of KYC can introduce additional friction, potentially causing customer dissatisfaction (Gill & Taylor, 2004). Many clients have voiced frustration over the lack of flexibility when firms require additional documents to comply with regulatory mandates for identity validation. Consequently, financial practitioners are increasingly exploring customer profiling as a risk-based strategy to complement identity verification within the KYC process (Gill & Taylor, 2004). Risk profiles offer insights into individuals' financial behaviour, enabling firms to allocate resources more effectively by closely monitoring high-risk areas and targeting vulnerabilities while minimizing costs elsewhere. This risk-based approach presents an alternative method to streamline the onboarding process, as nuanced profile analysis can supplement the information gathered during identity verification and provide deeper insights into customer needs even at the initial stage of onboarding.

The risk-based approach, aimed at identifying both typical and atypical profiles, has proven to be an effective monitoring system to facilitate the reporting of suspicious activities in KYC protocols. The definition of 'suspicious' varies depending on the specific context of financial crime. For example, in the case of money laundering, suspicious activities revolve around patterns of money transfers designed to obfuscate the origins of illegally acquired funds, making them appear legal and legitimate (Savage et al., 2016). This illicit money movement typically involves three phases: firstly, the placement of funds into financial institutions through complex deposits across various accounts; secondly, the layering of criminal proceeds to conceal their true source; and finally, the integration of the money back into legitimate channels to disguise the illicit transactions.

Regarding credit card fraud, 'suspicious' behaviours may include activities such as the exploitation of stolen or lost cards for unauthorized purchases, or instances of bankruptcy fraud where individuals deliberately use a credit card with no intention of repaying the balance, subsequently filing for personal bankruptcy (Hilal et al., 2022). In the realm of investment fraud, schemes often entail the illegal sale or misrepresentation of financial instruments. These schemes typically involve purported low-risk investments with guaranteed or excessively high returns (Federal Bureau of Investigation, 2011). Additionally, market manipulation in trading activities, involving illicit buying and selling of securities after regular market hours, is another form of investment fraud (Federal Bureau of Investigation, 2011).

Nevertheless, identity verification remains the core component to onboard customers. The risk-based approach to distinguish normal and abnormal profiles have been utilised only for transaction monitoring but not utilised to support KYC customer onboarding. This is primarily due to the fact that transactional data becomes accessible only after a business relationship has been established.

4.2.2 Transactional data

The regulatory framework governing OB data sharing has shifted control of data back into the hands of customers, thereby enhancing their options and convenience. Furthermore, OB serves as a valuable resource for bolstering the detection of financial crimes, thereby fostering greater confidence in financial services (Podder, 2022). Podder (2022) delineate three key advantages of leveraging OB in the fight against financial crimes. Firstly, it enables the creation of a comprehensive personal financial profile, facilitating the identification of anomalies through the analysis of aggregated transaction histories across multiple institutions. Secondly, OB streamlines KYC data sharing for customer onboarding, reducing redundancy by allowing OB participants to rely on CDD conducted by third parties when applying for new products across different institutions, contingent upon regulatory provisions facilitating such data sharing. Thirdly, the convenience of utilizing KYC checks conducted by customers' existing providers can lead to reduced transaction costs, saving both time and effort in re-submitting information.

As the implementation of OB is still in its early stages, integrating OB into KYC procedures remains a mere conjecture at the time being (Podder, 2022). The common transactional data used in previous KYC research has involved credit card purchases (Bhattacharyya et al., 2011; Van Vlasselaer et al., 2015; Carneiro et al., 2017; Van Belle et al., 2023), business-to-business transactions (Paula et al., 2016; Colladon & Remondi, 2017), and interbank transfers (Drężewski et al., 2012, 2015; Demetis, 2018; Segovia-Vargas et al., 2022). However, OB bank transactions serve distinct purposes compared to the transactional data examined in traditional KYC literature. OB data primarily comprises connections to current accounts, reflecting both income and expenditure behaviour. These transactional patterns present new challenges for KYC assessment. Unlike specific contexts such as card purchases or business-related transfers, the multipurpose nature of current accounts, utilised for receiving income, meeting financial obligations, and managing day-to-day expenses, poses unique challenges. Identifying KYC risk profiles within OB data necessitates a thorough examination of diverse personal financial behaviours to detect irregularities. In this chapter, we initiate utilising OB data for CDD during onboarding by analysing earning and spending patterns to assess exposure risk to financial crimes.

The Recency, Frequency, Monetary (RFM) framework has emerged as a practical strategy for characterising transactional patterns. Widely utilized in marketing decision-making, the RFM framework has proven to be an effective tool for understanding customer behaviour (Chang & Tsai, 2011; Alboukaey et al., 2020; Olson & Chae, 2012). Chang & Tsai (2011) calculate RFM values to assess customers' purchase potential based on their specific buying patterns, including factors such as unit price and lifetime. These RFM values are closely tied to the products purchased, offering insights into customers' actual consumption behaviour. By clustering purchase records according to RFM attributes, they devise various sales policies to develop personalized management systems and enhance inventory control. Alboukaey et al. (2020) generate RFM features from time series data of mobile consumption, such as call volumes, SMS usage, data usage, and top-ups, to predict churn on a daily basis. Their model, leveraging RFM features, demonstrates superior predictive accuracy compared to statistical-based attributes, showcasing the effectiveness of RFM in capturing mobile usage behaviour. Olson & Chae (2012) compare RFM-based predictive models with classical data mining classification techniques in modelling customer response across two scenarios: individual purchases in catalogue sales and donor contributions. They refine RFM methods by adjusting the limits of RFM values to balance the density of consumer behaviour groups and compressing variables into a value function. While RFM methods lean toward being prescriptive rather than predictive, they offer a simpler means of providing well-organized descriptions of individuals based on their past behaviours, aiding marketers in effectively identifying valuable clients.

In addition to its role as a useful marketing tool, RFM variables find widespread application in financial crime detection. These attributes have been effective in extracting suspicious money laundering patterns from fund transfer transactions (Le Khac & Kechadi, 2010; Larik & Haider, 2011; Cao & Do, 2012). Le Khac & Kechadi (2010) aggregate the frequency and amount of transactions across six investment funds on a daily, weekly, and monthly basis to analyse the behaviours of corporate customers in subscribing to and redeeming these funds. Initially, these attributes serve as inputs for a clustering algorithm to identify normal investment patterns. Subsequently, a genetic algorithm is employed to enhance the detection of suspicious instances, followed by scoring new patterns using a neural network to flag potential cases of concern. Larik & Haider (2011) compute the average monthly frequency and amount for both credit and debit transactions to capture money transfer behaviours. They establish customers' normal groups of bank transfer behaviour by clustering the frequency and monetary attributes, subsequently computing an anomaly index to gauge the deviation of new transactions from the corresponding group behaviour, thereby ranking anomalous transactions. To implement a money laundering detection system, Cao & Do (2012) characterise money transfer patterns by calculating the sum and number of money sending or receiving relationships. These attributes are then subjected to a clustering algorithm to identify unusual patterns. The proposed system necessitates a combined manual and data-driven approach wherein analysts have to provide a criteria set to validate the output clusters.

RFM serves as an effective framework for detecting credit card fraud as well. Weston et al. (2008) compute the amount and number of transactions as input attributes for unsupervised peer group analysis. They discovered that tracking three-month representations of frequency (F) and monetary (M) transactions adequately projects behaviour one month ahead. Transactions significantly deviating from their peer group are flagged as potentially fraudulent. Bhattacharyya et al. (2011) construct F and M attributes for credit card transactions over one-month and three-month time windows, employing them as predictors in classification models to forecast fraudulent transactions. Their reported high predictive performance indicates that these derived attributes effectively represent recent buying behaviour with sufficient ability to model fraud. Baesens et al. (2021) propose an RFM-based data engineering process to capture card payment patterns. They outline various criteria for aggregating RFM features, including payment channel, authentication method, beneficiary country, and communication type. They advocate for the use of RFM features due to their straightforward interpretations and significant improvements in predictive performance and cost savings. Thompson et al. (2021) develop KYC profiles to analyse retail investors' trading behaviour based on engineered RFM features extracted from trading transactions. They formulate categorical RFM attributes to quantify trading volume and patterns, grouping clients using the k-prototype clustering algorithm. The resulting profiles reveal distinct personas based on trade sizes and investment strategies.

While RFM generally has the capability to quantify underlying behaviour from transactional data, it may fall short in fully describing all crucial aspects of data in certain domains. For instance, in quantifying bank card transactions, Huang et al. (2018) point out a limitation where two applicants with different consumption behaviours may have identical RFM scores. To address this, they propose extending RFM with a standard deviation (S) of the amount spent. By aggregating RFMS features over a year, they summarize consumer card spending behaviour, with the inclusion of the S dimension capturing spending fluctuations and improving credit default prediction. In the online retail business, Zhang et al. (2015) highlight a deficiency of RFM leading to significant increases in micro-level prediction errors due to mis-ranking of customer value. Data compression of online retail visits and purchases with RFM fails to account for spacing between these transactions. To address this, they propose a clumpiness (C) metric to measure the degree of nonconformity to equal spacing, augmenting the RFM framework. The augmented C metric captures groups of customers with clumpy purchasing patterns, indicating both high potential and high risk, as these patterns may suggest dormant activities that could potentially resume. Hence, the extended C dimension assists in better management and targeted promotions to increase company profits, capturing insights that might otherwise be overlooked. Baesens et al. (2021) suggest measuring the distribution of timestamps in transactional data, a component left out in traditional RFM analysis. They estimate the periodic mean and standard deviation for transaction times based on a periodic normal distribution. By investigating whether transaction times fall within the confidence interval, transactions can be flagged as regular or irregular with respect to customer transaction history. Thus, in domains with additional key characteristics to represent underlying behaviour in transactional data, an extended framework beyond the fundamental RFM components should be considered.

In the OB context, the presence of recurring transactions represents a distinctive characteristic within cash flow activities. Day-to-day financial activities encompass regular income and expenditure patterns arising from sources such as salary deposits, subscriptions, debt repayments, and habitual spending. However, this aspect of regularity is often overlooked in the traditional RFM framework, as none of its three components account for the distribution of intervals between transactions. To address this gap and capture periodic transactional patterns more accurately, it is essential to quantify and incorporate regularity as an additional feature within the RFM framework.

The persistence measure, as developed by (Belth et al., 2020), serves as a potential metric for quantifying regularity from timestamped transactions, particularly within an evolving network. Measuring persistence in cash flow transactions is paramount for KYC assessment as it helps uncover two distinct types of anomalies—subtle persistence and bursty patterns—that might otherwise be overlooked when solely relying on RFM (Belth et al., 2020). Both types of anomalies enhance the ability to detect repeated transactions representing normal behaviour, thus ensuring a more accurate depiction of KYC risk profiles and preventing instances from being erroneously flagged as abnormal behaviour.

Subtle persistence denotes rare events with low frequencies that may be disregarded or misinterpreted as anomalies when relying solely on RFM. However, these instances of low persistence could also represent activities such as quarterly subscriptions or credit payment schemes, which are challenging to capture due to their infrequency. Subtle persistence has been recognised as a critical indicator of slow-moving threats in the cybersecurity domain. Friedberg et al. (2015) track system events to detect advanced persistent threats, deliberate, slow-moving cyberattacks designed to silently compromise interconnected information systems without detection. Dai et al. (2016) identify harmful trends in data streams by identifying items that persist and occur more frequently compared to others over an extended period. Giroire et al. (2009) leverage temporal persistence to detect suspicious communications that are lightweight and sporadic over a prolonged period, making them stealthy and difficult to uncover. Xiao et al. (2014) uncover malicious activities characterised by suppressed traffic volumes, where the attack avoids overwhelming external requests but exhibits a coordinated pattern to evade detection. These deliberate and harmful persistent patterns serve as cyber-attack strategies aimed at mimicking normal behaviour, thereby underscoring the importance of identifying potentially suspicious behaviour from recurring cash flow transactions.

On the other hand, the second type of anomaly identified by the persistence measure, known as bursty patterns, highlights abrupt spikes in high-volume transactions occurring within a short timeframe. This anomaly type also effectively facilitates the detection of unusual behaviour. For example, temporal spikes in social media platforms, such as sudden bursts or drops in activity, may result in a higher degree of suspicion for two users with the same number of retweets to a message

(Liu et al., 2008). Bursts of reviews indicate sudden surges in the popularity of products or instances of spam attacks, raising concerns among potential review spammers (Fei et al., 2013). Li et al. (2016) investigate co-bursting behaviour to identify spammer groups likely to be actively involved in writing reviews not only for one but for multiple products. These accounts typically behave normally initially to gain credibility before being utilized to post fake reviews, ultimately leading to higher review ratings and collectively launching spam campaigns. Xie et al. (2012) concentrate on an abnormally correlated temporal pattern detection technique to uncover spam attacks, which are typically bursty and either positively or negatively correlated with review ratings. Both subtle persistence and bursty patterns significantly complement the information provided by traditional RFM features. This broader perspective contributes to the identification of potentially suspicious activities to support KYC risk assessment using OB data.

4.2.3 Clustering on bank transactions

Clustering bank transactions is a widely adopted technique for detecting fraud (Kian & Obaid, 2022; Kargari & Eshghi, 2018; Oluwafolake & Solomon, 2017; Bharati et al., 2018; Aghabozorgi & Wah, 2014) marketing analysis (Zhu et al., 2021; Singh et al., 2014; Farajian & Mohammadi, 2010; Aliyev et al., 2020; Nofal, 2024; Abbasimehr & Shabani, 2021; Abbasimehr & Bahrini, 2022; Hossain et al., 2020; Abbasimehr & Baghery, 2022; Barkhordar et al., 2021) and gaining insights into emerging developments within specific areas (Reijerink, 2024; Bartels, 2022; Ghaharian et al., 2023). The literature typically highlights three main clustering methodologies: partitional-based algorithms (Kian & Obaid, 2022; Kargari & Eshghi, 2018; Zhu et al., 2021; Farajian & Mohammadi, 2010; Reijerink, 2024; Bartels, 2022; Ghaharian et al., 2023), density-based algorithms (Aliyev et al., 2020; Nofal, 2024; Oluwafolake & Solomon, 2017; Bharati et al., 2018), and time series clustering algorithms (Abbasimehr & Shabani, 2021; Abbasimehr & Bahrini, 2022; Hossain et al., 2020; Aghabozorgi & Wah, 2014; Abbasimehr & Baghery, 2022; Barkhordar et al., 2021; Bakoben et al., 2020).

Partitional-based clustering: Partitional-based clustering divides data points into homogeneous groups based on their similarities. Among these, the k-means clustering algorithm is widely used in the literature for identifying customer groupings due to its simplicity and efficiency. For example, Kian & Obaid (2022) employ this approach to detect unusual behaviours in ATM card transactions by analysing the frequency and monetary value of transactions. They identify suspicious patterns, such as repeated use of a card at a single ATM over a short period or large cash withdrawals from an ATM, to help establish fraud prevention rules. Additionally, Kargari & Eshghi (2018) enhance fraud detection by combining association rule mining with fuzzy k-means clustering. They extract rules based on demographic attributes, payment channels, transaction timing, and amounts. Transactions that deviate from these established rules are flagged as risky. Fuzzy K-means clustering is then used to analyse trends by grouping aggregated transactions across various time intervals (e.g., 1 hour, 3 hours, up to 168 hours) for different payment channels. Transactions with a low degree of membership in all clusters are considered high risk. This combined approach categorises transactions into low, medium, or high risk based on their likelihood of being fraudulent.

Several studies have explored the use of k-means clustering to support marketing decisions. Zhu et al. (2021) introduce a fast Kolmogorov-Smirnov k-means clustering algorithm (KSKC) designed to identify merchant groups with varying behaviours in online card payment services. They develop normalized KS statistics to measure dissimilarity between empirical distributions of transaction amounts, using this metric as the distance function for k-means clustering. This approach effectively distinguishes between merchants with unusual behaviours, such as cash-out activities potentially linked to credit card fraud, and those engaging only in promotional activities without active business operations. Singh et al. (2014) extend k-means clustering to handle large datasets by parallelising the algorithm to cluster 1 million retailers based on electronic fund transfers. By focusing on an 18-day window of point-of-sale transactions, their analysis reveals distinct cluster behaviours that necessitate different marketing strategies. These strategies include retention efforts for profitable customers, incentives to boost business, and consideration of attrition risk for inactive retailers. Farajian & Mohammadi (2010) apply k-means clustering to profile data and RFM attributes aggregated over 12 months of debit card transactions. They then extract association rules within each cluster to characterise customer profiles, informing targeted marketing strategies.

For new developments in the financial industry, clustering has been a useful tool utilised to explore new consumer behavioural insights. Reijerink (2024) investigate various user types within the payment domain to understand the transition from cash to online digital payments in the Netherlands. By applying k-means clustering to payment transactions and demographic data, they identify diverse payment behaviours and reveal the connections between socioeconomic factors and payment preferences. Their profiles show that middle-to-high-income, highly educated individuals predominantly use internet and mobile banking, with minimal reliance on cash. Conversely, senior citizens and lower-income individuals, including those facing financial difficulties, tend to use cash more frequently. Additionally, younger consumers, who engage in infrequent and low-value transactions, use both digital and cash methods for their payments.

Following the recent implementation of OB, Bartels (2022) perform cluster analysis to segment customers based on OB data and analyse their financial behaviours in current account transactions. By examining three months of transaction data, they compute RFM attributes from specific inflow and outflow subcategories—such as recurrent and non-recurrent income, as well as basic, discretionary, and luxury expenses. These RFM values are discretised into five quantiles. The k-means clustering then groups the customers into the most valuable, medium valuable, and least valuable segments, highlighting groups that are potentially financially vulnerable or valuable. Ghaharian et al. (2023) explore the relationship between digital payment trails and specific activities, such as gambling. They identify subgroups of gamblers to gain preliminary insights into behavioural markers associated with gambling harm by analysing payment transaction data. This data includes deposits and withdrawals between (i) bank accounts and digital wallets, and (ii) digital wallets and gambling operator betting accounts. The k-means algorithm demonstrates superior stability compared to other methods (partitional around medoids, Gaussian mixture models, and hierarchical clustering). Their findings reveal two major clusters of low-risk accounts and three smaller clusters of potential risk accounts. These risk clusters exhibit harmful behaviours related to (1) frequent and intense payments, (2) a lower number of withdrawals relative to deposits, (3) substantial variation in deposited amounts, and (4) a higher incidence of transaction declines, particularly when accompanied by small deposit amounts.

Density-based clustering: Density-based clustering identifies clusters where data points are grouped in dense regions and separated by sparser areas. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm offers two main advantages: (i) it can identify clusters without requiring the number of clusters to be predetermined and (ii) it can detect noise points as outliers, which helps in identifying anomalous patterns. These DBSCAN noise points can reveal valuable customer characteristics for relationship management. Aliyev et al. (2020) use DBSCAN to segment bank card transactions based on RFM features to identify profitable customer groups.

Initially, they label noise points with DBSCAN and then apply K-means clustering to further analyse these outliers. The DBSCAN outliers highlight customers with high transaction volumes in terms of frequency and monetary value, which are deemed valuable and targeted by the marketing team. Nofal (2024) focus on identifying high-value bank customers by analysing the frequency and amount of transactions across various payment channels (e.g., cash deposits, bills, credit cards, debit cards). Their goal is to distinguish between special, high-value customers who should receive targeted offers and treatments, and standard accounts. They use density-based and spectral clustering algorithms to classify accounts as either special or standard, which then serves as a target variable for classification models. Special customers are characterized by frequent transactions of relatively large amounts, and customized offers are developed to enhance customer retention for this valuable group.

DBSCAN has been effectively combined with other algorithms to enhance fraud detection. Oluwafolake & Solomon (2017) integrate DBSCAN with a rule-based engine to reduce false alarms in credit card fraud detection. After identifying outliers using DBSCAN, they apply three additional rules before flagging a transaction as fraudulent: (i) the transaction amount exceeds 200% of the total outflow in the past three months, (ii) the transaction country code matches the customer's country code, and (iii) a check is performed to determine if the transaction originates from a stolen card or if the billing address of an online transaction differs from the shipping address. Bharati et al. (2018) approach credit card fraud detection in two stages. In the first stage, they use a hidden Markov model to track the sequence of low, medium, and high-risk spending profiles over time, identifying latent groups of fraud and non-fraud. In the second stage, they apply DBSCAN to the non-fraudulent latent group to detect noise points, providing an additional layer of screening to flag suspicious transactions.

Time series clustering: Time series clustering groups sequences of time series based on their temporal patterns. There are three primary approaches to time series clustering: distance-based, feature-based, and model-based (Aghabozorgi et al., 2015). The distance-based approach relies on defining distance metrics to measure similarity between time series. Euclidean Distance (ED) and Dynamic Time Warping (DTW) are two commonly used metrics in this context (Holder et al., 2024). ED calculates the sum of squared differences between corresponding points in the time series, providing a straightforward measure of similarity. In contrast, DTW is a more specialized metric that assesses the similarity between time series by finding the optimal alignment that minimizes the distance between them, even if the series vary in length or speed. DTW is often preferred for its ability to handle distortions and differences in timing between series, making it a valuable tool for clustering and comparison in time series analysis (Holder et al., 2024).

Several studies have examined the segmentation of point-of-sale (POS) purchase transactions into low, medium, and high-value customer groups. Abbasimehr & Shabani (2021) aggregate daily transactions to create monthly RFM time series data. They then perform hierarchical clustering on these time series using various distance metrics, including ED, DTW, correlation-based measures, and temporal correlation coefficients, to identify the final clusters. Abbasimehr & Bahrini (2022) also focus on POS transactions within grocery and appliance retail sectors, but using weekly-aggregated RFM time series. They employ DTW as a baseline metric, alongside other time series metrics, and compare the results using spectral and hierarchical clustering algorithms. Both studies analyse time series trends—whether stable, upward, or downward—along with monetary values, to categorize customers into low-value (potential churners), middle-value, and high-value (growing customers) groups. These insights are then used to develop targeted marketing strategies.

In analysing banking customer segments based on daily transaction counts for retail and business transactions, Hossain et al. (2020) use ED and DTW as time series metrics within k-means, k-medoids, and Self-Organizing Maps (SOM) clustering algorithms. They further characterise these customer segments using association rule mining, generating rules from contextual information such as post-tax income, product count, and age. Their findings reveal new opportunities for banks to refine their customer targeting strategies: (i) On the retail side, they identify a group of elderly, low-income individuals with few accounts who rarely transact, which challenges the conventional expectation that older individuals have larger savings portfolios. (ii) On the business side, they uncover a prominent cluster of businesses with unusually low transaction volumes, indicating that while the bank has successfully attracted a large number of these businesses, it has not fully capitalised on them for profitability. In contrast, Aghabozorgi & Wah (2014) focus on developing a new methodology to enhance computational efficiency in time series clustering, rather than on consumer profiling. They propose a two-level time series clustering algorithm designed to improve computational performance in segmenting credit card outstanding amount time series. In the initial phase, they use hierarchical clustering based on the longest common subsequence similarity measure. The resulting prototypes (centroids) from this phase are then used as initial centroids for the fuzzy K-means algorithm in the second phase, optimising the clustering process.

The feature-based approach for clustering time series is also employed in the literature. Abbasimehr & Baghery (2022) build on their previous work (Abbasimehr & Shabani (2021), Abbasimehr & Bahrini (2022)) to analyse customer behaviour based on weekly-aggregated POS transactions from grocery, home appliance, and supermarket merchants. They propose a feature-based time series clustering approach that involves calculating various time series features—such as trend, linearity, curvature, and seasonality—to serve as inputs for clustering algorithms like k-means, k-medoids, fuzzy C-means, and SOM. Consistent with their earlier studies, they characterize the resulting clusters by their trends—decreasing, increasing, or stable—to guide the development of targeted marketing strategies. Additionally, they forecast future customer behaviour for each cluster. Barkhordar et al. (2021) employ long short-term memory autoencoders to extract features from credit card transaction amount time series. Their results show that k-means clustering based on these extracted latent features achieves better separability compared to clustering using DTW time series representations and RFM variables.

The model-based approach represents a more sophisticated method for clustering, as it leverages parameters from other models to run the clustering algorithm. Bakoben et al. (2020) distinguish between low and high-risk customer groups based on their monthly credit behaviours using a two-stage clustering process. In the first stage, they train a vector autoregression (VAR) model on time series data for monthly repayment amounts and utilisation rates. To address statistical uncertainty arising from changing patterns, they introduce a novel distance metric that measures dissimilarity between the confidence regions of the VAR coefficients. The k-medoids clustering is then applied using this distance measure. The inclusion of cluster assignments in the predictive model improves the accuracy of credit default forecasts. Bester & Rosman (2024) focus on segmenting customers into homogeneous groups based on their spending behaviour, as encoded within transaction data. They propose a deep representation learning approach to partition customers into clusters. Their deep representation clustering algorithm comprises four key components: (i) an autoencoder architecture that generates an alternative representation of raw transaction amount time series, (ii) dimensionality reduction to extract important features, (iii) a pretext loss function to ensure that the latent features effectively represent the original data, and (iv) a clustering loss function to achieve optimal clustering results.

We identify three key limitations in the reviewed clustering studies (on bank transactions) that this research aims to address.

Data used: Despite the potential of OB data as a new source of information on current account transactions—including money inflows, outflows, and available balances—there has been limited exploration of consumer behaviour using OB data. Previous studies predominantly focus on money outflows, such as credit card repayments (Bakoben et al., 2020), credit or debit card purchases (Farajian & Mohammadi, 2010; Aliyev et al., 2020; Oluwafolake & Solomon, 2017; Bharati et al., 2018; Barkhordar et al., 2021), POS expenditures (Abbasimehr & Shabani, 2021; Abbasimehr & Bahrini, 2022; Abbasimehr & Baghery, 2022), online retail payments (Zhu et al., 2021; Singh et al., 2014; Ghaharian et al., 2023), etc. This focus on expenditure alone does not fully capture personal financial characteristics, as it overlooks the income side of financial behaviour. The study by Bartels (2022) is one of the few that utilizes OB data to segment customers. However, it primarily concentrates on specific income and expenditure subcategories, neglecting other important aspects such as savings, debt, and transfers. As a result, the segments derived from their study do not provide a comprehensive view of personal financial behaviours. This chapter aims to address these gaps by exploring distinct segments of personal financial behaviour using OB data. It will incorporate a broader range of financial aspects, including income, discretionary and non-discretionary expenses, transfers, balances, debt, and savings, to provide a more complete picture of individual financial behaviours.

Purpose of cluster analysis: Cluster analysis of OB transactions for KYC risk assessment represents a significant gap in the literature. Leveraging OB data to conduct KYC checks at the onboarding stage, rather than after the banking relationship has commenced, could provide valuable insights. Cluster analysis is essential for identifying distinctive financial behaviours, enhancing the understanding of customer profiles, and facilitating smooth onboarding. It also aids in distinguishing between normal and abnormal patterns, which is crucial for flagging suspicious activities early in the business relationship, even before transaction data becomes available for monitoring. However, current frameworks for interpreting clusters from bank transactions are not adequately suited for KYC purposes within the OB context.

First, some studies focus on the methodological development of clustering large-scale bank transactions (Aghabozorgi & Wah, 2014; Barkhordar et al., 2021). However, the clustering outputs from these studies may lack robustness for KYC profiling due to insufficient interpretation of the identified clusters. Second, customer relationship management (CRM) is another significant application of bank transaction clustering. In this context, clusters are characterised to create tiered consumer profiles—from low to high risk—to inform marketing strategies (Zhu et al., 2021; Singh et al., 2014; Farajian & Mohammadi, 2010; Bartels, 2022; Ghaharian et al., 2023; Nofal, 2024; Abbasimehr & Shabani, 2021; Abbasimehr & Bahrini, 2022; Abbasimehr & Baghery, 2022). While this approach effectively segments different financial behaviours from OB data, CRM clusters typically do not address anomalous instances that could signal exposure to financial crimes. This omission could result in missing crucial alerts necessary for KYC checks, potentially leaving gaps in risk assessment.

In contrast, several studies focus on clustering to identify noise points from bank transactions for fraud detection (Oluwafolake & Solomon, 2017; Bharati et al., 2018). These anomalies are then evaluated using domain-specific rules to assess their risk of credit card fraud or money laundering. However, open banking (OB) is a relatively new field and lacks established expertise for defining what constitutes suspicious activity. Without standardized criteria for identifying suspicious behaviour, it is crucial not to concentrate solely on noise points. Instead, analysing all identified segments becomes essential for distinguishing between 'normal' and 'suspicious' activities. This chapter highlights two key approaches for interpreting OB-based clusters for KYC assessment: (1) characterising the centroids to build detailed consumer profiles that can serve as a digital identity derived from transaction records, and (2) using these profiles as a baseline to differentiate between 'normal' and 'abnormal' behaviours. This comprehensive approach enhances the ability to justify and flag suspicious patterns more effectively.

Clustering methodology: Partitional-based clustering using aggregated RFM features is the most prevalent method for clustering bank transactions. K-means is widely favoured for its simplicity and efficiency, while hierarchical clustering is also used for its advantage of not requiring a predetermined number of clusters. Density-based clustering is also a commonly used technique, typically employed for outlier detection. Time series clustering identifies customer segments directly from time series, most commonly with the F and M values. This approach is particularly useful for dynamic segmentation, where tracking trends is crucial for characterising customer profiles. Despite the advancements in these methods, the application of clustering techniques to OB transactions for exploring personal finance behavioural profiles remains unexplored.

Clustering of consumer behaviours from banking transactions has been conducted in the context of retail or merchant payments, credit or debit card purchases, cash withdrawals or transfers, online digital payments, and gambling engagements. While Bartels (2022) is the only research that has applied clustering on OB transactions, the segmentation is focused on spending behaviours in the expenditure subcategories i.e., basic, discretionary and luxury, to complement identifying profitable customers for marketing purposes, they do not explore profiles to portray personal financial behaviours i.e., the habits in managing financial resources. While the literature has effectively use the RFM representation to find customer clusters, RFM could not capture the recurrent patterns in OB transactions (as discussed in Section). As an initial exploratory study, we focus on aggregated features to understand the overall financial behaviour, rather than investigation on trend analysis. This chapter initiates an OB-specific clustering setup, using the conventional clustering algorithm (partitional and density based) on aggregated RFMP features to explore the personal financial behavioural groups.

4.2.4 Anomaly detection in KYC

Supervised machine learning remains the conventional method for KYC risk scoring, where the primary approach involves classifying fraudulent data (Ngai et al., 2011). However, applying classification techniques to conduct KYC with OB data presents two challenges. Firstly, labelling suspicious behaviour in OB data proves difficult. Traditional financial crime detection has focused on specific contexts such as insurance fraud, credit card fraud, and money laundering (Hilal et al., 2022), where fraud labelling has relied on past event occurrences or expert knowledge. In contrast, integrating OB into KYC during onboarding represents a broader adaptation without specific ties to particular criminal types. Consequently, there lacks clear guidance or definition on how to identify suspicious activity from OB data. Secondly, manual labelling is cost-intensive (Ngai et al., 2011; Hilal et al., 2022). As OB adoption rates rise, manually investigating suspicious behaviour to provide labels for a classification framework becomes increasingly time-consuming and resource-intensive. To address these challenges and reduce costs, the fraud detection literature has witnessed a notable shift from supervised techniques to unsupervised anomaly detection methods (Ngai et al., 2011; Hilal et al., 2022). In the absence of domain-specific expertise or prior OB-related crime incidents, unsupervised anomaly detection emerges as a reliable tool for assessing suspiciousness based on deviations from normal behaviour.

Common unsupervised techniques utilised in KYC include network-based models and outlier detection methods. Network-based models establish connections between nodes (such as senders or recipients) via edges (transactions) within a graph network. These models compute network centrality measures to identify important nodes within the network, thereby revealing both normal and abnormal patterns (Dreżewski et al., 2012; Colladon & Remondi, 2017; Segovia-Vargas et al., 2022; Van Vlasselaer et al., 2015; Van Belle et al., 2023). Dreżewski et al. (2012) conduct social network analysis on bank statements and national court registers to investigate money laundering cases. They introduce a node proximity module to assign roles to the nodes, followed by analysing the connections between these roles to compute the proximity of entities. This method helps detect bank accounts owned by the same individuals, automating the exploration of interactions between offenders' bank transfers to identify their roles within criminal groups. Colladon & Remondi (2017) evaluate business transactions within a factoring company to highlight risky clients and visualize implicit crime links among different companies sharing the same owner. By computing network metrics, they organize and map relational data between companies to develop a predictive model for assessing the risk profiles of clients involved in the factoring business. These metrics reveal that the most risky social actors tend to be more peripheral in the transactions network and often engage in financial operations of larger amounts.

Segovia-Vargas et al. (2022) assess the risk of interactions among peer companies in a dynamic financial system to identify transactions involving shell companies. They compute vectors of risk metrics from transactions to gauge the intensity of connections between nodes, subsequently establishing thresholds to identify suspicious connections. By incorporating business transactions and legal person attributes, the dynamic social network allows for self and group comparisons of companies, considering variations in interactions over time. Van Vlasselaer et al. (2015) construct a network to link transactions between credit card holders and merchants. They propagate fraud from network edges to all components within the network to derive the network object exposure score over the long, medium, and short terms. A higher exposure score for a network object indicates that the node and edge are surrounded by more fraud within the neighbourhood. Van Belle et al. (2023) replace the rule-based system with a real-time application of network representation learning. They propose a tripartite graph design to extract individual embeddings for each transaction. This graph representation allows for the injection of fraud to augment the network, enhancing the distinction between fraudulent and genuine transactions. While network-based models prove effective in exploring fraud through transaction interactions, constructing networks is challenging with OB data, as cash flow transactions only record amounts and descriptions without access to exact sender and recipient information.

On the contrary, outlier detection (OD) is another unsupervised method prevalent in the literature for assessing KYC risk, focusing on scoring the degree of deviation from the underlying data distribution (Luna et al., 2018; Stripling et al., 2018; Vandervorst et al., 2022; Carcillo et al., 2021; Nian et al., 2016). Given the absence of explicit sender-recipient information in OB cash flow transactions, the OD technique becomes pertinent. It enables the computation of outlier scores solely based on transactional patterns' attributes, eliminating the need for sender-recipient records and facilitating the identification of anomalies within the OB context. The following paragraphs provide an overview of previous applications of OD techniques in detecting financial crimes, serving as inspiration for establishing KYC protocols within the OB framework.

OD serves as a KYC mechanism for ranking abnormalities, with each OD method employing a specific approach to represent the underlying distribution of normal data points as the baseline for distinguishing distant abnormal points. Luna et al. (2018) demonstrate that the density-based local outlier factor outperforms distance and statistical-based OD algorithms in detecting shell companies when evaluated with synthetic bank transactions simulated from real-world fraud cases. These OD models compute outlier scores from both incoming and outgoing transactions, allowing for the ranking of bank accounts and the creation of an initial list of potential shell company candidates for further investigation by financial experts. Among the detected shell companies, some common fraud patterns include an excessive number of incoming transactions in small amounts or cash, frequent large transactions with a specific overseas company, or a substantial volume of transactions involving the same set of shell accounts. Stripling et al. (2018) adapt the tree-based isolation forest for the conditional estimation of anomaly scores relative to a reference group of interest. This approach can identify fraudulent worker compensation claims from a combination of nominal and numeric attributes that remain undiscovered in the unconditional version. The tree-based OD technique offers a data-driven strategy to focus on specific subgroups, such as individuals working in construction, and identify abnormal durations of incapacity conditional on a group of workers sharing the same nominal attributes.

Given that data misrepresentation heightens the exposure risk to insurance premium fraud, Vandervorst et al. (2022) explicitly model the distribution of self-reported variables and their correlation with pricing policies to compute anomaly scores for user-filled information. This allows for the evaluation of data misrepresentation risk and pricing accuracy. They propose a conditional density estimation algorithm capable of handling mixed data types and utilize Shapley additive explanations to interpret the estimated conditional densities, thus justifying contract refusals. To compute outlier scores from credit card transaction behaviour, Carcillo et al. (2021) employ statistical techniques such as z-score, principal component analysis, Gaussian mixture estimation, and the non-parametric tree-based isolation forest method. Outlier scores are computed at three levels of granularity: global, local, and cluster. The global level considers all transactions collectively and identifies anomalies that deviate from the overall multivariate pattern. The local level detects anomalies at the card level, flagging transactions that significantly differ from past behaviour by the same card. The cluster level identifies group-specific anomalies. Incorporating these outlier scores as additional feature vectors in a supervised credit card fraud model demonstrates promising results, particularly with cluster-level outlier scores, which enhance fraud detection accuracy and exhibit high feature importance ranking, thereby adding value to risk prediction. Nian et al. (2016) derive anomaly rankings for auto insurance fraud using a graph-based spectral method. By focusing on the first non-principal eigenvector of a Laplacian matrix, they discern the data distribution for normal instances and assess the suspiciousness of insurance claims. This unsupervised spectral ranking method effectively explores multiple major normal patterns, generating fraud score rankings based on deviations from the majority.

While OD has proven effective in financial crime detection, there is no universally applicable technique for fitting data distributions to represent normal patterns. The selection of OD algorithms and the configuration of anomaly score rankings vary depending on the specific application domain. These specialised fraud models have been developed to detect a particular fraud type to a high degree of accuracy. Nonetheless, OB current accounts are susceptible to different financial crime context, where using the specialised fraud detection model may leave out the identification of the other types of suspicious activities. A generalised OB-based fraud detection model is yet to be developed. Implementing OD as a KYC instrument based on OB data necessitates tailored adaptation to the characteristics of cash flow data.

4.2.5 Limitation

Drawing from KYC-related studies aimed at identifying potential suspicious patterns related to financial crimes, adapting OB data for KYC purposes remains an exploratory task, with three significant unaddressed challenges: (i) the definition of 'suspicious' within the context of OB cash flow transactions is unknown, lacking specific definitions or guidelines; (ii) the exploration of attribute representations derived from transactional data in existing literature lacks exploitation of bank current account cash flow transactions, thus limiting the exploitation of personal financial behaviour for KYC purposes; (iii) model selection for identifying suspicious activities has predominantly focused on domain-specific solutions, neglecting to account for the unique properties inherent in cash flow data. These limitations are further detailed as follows.

Defining 'suspicious': Previous KYC-related studies have been tailored to specific contexts, such as identifying suspicious behaviour related to money laundering (Dreżewski et al., 2012; Colladon & Remondi, 2017; Segovia-Vargas et al., 2022; Luna et al., 2018), credit card fraud (Van Vlasselaer et al., 2015; Van Belle et al., 2023), insurance fraud claims (Stripling et al., 2018; Vandervorst et al., 2022), or irregular investment trading (Thompson et al., 2021). These financial crimes are typically characterised by previous occurrences and validated manually by experts, serving as reference points for determining true 'suspicious' events. Moreover, transactions within these contexts are typically focused on specific purposes. For example, in the case of money laundering, transactions typically involve fund transfers between entities, while credit card fraud transactions usually consist of card purchases. In contrast, establishing what constitutes 'suspicious' in OB cash flow transactions poses challenges due to the multifaceted nature of OB-connected accounts, which are utilized for various day-to-day personal finance activities such as receiving income, bill payments, necessary expenses, credit repayments, and more. This diversity in transaction purposes complicates the definition of 'suspicious' because expected normal behaviours can vary significantly compared to other financial crime contexts. In the absence of expert knowledge, the task of detecting potential suspicious activities from OB transactions remains largely unexplored in the literature.

Data representations: Transactional data has been represented using the recency, frequency, and monetary (RFM) attributes, serving as a method to summarise fundamental transactional patterns (Chang & Tsai, 2011; Alboukaey et al., 2020; Olson & Chae, 2012). However, the RFM framework proves inadequate for capturing the distinctive characteristics of OB cash flow transactions, such as recurring transactions for consistent income or fixed-scheduled payment commitments in personal finance contexts. This limitation arises because none of the three RFM dimensions considers the time gap between transaction occurrences, thus failing to capture regularity effectively. To address this gap, the literature has introduced the persistence measure as a means to quantify repeated patterns in evolving transactions (Belth et al., 2020). The persistence measure can detect subtle anomalies and bursty patterns that could indicate potential suspicious activities, which may go unnoticed when relying solely on RFM. However, to date, the integration of the persistence measure with RFM to represent repeated cash flow patterns remains unexplored.

Model selection: Recent advances in financial crime detection have shifted towards unsupervised techniques, notably graph-based network models and OD methods. The graph-based network approach offers the advantage of uncovering abnormalities through the interconnections between transactions (Dreżewski et al., 2012; Colladon & Remondi, 2017; Segovia-Vargas et al., 2022; Van Vlasselaer et al., 2015; Van Belle et al., 2023). However, this method necessitates explicit information on sender and recipient records, a requirement not met by OB data. Consequently, we turn our attention to the outlier detection approach for spotting suspicious activities within OB data. Nevertheless, existing applications of OD in KYC-related works present three limitations that must be addressed when adapting to the OB context.

First, the conventional approach to identifying suspicious activities typically involves establishing a binary profile to distinguish abnormal from normal patterns (Dreżewski et al., 2012; Colladon & Remondi, 2017; Segovia-Vargas et al., 2022; Van Vlasselaer et al., 2015; Van Belle et al., 2023; Luna et al., 2018; Stripling et al., 2018; Vandervorst et al., 2022; Carcillo et al., 2021). However, this method often overlooks the existence of multiple segments of normal patterns, especially within the context of KYC profiling. Exploring various profiles of current accounts with diverse income and spending levels is crucial not only to differentiate between typical and atypical accounts based on varying degrees of financial risk but also to provide guidance in understanding the characteristics of minority abnormalities. This ensures that rare anomalous cases are not mistaken for true 'suspicious' activities. Second, the selection of an OD algorithm tends to be highly subjective and dependent on domain-specific solutions. While some studies have focused solely on enhancing a single technique for detecting financial fraud (Stripling et al., 2018; Vandervorst et al., 2022; Van Vlasselaer et al., 2015), these approaches are often grounded in past literature that justifies their specific choice for addressing the problem at hand. However, the field of OB is still relatively new and lacks established data patterns. Consequently, relying solely on a single method carries the risk of producing biased estimations when attempting to model normal behaviour and identify abnormalities. During the exploratory phase, it is essential for OD technique selection to encompass a range of potential data patterns. Third, existing fraud detection practices have predominantly focused on identifying suspicious activities in a data-driven manner, without delving into explicit explanations to validate the factors driving such peculiar behaviour. Among the reviewed studies, only Vandervorst et al. (2022) have provided justifications for flagging suspicious instances based on model explanations. The incorporation of feature interpretation, interpreting the influences of various factors on the model's predictions, remains under-researched within the domain of KYC.

4.3 Methodology

This section outlines the three-phase KYC framework for OB. Phase 1 prepares the RFMP attributes to represent financial behaviour from cash flow transactions. Phase 2 clusters the accounts into distinctive groups to understand the profiles of typical vs atypical behaviour. Phase 3 derives outlier scores to risk-rank the suspiciousness level of the accounts and label them as outlier (high-risk) vs non-outlier (low-risk), followed by interpreting the characteristics of high-risk accounts to highlight potential suspicious activities.

4.3.1 Phase 1: Data and RFMP attributes

The OB data comprises 10,535 connected UK current accounts with 12 months of transactional records[16]. Each transaction carries the information of an account identifier, a date, a description, a credit or debit indicator, an amount, a category label, and the remaining balance after the transaction. For the purpose of methodological description in the coming section, the raw transactional records are given the following notations, listed in Table 4.1.

To extract financial behaviour from cash flow transactions, we focus on eight financial aspects of inflow, outflow, balance, and expenditures in fixed cost, living cost, debt payment, saving, and

[16] Details of the data can be found in Table 1.2

Table 4.1: Notations of raw transactional records

Variable	Abbreviation
Account Identifier	ID
Transaction Date	t
Description	r_1
Transaction Type	r_2
Amount	r_3
Category	r_4
End balance	r_5

transfer.

We introduce a Recency, Frequency, Monetary, Persistence (RFMP) framework to depict financial activities in each aspect. While RFM attributes have proven effective in marketing strategies for summarising customer purchase transactions and exploring spending patterns, we extend their use to represent income and expenditure behaviours from cash flow transactions. However, RFM neglect transaction regularity, where these three components do not incorporate the distribution of intervals between transaction occurrences to accurately quantify repeated and consistent patterns. The addition of the P dimension to RFM is motivated by several factors. Firstly, the P dimension serves to quantify regularity, which holds particular significance in the context of bank account transactions for capturing agreed recurring payments to/from a merchant and understanding account holders' habitual behaviour. Furthermore, the P dimension is crucial for detecting various types of anomalies, such as subtle (infrequent yet highly regular over an extended period) and bursty (numerous occurrences with low persistence) activities, which deviate from typical patterns. These anomalies provide valuable insights beyond what RFM alone can offer, thereby enhancing the comprehension of KYC profiles and facilitating the identification of potential suspicious activities.

Persistence extends the idea of temporal motif, by introducing activity snippets (Belth et al., 2020). Temporal motifs are a collection of directed timestamped edges conforming to a specified patterns within a duration of time in which the events occur. Temporal motifs focus on representing the general activity among nodes, by calculating the counts of the subgraph patterns in an evolving network. On the other hand, activity snippets in Persistence capture exact activity i.e., specific edge in the network, which consider the network activity as a stream of edge updates (insertions of deletions) over time. Apart from the frequency of the activity occurrences, persistence encodes the dynamics of the evolving networks, by measuring how long is the occurrence, and how uniformly is the spread out of the occurrences, together with its frequency.

The following paragraphs describe the creation of the RFMP components. The attributes from RFM dimensions (for each financial aspect) are quantified as follows.

- Obtain the monthly-aggregated RFM values (gives a total of 24 RFM time series across the eight aspects).
  - Recency (R): the number of days since the previous transaction occurred (measured on the last day of the month).
  - Frequency (F): the total number of transactions per month.
  - Monetary (M): the total amount of transactions per month
- Compute the attributes based on the summary of the RFM dimensions over 12 months to depict the overall behaviour. Through this approach, we compute the average monthly income for an account that receives a salary solely in the first three months, by dividing the sum of the three-month earnings by 12 months. This computation effectively portrays the income loss scenario for the remaining nine months by evenly distributing the total income across all 12 months, to prevent over-representation of income based on the few high-earning months. The summary attributes are listed below.
  - Standard statistical metrics to represent the overall status: minimum, maximum, mean and standard deviation.
  - Functional data principal component scores to summarise changing patterns across the 12 months.
  - Metrics to capture evolving patterns: the proportion of months with values below the mean, the proportion of months with values lower than the preceding month, the proportion of months with extreme outliers, the entropy of inflow versus outflow, and coefficient derived from linear regression of the time series against the 12-month time points to indicate trend (no trend, increasing or decreasing).

The Persistence component is calculated from daily transactions spanning 12 months, with a higher value indicating a higher degree of regularity. Let R denotes a set of J raw transactions.

Each raw transaction is represented as $R^j = (ID^j, t^j, r_1^j, r_2^j, r_3^j, r_4^j, r_5^j)$ with the information listed in Table 4.1 for $j = 1, 2, ..., J$ transactions. The bank account transactions can be denoted as an evolving network $G = (V, E)$ with a set of nodes V and a set of edges $E \subset V \times V$. The following steps describe the formulation of the Persistence attribute (Belth et al., 2020).

- Form an activity snippet z from the collection of the sequence of edge updates.
  - For each account with a set of transactions in the time interval $[t_s, t_e]$ where t_s and t_e represent the date where first and last transaction occurs, each transaction occurrence represents an edge update.
  - An edge update $u = (v_1, v_2)$ connects the account identifier, ID (node v_1) and the category label, r_4 (node v_2), representing an insertion of a new transaction at timestamp t into the network G.
- Compute Persistence for each activity snippet that occurs between $[t_s, t_e]$ with an exponential function combining the three components of:
  - Width measured from the percentage of interval width covered by the occurrence of z,
  - Frequency measured from the logarithm of the number of occurrences in the interval, and
  - Spread measured from the distribution gap time between the occurrences.

4.3.2 Phase 2: Clustering

This section describes the implementation of clustering on the RFMP attributes to identify different groups of account behaviour. The input data is denoted as an $n \times p$ matrix X to represent the p-dimensional RFMP attributes for n accounts. We use the notation x_i to indicate the i-th data instance, for $i = 1, 2, ..., n$. First, we outline the clustering algorithms employed in this experiment. Focusing on k-means and hierarchical clustering, we further explore the possibility of discovering non-identical clusters by comparing the cluster outputs run on the original RFMP attributes with two other different data representations from a 50×50 self-organising map and a 10-dimensional autoencoder bottleneck layer. To determine the optimal number of clusters, the elbow method and silhouette scores are examined together. Second, we propose a clustering setup which first filters out 'strong' outlier accounts, followed by clustering on the remaining accounts. This is to directly spot accounts with high risk of being anomalous and ensure their noisy behaviour do not distort the data space of the standard ones. Third, we conduct cluster profiling to analyse the representative financial circumstances in each cluster and distinguish between typical vs atypical behaviour. Clustering algorithms are run with the Python `scikit-learn` package (k-means and hierarchical), `minisom` package (SOM) and `pyod` package (AE).

4.3.2.1 Clustering algorithms

4.3.2.1.1 k-means clustering

k-means clustering (k-means) is a technique to partition all data points into k clusters (MacQueen, 1967). Each observation is assigned to the closest cluster centroid i.e., mean of all observations that belong to the cluster. The cluster assignment process minimises the sum of squared Euclidean distances between the data points and centroids to obtain homogeneous clusters with minimum within cluster variances. The k-means process is outlined below (Aggarwal Charu & Reddy Chandan, 2013).

1. Randomly select k data points as the initial centroids.

2. Compute the Euclidean distance of each data point with the centroids.

3. Given n data points denoted as $D = \{x_1, x_2, ..., x_n\}$ and k clusters denoted as $C = \{C_1, C_2, ..., C_K\}$, the objective function of the cluster assignment process is to minimise the sum of squared error (SSE), equivalently the within-cluster variance,

$$SSE(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - c_k||^2, \tag{65}$$

   such that c_k represents the cluster centroid i.e., mean of each cluster obtained by dividing the attribute values with the cluster size.

4. Repeat steps 2-3 until there is no change in the centroids.

4.3.2.1.2 Hierarchical clustering

Agglomerative hierarchical clustering (HIER) starts from each data point as its own cluster and successively merges the most similar cluster pair until all points form a single large cluster (Kaufman & Rousseeuw, 2009). The merging requires a linkage criteria to measure the dissimilarity between the sets of data points. This study uses the Ward linkage such that the choice of cluster pair to merge leads to a minimum increase in total within-cluster variance (squared of Euclidean distance between the centroids) after the merge. The recursive merging process can be represented as a binary dendogram tree. The leaves at the bottom of the tree are the single data points. The branches join the pairwise data points (or subclusters) moving up the tree where the height of the branch is equivalent to the dissimilarity measure of the joined data points. The root at the top of the tree takes all data points. The agglomerative hierarchical clustering algorithm with Ward's criterion is detailed as follows (Aggarwal et al., 2015).

1. Construct the dissimilarity matrix (Euclidean distance) between all data points.

2. Compute the Ward's criterion to measure the merging cost $\Delta(A, B)$ for any two clusters C_A and C_B with centroids c_A and c_B, and cardinality N_A and N_B,

$$\begin{aligned}\Delta(A,B) &= \sum_{i \in A \cup B} ||x_i - \bar{x}_{A \cup B}||^2 - \sum_{i \in A} ||x_i - \bar{x}_A||^2 - \sum_{i \in B} ||x_i - \bar{x}_B||^2 \\ &= \frac{N_A N_B}{N_A + N_B} ||c_A - c_B||^2.\end{aligned} \tag{66}$$

   Merge the cluster pair with the minimum increase in the total within-cluster variance (minimum $\Delta(A, B)$).

3. Update the dissimilarity matrix by inserting a new row and column with the distances between the newly-merged cluster $C_{A \cup B}$ and the other clusters.

4. Repeat steps 2-3 until only one cluster remains.

4.3.2.1.3 Self organising map

A Self Organising Map (SOM) is a type of artificial neural network trained with competitive learning to produce a low-dimensional representation of the input data, typically a two-dimensional map, while retaining the underlying topological structure of the original data (Kohonen, 1990). The original RFMP attributes are now represented as neurons in the SOM map. The clustering of the accounts, using k-means and hierarchical clustering, is now conducted on these extracted features (projected data dimension in the SOM topological structure). Given an input data with p-dimensional vectors of $x = [\xi_1, ..., \xi_p]$, each of i neuron in the $(m \times l)$ grid map is a p-dimensional weight vector denoted as $w_i = [\mu_{i1}, ..., \mu_{ip}]$. The SOM training is as follows.

1. Initialisation of the neurons' weights in the mapping layer.

2. Select a data instance x as the training input vector.

3. Examine the similarity of all neurons with the input vector. The neuron with the minimum Euclidean distance to the input vector is the Best Matching Unit (BMU).

4. The BMU and its neighbouring neurons are activated and adjusted to become closer to the input vector with
$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)], \tag{67}$$
where t is the time step, $h_{ci}(t) = \alpha(t) \cdot \exp\{-d_{ci}^2/2\sigma_i^2(t)\}$ is the Gaussian neighbourhood function depending on the distance between BMU and the respective neuron c at time step t (d_{ci}), and $\sigma_i(t)$ is the kernel width which decays over time such that the weight updating starts from a global scale and gradually focus locally. For a similar reason, the learning rate ($\alpha(t)$) also gradually decreases over iterations.

5. Terminate SOM training until maximum iteration is reached.

4.3.2.1.4 Autoencoder

Autoencoder (AE) is an unsupervised artificial neural network with the goal to reconstruct the input data. The key components in AE are the encoder, bottleneck and decoder (Ng, 2011; Bank et al., 2020). Autoencoders (AEs) have been employed for clustering in two primary ways. The first method leverages AEs for feature reconstruction, aiming to uncover latent patterns by compressing the original data into a lower-dimensional feature space. This compressed representation is then used as input for the clustering algorithm (Tavakoli et al., 2020; Song et al., 2013; Ryu et al., 2019; Amarbayasgalan et al., 2018). The second method contributes to methodological advancements by integrating the clustering task directly into the AE framework, utilizing a clustering loss function to optimize the updates of cluster centroids (Lim et al., 2020; Zhang & Qian, 2021; Zamini & Montazer, 2018).

In this chapter, we adopt the first approach. Our goal is to explore and incorporate potential hidden features that might be overlooked if clustering were performed directly on the original RFMP features. The bottleneck layer compresses the input data, providing a latent feature representation of the original RFMP attributes. Both the k-means and hierarchical clustering are run on the bottleneck latent features, instead of the original data, to group the accounts. Both the encoder and decoder are neural network functions represented in Equations (68) and (69) respectively,
$$H = f_1(W_1X + b_1), \tag{68}$$

$$\hat{x} = f_2(W_2 H + b_2), \tag{69}$$

where $f(\cdot)$ is the rectified linear unit activation function, W and b are the weight matrix and bias for the encoder (W_1, b_1) and decoder (W_2, b_2). H is the latent features which is the output from the encoder (f_1) and input for the decoder (f_2) to reconstruct the original data as $\hat{x}$. AE training aims to find the functions f_1 and f_2 that minimises the reconstruction error i.e., smallest difference between the original input and the reconstructed data instances. While Equations (68) and (69) represent a single layer neural network, they can also be feedforward multilayer perceptrons with several layers of neurons.

To extract essential features from AE, the latent representation H is made the bottleneck layer with the constraint to have a smaller dimension than the input data X. The AE network architecture in this study has a total of five layers and is symmetrical with respect to the bottleneck layer in the middle. This multilayer network setup is to capture potential non-linearities from the attributes. The first layer has the same dimension as X, the second layer has a smaller dimension of $(n \times (p/2))$, and the bottleneck layer is restricted to $(n \times 10)$ dimension to compress the input attributes. The next two decoder layers are the mirrored version of the layers before the bottleneck.

4.3.2.1.5 DBSCAN

Density based spatial clustering of applications with noise (DBSCAN) is a clustering algorithm to group points that are closely-packed together (high density with many nearby neighbouring points) and assign points in low-density regions as outliers (nearest neighbours are too far away) (Ester et al., 1996). The key idea of DBSCAN is that the density in the neighbourhood has to exceed certain threshold, such that for each point in a cluster, the neighbourhood of a given radius has to contain at least a minimum number of points. The notion of "clusters" and "noise" given a dataset D containing points $p \in D$ is formalised in the following definitions.

Definition 1: ϵ- neighbourhood. The ϵ- neighbourhood of a data point p, $N_\epsilon(p)$, is defined as the set of points within a radius ϵ around p,

$$N_\epsilon(p) = q \in D | d(p, q) \leq \epsilon, \tag{70}$$

where d is a distance measure and $\epsilon \in \mathbb{R}^+$ represents the specified radius. Note that $p \in N_\epsilon(p)$, p

is always part of its own ϵ- neighbourhood. The dense regions are detected based on $N_\epsilon(p)$ and a threshold $minPts$ (the minimum number of points in a neighbourhood where $minPts \in \mathbb{Z}^+$ is a user-specified threshold). The points are then classified as core, border, or noise points.

Definition 2: Point classes. Each point $p \in D$ is classified as

- core point: if $N_\epsilon(p)$ has high density where $|N_\epsilon(p)| \geq minPts$, or
- border point: if p is not a core point, but in the neighbourhood of a core point i.e., $p \in N_\epsilon(q)$ where $q \in D$, or
- noise point: otherwise.

Definition 3: Directly density-reachable. A point $q \in D$ is directly density-reachable from a point $p \in D$ with respect to ϵ and $minPts$, if and only if,

- p is a core point: $|N_\epsilon(p)| \geq minPts$, and
- q is in its $\epsilon-$ neighbourhood: $q \in N_\epsilon(p)$.

Definition 4: Density-reachable. A point p is density-reachable from a point q with respect to ϵ and $minPts$ if there exists in D a chain of points $(p_1, \ldots, p_n)$ where $q = p_1$ and $p = p_n$, such that p_{i+1} is directly density-reachable from p_i with $i \in 1, 2, \ldots, n-1$.

Definition 5: Density-connected. A point p is density-connected to a point q with respect to ϵ and $minPts$ if there exists a point o, where both p and q are density-reachable from o.

Definition 6: Density-based cluster. A density-based cluster C is a non-empty subset of D which satisfies the following criteria:

- Maximality: $\forall p, q$, if $p \in C$ and q is density-reachable from p, then $q \in C$.
- Connectivity: $\forall p, q \in C$, p is density-connected to q.

The DBSCAN algorithm works as follows. Starting with an arbitrary point p,

1. Retrieve the ϵ- neighbourhood for point p.

2. Cluster formation: If p is a core point, start a new cluster by assigning all points in its neighbourhood to the cluster. If another core point is found in the neighbourhood, expand the search to include all points in its neighbourhood. The cluster is complete if no more core points are found in the expanded neighbourhood.

3. Search the remaining points to examine whether there is another core point to be found to start a new cluster.

4. Assign points which are not assigned to a cluster as noise points.

4.3.2.2 Clustering setup

We propose a preliminary step to extract 'strong' outlier accounts as a separate high-risk group and run clustering on the remaining accounts. This is to ensure the extremely skewed data points do not mislead the clustering process in identifying standard behaviour. Such extra step leads to two clustering setups where clustering is run on (i) full accounts (Setup 1) and (ii) remaining accounts after filtering out 'strong' outlier accounts (Setup 2). Table 4.2 summarises the the 12 clustering algorithms run across both setups.

Table 4.2: Clustering setups

	Setup 1	Setup 2
kmeans	✓	✓
hierarchical	✓	✓
SOM + kmeans	✓	✓
SOM + hierarchical	✓	✓
AE + kmeans	✓	✓
AE + hierarchical	✓	✓

Setup 1: before filtering out the strong outliers

Setup 2: after filtering out the strong outliers

An account is defined as a 'strong' outlier when it satisfies both outlier checks as follows.

1. Functional boxplot (FBP) is used to examine the 12-month RFM[17] time series over the eight aspects. While the conventional boxplot identifies the interquartile range from a single data

[17] Outlier check with FBP serves as the initial screening step. It is only run with the raw time series values without including other summarised attributes i.e., RFM summary attributes and Persistence (they are used for outlier check with the OD algorithms in point number 2.)

attribute, FBP identifies a 50% central envelope region from functional data, which is the 12-month time series in our context. Then, in analogy to the empirical rule of 1.5 times interquartile range, FBP detects a time series as an outlier when there are any observations lying out of the 1.5 times 50% central envelope region. Given a total of 24 RFM time series from the eight aspects, an account is an outlier if there are more than 6 (a quarter) of the 24 time series are outliers.

2. Outlier detection algorithms (detailed in Section 4.3.3) are used to compute the anomaly score for each account. The average anomaly score represents the final score for each account. Accounts with the top 5% highest scores are outliers.

We use the elbow method to determine the optimal number of clusters. All the clustering techniques across both setups are evaluated with the most commonly used internal clustering metrics, i.e., silhouette score (Rousseeuw, 1987), Calinski-harabasz index (ch-score) (Caliński & Harabasz, 1974), and Davies bouldin score (db-score) (Davies & Bouldin, 1979).

4.3.2.3 Cluster profiling

The cluster profiles are defined from the RFMP centroids (Decker et al., 2006; Tuma et al., 2011). We compare the relative difference (higher or lower values) of the centroids to infer the representative financial behaviour in each cluster. Due to the unsupervised nature of this study, there is no way to evaluate the correctness of the identified account behaviour. Thus, we conduct a word n-gram analysis on the transactions' descriptions to characterise the cluster-specific key activities as additional details to validate the RFMP profiles. The n-gram tokenises the descriptions into consecutive sequence of n words. Note that we focus on trigram terms because they give us better understanding on the purpose of the transactions, e.g., grocery shopping, clothing, retail, and etc.

Prior to obtaining the trigram terms, the descriptions have to be pre-processed by removing stopwords and alphanumeric characters, lower casing for all (except the first letter of each word), replacing digits with the capital letter 'X' and deleting blank descriptions. Using the tagged category, we group the descriptions into six themes, i.e., income, essential cost, non-essential cost, transfer, cash and cheque, and others. The procedure to characterise the activities in each cluster is summarised below.

1. Obtain all trigram terms.

2. Compute the frequency (number of occurrences) of the trigram terms.

3. Rank the trigram terms frequency in descending order from highest to lowest occurrences.

4. Focus on the top 50 trigram terms and manually examine the related financial activity of each trigram term e.g., 'Tesco Pump XXX' indicates fuel expenses.

5. Compute the proportion of trigram term occurrences with respect to the financial activity by dividing the sum of trigram terms occurrence in each financial activity with the frequency sum over the top 50 trigram terms.

6. Reveal the dominant financial activities from those with high proportion of trigram term occurrences.

7. Repeat this in every description theme and for all clusters.

4.3.3 Phase 3: Anomaly risk assessment

Anomaly risk assessment measures the level of deviation/being an outlier of an account as a proxy to evaluate the exposure risk to financial crime. We propose an outlier detection (OD) scheme which incorporates several state-of-the-art algorithms to compute the outlier score where a higher value implies a higher degree of relation to suspicious behaviour. Based on the outlier scores derived from the scheme, we then self-label the accounts as outliers to further examine the characteristics driving the high anomaly risk and highlight peculiar attributes relating to potential suspicious financial activities. In the coming paragraphs, we first present the selected OD algorithms, followed by providing details on the proposed scheme to label outliers and outline the explainability tool used to reveal the underlying drivers of anomalies. The OD algorithms are run using Python `pyod` package.

We employ the following OD algorithms to capture both local and global outliers that could signal anomalous patterns.

Local Outlier Factor (LOF) detects outliers based on the relative density of a data point with its neighbouring points. Outliers are points that having low density which indicate they have high deviation from the local neighbourhood (Breunig et al., 2000). LOF determines locality from the density estimated with k-nearest neighbours. Given a data point x, the set of k-nearest neighbours of x is denoted as $N_k(x)$ and the local reachability density (LRD) of x is defined as the inverse of the average reachability distance (RD) from point x to all its k-neighbours,

$$LRD_k(x) = \left(\frac{\sum_{y \in N_k(x)} RD_k(x,y)}{|N_k(x)|} \right)^{-1}, \tag{71}$$

where $RD_K(x, y)$ is the maximum of (i) the distance between x and y or (ii) distance between x with the k-th neighbour. The LOF score is defined as the ratio of the average LRD of the k-neighbours of x to the LRD of x,

$$LOF_k(x) = \frac{\sum_{y \in N_k(x)} LRD_k(y)}{|N_k(x)|} \cdot \frac{1}{LRD_k(x)}. \tag{72}$$

Isolation forest (IForest) is inspired by the idea that outliers are few and different from the regular data instances, thus easier to be isolated. IForest is an ensemble of isolation trees (ITree) (Liu et al., 2008). Sub-sampling via random selection of data instances without replacement is the key strategy used in IForest to effectively isolate anomalies by alleviating the swamping (subsamples control the data size to reduce to risk of mistakenly detect normal instances as anomalies) and masking (each subsample specialise the ITree by including different set of anomalies or even no anomaly to avoid anomalies being concealed even where there are plenty of them). IForest follows a two-stage process: (i) training and (ii) scoring to detect anomalies. The training stage constructs ITrees from subsamples to form IForest with the following steps.

1. For an input data X, determine the number of trees in the forest (T), set the subsampling size (ψ) and the height limit for ITree at $l = ceiling(log_2\psi)$.
2. For each Itree,
   - Obtain a subsample of X with size ψ.
   - Recursively partition the data by randomly select an attribute and randomly split it at a value between the maximum and minimum.
   - Continue the recursive binary splits until either reaching the height limit (l), only one instance left in the node, or all instances in the node have the same values.
3. Form the IForest with the collection of all T ITrees.

The scoring stage computes an outlier score for a new data instance based on the average path length i.e., number of edges the data instance traverse from the root to the terminal node, across all trees.

Cluster based local outlier factor (CBLOF) defines outliers from (i) the size of the cluster which the data point belongs to and (ii) the distance of the data point to the closest large cluster (He et al., 2003). Let $C = c_1, c_2, ..., c_k$ be the set of k clusters obtained from the input data X with a clustering algorithm sorted in descending order of the cluster size, a cluster is defined as the boundary cluster C_b when either one of the following two conditions holds,

1. $\sum_{i=1}^{b} |c_i| \geq \alpha |D|$ or
2. $\frac{|C_b|}{|C_{b+1}|}$,

then large clusters is defined as $LC = \{c_i | i \leq b\}$ and small clusters is defined as $SC = \{c_j | j > b\}$, where α and β are two numeric parameters to denote the proportion of data in LC and the ratio of the size of any cluster in LC relative to the size of SC. The outlier score for data point x for which $x \in c_i$ is then computed by the size of its cluster and the distance to its centroid (if c_i is large cluster) or the distance to the closest large cluster centroid (if c_i is small cluster and c_j is a large cluster where $x \notin c_j$),

$$CBLOF(x) = \begin{cases} |c_i| \cdot min(distance(x, c_j)), & \text{if } c_i \text{ is small cluster,} \\ |c_i| \cdot distance(x, c_i), & \text{if } c_i \text{ is large cluster.} \end{cases} \tag{73}$$

Scalable unsupervised outlier detection algorithm (SUOD) is an ensemble of heterogeneous unsupervised outlier detection algorithm in an efficient manner (Zhao et al., 2021). Due to the absence of ground truth, the reliance on a single algorithm poses risk of having wrong assumption in searching for outliers. SUOD provides a three-module acceleration framework to leverage the use of multiple models. The first module induces diversity into model training through the random projection of higher dimensional data into a lower dimensional space. The data compression scheme follows a linear transformation with the Johnson-Linderstraus (JL) projection (Johnson, 1984). The input data $X \in \mathbb{R}^{n \times p}$ will be projected to a subspace $S \in \mathbb{R}^{n \times d}$ at a lower dimension of $d = \frac{p}{2}$. For a basic JL projection, each entry in the subspace S is independently drawn from a standard Gaussian $N(0, 1)$ distribution and the projected space preserves the distance relationship between the data instances to ensure no heavy distortion on the original data while injecting diversity via the transformation.

The second module handles the training with multiple models in a balanced parallel scheduling mechanism to enable all workers finish the scheduled tasks within similar duration. The authors proposed a balanced parallel scheduling heuristic to impose nearly equal resources to each worker based on the sum of the rank on running time, instead of the time itself. The outlier score is computed from the average of the output from the parallel runs. The third module predicts on unseen test instances where the output outlier score from the second module is used as the pseudo target and the prediction is done with an efficient supervised regressor. We utilise the first two modules to facilitate the diversification of three unsupervised base detectors (LOF, CBLOF, IForest) to leverage the underlying assumptions in each technique to obtain the final outlier score.

The proposed outlier detection scheme comprises two parts to risk-rank and define the outlier (high-risk) vs non-outlier (low-risk) accounts. An account is tagged as outlier when its score is ranked within the top five percentile in *either one part* of the scheme. The score computation process is detailed below.

- Part 1: Compute a single outlier score with all attributes by averaging across the outlier detection (OD) algorithms.
- Part 2: Compute multiple outlier scores across specific two-dimensional spaces (2D). The inclusion of all attributes to detect outliers may mask the outlying effects that is only obvious in 2D spaces. Thus, this approach ensures the capability to capture such 2D-specific anomalies. The outlier score with respect to each 2D space is the average of the OD algorithms. Final score from Part 2 is the average score across all 2D spaces summarised in Table 4.3.

Using the self-defined outlier labels resulting from the proposed scheme, we proceed to investigate the relationship between the RFMP attributes (features) with outlier accounts (target variable) by fitting an Extreme Gradient Boosting Classifier. We employ SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) to explain the outputs from the fitted classifier. SHAP assigns a contribution value for each input attribute on a specific prediction to reveal how each feature impacts the model output i.e., being an outlier, and point out unusual attributes that might be suspicious.

Table 4.3: Specific two-dimensional spaces to compute outlier scores

Descriptions of the specific 2D spaces	Total number of 2D spaces
1. All possible combination of RFMP within each aspect	48
e.g. inflow R vs inflow F/ inflow R vs inflow M/ inflow R vs inflow P, and etc.	
2. Mean vs standard deviation (std) for F/M in each aspect	16
e.g. inflow F mean vs inflow F std/ inflow M mean vs inflow M std/	
outflow F mean vs outflow F std, and etc.	
3. Cross combination of financial aspects	
a) inflow vs outflow	4
e.g. inflow R vs outflow R/ inflow F vs outflow F/	
inflow M vs outflow M/ inflow P vs outflow P, and etc.	
b) fixed cost vs transfer	4
c) living cost vs transfer	4

4.4 Results and discussions

4.4.1 Cluster evaluation

Table 4.4 assesses the clustering performance on all accounts (Setup 1) and the remaining accounts after excluding strong outliers as a separate cluster (Setup 2). According to the elbow method, the optimal number of clusters for all clustering techniques is identified as k=4. However, Setup 2, which incorporates the strong outlier cluster, comprises a total of k=5 clusters.

Table 4.4: Evaluation of clustering algorithms

	Setup 1			Setup 2		
	silhouette	ch	db	silhouette	ch	db
kmeans	**0.09055**	**1118**	**2.732**	***0.09728***	**1116**	***2.625***
hierarchical	0.08210	956.5	2.843	0.07688	948.7	2.849
SOM + kmeans	0.07655	1006	2.867	*0.07707*	1005	2.921
SOM + hierarchical	0.05809	799.5	3.077	*0.06387*	628.8	3.753
AE + kmeans	0.08747	1006	2.795	*0.08825*	1004	2.791
AE + hierarchical	0.06634	791.4	3.250	*0.07044*	816.8	3.107
Preliminary: DBSCAN*	0.9539	320.2	2.013	-	-	-

Setup 1: before filtering out the strong outliers, Setup 2: after filtering out the strong outliers

k=2 clusters automatically detected by DBSCAN

The clustering metrics in Table 4.4 provide three findings. First, k-means performs better than hierarchical clustering with lesser cluster overlapping (higher silhouette score), larger cluster separation (higher ch-score), and more distinct clusters (lower db-score) across all scenarios. The inferior performance of the hierarchical clustering process could be due to its susceptibility to the sequence of input instances, which leads to improper merges and cluster centroid distortion. Second, SOM and AE-based approaches consistently perform worse than the standalone k-means and hierarchical clustering algorithms. Therefore, having good separation in the projected SOM and AE spaces does not necessarily relate to an equally good separability in the original data space. Third, the clustering performance is enhanced by isolating strong outliers into a single cluster (in *italic*), especially the silhouette score, in almost all techniques. This implies that the filtering procedure effectively diminishes noise in the data instances to result in lower cluster intersection. Based on these three findings, filtering out strong outliers, followed by conducting k-means clustering on the original data space, gives the best clustering results (in **bold**).

DBSCAN, a density-based clustering method introduced by Hinneburg (1996), is widely employed for the simultaneous identification of clusters and isolation of outliers. Table 4.4 presents the initial outcomes obtained with DBSCAN when applied to all accounts, serving as a benchmark against the proposed clustering techniques. DBSCAN identifies k=2 clusters after a cautious hyperparameter tuning procedure as detailed in Tiukhova et al. (2022), with several accounts flagged as outliers while the majority are grouped within a single cluster. Despite reporting superior silhouette and db-score metrics, indicating minimal cluster overlap and well-defined clusters, DBSCAN exhibits a significantly lower ch-score compared to the proposed clustering methods, suggesting that the resulting two clusters are relatively close to each other. Furthermore, the formation of only two clusters underscores the limitation of lumping nearly all accounts into a dominant cluster without the ability to discern distinct characteristics.

The suboptimal performance of DBSCAN may stem from the interdependence among financial attributes (inflow, outflow, and balance). These attributes often lack distinct, dense regions in the data space that are conducive to cluster formation. The RFMP attributes are interrelated, leading to underlying similarities in different behaviours. For example, accounts may differ in M but exhibit similar F/P due to individuals receiving regular monthly salary deposits at different income levels. Other potential scenarios include: (i) similar total grocery expenses but varying shopping patterns, such as consistent monthly bulk purchases versus random, need-based buying; (ii) similar income levels but differing degrees of credit utilisation. These interdependencies can make it challenging for DBSCAN to identify distinct clusters.

The distribution of accounts and outliers across each cluster, with and without the inclusion of strong outliers, is presented in Table 4.5. The resulting clusters are derived from the stand-alone k-means method, since it is identified as the top-performing clustering technique as reported in Table 4.4. According to the outlier detection procedure outlined in Section 4.3.3, approximately 9% of the accounts in this sample are labelled as outliers.

Table 4.5: Distribution of accounts and outliers

	Setup 1				Setup 2			
	Account		Outlier		Account		Outlier	
Cluster	number	proportion	number	proportion	number	proportion	number	proportion
1	3399	0.32	238	0.07	4474	0.42	199	0.04
2	3343	0.32	174	0.05	2267	0.22	153	0.07
3	2155	0.20	266	0.12	2127	0.20	238	0.11
4	1638	0.16	220	0.13	1493	0.14	197	0.13
5	-	-	-	-	174	0.02	111	0.63
Total	10535	1.00	898	1.00	10535	1.00	898	1.00

The distribution of accounts and outliers is similar across both setups among the clusters. With less than 10% of the accounts being outliers, Clusters 1 and 2 account for the majority, around 60% of the total accounts. These clusters reflect 'standard' behaviour with minimal noise. Clusters 3 and 4 together contain around 35% of the accounts, with a moderate level of outliers at a proportion of 10% to 15%. These clusters correspond to less common account behaviour. The primary distinction between the two experimental setups is the existence of Cluster 5, which results from the isolation of strong outliers. Cluster 5 comprises 2% of the accounts and its high outlier composition of 64% indicate a collection of rare accounts depicting atypical behaviours with high anomaly risk.

There are two main benefits to filtering strong outliers in clustering for KYC profiling. In order to create a more compact clusters that better denote typical behaviours, it first isolates rare patterns to reduce data noise. Secondly, it effectively identifies atypical profiles to reduce the number of accounts needed for additional examination. Subsequent discussions will focus on the clusters derived from the top-performing technique in the second experimental setup, comprising k=4 clusters along with an additional isolated strong outlier cluster. To understand the financial behaviour for each cluster and differentiate between typical and atypical behaviours, the next section characterises the clusters.

The identified outliers represent extreme deviations from the typical behaviours observed in the resulting clusters. Instead of focusing solely on cluster-specific outliers, the anomaly detection approach can be broadened to identify accounts that are outliers across all (or most) clusters, thereby enhancing the detection of potential suspicious activities. To implement this approach, compute the distance of each account from the cluster centroids to derive additional anomaly scores. For each cluster, select the top 5% (or another appropriate threshold) of accounts with the greatest distance (highest anomaly scores). Then, identify accounts that are distant from all (or nearly all) of the five clusters. This strategy shifts the focus from cluster-specific outliers to a more comprehensive identification of overall outliers. While this chapter concentrates on cluster-specific outliers, the described method for detecting generic outliers could serve as a valuable extension of the current approach, providing an additional layer of scrutiny for suspicious activities.

4.4.2 Cluster profiling: The RFMP profiles

To depict the clusters, our focus lies on the Recency, Frequency, Monetary, and Persistence (RFMP) centroids, computed by averaging the attributes for all accounts within each cluster. These attributes encompass inflow, outflow, balance, fixed cost, living cost, debt, saving, and transfer, as outlined in Table 4.6. Containing 40% of the accounts, Cluster 1 is the largest group and is regarded as the 'baseline' group signifying 'standard' behaviour. To analyse the remaining clusters, we highlight their relative deviations from Cluster 1 (marked in **bold** in Table 4.6) to comprehend their respective characteristics. We also point out that P has the advantage of giving additional cluster features that RFM alone might miss. The cluster profiles are elaborated in the subsequent paragraphs.

Cluster 1 represents accounts primarily utilised for day-to-day personal finances. These accounts exhibit active cash flows, indicated by low R for both inflow and outflow, and demonstrate a tendency to spend in alignment with earnings (similar M for inflow and outflow). Notably, the F and P dimensions of expenses (outflow) surpass those of income (inflow) in these accounts. This pattern aligns with typical personal finance behaviour, where monthly expenses often exceed the salary which is usually deposited once a month. These accounts maintain sufficient balances (higher M than inflow and outflow), enabling them to handle additional spending without resorting to overdrafts (low F/P in negative balance). Among the spending categories, living costs take precedence as the top priority, as indicated by the lowest R and the highest F/M/P. Fixed costs follow as the next priority. Transfer payments have regular patterns (relatively lower F/M but higher P compared to fixed costs), while savings display irregular spikes with large accumulations over time (similar M but lower F/P compared to transfer). Debt is minimal, with the lowest F/M/P, indicating that these accounts do not heavily rely on credit facilities to manage their finances.

Cluster 2 also consists of accounts intended for day-to-day personal finance usage, yet it distinguishes itself with active credit utilization. While having RFMP values akin to those of Cluster 1 in most aspects, Cluster 2 demonstrates a significant divergence in the debt category. Here, Cluster 2 presents the lowest R and the highest M/P. Consequently, this cluster displays the heaviest reliance on credit utilisation among all clusters. The elevated P in the debt category suggests a well-managed debt schedule characterised by on-time repayments.

Cluster 3 characterises accounts utilised for occasional payments, portraying minimal cash flow activity compared to other clusters. These accounts typically have the highest R and the lowest F/P. Despite featuring infrequent outflows, Cluster 3 shows a similar outflow M to Cluster 1, indicating occasional substantial outflows across only a few transactions. Upon closer inspection of spending categories, outflow transactions in Cluster 3 primarily pertain to transfers, with no discernible spending priority observed in fixed and living costs, where fixed and living costs have high R and low F across all clusters.

Cluster 4 denotes accounts that are susceptible to becoming over-indebted. This cluster highly resembles Cluster 2; however, the accounts within it are in a more precarious indebted state. Within this cluster, the balance M is at its lowest, while the F/P are the highest amongst the clusters, indicating a greater susceptibility to financial strain. These accounts have a lower balance M compared to the inflow M, implying a continual deficit scenario. Furthermore, the consistent use of overdrafts, surpassing even inflow and outflow, indicates a persistent and prolonged financial shortfall. Although both Clusters 4 and 2 display high F/M/P in the debt category, the relatively higher F but lower P in Cluster 4 suggests riskier financial circumstances, such as borrowing from multiple credit facilities, likely involving short-term high-cost loans with irregular repayment schedules. Despite the low balance and heavy reliance on debt, these accounts remain active in daily expenditure (fixed and living costs), indicating a significant dependence on overdrafts or other forms of debt to meet financial obligations.

Cluster 5 comprises accounts depicting a high volume of transactions, showcasing several distinct financial attributes. Having habitual saving practices (lowest R and highest F/P) with substantial accumulated savings (highest M) is one of the financial patterns discovered in this cluster. The remaining results discovered from this cluster unveil characteristics diverged from conventional personal finance behaviour. First, all aspects except for living costs, exhibit the highest M with a significant 10-fold difference compared to the other clusters. Despite the unusually large M, there is no proportional increase in F/P, thus signifying atypical scenarios where each transaction involves a huge money flow. Second, this cluster reports a higher F of inflow than outflow. Having more income transactions than expenses contradicts standard personal finance behaviour. Lastly, this cluster indicates the highest F/M/P in transfer payments, implying that large transfer amounts occur regularly, at a nearly equivalent level to living costs, which is also uncommon.

In addition to RFM, the inclusion of the P dimension reveals several additional key characteristics of the cluster profiles. For Clusters 1 and 2, the subtle regularities (low F/M and high P) pick up recurring merchant payments as one of the primary expenditures, alongside fixed and living costs. Such rare and low-value repetitive expenses would have gone unnoticed without the incorporation of the P dimension. The low P/F for fixed and living costs in Cluster 3 confirms that the low transaction counts represent insignificant random behaviour with no discernible habitual spending. Moreover, a high P/F/M for the transfer category in Cluster 5 indicates that the frequent large-value transactions are consistent practices rather than bursty behaviour with transaction spikes during certain periods. Similarly, a high P/F for overdraft use in Cluster 4 represents a consistent negative situation evenly spread over the entire period, rather than an abrupt use of overdrafts in response to financial shock. The relative comparison of P in the debt category between Cluster 2 and 4 also provides insights into differentiating regularly scheduled versus fluctuating credit repayment patterns. The supplementary information from P not only enriches the understanding of cluster profiles but also provides evidence to alert potential suspicious activities (discussed in Section 4.4.4).

We further relate these cluster profiles to the findings in previous chapters on cash flow volatility (Chapter 2) and financial transitions (Chapter 3). Accounts in Cluster 1 exhibit stable day-to-day finances, suggesting low risk and low volatility (in absence of vulnerable characteristics). These accounts are likely to show long-term stability with minimal exposure to financial distress. For Cluster 2, the accounts experience moderate, non-threatening cash flow fluctuations. They might show unstable cash flows due to limited spending capacity, but active credit use with timely repayments can mitigate this. Consequently, these accounts may frequently transition between low and medium distress risks. Accounts in Cluster 3 have occasional payments that indicate medium cash flow volatility, potentially relating to transitions between low-to-medium and medium-to-low distress risks. Cluster 4 reflects over-indebted accounts with medium to high volatility risk, due to irregular credit repayments or insufficient funds. Accounts in Cluster 4 are likely to face persistent high distress risk throughout the year and may struggle to recover.

Assessing the cluster outputs and the detected outliers is challenging in the absence of data labels. In the upcoming sections, we conduct additional analyses to validate the identified clusters and explore the significant distinctions between outliers and non-outliers.

Table 4.6: Cluster centroids of RFMP attributes

Bold figures represent significant (lower/higher) deviation relative to the baseline Cluster 1

	Inflow				Outflow				Balance				Fixed cost			
Cluster	R	F	M	P	R	F	M	P	R*	F*	M	P*	R	F	M	P
1	0.0999	0.0499	0.0470	0.2181	0.0357	0.0626	0.0046	0.2627	0.8675	0.0241	0.0082	0.0241	0.4516	0.0650	0.0023	0.0978
2	0.0929	0.0425	0.0051	0.2026	0.0317	0.0640	0.0050	0.2787	0.8554	0.0259	0.0081	0.0259	0.3476	0.0924	0.0026	0.1374
3	**0.3704**	0.0122	0.0046	**0.0656**	**0.3884**	0.0080	0.0043	**0.0353**	0.8944	0.0356	**0.0115**	0.0356	**0.8839**	**0.0068**	**0.0006**	**0.0102**
4	0.0976	0.0505	0.0040	0.2246	0.0372	0.0770	0.0040	0.3210	**0.1440**	**0.6284**	**0.0067**	**0.6284**	0.3699	0.0955	0.0022	0.1381
5	0.1215	0.0780	**0.0487**	0.2479	0.0602	0.0687	**0.0489**	0.2647	0.8245	0.0824	**0.0222**	0.0824	0.4756	0.0749	**0.0153**	0.1084

	Living cost				Debt				Saving				Transfer			
Cluster	R	F	M	P	R	F	M	P	R	F	M	P	R	F	M	P
1	0.1434	0.1391	0.0113	0.2540	0.9528	0.0004	0.0001	0.0009	0.9178	0.0041	0.0024	0.0081	0.3172	0.0394	0.0016	0.1419
2	0.1428	0.1433	0.0122	0.2612	**0.5769**	**0.0386**	**0.0025**	**0.0682**	0.9424	0.0008	0.0007	0.0023	0.2657	0.0380	0.0018	0.1471
3	**0.7942**	**0.0096**	**0.0019**	**0.0191**	0.9231	0.0025	0.0002	0.0025	0.9436	0.0002	0.0002	0.0005	**0.6358**	**0.0158**	0.0020	**0.0571**
4	0.1530	0.1445	0.0107	0.2562	**0.5564**	**0.1364**	**0.0013**	**0.0406**	0.9482	0.0006	0.0003	0.0014	0.2740	0.0425	0.0013	0.1557
5	0.3162	0.1181	0.0291	0.1972	0.6644	0.0364	**0.0147**	0.0540	**0.7149**	**0.0282**	**0.0252**	**0.0640**	0.2866	**0.0570**	**0.0258**	**0.1696**

*R/F/P of being in negative balance

4.4.3 Validation of cluster profiles

Analysing n-grams in transaction descriptions outlines the main activities within each cluster, aiding in the validation of the RFMP profiles. Transaction descriptions are categorised into six themes: income, essential costs, non-essential costs, transfers, cash or cheque, and other. Focusing on trigram terms (sequences of three words), we extract the top 50 most frequently occurring terms to gain insight into the key activities in each cluster.

Table 4.7 reports the proportions of trigram word frequencies for each description theme across all clusters. These proportions indicate the relative significance of one theme compared to another within each cluster. The findings in Table 4.7 correspond with the RFMP profiles (Table 4.6). For Clusters 1, 2, and 4, expenditures are more prevalent than income, and the primary activities revolve around essential and non-essential costs. For Clusters 3 and 5, earning activities surpass expenses and transfer payments account for a higher proportion than essential costs.

Table 4.7: Proportion of trigram terms

	Cluster				
Description group	1	2	3	4	5
income	0.2042	0.1661	0.3008	0.1211	0.4096
essential	0.2666	0.3030	0.1441	0.3274	0.1382
non-essential	0.2068	0.1960	0.0877	0.2398	0.1722
transfer	0.1573	0.1681	0.3064	0.0984	0.1668
cash & cheque	0.0404	0.0367	0.0616	0.0298	0.0206
others	0.1246	0.1301	0.0994	0.1835	0.0925

The breakdown of key activities noticed from the top 50 highest frequency trigram terms within each description theme is presented in Table 4.8. A greater proportion denotes a higher number of trigram terms associated with a particular activity. Many of these trigram terms, however, only provide limited information regarding the true purpose of the transaction, since they relate to payment details such as dates, amounts, or reference numbers. As such, these details are not regarded as representative activities. The subsequent analysis explores the differences in activities between the typical and atypical profiles.

Clusters 1, 2, and 4 refer to profiles related to day-to-day personal finance across different scenarios. The discovered major activities revealed from the most frequent trigram terms add to the profile characterisation. Income sources in these clusters are received from company payments of government benefits, via different payment channels i.e., fast payment, bank giro credit, or fund transfers. Essential costs are dominated by groceries, transportation, and fuel expenses. In addition, the essential theme also identifies regular direct debits, potentially representing fixed expenses, as the main activity. Notably, interest-related trigram terms emerge as a significant activity solely in Clusters 2 and 4, with Cluster 4 exhibiting a notably higher percentage of interest-related terms at 22.7%.

The predominant activities observed in the remaining description themes are as follows: (i) Non-essential costs refer to product or service subscriptions, goods purchases, and transaction fees or charges; (ii) Transfers are typically mandated payments and money transfers to other individuals in the form of fast payment, mobile or fund transfers; (iii) The 'others' theme depicts living expenses in alternative food or transport options, such as delivery or ride-hailing services. Cluster 4 diverges from Clusters 1 and 2 in several aspects. In the non-essential cost theme, leisure and entertainment activities, particularly gambling and gaming, carry almost equal weight to subscription and purchase activities. Moreover, credit-related activities, such as buy-now-pay-later or overdraft fees, constitute a significant component of the 'others' theme. These disparities highlight the unhealthy gambling habits and high risk of over-indebtedness associated with Cluster 4.

Cluster 3 comprises accounts used as secondary accounts rather than for daily personal finance purposes. Income sources involve 7.6% of transfers via financial services and 1.5% of cash deposits, suggesting occasional significant third-party inflows. The non-essential cost category is dominated by fees and charges for account maintenance, which even surpass the non-essential living expenses. Additionally, transfer payments to individuals are prevalent in this cluster, where we observe 14.6% of the trigram terms are associated with people's names.

Unlike the other profiles, Cluster 5 exhibits atypical activities deviating from typical personal finance purposes. We highlight the activities that set this cluster apart from the others throughout the description themes. Inflow via financial services emerges as the primary channel for receiving income, rather than conventional bank payment methods such as bank giro credit, fund transfer, or fast payment. Essential cost is dominated by service payments (4.3%), transfers (7.7%) and others (11.7%). Rare characteristics, such as payroll (5.9% in essential costs), cheque transactions

(10.5% in cash & cheque), and saving-related activities (11.3% in others) are also observed solely in this cluster. Therefore, the unusually large volume of transactions in this cluster may be related to current account usage for business or saving purposes.

Table 4.8: Key activities observed from trigram terms

Description group	Cluster 1	2	3	4	5
income	payment method: 50.1% company: 19.6% benefit: 7.6%	payment method: 47.8% company: 25.3% benefit: 9.9%	payment method: 65% company: 8.1% financial service: 7.6% cash deposit: 1.5%	payment method: 60.7% company: 12.8% benefit: 7.4% cash deposit: 0.5%	financial service: 48.6% payment method: 33% company: 4.1%
essential	grocers/transport/fuel: 55.4% regular payment: 1.7%	grocers/transport/fuel: 54.6% regular payment: 2% interest: 0.9%	grocers/transport/fuel: 39.8% regular payment: 6.9% tax: 6.3%	grocers/transport/fuel: 34.9% interest: 22.7% regular payment: 2%	grocers/transport/fuel: 37.0% regular payment: 11.5% payroll: 5.9% services: 4.3%
non-essential	subscriptions: 9% purchases: 3.5% fees or charges: 3.2% leisure and entertainment: 2.4%	subscriptions: 11.7% purchases: 6.7% fees or charges: 5.1% regular payment: 2.1%	fees or charges: 20.9% subscriptions: 14.2% purchases: 3.0% regular payment: 2.8%	subscriptions: 9.8% purchases: 8.2% leisure and entertainment: 6.4% fees or charges: 2.3%	purchase: 10.9% fees or charges: 4.3% regular payment: 2.6%
transfer	payment method: 66.7% names: 8.2% regular payment: 7.4%	payment method: 55.3% regular payment: 6% names: 2.7%	payment method: 65.9% names: 14.6% regular payment: 10.7%	payment method: 63.6% regular payment: 14.8% names: 5.9%	payment method: 41.3% regular payment: 25.9% names: 19.9%
cash & cheque	atm: 73.5% post office: 1.3%	atm: 70.5%	atm: 48.2% post office: 6.2%	atm: 52% post office: 6.2%	atm: 59.6% cheque: 10.5%
others	food/transport: 14.2% payment method: 10.8% purchase: 2% credit use: 0.9%	food/transport: 10.7% payment method: 10.7% names: 3% saving: 2.5% credit use: 2.4%	payment method: 25.7% food/transport: 10% names: 3.8% saving: 3.5% credit use: 2.9%	credit use: 34.7% food/transport: 5.8% payment method: 4.7%	payment method: 14.4% services: 11.7% saving: 11.3% food/transport: 8.2% credit use: 3% names: 1.5%

4.4.4 Interpretation of outlier characteristics

Although the RFMP profiles reveal both normal and abnormal behaviours, they are insufficient on their own to relate to suspicious activities. To investigate the characteristics of outliers and their possible relationship to suspicious behaviour, we use the RFMP attributes (average values over 12 months in all eight aspects) and three calculated ratios (Transfer P to Living P, Transfer M to Living M, and Outflow M to Inflow M) as input features for an XGBoost (XGB) Classifier. The detected outliers are set as the target variable to train the classifier and SHAP (SHapley Additive exPlanations) values are obtained to comprehend attribute contributions to outlier prediction. The distribution of SHAP values is illustrated in Figure 4.1. The distinct groupings of red and blue points indicate that low and high attribute values have discernible effects on outlier prediction.

While evaluating the accuracy of outlier prediction is unfeasible due to the absence of data labels, the distinct segregation of red and blue groups depicted in Figure 4.1 validates the efficacy of the proposed outlier detection framework in distinguishing between outlier and non-outlier accounts. Each point in the figure denotes an attribute's contribution to the model's prediction, enabling us to explain the relationships between attributes and the target variable. This insight informs whether an attribute exerts a huge or small influence and whether it has a positive or negative impact on the likelihood of an account being an outlier, thus identifying peculiar characteristics that may warrant suspicion.

The main characteristics of outlier accounts can be summarized as follows. Accounts exhibiting an underspending budget are more likely to be outliers. These accounts allocate their funds towards necessities, resulting in minimal activity in living costs (high R and low P), as well as in transfer and outflow categories (high R). Overspending behaviour is another rare pattern. The accounts are characterised by higher outflows accompanied by increased fixed and living expenses (high M), along with frequent fund transfers (high F) rather than sporadic miscellaneous payments. Extreme financial difficulties are also indicative of abnormal behaviour, wherein accounts carry substantial debt commitments (high M, low R) and heavily rely on overdrafts (high F in negative balance). Furthermore, saving practices (low R, high M) are uncommon. These outlier characteristics signify extreme digression from the typical RFMP profiles, suggesting potential financial challenges rather than indicating suspicious activity.

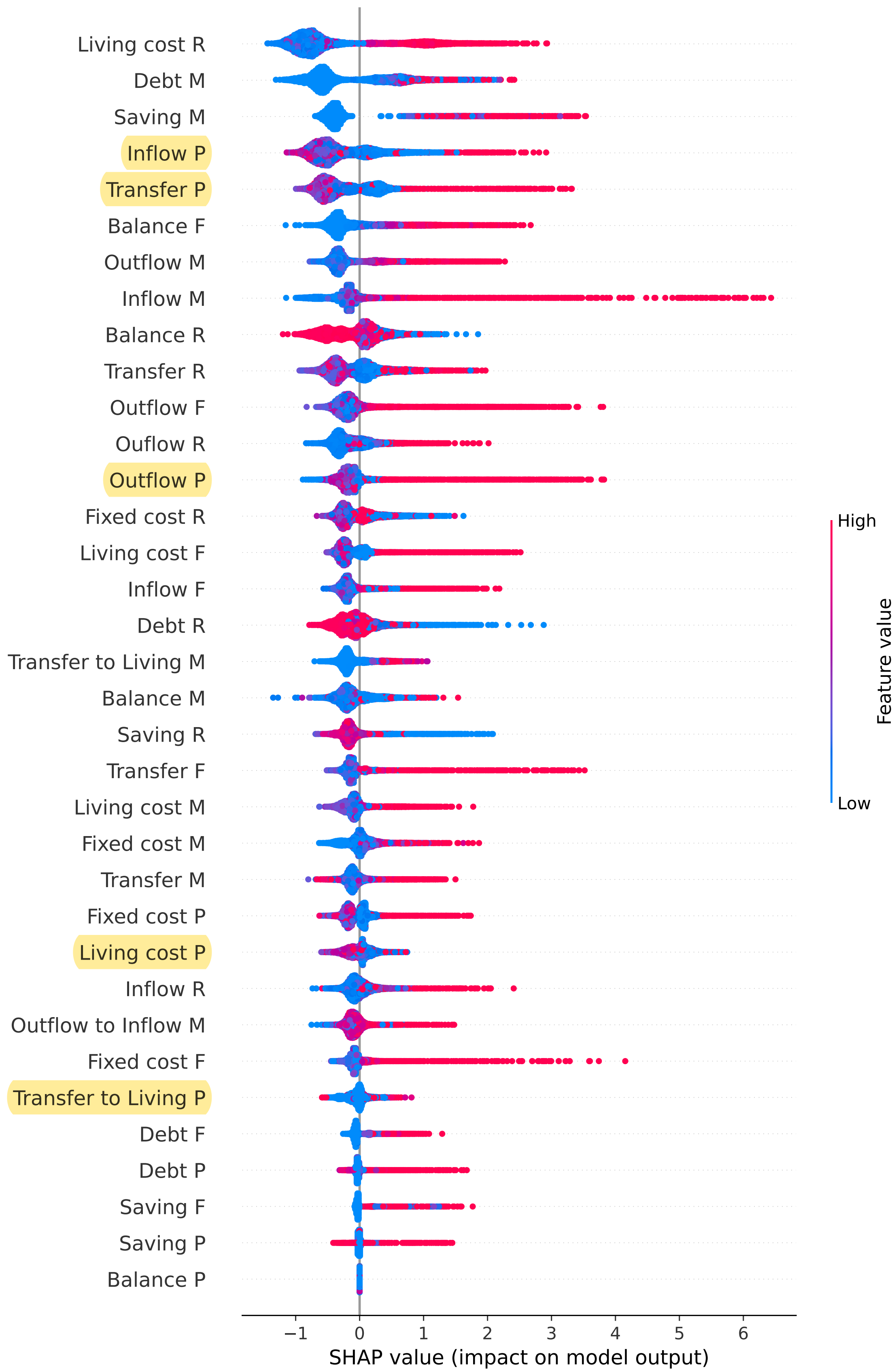


Figure 4.1: SHAP summary plot for XGB Classifier

We would like to highlight the significance of the P attributes in flagging suspicious accounts (highlighted in yellow in Figure 4.1). For instance, an over-persistent pattern of large inflow, outflow, and transfer (high P/M) may be indicative of money laundering tactics, such as consistently transferring funds to the same third party or conducting small, regular transfers that accumulate into enormous sums. While large volumes of transfers could be legitimate business-to-business transactions, an unusually high level of persistence might suggest transfers to non-legitimate shell accounts. Regarding living costs, a combination of exceptionally high P and high F could imply the concealment of non-legitimate payments under the guise of living expenses. On the other hand, simultaneously having unusually low P and high F may suggest isolated or periodic bursts of unexpected spending on living costs, potentially linked to the use of lost or stolen debit cards. Moreover, a disproportionate ratio of transfer payments to living expenses (high P/M) could signal the misuse of the account for illicit transactions. These insights indicate the importance of leveraging the P attributes to raise potentially suspicious accounts for further investigation.

4.5 Conclusion

This chapter designs a KYC tool that leverages the use of OB transactions to streamline CDD for customer onboarding. Based on the RFMP representation of bank transactions, we evaluate the financial crime exposure risk of bank accounts in two steps. Cluster analysis is first conducted to reveal profiles with varying financial risk levels with respect to the accounts' behaviour, followed by assessment with outlier scores as the proxy for suspiciousness measures and interpretation of the underlying drivers of high anomaly risk to justify the decision to flag potentially suspicious accounts.

The cluster analysis discovers five RFMP profiles. The typical profiles relate to day-to-day personal finance activities under different circumstances of (i) being non-indebted with sufficient money to keep up with living, (ii) lightly or (iii) over-indebted with low or high reliance on credit facilities to manage expenses. These clusters prioritise spending on regular direct debits or essential living costs, but the indebted profiles have debt payments as the additional key activity. Compared to the less-indebted cluster, the over-indebted group has a relatively higher proportion of transactions towards interest payment, overdraft fees and buy now pay later. We also found a less dominant profile of subsidiary accounts which only have occasional large third-party cashflow transactions. On the other hand, the atypical profile depicts deviating behaviour from daily personal finance activities with a large volume of transactions prioritised on savings or business activities such as service payments and payroll. These KYC risk profiles reveal clients-owned accounts' behaviour to better understand their needs at the start of the business relationship.

Moreover, the Persistence dimension discovers additional cluster-specific characteristics which would have been unexploited with only the RFM dimensions. These characteristics are infrequent yet regular recurring merchant payments (non-indebted profile), timely-scheduled credit repayments (lightly indebted profile), inconsistent debt payments alongside persistent overdraft reliance (over-indebted profile), random fixed and living expenditure (subsidiary profile) and large transfer amounts are consistent rather than transaction spikes over the year (large volume profile). The extended RFMP framework fully exploits regularity in cash flow transactions in obtaining more distinctive KYC risk profiles.

The anomaly risk assessment identifies 9% of the accounts as outliers yet not all of them are qualified to be escalated for further investigation. We discover several non-suspicious outlier characteristics denoting extreme deviation from the RFMP profiles i.e., underspending, overspending, unusually large debt commitments and active saving practice. These non-suspicious anomalies demonstrate how the RFMP profiles serve as the guideline to explore 'standard' accounts before being able to detect the 'suspicious' ones.

We highlight the benefit of the Persistence attributes to imply linkage to suspicious activities. For instance, over-persistent large money flow might be a money laundering strategy, the pairing of high P and high F in living cost could be illegal transactions faked under the name of daily living activities, bursty patterns with low Persistence alongside high F may signal anomalous stolen debit card purchasing behaviour, and the disproportionate ratio of transfer payments relative to living expenses might indicate illicit use of accounts. The ability to identify a set of suspicious properties provides clear justifications to escalate the accounts for further investigations.

While the developed system can effectively support KYC onboarding with OB data, there are several limitations to note on. First, OB data used in this study provides account-level rather than individual-level transactions. Thus, this study can be extended to the individual level based on the account(s) owned by the same person. Second, while the current framework is operated in an unsupervised context, a semi-supervised active learning implementation guided by experts could be a future work to improve the detection of suspicious accounts. Third, the account-based KYC assessment in this research is restricted to a static evaluation of financial risk exposure from the 12-month time frame. Potential extension could consider time series clustering, using specific distance metrics i.e., Dynamic Time Warping or other shaped-based metrics, could be employed to run clustering on the RFM time series and track the cluster-specific trend to characterise account profiles.

5 Conclusion

This thesis develops three personal financial risk assessment frameworks in the OB context. The methodological development of these OB-driven financial risk models showcases potential new ideas to deploy OB analytics in practice. After the Introduction, Chapters 2, 3 and 4 detail the empirical works to unleash the potential of OB data for risk assessment in three different ways i.e., cashflow instability, transitions of financial status, and exposure to suspicious activities. In each chapter, we introduce the methodological contributions by specifying *what* and *how* to evaluate financial risk with OB data as well as outline the empirical contributions to utilise the research findings for practical decision analytics applications. In the subsequent subsections, we conclude the contributions to the literature, outline practical implications for regulators and practitioners, and identify limitations to highlight potential future research.

5.1 Academic contribution

Chapter 2 sets up regression models to forecast future account-level volatility by incorporating past volatility, transactional patterns related to vulnerability, and a vulnerability index. This chapter is the first to introduce a novel composite volatility score, offering a comprehensive assessment of uncertainties in cash flow fluctuations. Additionally, the proposed vulnerability index objectively indicates financial struggles by synthesising four key pillars of vulnerable characteristics: low resilience, poor health, negative life events, and low capacity. Furthermore, this chapter initiates the exploration of the non-linear relationship between vulnerability and volatility, to investigate the varying risk impact of vulnerable patterns on cash flow transactions and to alert against harmful unstable circumstances.

Chapter 3 employs a multivariate latent Markov model to estimate and predict transition probabilities of distress risk over time from vulnerable-related transactional patterns. This is the initial deployment of the latent Markov model to extract underlying distress risk states from observed account balances, pioneering a statistical, data-driven approach to formalize the quantification of distress as a scaled indicator. Moreover, we incorporate previous vulnerabilities as covariates to forecast future state transitions, thereby introducing a novel framework to capture the temporal dynamics of vulnerable characteristics and project evolving exposure risk to distress. The proposed model contributes to a conceptualised statistical estimation framework for tracking and monitoring dynamic vulnerability, facilitating the assessment of distress risk transitions.

Chapter 4 develops a KYC mechanism utilising OB data to streamline customer due diligence during new client onboarding. We introduce a novel three-phase unsupervised approach for detecting suspicious bank accounts, which includes (i) employing RFMP (Recency, Frequency, Monetary, Persistence) representations of cash flow attributes, (ii) conducting cluster analysis to establish RFMP profiles as the baseline reference of 'normal' account behaviour, and (iii) implementing an outlier detection scheme to assess the degree of suspiciousness and flag accounts associated with unusual activities. The RFMP framework represents an OB-specific adaptation tailored to characterise the unique properties of cash flow data, particularly recurring transactions. The persistence dimension effectively leverages regular patterns to highlight additional cluster characteristics and identify dubious activities. This chapter serves as an initial exploratory study to delineate segments of 'normal' account behaviour in OB data before proceeding to identify 'suspicious' activities. The outlier detection phase contributes to establishing an automated anomaly risk ranking system, alongside interpreting attributes related to high-risk accounts, to efficiently justify flagging accounts for due diligence considerations.

5.2 Practical implications

The volatility forecast model introduced in Chapter 2 serves as a powerful strategic risk management tool, offering practitioners insights into the sources of financial instability to guide risk mitigation efforts. The predicted volatility score facilitates a useful ranking of accounts based on their unpredictability, enabling practitioners to identify accounts requiring special attention. Interpreting the association between vulnerability predictors and volatility unveils specific characteristics contributing to unstable future cash flow, providing early warnings of potential financial difficulty. Additionally, the non-monotonic vulnerability-volatility effect allows for the segmentation of accounts into low, medium, and high-risk categories, facilitating tailored treatment for each target risk group. Accounts categorised as low-risk can be monitored without interventions, while those in the medium-risk group require close surveillance to prevent deterioration. High-risk accounts necessitate focused inspections of potentially harmful transactional behaviour to alleviate financial hardships.

In Chapter 3, modelling the temporal dimension of FV is a valuable tool to strategise support for vulnerable consumers. The statistical framework facilitates a quantified assessment of distress risk across a spectrum of magnitudes, ranging from low to high, based on account balance situations. Additionally, the proposed framework serves as an effective diagnostic tool for identifying problematic behaviour. The model explores how vulnerability correlates with financial improvement and deterioration, highlighting specific behaviours that contribute to higher levels of distress risk. Furthermore, practitioners can monitor the progression of distress risk states to identify both short and long-term financial stress. Lastly, the model can detect early warning signals of detrimental deterioration, enabling the provision of timely support and preventing temporary vulnerability from evolving into distress.

In Chapter 4, the unsupervised anomaly risk assessment framework bolsters KYC processes using OB-connected current accounts. The RFMP representation of cash flow attributes aids practitioners in analysing personal financial behaviour objectively. Particularly, the expanded persistence dimension offers additional insights into regular recurring transactions, as well as unusually stable or irregular spiky patterns. Furthermore, firms can construct KYC risk profiles based on the derived RFMP clusters, enhancing existing identity verification checks with an additional 'financial identity' rooted in clients' financial behaviour and intended use of financial products at the onboarding stage. The anomaly risk scoring framework serves as an automated tool to enhance efficiency in screening current accounts. Practitioners can concentrate on high-risk anomalous accounts for due diligence considerations, with the flexibility to adjust the focus according to their risk appetite. Moreover, the interpretation of specific peculiar attributes associated with high-risk accounts justifies the decision to flag suspicious cases.

5.3 Limitation

Nonetheless, the proposed OB risk models have several limitations that shall be considered for future research. In this thesis, financial risk is evaluated at the account level rather than the individual level. Consumers owning multiple accounts might have over or under-estimated risks due to the inability to simultaneously capture financial behaviour in the other accounts. Hence, the presented frameworks could be extended to individual-level consolidated bank account data. Using solely OB current account data for personal financial risk assessment lose a holistic view of financial lives due to missing information on financial asset ownership, records of all existing credit commitments, investments etc. Potential future work shall complement this information with OB data for a more accurate risk evaluation.

The insights presented in this thesis are derived from aggregated data, compiled quarterly or annually. This aggregation process results in the loss of detailed transactional information, missing real-time trend patterns and seasonal effects. Future work shall focus on modelling daily granular data and incorporating seasonal checks to enable more timely and accurate risk assessments.

This thesis works on account transactions that span over twelve months. The risk projection in Chapters 2 and 3 only allows short-term quarterly prediction without the ability to estimate long-term behaviour to forecast further time points. If OB data is available for a longer duration (up to years), extending the volatility forecast model (Chapter 2) to track long-term trends and enabling out-of-time risk monitoring (Chapter 3) could be the potential future research direction. The KYC risk assessment framework in Chapter 4 is completely unsupervised without expert validation of the outcomes i.e., suspicious accounts. This KYC methodology can be extended with a semi-supervised active learning approach to include expert labels for corrections on mistaken outcomes so that KYC screening can proceed with only a low number of labels and gradually improve over time as more accounts are assessed. The KYC framework is also restricted to a static evaluation of financial risk exposure from the 12-month time frame. Dynamic tracking of the incoming data stream could be a potential extension for the timely monitoring of anomalous transactions.

References

Abbasimehr, H., & Baghery, F. S. (2022). A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis. *Expert Systems with Applications*, *204*, 117584.

Abbasimehr, H., & Bahrini, A. (2022). An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation. *Expert Systems with Applications*, *192*, 116373.

Abbasimehr, H., & Shabani, M. (2021). A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers. *Kybernetes*, *50*(2), 221–242.

Aggarwal, C. C., et al. (2015). *Data mining: the textbook* (Vol. 1). Springer.

Aggarwal Charu, C., & Reddy Chandan, K. (2013). *Data clustering: algorithms and applications.* CRC press.

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information systems*, *53*, 16–38.

Aghabozorgi, S., & Wah, T. Y. (2014). Effective clustering of time-series data using fcm. *Int. J. Mach. Learn. Comput*, *4*(2), 170–176.

Ala'raj, M., Abbod, M. F., Majdalawieh, M., & Jum'a, L. (2022). A deep learning model for behavioural credit scoring in banks. *Neural Computing and Applications*, 1–28. doi: 10.1007/s00521-021-06695-z

Albacete, N., & Lindner, P. (2013). Household vulnerability in austria–a microeconomic analysis based on the household finance and consumption survey. *Financial stability report*, *25*, 57–73.

Alboukaey, N., Joukhadar, A., & Ghneim, N. (2020). Dynamic behavior based churn prediction in mobile telecom. *Expert Systems with Applications*, *162*, 113779. doi: 10.1016/j.eswa.2020.113779

Aliyev, M., Ahmadov, E., Gadirli, H., Mammadova, A., & Alasgarov, E. (2020). Segmenting bank customers via rfm model and unsupervised machine learning. *arXiv preprint arXiv:2008.08662*.

Al Mamun, A., Hasan, S. R., Bhuiyan, M. S., Kaiser, M. S., & Yousuf, M. A. (2020). Secure and transparent kyc for banking system using ipfs and blockchain technology. In *2020 ieee region 10 symposium (tensymp)* (pp. 348–351).

Amarbayasgalan, T., Jargalsaikhan, B., & Ryu, K. H. (2018). Unsupervised novelty detection using deep autoencoders with density based clustering. *Applied Sciences*, *8*(9), 1468.

Ampountolas, A., Nyarko Nde, T., Constantinescu, C., et al. (2021). A machine learning approach for micro-credit scoring. *Risks*, *9*(3), 50.

Apergis, N., Baruník, J., & Lau, M. C. K. (2017). Good volatility, bad volatility: What drives the asymmetric connectedness of australian electricity markets? *Energy Economics*, *66*, 108–115.

Aspen Financial Security Program, A. I. (2016). *Income volatility a primer* (Tech. Rep.). Retrieved from `https://www.aspeninstitute.org/wp-content/uploads/files/content/docs/pubs/EPIC+Volatility+Primer+(May).pdf`

Baesens, B., Höppner, S., & Verdonck, T. (2021). Data engineering for fraud detection. *Decision Support Systems*, *150*, 113492. doi: 10.1016/j.dss.2021.113492

Bakoben, M., Bellotti, T., & Adams, N. (2020). Identification of credit risk based on cluster analysis of account behaviours. *Journal of the Operational Research Society*, *71*(5), 775–783.

Bania, N., & Leete, L. (2009). Monthly household income volatility in the us, 1991/92 vs. 2002/03. *Economics Bulletin*, *29*(3), 2100–2112.

Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. *arXiv preprint arXiv:2003.05991*.

Barkhordar, E., Shirali-Shahreza, M. H., & Sadeghi, H. R. (2021). Clustering of bank customers using lstm-based encoder-decoder and dynamic time warping. *arXiv preprint arXiv:2110.11769*.

Bartels, C. (2022). Cluster analysis for customer segmentation with open banking data. In *2022 3rd asia service sciences and software engineering conference* (pp. 87–94).

Bartolucci, F., & Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent markov heterogeneity structure. *Journal of the American Statistical Association*, *104*(486), 816–831.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2012). *Latent markov models for longitudinal data.* CRC Press.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Latent markov models: a review of a general framework for the analysis of longitudinal data with covariates. *Test*, *23*, 433–465.

Bartolucci, F., Montanari, G. E., & Pandolfi, S. (2015). Three-step estimation of latent markov models with covariates. *Computational Statistics & Data Analysis*, *83*, 287–301.

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). Lmest: An r package for latent markov models for longitudinal categorical data. *Journal of Statistical Software*, *81*, 1–38.

Barunik, J., Krehlik, T., & Vacha, L. (2016). Modeling and forecasting exchange rate volatility in time-frequency domain. *European Journal of Operational Research*, *251*(1), 329–340.

Bastos, J. A., & Matos, S. M. (2022). Explainable models of credit losses. *European Journal of Operational Research*, *301*(1), 386–394.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, *41*(1), 164–171.

Belth, C., Zheng, X., & Koutra, D. (2020). Mining persistent activity in continually evolving networks. In *Proceedings of the 26th acm sigkdd international conference on knowledge discovery & data mining* (pp. 934–944).

Bester, T., & Rosman, B. (2024). Towards financially inclusive credit products through financial time series clustering. *arXiv preprint arXiv:2402.11066*.

Bharati, H., Bastikar, S., Gavade, M., & Samota, S. (2018). Credit card fraud detection using hmm and dbscan. *International Journal of Science and Research (IJSR)*, *7*(10), 1086–1090.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*(3), 602–613. doi: 10.1016/j.dss.2010.08.008

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford university press.

Blümke, O. (2022). A structural hidden markov model for forecasting scenario probabilities for portfolio loan loss provisions. *Knowledge-Based Systems*, *249*, 108934.

Bocchio, C., Crook, J., & Andreeva, G. (2022). The impact of macroeconomic scenarios on recurrent delinquency: A stress testing framework of multi-state models for mortgages. *International Journal of Forecasting*.

Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). Machine learning explainability in finance: an application to default risk analysis. *Bank of England Working Paper*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Breiman, L. (2017). *Classification and regression trees.* Routledge.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 acm sigmod international conference on management of data* (pp. 93–104).

Brodsky, L., & Oakes, L. (2017). Data sharing and open banking. *McKinsey & Company*, *1105*.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, *57*, 203–216.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1–27.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications.* Cambridge university press.

Cao, D. K., & Do, P. (2012). Applying data mining in money laundering detection for the vietnamese banking industry. In *Intelligent information and database systems: 4th asian conference, aciids 2012, kaohsiung, taiwan* (pp. 207–216).

Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, *557*, 317–331. doi: 10.1016/j.dss.2022.113798

Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, *95*, 91–101. doi: 10.1016/j.dss.2017.01.002

Chang, H. C., & Tsai, H. P. (2011). Group rfm analysis as a novel framework to discover better customer consumption behavior. *Expert Systems With Applications*, *38*(12), 14499–14513. doi: 10.1016/j.eswa.2011.05.034

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... Zhou, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1–4.

CMA. (2019). *Consumer vulnerability: Challenges and potential solutions.* Competition and Market Authority London.

Colladon, A. F., & Remondi, E. (2017). Using social network analysis to prevent money laundering. *Expert Systems with Applications*, *67*, 49–58. doi: 10.1016/j.eswa.2016.09.029

Collard, S. (2011). Understanding financial difficulty: Exploring the opportunities for early intervention. *Barclays and Money Advice Trust*.

Consumer Financial Protection Bureau, C. (2017). Cfpb financial well-being scale: scale development technical report. *Washington, DC: CFPB*.

Consumer Financial Protection Bureau, C. (2023). Overdraft and nonsufficient fund fees. *Washington, DC: CFPB*.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157–175.

Dai, H., Shahzad, M., Liu, A. X., & Zhong, Y. (2016). Finding persistent items in data streams. *Proceedings of the VLDB Endowment*, *10*(4), 289–300.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, *PAMI-1*(2), 224–227.

Decker, R., Scholz, S. W., & Wagner, R. (2006). Growing clustering algorithms in market segmentation: defining target groups and related marketing communication. In *Data analysis, classification and the forward search: Proceedings of the meeting of the classification and data analysis group (cladag) of the italian statistical society, university of parma, june 6–8, 2005* (pp. 23–30).

Demetis, D. S. (2018). Fighting money laundering with technology: A case study of bank x in the uk. *Decision Support Systems*, *105*, 96–107. doi: 10.1016/j.dss.2017.11.005

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, *39*(1), 1–22.

Djeundje, V. B., & Crook, J. (2018). Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards. *European Journal of Operational Research*, *271*(2), 697–709.

Djeundje, V. B., & Crook, J. (2019a). Dynamic survival models with varying coefficients for credit risks. *European Journal of Operational Research*, *275*(1), 319–333.

Djeundje, V. B., & Crook, J. (2019b). Identifying hidden patterns in credit risk survival data using generalised additive models. *European Journal of Operational Research*, *277*(1), 366–376.

Doyle, M. D., Sharma, R., & Ross, C. (2023). *How to flourish in an uncertain future Open banking and PSD2* (Tech. Rep.). Retrieved from `https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/financial-services/deloitte-uk-open-banking-how-to-flourish-in-an-uncertain-future.pdf`

Dreżewski, R., Sepielak, J., & Filipkowski, W. (2012). System supporting money laundering detection. *Digital Investigation*, *9*(1), 8–21. doi: 10.1016/j.diin.2012.04.003

Dreżewski, R., Sepielak, J., & Filipkowski, W. (2015). The application of social network analysis algorithms in a system supporting money laundering detection. *Information Sciences*, *295*, 18–32. doi: 10.1016/j.ins.2014.10.015

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of financial economics*, *83*(3), 635–665.

Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative finance*, *1*(2), 237.

Equifax. (2021). *Financial health index* (Tech. Rep.). Retrieved from `https://assets.equifax.com/assets/unitedkingdom/financial-health-index-open-banking.pdf`

Equifax. (2022). *Equifax financial durability index* (Tech. Rep.). Retrieved from `https://assets.equifax.com/marketing/US/assets/financial-durability-index.pdf`

Ernst & Yong, E. (2018). *New definition of default* (Tech. Rep.). Retrieved from https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/emeia-financial-services/ey-new-definition-of-default.pdf

Escobar-Anel, M., Rastegari, J., & Stentoft, L. (2021). Option pricing with conditional garch models. *European Journal of Operational Research*, *289*(1), 350–363.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). Density-based spatial clustering of applications with noise. In *Int. conf. knowledge discovery and data mining* (Vol. 240).

Experian. (2019). *Helping Identify and Pre-empt Financial Vulnerability* (Tech. Rep.). Retrieved from https://www.experian.co.uk/assets/ebook/financial-vulnerability-v2.pdf

Experian. (2020a). *Experian Affordability* (Tech. Rep.). Retrieved from https://www.experian.co.uk/content/dam/marketing/uki/uk/en/pdf/affordability-check/Experian-Affordability-Guide.pdf

Experian. (2020b). *Experian Affordability: Affordability Passport* (Tech. Rep.). Retrieved from https://www.experian.co.uk/content/dam/marketing/uki/uk/en/pdf/affordability-check/Experian-Affordability-Passport.pdf

Farajian, M. A., & Mohammadi, S. (2010). Mining the banking customer behavior using clustering and association rules methods. *International Journal of Industriial Engineering & Production Research*.

Farrell, D., & Greig, F. (2015). Weathering volatility: big data on the financial ups and downs of us individuals. *New York: JP Morgan Chase Institute*.

Farrell, D., & Greig, F. (2016). Paychecks, paydays, and the online platform economy. In *Proceedings. annual conference on taxation and minutes of the annual meeting of the national tax association* (Vol. 109, pp. 1–40).

Federal Bureau of Investigation, F. (2011). *Financial Crimes Report to the Public* (Tech. Rep.). Retrieved from https://www.fbi.gov/file-repository/stats-services-publicationsfinancial-crimes-report-2010-2011-financial-crimes-report-2010-2011.pdf/view.

Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. In *Proceedings of the international aaai conference on web and social media* (pp. 175–184).

Ferretti, C., Gabbi, G., Ganugi, P., Sist, F., & Vozzella, P. (2019). Credit risk migration and economic cycles. *Risks*, *7*(4), 109.

Financial Conduct Authority, F. (2015). *Consumer vulnerability* (Tech. Rep.). Retrieved from https://www.fca.org.uk/publication/occasional-papers/occasional-paper-8.pdf

Financial Conduct Authority, F. (2018). *FCA Mission: Approach to Consumers* (Tech. Rep.). Retrieved from https://www.fca.org.uk/publication/corporate/approach-to-consumers.pdf

Financial Conduct Authority, F. (2021a). Financial lives 2020 survey: the impact of coronavirus. *Financial Conduct Authority*, *18*.

Financial Conduct Authority, F. (2021b). *Guidance for firms on the fair treatment of vulnerable customers* (Tech. Rep.). Retrieved from https://www.fca.org.uk/publication/finalised-guidance/fg21-1.pdf

Financial Conduct Authority, F. (2023a). *Financial Crime Guide: A firm's guide to countering financial crime risks (FCG)* (Tech. Rep.). Retrieved from https://www.handbook.fca.org.uk/handbook/FCG.pdf

Financial Conduct Authority, F. (2023b). *Financial Lives January 2023: Consumer experience of the rising cost of living – the burden of bills and ways to get support* (Tech. Rep.). Retrieved from https://www.fca.org.uk/publications/financial-lives/financial-lives-january-2023-consumer-experience

Finlay, S. M. (2006). Predictive models of expenditure and over-indebtedness for assessing the affordability of new consumer credit applications. *Journal of the Operational Research Society*, *57*(6), 655–669.

Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: From network event correlation to incident detection. *Computers & Security*, *48*, 35–57.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Frydman, H., & Matuszyk, A. (2018). Estimation and status prediction in a discrete mover-stayer model with covariate effects on stayer's probability. *Applied Stochastic Models in Business and Industry*, *34*(2), 196–205.

Frydman, H., Matuszyk, A., Li, C., & Zhu, W. (2019). Mover-stayer model with covariate effects on stayer's probability and mover's transitions. *Applied Stochastic Models in Business and Industry*, *35*(5), 1171–1184.

Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.

Gathergood, J., & Guttman-Kenney, B. (2016). Can we predict which consumer credit users will suffer financial distress? *FCA Occasional Paper*(20).

Ghaharian, K., Abarbanel, B., Kraus, S. W., Singh, A., & Bernhard, B. (2023). Players gonna pay: Characterizing gamblers and gambling-related harm with payments transaction data. *Computers in Human Behavior*, *143*, 107717.

Gill, M., & Taylor, G. (2004). Preventing money laundering or obstructing business? financial companies' perspectives on 'know your customer' procedures. *British Journal of Criminology*, *44*(4), 582–594.

Giordani, P., Jacobson, T., Von Schedvin, E., & Villani, M. (2014). Taking the twists into account: Predicting firm bankruptcy risk with splines of financial ratios. *Journal of Financial and Quantitative analysis*, *49*(4), 1071–1099.

Giroire, F., Chandrashekar, J., Taft, N., Schooler, E., & Papagiannaki, D. (2009). Exploiting temporal persistence to detect covert botnet channels. In *International workshop on recent advances in intrusion detection* (pp. 326–345).

Gladstone, J., Nieboer, J., & Raghavan, K. (2020). *Understanding consumer financial wellbeing through banking data* (Tech. Rep.). Retrieved from `https://www.fca.org.uk/publication/occasional-papers/occasional-paper-58.pdf`

Gozman, D., Hedman, J., & Olsen, K. S. (2018). Open banking: Emergent roles, risks & opportunities. *Research Paper*.

Granger, C. W., & Poon, S.-H. (2001). Forecasting financial market volatility: A review. *Available at SSRN 268866*.

Grennepois, N., & Robin, E. (2019). Explain artificial intelligence for credit risk management. *Deloitte Risk Advisory, juillet*.

Gunnarsson, B. R., Vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: Do or don't? *European Journal of Operational Research*, *295*(1), 292–305. doi: 10.1016/j.ejor.2021.03.006

Hannagan, A., & Morduch, J. (2015). Income gains and month-to-month income volatility: Household evidence from the us financial diaries. *NYU Wagner research paper*(2659883).

Harrison, T., & Andreeva, G. (2021). *Financial Health of NHS Workers* (Tech. Rep.). University of Edinburgh Business School. Retrieved from `https://www.imf.org/-/media/Websites/IMF/imported-flagship-issues/external/pubs/ft/GFSR/2003/02/pdf/_chp3pdf.ashx`

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

He, Z., Xu, X., & Deng, S. (2003). Discovering cluster-based local outliers. *Pattern recognition letters*, *24*(9-10), 1641–1650.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud:: A review of anomaly detection techniques and recent advances. *Expert Systems With Applications*, *193*, 116429. doi: 10.1016/j.eswa.2021.116429

Hinneburg, A. (1996). A density based algorithm for discovering clusters in large spatial databases with noise. In *Kdd conference, 1996* (p. 226-231).

Hinton, G. E. (2009). Deep belief networks. *Scholarpedia*, *4*(5), 5947.

Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. *Neural Networks: Tricks of the Trade: Second Edition*, 599–619.

Hjelkrem, L. O., De Lange, P. E., & Nesset, E. (2022). The value of open banking data for application credit scoring: case study of a norwegian bank. *Journal of Risk and Financial Management*, *15*(12), 597.

Holder, C., Middlehurst, M., & Bagnall, A. (2024). A review and evaluation of elastic distance functions for time series clustering. *Knowledge and Information Systems*, *66*(2), 765–809.

Hossain, M. M., Sebestyen, M., Mayank, D., Ardakanian, O., & Khazaei, H. (2020). Large-scale data-driven segmentation of banking customers. In *2020 ieee international conference on big data (big data)* (pp. 4392–4401).

Huang, D., Zhou, J., & Wang, H. (2018). Rfms method for credit scoring based on bank card transaction data. *Statistica Sinica*, *28*(4), 2903–2919.

International Monetary Fund, I. (2003). *Global Financial Stability Report Chapter III Financial asset price volatility: a source of instability?* (Tech. Rep.). Retrieved from `https://www.imf.org/-/media/Websites/IMF/imported-flagship-issues/external/pubs/ft/GFSR/2003/02/pdf/_chp3pdf.ashx`

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Johnson, W. B. (1984). Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, *26*, 189–206.

Kapsoulis, N., Psychas, A., Palaiokrassas, G., Marinakis, A., Litke, A., & Varvarigou, T. (2020). Know your customer (kyc) implementation with smart contracts on a privacy-oriented decentralized architecture. *Future Internet*, *12*(2), 41. doi: 10.3390/fi12020041

Kargari, M., & Eshghi, A. (2018). A model based on clustering and association rules for detection of fraud in banking transactions. In *Proceedings of the 4th world congress on electrical engineering and computer systems and sciences* (pp. 21–23).

Kassab, M., & Laplante, P. A. (2022). Open banking: What it is, where it's at, and where it's going. *Computer*, *55*(1), 53–63.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kelly, R., & O'Malley, T. (2016). The good, the bad and the impaired: A credit risk model of the irish mortgage market. *Journal of Financial Stability*, *22*, 1–9.

Keys, B. J., Mahoney, N., & Yang, H. (2023). What determines consumer financial distress? place-and person-based factors. *The Review of Financial Studies*, *36*(1), 42–69.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787.

Kian, R., & Obaid, H. S. (2022). Detection of fraud in banking transactions using big data clustering technique customer behavior indicators. *Journal of applied research on industrial engineering*, *9*(3), 264–273.

Kim, S. D., Andreeva, G., & Rovatsos, M. (2023). The double-edged sword of big data and information technology for the disadvantaged: A cautionary tale from open banking. *arXiv preprint arXiv:2307.13408*.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, *15*(4), 143–156.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1461–1480.

Kriebel, J., & Stitz, L. (2022). Credit default prediction from user-generated text in peer-to-peer lending using deep learning. *European Journal of Operational Research*, *302*(1), 309–323. doi: 10.1016/j.ejor.2021.12.024

Lando, D., & Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of banking & finance*, *26*(2-3), 423–444.

Larik, A. S., & Haider, S. (2011). Clustering based anomalous transaction reporting. *Procedia Computer Science*, *3*, 606–610. doi: 10.1016/j.procs.2010.12.101

Le Khac, N. A., & Kechadi, M. T. (2010). Application of data mining for anti-money laundering detection: A case study. In *2010 ieee international conference on data mining workshops* (pp. 577–584).

Leow, M., & Crook, J. (2014). Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, *236*(2), 685–694.

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. doi: 10.1016/j.ejor.2015.05.030

Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., & Shao, J. (2016). Modeling review spam using temporal patterns and co-bursting behaviors. *arXiv preprint arXiv:1611.06625*.

Lim, K.-L., Jiang, X., & Yi, C. (2020). Deep clustering with variational autoencoder. *IEEE Signal Processing Letters*, *27*, 231–235.

Lin, J. T., Bumcrot, C., Mottola, G., Valdes, O., Ganem, R., Kieffer, C., ... Walsh, G. (2021). *Financial Capability in the United States: Highlights from the FINRA Foundation National Financial Capability Study (5th Edition)* (Tech. Rep.). Retrieved from `www.FINRAFoundation.org/NFCSReport2021`

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *Eighth ieee international conference on data mining* (pp. 413–422).

Luna, D. K., Palshikar, G. K., Apte, M., & Bhattacharya, A. (2018). Finding shell company accounts using anomaly detection. In *Proceedings of the acm india joint international conference on data science and management of data* (pp. 167–174).

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Luo, S., Kong, X., & Nie, T. (2016). Spline based survival model for credit risk modeling. *European Journal of Operational Research*, *253*(3), 869–879.

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th berkeley symp. math. statist. probability* (pp. 281–297).

Malik, M., & Thomas, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, *28*(1), 261–272.

Medina-Olivares, V., & Calabrese, R. (2023). Detecting consumers' financial vulnerability using open banking data: Evidence from uk payday loans. *arXiv preprint arXiv:2306.01749*.

Mitchell, D. S. (2017). *Stable and predictable scheduling as antidote to income volatility* (Tech. Rep.). Retrieved from `https://www.aspeninstitute.org/wp-content/uploads/2017/02/Predictable-Schedling.pdf`

Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, *165*, 113986.

Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability*, *11*(3), 699.

Ng, A. (2011). Sparse autoencoder. *CS294A Lecture notes*, *72*(2011), 1–19.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), 559–569. doi: 10.1016/j.dss.2010.08.006

Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *The Journal of Finance and Data Science*, *2*(1), 58–75. doi: 10.1016/j.jfds.2016.03.001

Noerhidajati, S., Purwoko, A. B., Werdaningtyas, H., Kamil, A. I., & Dartanto, T. (2021). Household financial vulnerability in indonesia: Measurement and determinants. *Economic Modelling*, *96*, 433–444.

Nofal, S. (2024). Identifying highly-valued bank customers with current accounts based on the frequency and amount of transactions. *Heliyon*.

O'Connor, G. E., Newmeyer, C. E., Wong, N. Y. C., Bayuk, J. B., Cook, L. A., Komarova, Y., ... Warmath, D. (2019). Conceptualizing the multiple dimensions of consumer financial vulnerability. *Journal of Business Research*, *100*, 421–430.

Olson, D. L., & Chae, B. K. (2012). Direct marketing decision support through predictive customer response modeling. *Decision Support Systems*, *54*(1), 443–451. doi: 10.1016/j.dss.2012.06.005

Oluwafolake, A., & Solomon, O. A. (2017). A multi-algorithm data mining classification approach for bank fraudulent transactions. *African Journal of Mathematics and Computer Science Research*, *10*(1), 5–13.

Open Banking Implementation Entity, O. (2023a). *From Open Banking to Open Finance and Beyond* (Tech. Rep.). Retrieved from `https://openbanking.foleon.com/live-publications/openbanking-to-openfinance/`

Open Banking Implementation Entity, O. (2023b). *THE OPEN BANKING IMPACT REPORT: MARCH 2023* (Tech. Rep.). Retrieved from `https://www.openbanking.org.uk/wp-content/uploads/Open-Banking-Impact-Report-4-30-March-2023.pdf`

Paula, E. L., Ladeira, M., Carvalho, R. N., & Marzagão, T. (2016). Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering. In *Ieee international conference on machine learning and applications (icmla)* (pp. 954–960).

Peetz, J., Robson, J., & Xuereb, S. (2021). The role of income volatility and perceived locus of control in financial planning decisions. *Frontiers in Psychology*, *12*, 638043.

Podder, S. (2022). Leveraging the provisions of open banking to fight financial crimes. In *Financial technology and the law: Combating financial crime* (pp. 19–46).

Poon, S.-H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, *41*(2), 478–539.

Prautzsch, H., Boehm, W., & Paluszny, M. (2002). *Bézier and b-spline techniques* (Vol. 6). Springer.

Racine, J. S. (2014). *A primer on regression splines* (Tech. Rep.). Retrieved from `http://cranrprojectorg/web/packages/crs/vignettes/splineprimerpdf`

Reijerink, J. (2024). *Distinguishing payment user groups by clustering transaction data* (Tech. Rep.). DeNederlanscheBank. Retrieved from `https://www.dnb.nl/media/wkgoelxg/dnb-analysis-distinguishing-payment-user-groups-by-clustering-transaction-data.pdf`

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65.

Ryu, S., Choi, H., Lee, H., & Kim, H. (2019). Convolutional autoencoder based feature extraction and clustering for customer load analysis. *IEEE Transactions on Power Systems*, *35*(2), 1048–1060.

Santos, J. P. N. (2018). *Credit risk modelling using multi-state markov models* (Unpublished doctoral dissertation). Universidade Nova de Lisboa.

Savage, D., Wang, Q., Chou, P., Zhang, X., & Yu, X. (2016). Detection of money laundering groups using supervised learning in networks. *arXiv preprint arXiv:1608.00708*.

Schechtman, R. (2013). Default matrices: A complete measurement of banks' consumer credit delinquency. *Journal of Financial Stability*, *9*(4), 460–474.

Schuermann, T., Jafry, Y., et al. (2003). Measurement and estimation of credit migration matrices. *Center for Financial Institutions Working Papers*, *3*(8).

Segovia-Vargas, M.-J., et al. (2022). Detection of shell companies in financial institutions using dynamic social network. *Expert Systems with Applications*, *207*, 117981.

Singh, A., Rumantir, G., South, A., & Bethwaite, B. (2014). Clustering experiments on big transaction data for market segmentation. In *Proceedings of the 2014 international conference on big data science and computing* (pp. 1–7).

Song, C., Liu, F., Huang, Y., Wang, L., & Tan, T. (2013). Auto-encoder based data clustering. In *Progress in pattern recognition, image analysis, computer vision, and applications: 18th iberoamerican congress, ciarp 2013, havana, cuba, november 20-23, 2013, proceedings, part i 18* (pp. 117–124).

Stefanelli, V., & Manta, F. (2022). Digital financial services and open banking innovation: Are banks becoming 'invisible'? *Global Business Review*, 09721509231151491.

Stripling, E., Baesens, B., Chizi, B., & Vanden Broucke, S. (2018). Isolation-based conditional anomaly detection on mixed-attribute data to uncover workers' compensation fraud. *Decision Support Systems*, *111*, 13–26. doi: 10.1016/j.dss.2018.04.001

Šubová, N., Mura, L., & Buleca, J. (2021). Determinants of household financial vulnerability: Evidence from selected eu countries. *E&M Economics and Management*, *24*(3), 186–207.

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, *39*(1), 43–62.

Tavakoli, N., Siami-Namini, S., Adl Khanghah, M., Mirza Soltani, F., & Siami Namin, A. (2020). An autoencoder-based deep learning approach for clustering time series data. *SN Applied Sciences*, *2*, 1–25.

Tescher, J., & Schneider, R. (2015). The real financial lives of americans. *What it's worth: Strengthening the financial future of families, communities and the nation*, 53–67.

Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications.* Society for Industrial and Applied Mathematics.

Thompson, J. R., Feng, L., Reesor, R. M., & Grace, C. (2021). Know your clients' behaviours: a cluster analysis of financial transactions. *Journal of Risk and Financial Management*, *14*(2), 50. doi: 10.3390/jrfm14020050

Tiukhova, E., Reusens, M., Baesens, B., & Snoeck, M. (2022). Benchmarking conventional outlier detection methods. In *Research challenges in information science: 16th international conference, rcis 2022, barcelona, spain, may 17–20, 2022, proceedings* (pp. 597–613).

Trusts, P. C. (2017). *How income volatility interacts with american families' financial security: An examination of gains, losses, and household economic experiences.*

Tuma, M. N., Decker, R., & Scholz, S. W. (2011). A survey of the challenges and pifalls of cluster analysis application in market segmentation. *International Journal of Market Research*, *53*(3), 391–414.

Upadhya, V., & Sastry, P. (2019). An overview of restricted boltzmann machines. *Journal of the Indian Institute of Science*, *99*, 225–236.

Van Belle, R., Baesens, B., & De Weerdt, J. (2023). Catchm: A novel network-based credit card fraud detection method using node representation learning. *Decision Support Systems*, *164*, 113866. doi: 10.1016/j.dss.2022.113866

Vandervorst, F., Verbeke, W., & Verdonck, T. (2022). Data misrepresentation detection for insurance underwriting fraud prevention. *Decision Support Systems*, *159*, 113798. doi: 10.1016/j.dss.2022.113798

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, *75*, 38–48. doi: 10.1016/j.dss.2015.04.013

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, *13*(2), 260–269.

West, D. (2000). Neural network credit scoring models. *Computers & operations research*, *27*(11-12), 1131–1152. doi: 10.1016/S0305-0548(99)00149-5

Weston, D. J., Hand, D. J., Adams, N. M., Whitrow, C., & Juszczak, P. (2008). Plastic card fraud detection using peer group analysis. *Advances in Data Analysis and Classification*, *2*, 45–62.

Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data mining and knowledge discovery*, *18*, 30–55.

Wiggins, L. M. (1973). *Panel analysis: Latent probability models for attitude and behavior processes.* Jossey-Bass.

Xiao, L., & Aydemir, A. (2007). Volatility modelling and forecasting in finance. In *Forecasting volatility in the financial markets* (pp. 1–45). Elsevier.

Xiao, Q., Qiao, Y., Zhen, M., & Chen, S. (2014). Estimating the persistent spreads in high-speed networks. In *2014 ieee 22nd international conference on network protocols* (pp. 131–142).

Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012). Review spam detection via temporal pattern discovery. In *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining* (pp. 823–831).

Yadav, P., & Chandak, R. (2019). Transforming the know your customer (kyc) process using blockchain. In *2019 international conference on advances in computing, communication and control (icac3)* (pp. 1–5).

Yu, F.-H., Lu, J., Gu, J.-W., & Ching, W.-K. (2019). Modeling credit risk with hidden markov default intensity. *Computational Economics*, *54*, 1213–1229.

Zamini, M., & Montazer, G. (2018). Credit card fraud detection using autoencoder based clustering. In *2018 9th international symposium on telecommunications (ist)* (pp. 486–491).

Zhang, B., & Qian, J. (2021). Autoencoder-based unsupervised clustering and hashing. *Applied Intelligence*, *51*(1), 493–505.

Zhang, Y., Bradlow, E. T., & Small, D. S. (2015). Predicting customer value using clumpiness: From rfm to rfmc. *Marketing science*, *34*(2), 195–208. doi: 10.1287/mksc.2014.0873

Zhao, Y., Hu, X., Cheng, C., Wang, C., Wan, C., Wang, W., . . . others (2021). Suod: Accelerating large-scale unsupervised heterogeneous outlier detection. In *Proceedings of machine learning and systems* (pp. 463–478).

Zhu, Y., Deng, Q., Huang, D., Jing, B., & Zhang, B. (2021). Clustering based on kolmogorov–smirnov statistic with application to bank card transaction data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *70*(3), 558–578.

A Descriptive statistics for category variable

Category	Raw data A Number of transactions	Raw data B Number of transactions
Accommodation	17049	15738
Beauty treatments	26716	16777
Cash	270941	397915
Charitable giving	9084	9439
Checks payments	2125	4147
Child tax credit	63727	12815
Childcare	5544	6354
Clothing	86499	256944
Convenience store	275076	381117
Credit card	56980	69903
Debt recovery	18020	10770
Deposit	1774074	1508541
Disability	2764	1296
Discount store	63387	77160
Eating out	892359	2176350
Electricity and gas	59961	90746
Electronics	12165	3526
Expenses	8649	436
Fines and penalties	602	0
Flight	3807	3
Gambling and gaming	225408	202419
Gifts	12591	279
Gym and fitness	7227	17223
Healthcare and medical	50585	68705
Holidays	9156	6959
Home improvement	86771	122906
Insurance	79024	151780
Interest	113479	42738
Investments	851	855
Legal expenses	689	552
Leisure and entertainment	269330	336437
Loan	97199	93766
Mortgage payment	4668	1253
Office expenses	19	0
Online retail	440145	498304
Other	91116	45263
Other benefits	43782	23497
Other transport	104433	148964
Other utilities	19427	29507
Overdraft charges	13508	13093
Pension	2242	379
Pet care	11676	10041
Postage and courier	27308	24746
Property and council tax	24982	37115
Public transport	74780	64083
Refunds and adjustments	76197	5338
Rent payment	30819	32594
Repairs and maintenance	940	20
Retail	192478	183921
Rewards	780	0
Salary	29084	45542
Savings	74108	96968
School fees	477	0
Service charges and fees	68768	120157
Services	26076	4273

Shoes	7106	21052
Subscriptions and renewals	1785	9
Supermarket	892775	1601061
Taxes	11358	0
Telephone and TV	197526	232016
Train	16215	57989
Training	318	3
Transfer payments	949590	599762
Uncategorised	1513358	9273784
Unemployment benefit	1334	278
Union membership	2366	854
University fees	62	0
Vehicle expenses	83716	80940
Vehicle fuel	315740	512826
Vehicle loan	5303	2225

B List of covariates

Table B.1: List of covariates

Covariate	Short name	Description
		All covariates are calculated as proportion except for X93 and X94 Depending on how the covariate's information is aggregated (in bold), each covariate below is divided by either one of the following: **total number of days (in six months) / total number of months (six months)/ total number of three-month rolling windows (four windows in six months)/ total amount spent on full expenses (in six months)**
(1) Balance		
(a) overall summary		
day balance		
X1	in negative balance	**number of days** in negative balance
X2	consecutive days in same balance amount	**number of days** with consecutive same balance amount
X3	move towards a worse negative balance	**number of days** moving towards a worse negative balance
X4	extreme high or low balance amount	**number of days** the balance amount (lower/higher) than three sigma from the mean
X5	transition from positive to negative balance	**number of days** a positive balance moves to a negative
month balance		
X6	month inflow less than month outflow	**number of months** total inflow less than total outflow amount
X7	month end balance less than month inflow	**number of months** the end balance less than total inflow amount
X8	month end balance less than month open balance	**number of months** the end balance less than opening balance
distress situation		
X9	in negative balance at least once	**number of months** the account has been in negative balance at least once
X10	negative month end balance	**number of months** the account has negative month end balance
X11	larger negative month end balance than a negative month opening balance	**number of months** the account has a larger negative month end balance than a negative opening balance
X12	negative month end balance where more than half of the month is in negative	**number of months** the account has negative month end balance where more than half of the month is in negative
(b) decreasing trend		
X13	decrease in month largest negative balance amount	**number of months** the current month's largest negative balance is lower than the previous month
X14	continuous decrease in month largest negative balance amount	**number of three-month rolling windows** where the monthly largest negative balance has a negative linear regression slope
X15	decrease in minimum balance amount	**number of months** the current month's minimum balance amount is lower than the previous month
X16	continuous decrease in minimum balance amount	**number of three-month rolling windows** where the monthly minimum balance amount has a negative linear regression slope

(2) Inflow		
(a) overall summary		
X17	inflow lower than usual	**number of months** the inflow amount is lower than the mean
X18	extreme high or low inflow amount	**number of months** the inflow amount (lower/higher) than three sigma from the mean
X19	days with inflow transactions	**number of days** with inflow transactions
(b) decreasing trend		
X20	decrease in inflow amount	**number of months** the current month's total inflow amount is lower than the previous month
X21	continuous decrease in inflow amount	**number of three-month rolling windows** where the monthly total inflow amount has a negative linear regression slope
X22	decrease in number of inflow transactions	**number of months** the current month's total number of inflow transactions is lower than the previous month
X23	continuous decrease in number of inflow transactions	**number of three-month rolling windows** where the monthly total number of inflow transactions has a negative linear regression slope
X24	decrease in days with inflow	**number of months** the current month's total days with inflow transactions is lower than the previous month
X25	continuous decrease in days with inflow	**number of three-month rolling windows** where the monthly total days with inflow transactions has a negative linear regression slope
(3) Outflow		
(a) overall summary		
X26	outflow higher than usual	**number of months** the outflow amount is higher than the mean
X27	extreme high or low outflow amount	**number of months** the outflow amount (lower/higher) than three sigma from the mean
X28	days with outflow transactions	**number of days** with outflow transactions
(b) decreasing trend		
X29	decrease in outflow amount	**number of months** the current month's total outflow amount is lower than the previous month
X30	continuous decrease in outflow amount	**number of three-month rolling windows** where the monthly total outflow amount has a negative linear regression slope
X31	decrease in number of outflow transactions	**number of months** the current month's total number of outflow transactions is lower than the previous month
X32	continuous decrease in number of outflow transactions	**number of three-month rolling windows** where the monthly total number of outflow transactions has a negative linear regression slope
X33	decrease in days with outflow	**number of months** the current month's total days with outflow transactions is lower than the previous month
X34	continuous decrease in days with outflow	**number of three-month rolling windows** where the monthly total days with outflow transactions has a negative linear regression slope
(4) Spending categories		
(a) overall summary		
fixed costs		
X35	spending on rent payment	**amount spent** on rent payment
X36	spending on telephone and tv	**amount spent** on telephone and tv

X37	spending on insurance	**amount spent** on insurance
X38	spending on electricity and gas	**amount spent** on electricity and gas
X39	spending on property and council tax	**amount spent** on property and council tax
X40	spending on subscriptions and renewals	amount spent on subscriptions and renewals
living costs		
X41	spending on eating out	**amount spent** on eating out
X42	spending on supermarket	**amount spent** on supermarket
X43	spending on retail	**amount spent** on retail
X44	spending on online retail	**amount spent** on online retail
X45	spending on convenience store	**amount spent** on convenience store
X46	spending on clothing	**amount spent** on clothing
X47	spending on vehicle fuel	**amount spent** on vehicle fuel
X48	spending on public transport	**amount spent** on public transport
debts		
X49	spending on mortgage payment	**amount spent** on mortgage payment
X50	spending on vehicle loan	**amount spent** on vehicle loan
X51	spending on loan	**amount spent** on loan
X52	spending on credit card	**amount spent** on credit card
X53	spending on interest	**amount spent** on interest
X54	spending on overdraft charges	**amount spent** on overdraft charges
others		
X55	spending on gambling and gaming	**amount spent** on gambling and gaming
X56	spending on savings	**amount spent** on savings
X57	spending on healthcare and medical	**amount spent** on healthcare and medical
X58	spending on transfer payments	**amount spent** on transfer payments
X59	spending on leisure and entertainment	**amount spent** on leisure and entertainment
X60	cash withdrawal amount	**amount spent** on cash withdrawal
(b) decreasing trend		
fixed costs		
X61	decrease in rent payment	**number of months** the current month's amount spent on rent payment is lower than the previous month
X62	continuous decrease in rent payment	**number of three-month rolling windows** where the monthly amount spent on rent payment has a negative linear regression slope
X63	decrease in electricity and gas payment	**number of months** the current month's amount spent on electricity and gas is lower than the previous month
X64	continuous decrease in electricity and gas payment	**number of three-month rolling windows** where the monthly amount spent on electricity and gas has a negative linear regression slope
X65	decrease in insurance payment	**number of months** the current month's amount spent on insurance is lower than the previous month

X66	continuous decrease in insurance payment	**number of three-month rolling windows** where the monthly amount spent on insurance has a negative linear regression slope
X67	decrease in property and council tax payment	**number of months** the current month's amount spent on property and council tax payment is lower than the previous month
X68	continuous decrease in property and council tax payment	**number of three-month rolling windows** where the monthly amount spent on property and council tax has a negative linear regression slope
X69	decrease in telephone and tv payment	**number of months** the current month's amount spent on telephone and tv is lower than the previous month
X70	continuous decrease in telephone and tv payment	**number of three-month rolling windows** where the monthly amount spent on telephone and tv has a negative linear regression slope
X71	decrease in subscriptions and renewals payment	**number of months** the current month's amount spent on subscriptions and renewals is lower than the previous month
X72	continuous decrease in subscriptions and renewals payment	**number of three-month rolling windows** where the monthly amount spent on subscription and renewals has a negative linear regression slope
debts		
X73	decrease in mortgage payment	**number of months** the current month's amount spent on mortgage payment is lower than the previous month
X74	continuous decrease in mortgage payment	**number of three-month rolling windows** where the monthly amount spent on mortgage payment has a negative linear regression slope
X75	decrease in loan payment	**number of months** the current month's amount spent on loan payment is lower than the previous month
X76	continuous decrease in loan payment	**number of three-month rolling windows** where the monthly amount spent on loan payment has a negative linear regression slope
X77	decrease in vehicle loan payment	**number of months** the current month's amount spent on vehicle loan is lower than the previous month
X78	continuous decrease in vehicle loan payment	**number of three-month rolling windows** where the monthly amount spent on vehicle loan has a negative linear regression slope
X79	decrease in credit card payment	**number of months** the current month's amount spent on credit card payment is lower than the previous month
X80	continuous decrease in credit card payment	**number of three-month rolling windows** where the monthly amount spent on credit card payment has a negative linear regression slope
X81	decrease in interest payment	**number of months** the current months' amount spent on interest payment is lower than the previous month
X82	continuous decrease in interest payment	**number of three-month rolling windows** where the monthly amount spent on interest payment has a negative linear regression slope
X83	decrease in overdraft charges payment	**number of months** the current months' amount spent on overdraft charges payment is lower than the previous month
X84	continuous decrease in overdraft charges payment	**number of three-month rolling windows** where the monthly amount spent on overdraft charges has a negative linear regression slope
others		
X85	decrease in gambling and gaming	**number of months** the current month's amount spent on gambling and gaming is lower than the previous month

X86	continuous decrease in gambling and gaming	**number of three-month rolling windows** where the monthly amount spent on gambling and gaming has a negative linear regression slope
X87	decrease in savings	**number of months** the current month's amount spent on savings is lower than the previous month
X88	continuous decrease in savings	**number of three-month rolling windows** where the monthly amount spent on savings has a negative linear regression slope
X89	decrease in healthcare and medical expenses	**number of months** the current month's amount spent on healthcare and medical is lower than the previous month
X90	continuous decrease in healthcare and medical expenses	**number of three-month rolling windows** where the monthly amount spent on healthcare and medical has a negative linear regression slope
X91	decrease in transfer payments	**number of months** the current month's amount spent on transfer payments is lower than the previous month
X92	continuous decrease in transfer payments	**number of three-month rolling windows** where the monthly amount spent on transfer payments has a negative linear regression slope
(6) vulnerability		
X93	vulnerability risk index	a composite index constructed by PCA from the following numeric attributes: (i) total healthcare expenditure (ii) total income loss (total decreased inflow amount relative to the previous month) (iii) total amount of benefits received (iv) debt-to-inflow ratio (total debt amount divided by total inflow amount) (v) debt amount (sum of loan and overdraft amounts) (vi) outflow-to-inflow ratio (total outflow amount divided by total inflow amount) (vii) buffer-to-outflow ratio (sum of available balance and savings divided by mean monthly outflow amount) (viii) expenses on social activities or hobbies