# HOUSEHOLD LIQUIDITY NEEDS AND DEPOSIT STABILITY IN IRRBB: EVIDENCE FROM THE HOUSEHOLD FINANCE AND CONSUMPTION SURVEY

Diogo Gonçalo da Silva Vasconcelos Dores

Dissertation submitted for partial fulfilment of the requirements for the degree of "Master in Finance" (Mestrado em Finanças) at Universidade Católica Portuguesa.

January of 2026

Dissertation written under the supervision of Professor Diana Bonfim.

## Acknowledgements

I would like to express my gratitude to all the people who contributed to this work. I am thankful to the Professors of the Master in Finance at Universidade Católica for the intellectual challenges that shaped this thesis, and to my classmates who accompanied me throughout this journey. In particular, I am grateful to Duarte, Gustavo, and Rui, who always offered words of encouragement during the most challenging moments and patiently supported me both during the programme and the development of this thesis.

I am especially grateful to my supervisor, Professor Diana Bonfim, for her dedication, advice, and insightful suggestions, which allowed this thesis to take shape even when doubts and difficulties developed along the way.

I would like to dedicate this thesis to Miguel and Beatriz, as it was also for them that their father embarked on this adventure.

Finally, I would like to thank Joana, the pillar of my life. There are no words to fully express how important she has been throughout this journey, nor enough ways to express how grateful I am for her support in helping me achieve yet another personal goal. None of this would have been possible without her. Thank you.

# Abstract


Title: Household liquidity needs and deposit stability in IRRBB: Evidence from Household Finance and Consumption Survey

Author: Diogo Gonçalo da Silva Vasconcelos Dores

This thesis studies how household financial characteristics influence the behavioural risks associated with bank deposit stability under the Interest Rate Risk in the Banking Book (IRRBB) framework. Traditional behavioural assumptions for non-maturity deposits often rely on historical patterns or experts' decisions. This research provides an empirical basis for modelling depositor behaviour.

Using the 2021 wave of the Household Finance and Consumption Survey (HFCS), two models are estimated. The first predicts the probability that a household experiences a liquidity problem under adverse conditions. The second estimates, conditional on distress, the probability that households rely on deposit drawdowns rather than credit. These predictions are aggregated across relevant demographic and economic groups to produce behavioural run-off indicators for IRRBB.

The results show that liquidity pressure is more prevalent among lower-income households, those with higher debt-service ratios and those with weaker links to the labour market. On the other hand, households with greater liquidity buffers and higher wealth are more likely to use deposits when shocks occur, resulting in differentiated run-off risk across segments.

The findings highlight the value of incorporating microdata into deposit modelling practices and suggest that behavioural run-off risk is shaped by both the likelihood of households entering distress and their financial adjustment path once stress happens. This empirical evidence can support more informed calibration of behavioural assumptions for IRRBB stress testing and risk management.



Keywords: IRRBB; HFCS; deposit behaviour; household finance; liquidity risk.

## Resumo

Título: Household liquidity needs and deposit stability in IRRBB: Evidence from Household Finance and Consumption Survey

Autor: Diogo Gonçalo da Silva Vasconcelos Dores

Este trabalho analisa a forma como as características financeiras das famílias influenciam o risco comportamental associado aos depósitos bancários no contexto do *Interest Rate Risk in Banking Book* (IRRBB). Constata-se que os modelos tradicionais de estabilidade dos depósitos dependem, em larga medida, de pressupostos históricos ou de decisões discricionárias. Assim, este estudo procura fornecer uma base empírica fundamentada para a calibração de pressupostos comportamentais.

Utilizando dados de 2021 do Inquérito à Situação Financeira das Famílias (HFCS), estimaram-se dois modelos. Um para a probabilidade de uma família enfrentar restrições de liquidez e, condicionalmente, um segundo modelo para a probabilidade de recorrer ao levantamento de depósitos em detrimento da utilização de crédito. As probabilidades previstas são agregadas em indicadores de saídas de depósitos por segmentos económicos e demográficos relevantes para o IRRBB.

Os resultados demonstram que o risco de liquidez é maior em famílias com rendimentos reduzidos, rácios de dívida elevados ou emprego mais instável. Por outro lado, famílias com maiores reservas de liquidez e riqueza recorrem mais frequentemente à utilização de depósitos quando enfrentam choques financeiros. Estes padrões resultam na concentração do risco comportamental de saídas de depósitos em segmentos específicos.

O estudo evidencia o valor na incorporação de dados microeconómicos no processo de modelização de depósitos e sugere que a compreensão das vulnerabilidades das famílias pode reforçar a gestão do IRRBB, reconhecendo que o risco comportamental é determinado simultaneamente pela probabilidade de stress e pela forma de ajuste financeiro.

Palavras-chave: IRRBB; HFCS; depósitos; finanças das famílias; risco de liquidez

# Table of Contents

# List of Tables

# 1

## 1 - Introduction

Even nowadays retail deposits remain the foundation of bank funding. They are considered a stable, low-cost source of liquidity, supported by customer relationships and, in most countries, by the presence of deposit guarantee schemes. For these reasons, banks and supervisors treat retail deposits as a buffer that smooths balance sheet volatility and reduces funding issues. However recent episodes have shown that deposit behavior can change rapidly when they come across economic turmoil or when their own financial positions deteriorate. So, understanding behavioral drivers of deposit stability has become an important component of risk management, especially in the Interest Rate Risk in the Banking Book (IRRBB) framework.

Under IRRBB, banks must model how deposits volumes and margins respond to changes in interest rates. This requires making some assumptions about the stability of non-maturity deposits (NMD), including how they behave in adverse scenarios. The reality is that many of these assumptions have to rely on historical averages or internal behavioral models. While useful, these approaches do not incorporate household-level vulnerabilities and heterogeneity. Deposit outflows are ultimately driven by the decisions of individual households, whose liquidity constraints, tendencies and buffers vary. A more micro understanding of household behavior can complement existing exercises by providing knowledge on why certain depositor group might be more prone to withdrawals under stress.

This thesis contributes to this debate by using detailed household-level data to model the behavioural mechanisms that could lead to deposit drawdowns under difficult economic or financial conditions. The household information was collected with the 2021 wave of the Household Finance and Consumption Survey (HFCS), a rich, harmonised micro dataset which covers household income, assets, debts, expenditures, demographics and subjective financial assessments across European countries. This detailed structure and multiple imputation framework showed has the most suitable setting for capturing financial heterogeneity that represents behavioural risk in retail deposits.

The central objective of the thesis is therefore defined in two core research questions, which represent also two critical dimensions of deposit-related behaviour:

RQ1: Which household financial and demographic characteristics predict the likelihood of experiencing a liquidity shortfall in an adverse scenario?

RQ2: Conditional on liquidity pressure, which characteristics determine whether households rely primarily on deposit drawdowns (rather than credit) to absorb the shock?

By answering these questions, this thesis tries to provide a behavioural foundation for deposit segmentation, identifying which groups of households are most likely to contribute for deposit outflows during stress periods. These results are relevant for IRRBB, where correct assumptions about deposit stability can influence interest rate sensitivity, internal stress tests and the management of structural balance-sheet risk. The chosen strategy involves estimating two weighted logistic models using the HFCS. The first model predicts the probability of liquidity pressure. Throughout this thesis, a household is considered "under pressure" when it shows indicators of liquidity stress, namely insufficient saving capacity, a high debt-service ratio, or recent payment arrears, following the criteria presented in the HFCS.

The second model predicts coping strategy among households facing pressure, distinguishing who draws on deposits and those who turn to credit. Both models are estimated using survey weights and HFCS multiple-imputation structure to ensure representativeness and the treatment of missing data. Average Marginal Effects (AME) are used to quantify how each variable contributes to behavioural outcomes.

There is also a household-level predictions output derived from these models that determine two probabilities: Pliq, probability of liquidity pressure and Pdepdd, conditional probability of deposit drawdown. These combined probabilities determine a joint indicator that tries to approximate behavioural run-off risk. To make the outputs more IRRBB friendly, those probabilities were aggregated across meaningful economic segments, providing a structured view on how vulnerabilities and adjustment capacities are distributed across the retail depositor base.

Finally, from the aggregated segmentation was created a two-dimensional buffer-debt pivot table to give a simpler picture of the studied problem. Behavioural run-off risk arises from the interaction between financial vulnerability and coping capacity. These dynamics determine which depositor segment contributes most to retail outflows.

With this line of work this thesis wants to contribute by providing a micro-founded analysis of household liquidity vulnerability and deposit behaviour and offering a segmentation that can be used by IRRBB teams to benchmark behavioural assumptions and identify vulnerable depositor groups. It will be relevant as a complement for the regulatory run-off assumptions or banks' specific models by offering empirical evidence on the key drivers of deposit stability.

The structure of the thesis follows a logical progression. Chapter 1 introduces the relevant motivation for the study, outlines the research questions, and situates the problem within the broader context of deposit stability. Chapter 2 reviews the relevant literature, covering household financial vulnerability, coping behaviour, and the empirical and regulatory perspectives on deposit modelling. Chapter 3 provides the conceptual and regulatory background on IRRBB, explaining the risk categories, supervisory framework, embedded behavioural optionalities, and the relevance of depositor behaviour for balance-sheet risk management. Chapter 4 sets out the empirical methodology, describing the HFCS dataset, the construction of key variables, and the econometric models used to estimate liquidity pressure and coping behaviour. Chapter 5 presents the results, including the Average Marginal Effects, the household-level predictions, and the segment-level behavioural indicators relevant for IRRBB. Finally, Chapter 6 concludes by summarising the main findings, discussing their implications for deposit modelling and risk management, and outlining limitations and avenues for future research.

# 2

## 2 - Literature Review

Banks operate in a scenario defined by an array of risks, each of which can have a significant impact on their stability, profitability, and long-term viability. Among the most relevant risks are credit risk, market risk, liquidity risk, and interest rate risk. Operational and environmental risks, including cybersecurity concerns, are also increasingly relevant in today's financial environment. As the banking sector has become more complex and subject to increased regulation, the academic research has expanded to reflect this new reality, moving beyond traditional credit risk to explore how banks manage a wider diversity of risks to their stability and profitability.

Much of the research in this area has focused on how the structure of a bank's balance sheet makes it more sensible to changes in interest rates. Staikouras (2003) highlights the importance of the duration gap - the difference between the durations of a bank's assets and liabilities - based on foundational work by Samuelson and Hicks. When this gap is large, even smaller shifts in interest rates can have a significant impact on a bank's equity value. This paper laid critical work by trying to understand why bank's profitability depend on yield curve dynamics. On these insights, Fraser, Madura and Weigand (2003) provided an analysis of the factors that drive commercial banks' exposure to interest rate risk. A sample of U.S. banks balances were analyzed and revealed that the primary drivers of interest rate risk are not simply the size of the bank, but rather specific characteristics such as capital adequacy, dependency on non-interest income -fees and trading revenue - and the structure of both assets and liabilities. For example, banks with a higher proportion of floating-rate loans or stable current accounts are generally less exposed to interest rate swings, since it allows faster repricing. Inversely, banks that rely on non-interest income tend to exhibit great interest rate risk, reflecting the cyclical nature of those income streams. All things considered, risk exposure stems not from size, but from strategic choices in asset-liability structure. Managers and regulators can use basic balance sheet and income statement information to assess and manage the interest rate risk exposure of individual banks.

More recent research has also focused on how interest rate risk can influence the transmission of monetary policy. Gomez et al (2020), using a panel of U.S. banks found that financial institutions with a larger income gap – their assets reprice faster than their liabilities – have a better positioning when rates rise. These banks tend have stronger earnings growth and are less likely to cut back on lending during periods of monetary tightening. This study also showed that this effect is especially noticeable among smaller and more financially constrained banks, suggesting that the structure of the balance sheet can influence the effects of monetary policy on the real economy.

Supporting this, Beutler et al. (2020) investigated how banks' exposure to interest rate risk in the banking book (IRRBB) affects their lending behavior, particularly in response to changes in monetary policy. Using a panel of Swiss banks from 2005 to 2017, the authors construct a detailed measure of each bank's interest rate risk exposure based on the maturity mismatch between assets and liabilities. This research showed that banks with greater gap between maturities are more likely to reduce lending when interest rates rise. This not only affects bank profitability and solvency, but also has real macroeconomic consequences by influencing the overall availability of credit.

Both studies exposed that interest rate risk in banking book is a key channel through which monetary policy shocks are transmitted to the real economy primarily by shaping the lending behavior of banks.

In light of the growing significance of interest rate risk in the banking book, the development and implementation of risk management practices have become fundamental for banks seeking to preserve their stability and profitability. Esposito, Nobili, and Ropele (2015) examined some Italian banks during the global financial crisis and found that most maintained a cautious approach, keeping their interest rate risk below regulatory thresholds. Banks used several strategies, including adjusting the maturity structure of their assets and liabilities and using derivatives to hedge against shift in rates. Smaller banks favored integrated strategies that offset interest rate and credit risk simultaneously, while larger banks were more willing to take on additional risk when liquidity was abundant. This diversity in approach challenges the "one size fits all" risk models.

Another interesting point of debate in the literature is the whether banks' dependency on retail deposits helps protect them from interest rate risk. Drechsler, Savov, and Schnabl (2021) discuss that banks' ability to pay depositors rates that are less sensitive to market fluctuations act as a natural hedge, allowing them to hold long-term assets without taking on so much risk. This study challenges the conventional wisdom that banks' core function of maturity transformation – borrow short-term through deposits and lending long term – exposes them to interest rate risk. According to the study's results, U.S. banks, especially the large ones, are able to keep net interest margins and profitability stable across different interest rate environments.

However, DeMarzo, Krishnamurthy, and Nagel (2024) challenge this conclusion using also U.S. banks data. They found that while deposits with slow repricing do provide some stability, they do not create a negative duration for the bank's present value of future net interest income. While banks with low deposit rate beta do earn higher spreads as rate rises, the present value of future profits actually decline when interest rate increases. This refutes the common regulatory practice of treating deposits with low-rate sensitivity as a hedge against interest rate risk, concluding that, in fact, for most banks, rising interest rates reduce bank's value, increasing vulnerability. The authors recommend that internal risk models should be recalibrated to reflect these realities instead of relying on incorrect assumptions about deposits sensitivity and negative duration.

At a macro level, Gentil, Ray and Toader (2023) provided a model to forecast the impact of interest rate changes on banks' net interest income (NII) across the euro area. The context of this research is the recent period of high-interest rate volatility and the importance of bank profitability for global financial stability, as NII affects bank's ability to absorb losses and build capital through retained earnings. Applying the model to the euro area banking sector, the authors found that the structure of the banks' balance sheets as off 2022 – surplus of long fixed-rate assets and short or variable rate assets funded by non-interest-bearing liabilities – results in a positive sensitivity of aggregate NII to rising interest rates. In their scenario, high interest rates in the period 2023 – 2027 are projected to boost NII, reflecting the favorable structural position of euro area banks. This study also suggests that, while high-interest rate environment can boost profitability, the benefits depend on the stability of core deposits and the broader policy environment

All things considered, banking exposure to interest risk is such an important topic that it justifies rigorous investigation. Monetary policy, balance sheet structure, funding sources and the strategies used to manage risk all play a role in the final results. Academic research in this area has grown significantly over the past decades, reflecting both the complexity of the subject and its practical importance.

# 3

## 3 - IRRBB Framework and Behavioural Relevance

This chapter provides a structured overview of Interest Rate Risk in the Banking Book (IRRBB), beginning with basic definitions and information on its growing importance for both banks and regulators under the Basel capital framework. It explains how IRRBB arises from mismatches in the maturities and repricing of assets and liabilities, as well as from embedded options in banking products, and highlights the risks posed by customer behaviors and prolonged periods of low or negative interest rates. The chapter then outlines the main subtypes of IRRBB and discusses their implications for bank profitability and stability. It reviews the evolution of the regulatory framework, focusing on the roles of the Basel Committee on Banking Supervision and the European Banking Authority, and describes how banks are required to measure and manage IRRBB using both earnings-based and economic value-based metrics, such as gap analysis, ΔNII, and ΔEVE.

### 3.1 - IRRBB

Interest Rate Risk in Banking Book (IRRBB) became such an important topic for banks and also regulators, it is now part of the Basel capital framework's Pillar 2. In their IRR Principles, The Basel Committee on Banking Supervision defines it as the current or prospective risk the bank's capital and earnings arising from adverse movements in interest rates that affect the bank's banking book positions (BCBS, 2016). The fluctuation of interest rates impact both the present value and the timing of a bank's future cash flows, which affects the value of bank's assets, liabilities and off-balance sheet positions, ultimately influencing the overall economic value.

IRRBB is a critical risk for all banks, arising from the very nature of how they operate. Banks typically make money by acting as intermediaries: borrowing short-term (like deposits) to lend long-term (think mortgages). This creates two key mismatches:

- maturity mismatch - Using short-term liabilities (e.g., customer’s savings accounts) to fund long-term assets (e.g., 30-year mortgages)

- rate mismatch: Offering fixed-rate loans (which lock in a rate) but relying on variable-rate deposits (which fluctuate with market rates)

Adding to this, many everyday banking products like non-maturity deposits, term deposits, and fixed-rate loans have built-in options that customers can exercise when rates change. For example, borrowers might refinance a fixed-rate loan if rates drop, or depositors could withdraw funds early to chase higher yields elsewhere. These behaviours add layers of complexity to managing IRRBB.

If not properly managed, excessive IRRBB can threaten a bank’s capital and future profitability. For example, after the financial crisis, when interest rates began to fall, many banks saw their profit margins shrink because they did not hedge their banking book against declining rates. So, one of the biggest factors for IRRBB and its regulation has been the prolonged period of low and even negative interest rates, which also challenges the basic financial principle that money should have a positive time value.

Lubinska (2021) refers that when interest rates go negative, banks with a lot of extra cash end up facing higher costs. To make up for this, they often look for ways to recover those expenses-sometimes by increasing account fees or, in more extreme cases, by pulling back on lending to businesses and households. The way interest rates move is hugely important because it can influence both a bank’s profits and its overall risk in different ways. That is why regulators insist on banks using accurate and reliable methods to measure interest rate risk, so they can identify all the major sources of risk and understand how it might affect the bank’s day-to-day operations.

Profitability is still a major issue for banks. In some cases, profits are not only low but also vary widely, and high operating costs make profits even more difficult. This is especially true for banks in Europe with shrinking profits, increasing regulation, and ongoing financial risks. Baldan et al. (2012) proposed that cutting exposure to liquidity risk might also lower interest rate risk. Their study focused on a small Italian bank in 2009–2010, which had to adjust its liquidity strategy to meet Basel III requirements. Interestingly, this shift not only improved liquidity but also saw its interest rate risk drop. The researchers argue that banks should adopt integrated risk management, where handling one risk (like liquidity) directly informs strategic decisions and influences other risks (such as interest rate exposure). Their work highlights the need for a holistic approach to risk, rather than treating each type in isolation

## 3.2 - Regulatory framework

Since IRRBB has been getting a lot of attention in recent years, it is clear that the banking industry and its regulators are moving toward more standardized ways of measuring, modelling and monitoring this risk. There are two main regulatory bodies overseeing interest rate risk in the banking book (IRRBB): the Basel Committee on Banking Supervision (BCBS) and the European Banking Authority (EBA). The BCBS is an international body that sets global standards for banking regulation, aiming to strengthen the regulation, supervision, and practices of banks worldwide to enhance financial stability. The EBA, on the other hand, is an independent EU authority established to ensure effective and consistent prudential regulation and supervision across the European banking sector.

The first regulatory guidelines were introduced in 2004, when the Basel Committee on Banking Supervision (BCBS) published the "Principles for the Management and Supervision of Interest Rate Risk." This document defined the sources and effects of IRRBB and established 15 principles for its management. In 2006, the Committee of European Banking Supervisors (CEBS) issued "Technical aspects of the management of interest rate risk arising from non-trading activities under the supervisory review process," which further specified the BCBS guidelines for European application. CEBS guidelines were non-binding, but this changed with the transition to the European Banking Authority (EBA) in 2011, which was granted the authority to develop legally binding technical standards (Seidel, 2018).

The regulatory process intensified in May 2015, when the EBA published its "Guidelines on the management of interest rate risk arising from non-trading activities". This was a notable departure from the usual process, as the EBA released updated guidelines before the BCBS had finalized its global standards. In June 2015, BCBS issued another consultation document on IRRBB, and in April 2016 published its final standards, replacing the 2004 Principles. The new BCBS standards set out expectations for the identification, measurement, monitoring, control, and supervision of IRRBB, reflecting changes in market conditions and supervisory practices, especially in light of the prolonged low interest rate environment (Lubinska, 2021).

The final EBA IRRBB guidelines, known as the "Final Report," were issued on 19 July 2018, posing significant challenges for banks in terms of software solutions, metrics revision, and model validation frameworks. With these evolving guidelines and more restrictive supervisory expectations, banks have had to adapt their IRRBB strategies and implementation measures to comply with both global and European regulatory standards.

### 3.3 - IRRBB Risk management

According to (BCBS, 2016) there are three main subtypes of IRRBB that potentially change the price/value or earnings/cost of interest rate-sensitive assets, liabilities and/or off-balance sheet items:

- Gap Risk emerges from the term structure of banking book instruments, and describes the risk arising from the timing of instruments' rate changes. For example, a bank may hold an asset that reprices based on the 12-month Euribor, while funding it with a liability that pays a fixed interest rate. Because interest rate resets on various instruments occur at different tenors, a bank faces risk if the interest paid on liabilities rises before the interest received on assets adjusts, or if asset yields decrease before liability costs do. Without appropriate hedging of both tenor and amount, the bank may be exposed to periods of compressed or even negative interest margins, as well as fluctuations in the relative economic values of its assets and liabilities. Seidel (2018) adds another layer to this definition by referring that gap risk comprises two other risk which are yield curve risk and repricing risk. Yield curve risk results from changes in the shape of the interest rate term structure, which can occur as both parallel and nonparallel shifts. Repricing risk, on the other hand, occurs when the timing of interest rate adjustments differs between assets and liabilities.

- Basis Risk describes the impact of relative changes in interest rates for financial instruments that have similar tenors but are priced using different interest rate indices, i.e. this risk occurs when there is imperfect correlation between reference rates used to set interest payments on different instruments-despite having similar repricing schedules. For example, considering a bank that funds a loan by borrowing at the 3-month Euribor rate, while the interest earned on the loan is indexed to the 3-month €STR (Euro Short-Term Rate). Although both the asset and the liability have matching three-month repricing intervals, they are referenced to different benchmark rates. Should the spread between 3-month Euribor and 3-month €STR fluctuate due to market dynamics, the bank's interest income and expense may diverge, despite the alignment in repricing frequency. This lack of perfect correlation between similar-tenor reference rates constitutes basis risk.

- Option Risk arises from option derivative positions or from optional elements embedded in a bank's assets, liabilities and/or off-balance sheet items, where the bank or its customer can alter the level and timing of their cash flows. From a financial perspective, an option is an instrument that provides the holder the right but not the obligation to buy or sell the underlying asset. In the context of IRRBB, options commonly refer to features such as loan prepayment rights or deposit withdrawal rights, which allow customers or the bank to alter the timing or amount of cash flows. This risk can be categorized into two subtypes:

  - Automatic option risk that refers to risk arising from explicit options, either as standalone instruments (such as exchange-traded or over-the-counter option contracts) or embedded within the contractual terms of a standard financial product (for example, a capped rate loan). In these cases, the option is likely to be exercised whenever it is financially advantageous for the holder.

- Behavioural option risk arises from implicit or explicit flexibility within financial contracts, where changes in interest rates may influence client behaviour. Examples include a borrower's right to prepay a loan, with or without penalty, or a depositor's ability to withdraw funds in pursuit of higher returns elsewhere.

Automatic options, in contrast to behavioural options, are contractual features that are exercised strictly according to predefined conditions, following a rational pattern. Examples include interest rate floors, caps, or collars embedded in financial instruments.

Behavioural optionality is the most common in banking book. For example, current accounts present risk due to their uncertain maturities; depositors can withdraw funds at any time, which may differ significantly from the bank's assumed maturity profile. This represents behavioural optionality from the client's side. On the other hand, banks may retain the right to change the interest rates offered on these products, sometimes with little or no notice, introducing optionality from the bank's side.

Behavioural risk also exists on the asset side, particularly through early redemption (full or partial prepayment) options on mortgages and other customer loans. Borrowers may repay their loans ahead of schedule. This prepayment behavior tends to increase in low interest rate environments, as clients seek to refinance at more favorable terms.

Because customers may exercise these options in response to changes in market interest rates, they introduce a type of interest rate risk known as optionality, already discussed in the previous chapter. Managing this risk is particularly challenging in the retail banking sector, largely because customers behave so differently from one another when deciding whether to use these options. This diversity makes it crucial to develop behavioural models that link customer actions to movements in interest rates and that accurately reflect the variety of behaviours observed in practice.

Not all borrowers exhibit irrational behaviour, but patterns can be tracked that influence risk measurement. For example, early repayment is most common when interest rates fall, allowing borrowers to secure new loans at more favourable rates, while fixed-rate loans are rarely refinanced when rates rise.

Behavioural models play a pivotal role in asset and liability management as they allow banks to analyse how future cash flows and net interest income might change under different financial scenarios, including stress situations. Also, regulators now require banks to demonstrate this capability, as mentioned in the Basel (2001). If future cash flows are not estimated accurately, treasury and risk managers may end up relying on misleading indicators.

Since behavioural patterns are uncertain and subject to change, these models must be updated frequently. Therefore, risk metrics based on economic value and earnings should be regularly reviewed and adjusted, ensuring that calculations reflect any changes in expected customer behaviour.

Banks combine statistical data with behavioural assumptions to analyse customer behaviour. What makes these assumptions particularly challenging is that they aren't always aligned with market conditions. Sometimes customers act against their own financial interests to fulfil psychological needs, while other behaviours arise from rumours or lost confidence in financial institutions.

There are two events that are especially important for retail banks: **prepayment** on the asset side and **redemptions** on the liability side. Both are customer-exercised options that modify the cash flows, and both are triggered by a mix of price incentives (interest-rate moves, refinancing opportunities) and non-price factors (habits, confidence, convenience). Essentially, they push risk in opposite directions—prepayments shorten asset duration when rates fall, while redemptions shorten liability duration when funding is most valuable. The next sections therefore isolate these two behaviours.

### 3.3.1 - Prepayment risk

Prepayment risk describes the unpredictability regarding both the timing and amount of principal repayments on loans and mortgages. These products typically include an option for early repayment. From a standard point of view, borrowers agree to make monthly payments over the lifetime of the loan or mortgage. However, the contract often allows the borrower to pay off the balance ahead of schedule, sometimes without a penalty.

Borrowers might choose to prepay for a variety of reasons, including shifts in interest rates, the opportunity to refinance at a lower rate, changes in payment obligations, the sale of the property, or personal circumstances. Equally, the lender also has the right to terminate the contract if the borrower defaults on payments. Lenders face considerable interest rate risk when loans are prepaid and credit risk when defaults occur.

Prepayments disrupt the expected stream of cash flows for lenders. For example, when interest rates fall, prepayment activity tends to increase as borrowers seek better terms, forcing lenders to reinvest at lower rates. On the other hand, rising interest rates typically lead to fewer prepayments, as borrowers are less inclined to refinance. As a result, being able to accurately forecast prepayment behaviour is crucial for projecting future cash flows.

### 3.3.2 - Redemption risk

Redemption risk refers to the flexibility built into deposits. In the case of non-maturing deposits, this risk comes from the fact that customers can withdraw their funds at any time, since these accounts don't have a set maturity date. These deposits are the main funding source of retail banks and generally offer the lowest cost among funding sources.

Term deposits, on the other hand, present a different kind of redemption risk. They typically offer a fixed interest rate for a defined period. Although they are designed to be held until a specific maturity date, many allow customers to withdraw their money early, usually for a penalty. This early withdrawal option introduces another layer of uncertainty for banks. According to Basel (2001) It is essential to specify whether a term deposit is subject to redemption penalties or controlled with other contractual protections that help maintain the expected cash flow characteristics of the instrument.

A brief note to highlight that in most cases, penalties for early withdrawal are not calculated to match the true economic impact of breaking a deposit's contract. Instead, banks use more simplified approaches, such as part of the accrued interest, regardless of the actual economic cost incurred by the institution. This practice may lead to a disconnect between the penalty imposed and the real financial impact suffered by the bank.

Several factors can influence how embedded options in term deposits are exercised. These include the size of the deposit, characteristics of the depositor, the channel through which the funds are placed, the applicable interest rates, as well as seasonal influences, geographical considerations, competitive dynamics, the time remaining until maturity, and past behavioural patterns. Because both non-maturing and term deposits play such a central role in bank funding, it is essential to model their behaviour accurately. This is was the main driver to develop this work.

# 4

## 4 - Methodology

Building on the IRRBB concepts discussed in the previous chapter, the established goal for this thesis was to link household's liquidity risk and coping behaviour to bank deposit stability assumptions under IRRBB. With this mind two research questions were formulated:

1) Which households are more likely to face liquidity pressure?
2) Conditional on being under liquidity pressure, how do households tend to cope?

These two questions capture the behavioural mechanisms that matter for deposit stability under IRRBB. The first identifies which households are at risk of becoming stressed, which determines the size of the population that may need to withdraw funds in adverse conditions. The second focuses on how these households adjust once stress occurs, distinguishing between those who rely on deposits drawdowns and those who turn to credit.

Answering these questions requires a dataset that contains detailed information on households' balance sheets, income and indebtedness. The first stage in the methodology was therefore identifying some suitable public data to work with. In that line of investigation, we choose the Household Finance and Consumption Survey (HFCS) which is a euro-area harmonized survey, coordinated by the European Central Bank (ECB). It collects nationally representative microdata on households' financial information on all euro-area countries plus Croatia, Czech Republic and Hungary. In the 2021 wave this report took a sample of over 83,000 households providing structural information on asset and liabilities that can be used to analyse investment and consumption decisions.

The HFCS has two parts: (i) the household-level modules on real assets and their financing, other liabilities/credit constrains, financial assets, consumption and (ii) person level modules on demographics, employment, pensions and income. This kind of information suits the problem we wish to address. We will be able to connect who households are, how they tend to behave under pressure so that, in some extent, be able to duplicate the behavioural optionality that banks face, which is crucial for deposit stability and IRRBB.

From a simple perspective, the analysis was made in two steps. First estimate the probability that a household experienced some liquidity distress using the HFCS data with survey weights and multiple imputations. A pressure indicator was used to flag households with "no saving room" or a "debt-service ratio" higher than 40%. Some arrears measures were included as an additional trigger. Second, now that we have the households flagged as "under pressure", the next step was to estimate the probability that they covered this deficit by withdrawing deposits rather than borrowing, using the HFCS coping variables, distinguishing savings-leaning from credit-leaning households.

The explanatory variables are grouped into several blocks: liquidity buffers and deposit mix (net liquid assets-to-income and share of sight deposits), debt service (mortgage debt service ratio and total debt service ratio), income level and demographic (employment status, education, household size and education). Country fixed effects are included to mitigate differences across economies and banking markets.

For the estimation and outputs phase, weighted logistic regressions were used for liquidity pressure and multinomial regressions for the coping choice. The outputs are documented in several reports with segment-level probabilities of (i) facing a liquidity distress and (ii) solving it by withdrawing deposits and average marginal effects. These estimates are translated into IRRBB inputs: run-off propensities and deposit betas by household segment in order to support stress testing and behavioural assumptions in the banking book.

### 4.1 - Variable definitions and constructions

The model's several variables will now be detailed. Beginning with the **dependent variables:** first, whether the household faced a liquidity pressure; second, for those under pressure, how they coped with it. The **explanatory or independent variables** that describe each household's situation: how much money they hold and how easily it can be accessed, overall debt load, earning capacity and stability. Also important is the **household structure** and **housing tenure** to explain differences in fixed costs and financial flexibility. Finally, **country**

**indicators** are used to absorb time-varying cross-country differences. Table 4.1 lists names, HFCS codes and brief definitions of the variables.

### 4.2.1 - Dependent variables

Liquidity pressure (liq_press) is the primary outcome, being defined as a binary variable for whether the household was under liquidity distress when at least one condition is true: no saving room (doabletosave = 0), debt-service ≥ 40% (dodstotal40 =1) and where available, arrears (arrears_any = 1).

Coping channel under pressure explains the household's first line of defence when liquidity pressure exists. So, among households flagged as pressured, the coping actions will be proxied with a tri-category indicator (cope_cat) built from balances and credit signals. Households will be classified as savings-leaning (cope_cat =1) if deposits exist (dnnlaratio ≥ 1 or da2101 ≥ 0) and no financial constraint due to credit; credit-leaning (cope-cat =2) if any of dl1220i, dl220i, docredic, docreditrefused = 1; other (cope_cat = 0).

### 4.2.2 - Explanatory variables

#### 4.2.2.1 - Cash buffers and deposits

These variables will reflect the household's balances in bank accounts. Larger balances should reduce the likelihood of stress and raise the probability that a pressured household draws on savings. The dnnlaratio_cap (net liquid assets / income, capped [0,10] to remove extreme outliers) reflects how the household buffer covers its budget needs. The sight_share variable. (da21011/da21012) captures the allocation of sight deposits in total deposits. A higher sight-share indicates easier access to spendable funds and therefore more willingness to withdraw deposits under stress.

#### 4.2.2.2 - Debt burden and leverage

These variables summarise claims on income and leverage, both of which amplify stress when income lowers or rates rises. The variable dodsmortg (mortgage service / income) reflects the impact of scheduled mortgage payment on income.

Leverage or debt-to-income (dodiratio) tracks total outstanding debt relative to income. High leverage makes households more exposed to rate shocks, increasing pressure risk and, in coping, may limit access to credit.

#### 4.2.2.3 - Income, employment, and demographics

This group of variables summarise a household's earning capacity, income risk and human capital. This is the way the model can attribute profiles, making it explicit which types of households do what when they suffer liquidity issues. Gross household income (di2000) as an overall budget capacity, Employment status (dhemph_1) to capture earnings stability and access to credit, as unemployment or precarious work is typically associated with higher liquidity risk and a greater dependency on credit or payment deferrals if savings are insufficient. Education (dheduh1_p) is also a proxy for permanent income and financial resilience since higher education tends to be correlated with lower risk. The equivalence scale (dh0002) controls size and needs, making comparisons fairer between family types, since larger or more dependent households face, for example, higher routine expenses. With these variables it is possible to map which family profiles are more vulnerable to pressure and how they tend to respond when it happens.

The last variables are housing tenure (dhhst_p) which serves as a proxy for fixed housing costs, collateral and access to credit. Mortgagors face regular debt service and owners often have the lowest liquidity pressure risk. On the other hand, tenants can face higher liquidity risk due to rent obligations and have less collateral. The degree of urbanisation (dhdegurba_p) is also important to capture differences in living cost and labour markets. Country fixed effects (sa0100) are also included in the model to absorb structural differences across banking markets and make cross-country estimations comparable.

Table 4.1 - Variables used in the study - definition, role and interpretation

| Group | Variable | Description | Source/Construct |
|---|---|---|---|
| Dependent / Prediction | liq_press | Liquidity pressure flag: 1 if household had no saving room OR debt-service ≥40% (plus arrears if available). | Built from doabletosave, dodstotal40p, arrears_any. |
| Dependent / Prediction | p_liq | Predicted probability of liquidity pressure. | Fitted per-imputation, averaged across m. |
| Dependent / Prediction | depdd_bin | Deposit drawdown among pressured (1 = withdrew savings; 0 = otherwise). | Constructed from cope_cat==1 among liq_press==1. |
| Dependent / Prediction | p_depdd | Predicted probability of deposit drawdown given pressure. | Fitted among liq_press==1, averaged across m. |
| Regressor (continuous) | dnnlaratio_cap | Net liquid assets / income (capped 0..10). | From dnnlaratio with capping rules. |
| Regressor (continuous) | sight_share | Sight deposits share of total deposits. | da21011 / da2101 (with robust fallbacks). |
| Regressor (continuous) | dodsmortg | Mortgage debt service ratio. | From HFCS debt-service items. |
| Regressor (continuous) | dodiratio | Total debt-to-income. | From HFCS debt stock and income. |
| Regressor (continuous) | di2000 | Gross household income. | HFCS income aggregate. |
| Regressor (continuous) | dh0002 | Equivalence scale (household needs adjustment). | HFCS equivalence factor. |
| Regressor (categorical) | dhemph1_p | Employment status (cleaned for factor use). | From dhemph1; negatives set to missing; value labels preserved. |
| Regressor (categorical) | dheduh1_p | Education of reference person (cleaned). | From dheduh1; negatives set to missing; labels preserved. |
| Regressor (categorical) | dhhst_p | Housing tenure (cleaned). | From dhhst; negatives set to missing; labels preserved. |
| Regressor (categorical) | dhdegurba_p | Degree of urbanisation (cleaned). | From dhdegurba; negatives set to missing; labels preserved. |
| Regressor (categorical) | ctry | Country (encoded). | Encoded from sa0100. |
| Builder / component | not_saving_room | No saving room flag (doabletosave==0). | From doabletosave. |

| | | | |
|---|---|---|---|
| Builder / component | dsr40 | Debt service ≥40% (all indebted). | From dodstotal40p==1. |
| Builder / component | arrears_any | Any arrears available in delivery; else set to 0. | From hc1250 or hcarears, else 0. |
| Builder / component | cope_sav_proxy | Signals savings-leaning (buffers/deposits). | dnnlaratio_cap>=1 or da2101>0. |
| Builder / component | cope_cred_proxy | Signals credit-leaning (line/card/constraints/refusal). | dl1210i==1 or dl1220i==1 or docreditc==1 or docreditrefused==1. |
| Builder / component | cope_cat | Coping tendency under pressure (tri-category). | Deterministic from proxies: 0 other, 1 savings-leaning, 2 credit-leaning. |
| ID / Weight | sa0100 | Country identifier (string in raw HFCS). | HFCS ID; encoded to ctry for factors. |
| ID / Weight | sa0010 | Household identifier. | HFCS ID. |
| ID / Weight | hw0010 | Sampling weight. | HFCS design weight. |
| Builder (deposits) | da2101 | Total deposits (sight + savings). | HFCS deposit totals (with fallbacks). |
| Builder (deposits) | da21011 | Sight deposits. | HFCS deposit detail. |
| Builder (deposits) | da21012 | Savings/time deposits. | HFCS deposit detail. |
| Builder (deposits) | da2100 | Any deposits (fallback denominator). | HFCS deposit any. |

### 4.2 - Stata code walkthrough

All the code was written and run in Stata (check appendix for more details) with a working sample of 83,162 unique households with multiple imputation (m=4) used in modelling steps. HFCS methodological documentation states that multiple imputation (MI) is recommended to handle non-response on balance-sheet items and therefore avoid biased estimates and coefficients and minimize standard errors. Since HFCS is a complex survey, where countries oversample wealthier households to ensure adequate coverage of wealth in each country, a calibrated household weight was used to fix those design features and avoid overoptimistic standard errors.

The code begins by assembling the HFCS microdata into a MI dataset. The household identifiers are used as MI panel keys and the survey weight is also preserved as mentioned earlier. The next stage is just basic harmonisation: country is encoded into a numeric factor and common HFCS special codes are cleaned so that categorical regressors can be used. This data is place under Stata's survey framework with survey weight adjustment, assuring that descriptive statistics and estimates reflect correct population weights.

With all the data in the framework, the code constructs the financial covariates and outcome variables used in the estimation. All the balance sheet, debt and demographic variables are created and the country fixed effects are added to the mix.

Outcomes are built following a predetermined rule-based process. The primary dependent variable is liquidity pressure, while a second construct classifies coping behaviour conditional on pressure, distinguishing between a savings-leaning proxy and leaning proxy. For estimation stability and direct relevance to IRRBB, the code also forms a binary variable among pressured households, depdd_bin, which equals one when the coping channel is deposit drawdown.

With all the variables in place, the script estimates two models: RQ1 which is a survey-weighted logistic regression for the probability of liquidity pressure with all the financial covariates and demographic factor, and RQ2 that focuses on behaviour among pressured households. It estimates a survey-weighted logistic regression for the probability of deposit drawdown vs other coping actions. Predictions are now generated at the household level per implicate to produce two final behavioural measures for every household: Pliq and Pdepdd.

To help the interpretation, the code produces two Average Marginal Effects files. Associated with RQ1, quantifies how a one-unit change in the variable shifts the average probability of distress, and how each demographic category differs from its base group. For RQ2, the AME show which characteristics move a pressured household towards using deposits rather than credit. Results are saved in Excel workbooks.

After this routine, there is a dedicated data validation routine, reporting the MI structure and inspecting survey weights, exporting weighted means overall and by country, building income quintiles to show how distress and covariates move in the distribution.

In addition, the code compiles the household-level predictions in a file used for IRRBB translation. It takes the non-imputed extract (m=0) as a carrier of identifiers and weights, merges in the averaged predictions Pliq and Pdepdd and saves both an Excel version and a Stata dataset. These two probabilities are the key outputs because, when combined and aggregated over segments, they can give run-off propensities that can feed directly into banking-book stress tests and behavioural deposit assumptions.

Finally, the analysis produces a complementary workbook which aggregates these predictions into run-off indicators. This file groups households into relevant segments and delivers weighted averages of the predict probabilities of liquidity stress and deposit drawdown. Visually these data are displayed in tables with a “heat map” colour scheme to highlight relative risk intensities, providing an intuitive way to compare stability profiles across segments and therefore showing itself useful for the calibration of behavioural assumptions in banking-book stress tests.

# 5

## 5 - Outcomes and analysis

The results of this study are compiled in average marginal effects for liquidity pressure (RQ1), and the coping option (RQ2) and also the household-level predicted probability. The AME workbooks manage to quantify how liquidity buffers, debt and demographics influence risk and coping decisions. The predictions file provides, for each household, the estimated probability of liquidity pressure and, conditional on that pressure, the probability of covering that issue by withdrawing deposits. Combining these two numbers, it is possible to obtain a run-off propensity that can be relevant for IRRBB stress testing.

### 5.1 - Average Marginal Effects for RQ1

This first table answers the question "Who is under pressure?" reporting average marginal effects, on probability scale: how many percentage points the risk changes, when the variable moves by one unit, holding the rest fixed.

Table 5.1.1 - RQ1 AME for financial variables

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Liquidity buffer | 0.0039 | 0.0085 | 0.4587 | 0.6464 | -0.0128 | 0.0206 |
| Sight in total deposits | -0.0415 | 0.0235 | -1.7622 | 0.0781 | -0.0876 | 0.0047 |
| Mortgage debt-service | 0.7394 | 0.0721 | 10.2530 | 0.0000 | 0.5980 | 0.8807 |
| Debt-to-income | 0.0064 | 0.0043 | 1.4916 | 0.1358 | -0.0020 | 0.0148 |
| Gross household income | 0.0000 | 0.0000 | 1.0606 | 0.2889 | 0.0000 | 0.0000 |
| Equivalence index | -0.0001 | 0.0137 | -0.0066 | 0.9947 | -0.0270 | 0.0268 |

*Notes: Average Marginal Effects from the RQ1 survey-weighted logistic regression estimating the probability of liquidity pressure (Pliq), Pliq is defined under the rule: no saving room OR DSR ≥40% OR arrears. AMEs reflect 1-unit changes in financial covariates.*

The key factor is mortage debt-service (ddodsmortg). The AME is +0.739 with high significance which means that a 10% increase in the ratio is associated with a 7.4% probability of liquidity pressure. This confirms the idea that high, regular, debt payments reduce the financial cushion to face unexpected expenses or cash-flows fluctuations.

There is suggestive evidence that the proportion of sight deposits (sight_share) plays a helpful role in liquidity issues. Its AME is about -0.041 with a p-value close to 0.08. This seems to imply that a household with a more ready to spend funds lowers the pressure risk, which makes sense as households holding more easily access liquidity are generally better able to handle financial shocks.

On the other hand, the other variables like liquidity buffer level (dnnlaratio_cap), debt-to-income ratio (dodiratio), income (di2000) or equivalence index (dh0002) do not show a clear independent effect on whether a household reports pressure.

Bottom line, households' cash-flow issues are associated to the mortgage debt, with a higher amount of available associated with slight lower pressure risk. This pattern is consistent with the stress mechanism driven by monthly payments obligations and immediate access to cash that is consistent with IRRBB run-off logic.

The next tables show which household demographics characteristics are associated with a higher or lower chance of being under liquidity pressure. The variables are presented by level relative to a baseline category - first row in each table, since these are categorical variable.

Table 5.1.2 - RQ1 AME for employment

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Employee | 0.0000 | | | | | |
| Self-employed | 0.0205 | 0.0220 | 0.9309 | 0.3519 | -0.0227 | 0.0637 |
| Unemployed | 0.0710 | 0.0609 | 1.1657 | 0.2438 | -0.0484 | 0.1904 |
| Retired | 0.0067 | 0.0336 | 0.2007 | 0.8409 | -0.0591 | 0.0725 |
| Other inactive | 0.1918 | 0.0582 | 3.2984 | 0.0010 | 0.0778 | 0.3058 |

*Note: AMEs for categorical employment status from RQ1 logistic regression. Effects reflect differences relative to the base category ("Employee").*

Using Employee as the reference category, the results show a clear vulnerability among households where the reference is "Other inactive" (e.g. students, homemakers, long-term sick). The probability of reporting liquidity pressure is 19.2% higher than for employees. The other variables are positive but statistically not relevant. This means that a household with weak, irregular or lacking earnings capacity are more vulnerable.

Table 5.1.3 - RQ1 AME for education

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Primary or below | 0.0000 | | | | | |
| Lower secondary | 0.1133 | 0.0518 | 2.1886 | 0.0286 | 0.0118 | 0.2148 |
| Upper secondary | 0.0524 | 0.0424 | 1.2362 | 0.2164 | -0.0307 | 0.1354 |
| Tertiary | 0.0537 | 0.0416 | 1.2907 | 0.1968 | -0.0279 | 0.1354 |

*Note: AMEs for education levels from RQ1 logistic regression. Effects represent differences relative to the base group ("Primary or below").*

Regarding the education of the household reference person, with "Primary or below" as the base, "Lower secondary" shows statistic association with liquidity pressure. The other two variables are not statistically relevant, suggesting that the lower the education the more households are exposed to liquidity pressure.

Table 5.1.4 - RQ1 AME for housing tenure

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Owner outright | 0.0000 | | | | | |
| Owner w/ mortgage | -0.0652 | 0.0236 | -2.7561 | 0.0059 | -0.1115 | -0.0188 |
| Tenant/other | -0.0141 | 0.0417 | -0.3390 | 0.7346 | -0.0958 | 0.0676 |

*Note: AMEs for housing tenure from RQ1 logistic regression. Effects represent differences relative to the base group ("Owner outright").*

Relative to housing tenure, the base variable is "Owner outright" and the "Owner with mortgage" shows negative AME, while "Tenant/other" has no significance. This result looks counterintuitive at first. However, mortgaged households tend to be younger, strongly connected to the labour market and often with higher income than outright owners, a group that includes retirees on fixed or lower incomes. With this in consideration, the lower vulnerability of mortgaged households becomes more understandable.

Table 5.1.5 RQ1 AME for urbanisation

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| City | 0.0000 | | | | | |
| Towns & suburbs | 0.0007 | 0.0190 | 0.0380 | 0.9697 | -0.0365 | 0.0379 |
| Rural | 0.0024 | 0.0191 | 0.1234 | 0.9018 | -0.0350 | 0.0397 |

*Notes: AMEs for Urbanisation category from RQ1 logistic regression. Effects represent differences relative to the base group ("City").*

Using "City" as base, the differences for the other variables are essentially zero and statistically irrelevant, meaning that living in the suburbs or in a rural area does not meaningfully change the odds of being pressured relative to a city household. Geography does not seem to be a useful segmentation for IRRBB behaviour, at least in this dataset.

## 5.2 - Average Marginal Effects for RQ2

The second output focuses on households that are already pressured and determines who can solve it by drawing deposits rather than turning into credit or delaying payments. The structure mirrors the one displayed in RQ1.

Table 5.2.1 - RQ2 AME for financial variables

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Liquidity buffer | 0.0653 | 0.0219 | 2.9827 | 0.0029 | 0.0224 | 0.1082 |
| Sight in total deposits | -0.0028 | 0.0275 | -0.1015 | 0.9191 | -0.0566 | 0.0510 |
| Mortgage debt-service | -0.0006 | 0.0003 | -2.1649 | 0.0304 | -0.0011 | -0.0001 |
| Debt-to-income | 0.0000 | 0.0000 | -2.6251 | 0.0087 | 0.0000 | 0.0000 |
| Gross household income | 0.0000 | 0.0000 | 3.6751 | 0.0002 | 0.0000 | 0.0000 |
| Equivalence index | -0.0197 | 0.0132 | -1.4954 | 0.1348 | -0.0455 | 0.0061 |

*Notes: Average Marginal Effects from the RQ2 logistic regression, estimated on the subsample of households flagged as pressured. The dependent variable is deposit drawdown (Pdepdd). AMEs reflect 1-unit changes in financial covariates.*

Two results are clear. First, larger liquidity buffers favour behaviours toward using savings, as one unit rise in the liquidity buffer ratio (ddnlaratio_cap) raises the probability of a savings-leaning reaction by 6.5%. This points that household with bigger liquidity cushions are more likely to cover the problems with their own money.

Second, the indicators of debt pressure point the other direction as the higher mortgage burden (dodsmortg) is associated with a lower chance of drawing savings. The debt-to-income ratio (dodiratio) also shows a negative AME. These patterns confirm the idea that more indebted households lack sufficient buffers and are pushed toward credit or payments deferrals rather than drawing deposits.

The remaining covariates are less significant for the study. In the essence, among pressured households the coping choice is like a balance sheet decision: more money relative to income, more deposits drawdowns; more debt pressure, less deposits drawdown.

As in RQ1, the next tables show how household demographic characteristics influence the choice to drawdown deposits.

Table 5.2.2 - RQ2 AME for employment

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Employee | 0.0000 | | | | | |
| Self-employed | -0.0618 | 0.0248 | -2.4961 | 0.0126 | -0.1103 | -0.0133 |
| Unemployed | 0.0068 | 0.0445 | 0.1518 | 0.8793 | -0.0805 | 0.0940 |
| Retired | 0.0246 | 0.0292 | 0.8425 | 0.3995 | -0.0326 | 0.0818 |
| Other inactive | -0.0527 | 0.0607 | -0.8680 | 0.3854 | -0.1718 | 0.0663 |

*Note: AMEs for employment categories from RQ2 logistic regression. Base category is "Employee". Effects indicate differences in the conditional probability of drawing deposits among pressured households.*

Employment status is measured relative to "Employee". The only relevant difference between status is for the "Self-employed", with an AME of 6% less likely than employees to use deposits to cope to liquidity pressure. That fits the idea that business owners often rely more on credit lines. The results for "Unemployed", "Retired" and "Other inactive" do not show evidence that these groups differ from employees in their deposit drawing behaviour.

Table 5.2.3 - RQ2 AME for education

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Primary or below | 0.0000 | | | | | |
| Lower secondary | 0.0301 | 0.0472 | 0.6371 | 0.5241 | -0.0624 | 0.1225 |
| Upper secondary | 0.0031 | 0.0446 | 0.0684 | 0.9454 | -0.0844 | 0.0906 |
| Tertiary | 0.0153 | 0.0434 | 0.3539 | 0.7234 | -0.0696 | 0.1003 |

*Note: AMEs for Education-level from RQ2 logistic regression. Base category is "Primary or below". Effects indicate differences in the conditional probability of drawing deposits among pressured households.*

Regarding education, with “Primary of below” as baseline, there were not detected reliable differences to the other variables in the probability of using deposits rather than other coping channels.

Table 5.2.4 - RQ2 AME for housing tenure

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| Owner outright | 0.0000 | | | | | |
| Owner w/ mortgage | -0.0439 | 0.0203 | -2.1666 | 0.0303 | -0.0837 | -0.0042 |
| Tenant/other | -0.1098 | 0.0499 | -2.2001 | 0.0278 | -0.2076 | -0.0120 |

*Note: AMEs for Housing tenure from RQ2 logistic regression. Base category is “Owner outright”. Effects indicate differences in the conditional probability of drawing deposits among pressured households.*

Housing tenure uses “owner outright” as base variable and the results show an interesting contrast. People who own their homes with a mortgage are about 4.4% less likely to withdraw deposits compared to those who own their homes outright. Tenants shows even less tendency to use their deposits with 11%. Intuitively, outright owners tend to hold bigger balances while mortgages and tenants may rely more on credit or payment delay.

Table 5.2.5 - RQ2 AME for urbanisation

| | AME | std error | t-stat | p-value | lower limit | upper limit |
|---|---|---|---|---|---|---|
| City | 0.0000 | | | | | |
| Towns & suburbs | -0.0054 | 0.0188 | -0.2892 | 0.7724 | -0.0423 | 0.0314 |
| Rural | -0.0190 | 0.0187 | -1.0190 | 0.3082 | -0.0556 | 0.0176 |

*Note: AMEs for urbanisation categories from RQ2 logistic regression. Base category is “City”. Effects indicate differences in the conditional probability of drawing deposits among pressured households.*

The degree of urbanisation doesn’t show a clear association with the propensity to use deposits as the first coping response.

## 5.3 - Household predictions

In order to translate the model’s outputs, intro a structure that helps shape IRRBB behavioural assumptions, the analysis begins with the Household_predictions file. The dataset contains each household’s observations and consolidates the two studied probabilities: likelihood of facing a liquidity issue and conditional on being on pressure, the probability of covering that limited liquidity with deposit drawdowns rather than credit. This file also has the main segmentation variables and with all this information, seven separate excel files were produced, each summarising implied run-off rates for the main segmentation variables.

With this analysis becomes possible to identify which household groups are more prone to liquidity stress and more likely to drawdown deposits, offering a structured way to calibrate behavioural assumptions and benchmark internal stress-testing parameters.

Table 5.3.1 - Segmentation by income

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| Q1 (low income) | 0.76 | 0.86 | 0.73 |
| Q2 | 0.63 | 0.91 | 0.67 |
| Q3 | 0.60 | 0.91 | 0.63 |
| Q4 | 0.59 | 0.93 | 0.62 |
| Q5 (high income) | 0.51 | 0.94 | 0.61 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to income quintiles.*

The income profiles show an obvious gradient in the probability of experiencing liquidity stress. Low-income households are 25% more likely to face a liquidity stress than high income households. This an obvious pattern, consistent with the idea that limited and more volatile resources are connected with poorer households, leaving them with less ability to absorb shocks without reaching for their liquidity buffers.

The conditional probability of drawing down deposits shows just the opposite. When high income households face liquidity pressure, they are more inclined to use deposits as the primary coping action, consistent with having larger and more accessible buffers.

Combining these two dimensions as joint run-off measure shows that overall run-off propensity is still highest among low-income households. From an IRRBB perspective, this means that deposit portfolios with a greater concentration of low-income households are structurally more exposed to outflows under stress. On the other hand, deposits held by higher-income households show more stability because they are are unlikely to reach a point of liquidity distress.

Table 5.3.2 - Segmentation by wealth

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| Q1 (low wealth) | 0.58 | 0.88 | 0.57 |
| Q2 | 0.56 | 0.90 | 0.59 |
| Q3 | 0.57 | 0.92 | 0.63 |
| Q4 | 0.59 | 0.94 | 0.65 |
| Q5 (high wealth) | 0.62 | 0.95 | 0.70 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to wealth quintiles.*

When households are grouped by wealth quintiles, the pattern is different from the income gradient as the probability of a liquidity distress is relatively flat across the wealth distribution. This suggests that higher net worth does not shield households from cash-flow pressures. By contrast, the conditional probability of using deposits rises with wealth.

The joint run-off measure therefore increases strongly with wealth, implying that, even though liquidity distress should not be common among higher wealth households, when facing stress, they are more likely to respond by drawing deposits instead of using credit. This clearly points out a run-off issue concentrated at the top of the wealth distribution.

Table 5.3.3 - Segmentation by debt-service ratio

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| <20% of income | 0.43 | 0.92 | 0.53 |
| 20–40% of income | 0.48 | 0.92 | 0.57 |
| >40% of income | 0.60 | 0.92 | 0.64 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to DSR bands.*

The segmentation by debt-service ratio (DSR) shows a relationship between debt and liquidity stress. Households that allocate less than 20% of their income to debt have a smaller probability of distress. This is not surprising since heavy financial obligations are a key driver of vulnerability since a larger share of income is committed to debt. However, the conditional probability of adjusting through deposits is flat across the DSR bands. As a result, the run-off measure mirrors the Pliq outcome, showing more predominance in households whose debt payments exceed 40% of income. There is an obvious behavioural run-off risk in portfolios concentrated in highly indebted households as these borrowers are more prone to liquidity stress and are not less likely to drawdown deposits, since they must have limited credit access.

Table 5.3.4 - Segmentation by employment status

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| Employee | 0.57 | 0.93 | 0.63 |
| Self-employed | 0.64 | 0.90 | 0.66 |
| Unemployed | 0.61 | 0.95 | 0.65 |
| Retired | 0.62 | 0.92 | 0.62 |
| Other inactive | 0.74 | 0.83 | 0.71 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to employment status..*

The employment segmentation shows some heterogeneity. The highest value of predicted probability of liquidity distress is observed in the "Other inactive" group, suggesting fragile cash-flows for households outside standard labour market profiles (e.g. students or long-term inactive individuals). This pattern is consistent with the idea that irregular or absent income streams are associated with liquidity shocks.

The conditional probability shows less variation among employment groups, with the "Unemployed" households revelling the highest propensity to use deposits when under pressure followed by employees and retirees.

Combining these two dimensions, the joint run-off measure is highest for the "Other inactive" group. This indicates that retail portfolios with a strong presence of non-employed inactive households are more exposed to deposit outflows in stress scenarios.

Table 5.3.5 - Segmentation by education level

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| Primary or below | 0.64 | 0.92 | 0.62 |
| Lower secondary | 0.61 | 0.90 | 0.64 |
| Upper secondary | 0.57 | 0.90 | 0.60 |
| Tertiary | 0.58 | 0.94 | 0.65 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to education levels.*

Education level suggests that higher schooling is associated with lower exposure to liquidity stress but also a strong propensity to use deposits when stress occurs. This points to a mitigation of liquidity risk as education rises, likely reflecting more stable income and better financial planning. As a result, the joint run-off measure is highest in the "Tertiary" group.
It can be determined that households with low education are more vulnerable to liquidity shocks, but, on the other hand, tertiary-educated households are less likely to become distressed but may generate larger deposit outflows because they hold and actively use more deposit buffers.

However, the regression-based AME (table 5.1.3) indicates that these differences are modest and not statistically robust, so education should be considered as a secondary segmentation criteria.

Table 5.3.6 – Segmentation by urbanisation

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| City | 0.64 | 0.95 | 0.68 |
| Towns & suburbs | 0.51 | 0.91 | 0.58 |
| Rural | 0.54 | 0.89 | 0.58 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to urbanisation.*

The profiles by degree of urbanisation reveal a contrast between urban and non-urban households. City households display the highest predicted probability of liquidity stress, suggesting that urban households face more frequent cash-flow pressures, reflecting higher living costs and debt obligations associated with metropolitan areas. Since the highest conditional probability of responding to stress by drawing down deposits is also in city households, there is no surprise that the joint run-off measure is substantially higher in cities than in other areas.

However, this pattern should not be interpreted as evidence that geography itself as an independent driver of liquidity stress, since the AME (table 5.1.5) results show that once other variables are controlled, the marginal effects to urbanisation degrees are small and statistically insignificant. In other words, the higher distress and run-off probabilities observed for city households in the segment profiles reflect the characteristics of households living in cities, rather than a direct causal effect of urban location.

Table 5.3.7 - Segmentation by housing tenure

| | mean_Pliq | mean_Pdepdd | mean_Prunoff |
|---|---|---|---|
| Owner outright | 0.66 | 0.96 | 0.69 |
| Owner with mortgage | 0.57 | 0.92 | 0.62 |
| Tenant/other | 0.68 | 0.94 | 0.69 |

*Note: Values represent survey-weighted averages of model-predicted probabilities using HFCS 2021 microdata. Pliq = probability of liquidity pressure from the RQ1 model. Pdepdd = conditional probability of coping through deposit drawdown from the RQ2 model. Prunoff = joint run-off indicator, computed as Pliq × Pdepdd. Segments correspond to housing tenure.*

Housing tenures segmentation reveals that outright owners and tenants exhibit the highest predicted probabilities of liquidity problems. This is consistent with the demographic composition of the groups: outright owners typically include older households with more limited incomed and tenants often tend to have greater income volatility. The conditional probability of using deposits is lower for mortgaged owners, which have less flexible liquidity due to the mortgage constraints. The generated joint run-off probability confirms what was shown in the probability of stress information.

For IRRBB it is interesting that although mortgaged households carry leverage, they appear more stable in terms of deposit outflows, suggesting that a portfolio concentrated in outright owners or tenants may be more exposed to run-off volatility in stress scenarios.

All these segmentations show that behavioural run-off risk arises from the interaction between the likelihood of experiencing liquidity pressure and the capacity and inclination to cope by drawing down deposits. Income, debt, employment status and tenure are the most informative dimensions for IRRBB segmentation, while geography and education, once controls are included, play only a secondary role. This provides a structured basis for calibrating behavioural assumptions in stress testing, enabling banks to identify depositor segments that are more prone to outflows under stressful conditions.

## 5.4 - Segment run-off rates

This file can be used as complementary descriptive step. It uses the multiple imputation sample to group households in a matrix defined by quartiles of the liquidity-buffer ratio and quartiles of the mortgage debt-service ratio, reporting survey-weighted averages of the probability of being under liquidity pressure and the probability of coping with stress primarily by drawing down deposits. The output is displayed as pivot table with a “heat map” colour scheme.

Table 5.4.1 - Probability of liquidity pressure

| P(liquidity pressure) | Debt | | | | |
|---|---|---|---|---|---|
| **Liquidity** | **No debt** | **Q1 (low)** | **Q2** | **Q3 (high)** | **Grand Total** |
| No buffers | 0.551 | 0.583 | 0.602 | 0.809 | 0.636 |
| Q1 (low) | 0.394 | 0.347 | 0.484 | 0.658 | 0.471 |
| Q2 | 0.260 | 0.278 | 0.412 | 0.654 | 0.401 |
| Q3 (high) | 0.222 | 0.274 | 0.398 | 0.749 | 0.411 |
| **Grand Total** | **0.357** | **0.370** | **0.474** | **0.718** | **0.480** |

*Note: Survey-weighted segment-level averages of Pliq, the predicted probability of household liquidity stress from RQ1. Values aggregated by liquidity and debt.*

Table 5.4.2 - Probability of coping via deposit

| P(savings leaning) | Debt | | | | |
|---|---|---|---|---|---|
| **Liquidity** | **No debt** | **Q1 (low)** | **Q2** | **Q3 (high)** | **Grand Total** |
| No buffers | 0.571 | 0.598 | 0.536 | 0.576 | 0.570 |
| Q1 (low) | 0.883 | 0.848 | 0.867 | 0.864 | 0.865 |
| Q2 | 0.832 | 0.900 | 0.936 | 0.927 | 0.899 |
| Q3 (high) | 0.913 | 0.930 | 0.977 | 0.907 | 0.932 |
| **Grand Total** | **0.800** | **0.819** | **0.829** | **0.819** | **0.817** |

*Note: Survey-weighted segment-level averages of Pdepdd, the conditional probability of coping via deposit drawdown among pressured households (RQ2). Values aggregated by liquidity and debt.*

Liquidity stress is lowest in households with high buffers and low DSRs. and increase progressively as either buffers shrink or debt obligations rise. The highest distress levels cluster in the top-right corner of the matrix, where households combine no buffers with high debt-service ratios, producing values above 0.80. This confirms that low liquidity and high leverage interact multiplicatively: households with both characteristics are distinctly more exposed to liquidity shocks than those with only one of these vulnerabilities.

Regarding the second table, households with larger buffers exhibit systematically higher savings-leaning probabilities. This is consistent with the idea that households with more resources are both more able and more inclined to use deposits to endure shocks. By contrast, variation across DSR bands is much weaker, indicating that the debt-service obligations affect the likelihood of needing to adjust, but plays a secondary role in determining how adjustment occurs once stress is present.

All together, these tables provide a simple but powerful conclusion for IRRBB. The combination of small buffers and high debt-service obligations identifies a segment of households with both high distress risk and strong dependency on deposit drawdowns as a coping mechanism. These households represent the “high-run-off” group of the retail portfolio: they are significantly more likely to generate outflows when adverse conditions happen. Conversely, households with strong buffers, even if indebted, exhibit lower stress probabilities and higher resilience. This two-way segmentation therefore offers a concise behavioural map for stress-testing, enabling risk managers to focus on the parts of the depositor base where outflows are most likely to concentrate.

# 6

## 6 - Conclusions

The objective of this thesis was to analyse how household financial structures influence the possibility of families experience liquidity stress and how they respond when that stress occurs. Although retail deposits still are the backbone of banks' funding models, there is not very practical work done to understand which types of households are more likely to drawdown deposits in stressful times and how this behaviour could vary around socioeconomic groups.
In the IRRBB context, where behavioural assumptions regarding deposit stability can affect interest rate sensitivity and the outcome of stress tests, internal or regulatory, gaining an understanding of depositors' behaviour seems very relevant. With this in mind, this thesis tries to provide an evidence-based framework for understanding behavioural run-off risk using a household-level dataset.

To accomplish the proposed goals, this thesis uses the Household Finance and Consumption Survey (HFCS), which offers a detailed picture of household's finance across income, wealth, debt, employment and demographic variables.

Two empirical models were developed: the first estimates the probability that a household would face a liquidity stress if adverse economic conditions occur. The second model focus on how households choose to cope once such liquidity pressure arises, if they rely on deposit drawdowns or instead turn to credit. These models were estimated with survey weights and using multiple imputation structure of the HFCS, ensuring that all data is handled appropriately. The resulting outcomes provided the basis for a behavioural segmentation of the depositor base. The main conclusions can be separated into two dimensions. On the first dimension, the results show that liquidity pressure is most prevalent among households with lower income, heavy debt obligations and precarious employment status. Self-employed households, unemployed and those classified as "other inactive" showed the highest predicted probability of incurring into liquidity constraints. Housing tenure also played an important role, since mortgaged households were less likely to encounter liquidity pressure, consistent with their theoretically higher incomes and stable employment profiles. On the other hand, tenants and outright owners displayed higher predicted distress due to the rent obligations and lower income.

The second dimension addresses how households cope once pressure arises and the results point into an interesting direction. In this case, liquidity buffers and wealth dominate. Households with more substantial buffers are more likely to draw down deposits when facing a liquidity issue, while those with smaller liquidity positions are more likely to resort to credit. This implies that wealthier households, when shocks happen, have both the means and the preference to absorb the shock by using savings. Also, income plays a similar role as higher income households show stronger deposit leaning tendency when under pressure. Education and geography did not emerge as strong independent variables of coping behaviour once other characteristics were taken into account.

One final note is that, within the 2021 HFCS wave, this study found that 49.8% of households in the dataset experienced of liquidity distress, which means that they would face difficulty, for example, meeting an unexpected payment. The author considers that this number should be interpreted in the context of the COVID-19 period, during which many households experienced income reductions and uncertainty regarding employment. These conditions most likely increased the number of households at risk compared with a more typical economic environment. Even though the models in this thesis aimed at the estimation of structural behaviours instead of crisis effects, we will have to take this into consideration and reinforce the idea that, in periods of macroeconomic uncertainty, a larger share of depositors may be susceptible to run-off behaviours.

A useful contribution of this thesis is the translation of these behavioural predictions into segment run-off indicators. By aggregating the predicted probabilities across specific groups like income, debt, employment, education or housing tenure, it is possible to identify which types of households are likely to generate the largest deposit outflows in stressful situations.
Low-income households and those with heavy debt obligations show the highest likelihood of liquidity pressure but high buffer and high wealth households exhibit stronger deposit-base coping behaviour. This is interesting because shows that different segments contribute to run-

off risk in different ways, some are more frequently pushed into distress, others because they respond to distress by drawing deposits more repeatedly.

The segment_runoff_rates output show exactly this: a clear concentration of risk among households with low liquidity buffers and high mortgage debt-service ratios. Alternatively, households with strong buffers, even if indebted, display lower run-off risk. This two-dimensional view of depositors' behavior provides an intuitive tool for IRRBB teams identify and monitor their depositor base.

Altogether, there are three findings of this thesis that can be useful for the IRRBB practice. First, they show that household-level characteristics can predict the likelihood of distress and the coping option used under stress. Second, the segmentation results provide a structured view of depositor risk, helping banks understand how vulnerabilities are spread across household groups. Third, the flexibility of this methodology, meaning that the household-level probabilities can be aggregated with any segmentation useful for the bank, adapting the analysis to different retail deposit books.

Despite its contributions, this is thesis is also subject to some limitations. The HFCS is cross-sectional, which means that the analysis cannot capture any kind of dynamic behavioural adjustments, neither can observe how individual households respond to different economic conditions or interest rates environments. The liquidity pressure indicator and the coping responses are self-reported and hypothetical, as actual withdrawal behaviour may differ during real stress episodes. Also, the dataset does not incorporate any bank-level characteristics such as account types, interest rates or customer relationship depth, factors that could influence deposit stability in practice. Finally, the translation to IRRBB is partial. While the predicted run-off measures provide a behavioural response, they do not map directly onto regulatory buckets.

These limitations offer several opportunities for future research. Linking HFCS-style survey to a banking deposit data would allow validation of behavioral predictions and enable a better estimation of the run-off dynamics. A panel data would allow the modelling the transitions into and out of liquidity stress, as well as repeated deposit-drawdown events. Incorporating bank-specific features could offer a broader view of the interaction between households and banks.

It would be also very interesting to simulate macroeconomic shocks, such as inflation or big interest rate movements, to evaluate how behavioral run-off risk evolves under different stress environments.

In sum, this thesis provides a structured analysis on how household characteristics influence behavioral deposit risk under stress. Combining household microdata, survey econometrics and IRRBB aggregation offers a framework that helps identifying which depositor group is most likely to contribute to run-off in stressful conditions. The results highlight the importance of considering vulnerability and coping capacity when measuring retail deposit stability and offer a base for future analytical developments when connecting household finance with bank risk management.

# Appendix A – Stata Code

```
****************************************************************************
* HFCS Liquidity Model — FINAL (lowercase vars; robust to missing HI/HC items)
* - Keeps ALL vars from d1...d5; weights from D files (hw0010)
* - MI via mi import mlong (m_imp index)
* - Country encoded for factor use
* - Liquidity pressure = (no saving room) OR (DSR>=40%) [+ arrears if available]
* - Coping tendency among pressured: savings-leaning vs credit-leaning
****************************************************************************

version 17
clear all
set more off
#delimit cr

*==============================*
* 0) USER PATHS
*==============================*
global HFCSDATA "/Users/Dores/Desktop/TESE/ECB DATA/HFCS_UDB_4_0_Stata"   // EDIT
global OUT      "/Users/Dores/Desktop/TESE/OUTPUT"                        // EDIT
cap mkdir "$OUT"

*==============================*
* 1) STACK ALL IMPUTATIONS (KEEP EVERYTHING)
*==============================*
use "$HFCSDATA/d1.dta", clear

* Standardize key names (brace-free)
capture rename SA0100 sa0100
capture rename SA0010 sa0010
capture rename im0100 IM0100
capture rename DNNLARATIO dnnlaratio
capture rename HW0010 hw0010

* Ensure IM0100 exists
capture confirm variable IM0100
if _rc gen byte IM0100 = 1
label var IM0100 "Imputation number"

gen byte _mi_m = 1
tempfile BASE
save `BASE', replace

* Implicates 2..5 -> append (no filtering)
forvalues m = 2/5 {
   use "$HFCSDATA/d`m'.dta", clear
   capture rename SA0100 sa0100
   capture rename SA0010 sa0010
   capture rename im0100 IM0100
   capture rename DNNLARATIO dnnlaratio
   capture rename HW0010 hw0010

   capture confirm variable IM0100
   if _rc gen byte IM0100 = `m'
   label var IM0100 "Imputation number"

   gen byte _mi_m = `m'
   append using `BASE'
```

```
    save `BASE', replace
}
use `BASE', clear

*==============================*
* 2) ENTER MI MODE (mlong via import)
*==============================*
* Avoid clash with MI system var name
capture confirm variable _mi_m
if !_rc rename _mi_m m_imp
save, replace

* Convert long-stacked to MI (mlong)
mi import mlong, id(sa0100 sa0010) m(m_imp)

*==============================*
* 3) REGISTER MI VARIABLES (robust)
*==============================*
* Build imputed list: numeric vars excluding IDs, weight, helper, MI system vars
local imps
ds
foreach v of varlist _all {
    capture confirm numeric variable `v'
    if !_rc {
        if !inlist("`v'","sa0100","sa0010","hw0010","m_imp","IM0100","_mi_m","_mi_id","_mi_miss") {
            local imps `imps' `v'
        }
    }
}
mi register imputed `imps'
mi register regular sa0100 sa0010 hw0010 IM0100
mi describe

* Register remaining structural vars as REGULAR
capture confirm variable id
if !_rc mi register regular id

capture confirm variable dhregion
if !_rc mi register regular dhregion

capture confirm variable sb1000
if !_rc mi register regular sb1000

* Keep m_imp as a helper (unregistered), or drop if you prefer:
* drop m_imp

* Encode country (sa0100 is string) for factor models
capture confirm string variable sa0100
if !_rc {
    capture drop ctry
    encode sa0100, gen(ctry)
    label var ctry "Country (encoded from sa0100)"
    mi register regular ctry
}
```

```
*==============================*
* 4) SURVEY SETTINGS
*==============================*
mi svyset [pw=hw0010]
mi svyset

*==============================*
* 5) PREP: COUNTRY, DEPOSITS, BUFFERS
*==============================*

* Country (keep existing ctry)
capture noisily mi register regular ctry
capture label var ctry "Country (encoded from sa0100)"

* Build total deposits if needed
capture confirm variable da2101
if _rc mi passive: gen double da2101 = da21011 + da21012
label var da2101 "Total deposits (sight+savings)"

* Sight share in [0,1]
mi passive: gen double sight_share = da21011/da2101 if da2101>0
mi passive: replace sight_share = . if sight_share<0 | sight_share>1
label var sight_share "Sight deposits share of total deposits"

* Cap buffer ratio (stable & interpretable 0..10)
mi passive: gen double dnnlaratio_cap = dnnlaratio
mi passive: replace dnnlaratio_cap = 0  if dnnlaratio_cap<0  & !missing(dnnlaratio_cap)
mi passive: replace dnnlaratio_cap = 10 if dnnlaratio_cap>10 & !missing(dnnlaratio_cap)
label var dnnlaratio_cap "Net liquid assets / income (capped 0..10)"

*==============================*
* 6) OUTCOMES (pressure & coping)
*==============================*

* --- 6.1 Liquidity pressure components ---

* A) No saving room (HFCS derived: expenses < income == 1; so "no room" is 0)
mi passive: gen byte not_saving_room = (doabletosave==0 & !missing(doabletosave))
label var not_saving_room "No saving room (doabletosave==0)"

* B) High debt-service stress (all indebted households)
mi passive: gen byte dsr40 = (dodstotal40p==1)
label var dsr40 "Debt service >= 40% (all indebted)"

* C) Arrears (if present; otherwise set to 0 so the definition is explicit)
capture confirm variable hc1250
if !_rc {
   mi passive: gen byte arrears_any = (hc1250==1)
}
else {
   capture confirm variable hcarears
   if !_rc {
      mi passive: gen byte arrears_any = (hcarears==1)
   }
   else {
      mi passive: gen byte arrears_any = 0
      label var arrears_any "Arrears (not available in delivery; set to 0)"
   }
}
```

```
* Final pressure flag: any of the above
mi passive: gen byte liq_press = (not_saving_room==1 | dsr40==1 | arrears_any==1)
label var liq_press "Liquidity pressure: no saving room OR DSR>=40% [+ arrears if available]"

* --- 6.2 Coping tendency (proxy tri-category among pressured) ---

* Signals that household can/does lean on savings
mi passive: gen byte cope_sav_proxy  = (dnnlaratio_cap>=1 | da2101>0)
label var cope_sav_proxy "Signals savings-leaning (buffers/deposits)"

* Signals that household can/does lean on credit (line/card/constraints/refusal)
* If any of these vars are missing in your file, tell me and I'll swap a fallback.
mi passive: gen byte cope_cred_proxy = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)
label var cope_cred_proxy "Signals credit-leaning (line/card/constraints/refused)"

* 0 = other/unclear; 1 = savings-leaning; 2 = credit-leaning (only defined if pressured)
mi passive: gen byte cope_cat = .
replace cope_cat = 1 if liq_press==1 & cope_sav_proxy==1 & cope_cred_proxy!=1
replace cope_cat = 2 if liq_press==1 & cope_cred_proxy==1
replace cope_cat = 0 if liq_press==1 & missing(cope_cat)
label define cope_lbl 0 "other/unclear" 1 "savings-leaning" 2 "credit-leaning"
label values cope_cat cope_lbl
label var cope_cat "Coping tendency under pressure (proxy tri-cat)"



*==============================*
* 7) COPING TENDENCY (proxy, among pressured)
*==============================*
* Recompute (generate if missing, replace if exists)
capture confirm variable cope_credit
if _rc  mi passive: gen byte cope_credit = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)
else    mi passive: replace    cope_credit = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)
label var cope_credit "Signals of leaning on credit (line/card/constrained/refused)"

capture confirm variable cope_sav
if _rc  mi passive: gen byte cope_sav = (dnnlaratio_cap>=1 | da2101>0)
else    mi passive: replace    cope_sav = (dnnlaratio_cap>=1 | da2101>0)
label var cope_sav "Signals of leaning on savings (buffers/deposits)"

* Rebuild tri-category deterministically
capture confirm variable cope_cat
if _rc  mi passive: gen byte cope_cat = .
else    mi passive: replace    cope_cat = .

replace cope_cat = 1 if liq_press==1 & cope_sav==1 & cope_credit!=1
replace cope_cat = 2 if liq_press==1 & cope_credit==1
replace cope_cat = 0 if liq_press==1 & missing(cope_cat)

capture label drop cope
label define cope 0 "other/unclear" 1 "savings-leaning" 2 "credit-leaning"
label values cope_cat cope
label var cope_cat "Coping tendency among pressured (proxy)"
```

```
*===============================*
* 8) MODELS (with clean cats)
*===============================*


*--- Make sure we are in MI context before Block 8 models ---
capture mi describe
if _rc {
   di as err "Not in MI context. Trying to load your MI dataset from $OUT ..."
   capture noisily use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear
   capture mi describe
   if _rc {
      di as err "Still not in MI context. Re-run Blocks 1–3 (stack + mi import + register) and try again."
      exit 111
   }
}

* -- 8.0 Clean categorical variables for factor notation (no negatives) --
capture mi describe
local is_mi = (_rc==0)

foreach v in dhemph1 dheduh1 dhhst dhdegurba {
   capture confirm variable `v'_p
   if _rc {
      if `is_mi' {
         mi passive: gen `v'_p = cond(`v'<0,.,`v')
      }
      else {
         gen `v'_p = cond(`v'<0,.,`v')
      }
      local lbl : value label `v'
      if "`lbl'" != "" label values `v'_p `lbl'
   }
}


*---------------------------*
* 8.1 RQ1: Pr(liquidity pressure)
*---------------------------*
mi estimate, vceok esampvaryok post: ///
   svy: logit liq_press ///
      c.dnnlaratio_cap c.sight_share ///
      c.dodsmortg      c.dodiratio  ///
      c.di2000         i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
      i.ctry
estimates store RQ1

* ---- AMEs for RQ1 (continuous regressors) ----
local dx_rq1 dnnlaratio_cap sight_share dodsmortg dodiratio di2000 dh0002
mimrgns, dydx(`dx_rq1')

* (Optional) show tidy table of AMEs
* estimates table, b(%9.3f) se(%9.3f) star stats(N)

*---------------------------*
* 8.2 RQ2: Coping tendency among pressured
*---------------------------*
mi estimate, vceok esampvaryok post: ///
   svy, subpop(if liq_press==1): ///
   mlogit cope_cat ///
```

```
    c.dnnlaratio_cap c.sight_share ///
    c.dodsmortg     c.dodiratio  ///
    c.di2000        i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
    i.ctry
estimates store RQ2

* ---- AMEs for RQ2 by outcome ----
local dx_rq2 dnnlaratio_cap sight_share dodsmortg dodiratio di2000 dh0002
mimrgns, dydx(`dx_rq2') predict(outcome(1))   // savings-leaning
mimrgns, dydx(`dx_rq2') predict(outcome(2))   // credit-leaning


* After Block 8 models finish (RQ1/RQ2 just estimated)
* Make sure mimrgns is installed and MI is active (optional guards)
cap which mimrgns
if _rc ssc install mimrgns, replace
cap mi describe
* if _rc use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear

* ---------- RQ1 AMEs (probability scale) ----------
estimates restore RQ1
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(pr)
matrix A = r(table)'   // capture immediately

putexcel set "$OUT/RQ1_AMEs.xlsx", replace
putexcel A1=("var") B1=("dy/dx") C1=("se") D1=("t") E1=("p") F1=("ll") G1=("ul")
putexcel A2=matrix(A), names

* ---------- RQ2 AMEs by outcome ----------
estimates restore RQ2

* Outcome (1)
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(1))
matrix S = r(table)'   // capture right after outcome(1)

* Outcome (2)
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(2))
matrix C = r(table)'   // now capture outcome(2)

putexcel set "$OUT/RQ2_AMEs.xlsx", replace
putexcel A1=("Outcome: savings-leaning")
putexcel A2=("var") B2=("dy/dx") C2=("se") D2=("t") E2=("p") F2=("ll") G2=("ul")
putexcel A3=matrix(S), names
local base = rowsof(S) + 4
putexcel A`base'=("Outcome: credit-leaning")
putexcel A`= `base'+1' =("var") B`= `base'+1' =("dy/dx") C`= `base'+1' =("se") ///
    D`= `base'+1' =("t") E`= `base'+1' =("p") F`= `base'+1' =("ll") G`= `base'+1' =("ul")
putexcel A`= `base'+2' =matrix(C), names
```

```
* ============================================================
* Append categorical effects to the existing AME workbooks
*   - Keeps previous sheets (continuous AMEs)
*   - Adds new sheets per categorical variable
*   - RQ1: dy/dx on Pr(liq_press)
*   - RQ2: dy/dx on Pr(cope_cat==1) and Pr(cope_cat==2)
* ============================================================


* ---------- EMPLOYMENT STATUS (reference person)
capture label drop L_dhemph1
label define L_dhemph1 ///
   1 "Employee" ///
   2 "Self-employed" ///
   3 "Unemployed" ///
   4 "Retired" ///
   5 "Other inactive"
capture label values dhemph1_p L_dhemph1

* ---------- EDUCATION (ISCED-style bands)
capture label drop L_dheduh1
label define L_dheduh1 ///
   1 "Primary or below" ///
   2 "Lower secondary" ///
   3 "Upper secondary / Post-sec non-tertiary" ///
   5 "Tertiary"
capture label values dheduh1_p L_dheduh1

* ---------- HOUSING TENURE/STATUS
capture label drop L_dhhst
label define L_dhhst ///
   1 "Owner outright" ///
   2 "Owner with mortgage" ///
   3 "Tenant/other"
capture label values dhhst_p L_dhhst

* ---------- DEGREE OF URBANISATION
capture label drop L_dhdeg
label define L_dhdeg ///
   1 "City" ///
   2 "Towns & suburbs" ///
   3 "Rural"
capture label values dhdegurba_p L_dhdeg



* Make sure mimrgns is available
cap which mimrgns
if _rc ssc install mimrgns, replace

* --- Recreate/restore models if needed ---
cap estimates restore RQ1
if _rc {
   mi estimate, vceok esampvaryok post: ///
      svy: logit liq_press ///
         c.dnnlaratio_cap c.sight_share ///
         c.dodsmortg c.dodiratio c.di2000 ///
         i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
         i.ctry
   estimates store RQ1
```

```
}

cap estimates restore RQ2
if _rc {
   mi estimate, vceok esampvaryok post: ///
      svy, subpop(if liq_press==1): ///
      mlogit cope_cat ///
         c.dnnlaratio_cap c.sight_share ///
         c.dodsmortg c.dodiratio c.di2000 ///
         i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
         i.ctry
   estimates store RQ2
}

* --- Choose which categorical factors to export ---
local cats dhemph1_p dheduh1_p dhhst_p dhdegurba_p
* (Optional) also add country: ctry   -> many levels; uncomment if desired
* local cats `cats' ctry

* ==========================
* RQ1: categorical AMEs
* ==========================
* Which categorical factors to export
local cats dhemph1_p dheduh1_p dhhst_p dhdegurba_p

* Make sure the model is active
estimates restore RQ1

foreach c of local cats {
   quietly mimrgns, dydx(i.`c') predict(pr)
   matrix T = r(table)'

   * OPEN the workbook and SELECT the target sheet here (not on the write line)
   putexcel set "$OUT/RQ1_AMEs.xlsx", modify sheet("Cat_`c'")

   * Write header and table
   putexcel A1=("level") B1=("dy/dx") C1=("se") D1=("t") E1=("p") F1=("ll") G1=("ul")
   putexcel A2=matrix(T), names
}

* ==========================
* RQ2: categorical AMEs
* ==========================
* Same categorical blocks
local cats dhemph1_p dheduh1_p dhhst_p dhdegurba_p

* Activate RQ2 model
estimates restore RQ2

foreach c of local cats {
   * Outcome = 1 (savings-leaning)
   quietly mimrgns, dydx(i.`c') predict(outcome(1))
   matrix S = r(table)'

   putexcel set "$OUT/RQ2_AMEs.xlsx", modify sheet("Cat_`c'_out1")
   putexcel A1=("level") B1=("dy/dx") C1=("se") D1=("t") E1=("p") F1=("ll") G1=("ul")
   putexcel A2=matrix(S), names

   * Outcome = 2 (credit-leaning)
   quietly mimrgns, dydx(i.`c') predict(outcome(2))
```

```
    matrix C = r(table)'

    putexcel set "$OUT/RQ2_AMEs.xlsx", modify sheet("Cat_`c'_out2")
    putexcel A1=("level") B1=("dy/dx") C1=("se") D1=("t") E1=("p") F1=("ll") G1=("ul")
    putexcel A2=matrix(C), names
}

display as result "Categorical AMEs appended to:"
display as result " - $OUT/RQ1_AMEs.xlsx  (sheets: Cat_*)"
display as result " - $OUT/RQ2_AMEs.xlsx  (sheets: Cat_*_out1 / Cat_*_out2)"

*=============================*
* 9) SEGMENT EXPORTS (observed, weighted means)
*=============================*

* (Optional) ensure OUT exists (harmless if it already does)
capture noisily mkdir "$OUT"

* --- Build quartiles for buffers & DSR (MI-aware & re-runnable) ---
capture mi describe
local is_mi = (_rc==0)

if `is_mi' {
    capture confirm variable q_buf0
    if _rc mi passive: egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
    capture confirm variable q_dsr0
    if _rc mi passive: egen q_dsr0 = cut(dodsmortg),      group(4) icodes

    capture confirm variable q_buf
    if _rc mi passive: gen q_buf = q_buf0 + 1
    capture confirm variable q_dsr
    if _rc mi passive: gen q_dsr = q_dsr0 + 1
}
else {
    capture confirm variable q_buf0
    if _rc egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
    capture confirm variable q_dsr0
    if _rc egen q_dsr0 = cut(dodsmortg),      group(4) icodes

    capture confirm variable q_buf
    if _rc gen q_buf = q_buf0 + 1
    capture confirm variable q_dsr
    if _rc gen q_dsr = q_dsr0 + 1
}

* Consistent labels 1..4
label define q4 1 "Q1 (low)" 2 "Q2" 3 "Q3" 4 "Q4 (high)", replace
capture label values q_buf q4
capture label values q_dsr q4

* --- Export weighted cell means (observed rates) ---
preserve
    * Keep only needed vars and drop rows without quartiles
    keep liq_press cope_cat q_buf q_dsr hw0010
    keep if !missing(q_buf, q_dsr)

    * 0/1 indicators for coping channels (avoid == expression in collapse)
    gen byte sav  = (cope_cat==1) if !missing(cope_cat)
    gen byte cred = (cope_cat==2) if !missing(cope_cat)
```

```
   * Weighted means by (q_buf, q_dsr)
   collapse (mean) ///
      P_liq = liq_press ///
      P_sav = sav ///
      P_cred= cred [pw=hw0010], by(q_buf q_dsr)

   export excel using "$OUT/segment_runoff_rates.xlsx", replace firstrow(variables)
restore


*==============================*
* 10) SAVE MI SAMPLE
*==============================*
keep sa0100 sa0010 IM0100 hw0010 m_imp ctry ///
   liq_press not_saving_room dsr40 arrears_any ///
   cope_cat cope_sav cope_credit ///
   da2101 da21011 da21012 sight_share dnnla dnnlaratio dnnlaratio_cap ///
   dodiratio dodsmortg di2000 dheduh1 dhemph1 dh0002 dhhst dhdegurba ///
   q_buf q_dsr
compress
save "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", replace

display as res "Done. Outputs saved to $OUT"


*==============================*
* 11) FINAL EXPORT PACK (robust)
*==============================*

* -- Ensure OUT path exists (ignore if it already exists) --
capture noisily mkdir "$OUT"
display as res "Output folder: $OUT"

* -- Ensure the right data are loaded (avoid empty memory after 'clear') --
capture confirm variable dnnlaratio_cap
if _rc {
   use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear
}

* -- Detect whether current data are MI-managed --
capture mi describe
local is_mi = (_rc==0)

* -- Build quartiles safely (MI if possible, else plain) --
if `is_mi' {
   mi passive: egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
   mi passive: egen q_dsr0 = cut(dodsmortg),      group(4) icodes
   capture confirm variable q_buf
   if _rc mi passive: gen q_buf = q_buf0 + 1
   capture confirm variable q_dsr
   if _rc mi passive: gen q_dsr = q_dsr0 + 1
}
else {
   capture confirm variable q_buf0
   if _rc egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
   capture confirm variable q_dsr0
   if _rc egen q_dsr0 = cut(dodsmortg),      group(4) icodes
   capture confirm variable q_buf
   if _rc gen q_buf = q_buf0 + 1
   capture confirm variable q_dsr
```

```
    if _rc gen q_dsr = q_dsr0 + 1
}

label define q4 1 "Q1 (low)" 2 "Q2" 3 "Q3" 4 "Q4 (high)", replace
label values q_buf q4
label values q_dsr q4

*-----------------------------*
* 11.1 Segment runoff proxies (fixed)
*-----------------------------*
preserve
    keep liq_press cope_cat q_buf q_dsr hw0010
    keep if !missing(q_buf, q_dsr)

    gen byte sav  = (cope_cat==1) if !missing(cope_cat)
    gen byte cred = (cope_cat==2) if !missing(cope_cat)

    collapse (mean) ///
        P_liq = liq_press ///
        P_sav = sav ///
        P_cred= cred [pw=hw0010], by(q_buf q_dsr)

    export excel using "$OUT/segment_runoff_rates.xlsx", replace firstrow(variables)
restore

*-----------------------------*
* 11.2 Save tidy MI sample
*-----------------------------*
keep sa0100 sa0010 IM0100 hw0010 m_imp ctry ///
     liq_press not_saving_room dsr40 arrears_any ///
     cope_cat cope_sav cope_credit ///
     da2101 da21011 da21012 sight_share dnnla dnnlaratio dnnlaratio_cap ///
     dodiratio dodsmortg di2000 dheduh1 dhemph1 dh0002 dhhst dhdegurba ///
     q_buf q_dsr
compress
save "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", replace

display as res "Done. Core outputs saved in: $OUT"




*=====================================================================*
* THESIS DELIVERABLES PACK
* Requires: RQ1 & RQ2 already estimated and stored (est store RQ1/RQ2)
* Output folder: global OUT "/Users/Dores/Desktop/TESE/OUTPUT"
*=====================================================================*

* Housekeeping
capture noisily mkdir "$OUT"
capture which mimrgns
if _rc ssc install mimrgns, replace
```

```
*------------------------------------------------------------*
* Merge liq_press from tidy file into MI_CLEAN (safe + MI-aware)
*------------------------------------------------------------*

* 0) Open your clean MI dataset as master
clear
use "$OUT/hfcs2021_liquidity_model_MI_CLEAN.dta", clear

* Sanity: confirm MI context
capture mi describe
if _rc {
   di as err "Master file is not MI-managed. Open *_MI_CLEAN.dta first."
   exit 111
}

* 1) Build an implicate key in MASTER: m_imp_key = cond(_mi_m==0,1,_mi_m)
*    (Do NOT touch your existing m_imp; we use a separate key to avoid conflicts)
capture confirm variable m_imp_key
if _rc {
   mi passive: gen byte m_imp_key = cond(_mi_m==0, 1, _mi_m)
   capture mi register regular m_imp_key   // harmless if already registered
}

preserve
   * 2) Build USING file from tidy data, enforcing uniqueness per hh+implicate
   use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear

   * Normalize implicate marker to IM0100
   capture confirm variable IM0100
   if _rc {
      capture confirm variable m_imp
      if _rc {
         di as err "No implicate marker (IM0100 or m_imp) in tidy file."
         exit 111
      }
      rename m_imp IM0100
   }

   * Require liq_press in tidy file
   capture confirm variable liq_press
   if _rc {
      di as err "liq_press not found in tidy file. Re-run the block where you created it."
      exit 111
   }

   * Build same key name used in master and keep only needed vars
   gen byte m_imp_key = IM0100
   keep sa0100 sa0010 m_imp_key liq_press

   * ENFORCE UNIQUENESS: if multiple rows exist, take max(liq_press) per key
   collapse (max) liq_press, by(sa0100 sa0010 m_imp_key)

   tempfile TIDY_LIQ
   save `TIDY_LIQ', replace
restore

* 3) Now merge many(master):one(using) safely
merge m:1 sa0100 sa0010 m_imp_key using `TIDY_LIQ', keepusing(liq_press) nogenerate
```

```
* 1) Remove the problematic registration and drop the helper key (we don't need it anymore)
capture mi unregister passive  m_imp_key
capture mi unregister regular  m_imp_key
capture drop m_imp_key

* 2) Fill liq_press across implicates (do NOT use mi passive here)
mi xeq: bysort sa0100 sa0010: egen __liq_any = max(liq_press)
mi xeq: replace liq_press = __liq_any if missing(liq_press)
mi xeq: drop __liq_any

* 3) Optional: register liq_press as imputed (safe to repeat)
capture mi register imputed liq_press

* 4) Check there are no missing liq_press in any m
mi xeq: count if missing(liq_press)

* (optional) save
save "$OUT/hfcs2021_liquidity_model_MI_CLEAN.dta", replace


*==================================================*
* FIX BUFFERS + CREDIT FLAGS + BUILD COPE_CAT   *
* (Run with ..._MI_CLEAN.dta open; mi set)      *
*==================================================*

* 0) Sanity: we need liq_press already merged into MI_CLEAN
capture confirm variable liq_press
if _rc {
   di as err "liq_press missing in MI_CLEAN. Run the earlier merge step first."
   exit 111
}

* 1) Rebuild dnnlaratio_cap if missing (cap to [0,10])
capture confirm variable dnnlaratio_cap
if _rc {
   capture confirm variable dnnlaratio
   if !_rc {
      mi xeq: gen double dnnlaratio_cap = dnnlaratio
      mi xeq: replace dnnlaratio_cap = 0  if missing(dnnlaratio_cap) | dnnlaratio_cap<0
      mi xeq: replace dnnlaratio_cap = 10 if dnnlaratio_cap>10
      capture mi register imputed dnnlaratio_cap
   }
   else {
      di as txt "dnnlaratio also missing; will rely on da2101 only in cope_sav proxy."
      mi xeq: gen double dnnlaratio_cap = .
      capture mi register imputed dnnlaratio_cap
   }
}

* 2) Ensure credit flags exist (fallback to 0 if missing)
foreach v in dl1210i dl1220i docreditc docreditrefused {
   capture confirm variable `v'
   if _rc mi xeq: gen byte `v' = 0
}

* 3) Ensure da2101 exists (used in savings-leaning proxy)
capture confirm variable da2101
if _rc {
   di as err "da2101 missing; cannot build cope_sav proxy. Add/merge it first."
   exit 111
```

```
}

* 4) Build proxies in each implicate; label & register
capture mi unregister imputed cope_sav
capture mi unregister imputed cope_credit
capture mi unregister imputed cope_cat

mi xeq: capture drop cope_sav cope_credit cope_cat
mi xeq: gen byte cope_sav    = ( (dnnlaratio_cap>=1 & !missing(dnnlaratio_cap)) | da2101>0 )
mi xeq: gen byte cope_credit = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)

mi xeq: gen byte cope_cat = .
mi xeq: replace cope_cat = 1 if liq_press==1 & cope_sav==1    & cope_credit!=1
mi xeq: replace cope_cat = 2 if liq_press==1 & cope_credit==1
mi xeq: replace cope_cat = 0 if liq_press==1 & missing(cope_cat)

label define cope 0 "other/unclear" 1 "savings_leaning" 2 "credit_leaning", replace
mi xeq: label values cope_cat cope

capture mi register imputed cope_sav cope_credit cope_cat

* 5) Quick checks
mi xeq: tab cope_cat, missing
mi xeq: summarize dnnlaratio_cap if liq_press==1

* ----------------------------------------------------------------
* 2) If cope_cat missing, construct a proxy (defensive, MI-safe)
*    1 = savings-leaning (deposit drawdown), 2 = credit-leaning
* ----------------------------------------------------------------
capture confirm variable cope_cat
if _rc {
   * make sure required inputs exist
   foreach v in dnnlaratio_cap da2101 dl1210i dl1220i docreditc docreditrefused {
      capture confirm variable `v'
      if _rc di as err "Missing variable: `v'"
   }

   * create in each m
   mi xeq: gen byte cope_sav    = (dnnlaratio_cap>=1 | da2101>0)
   mi xeq: gen byte cope_credit = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)
   mi xeq: gen byte cope_cat    = .

   mi xeq: replace cope_cat = 1 if liq_press==1 & cope_sav==1    & cope_credit!=1
   mi xeq: replace cope_cat = 2 if liq_press==1 & cope_credit==1
   mi xeq: replace cope_cat = 0 if liq_press==1 & missing(cope_cat)

   label define cope 0 "other/unclear" 1 "savings_leaning" 2 "credit_leaning", replace
   label values cope_cat cope

   * register them (safe if already registered)
   capture mi register imputed cope_cat cope_sav cope_credit
}
```

```
*===============================*
* HOUSEHOLD-LEVEL PREDICTIONS *
*===============================*

* Safety: MI + SVY context
mi describe
capture mi svyset, clear
mi svyset [pw=hw0010]

* ---------- RQ1: logit(liq_press) -> p_liq ----------
* Fit per imputation and predict per imputation
mi xeq: quietly svy: logit liq_press ///
    c.dnnlaratio_cap c.sight_share ///
    c.dodsmortg c.dodiratio c.di2000 ///
    i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
    i.ctry

mi xeq: capture drop __p_liq_m
mi xeq: quietly predict double __p_liq_m, pr


preserve
tempfile P1 P2 P3 P4

forvalues m = 1/4 {
    mi extract `m', clear                  // now plain (non-MI) data for m=`m'
    keep sa0100 sa0010 hw0010 __p_liq_m
    rename __p_liq_m p_liq_`m'
    sort sa0100 sa0010
    save `P`m'', replace
    restore, preserve                      // go back to MI data for the next loop
}

* Merge the four files and average across m
use `P1', clear
foreach m of numlist 2/4 {
    merge 1:1 sa0100 sa0010 using `P`m'', nogenerate
}
egen double p_liq = rowmean(p_liq_1 p_liq_2 p_liq_3 p_liq_4)
keep sa0100 sa0010 hw0010 p_liq
tempfile LIQ
save `LIQ', replace
restore

* ---------- RQ2: logit(dep drawdown | pressure) -> p_depdd ----------
* If cope_cat missing, build a proxy once (safe no-op if it exists)
capture confirm variable cope_cat
if _rc {
    foreach f in dl1210i dl1220i docreditc docreditrefused {
        capture confirm variable `f'
        if _rc mi xeq: gen byte `f' = 0
    }
    capture confirm variable da2101
    if _rc {
        di as txt "RQ2 skipped: da2101 missing and cope_cat absent."
        local SKIP_RQ2 1
    }
    else {
        mi xeq: gen byte cope_sav    = ((dnnlaratio_cap>=1 & !missing(dnnlaratio_cap)) | da2101>0)
        mi xeq: gen byte cope_credit = (dl1210i==1 | dl1220i==1 | docreditc==1 | docreditrefused==1)
```

```
        mi xeq: gen byte cope_cat    = .
        mi xeq: replace cope_cat = 1 if liq_press==1 & cope_sav==1    & cope_credit!=1
        mi xeq: replace cope_cat = 2 if liq_press==1 & cope_credit==1
        mi xeq: replace cope_cat = 0 if liq_press==1 & missing(cope_cat)
        label define cope 0 "other/unclear" 1 "savings_leaning" 2 "credit_leaning", replace
        mi xeq: label values cope_cat cope
        local SKIP_RQ2 0
    }
}
else local SKIP_RQ2 0

* Fit per imputation (binary logit among pressured) and predict per imputation
if `SKIP_RQ2'==0 {
    mi xeq: capture drop depdd_bin
    mi xeq: gen byte depdd_bin = (cope_cat==1) if liq_press==1

    mi xeq: quietly svy: logit depdd_bin ///
        c.dnnlaratio_cap c.sight_share ///
        c.dodsmortg c.dodiratio c.di2000 ///
        i.dhemph1_p i.dheduh1_p c.dh0002 i.dhhst_p i.dhdegurba_p ///
        i.ctry if liq_press==1, iterate(80) nolog

    mi xeq: capture drop __p_depdd_m
    mi xeq: quietly predict double __p_depdd_m if liq_press==1, pr

    * Collect and average by household
    * Before this block you should already have created the proxy cope_cat if needed,
* and run per-imputation model + predict:
* mi xeq: gen byte depdd_bin = (cope_cat==1) if liq_press==1
* mi xeq: quietly svy: logit depdd_bin ... if liq_press==1
* mi xeq: quietly predict double __p_depdd_m if liq_press==1, pr

preserve
tempfile D1 D2 D3 D4

forvalues m = 1/4 {
    mi extract `m', clear
    keep sa0100 sa0010 __p_depdd_m
    rename __p_depdd_m p_depdd_`m'
    sort sa0100 sa0010
    save `D`m'', replace
    restore, preserve
}

use `D1', clear
foreach m of numlist 2/4 {
    merge 1:1 sa0100 sa0010 using `D`m'', nogenerate
}
egen double p_depdd = rowmean(p_depdd_1 p_depdd_2 p_depdd_3 p_depdd_4)
keep sa0100 sa0010 p_depdd
tempfile DEP
save `DEP', replace
restore

* ---------- Build the export file ----------
preserve
mi extract 0, clear
* Keep some useful columns to accompany the predictions (edit as needed)
keep sa0100 sa0010 hw0010 ctry liq_press cope_cat ///
     dnnlaratio_cap sight_share dodsmortg dodiratio di2000 dh0002 ///
```

```
    dhemph1_p dheduh1_p dhhst_p dhdegurba_p

* Merge in p_liq (always) and p_depdd (if available)
merge 1:1 sa0100 sa0010 using `LIQ', nogenerate
capture confirm file `DEP'
if !_rc merge 1:1 sa0100 sa0010 using `DEP', nogenerate

order sa0100 sa0010 p_liq p_depdd liq_press cope_cat
export excel using "$OUT/Household_predictions.xlsx", replace firstrow(variables)
save "$OUT/Household_predictions_m0_V2.dta", replace
restore

*==============================================*
* 5) SEGMENTS + IRRBB RUN-OFF (from m=0)
*==============================================*
clear
use "$OUT/Household_predictions_m0_V2.dta", clear

* Ensure buffer & DSR inputs exist in the extracted file
capture confirm variable dnnlaratio_cap
if _rc {
   capture confirm variable dnnlaratio
   if !_rc gen double dnnlaratio_cap = cond(dnnlaratio<0,.,dnnlaratio)
}
capture confirm variable dodsmortg
if _rc {
   di as err "dodsmortg missing in predictions file — cannot form DSR segments."
   exit 111
}

* Quartiles 1..4
egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
egen q_dsr0 = cut(dodsmortg),      group(4) icodes
gen  q_buf = q_buf0 + 1
gen  q_dsr = q_dsr0 + 1
label define q4 1 "Q1 (low)" 2 "Q2" 3 "Q3" 4 "Q4 (high)", replace
label values q_buf q4
label values q_dsr q4

* Segment propensities (weighted means of model predictions)
collapse (mean) Seg_P_liq = p_liq ///
              Seg_P_depdd = p_depdd [pw=hw0010], by(q_buf q_dsr)

* IRRBB translation (tune these if you want)
local theta_mild   = 0.25
local theta_severe = 0.60
gen Runoff_mild   = Seg_P_liq * Seg_P_depdd * `theta_mild'
gen Runoff_severe = Seg_P_liq * Seg_P_depdd * `theta_severe'

export excel using "$OUT/Segments_runoff_IRRBB_V2.xlsx", replace firstrow(variables)
save "$OUT/Segments_runoff_IRRBB_V2.dta", replace
```

```
*------------------------------------------------------------*
* Table A: Variables & constructions (data dictionary)
*------------------------------------------------------------*
preserve
tempfile dict
* Choose the variables you want documented (add/remove as needed)
local vlist  ///
   sa0100 sa0010 IM0100 hw0010 m_imp ctry ///
   liq_press not_saving_room dsr40 arrears_any ///
   cope_cat cope_sav cope_credit ///
   da2101 da21011 da21012 sight_share dnnla dnnlaratio dnnlaratio_cap ///
   dodiratio dodsmortg di2000 dheduh1 dhemph1 dh0002 dhhst dhdegurba ///
   q_buf q_dsr

* Build dictionary dataset
postutil clear
tempname mem
postfile `mem' str32 varname str12 type str80 label using `dict', replace

quietly {
   foreach v of local vlist {
      capture confirm variable `v'
      if !_rc {
         local L : variable label `v'
         if "`L'"=="" local L "(no label)"
         describe `v'
         local T = r(type)
         post `mem' ("`v'") ("`T'") ("`L'")
      }
   }
}
postclose `mem'
use `dict', clear
export excel using "$OUT/TableA_data_dictionary.xlsx", replace firstrow(variables)

*------------------------------------------------------------*
* Table B: Weighted descriptives by income & wealth quintile
*------------------------------------------------------------*
* Make sure the HFCS analysis data are back in memory
clear
use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear

* (Optional sanity)
describe di2000 hw0010




* Income quintiles (0..4)
capture drop inc_q
egen inc_q = cut(di2000), group(5) icodes
label define q5_04 0 "Q1 (low)" 1 "Q2" 2 "Q3" 3 "Q4" 4 "Q5 (high)", replace
label values inc_q q5_04

preserve
   keep inc_q hw0010 liq_press sight_share dnnlaratio_cap dodsmortg dodiratio da2101
   collapse (mean) liq_press sight_share dnnlaratio_cap dodsmortg dodiratio da2101 [pw=hw0010], by(inc_q)
   export excel using "$OUT/TableB_descriptives_by_income_q.xlsx", replace firstrow(variables)
restore
```

```
* Wealth quintiles only if dnnla exists
capture confirm variable dnnla
if !_rc {
   capture drop wealth_q wealth_q0
   egen wealth_q0 = cut(dnnla), group(5) icodes
   gen  wealth_q  = wealth_q0 + 1
   label define q5_15 1 "Q1 (low)" 2 "Q2" 3 "Q3" 4 "Q4" 5 "Q5 (high)", replace
   label values wealth_q q5_15

   preserve
      keep wealth_q hw0010 liq_press sight_share dnnlaratio_cap dodsmortg dodiratio da2101
      collapse (mean) liq_press sight_share dnnlaratio_cap dodsmortg dodiratio da2101 [pw=hw0010],
by(wealth_q)
      export excel using "$OUT/TableB_descriptives_by_wealth_q.xlsx", replace firstrow(variables)
   restore
}



/-------------------------------------------------------------*
* Table C: Logit RQ1 results with AMEs (probability scale)
*--------------------------------------------------------------*
* Load tidy MI file
use "$OUT/hfcs2021_liquidity_model_MI_pweight.dta", clear

* If not MI-managed, (re)create MI context — tolerate r(459)
capture mi describe
if _rc {
   * Guards
   capture confirm variable m_imp
   capture confirm variable sa0100
   capture confirm variable sa0010
   if _rc {
      di as err "m_imp / sa0100 / sa0010 missing — re-run Blocks 1–3 & 10."
      exit 111
   }

   * Clean slate (no error if nothing to unset)
   capture mi unset

   * Import MI from long; IGNORE r(459)
   capture noisily mi import mlong, id(sa0100 sa0010) m(m_imp)
   if _rc==459 {
      di as txt "Data already MI-set; continuing."
   }
   else if _rc {
      di as err "mi import failed (rc = " _rc ")"
      exit _rc
   }

   * Register regular vars (no-op if already done)
   local regs sa0100 sa0010 hw0010 IM0100
   foreach v in dhemph1_p dheduh1_p dhhst_p dhdegurba_p ctry {
      capture confirm variable `v'
      if !_rc local regs `regs' `v'
   }
   capture mi register regular `regs'

   * Register remaining numeric as imputed (skip MI system + regulars)
   ds, has(type numeric)
```

```
   local skip `regs' m_imp _mi_m _mi_id _mi_miss _mi_vno _mi_style
   local imps
   foreach v of varlist `r(varlist)' {
      local pos : list posof "`v'" in skip
      if !`pos' local imps `imps' `v'
   }
   capture mi register imputed `imps'
}

* Confirm MI context is active
mi describe

* Survey settings for MI data (safe to re-apply)
mi svyset, clear
mi svyset [pw=hw0010]
mi svyset

* ---------- AMEs EXPORT ----------
* Ensure mimrgns exists
capture which mimrgns
if _rc ssc install mimrgns, replace

* RQ1 AMEs
capture estimates restore RQ1
if _rc estimates use "$OUT/RQ1_logit_liqpress.ster"
estimates restore RQ1
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(pr)
matrix A = r(table)'
putexcel set "$OUT/TableC_RQ1_AMEs.xlsx", replace
putexcel A1=("var") B1=("dy/dx") C1=("se") D1=("t") E1=("p") F1=("ll") G1=("ul")
putexcel A2=matrix(A), names

* RQ2 AMEs
capture estimates restore RQ2
if _rc estimates use "$OUT/RQ2_mlogit_coping.ster"
estimates restore RQ2
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(1))
matrix S = r(table)'
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(2))
matrix C = r(table)'

putexcel set "$OUT/TableD_RQ2_AMEs.xlsx", replace
putexcel A1=("Outcome: savings-leaning")
putexcel A2=("var") B2=("dy/dx") C2=("se") D2=("t") E2=("p") F2=("ll") G2=("ul")
putexcel A3=matrix(S), names
local base = rowsof(S) + 4
putexcel A`base'=("Outcome: credit-leaning")
putexcel A`= `base'+1' =("var") B`= `base'+1' =("dy/dx") C`= `base'+1' =("se") ///
      D`= `base'+1' =("t") E`= `base'+1' =("p") F`= `base'+1' =("ll") G`= `base'+1' =("ul")
putexcel A`= `base'+2' =matrix(C), names



/* (Optional) Save the RQ1 estimates object (already done earlier, but safe here)
capture estimates save "$OUT/RQ1_logit_liqpress.ster", replace*/

*------------------------------------------------------------*
* Table D: Multinomial RQ2 results by outcome (AMEs)
*------------------------------------------------------------*
estimates restore RQ2
```

```
* Outcome(1) = savings-leaning
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(1))
matrix S = r(table)'

* Outcome(2) = credit-leaning
mimrgns, dydx(dnnlaratio_cap dodsmortg sight_share dodiratio di2000 dh0002) predict(outcome(2))
matrix C = r(table)'

* Write both to one Excel
putexcel set "$OUT/TableD_RQ2_AMEs.xlsx", replace
putexcel A1=("Outcome: savings-leaning")
putexcel A2=("var") B2=("dy/dx") C2=("se") D2=("t") E2=("p") F2=("ll") G2=("ul")
putexcel A3=matrix(S), names
local base = rowsof(S) + 5
putexcel A`base'=("Outcome: credit-leaning")
putexcel A`= `base'+1' =("var") B`= `base'+1' =("dy/dx") C`= `base'+1' =("se") ///
        D`= `base'+1' =("t") E`= `base'+1' =("p") F`= `base'+1' =("ll") G`= `base'+1' =("ul")
putexcel A`= `base'+2' =matrix(C), names

 (Optional) Save RQ2 model
capture estimates save "$OUT/RQ2_mlogit_coping.ster", replace

*=================================================================*
* PROFILES OF VULNERABLE DEPOSITORS USING Household_predictions *
*=================================================================*

* 0) Load the household-level predictions (m = 0 extract)
 use "/Users/Dores/Desktop/TESE/OUTPUT/Household_predictions_m0_V2.dta", clear

* Sanity check
describe p_liq p_depdd liq_press hw0010 di2000 dnnlaratio_cap dodiratio ///
         dhemph1_p dheduh1_p dhhst_p dhdegurba_p

*--------------------------------------------------------------*
* 1) Joint run-off probability (only defined where p_depdd exists)
*--------------------------------------------------------------*
capture confirm variable p_runoff
if _rc {
   gen double p_runoff = p_liq * p_depdd
   label var p_runoff "Joint prob: liq stress & deposit drawdown"
}

*--------------------------------------------------------------*
* Helper: little program to build profile tables for any segment
*   - varseg: segment variable (already created)
*   - outxlsx: Excel file name (within $OUT)
*--------------------------------------------------------------*
capture program drop make_profile
program define make_profile
   syntax varname [ , OUTfile(string) ]

   local seg `varlist'
   if "`outfile'" == "" {
      di as err "You must specify , outfile() in make_profile"
      exit 198
   }

   preserve
```

```
      * ---------- (a) Unconditional means: p_liq and p_runoff ----------
      tempname agg_uncond agg_cond
      tempfile T_uncond T_cond

      collapse (mean) mean_Pliq = p_liq ///
                   mean_Prunoff = p_runoff ///
                   [pw = hw0010], by(`seg')
      save `T_uncond'

      restore, preserve

      * ---------- (b) Conditional means: p_depdd | liq_press == 1 ----------
      keep if liq_press == 1 & !missing(p_depdd)
      collapse (mean) mean_Pdepdd = p_depdd [pw = hw0010], by(`seg')
      save `T_cond'

      * ---------- (c) Merge and export ----------
      use `T_uncond', clear
      merge 1:1 `seg' using `T_cond', nogenerate

      order `seg' mean_Pliq mean_Pdepdd mean_Prunoff
      export excel using "$OUT/`outfile'", replace firstrow(variables)

   restore
end

*==================================================================*
* 2) SEGMENT 1 — Income quintiles (Q1–Q5)
*==================================================================*

capture drop inc_q inc_q0
egen inc_q0 = cut(di2000), group(5) icodes
gen  inc_q  = inc_q0 + 1
drop inc_q0
label define inc5 1 "Q1 (low income)" 2 "Q2" 3 "Q3" 4 "Q4" 5 "Q5 (high income)", replace
label values inc_q inc5
label var inc_q "Income quintile (HFCS)"

make_profile inc_q, outfile("Profile_by_income_quintile.xlsx")

*==================================================================*
* 3) SEGMENT 2 — Wealth quintiles (Q1–Q5), if dnnla is available
*==================================================================*

capture confirm variable dnnla
if !_rc {
   capture drop wealth_q wealth_q0
   egen wealth_q0 = cut(dnnla), group(5) icodes
   gen  wealth_q  = wealth_q0 + 1
   drop wealth_q0

   label define wealth5 1 "Q1 (low wealth)" 2 "Q2" 3 "Q3" 4 "Q4" 5 "Q5 (high wealth)", replace
   label values wealth_q wealth5
   label var wealth_q "Net wealth quintile (HFCS)"

   make_profile wealth_q, outfile("Profile_by_wealth_quintile.xlsx")
}
else {
   di as txt "Note: dnnla not found in Household_predictions; wealth profiles skipped."
}
```

```
*==============================================================*
* 4) SEGMENT 3 — Debt service ratio bands (DSR)
*    Example bands: <20%; 20–40%; >40%
*==============================================================*

capture drop dsr_band
gen byte dsr_band = .
replace dsr_band = 1 if dodiratio < 0.20  & !missing(dodiratio)
replace dsr_band = 2 if dodiratio >= 0.20 & dodiratio < 0.40
replace dsr_band = 3 if dodiratio >= 0.40 & !missing(dodiratio)

label define dsrband 1 "<20% of income" 2 "20–40% of income" 3 ">40% of income", replace
label values dsr_band dsrband
label var dsr_band "Debt service ratio band"

make_profile dsr_band, outfile("Profile_by_DSR_band.xlsx")

*==============================================================*
* 5) SEGMENT 4 — Buffer quartiles (liquid assets / income)
*==============================================================*

capture drop q_buf0 q_buf
egen q_buf0 = cut(dnnlaratio_cap), group(4) icodes
gen  q_buf  = q_buf0 + 1
drop q_buf0

label define q4 1 "Q1 (low buffer)" 2 "Q2" 3 "Q3" 4 "Q4 (high buffer)", replace
label values q_buf q4
label var q_buf "Liquidity buffer ratio quartile"

make_profile q_buf, outfile("Profile_by_buffer_quartile.xlsx")

*==============================================================*
* 6) SEGMENT 5 — Employment status of household head
*==============================================================*

* dhemph1_p already coded & labelled in the main do-file
label var dhemph1_p "Employment status (HH head)"
make_profile dhemph1_p, outfile("Profile_by_employment_status.xlsx")

*==============================================================*
* 7) SEGMENT 6 — Education of household head
*==============================================================*

label var dheduh1_p "Education level (HH head)"
make_profile dheduh1_p, outfile("Profile_by_education_level.xlsx")

*==============================================================*
* 8) SEGMENT 7 — Housing tenure (owner vs renter etc.)
*==============================================================*

label var dhhst_p "Housing tenure (HH)"
make_profile dhhst_p, outfile("Profile_by_housing_tenure.xlsx")

*==============================================================*
* 9) SEGMENT 8 — Degree of urbanisation
*==============================================================*

label var dhdegurba_p "Degree of urbanisation"
```

```
make_profile dhdegurba_p, outfile("Profile_by_urbanisation.xlsx")

*=================================================================*
* 10) (Optional) SEGMENT 9 — Country-level profiles
*=================================================================*

label var ctry "Country (HFCS code)"
make_profile ctry, outfile("Profile_by_country.xlsx")

*=================================================================*
* End of profiling block
* All Excel files written under: global OUT
*=================================================================*
display as res "Done. Segment profiles created in $OUT:"
display as txt " - Profile_by_income_quintile.xlsx"
display as txt " - Profile_by_wealth_quintile.xlsx (if available)"
display as txt " - Profile_by_DSR_band.xlsx"
display as txt " - Profile_by_buffer_quartile.xlsx"
display as txt " - Profile_by_employment_status.xlsx"
display as txt " - Profile_by_education_level.xlsx"
display as txt " - Profile_by_housing_tenure.xlsx"
display as txt " - Profile_by_urbanisation.xlsx"
display as txt " - Profile_by_country.xlsx"
```

## References


BALDAN, Cinzia; ZEN, Francesco. Liquidity risk and interest rate risk on banks: are they related? The IUP Journal of Financial Risk Management, Forthcoming, 2012.

BCBS (BASEL COMMITTEE ON BANKING SUPERVISION). Interest rate risk in the banking book: standards. Basel: Bank for International Settlements, April 2016. [online]. Available from: https://www.bis.org/bcbs/publ/d368.htm [accessed: 2025-05-01].

BEUTLER, Toni, et al. The impact of interest rate risk on bank lending. Journal of Banking & Finance, 2020, 115: 105797.

DEMARZO, Peter M.; KRISHNAMURTHY, Arvind; NAGEL, Stefan. Interest rate risk in banking. National Bureau of Economic Research, 2024.

DRECHSLER, Itamar; SAVOV, Alexi; SCHNABL, Philipp. Banking on deposits: Maturity transformation without interest rate risk. The Journal of Finance, 2021, 76.3: 1091-1143.

ENGLISH, William B.; VAN DEN HEUVEL, Skander J.; ZAKRAJŠEK, Egon. Interest rate risk and bank equity valuations. *Journal of Monetary Economics*, 2018, 98: 80-97.

ESPOSITO, Lucia; NOBILI, Andrea; ROPELE, Tiziano. The management of interest rate risk during the crisis: Evidence from Italian banks. Journal of Banking & Finance, 2015, 59: 486-504.

FRASER, Donald R.; MADURA, Jeff; WEIGAND, Robert A. Sources of bank interest rate risk. Financial Review, 2002, 37.3: 351-367.

GENTIL, Thibaut; RAY, Sébastien; TOADER, Oana. Projecting Banks' Net Interest Income: an Asset-Liability Approach, Applied to the Euro Area. 2023.

GOMEZ, Matthieu, et al. Banks' exposure to interest rate risk and the transmission of monetary policy. *Journal of Monetary Economics*, 2021, 117: 543-570.

LUBINSKA, Beata. Interest Rate Risk in the Banking Book: A Best Practice Guide to Management and Hedging. John Wiley & Sons, 2021.

SEIDEL, Laura. *Interest rate risk in the banking book goes ballet: balancing between earnings and economic value volatilities*. 2018. Master's Thesis. University of Twente.

STAIKOURAS, Sotiris K. The interest rate risk exposure of financial intermediaries: A review of the theory and empirical evidence. Financial Markets, Institutions & Instruments, 2003, 12.4: 257-289.