CRC Credit Research Centre

CRC Working Papers

# Navigating dynamic financial risk: a latent Markov approach to safeguard vulnerable consumers in Open Banking

Rui Ying Goh, Galina Andreeva, Yi Cao

2024

Keywords: Decision analysis, Risk analysis, OR in banking

THE UNIVERSITY of EDINBURGH

Contact
Credit Research Centre
University of Edinburgh Business School
29 Buccleuch Place
Edinburgh, EH8 9JS
credit.research.centre@ed.ac.uk

# Navigating dynamic financial risk: a latent Markov approach to safeguard vulnerable consumers in Open Banking

Rui Ying Goh, Galina Andreeva, Yi Cao

* Business School, University of Edinburgh, 29 Buccleuch Pl, Edinburgh EH8 9JS, United Kingdom

## Abstract

In recent years, banks have exhibited an increased commitment to protecting vulnerable consumers. Their conventional approach remains primarily centred on detecting poor financial state, with relatively less emphasis on monitoring state transitions over time. Nonetheless, tracking the fluidity in financial vulnerability (FV) offers early warning indicators for banks and ex-ante protection for customers. In this paper, we introduce the latent Markov model for current account risk assessment by estimating accounts' financial states, evaluating the prob ability of transitions over time, and predicting the most likely sequences of future transitions. The model parameterisation, conditional on the distribution of account balance, quantifies dis tress into low, medium, and high severity states. The transition probabilities evaluate accounts' tendency to improve or deteriorate where low-risk accounts tend to maintain stable, medium risk accounts often shift to low-risk instead of progressing to high-risk, and high-risk accounts are more likely to improve rather than being trapped in high-risk. Additionally, the model interprets vulnerable characteristics and predicts risk migration, helping to monitor financial vulnerability tenacity and identify drivers of financial transitions, thereby flagging early alerts for detrimental deterioration. This approach presents a novel and easily implementable data driven tool to facilitate the financial risk decision-making process by accounting for evolving cash flow behaviour.

# 1 Introduction

Financial vulnerability (FV) serves as a precursor to financial distress, signifying various degrees of financial hardship that amplify the risk of severe adversity (O'Connor et al. 2019). Detecting FV is pivotal for early consumer engagement to mitigate potential harm (Collard 2011, Financial Conduct Authority 2021*a*). Regulators have placed the protection of vulnerable consumers at the forefront of their agenda, ensuring equitable access to financial products for all, including the underprivileged (CMA 2019, Financial Conduct Authority 2015). Therefore, identifying FV is a crucial step in preventing harm at an early stage.

The implementation of Open Banking (OB) provides a novel data source for early detection of FV. Built upon the foundation of the second payment service directive, OB mandates major banks to grant third-party service providers access to consumers' bank account data with their consent. Specifically, the current accounts connected through OB offer unique attributes for extracting insights into financial behaviour and identifying FV. First, the transactional records in these current accounts reflect cash flow dynamics (inflow and outflow) that cater to various day-to-day financial needs, unlike accounts with dedicated usage for mortgages or credit cards. Second, the data provides highly granular cash flow transactions, up to a daily frequency, enabling the tracking of short-term or immediate shifts in financial conditions. This data source empowers practitioners to issue timely warnings of FV, objectively assessing consumer risk without potential biases introduced by self-reported information obtained through interviews or surveys. Furthermore, this alternative FV risk assessment approach can supplement or replace the sole reliance on credit scores. This shift not only facilitates assessments based on the latest cash flow patterns rather than outdated bureau records at the time of loan applications but also broadens the risk evaluation for clients with limited credit histories, enhancing their accessibility to financial products.

However, such temporal aspect of cash flow behaviour has not been investigated in the personal finance context. Existing literature has primarily focused on examining poor financial status at specific time points (Gathergood & Guttman-Kenney 2016, Kim et al. 2023, Harrison & Andreeva 2021), lacking the capacity to trace risk movements swiftly in response to potential struggles. Leveraging the OB-connected current account cash flow transactions, our research addresses this critical gap by developing a comprehensive risk framework. This framework aims not only to identify but also to respond to evolving FV, providing a more dynamic and proactive approach to managing financial risk.

FV is associated with the likelihood of facing hardships (O'Connor et al. 2019), suggesting that possessing lower (higher) FV characteristics exposes individuals to a lower (greater) level of distress as their financial circumstances evolve over time. However, projecting the temporal impact of FV on potential distress remains a nuanced and unstructured challenge. While existing literature has outlined the dimensions of FV, it lacks a precise definition for quantifying financial risks. Guidelines have been proposed for the key dimensions of FV, yet there is a notable absence of specifications for objectively measuring the associated risk levels (Financial Conduct Authority 2021*b*). Kim et al. (2023) has compiled a comprehensive list of financial behaviours derived from OB data to indicate signs of FV. However, these individual measures are not interconnected, where the fragmented measures could not scale the multiple stages of financial troubles. Recognising the fluid nature of financial circumstances, our focus lies in conceptualising temporal FV to assess the dynamics across a spectrum of distress risk magnitudes.

Building on the well-established field of credit scoring, our study adopts the credit risk migration framework to assess financial risk dynamics in the context of OB. Modelling dynamic financial scenarios within bank transactions introduces intricacies beyond the mainstream credit risk migration framework in two key aspects. First, traditional methods typically model the transition among explicitly defined states, such as multiple corporate rating levels (Duffie et al. 2007), various stages of delinquency like arrears or 90 days past due to indicate default (Kelly & O'Malley 2016, Bocchio et al. 2022), or dividing the credit behavioural scores into bands to define credit states (Malik & Thomas 2012). However, there is no clearly defined target variable for quantifying varying degrees of hardship and determining the static, non-reverting state of distress from cash flow. Second, conventional methods often employing survival-based (Duffie et al. 2007, Kelly & O'Malley 2016, Bocchio et al. 2022, Leow & Crook 2014, Djeundje & Crook 2018) or Markov-based techniques with regression specifications (Malik & Thomas 2012, Frydman & Matuszyk 2018, Frydman et al. 2019) to model credit risk state transitions within a closed system, which include a permanent absorbing state (default). In contrast, account balances can move between positive and negative, and individuals may strategically use overdrafts as a cash flow turnover strategy rather than indicating a permanent financial difficulty.

These two aspects of modelling distress risk dynamics in cash flow underscore the necessity of employing the latent Markov (LM) model. Despite being a non-mainstream method for estimating risk transitions in the credit domain, latent state transitions effectively define hidden scenarios based on observed credit statuses, such as underlying economic scenarios from corporate ratings (Thomas et al. 2002, Yu et al. 2019,

Blümke 2022) and FV derived from the number of payday loans obtained (Medina-Olivares & Calabrese 2023). Notably, this method extracts the latent states as the target variable and does not impose the restriction of having an absorbing state in the transition process. Our work pioneers the application of the LM model to formalise the assessment of fluid financial states through a data-driven definition of distress risk states in observed cash flow balances. Additionally, we predict state transitions conditional on evolving FV characteristics, contributing to a nuanced understanding of financial dynamics.

We aim to evaluate the evolution of financial risk through current account cash flow transactions. Leveraging the LM model, we estimate implicit financial states through three facets of observed account balance information: the duration of negative balances, the significance of negative amounts, and the fluctuations in the balance amount. These discovered latent states act as the target variable, quantifying the various stages of distress risk on a quarterly basis. We employ vulnerable characteristics from the current quarter as covariates to estimate and predict the transitions of distress risks (improvement or deterioration) in the next quarter, allowing us to monitor the shifting distress risk over time. Furthermore, we analyse the marginal effects of significant covariates to understand the drivers of financial improvement and deterioration, as well as to differentiate between transient and permanent FV, thereby highlighting transitions related to financial instability.

This study contributes to both the fields of credit scoring and OB research. By extending the credit scoring methodology and embracing the latent Markov statistical framework, we formalise the setup to model evolving FV to assess financial risk transitions. We introduce a novel measurement scale for quantifying varying stages of distress and objectively diagnosing potential financial challenges. Moreover, our research unveils the relationship between FV and state transitions, shedding light on account-specific factors that influence changes in financial status. We can also identify early warning signs of detrimental deterioration by monitoring the tenacity of FV. These contributions are particularly valuable for consumer protection and can assist practitioners in designing effective, customer-centric strategies to mitigate existing risks and facilitate financial improvement.

This paper is structured as follows: Section 2 reviews relevant past studies, Section 3 outlines the data description and procedures for measuring distress, in addition to detailing the steps involved in estimating the evolution of financial states over time using a multivariate latent Markov model. Section 4 presents and discusses the findings, and Section 5 concludes the paper.

## 2 Related work

Existing approaches to support vulnerable consumers often depend on the consumers' proactive communication of their circumstances to seek financial assistance (Financial Conduct Authority 2015). However, these evaluations are primarily based on interviews or surveys, which can lead to misperceptions, as financially stressed individuals may not recognise their own vulnerability (O'Connor et al. 2019, Financial Conduct Authority 2015). Moreover, self-reported information tends to provide a summary of one's financial situation over a year, lacking the ability to capture the dynamic nature of their circumstances. This limitation hinders proactive support, making it difficult to pinpoint the exact onset of vulnerability (Collard 2011). To accommodate the evolving financial journey, firms have suggested the implementation of objective assessments as a complement or alternative to the current subjective framework (Experian 2019, Financial Conduct Authority 2018). This study focuses on the dynamic aspect of financial vulnerability (FV) to enable the ongoing monitoring of financial conditions.

Open Banking (OB) transactions represent a new source of data to detect FV. Despite the advantages of cash flow transactions in reflecting fluid financial situations, objective financial risk evaluation using banking data is underrepresented in the literature. Existing risk models have two notable limitations. First, they assess financial risk over fixed periods without considering evolving patterns. For example, Gladstone et al. (2020) examine the impact of current account transactional behaviour on financial well-being, while Kim et al. (2023) predict future FV based on current financial behaviour using OB data. Harrison & Andreeva (2021) analyse OB financial behaviour of high-cost short-term credit applicants over a year and the results detect a concerning proportion i.e., a quarter of the individuals in the sample, are in financial difficulty.

Second, there is a lack of specific guidelines for objectively quantifying FV and distress risk levels solely based on cash flow transactions. Unlike household FV, which is often measured in cases of overindebtedness, personal FV remains an underexplored area. While Kim et al. (2023) has proposed OB-specific FV indicators, these binary variables do not measure the magnitude of FV. Similarly, scaling financial difficulties to measure distress is rarely investigated in the literature. Previous studies have primarily focused on broader aspects of financial well-being, such as effective expense management, the ability to achieve financial goals, and the confidence to maintain control over finances (Bailey et al. 2022, Consumer Financial Protection Bureau 2017). Gathergood & Guttman-Kenney (2016) is one of the few works that has focused on defining distress within the context of overindebtedness, using objective credit arrears information and

subjective opinions to manage debt, rather than assessing day-to-day financial behaviour. Despite the significant impact of FV on distress, the connection between FV and distress risk transitions has not been explored. Therefore, the development of a methodology for modeling dynamic FV to assess evolving distress risks using OB data remains an unaddressed research challenge.

Assessing the dynamics of consumer risk has a well-established framework in credit scoring. We draw inspiration from the credit scoring methodology to establish an OB-specific model for assessing financial transition risks. Previous work has focused on the migration of corporate credit ratings to detect the timing of rating downgrades (Lando & Skødeberg 2002, Schuermann et al. 2003) and different severities of consumer credit delinquency (Schechtman 2013, Schuermann et al. 2003, Ferretti et al. 2019) or secondary mortgage commercial loans (Santos 2018) to monitor default and recovery. In addition to tracking transitions between credit states, several studies emphasise the importance of explaining the variables that influence these transitions as key insights for identifying harm. To investigate the link between FV and the migration of distress risk states, we further explore credit scoring methods that incorporate dynamic models with covariate effects.

Several studies utilise survival-based models with a Cox proportional hazard specification to estimate transition intensities that capture the time until the first event occurs. For instance, to predict the corporate bankruptcy process based on firm-specific and macroeconomic covariates, Duffie et al. (2007) combine a Cox-based duration model with a first-order vector autoregression model. This allows the simultaneous consideration of time-varying covariates and stochastic time series, facilitating the estimation of multiperiod survival probabilities. Hazard models are also used to express the transition intensities as functions of loan characteristics and time-varying macroeconomic variables. In residential mortgage credit risk assessment, Kelly & O'Malley (2016) estimate specific shifts, such as cure (from delinquency to performing) and default (from delinquency to default), using conventional intensity models. Bocchio et al. (2022) further examine nuanced transitions across different levels of delinquency severity and explicitly considers repeated events to forecast migrations within the lending portfolio. To monitor credit card delinquency risk, Leow & Crook (2014) predict transitions across varying delinquent states at future time periods, while Djeundje & Crook (2018) consider account-level random effects to capture unobserved heterogeneity in the intensity process.

On the other hand, the Markov model, which assumes that the future state depends solely on the current state, extended with a logit-based regression specification, is another common approach for incorporating covariate effects when estimating risk migrations. For example, Malik & Thomas (2012) predict the

movements of behavioral credit status based on factors like loan tenure, repayment behavior, and economic indicators using a cumulative logit specification. To estimate separate processes for "stayers" and "movers," loan application characteristics are included to parameterise the "stayer" probability with logit (Frydman & Matuszyk 2018) and the "mover" transition matrix with multinomial logit (Frydman et al. 2019). Generally, financial risk assessment in OB resembles the credit scoring framework, predicting transition probabilities across varying stages of distress risk (delinquency severities) based on changing FV (time-varying covariates or loan characteristics) to manage the risk of bank accounts (credit portfolios). To the best of our knowledge, this is the initial attempt to conceptualise the OB risk assessment framework based on the credit scoring methodology.

However, these mainstream credit scoring methods are constrained by their ability to estimate transitions only across defined states within a closed system, which includes a permanent absorbing state. This limitation proves challenging within the OB framework for two main reasons. First, there are no clearly defined distress states that can quantify varying levels of hardship solely from cash flow transactions. Varying stages of distress exposure risk are not directly observable from cash flow data but are implicitly represented by the account balance situations. Second, an irreversible state of distress is not readily identifiable. Continuous overdraft use (negative balance) does not necessarily indicate complete unaffordability, such as in the case of credit default, as it may serve as a strategy to manage daily expenses. The latent Markov model is employed when the focus is on assessing credit portfolio risk based on hidden states, rather than the observable credit states e.g., examining underlying economic scenarios, ranging from recession to robust economic growth, as inferred from corporate credit ratings (Thomas et al. 2002, Yu et al. 2019, Blümke 2022). To incorporate covariate effects in the transition probability distribution, Blümke (2022) have forecasted shifts in hidden economic scenarios based on observed credit ratings and macroeconomic variables. Medina-Olivares & Calabrese (2023) have included inflows and outflows as covariates in the latent Markov model to estimate transitions of hidden vulnerability states identified from the number of payday loans. Establishing a latent Markov statistical model to measure implicit distress risk in cash flow patterns within the OB context is a novel approach to monitor financial transitions based on the relationship between FV and distress.

# 3 Methodology

## 3.1 Data description

We analyse 19,447 personal current accounts with transaction data spanning 12 months. Each transaction includes details like account identifier, description, date, transaction type (credit/debit), amount, and post-transaction balance.

Our objective is to identify signs of distress within cash flow transactions. Distress, in general, signifies an inadequacy in current cash flow to meet immediate financial obligations. Specifically, personal distress has been defined in past study (Keys et al. 2023) as lasting hardships in fulfilling credit commitments. This is measured through early indicators such as credit card delinquency extending to 30 days or more past due. Moving along the spectrum of severity, the presence of debt in collection signifies a more critical level where the credit account is written off and sold. At the extreme end of the spectrum, personal bankruptcy represents the most severe case of financial distress. To portray enduring financial challenges arising from cash flow transactions, our emphasis lies in scrutinising the cash flow balance position across an extended time frame of several months. A persistent state of negative balance or a diminished buffer amount signifies a shortfall in financial resources. In contrast to the standard practice of monthly evaluations in conventional financial reporting for assessing firms' financial health, our approach centres on a quarterly time window. This focus ensures that our balance metrics, measured over several months, effectively capture prolonged periods (beyond a month) of financial hardship manifested in cash flow transactions, providing a more comprehensive indicator of potential distress.

Our goal is to understand how transaction patterns related to vulnerability influence the movement of distress risk states over quarters. Figure 1 depicts the predictive setup, where we project the change in distress risk state (explained in Section 3.2) based on transactional behaviour related to vulnerability (explained in Section 3.3) in the current quarter. The predictive model is trained on 80% of the accounts and tested on the remaining 20%. We use lagged covariates for one-month-ahead predictions. Therefore, the in-sample and out-of-sample predictions track the sequence of distress states from Q2 to Q4, influenced by transactions in the preceding quarters (Q1, Q2, and Q3).

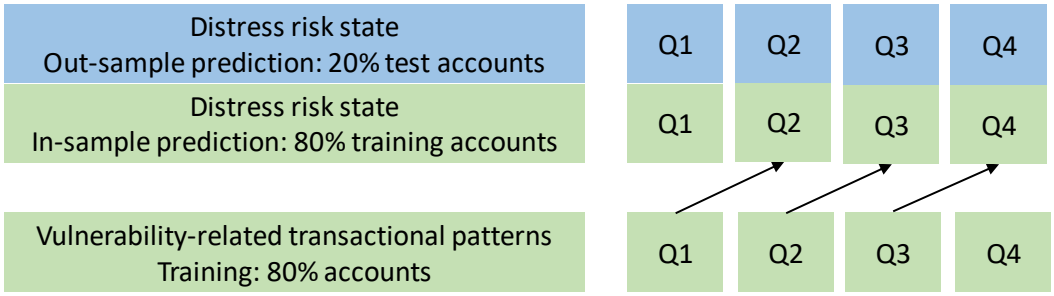



Figure 1: Prediction setup

## 3.2 Definition of distress

A measure of distress intends to capture varying levels of financial difficulty, ranging from minimal to extreme hardships. In the context of Open Banking (OB), we propose to quantify distress by examining overdraft use since negative balances indicate a deficit, essentially borrowing from the bank. In the absence of a concrete distress definition, We follow the logic adopted in credit scoring, where credit default is measured by the number of days past due and the overdue amount (Ernst & Young 2018). For financial distress, it has to be relative to the standard financial capacity of an individual. The same overdrawn amount may mean an insignificant hiccup for one person, and a major difficulty for another. It is also important to consider the changes in the financial situation from month to month. We follow Harrison et al (2022) and use the minimum monthly balance as a proxy for the worst situation during the month. Coefficient of variation (CV) allows to capture some changes in this worst position over three months. A low CV value suggests that the account has maintained a precarious state throughout the quarter. On the other hand, a notably high CV value indicates significant fluctuations in the minimum balance, reflecting either a substantial increase or a sharp decline over the quarter.

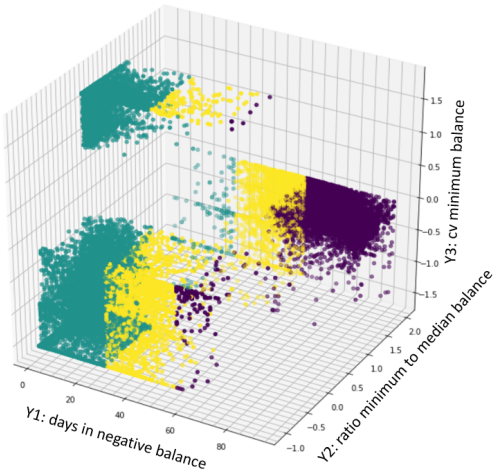
We incorporate three aspects of negative balances to measure distress: (i) the number of days in a negative balance ($y1$), (ii) the ratio of the (average) minimum balance to the (average) median balance ($y2$), and (iii) fluctuations in the minimum balance ($y3$), quantified using the CV. The CV is obtained by dividing the standard deviation of the minimum balance by its mean, allowing for a scaled comparison of between-account variations, regardless of the significant differences in minimum balance amounts. Frequent overdraft use is closely linked to subpar credit performance, evidenced by a lower average credit score, a

higher likelihood of possessing a subprime credit score, absence of available credit card balance, delinquent debt, and a diminished financial capacity to meet bill obligations (Consumer Financial Protection Bureau 2023). Consequently, detecting overdraft usage within cash flow balances serves as an effective means to discern ongoing financial challenges and measure the associated risks of distress. $y1$ represents how long an overdraft is used, while $y2$ reflects the magnitude of the overdraft amount, analogous to the definition of credit default, which considers the duration (90 days past due) and the materiality of the overdue amount (Ernst & Young 2018). However, $y1$ and $y2$ alone are insufficient to signify a constant pattern of negative balances, so we include $y3$ to account for this stagnant difficult situation.

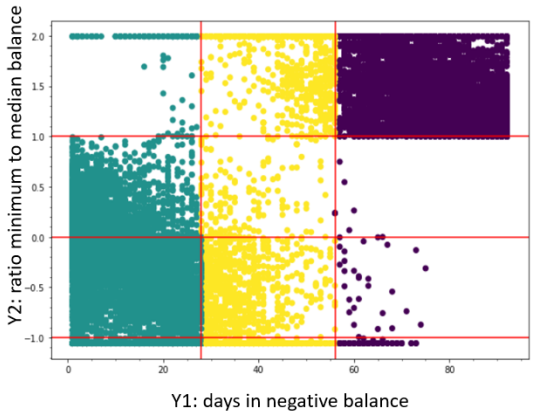
To categorise these variables into distinct levels of financial struggles, we propose a two-stage threshold selection framework. First, we use hierarchical clustering with $y1$, $y2$, and $y3$ to identify groupings. Then, we visualise these groupings in a 3D space (Figure 2a) and refine cutoff points in two-dimensional spaces (Figures 2b, 2c, 2d). We omit cases where y1 equals zero, indicating a positive balance, only during the threshold selection process. This exclusion is necessary due to the majority of such instances, which generate a sparse matrix and potentially cause memory-related issues. Given that $y1 = 0$ characterises the category of financially sound accounts, its exclusion does not compromise the identification of groups facing financial challenges. The threshold selection procedure then continues with the remaining 8940 accounts. Figure 2a shows clear thresholds to create three distinct groups.

Referring to Figure 2b, we set thresholds at $y1 = 28$ and $y1 = 56$ to create three regions in the data. This provides four categories of $y1$ after excluding the cases where $y1 = 0$. For $y2$, we use $y2 = -1$, $y2 = 0$ and $y2 = 1$ (horizontal lines) as cutoffs to divide $y2$ into four categories. Category 1 has a balanced mix of all three groups, category 3 has the fewest yellow points, while categories 2 and 4 are mostly blue and purple points, respectively. For $y3$, the selected thresholds are $y3 = 0$ and $y3 = -0.80$. As evident from Figures 2c and 2d, the cutoff at $y3 = 0$ clearly separates accounts with positive minimum balance fluctuations ($y3 > 0$), and $y3 = -0.80$ further divides accounts into those with low ($y3 < -0.80$) and high ($-0.80 \leq y3 < 0$) negative minimum balance fluctuations, with the high negative volatility category primarily containing purple points.

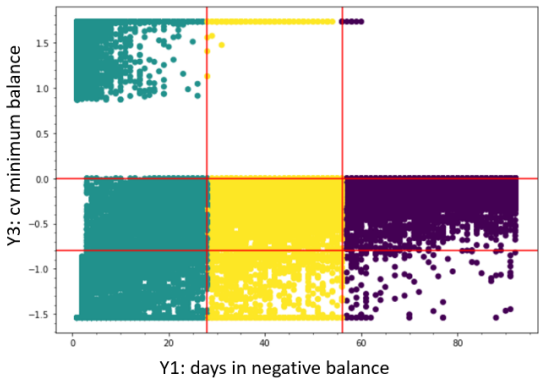
The proposed framework not only helps in selecting thresholds but also reveals multiple levels of distress through the clusters, exposing the non-linear relationship between the three distress measurement variables. This complexity arises because accounts may have many days of negative balance with a small overdraft



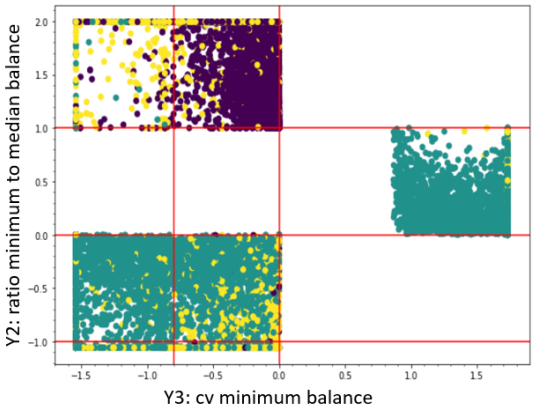

(a) Groupings based on the three aspects of negative balances




(b) $y1$ vs $y2$




(c) $y1$ vs $y3$




(d) $y2$ vs $y3$

Figure 2: Categories for the three aspects of negative balances

amount or vice versa. Accounts may also have frequent negative balance days, but their monthly minimum balance fluctuates between positive and negative. Such complications highlight the need for a measure of distress that can account for these non-linear relationships. We propose using a latent Markov model (detailed in Section 3.4) to uncover the underlying stages of distress using the three observed aspects of overdraft use.

## 3.3 Covariates

Individuals facing vulnerability are at a heightened risk of falling into financial distress. We gauge FV using income and expenditure patterns as covariates to predict potential movement towards or away from distress. The covariate computation draws inspiration from the four central aspects of FV outlined by the Financial Conduct Authority (2021*b*):

1. Low Financial Resilience: This is measured by approximating potential over-indebtedness, as seen in negative balance situations (X1-X2) as well as evaluating financial capacity based on the surplus available to handle additional expenses (X3) and the ability to manage spending within the given income (X4).

2. Negative Life Events: We account for unforeseen events like job loss, bereavement, or relationship breakdown by considering changes in income (X5), income fluctuations (X6), and variations in benefits (X7).

3. Poor Health: Expenses related to healthcare (X8) reflect costs associated with treating illnesses and health-related issues.

4. Low Capability: Covariates X9-X16 serve as proxy measures of financial literacy. They evaluate whether the accounts have the capability to afford debt payments with their existing income (X9), possess proper money management skills to prioritise essential expenses (X10), maintain healthy spending habits within a controlled budget, and have the confidence to handle changes in key expenses (X11-X14), exhibit harmful financial practices such as overspending on non-essential items (X15) or addictive gambling behaviour (X16).

A detailed list of these covariates is provided in Table 1.

Table 1: List of covariates

| Abbreviation | Vulnerability aspect | Covariate |
|---|---|---|
| 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 |

## 3.4 Modelling the evolution of distress risk

We employ the latent Markov model to estimate and predict the transition probabilities between various distress risk states. The progression of these different stages of distress risk represents a latent process since the defined distress measures (as discussed in Section 3.2) indirectly signify financial difficulties based on observed balance situations. For each account, we investigate how past vulnerable transactional patterns influence the shift to a new distress level, enabling us to predict a dynamic sequence of distress risks.

The Markov model characterises the stochastic process of a system as it sequentially transitions from one state to another. This process adheres to the Markov property, which means that the distribution of the future state depends solely on the current state (Gagniuc 2017). In our case, we utilise a latent variable model, specifically tailored to examine changes in an unobservable characteristic of interest, assuming that the system follows a Markov chain (Wiggins 1973). We focus on an extended multivariate latent Markov (LM) model, which incorporates covariates that influence the latent distribution (Bartolucci et al. 2017). The LM model provides valuable insights into how these covariates affect the unobservable characteristic, as measured by the response variable. This is crucial for distinguishing between the account-specific characteristics governing initial and transition probabilities. The following subsections provide a detailed outline

of the multivariate LM model, including the extension to incorporate covariates, parameters estimation, and sequence prediction procedures.

### 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 3.2) 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 $c_j = 0, ..., c_j - 1$, summarised in Table 2. 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.

Table 2: Observed target variables

| Observed target variable | Categories |
|---|---|
| $Y_1$: Days in negative balance | $c_1$: Categories in $Y_1$ |
| | 0: $y_1 = 0$ |
| | 1: $1 \leq Y_1 < 28$ |
| | 2: $28 \leq Y_1 < 56$ |
| | 3: $Y_1 \geq 56$ |
| $Y_2$: Ratio of minimum to median balance | $c_2$: Categories in $Y_2$ |
| | 0: $Y_2 < -1$ |
| | 1: $-1 \leq Y_2 < 0$ |
| | 2: $0 \leq Y_2 \leq 1$ |
| | 3: $Y_2 > 1$ |
| $Y_3$: Minimum balance CV | $c_3$: Categories in $Y_3$ |
| | 0: $Y_3 > -0.80$ |
| | 1: $-0.80 \leq Y_3 < 0$ |
| | 2: $Y_3 > 0$ |

The underlying distress risk states are represented by the latent process. 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

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

which are stored in a $c_j \times k$ matrices $\Phi_j$ 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

$$P(\mathbf{Y} = \mathbf{y}|\mathbf{U} = \mathbf{u}) = \prod_{t=1}^{T} \phi^{(t)}_{y^{(t)}|u^{(t)}}, \tag{2}$$

where due to the local independence assumption,

$$\phi^{(t)}_{y|u} = \prod_{j=1}^{J} \phi^{(t)}_{jy_j|u}, \tag{3}$$

$\mathbf{y}$ 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 3.3) 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

$$\pi_{u|x} = P(U^{(1)} = u|x^{(1)}), \qquad u = 1, ..., k, \tag{4}$$

and the transition probabilities are

$$\pi^{(t)}_{v|ux} = P(U^{(t)} = v|U^{(t-1)} = u, x^{(t)}), \qquad t = 2, ..., T, u, v = 1, ..., k. \tag{5}$$

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

$$P(\mathbf{U} = \mathbf{u}|\mathbf{X} = \mathbf{x}) = \pi_{u^{(1)}|x^{(1)}} \prod_{t=2}^{T} \pi_{u^{(t)}|u^{(t-1)}x^{(t)}}, \tag{6}$$

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 $\mathbf{X}$,

$$P(\mathbf{y}|\mathbf{x}) = \sum_{\mathbf{u}} \pi_{u^{(1)}|x^{(1)}} \pi_{u^{(2)}|u^{(1)}x^{(2)}} ... \pi_{u^{(T)}|u^{(T-1)}x^{(T)}} \times \phi_{y^{(1)}|u^{(1)}x^{(1)}} ... \phi_{y^{(T)}|u^{(T)}x^{(T)}}. \tag{7}$$

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 8) and transition (Equation 9) probabilities,

$$log\frac{P(U^{(1)}=u|x^{(1)})}{P(U^{(1)}=1|x^{(1)})}=\beta_{0u}+(x^{(1)})'\beta_{1u}, \qquad u\geq 2, \tag{8}$$

$$log\frac{P(U^{(t)}=v|U^{(t-1)},x^{(t)})}{P(U^{(t)}=u|U^{(t-1)}=u,x^{(t)})}=\gamma_{0uv}+(x^{(t)})'\gamma_{1uv}, \qquad t\geq 2, u\neq v, \tag{9}$$

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 $\beta$ and $\Gamma$ respectively.

### 3.4.2 Parameters estimation

This subsection describes the parameters estimation procedure with the Expected-Maximisation (EM) algorithm (Dempster et al. 1977). 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 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

$$\begin{aligned} l(\theta)=\sum_{i=1}^{n}[\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{u=1}^{k}\sum_{y=0}^{c_j-1}a_{ijuy}^{(t)}log\phi_{jy|u}^{(t)}+ \\ \sum_{u=1}^{k}b_{iu}^{(1)}logP(U_i^{(1)}=u|x_i^{(1)})+ \\ \sum_{t=2}^{T}\sum_{v=1}^{k}\sum_{u=1}^{k}b_{iuv}^{(t)}logP(U_i^{(t)}=v|U_i^{(t-1)}=u,x_i^{(t)})] \end{aligned} \tag{10}$$

where $\theta$ 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 10) 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 $\beta$ and $\Gamma$ by maximising the components of the complete data log-likelihood with the latent structure parameters i.e., fitting the two multinomial logits (Equation 5 and 9) based on the weighted log-likelihoods below (Bartolucci et al. 2015),

$$l(\beta) = \sum_i \sum_u \hat{b}_{iu}^{(1)} logP(U_i^{(1)} = u|\mathbf{x}_i^{(1)}), \tag{11}$$

$$log(\Gamma) = \sum_i \sum_{t\geq 2} \sum_u \sum_v \hat{b}_{iuv}^{(t)} logP(U_i^{(t)} = v|U_i^{(t-1)} = u, \mathbf{x}_i^{(t)}) \tag{12}$$

### 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

$$\hat{P}_i^{(1)}(u_i^{(1)}, y_i^{(1)}|x_i^{(1)}) = \hat{\pi}_{u_i|x_i}^{(1)} \hat{\phi}_{y_i^{(1)}|u_i^{(1)}}. \tag{13}$$

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

$$\hat{P}_i^{(t)}(v_i^{(t)}, y_i^{(t)}|x_i^{(t)}) = \hat{\phi}_{y_i^{(t)}|v_i^{(t)}} max_{u=1,...,k} \big[\hat{P}^{(t-1)}(u_i^{(t-1)}, y_i^{(t-1)}|x_i^{(t-1)}) \hat{\pi}_{v_i|u_i x_i}^{(t)}\big]. \tag{14}$$

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

$$\hat{u}_i^{*(T)}(y_i^{(T)}|x_i^{(T)}) = argmax_{u=1,...,k} \hat{P}_i^{(T)}(u_i^{(T)}, y_i^{(T)}|x_i^{(T)}). \tag{15}$$

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

$$\hat{u}_i^{*(t)}(y_i^{(t)}|x_i^{(t)}) = argmax_{u=1,...,k} \hat{P}_i^{(t)}(u_i^{(t)}, y_i^{(t)}|x_i^{(t)}) \hat{\pi}_{\hat{u}_i^* y_i|u_i}^{(t+1)}. \tag{16}$$

# 4 Results and discussions

Given that the number of latent states ($k$) is not known in advance, we estimate multiple models with varying values of $k = [2, 5]$. We then assess the models using AIC/BIC criteria and determine that the model with $k = 3$ best represents the data. This choice is based on its ability to minimise information loss, as indicated by the lowest AIC/BIC values, and it exhibits no convergence issues.

Consequently, the latent Markov model reveals three underlying states associated with potential distress, derived from the observed balance scenarios. These three latent states will be the focus of the subsequent discussions, including (i) the characteristics of these latent states, (ii) the distribution of the latent process to understand transitions between states, (iii) the influence of covariates on the latent distribution, and (iv) the prediction of state sequences for new accounts.

## 4.1 States of financial distress risk

Table 3 reports the estimated emission probability matrix, which illustrates the likelihood of observing specific balance scenarios given the latent states. These conditional response probabilities highlight key characteristics associated with each latent state (highlighted in bold) and help define their respective properties. The output from Table 3 enables the classification of the latent states as exhibiting varying degrees

Table 3: Estimated emission probability

| Observed target variables | Category | Latent states 1 | 2 | 3 |
|---|---|---|---|---|
| $y1$: 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 |

of risk, specifically as low, medium, and high risk with regard to facing financial distress.

Latent state 1 corresponds to a medium-risk state. Accounts in this state are most likely to experience short-term negative balances for 1-27 days during a quarter (77.4%). They typically maintain a positive buffer to handle relatively low-magnitude negative balances (66.3%) and encounter unexpected drops into negative minimum balances, as indicated by significant negative fluctuations over the months in a quarter (61.4%). Additionally, this state reveals secondary characteristics suggesting a regular overdraft practice, potentially as a strategy to manage day-to-day living expenses. Accounts in this state have a relatively lower likelihood of prolonged negative balances lasting 28-55 days (22.5%), inadequate positive funds to cover a large negative situation (29.6%), and consistent negative minimum balances with minimal fluctuations across the quarter (38.1%).

Latent state 2 depicts a low-risk state characterised by a consistently positive balance situation. Accounts in this state are highly likely to maintain positive balances throughout the entire quarter (91.3%), with sufficient funds to ensure the median minimum balance remains positive (100%) and positive fluctuations indicating that changes in the balance amount do not result in financial shortfalls (100%).

In contrast, latent state 3 signifies a high-risk state where accounts exhibit persistent dependence on overdraft facilities. Accounts in this state have the highest probability of experiencing negative balances for more than 56 days within a quarter (75.5%). They typically face a severe deficit position, with simultaneous negative median balances and significant negative minimum balances (90.2%). Accounts in this state consistently remain at the same level of negativity with minimal fluctuations throughout the quarter (93.2%).

## 4.2 Dynamics of the distress risk states

Table 4 provides a comprehensive breakdown of the distribution of distress risk states, divided into two key aspects. First, the estimated initial probabilities shed light on the proportions of accounts in each state at the onset of the observation period. Second, the estimated transition probabilities offer insights into the likelihood of accounts either remaining in their current states or transitioning to different states. From this reported latent process distribution (Table 4), we gain a valuable perspective on the dynamics of distress risk states across the quarters. Note that the covariates are lagged by one period to predict the risk state in the next quarter, thus making the observed time frame span from Q2 to Q4.

Table 4: Distribution of latent process

| | | **Initial probabilities** | | |
|---|---|---|---|---|
| | | medium | low | high |
| Q2 | | 0.1316 | 0.7227 | 0.1457 |
| | | **Transition probability matrix** | | |
| | | medium | low | 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 |

A prevalent financial pattern observed among many accounts is the consistent maintenance of a stable financial state with a positive balance throughout the year. Most accounts start in a low-risk state (72.2%),

and these accounts are likely to remain in this state. Approximately 84% of these accounts sustain a low risk profile from Q2 to Q4. Only a minor fraction, less than 10%, transitions to medium or high-risk states.

Conversely, a smaller proportion of accounts commence with medium (13.2%) or high (14.6%) risk at the beginning of the observation period. Medium-risk accounts demonstrate the most dynamic behaviour, with a 31.5% probability of remaining in the same state. Furthermore, these accounts frequently revert to a low-risk state (54%) rather than progressing to a high-risk state (15%). This indicates that accounts facing temporary financial challenges often utilise short-term overdrafts as a means to foster an optimistic recovery pattern. Even in high-risk situations, these accounts have a chance to escape rather than becoming entrenched in this state.

These findings reflect the adaptability of current accounts in traversing various risk states. Consequently, we can monitor the evolving stages of financial difficulty to assess the exposure risk to distress, although the model does not directly identify a severe and unrecoverable distress state.

## 4.3 Estimated regression results: insights into vulnerability driving potential distress

In addition to tracking the evolving financial positions over the quarters, we delve into the estimated regression parameters that impact both the initial probability (Table 5) and transition probability distribution (Tables 6 and 7) to uncover the factors influencing the latent process. The interpretation of the regression coefficients follows the logit model, whereby a positive/negative unit change in a covariate, holding others constant, is expected to increase/decrease the log odds of the model outcome (latent state) concerning the referent group.

Table 5 investigates the impact of covariates on the initial distress risk states, using the medium-risk state as the reference group. Significant covariates indicate the distinct characteristics associated with each distress state. Accounts are more likely to begin in a low-risk state when they portray financial stability in the preceding quarter, characterised by having adequate funds to cover expenses (X3) and a low incidence of negative balance occurrences (X1, X2). The low-risk state also correlates with higher inflow fluctuations (X5) when accounts receive supplementary income (X6, X7) to bolster their financial resources, as well as a wealthy status that allows servicing more debt (X9) and managing additional expenses (X4).

Table 5: Estimated parameters for initial probability distribution

| | | Latent states | |
|---|---|---|---|
| Covariate | Abbreviation | low | 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 $\alpha$ =0.10* 0.05** 0.01*** 0.001****

On the other hand, for the high-risk state, most estimated logits display opposite signs to those in the low-risk state, emphasising contrasting circumstances. Inadequate funds in the preceding quarter are the key driver of future high distress risk. These accounts often commence with a positive balance (X1) but frequently end up with a negative median balance (X2), signifying financial strain after essential expenses are met. Although the buffer-to-outflow ratio (X3) is not significant, the negative sign corroborates that high-risk accounts lack the financial cushion to absorb additional expenditures. Low earnings, without additional inflows, as denoted by the non-significance of X5, X6, and X7, increase the likelihood of high-risk states. Potential financial difficulties may arise from reliance on higher levels of debt at lower income (X9) or imprudent financial practices leading to high outflows, irrespective of limited inflow (X4).

In subsequent time periods, the model predicts transitions for each account. Tables 6 and 7 present the estimated parameters for two types of transitions: transitioning from higher to lower risk (improvement) and moving from lower to higher risk (deterioration) in the future. These transitions are associated with distinct sets of characteristics.

Financial improvements (Table 6) are generally linked to financial resilience, evident through the presence of substantial funds to buffer unforeseen expenses (X3) and infrequent negative balance occurrences (X5). Additionally, spikes in additional income sources (X6, X7) provide a larger financial cushion for transitioning toward relatively lower risk states. Notably, reduced transfer payments (X15), probably due to reduced miscellaneous expenses or a decrease in borrowing from friends and family, are associated with improvements.

Each type of improvement is linked to specific characteristics. The transition from high to medium risk is more likely when accounts practise effective financial management, including restricted spending (X4) to retain more funds and reduced dependence on debt (X9) to emerge from a prolonged negative state. The need to manage rising fixed and living costs (non-significant X11, X12) could explain why accounts often revert to a medium risk rather than fully recovering to a positive low-risk state.

Returning to a low-risk state (whether from high or medium risks) reveals contrasting behaviour compared to the transition from high to medium risk. The results indicate that X4 and X9 are significantly related but in opposite directions. A positive X4 suggests that accounts with a greater capacity and confidence to

Table 6: Estimated parameters for transition probability distribution (improvement)

| Covariate | Abbreviation | Latent states medium → low | high → medium | high → 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 $\alpha$ =0.10* 0.05** 0.01*** 0.001****

spend without difficulties are more likely to transition out of a negative status. The positive X9 implies that the probability of returning to a low risk increases when accounts can service more debt with their income. Alternatively, a full recovery from negative states, such as transitioning from high to low risk, is more likely when accounts increase their reliance on credit as a cash flow strategy to sustain a positive balance. The non-significant nature of X11 and X12 indicates that budgeting, potentially involving sacrifices in fixed or living expenses, increases the likelihood of regaining a healthy positive status.

Table 7: Estimated parameters for transition probability distribution (deterioration)

| | | Latent states | | |
|---|---|---|---|---|
| Covariate | Abbreviation | medium → high | low → medium | low → 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 $\alpha$ =0.10* 0.05** 0.01*** 0.001****

In contrast to financial improvements, deterioration (Table 7) is generally driven by a lack of financial resilience, often evident in frequent negative balance occurrences (X1, X2). However, for the low-to-high transition, the model could not capture significant covariates, probably due to the infrequent occurrences of this transition. For the other two transitions, low inflow fluctuations (X5) indicate a tendency for accounts to shift from a low to medium risk state. Accounts facing financial constraints and lacking extra earnings spikes struggle to accumulate funds, increasing the likelihood of overdraft usage. The significantly positive X12 explains that higher living costs increase the chance of transitioning from low to medium risk. The non-significant negative coefficients suggest that low-risk accounts are more inclined to shift to a medium-risk

state when they are compelled to reduce their fixed (X11) and discretionary expenses (X13) to prioritise essential spending (X10). This reflects struggling behaviour to maintain financial stability. On the other hand, the deterioration from medium to high risk is more likely to occur when accounts experience unexpected hardships, such as a dramatic income loss (X5, X6), and increased healthcare costs due to unforeseen life events (X8). The positive association between the medium-to-high transition and X9 signals the need for increased reliance on credit facilities when low income is insufficient to make ends meet.

In summary, our analysis highlights how distress risk states are influenced by various aspects of FV. Financial resilience is a pivotal factor impacting an account's distress risk. A deficit balance not only leads to a high distress risk initially but also results in transitions to worse financial states. In contrast, accounts with surplus balances usually begin in a low-risk state and can alleviate potential distress situations (medium/high risk) to low risk. Life events, whether positive (extra salaries) or negative (job loss, illness, accidents, etc.), play distinct roles in determining distress risk states. Imprudent spending habits despite low inflow increase the likelihood of high-risk states. Surprisingly, accounts that portray financial capability through proper money management practices may not necessarily revert to a low-risk state. This implies that budget planning, potentially requiring the reduction of certain fixed or living expenses, reflects a financially constrained scenario.

## 4.4 Predictions on test set accounts

### 4.4.1 Monitoring risks from the predicted sequence of states

The predicted sequence of states across the quarters, as presented in Table 8, serves as a valuable tool for risk management, enabling the monitoring of evolving vulnerable circumstances and potential distress. In the first part of Table 8, we observe the predicted number of test set accounts within each distress risk state for each quarter. Notably, there is a slight increase in the low-risk state, indicating a gradual shift towards financial stability. High-risk accounts exhibit a fluctuating trend, peaking in Q3 and subsequently decreasing in Q4, suggesting that some accounts experience short-term financial difficulties followed by a swift recovery. Meanwhile, the medium-risk state shows a consistent decreasing pattern, indicating a high tendency for accounts to transition into either a non-struggling or struggling state.

The second part of Table 8 summarises the transitions between states for the accounts in Q2-Q3 and Q3-Q4. Each cell in the transition matrix represents a shift from the previous state (row) to the current state

Table 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 → Q3 | medium | 331 | 149 | *92* |
| | low | *171* | 2555 | *38* |
| | high | 56 | 13 | 485 |
| Q3 → Q4 | medium | 307 | 189 | *62* |
| | low | *140* | 2558 | *19* |
| | high | 75 | 34 | 506 |

(column). Diagonal elements denote accounts that have remained in the same state, while the upper and lower triangular parts represent accounts that have moved to a different state in the subsequent quarter, indicating financial improvements or *deterioration*. These results offer insights into two strategies for tracking the tenacity of FV and assessing financial stress.

First, accounts that consistently remain in the same risk state across all quarters can be segmented into distinct stress levels: minimal exposure to vulnerability (low risk), financial restriction with a tight buffer (medium risk), or being trapped in critical money deficits (high risk) for an extended period. These stationary situations reflect long-term circumstances, and understanding these can inform tailored interventions for each group.

Second, the state transitions introduce complexity and provide insights into short-term financial adjustments and how accounts recover from or enter transient vulnerabilities. Monitoring these transitions, alongside the vulnerability drivers outlined in Section 4.3, can serve as a strategy to mitigate distress. For instance, improvements back into low risk (from high/medium risks) suggest the ability to achieve a stable financial status. However, it is essential to track the subsequent states following such transitions to ensure a lasting departure from temporary vulnerable scenarios. Examining the drivers behind these improvements could inspire ideas for short-term financial assistance, such as flexible credit schemes or interim incentives.

Deterioration signifies a downgrade in financial standards due to the occurrence of disruptive vulnerabilities. Transitioning to a high-risk state is a critical point of risk that warrants close monitoring in the following period. If accounts manage to recover, either immediately transitioning to low risk or gradually passing through medium and then low risk, this sequence reflects resilience in overcoming transient financial shocks. However, if accounts remain in a high-risk state, it indicates the presence of destructive and potentially permanent vulnerabilities that may eventually lead to genuine distress. Detecting the transient or permanent vulnerable characteristics can support more effective remedial actions to alleviate potential distress. Additionally, as the medium risk serves as an intermediary state that could worsen to high risk, a transition from low to medium risk should raise an alert to prevent transient vulnerability from becoming a permanent one. Understanding the specific vulnerable characteristics driving this transition can guide the establishment of preemptive strategies to guide temporary medium-risk accounts back to a stable position.

### 4.4.2 Predictive performance

Table 9 presents the predictions for the test set accounts at each quarter, with accuracy measures indicated in **bold**. In this table, the rows represent the labels of the latent states, while the columns show the model's predictions. Note that the model's objective is to predict the unobserved latent states, not the directly observed target variables.

Table 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** |

To derive the risk state labels for the test set, we rely on mappings between the observed and latent states, which are estimated using the emission probability matrix found in Table 3. Subsequently, we predict the sequence of latent states for each test set account, following the procedures outlined in Section 3.4.3, to obtain state sequences over the quarter.

The results indicate a high accuracy of approximately 97%. However, the confusion matrices reveal that the model tends to mistakenly predict high risk as medium risk. It is important to recognise that the high accuracy is partly attributed to the interplay between the computation of latent state target variables and the predictions, both of which heavily rely on the emission probability matrix.

# 5 Conclusion

We establish a comprehensive framework for assessing personal financial risk within the Open Banking domain. The latent Markov model is utilised to objectively evaluate the risk associated with current account transactions. Our approach focus on capturing dynamic financial vulnerabilities (FV) to predict the likelihood of an account transitioning to a less or more severe distress risk state in the upcoming quarter. These predictions are based on the vulnerable characteristics observed in the current quarter through cash flow transactions.

From the observed data on negative account balances, we have uncovered three latent states of distress, each representing a different degree of financial difficulty: low, medium, and high exposure risk to distress. The low-risk state signifies a stable financial position characterised by a sufficient positive balance and minimal overdraft usage. The medium-risk state indicates a minimal to mild level of financial difficulty. Accounts in this state rarely fall into low negative balances but usually maintain a positive balance (minimal) or employ low overdraft amounts as a strategic financial tool (mild). The high-risk state reveals a consistent reliance on overdrafts with low fluctuations involving high negative balances within the quarterly observation window. This data-driven approach enables an objective assessment of financial distress.

The relationship between vulnerability and financial transitions helps distinguish the different characteristics of improvement versus deterioration. Transitioning from a high-risk to a medium-risk state is associated with financial capability, including effective money management to restrict spending and reduce dependence on debt. Achieving a full recovery to a low-risk state requires financial resilience, with sufficient

income to service more debt or access to alternative sources of credit for cash flow turnover. Conversely, the deterioration from a low to a medium-risk state is driven by low financial resilience, where financially constrained accounts lack extra income to bolster their limited buffer. Unexpected hardships, such as sudden unemployment or the need for expensive medical treatments, are the primary vulnerability factors that lead to a transition toward a high-risk state. Understanding these vulnerability drivers helps increase or decrease the likelihood of transitions and provides valuable insights for effectively investigating specific issues that impact financial health.

The evolution of financial states alongside the vulnerability drivers reveals the tenacity of vulnerability in evaluating financial stress. Static financial states over the quarters depict long-term situations where accounts remain stable with minimal vulnerability (low risk), experience financial pressure with a limited buffer (medium risk), or face critical financial deficits (high risk). Temporary vulnerability is associated with improvements when an account completely moves away from high or medium risks to maintain a low-risk state thereafter. Permanent vulnerability refers to a non-reverting deterioration into a high-risk state. The transition from a low to a medium-risk state is not only the initial sign of temporary vulnerability leading to financial struggles but also the starting point for developing permanent vulnerability. A joint analysis of the predicted sequence of states and the vulnerability drivers empowers practitioners to implement remedial or preemptive actions tailored to specific short or long-term vulnerable characteristics, ultimately alleviating distress risk.

This research can be extended in several ways. While the primary focus here is on the out-of-sample prediction of financial transitions for new accounts, future work should consider out-of-time prediction if longer time-span data becomes available. The developed framework can also be applied to track long-term panel data, such as a five-year-long record, rather than solely relying on one year of transactional data. Additionally, the empirical analysis at the account level may not capture the holistic financial position of individuals with multiple accounts. Thus, the assessment could be extended to include consolidated account information for the same individual. Finally, the proposed framework can be integrated into credit risk models to support stress testing by incorporating time-varying vulnerable characteristics and distress risks.

# References


Bailey, D., del Pozo Mateos, L., Angulo Cardinale, J., van Woelderen, S. & Ahluwalia, S. (2022), 'Core indicators to measure financial health and inclusion', *United Nations Environment Programme Finance Initiative* .

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.

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.

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* .

CMA (2019), 'Consumer vulnerability: Challenges and potential solutions'.

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* .

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.

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

Ernst & Young, E. (2018), 'New definition of default'.

Experian (2019), 'Helping identify and pre-empt financial vulnerability'.

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', *Occasional Paper* **8**, 71–72.

Financial Conduct Authority, F. (2018), 'FCA mission: Approach to consumers'.

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

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

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* .

Gladstone, J., Nieboer, J. & Raghavan, K. (2020), Understanding consumer financial wellbeing through banking data, Technical report, Financial Conduct Authority.
**URL:** *https://www.fca.org.uk/publication/occasional-papers/occasional-paper-58.pdf*

Harrison, T. & Andreeva, G. (2021), 'Financial health of nhs workers', *https://static1.squarespace.com/static/5b9a99d8b98a78772c0d4ed6/t/601d6570f2af482b08fbcc98/1612539255770/uoe_r*
.

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.

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* .

Lando, D. & Skødeberg, T. M. (2002), 'Analyzing rating transitions and rating drift with continuous observations', *Journal of banking & finance* **26**(2-3), 423–444.

Leow, M. & Crook, J. (2014), 'Intensity models and transition probabilities for credit card loan delinquencies', *European Journal of Operational Research* **236**(2), 685–694.

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* .

O'Connor, G. E., Newmeyer, C. E., Wong, N. Y. C., Bayuk, J. B., Cook, L. A., Komarova, Y., Loibl, C., Ong, L. L. & Warmath, D. (2019), 'Conceptualizing the multiple dimensions of consumer financial vulnerability', *Journal of Business Research* **100**, 421–430.

Santos, J. P. N. (2018), Credit risk modelling using multi-state markov models, PhD thesis, Universidade Nova de Lisboa.

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).

Thomas, L. C., Allen, D. E. & Morkel-Kingsbury, N. (2002), 'A hidden markov chain model for the term structure of bond credit risk spreads', *International Review of Financial Analysis* **11**(3), 311–329.

Viterbi, A. (1967), 'Error bounds for convolutional codes and an asymptotically optimum decoding algorithm', *IEEE transactions on Information Theory* **13**(2), 260–269.

Wiggins, L. M. (1973), *Panel analysis: Latent probability models for attitude and behavior processes*, Jossey-Bass.

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.