The impact of macroeconomic scenarios on recurrent delinquency

ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The impact of macroeconomic scenarios on recurrent delinquency: A stress testing framework of multi-state models for mortgages

Cecilia Bocchio, Jonathan Crook, Galina Andreeva *

Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK

ARTICLE INFO

Keywords:
Multi-state models
Stress testing
Retail credit risk
Mortgages
Transition probabilities

ABSTRACT

Transition probabilities between delinquency states play a key role in determining the risk profile of a lending portfolio. Stress testing and IFRS9 are topics widely discussed by academics and practitioners. In this paper, we combine dynamic multi-state models and macroeconomic scenarios to estimate a stress testing model that forecasts delinquency states and transition probabilities at the borrower level for a mortgage portfolio. For the first time, a delinquency multi-state model is estimated for residential mortgages. We explicitly analyse and control for repeated events, an aspect previously not considered in credit risk multi-state models. Furthermore, we enhance the existing methodology by estimating scenario-specific forecasts beyond the lag of time-dependent covariates. We find that the number of previous transitions have a significant impact on the level of the transition probabilities, that severe economic conditions affect younger vintages the most, and that the relative impact of the stress scenario differs by attributes observed at origination.

1. Introduction

The implementation of IFRS9[1] brought a substantial change in the way banks calculate their provisions. For the first time, financial institutions needed to consider not only past and current information but also forward-looking components when building their impairment reports. Expectations became a key element. Moreover, stress testing under IFRS9 resulted in a new challenge for banks, as forecasts of the staging criteria[2] should contemplate alternative scenarios. To calculate stage allocation distributions and expected credit losses, a lender needs to know the time an account is expected to keep the contractual repayments up to date before missing any payment that could ultimately lead to default. This expectation may be influenced not only by the borrower's behaviour but also by the economic environment. The existing practice of estimating losses for just one state (90 days past due or three months in arrears) is no longer sufficient. There is a need for more nuanced insights into the movements of the account between different levels of days past due. For these reasons, transitions between delinquency states matter. These transitions can occur more than once throughout the life of a loan, and the methodology followed to estimate them needs to take this into consideration.

* Correspondence to:
E-mail addresses: j.crook@ed.ac.uk (J. Crook), galina.andreeva@ed.ac.uk (G. Andreeva).

1 International Financial Reporting Standard 9.

2 Impairments are recognised under three stages: Stage 1, for financial assets not experiencing a significant increase in credit risk (SICR) since initial recognition; stage 2, for financial assets experiencing SICR since initial recognition; and stage 3, for credit-impaired financial assets. Twelve-month expected credit losses (ECLs) are recognised under stage 1 and lifetime ECLs under stages 2 and 3.

https://doi.org/10.1016/j.ijforecast.2022.08.005
0169-2070/

Please cite this article as: C. Bocchio, J. Crook and G. Andreeva, The impact of macroeconomic scenarios on recurrent delinquency: A stress testing framework of multi-state models for mortgages. International Journal of Forecasting (2022), https://doi.org/10.1016/j.ijforecast.2022.08.005.

In the last two decades, a large body of the literature has concentrated on the application of survival (or single-event) analysis to estimate credit risk metrics such as transitions into default (Banasik, Crook, & Thomas, 1999; Bellotti & Crook, 2009, 2013a, 2013b; Calabrese & Crook, 2020; Djeundje & Crook, 2019; Malik & Thomas, 2012) or recovery (Zhang & Thomas, 2012)[3]. Even though their contribution was to build the first set of applications of survival analysis in credit risk modelling, more complex specifications are needed to represent the actual dynamics of these events. A borrower is subject to transit alternative paths before defaulting, and anticipating these "pre-default" movements is crucial for risk management. This complexity can be captured by multi-state (or intensity) models which are able to characterise a subject experiencing (or not) different types of events across time. Our research not only aims to estimate and forecast transition probabilities but also "jumps" between delinquency states. For this reason, first-order Markov transition matrices are disregarded, as they compute average transition probabilities for all the accounts in the portfolio without distinguishing them based on subject-specific characteristics. This implies that it is not possible to infer which accounts will experience a specific transition and which will stay in the current state given that all the individuals face the same probability.

This paper makes the following three contributions to knowledge. First, we apply a multi-state delinquency model to a residential mortgage portfolio with the purpose of estimating and predicting six alternative transition probabilities between delinquency states. There is some research that has analysed transitions towards default and cure states for mortgages, such as Kelly and O'Malley (2016); however, this paper uses very few states. The events we analysed in this study have only been modelled for credit cards (Djeundje & Crook, 2018; Leow & Crook, 2014). Second, we connect the concept of multiple transition intensities to the estimation of account-level transition probabilities while contemplating repeated events and discontinuous risk intervals. This is the first paper in which recurrent delinquency is estimated for a mortgage portfolio using intensity processes. Third, the incorporation of macroeconomic scenarios into the estimation enabled us to enhance the existing methodology. This permitted us to forecast scenario-conditional transition probabilities and delinquency states beyond the lag of time-varying covariates, making this type of model suitable for both stress testing and IFRS9. Banks have found it challenging to predict provisions as they are required to do under the new Standard, and this paper enhances the available methodology still further compared with the existing literature. Overall, we bring a complete solution that overcomes the statement by Cope et al. (2022) that it is "complex to estimate and execute forecasts" with the state transition model approach.

Research regarding default probability in mortgage portfolios has a long history. One of the first studies was developed by Kau, Keenan, and Kim (1994), who indicated that the rationale behind the decision to enter into default is driven by house prices falling well below the value of the mortgage's termination option. Since the publication of Basel II in 2004, credit scoring has gained significant attention in default risk modelling, especially for mortgage portfolios which significantly benefited from the implementation of IRB models via capital savings. In recent years, there has been an increasing amount of literature on the estimation of default probabilities for mortgages based on survival analysis. These studies have introduced different characteristics to better represent movements into default. For example, Beran and Djaïdja (2007) built a mixture exponential model with immunity to analyse extreme censoring in default risk, finding that numerous mortgages may actually not be subject to default. McDonald, Matuszyk, and Thomas (2010) created a pricing model for UK mortgages, incorporating default probabilities based on Cox's model, that used application and macroeconomic data. They also performed Monte Carlo simulations to produce a distribution of cash flow forecasts for different economic scenarios. Park (2016) developed a competing risk survival model to analyse the probability of default or prepayment for uninsured, FHA-insured, and privately insured US mortgages, finding that adverse selection was present. Even though application and behavioural information was included in the estimation, macroeconomic data was not considered. Kiefer and Larson (2015) also examined the association between defaults and insured mortgages but concentrated their analysis on mortgages financed by mortgage insurance or a second mortgage. Calabrese and Crook (2020) incorporated spillover effects in a survival model that also allowed for variations in the coefficient estimates, both over time and over space.

The application of survival analysis for mortgages is not limited to transitions into default or default probabilities. Competing risk survival models have been applied to post-default mortgage data to analyse loss given default. Wood and Powell (2017) applied Fine and Gray's competing risk regression to estimate the probability of possession and the probability of cure using the LTV and the applicant's age at origination.

Multi-state models are not yet widely explored in credit risk modelling. The first set of empirical analysis was developed on credit rating migrations (Duffie, 2011; Duffie, Saita, & Wang, 2007; Lando & Skødeberg, 2002; Schuermann et al., 2003). Only in the past decade have studies directly analysed multiple delinquency events and their respective transition probabilities using intensity models for retail lending portfolios. Schechtman (2013) compared empirical transition matrices between delinquency states for consumer credit, applying cohort and survival approaches. Leow and Crook (2014) developed the first multi-state delinquency model for credit cards using account-specific application and behavioural data while considering six plausible types of events. Djeundje and Crook (2018) enhanced the methodology by incorporating macroeconomic indicators, flexible baseline hazards, and random effects to control for repeated events. Kelly and O'Malley (2016) used a portfolio of Irish residential mortgages to build a two-state model considering

3 More examples of seminal papers in this area can be found in Dirick, Claeskens, and Baesens (2017).

movements into and out of default. The results were then translated into transition probabilities assuming time homogeneity.

On the other hand, recurrent events within the survival analysis framework have been extensively analysed in medical studies (e.g. Amorim and Cai (2015), Guo, Gill, and Allore (2008), Sagara, Giorgi, Doumbo, Piarroux, and Gaudart (2014), Ullah, Gabbett, and Finch (2014), Wei and Glidden (1997)); however, researchers have not treated this aspect in much detail in the context of credit risk. By applying random effects into the intensity processes, only Djeundje and Crook (2018) analysed a credit card portfolio considering the inherent dependency between recurrent delinquency events, finding that these random effects were "strongly" significant. However, the authors did not expand on the relative impact recurrent events have on the transition probabilities. Previous studies have also failed to incorporate the effect of macroeconomic scenarios in multi-state intensity processes. Single-event models (Bellotti & Crook, 2009, 2013a, 2013b; Calabrese & Crook, 2020; Djeundje & Crook, 2019) and multi-state models (Djeundje & Crook, 2018; Kelly & O'Malley, 2016) applied to credit risk have successfully incorporated macroeconomic data in their estimations, but no study has dealt with the incorporation of scenarios to build forward-looking intensity processes in a multi-state framework. Bellotti and Crook (2013a, 2013b) simulated severe macroeconomic conditions to forecast time to default for a credit card portfolio using survival analysis, but the forecast horizon was restricted to the lag of the time-dependent covariates.

An understanding of the dynamics of transitions between delinquency states is highly important. The subprime financial crisis shed light on how much an economy depends on the banking system, and its soundness is heavily determined by the risk profile of lending portfolios. A major failure of risk assessment was to base it on a mortgage's past performance without considering economic impacts (Hott, 2015). Therefore, the determination of delinquency transitions based on macroeconomic scenarios would help banks to better manage their credit risk, especially under severe economic conditions. Regulators are particularly concerned about the effect of adverse scenarios, as "forecasts of significant losses can trigger remedial supervisory actions" (Kupiec, 2018). Moreover, stress testing is expected to continue being a relevant topic, either because of regulator requirements that change from time to time or because it is considered good risk management practice (Schuermann, 2014).

The remainder of this article is organised as follows. Section 2 presents the methodology to estimate transition probabilities for a mortgage portfolio using multi-state models and considering both recurrent events and discontinuous risk intervals. Section 3 relies on the estimation results to build scenario-conditional forecasts of transition probabilities. Finally, Section 4 concludes.

2. Modelling recurrent delinquency and recovery events

We aim to estimate and predict transition probabilities among different delinquency states for a UK mortgage portfolio applying multi-state (or intensity) models which rely on survival analysis but add the complexity of handling multiple types of failures or events. By the delinquency state, we refer to how many payments a borrower has missed by the contractual due date. We are interested in analysing the time an individual spends in a given delinquency state before moving into an alternative one, and to use this information to predict dynamic probabilities of movement among the states.

In survival analysis, the focus is on the time to a single event, and the variable of interest is duration time, which reflects the time a subject spends in a given state. Multi-state models are an extension of single-event models where different types of events are analysed (e.g. the same borrower can move from up to date to one payment down and then to two payments down). At the same time, repeated events can be explicitly considered. This feature is particularly important given that it avoids making the strong assumption that the attributes characterising the first event also characterise consecutive events of the same type. Ignoring recurrence when estimating delinquency or default risk leads to biased coefficient estimates, as failure times within the same subject are correlated. Therefore, a methodology that handles this absence of independence is needed. Finally, when considering multiple events, the objective is to analyse and understand the different paths an individual might undergo before reaching a certain state. It is then possible to build transition probabilities to predict future movements or migrations.

We define T to be a non-negative random variable which represents the survival time, and we assume that T is continuous. We consider those individuals who have not experienced the event of interest by time t. Specifically, we are interested in knowing the probability that the event will occur in some small interval $[t, t + dt)$ provided that it did not happen before t. This probability is given by $\alpha(t)dt$, where $\alpha(t)$ is a hazard function defined as

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t). \tag{1}$$

The most common survival model, known as the proportional hazard model, was introduced by Cox (1972). It states that the hazard function is related to a vector of covariates in the following way:

$$\alpha(t) = \alpha_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{x}(t)), \tag{2}$$

where $\alpha_0(t)$ is the (unspecified) baseline hazard function, $\exp(\boldsymbol{\beta}'\boldsymbol{x}(t))$ is a relative risk function, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of p coefficients that shows the effects of the covariates $\boldsymbol{x}(t)$, which may or may not vary over time t.

2.1. Transition intensities

Cox's model can be extended to the multi-state framework. The generalisation of the multiplicative intensity function considering multiple events and time-varying covariates is

$$\alpha_{hji}(t) = Y_{hi}(t) \alpha_{hj0}(t) \exp(\boldsymbol{\beta}'_{hj} \mathbf{Z}_{hji}(t)), \tag{3}$$

where $Y_{hi}(t)$ is the at-risk indicator process that assumes the value 1 if the individual i is in state h at time t, and 0 otherwise; $\alpha_{hj0}(t)$ is the (unspecified) non-negative baseline transition intensity from state h to state j; and $\exp(\boldsymbol{\beta}'_{hj}\mathbf{Z}_{hji}(t))$ is the relative risk function with unknown parameters $\boldsymbol{\beta}_{hj}$ for the individual-specific covariates $\mathbf{Z}_{hji}(t)$ (see Aalen, Borgan, and Gjessing (2008), Andersen, Borgan, Gill, and Keiding (1993)).

The estimators for the unknown parameters ($\boldsymbol{\beta}_{hj}$) are obtained by maximising the logarithm of a partial likelihood function derived from Cox's model. This estimation is performed for each transition $h \to j$, separately. Suppose we observe n failure times $t_1 < t_2 < \cdots < t_n$, where t_i is the failure time for the ith individual, and define $R(t_i)$ as the set of individuals at risk of experiencing the event at t_i^- (just before t_i). Then, the partial likelihood function (assuming no ties) is given by

$$L_{hj}(\boldsymbol{\beta}_{hj}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}'_{hj}\mathbf{Z}_{hji}(t_i))}{\sum_{l\in R(t_i)} \exp(\boldsymbol{\beta}'_{hj}\mathbf{Z}_{hjl}(t_i))}. \tag{4}$$

As the partial likelihood function is based on the ordered failure times, if tied events are observed, i.e. more than one failure at the same time t, then the partial likelihood function needs to be adjusted. The problem is that at a given failure time t where tied events are observed, it is not possible to determine the risk set composition. There are three common approaches to handle ties: the exact method and two approximations developed by Breslow (1974) and Efron (1977). In this research, the Efron approximation is used given that it considers how the risk set changes depending on the order of the tied events. If two subjects experience the event at the same time t, Efron assumes that the probability of each individual failing first is the same.[4] Then, the partial likelihood function is adjusted accordingly:

$$L_{hj}(\boldsymbol{\beta}_{hj})_{ef} = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}'_{hj}s_i(t_i))}{\prod_{r=0}^{d_i-1}\left[\sum_{l\in R(t_i)} \exp(\boldsymbol{\beta}'_{hj}\mathbf{Z}_{hjl}(t_i)) - rd_i^{-1}\sum_{l\in D(t_i)} \exp(\boldsymbol{\beta}'_{hj}\mathbf{Z}_{hjl}(t_i))\right]}, \tag{5}$$

where d_i is the number of individuals experiencing the event at the same time t_i and grouped in the set $D(t_i)$, and $s_i(t_i) = \sum_{l=1}^{d_i} Z_{hji_l}(t_i)$ measures the sum of covariates of subjects failing at time t_i.[5]

Transition intensities can be defined in different ways depending on the underlying population and the characteristics of the events. In this study, we analyse transitions (or migrations) among alternative delinquency states considering that the subjects might transit the same path more than once. There are several Cox-based models for repeated events (Kelly & Lim, 2000; Thenmozhi, Jeyaseelan, Jeyaseelan, Isaac, & Vedantam, 2019), e.g. the Andersen–Gill (AG) counting process (Andersen & Gill, 1982), the Prentice–Williams–Peterson gap time (PWP-GP) and total time (PWP-CP) models (Prentice, Williams, & Peterson, 1981), the Wei–Lin–Weissfeld (WLW) and Lei–Wei–Amato (LWA) marginal models (Lee, Wei, Amato, & Leurgans, 1992; Wei, Lin, & Weissfeld, 1989), and the Cox frailty model. Additionally, there are four components to be considered when analysing recurrent events (Kelly & Lim, 2000): (i) the definition of the risk interval, (ii) common versus event-specific baseline hazards, (iii) the definition of the risk sets, and (iv) within-subject correlation.

A risk interval is defined as the period of time during which an individual may experience the event. There are three alternative definitions:

- Counting process: the risk interval depends on previous events. In order to experience the kth event, first the subject has to experience the $(k-1)$-th event.
- Total time: the risk interval for the kth event is independent of the risk interval of any other event. This means the analysis time starts at zero and finishes at the time the event is observed.
- Gap time: the risk interval measures the time from the previous event, re-setting the clock after each event.

The determination of the baseline hazard is constrained to the selection of the risk set, as it is possible to differentiate between restricted, unrestricted, and semi-restricted risk sets. In the first case, the risk set for the kth event will be based on the information provided by all the subjects that experience $(k-1)$ events. In the second case, overlaps are allowed and two individuals can contribute to the same risk interval regardless of whether one subject is at risk of the kth event and the other subject is at risk of the mth event. The assumption of restricted risk sets implies the distinction of event-specific baseline hazards (stratification). On the other hand, a common baseline hazard is estimated under an unrestricted risk set. Semi-restricted risk sets are characterised by event-specific baseline hazards but allow subjects to be at risk of the kth event even if they experience less than $(k-1)$ events.

There are two methods to handle within-subject correlation. Variance-corrected approaches which estimate the model and correct the variance to take into consideration the dependency in the observations, and frailty or random effects which introduce a random variable to account for correlation between repeated events for the same subject by accounting for unmeasured heterogeneity not captured by covariates.

For the purpose of this study, we selected three model specifications: the Andersen–Gill counting process model, the PWP-CP model, and the PWP-GT model. These specifications are based on a variance-corrected approach to handle within-subject correlation. The AG model follows the structure presented in Eq. (3). It is characterised by an unrestricted risk set with a common baseline hazard for repeated events, and the risk interval is described by a counting process. The standard approach assumes independence between repeated events observed for the same subject, but in the AG model this assumption is relaxed

[4] For more information regarding the alternative methods, refer to Kalbfleisch and Prentice (2002) and Box-Steffensmeier and Jones (2004).

[5] For more details regarding the derivation of the partial likelihood function, refer to Andersen et al. (1993), Kalbfleisch and Prentice (2002), and Efron (1977).

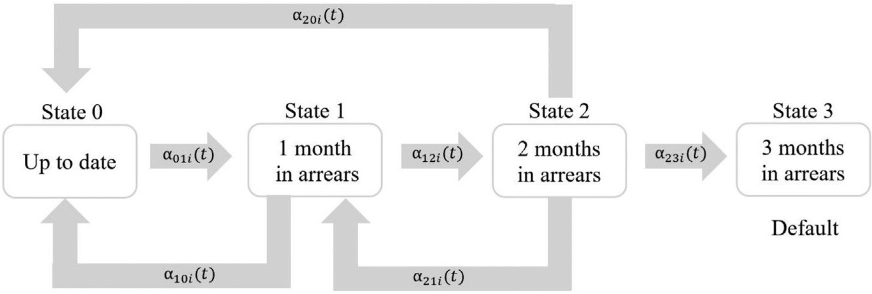


Fig. 1. Monthly transitions between delinquency states.

by incorporating the number of previous occurrences of the event for that subject (Kelly & Lim, 2000). Therefore, information regarding the times an account has been at risk of moving out of a given state is also considered. The PWP-CP model differs from the AG model in the determination of the risk set, which is a restricted one leading to event-specific baseline hazards, $\alpha_{hj0k}(t)$, where

$$\alpha_{hjik}(t) = Y_{hi}(t)\alpha_{hj0k}(t)\exp(\beta_{hj}'\mathbf{Z}_{hji}(t)), \tag{6}$$

and where k is the number of previous experienced events by time t.

As in the PWP-CP model, the PWP-GT method is defined by a restricted risk set, but the duration time is now defined as the time since the previous event $(t - t_{k-1})$. Therefore, the event-specific baseline hazard is given by $\alpha_{hj0k}(t - t_{k-1})$, where

$$\alpha_{hjik}(t) = Y_{hi}(t)\alpha_{hj0k}(t - t_{k-1})\exp(\beta_{hj}'\mathbf{Z}_{hji}(t)). \tag{7}$$

The AG model is a simpler specification, as it handles the dependence between subsequent events through time-varying covariates (such as the number of previous events), and the interest is in the overall effect on the intensity of the occurrence of a recurrent event (Amorim & Cai, 2015). On the other hand, the PWP-CP model and the PWP-GT model enable us to understand the impact of recurrent events at different duration times via stratification. However, the data need to be restricted to a maximum number of occurrences to ensure the stability of the coefficient estimates. Moreover, the PWP-GT model is convenient for repeated events observed at a low frequency or when the interest is in estimating the subsequent event (Amorim & Cai, 2015).

The methods described in this paper assume that censoring is non-informative, meaning that an individual who is censored at time c should be representative of all those subjects with the same values of explanatory variables who survive to c (Cox & Oakes, 1984). It would be inappropriate to exclude censored cases, as this can bring selection bias when censored data provide different information than uncensored data. Assuming non-informative or independent censoring still provides valid statistical inference (Aalen et al., 2008).

In line with Leow and Crook (2014) and Djeundje and Crook (2018), four delinquency states are defined:

- State 0 = Performing, i.e. the borrower is up to date.
- State 1 = One month in arrears, i.e. the borrower is 30 days past due.
- State 2 = Two months in arrears, i.e. the borrower is 60 days past due.
- State 3 = Three months in arrears, i.e. the borrower is 90 days past due. This is also the default definition.

Throughout the life of the mortgage, a borrower can transit between different states. Fig. 1 depicts transitions among the delinquency states defined above where the default definition is considered an absorbing state. The jumps are those possible in a one-month period, which results in six plausible types of monthly transitions $h \to j$: $0 \to 1$, $1 \to 2$, $2 \to 3$, $2 \to 1$, $1 \to 0$, and $2 \to 0$.

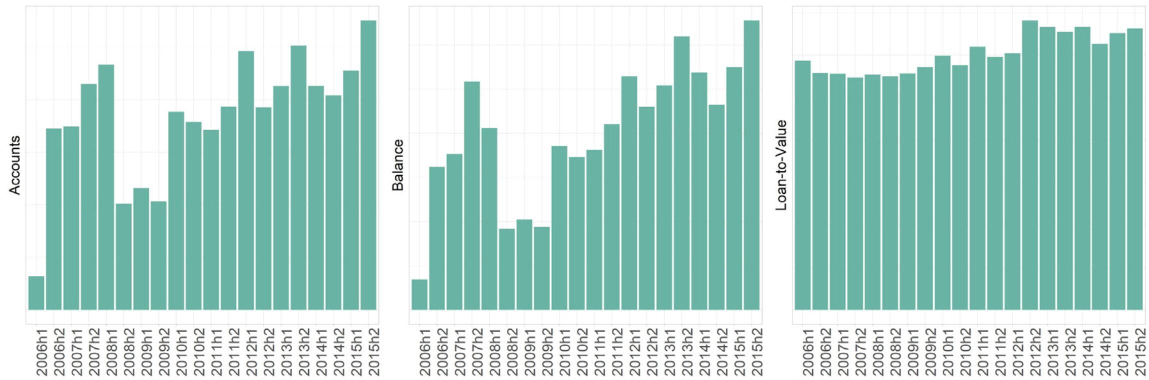
The intensity processes are structured such that they account not only for repeated events but also for discontinuous risk intervals. This means that an individual can transit from state h to state j more than once (repeated events), and it is possible that a subject is not at risk of moving into state j for a given period of time (discontinuous risk interval).

2.2. Data

2.2.1. Portfolio characteristics

Data on 67,827 mortgages originated from June 2006 to December 2015 from a UK financial institution were gained.[6] The dataset consists of both application data and monthly records of behavioural data. For each account, information from the origination date is available. Fig. 2 provides an overview of the portfolio's characteristics. The volume of new originations fluctuated over the time horizon, with new lending peaking in the first half of 2008. However, in 2009 the amount of new business reduced significantly, which suggests that the lender restricted the exposure during the international financial crisis (see panels (a) and (b)). At the same time, the average loan-to-value (LTV), a metric that represents how much collateral is kept behind the loan, decreased during this recession period (panel (c)). This is expected, as financial institutions tend to lend money to less risky customers under adverse macroeconomic environments. From 2010 onwards, the portfolio continued to grow.

[6] The portfolio data are not available for reasons of commercial confidentiality.



(a) Originated accounts (b) Originated balance (c) Average LTV by orig. date

Fig. 2. Portfolio data.

The portfolio comprises two types of mortgages: buy-to-let (BTL) and owner-occupied. Under a BTL contract, the borrower acquires the property with the objective of renting it out, while a non-BTL mortgage is acquired by those who buy a property to make it their residence. This portfolio is mainly characterised by owner-occupied mortgages (73.4% of the total lending). At the same time, we observe different repayment types: interest-only, repayment, and split. Under an interest-only scheme, the borrower commits to regular payments covering only the interest on the borrowed balance. This means that the principal does not change over time and the borrower has to pay it back at the end of the contract (or renegotiate new terms, if possible). Repayment mortgages are more common than interest-only mortgages and they consist of regular payments covering not only interest but also a proportion of the outstanding principal. If the borrower follows the repayment schedule, then the mortgage will be fully repaid by the end of the contract. Finally, split mortgages are a combination of interest-only and repayment mortgages. In this study, the portfolio presents 30.8% of interest-only mortgages, 64.5% of repayment mortgages, and 4.7% of split mortgages.

It is common to observe more than one borrower behind a mortgage contract. This could be the case when couples buy their first house, where a joint income enables them to access better contractual conditions. The portfolio shows mortgages with up to four applicants, with two applicants being the most common case (54.1%) followed by single applicants (45.3%).

2.2.2. Macroeconomic data

Macroeconomic variables for the UK economy are also analysed as potential covariates of the intensity processes given that when the economy deteriorates, borrowers find it more difficult to keep their financial commitments up to date (Crook & Banasik, 2012). The pool of potential candidates is selected based on economic theory and have been extensively used in the literature (Bellotti & Crook, 2009, 2013a, 2013b; Djeundje & Crook, 2018; Kiefer & Larson, 2015; Leow & Crook, 2016; Park, 2016). The following seven macroeconomic variables are included in the analysis:

- Unemployment rate. In a downturn, the percentage of unemployed individuals increases, worsening the repayment affordability, as an account holder is more likely to be unemployed and unable to repay as scheduled. Consequently, we would expect to see higher delinquency rates when the unemployment rate rises.
- Consumer price index (CPI). An increase in consumer prices can be seen as a deterioration of purchasing power, making debt repayments more difficult to afford. On the other hand, inflation might be caused by economic growth if the aggregate demand rises faster than the aggregate supply, in which case it would have a positive impact on repayments.
- House price index (HPI). The valuation of the property (the collateral) is reflected in the reported LTV, an indicator used by lenders to define the portfolio's risk profile. An increment of the property value is translated into a lower LTV and, therefore, into a less risky mortgage. This is the case because higher valuations are seen as increased capital value (equity) which the borrowers are less keen to lose, generating extra incentives to repay the mortgage and greater collateral for a bank to take over if the mortgage is not repaid.
- Five-year mortgage rate. Increases in the mortgage rate make payment commitments more expensive, affecting the individual's ability to repay a debt. Moreover, banks tend to offer higher interest rates to customers that are considered riskier, to cover the cost of default.
- Industrial production index (IPI). This index compiles information regarding the volume of production for several industries.[7] We used it as a proxy for the gross domestic product (GDP), which is only available on a quarterly frequency. Positive changes represent economic growth and higher incomes; therefore, we expect a reduction in transitions towards delinquency and an increase in recovery rates when the index of production improves.

[7] Manufacturing, mining and quarrying, energy supply, and water and waste management.

Table 1
Potential covariates.

Variable type	Variable	Units
Borrower's application data (time-invariant covariates)	Loan-to-value (LTV) at origination	%
	Number of applicants	1 to 4
	Marital status	Categorical
	Repayment type	Categorical
	Buy-to-let (BTL) indicator	Categorical
	Origination balance	£000
	Origination interest rate	%
Borrower's behavioural data (time-variant covariates)	Balance	£000
	Remaining balance (% of origination balance)	Basic points
	LTV	%
Other portfolio data (time-variant covariates)	Property value	£000
	Contractual interest rate	%
Macroeconomic data (time-variant covariates)	Unemployment rate	%
	Consumer price index	Index
	House price index	Index
	5-year mortgage rate	%
	Industrial production index	Index
	Consumer confidence index	Index
	FTSE-100	Index

- Consumer confidence index. This index provides an indication of expected future developments in household consumption and saving, based upon answers regarding their expected financial situation and their sentiments about the general economic situation, unemployment, and capability of saving. An increase in the index implies optimistic expectations regarding the economy, which can be translated into better recovery rates and lower levels of missed payments.
- FTSE-100. This index is a share price index of the largest 100 companies listed on the London Stock Exchange that summarises the financial market state for the UK. Positive variations in the FTSE-100 are associated with better financial conditions and, therefore, with lower delinquency and default rates.

The complete list of the variables considered is detailed in Table 1.

2.3. Estimation results

As one of the objectives is to predict transition probabilities, two independent samples were created: a training sample consisting of mortgage data observed between June 2006 and December 2011, and a test sample containing information relating to accounts originated from January 2012. All the estimations are developed on the training sample and validated on the test sample. To ensure independence between the two samples, information relating to accounts that originated before January 2012 but observed from January 2012 is excluded. We acknowledge that the selection of the time frames implies that the estimations are performed on data observed during the financial crisis, and tested on post-crisis information. Therefore, extra analyses considering alternative time frames were performed to understand the stability of the results. More details are available in Appendix.

To account for recurrent events, we define a stratum as the number of previous events experienced by the borrower at each duration time t. The low number of events in higher strata needs to be taken into consideration when estimating the transition intensities, as the reduced number of events of a given stratum might bring instability in the coefficient estimates (Amorim & Cai, 2015). There are at least three data manipulations that can be performed to mitigate this problem (Therneau & Hamilton, 1997). First, we could leave the information as it is and accept this instability for higher risk intervals. Second, we could disregard data with low events per stratum. Third, we could combine the events in higher risk intervals into one stratum. The last option was selected to estimate the transition intensities, as it avoids excluding data from the sample while bringing more robustness to the coefficient estimates. Table 2 presents the number of accounts by stratum after grouping. For example, we observe 2,269 accounts experiencing one payment down just once, and 441 accounts experiencing this event at least three times. The majority of the events are concentrated in movements out of state 0, where we observe accounts "failing" several times throughout the sample horizon. Moreover, the assumption of considering three months in arrears as an absorbing state is reflected in the definition of the stratum, as no account has previously moved out of state 3. This is also consistent with the standard definition of default as 90 days past due.

Table 2
Accounts by stratum.

Stratum	$0 \to 1$	$1 \to 2$	$2 \to 3$	$1 \to 0$	$2 \to 1$	$2 \to 0$
0	64,640	3,343	1,244	2,771	988	1,126
1	2,269	520		592	256	118
2	477			500		
3	441					
Total	67,827	3,863	1,244	3,863	1,244	1,244

Table 3 presents the estimated parameters for the six transition intensities under the AG and PWP models. For time-dependent variables the first, second, third, and sixth lags were considered to allow for delayed effects. Transformations of the macroeconomic variables were also used (annual growth rates and first differences). As

Table 3
Regression outputs.

Covariate	Transition intensity 0→1			Transition intensity 1→2			Transition intensity 2→3		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	1.519***			0.643***			n.a.		
LTV at Orig.	0.016***	0.016***	0.015***						
Repayment[a]	−0.218***	−0.206***	−0.210***						
One Applicant[b]							0.297*	0.297*	0.236**
Marital Status = Single[c]				0.413***	0.416***	0.349***			
Balance at Orig. (Log)	−0.162***	−0.170***	−0.157***						
Balance left (%), Lag 3	0.168***	0.172***	0.164***	0.210**	0.202*	0.175*	0.249***	0.249***	0.187***
LTV, Lag 3				0.003*	0.003*	0.005***			
Int. Rate, Lag 3	0.149***	0.146***	0.133***	0.090**	0.088*		0.311***	0.311***	0.256***
Unemp. Rate (D1)	0.560***	0.537***	0.549***						
Unemp. Rate (YoY)				0.954***	0.961***	0.558**	1.027**	1.027**	

Covariate	Transition intensity 1→0			Transition intensity 2→1			Transition intensity 2→0		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	−0.123***						0.638*		
One Applicant[b]						−0.210*			
Marital Status = Single[c]	−0.202***	−0.202***	−0.139***						−0.350**
BTL[d]			0.185***	0.343*	0.415**				
Interest Rate at Orig.									−0.302***
Balance at Orig. (Log)	0.123**	0.128**	0.144***						
LTV, Lag 1				−0.007**	−0.007**				
LTV, Lag 3	−0.008***	−0.008***	−0.008***			−0.007**			−0.007**
Property Value, Lag 3							0.135**	0.133**	
Int. Rate, Lag 3							−0.289***	−0.255*	
IPI (YoY), Lag 3	1.286**	1.232*							
IPI (D1)							0.225**	0.228**	0.253**
Unemp. Rate (D1)			−0.490***						
Mortgage Rate (D1)				−0.905*	−0.938*				

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
D1 = first difference; YoY = year-over-year growth rate.
[a]Repayment type: 1 = Repay, 0 = Interest-only.
[b]Number of applicants: 1 = One applicant, 0 = More than one applicant.
[c]Marital Status: 1 = Single, 0 = Non-single.
[d]BTL: 1 = Buy-to-let, 0 = Owner-occupied.

one of the purposes of this study is to build out-of-time forecasts of transition probabilities, only model specifications with significant coefficient estimates are considered. This implies using different combinations (and lags) of covariates to explain the dynamics behind each intensity process. We performed a pre-assessment based on univariate analysis by estimating each transition intensity on one covariate at the time, considering different lag specifications for both the behavioural variables and the macroeconomic factors in order to select a first pool of potential explanatory variables. We then combined several of these explanatory variables, and the final model specification was selected such that all covariates were significant at the 10% level. We also controlled for multicollinearity (by removing all pairs with correlation above 75%), for autocorrelation and heteroscedasticity (by implementing robust standard errors).

The covariates for the counting process models (AG and PWP-CP) are the same, and the respective coefficient estimates are similar, showing that the marginal effect of both the account-specific metrics and the macroeconomic environment is not impacted by the number of previously experienced events. The stratum (or number of previous events) is an explicit covariate of the AG model and it is significant in each estimation, except for transitions $2 \to 1$. Only the transition $1 \to 0$ shows a negative coefficient estimate, implying that the chances of recovering from one payment down decelerate if the account has experienced this event in the past. On the other hand, accounts that have previously been in the delinquent state face an increased risk of missing one payment. This risk is relatively large for movements out of state 0, as having transited once in the past increases the intensity rate by 356.8% ($exp(1.519) - 1$).

The different selection of covariates for the PWP-GT model is triggered by the definition of duration time, which is defined as the time since the loan was originated for the AG and PWP-CP models, and as the time since the previous event for the PWP-GT model. This selection of explanatory variables differs especially for intensities towards recovery ($1 \to 0$, $2 \to 1$, and $2 \to 0$).

Accounts with high LTV are more likely to miss one payment throughout the life of the contract. The LTV represents the proportion of exposure that is collateralised against the property, and it is a key element to be considered in any delinquency or loss model for a retail mortgage portfolio. The LTV at origination is determined by the bank's risk appetite. It describes how much risk the financial institution is willing to take. The behavioural or current LTV is affected by the payments made by the borrower and by the property's revaluation. Accounts with low LTV are considered less risky as they have lower incentives to default given that they would face a higher financial loss. Similarly, delinquent accounts with high current LTV are less prone to recover. These relationships are analogous to those of Kelly and O'Malley (2016).

The higher the percentage of balance outstanding, the higher the intensity rate of accounts moving into delinquency and into default, implying that borrowers tend to miss payments in the early stages of the repayment schedule. Moreover, accounts with a large absolute value of outstanding exposures at the origination date of the contract are less likely to miss a payment, as it is expected that low-risk accounts will have access to larger lending. Beran and Djaïdja (2007) found the opposite relationship in their survival model, concluding that large mortgages are at higher risk of default.

Repayment mortgages show lower transition rates towards delinquency (state 1) compared to interest-only (IO) accounts. Under an interest-only scheme, the borrower commits to regular payments covering only the interest on the principal. This means that the balance does not change over time and the borrower has to pay it back at the end of the contract (or re-negotiate new terms, if possible). Repayment mortgages consist of regular payments covering not only interest but also a proportion of the outstanding principal. IO accounts are more likely to default at the end of the contract when they need to make the final capital payment than throughout the life of the loan when the contractual payments are relatively low (compared to a repayment mortgage).

Banks charge higher interest rates to riskier accounts and this is reflected in the positive sign on the interest rate obtained for the transition towards delinquency and default. At the same time, increases in interest rates are translated into higher costs for borrowers under a flexible interest rate contract, meaning higher chances of missing a payment. Overall, mortgage accounts with single borrowers or one applicant are more likely to miss two or three payments, while married borrowers (or those in a partnership) have better chances of recovering.

Increases in delinquency and default rates are expected during downturns. For instance, poor macroeconomic conditions are reflected in a higher unemployment rate and in reductions of GDP growth. In economic recessions, individuals experience more difficulties keeping their finances up to date, resulting in an overall increase of missed payments. The opposite is expected under a benign economy, in which access to lending is easier and individuals have more chances of making repayments. A good economic environment helps borrowers in the up-to-date state to stay in that position, while delinquent accounts have better opportunities to recover. On the macroeconomic side, the change in the unemployment rate helps to explain transitions towards delinquency and default. As an economy deteriorates, the percentage of unemployed individuals increases, worsening the ability to repay. Consequently, we expect to see higher delinquency rates when the unemployment rate rises. The industrial production index and the five-year mortgage rate help to explain recovery movements, showing expectations aligned with those discussed in Section 2.2.

Finally, when the same covariate is selected for transition intensities to move in the opposite direction, the sign of the coefficient estimates also changes direction (e.g. marital status and current LTV for intensities $1\rightarrow2$ and $1\rightarrow0$), showing consistency in the impact of the selected drivers.

2.4. Transition probabilities

The intensity processes described by Eqs. (3), (6), and (7) represent the instantaneous risk of moving from one delinquency state to another one. However, practitioners and regulators are more interested in understanding the dynamics of the delinquent subjects. To do so, the results presented for each intensity process can be translated into account-specific transition probability matrices. These matrices show the probabilities $P_{hji}(s,t)$ of moving from any state h to any state j between any two points in time s and t, where $s < t$. Note that s and t are treated as non-negative integers in subsequent formulae. Specifically, for each borrower i and duration times s and t, it is possible to predict

$$\mathbf{P}_i(s,t)=\begin{pmatrix} P_{00i}(s,t) & P_{01i}(s,t) & P_{02i}(s,t) & P_{03i}(s,t)\\ P_{10i}(s,t) & P_{11i}(s,t) & P_{12i}(s,t) & P_{13i}(s,t)\\ P_{20i}(s,t) & P_{21i}(s,t) & P_{22i}(s,t) & P_{23i}(s,t)\\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (8)$$

where each row sums to 1, and the last row shows state 3 (the default state) as an absorbing state.

These transition probabilities are characterised by a non-homogeneous Markov process and can be estimated using the Aalen–Johansen estimator (Aalen & Johansen, 1978), also known as the product-integral estimator:

$$\mathbf{P}_i(s,t)=\prod_{u\in(s,t]}\{\mathbf{I}+d\mathbf{A}_i(u)\}\cong\prod_{u\in(s,t]}\{\mathbf{I}+\mathbf{A}_i(u)-\mathbf{A}_i(u-1)\}, \quad (9)$$

where $\mathbf{A}_i(t)$, known as the generator matrix, contains subject-specific cumulative transition intensities. The elements of $\mathbf{A}_i(t)$ are given by

$$\begin{aligned} A_{hji}(t,\boldsymbol{\beta}_{hj}) &= \int_0^t \alpha_{hji}(u)du \\ &\cong \sum_{\tau=0}^{t} Y_{hi}(\tau)\alpha_{hj0}(\tau)\exp(\boldsymbol{\beta}_{hj}'\mathbf{Z}_{hji}(\tau)), \quad (10)\\ A_{hhi}(t,\boldsymbol{\beta}_{hj}) &= -\sum_{h\neq j} A_{hji}(t,\boldsymbol{\beta}_{hj},\mathbf{Z}_{hji}(t)). \quad (11) \end{aligned}$$

The estimator of the cumulative hazard function is

$$\hat{A}_{hji}(t,\hat{\boldsymbol{\beta}}_{hj})=\sum_{\tau=0}^{t} Y_{hi}(\tau)\exp(\hat{\boldsymbol{\beta}}_{hj}'\mathbf{Z}_{hji}(\tau))d\hat{A}_{hj0}(\tau), \quad (12)$$

where the estimate for the increment of the cumulative baseline hazard, $d\hat{A}_{hj0}(t)$, when considering tied

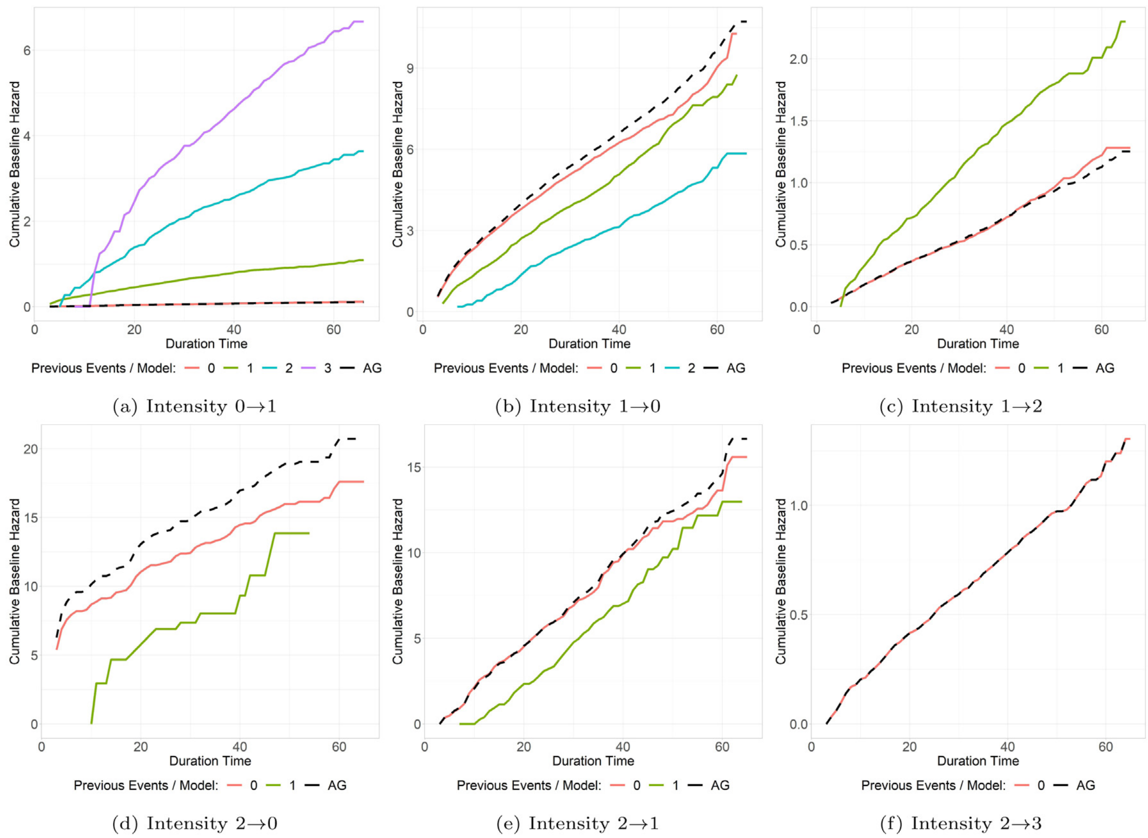


Fig. 3. Estimated cumulative baseline hazards for the AG and PWP-CP models.

events is[8]

$$d\hat{A}_{hj0}(t) = \sum_{r=1}^{d_i} \frac{1}{\sum_{i=1}^{n} \exp(\hat{\boldsymbol{\beta}}_{hj}'\mathbf{Z}_{hji}(t)) - \frac{r-1}{d_i}\sum_{i=1}^{d_i} \exp(\hat{\boldsymbol{\beta}}_{hj}'\mathbf{Z}_{hji}(t))}. \quad (13)$$

The estimated cumulative baseline hazards obtained from the Andersen–Gill model ($\hat{A}_{hj0}(t)$) and from the PWP-CP model ($\hat{A}_{hj0k}(t)$) are depicted in Fig. 3. The AG cumulative baseline hazards represented by the black dashed lines are common to all subjects, regardless of the number of previous events. On the other hand, the PWP-CP cumulative baseline hazards are stratified by the number of previous k events. The curve for the first stratum (i.e. for previous events = 0) uses information from all the accounts that are at risk of experiencing the event for the first time, while, for instance, the third stratum uses information of accounts that have experienced the events twice and are at risk for the third time. As the clock continues ticking between events, the cumulative baseline hazards for the different risk intervals are "shifted" towards the right given that an account has to first experience the event before being part of a higher stratum. The baseline cumulative hazards for movements towards delinquency are higher with a higher number of repeated events (see panels (a) and (c)). This means that those accounts that have already missed one or two payments in the past have a higher probability of doing it again. In contrast, borrowers are more likely to show a recovery if they miss one or two payments for the first time. However, this probability decreases if the account recurrently enters into a delinquency state (see panels (b), (d), and (e)), implying that even though the subject might have been able to totally or partially pay back the amount due, she struggles to repeat this action.

The common AG cumulative baseline hazard is generally aligned with the results obtained for the first stratum ($k = 0$) of the PWP-CP model. Given that default is assumed to be absorbent, the cumulative baseline hazard under both approaches is the same. In all cases, we observe that accounts in state 1 are more prone to recover than to move into state 2. At the same time, those borrowers that miss two payments are also more likely to recover than to move into default.

Similar to the PWP-CP model, the PWP-GT approach generates risk interval-specific baseline hazards. The difference relies on the treatment of the duration time. While the PWP-CP considers the exact month an event occurs (the time from entry), the PWP-GT looks at the time from the previous event. Fig. 4 shows that the PWP-GT results are less smooth, and this pattern is driven by

8 See Aalen et al. (2008), Ozenne, Sørensen, Scheike, Torp-Pedersen, and Gerds (2017).

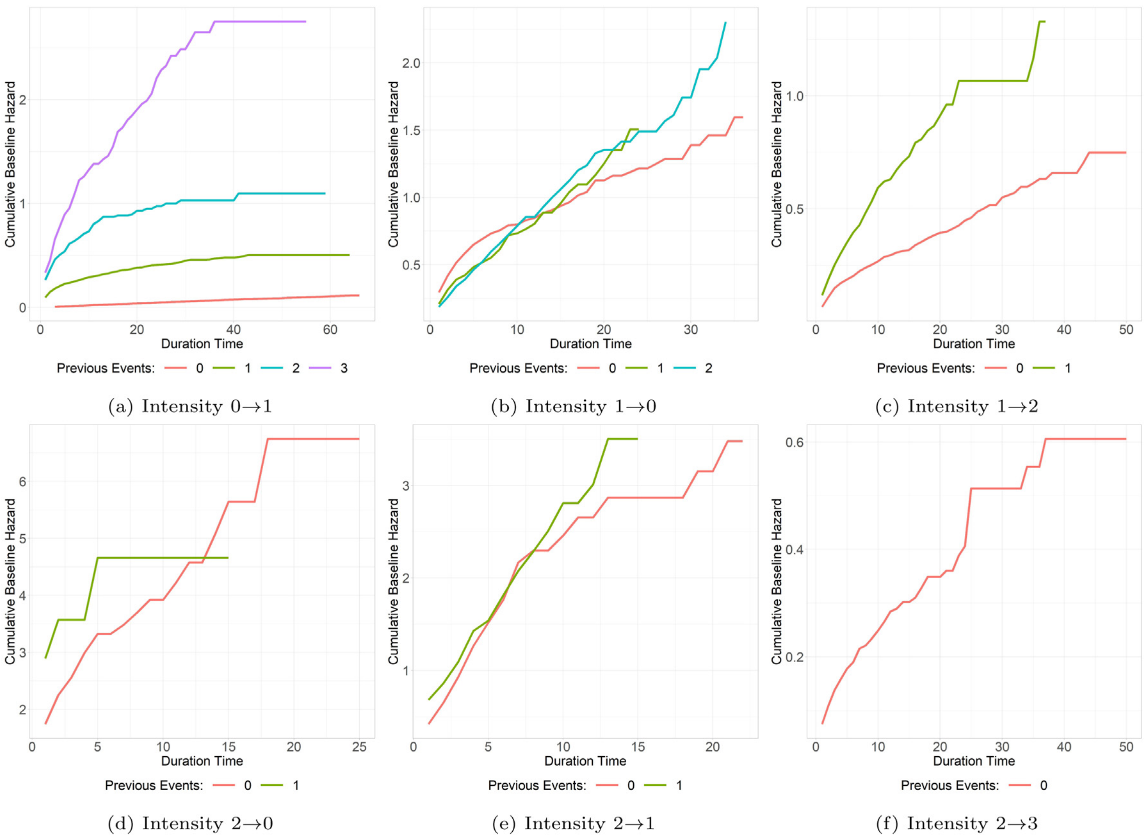


Fig. 4. Estimated cumulative baseline hazards for the PWP-GT model.

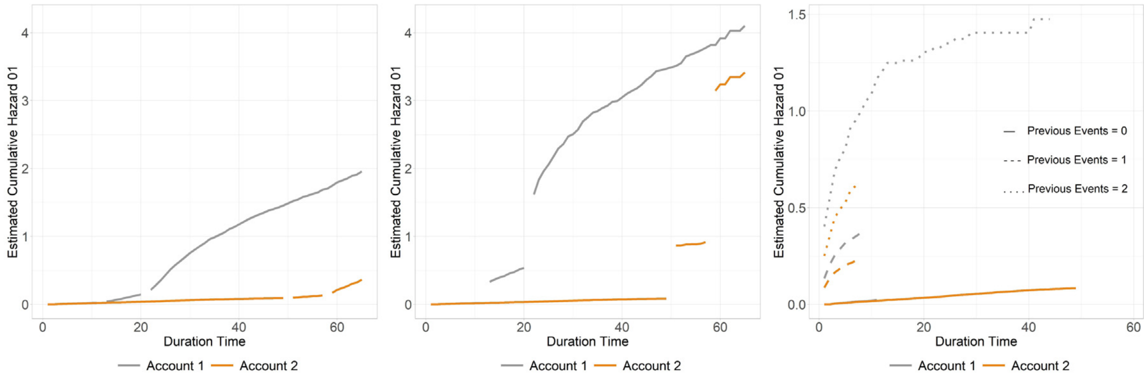
the concentration of observations for a given stratum. For instance, the recurrence of transition $2 \to 0$ is observed at different times since origination under the PWP-CP model (from $t = 10$ to $t = 50$), while under the gap time model the same observations are concentrated in the first six months since the previous event, with a maximum horizon of 16 months.

Fig. 5 presents the estimated cumulative hazard for transitions $0 \to 1$ (i.e. $\hat{A}_{01i}(t, \hat{\boldsymbol{\beta}}_{01})$) for two selected accounts experiencing repeated events. Account 1 moved from state 0 to state 1 at duration times $t = 12$ and $t = 21$, and moved back to state 0 at times $t = 13$ and $t = 22$. Account 2 transitioned at times $t = 50$ and $t = 58$, returning to state 0 at times $t = 51$ and $t = 59$. These "jumps" are represented by the discontinuity in the cumulative hazards. The discontinuous risk interval appears because the subject is no longer at risk of making the transition $0 \to 1$ at those specific duration times. When comparing the AG and PWP-CP models, Fig. 5 shows that every time an account misses one payment, the slope of the estimated cumulative hazard significantly increases, especially if the event is observed in the early months since origination. Moreover, the estimated impact is larger when we stratify the baseline hazard by the number of previous events (panel (b)). This characteristic is also observed under the PWP-GT model, where the duration time resets to zero after each observed transition (panel (c)).

The following step is to translate the estimated cumulative intensities into account-specific transition probabilities. In order to bound these probabilities to the space [0, 1], the Aalen–Johansen estimator (Eq. (9)) can only be applied to constant or increasing cumulative hazards. The characteristics behind the results obtained from the PWP-CP and the PWP-GT models (stratified hazards with lower or crossing cumulative baseline hazards for recurrent recovery events; see panels (b), (d), and (e) of Figs. 3 and 4) imply that we face a limitation to obtain the corresponding transition probabilities. However, these model specifications shed light on the impact recurrent events have on the instantaneous risk of transitioning. Consequently, in the remainder of this study, we concentrate on results obtained from the AG model (see Fig. 6).

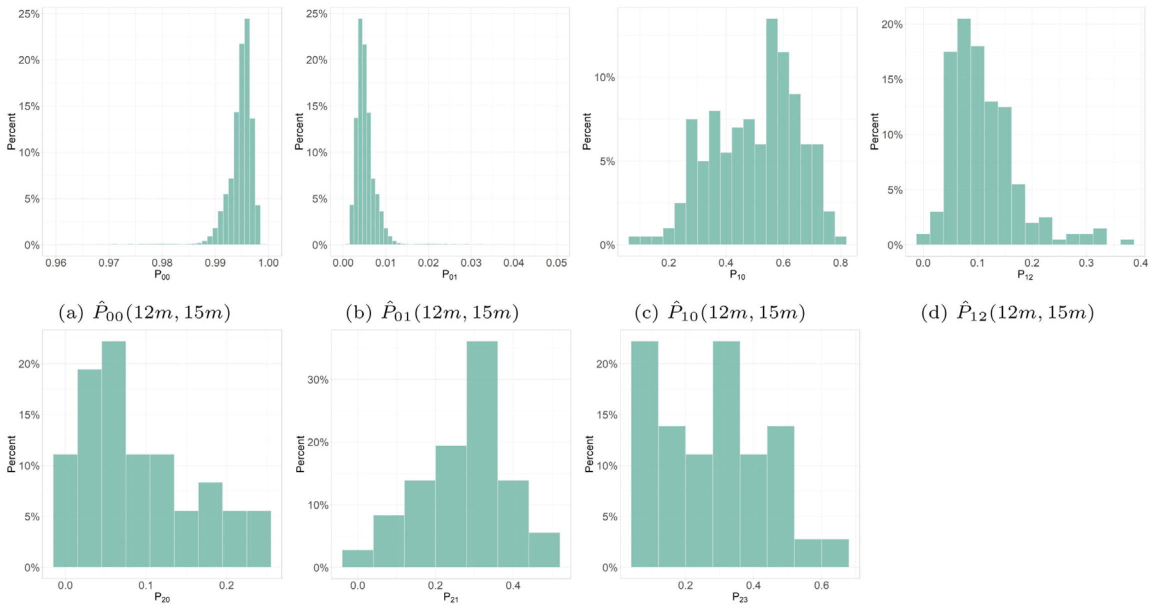
We assess the accuracy of the model by applying the methodology developed by Leow and Crook (2014) to test the results obtained through the estimation process of the transition intensities. The coefficient estimates from the AG model are applied to the accounts observed in the test sample, and transition probabilities $\hat{P}_{hji}(s, t)$ are computed for each account. These probabilities are then used to predict transitions from state h to state j. Finally, the predictions are compared against actuals (observed transitions) to validate the estimation results.

The first step is to define a time horizon, as the probabilities can be computed for any combination of duration times s and t. For this exercise, we are interested in three



(a) Andersen-Gill (b) PWP-CP (c) PWP-GT

Fig. 5. Estimated cumulative hazard, $\hat{A}_{01i}(t, \hat{\boldsymbol{\beta}}_{01})$. Example.



(a) $\hat{P}_{00}(12m, 15m)$ (b) $\hat{P}_{01}(12m, 15m)$ (c) $\hat{P}_{10}(12m, 15m)$ (d) $\hat{P}_{12}(12m, 15m)$

(e) $\hat{P}_{20}(12m, 15m)$ (f) $\hat{P}_{21}(12m, 15m)$ (g) $\hat{P}_{23}(12m, 15m)$

Fig. 6. Distribution of estimated transition probabilities between 12 and 15 months on book.

predictions: $\hat{P}_{hji}(s, s+1)$, $\hat{P}_{hji}(s, s+3)$, and $\hat{P}_{hji}(s, s+6)$. The aim is to study different time lengths, e.g. predictions one, three, and six months apart from the starting point. Given that any starting point could have been selected, we concentrated on accounts observed between six and 18 months since the origination date, i.e. $s = \{6, \ldots, 18\}$. The second step is to compute cut-off values (c_{hj}) to predict whether a transition takes place. These cut-offs are selected such that the proportion of accounts observed to go through the transition $h \to j$ in the training sample is the same proportion predicted to transit in the test sample. For each individual i in the test sample, the prediction to move to state j at time t given an initial state h at time s is defined as

$$j = \begin{cases} 0 & \text{if } \hat{P}_{h0i}(s,t) > c_{h0} \\ 1 & \text{if } \hat{P}_{h0i}(s,t) \le c_{h0} \text{ and } \hat{P}_{h1i}(s,t) > c_{h1} \\ 2 & \text{if } \hat{P}_{h0i}(s,t) \le c_{h0} \text{ and } \hat{P}_{h1i}(s,t) \le c_{h1} \\ & \quad \text{and } \hat{P}_{h2i}(s,t) > c_{h2} \\ 3 & \text{otherwise.} \end{cases} \tag{14}$$

The first step is to note the observed transitions (proportions) in the training sample between each pair of states. For instance, if we concentrate on movements out of state 0, these observed proportions are calculated as

$$P_{0j}^{train}(s,t) = \frac{N_{0j}^{train}(s,t)}{N_{00}^{train}(s,t)+N_{01}^{train}(s,t)+N_{02}^{train}(s,t)+N_{03}^{train}(s,t)}, \quad (15)$$

where $j = \{0, 1, 2, 3\}$.

We then estimate cut-off probabilities, c_{0j}, such that the proportion of accounts predicted to transit in the test sample matches the observed proportion in the training sample obtained in Eq. (15). To do so, the target number of accounts predicted to undergo the transition $0 \to 0$ in the test sample is defined as

$$\hat{N}_{00}^{test}(s,t) = P_{00}^{train}(s,t) * \left(N_{00}^{test}(s,t)+N_{01}^{test}(s,t)+N_{02}^{test}(s,t)+N_{03}^{test}(s,t)\right). \quad (16)$$

Once we define this target, we order the predicted transition probabilities $\hat{P}_{00}^{test}(s,t)$ from the highest to the lowest value. The first $\hat{N}_{00}^{test}(s,t)$ accounts obtained from this list are predicted to stay in state 0 ($0 \to 0$). The estimated transition probability at which this occurs is the cut-off value, c_{00}. We then repeat this exercise for the remaining transitions ($0 \to 1$, $0 \to 2$, and $0 \to 3$), and the same approach is followed for movements out of states 1 and 2.

This methodology possesses the disadvantage of disregarding the natural competing risk of the states, where the order of selecting the accounts to move to any delinquency state has an impact on the final results. For example, those accounts first selected to be in state 0 can no longer be selected to be in state 1 at time t. Instead of comparing all the estimated transition probabilities at the same time, a step-process is specified. Given that this method focuses on the observed proportions of accounts undergoing the different states, it ensures predictions among all the states, and therefore it was selected for this analysis.[9]

Table 4 summarises the prediction results obtained for accounts in the test sample. The table presents the average proportion of accounts in the test sample that are predicted to be in state j at time t after applying the cut-offs defined by Eq. (14) relative to the observed accounts in state j, i.e. $\frac{\text{predicted accounts in state } j}{\text{observed accounts in state } j}$. Moreover, bootstrapping analysis was performed to determine confidence intervals around these accuracy ratios. Specifically, for each combination of s and t analysed here, we generated 1000 random samples with replacement and calculated the 95th percentile confidence intervals. Notice that at higher states, the confidence interval is much larger than at lower states, and this is probably due to the smaller number of observations in those states.

Predictions for the up-to-date portfolio, that is for accounts in state 0, are very accurate across the three windows considered, as an accuracy ratio very close to 1 was obtained. On the other hand, the test sample has few cases in the delinquency states (less than 0.5% of the portfolio); therefore, it is reasonable to observe less accuracy in the results for the under-performing accounts. Except for transitions towards state 2 at time $t = s + 1$, the model tends to overestimate the number of accounts missing payments, especially for the defaulted population three and six months after the starting reference date, where the predicted number of accounts in state 3 is on average 2.61 times and 2.68 times, respectively, the number of observed accounts. To understand whether this overestimation is produced by model bias or by population drifts, we repeated this exercise on an 80% random sample obtained from the training sample (in-sample analysis). The difference between the predicted and observed values decreased compared to the results obtained for the test sample (see Table 5). This suggests that the overestimation might be explained by population drift and not by estimation biases. Note that population change is a known problem in credit risk modelling (Leow & Crook, 2016).

3. Scenario-conditional transition probabilities

Scenario-conditional forecasts for the six transition intensities are made using the parameter estimates obtained from the AG model. This implies that each component of Table 3 needs to be projected across a pre-defined horizon window. There are three types of covariates for which we need forecasts: application variables, behavioural variables, and macroeconomic variables. Application variables are, by definition, static across time, and the macroeconomic scenarios were obtained from the Bank of England.[10] A challenge is then to consistently forecast the account-level behavioural data considering that the economy and the delinquency state of the account that are unobserved at each future time t would have an impact on the account's behaviour. We present a method to do this in Section 3.2, below.

3.1. Macroeconomic scenarios

Every year, the Prudential Regulation Authority (PRA) publishes two stress testing scenarios (one baseline and one severe) that UK financial institutions are meant to use for the Internal Capital Adequacy Assessment Process (ICAAP). The scenarios are given such that consistency in the correlation among the different macroeconomic factors is retained. The PRA provides macroeconomic scenarios on a quarterly frequency. However, the transition intensities in our paper are estimated and projected on a monthly frequency, as movements across delinquency states occur from one month to another. This implies that an interpolation process to obtain monthly macroeconomic series is necessary. A simple approach could involve assuming a constant growth rate for the months belonging to the same quarter; nevertheless, this process would dismiss any volatility within a given quarter. To avoid this problem, a natural cubic spline interpolation was implemented by assuming that the information provided by the PRA matches the quarter-end values.

9 Refer to Djeundje and Crook (2018) for an alternative approach to define the cut-off values.

10 https://www.bankofengland.co.uk/stress-testing/2016/stress-test-scenarios-2016

Table 4
Prediction results: Out-of-time and out-of-sample average accuracy ratios.

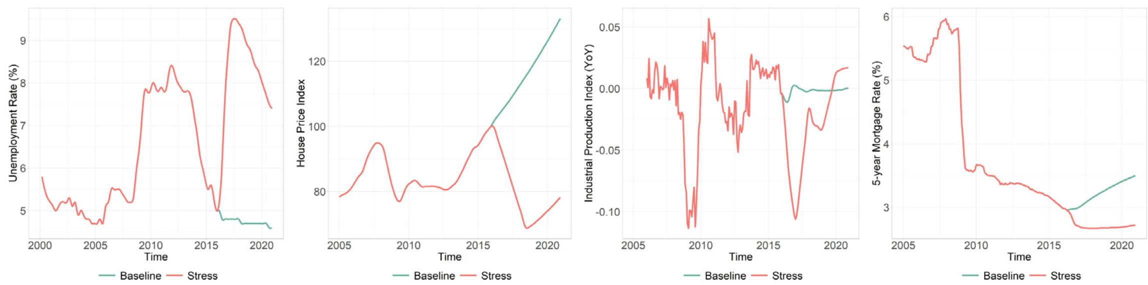
State	$t = s + 1$		$t = s + 3$		$t = s + 6$	
	Accuracy	(Conf. interval)	Accuracy	(Conf. interval)	Accuracy	(Conf. interval)
0	0.9989	(0.9988, 0.9991)	0.9984	(0.9982, 0.9986)	0.9979	(0.9977, 0.9982)
1	1.4249	(1.3588, 1.5309)	1.4754	(1.4048, 1.5929)	1.5449	(1.4734, 1.6681)
2	0.9926	(0.9127, 1.2778)	1.2930	(1.1826, 1.6537)	1.2300	(1.1340, 1.5371)
3	1.7179	(1.2778, 2.7969)	2.6090	(1.8078, 3.7287)	2.6778	(1.8922, 4.2445)

Note: The table presents the number of predicted accounts relative to the number of observed accounts in each state j for borrowers in the test sample. The ratios are mean values across duration times $s = \{6, \dots, 18\}$. We present the 95th percentile bootstrapped confidence intervals in parentheses.

Table 5
Prediction results: In-sample average accuracy ratios.

State	$t = s + 1$		$t = s + 3$		$t = s + 6$	
	Accuracy	(Conf. interval)	Accuracy	(Conf. interval)	Accuracy	(Conf. interval)
0	0.9996	(0.9994, 0.9998)	0.9993	(0.9991, 0.9995)	0.9990	(0.9987, 0.9992)
1	1.0994	(1.0583, 1.1573)	1.1387	(1.0930, 1.2044)	1.1672	(1.1267, 1.2434)
2	0.9725	(0.9043, 1.1229)	1.1326	(1.0517, 1.3551)	1.2032	(1.1029, 1.4551)
3	0.8751	(0.8002, 1.3647)	1.0502	(0.9083, 1.7036)	1.2152	(1.0444, 2.1550)

Note: The table presents the number of predicted accounts relative to the number of observed accounts in each state j for borrowers in the training sample. The ratios are mean values across duration times $s = \{6, \dots, 18\}$. We present the 95th percentile bootstrapped confidence intervals in parentheses.



(a) Unemployment rate (b) House price index (c) Indust. prod. index (YoY) (d) 5-year mortgage rate

Fig. 7. UK macroeconomic scenarios used for forecasting transition intensities.

Based on the results presented in Table 3, the macroeconomic factors determining the alternative transition intensities are the house price index (HPI) used to estimate the borrower's property value, the unemployment rate, the industrial production index, and the five-year mortgage rate. Given that the PRA only provides scenario-specific forecasts for the HPI and the unemployment rate, it was necessary to extend the PRA's assumptions to obtain scenario-specific forecasts for the industrial production index and for the mortgage rate. To do so, the PRA's available information was used as a proxy to expand the scenarios. We performed regression analyses to select the combination of PRA variables that better represent the movements of the target variables based on in-sample fit. We applied the estimated coefficient estimates to the selected explanatory variables to obtain estimated baseline and stressed values for the industrial production index and mortgage rate.[11]

The macroeconomic series used to forecast the transition probabilities under the alternative scenarios are presented in Fig. 7.

[11] Refer to Appendix A.3 in the Appendix for more details relating this methodology.

3.2. Forecasting scenario-conditional transition intensities and transition probabilities

The UK mortgage data show account-level application and monthly behavioural information until December 2015 for accounts originated between June 2006 and December 2015. A two-year forecast starting in January 2016 was chosen to allow for enough spread between the baseline and stress scenarios. All non-defaulted accounts that originated between July 2014 and June 2015 and were still open as of December 2015 were selected to produce the forecasts. The selection of alternative origination dates implies that different duration times are considered at the starting point (i.e. January 2016).

The first step to forecast transition probabilities is to produce scenario-conditional cumulative hazard functions. For each scenario *scen*, the cumulative hazard is estimated following Eq. (17):

$$\hat{A}^{scen}_{hji}(t, \hat{\boldsymbol{\beta}}_{hj}) = \sum_{\tau=0}^{t} Y^{scen}_{hi}(\tau) \exp(\hat{\boldsymbol{\beta}}'_{hj} \mathbf{Z}^{scen}_{hji}(\tau)) d\hat{A}_{hj0}(\tau). \tag{17}$$

The cumulative baseline hazard ($\hat{A}_{hj0}(t)$) is not risk interval-specific, meaning that each borrower receives the same specification regardless of how many times she has experienced the event of interest. However, this variance-corrected approach uses the risk interval information to account for dependency on repeated events for the same subject, implying that we need to know how many times an account has transited by any future time t.[12]. A one-step process was developed:

1. Select all non-defaulted accounts as of the latest available reporting date. Detect the final delinquency status to allocate the accounts into the corresponding risk set considering if the accounts have already experienced the event.
2. For each account, estimate the cumulative intensity since the origination date (in-sample estimation).
3. For each account and for each macroeconomic scenario, forecast the six transition intensities one month out-of-time, i.e. from t to $t+1$.
4. For each account and for each macroeconomic scenario, estimate the one-month transition probability using the Aalen–Johansen estimator:

$$\hat{\mathbf{P}}_i^{scen}(t, t+1) = \mathbf{I} + \hat{\mathbf{A}}_i^{scen}(t+1) - \hat{\mathbf{A}}_i^{scen}(t). \quad (18)$$

5. Based on the transition probabilities estimated in step 4, decide whether each account undergoes the transition $h \rightarrow j$. A series of cut-off values was defined by applying Eq. (14) to each single one-month transition using the training and test samples. This meant obtaining different cut-off values for different duration times. Because the cut-off values were determined in-sample, these are not scenario-dependent. The comparison between the estimated $\hat{\mathbf{P}}_i^{scen}(t, t+1)$ and the cut-off values determines whether a transition takes place between t and $t+1$. The estimated delinquency status obtained at this stage becomes the starting point for forecasting the transition in the following month, i.e. for $\hat{\mathbf{P}}_i^{scen}(t+1, t+2)$.
6. Based on the delinquency status estimated in step 5, determine the current value for the behavioural variables.
7. Reset the at-risk indicator $Y_{hi}(t)$ based on the results from step 5, and repeat from step 3 until reaching the end of the forecasting window.

The steps listed above involve different levels of sophistication. The challenge appears in the forecast of the behavioural variables as the application variables remain constant across time and the macroeconomic scenarios are given by the series depicted in Fig. 7. The time-varying portfolio data needed for the estimation process are account-level information for the exposure (or balance), the property value (to determine the current LTV), and the interest rates at each future time t. As the behavioural data used in the estimation processes are lagged, we first decide whether the account experiences a transition and we then estimate the *current* value of the behavioural information based on this decision.

Balance: if the account is selected to be performing (i.e. the account is selected to be in state 0) the balance is amortised following the expected contractual repayment:

$$\begin{aligned} Balance_i^{scen}(t) &= Balance_i^{scen}(t-1) \\ &\quad - \left(ContractualRepayment_i - Interest_i^{scen}(t)\right), \end{aligned} \quad (19)$$

where $ContractualRepayment_i$ is the account-specific contractual repayment covering both balance amortisation and interest payments, and $Interest_i^{scen}(t)$ is the interest paid on the remaining balance, defined as

$$Interest_i^{scen}(t) = Balance_i^{scen}(t-1) \times MonthlyInterestRate_i. \quad (20)$$

We assume that the contractual repayment remains constant across time, i.e. that the borrower does not make significant overpayments and the contractual interest rate does not change. If the account is selected to be underperforming (i.e. in state 1 or 2), the balance observed in the previous month is increased by accrued interest:

$$\begin{aligned} Balance_i^{scen}(t) &= Balance_i^{scen}(t-1) \\ &\quad \times (1 + MonthlyInterestRate_i). \end{aligned} \quad (21)$$

Property value: the collateral value is indexed using the scenario-specific house price index (HPI):

$$\begin{aligned} CollateralValue_i^{scen}(t) &= CollateralValue_i^{scen}(t-1) \\ &\quad \times \frac{HPI^{scen}(t)}{HPI^{scen}(t-1)}. \end{aligned} \quad (22)$$

This information is used to forecast the borrower's LTV at each future reporting time t:

$$CurrentLTV_i^{scen}(t) = \frac{Balance_i^{scen}(t)}{CollateralValue_i^{scen}(t)}. \quad (23)$$

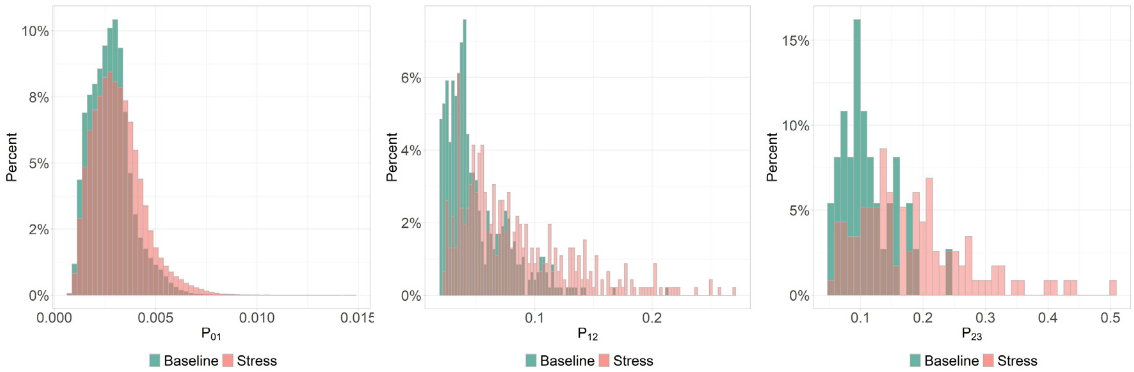
Contractual interest rate: assumed to be fixed throughout the life of the mortgage.

This methodology enabled us to build a full path of transition probabilities considering, for example, a two-year forecast (from January 2016 to December 2017) under alternative macroeconomic scenarios. Fig. 8 compares the distributions of the estimated transition probabilities for movements towards delinquency and default. In line with economic intuition, transition probabilities under stressed scenarios are higher compared to the baseline. This is evidenced by the fatter tail of the stressed distributions.

As the vast majority of the portfolio does not show missed payments, the remainder of this analysis is concentrated on transitions out of state 0. Fig. 9 presents the average relative spread between stress and baseline cumulative hazards for transitions $0 \rightarrow 1$ by origination vintage.[13] Panel (a) shows that while the spread follows a similar shape across all origination vintages, the level of the peak is significantly different. By December 2016, the average cumulative hazard under stress is 12.1% higher than the cumulative hazard under baseline for the 2015m6 vintage, while for the 2014m7 vintage,

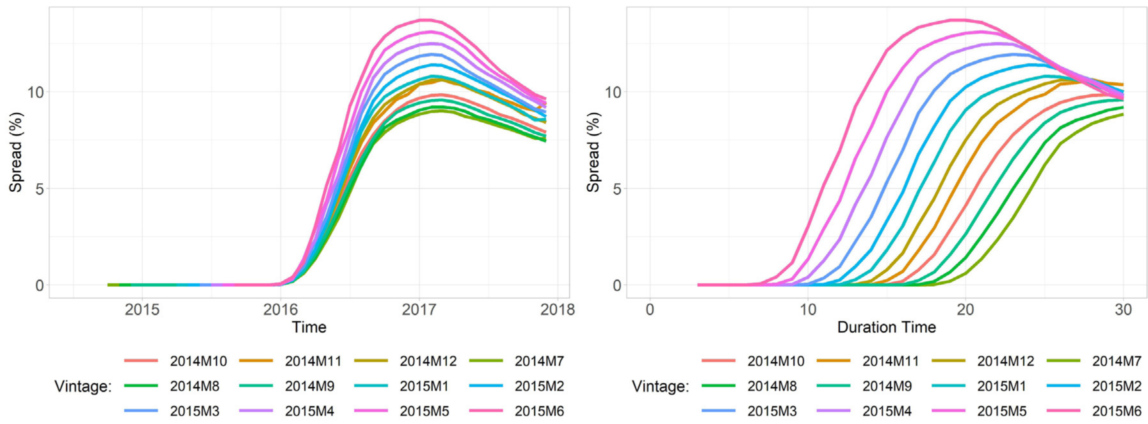
12 The stratum is a component of $\mathbf{Z}_i^{scen}(t)$ in Eq. (17).

13 The relative spread is calculated as $\left(\frac{A_{01}^{st}(t)}{A_{01}^{bl}(t)} - 1\right) \times 100$, where bl = baseline and st = stress.



(a) $\hat{P}_{01}^{scen}$ (b) $\hat{P}_{12}^{scen}$ (c) $\hat{P}_{23}^{scen}$

Fig. 8. Distributions of scenario-conditional transition probabilities.



(a) By calendar time (b) By duration time

Fig. 9. Average cumulative hazard $0 \rightarrow 1$. Relative spread stress/baseline.

this spread is approximately 7.5%. This implies that the relative risk of moving into one payment down under stress decreases the longer the account has been on book.

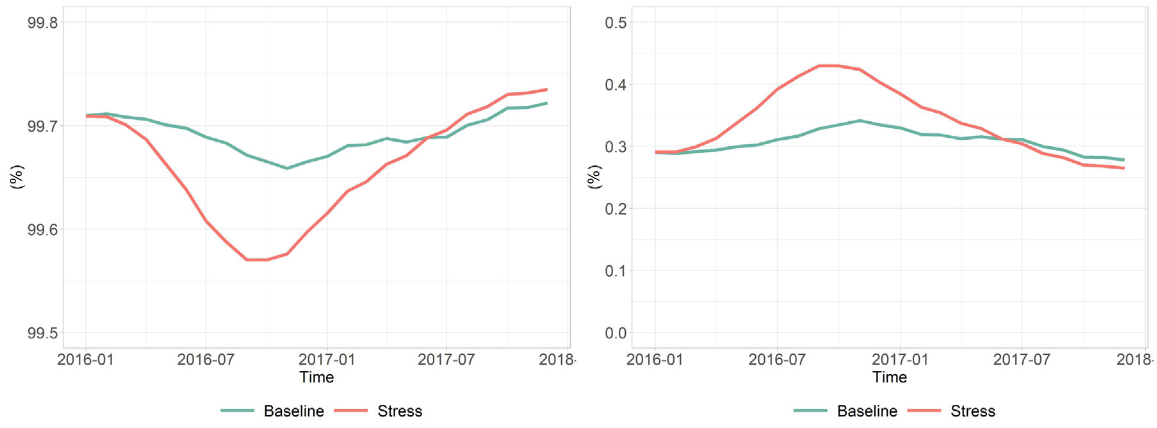
Fig. 10 presents the average transition probabilities for both baseline and stress scenarios. In line with economic theory, the probability of staying in state 0 is lower under stressed conditions. Similarly, movements into delinquency are more likely in a downturn.

The scenario-conditional transition probabilities also differ by segments. Repayment accounts and single applicants show a higher risk of entering into delinquency than repayment and multiple/married applicants. This line of reasoning is also reflected in the analysis by scenario. Fig. 11 shows average transition probabilities $\hat{P}_{01}^{scen}(t)$ by scenario and origination segment. While the spread between stress and baseline figures for a given segment is similar, the level of the transition probability is significantly different. For example, by October 2016, the average stressed $\hat{P}_{01}^{scen}(t)$ is 0.38% for IO accounts and 0.45% for repayment accounts, while the spread between stress and baseline for both segments remains around 28.4% (panel (b)). Similarly, single applicants face an average transition probability of 0.47% in October 2016, while married applicants have a lower probability of 0.40%, but in both cases the spread between stress and baseline scenarios is similar (panel (c)).

It is possible to apply the Aalen–Johansen estimator to obtain transition probabilities between any duration times under each macroeconomic scenario by applying Eq. (24):

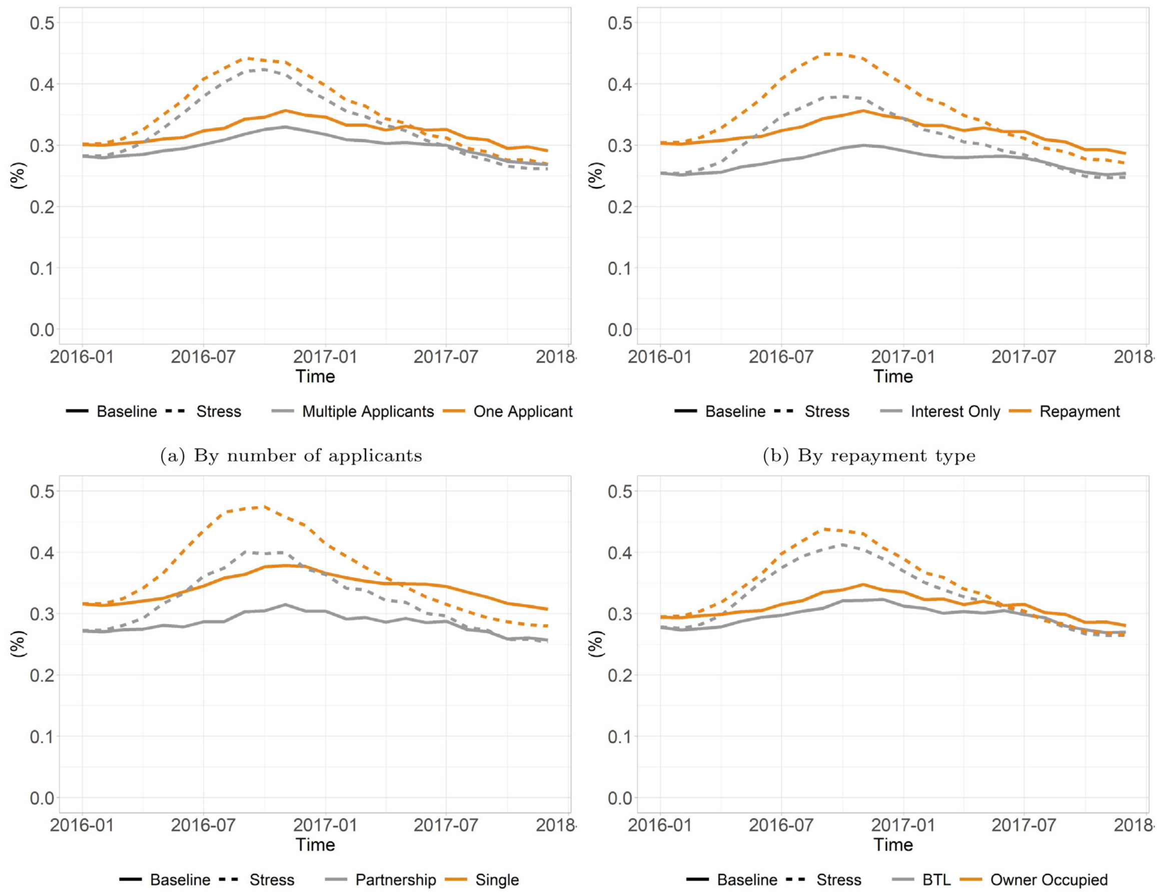
$$\hat{\mathbf{P}}_i^{scen}(s,t) = \prod_{(s,t]} \{\mathbf{I} + \hat{\mathbf{A}}_i^{scen}(u) - \hat{\mathbf{A}}_i^{scen}(u-1)\}. \tag{24}$$

Fig. 12 presents the average one-, three-, and six-month transition probabilities, where, considering the up-to-date book as the starting state, we plot the probability of being in state 0 (panel (a)) or in state 1 (panel (b)) one, three, and six months after. Because these are forward transition probabilities, we expect to observe the peak of the stressed scenario at different calendar times. For



(a) $\hat{P}_{00}^{scen}(t, t+1)$ (b) $\hat{P}_{01}^{scen}(t, t+1)$

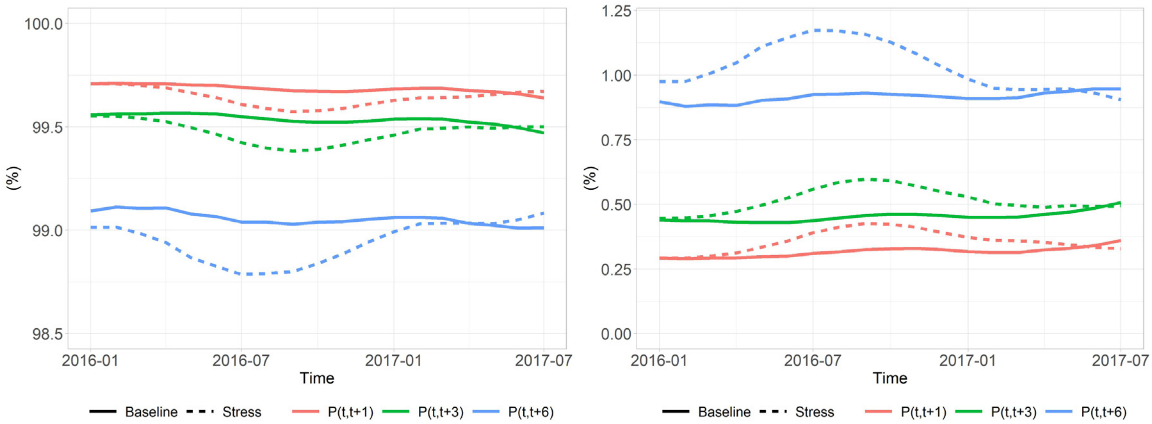
Fig. 10. Average one-month transition probabilities by scenario.



(c) By marital status (d) By occupancy type

Fig. 11. Average one-month transition probabilities $\hat{P}_{01}^{scen}(t, t+1)$ by scenario and segment.

instance, $\hat{P}_{01}^{scen}(t, t+1)$ incorporates transition intensity information of the following month, while $\hat{P}_{01}^{scen}(t, t+6)$ considers transition intensity information of the next six months. Furthermore, the results show that the levels of the average transition probabilities change together with the forecast time horizon, as it is more probable that an account that is in state 0 at any time t will stay in the same state the following month, but it has more chances



(a) $\hat{P}^{scen}_{00}(t, t+1)$, $\hat{P}^{scen}_{00}(t, t+3)$ & $\hat{P}^{scen}_{00}(t, t+6)$ (b) $\hat{P}^{scen}_{01}(t, t+1)$, $\hat{P}^{scen}_{01}(t, t+3)$ & $\hat{P}^{scen}_{01}(t, t+6)$

Fig. 12. Average one-, three-, and six-month transition probabilities by scenario.

of moving towards delinquency in the next six months (under any scenario).

3.3. *Model usage: Practical implications*

An added complexity that IFRS9 brought to financial institutions is forecasts of forward-looking expected credit losses when performing stress testing exercises. IFRS9 requires a bank to allocate accounts into stages if a significant increase in credit risk since the initial recognition has been observed. The criteria to trigger a change in stage allocation can be quantitative (based on default risk) or qualitative. The repayment performance or delinquency status is an example of a qualitative criterion. Therefore, being able to estimate and predict transitions between delinquency states for any macroeconomic scenario enables the lender to better reflect the impact of default risk in future provisions.

For instance, predicting the timing of a movement from up to date (state 0) to one payment down (state 1) translates into triggering a staging criterion to allocate the loan into stage 2, resulting in the calculation of lifetime expected credit losses. Multi-state models allow one to infer the time an account will spend in a given state before transitioning to an alternative delinquency state. If the loan is then predicted to move back to state 0, then 12-month expected credit losses will be calculated instead. This might impose a significant reduction in provisions, depending on the loan's characteristics and maturity date. However, this might not be the case under a stressed scenario. Understanding the probable path this account will experience under different economic environments helps forecast impairment losses at any future point.

4. Conclusions

The literature studying the applications of survival models and intensity models to credit risk portfolios has increased rapidly in recent years. At the same time, stress testing analysis is a vitally important activity for banks to protect depositors and the wider public from losses in the event of severe but plausible adverse economic events. Banks need to provide stress testing results to regulators on a regular basis and carry out stress tests to compute economic capital. This study presented a new approach to forecast delinquency states and transition probabilities among alternative macroeconomic scenarios based on a dynamic multi-state model, providing a suitable framework under IFRS9.

The contribution of this research is threefold. First, we estimated a delinquency multi-state model for residential mortgages. Previous studies have focused on similar specifications for credit cards (Djeundje & Crook, 2018; Leow & Crook, 2014), but ours is the first study to consider six plausible types of transitions for a mortgage lending portfolio. (Kelly and O'Malley (2016) based their study on movements into and out of default only.) Our estimation process considered both account-specific data (application and behavioural information) and macroeconomic series. The signs obtained for the coefficient estimates were aligned with expectations and economic intuition.

Second, the estimation and forecast of transition probabilities accounted for recurrent events. We considered three types of variance-corrected approaches: the Andersen–Gill model and the PWP-CP and PWP-GT models

Table 6
Macroeconomic data sources.

Variable name	Source
Consumer Confidence Index	OECD
Consumer Price Index (CPI)	ONS
FTSE-100	Yahoo Finance
House Price Index (HPI)	HM Land Registry
Industrial Production Index	OECD
Unemployment Rate	ONS
5-year Mortgage Rate	Bank of England

Table 7
Stability of coefficient estimates, transition intensity 0 → 1.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	1.519***			1.495***			1.480***			1.494***		
LTV at Orig.	0.016***	0.016***	0.015***	0.014***	0.014***	0.014***	0.013***	0.013***	0.012***	0.011***	0.010***	0.010***
Repayment[a]	−0.218***	−0.206***	−0.210***	−0.166***	−0.146***	−0.156***	−0.108***	−0.087**	−0.100***	−0.058	−0.034	−0.064*
Balance at Orig. (Log)	−0.162***	−0.170***	−0.157***	−0.127***	−0.132***	−0.125***	−0.095**	−0.100**	−0.094**	−0.077*	−0.077*	−0.071*
Balance left (%), Lag 3	0.168***	0.172***	0.164***	0.171***	0.174***	0.165***	0.174***	0.177***	0.167***	0.166***	0.169***	0.156***
Current Int. Rate, Lag 3	0.149***	0.146***	0.133***	0.143***	0.133***	0.125***	0.148***	0.130***	0.126***	0.230***	0.207***	0.210***
UK Unemp. Rate (D1)	0.560***	0.537***	0.549***	0.570***	0.497***	0.469***	0.587***	0.482***	0.428***	0.385**	0.290*	0.218

***p < 0.01, **p < 0.05, *p < 0.1.
D1 = first difference.
[a]Repayment type: 1 = Repay, 0 = Interest-only.

Table 8
Stability of coefficient estimates, transition intensity 1 → 2.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	0.643***			0.679***			0.692***			0.672***		
Marital Status = Single[a]	0.413***	0.416***	0.349***	0.364***	0.366***	0.314***	0.247***	0.245***	0.205***	0.231***	0.217***	0.186***
Balance left (%), Lag 3	0.210**	0.202*	0.175*	0.173	0.173	0.152	0.181	0.176	0.147	0.334***	0.316***	0.200**
Current LTV, Lag 3	0.003*	0.003*	0.005***	0.003*	0.003*	0.005***	0.003**	0.003**	0.005***	0.002	0.002	0.003***
Int. Rate, Lag 3	0.090**	0.088*		0.056	0.053		0.032	0.028		0.0004	−0.004	
Unemp. Rate (YoY)	0.954***	0.961***	0.558**	0.939***	0.935***	0.605**	0.961***	0.951***	0.470**	1.044***	1.039***	0.391*

***p < 0.01, **p < 0.05, *p < 0.1.
YoY = year-over-year growth rate.
[a]Marital Status: 1 = Single, 0 = Non-single.

Table 9
Stability of coefficient estimates, transition intensity 2 → 3.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
One Applicant[a]	0.297*	0.297*	0.236**	0.308**	0.308**	0.243**	0.222*	0.222*	0.177**	0.293**	0.293**	0.207**
Balance left (%), Lag 3	0.249***	0.249***	0.187***	0.248***	0.248***	0.181***	0.225***	0.225***	0.178***	0.233***	0.233***	0.146***
Int. Rate, Lag 3	0.311***	0.311***	0.256***	0.216***	0.216***	0.191***	0.253***	0.253***	0.231***	0.151**	0.151**	0.154***
Unemp. Rate (YoY)	1.027**	1.027**		0.962**	0.962**		1.229***	1.229***		1.334***	1.334***	

***p < 0.01, **p < 0.05, *p < 0.1.
YoY = year-over-year growth rate.
[a]Number of applicants: 1 = One applicant, 0 = More than one applicant.

with event-specific baseline hazards. To the best of our knowledge, only Djeundje and Crook (2018) controlled by repeated events using a multi-state frailty model for credit cards. By implementing the AG model, we found that the number of times an account has experienced the event is significant, while the estimation of the PWP models explicitly showed the impact on stratified risks. We also presented how recurrent events have a direct impact on the level of the cumulative intensities.

Third, this study adds to the existing literature on stress testing in that scenario-conditional transition probabilities based on dynamic multi-state models were not analysed before. The methodology we proposed allowed us to forecast delinquency states and transition probabilities consistent with changes in the underlying economic conditions and considering alternative macroeconomic scenarios. By forecasting each component of the intensity processes, we were able to predict transition probabilities for any future and unobserved time, i.e. beyond the lag of the time-varying covariates. We conclude that stress scenarios have a larger impact on younger vintages than on older vintages, suggesting that more recent borrowers are more likely to become delinquent under severe economic conditions. Moreover, the relative impact of the scenario also differs by origination characteristics, such as the number of applicants, the repayment type, or the occupancy type. We also found that both the level of the transition probabilities and their spread between stress and baseline scenarios change depending on the selected horizon $(t, t + \tau]$.

We believe there is space for more research to be explored. This methodology captures the impact of recurrent events either by considering the number of previous events as a covariate or through stratified baseline hazards. It would be interesting to analyse the impact that recurrence has on the coefficient estimates of other covariates, such as loan characteristics and macroeconomic data. On the other hand, we briefly mentioned the implications the selection of the development sample has on coefficient estimates, and we showed (in the Appendix) this impact for different samples. However, when stability of parameters is not observed, it would be interesting to analyse the effect on the level of transition probabilities and jumps towards alternative delinquency states. Although model re-calibrations and model re-estimations are expected in practice, understanding how the estimated impact of covariates changes across time would bring a better comprehension of the dynamics of these models. Finally, the forecast horizon for the prediction of scenario-conditional transition intensities is restricted to the length of the cumulative baseline hazards estimated in-sample. Therefore, a model that extends these baseline hazards beyond the last observation would be needed for longer forecasting windows.

Table 10
Stability of coefficient estimates, transition intensity 1 → 0.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	−0.123***			−0.141***			−0.145***			−0.163***		
Marital Status = Single[a]	−0.202***	−0.202***	−0.139***	−0.165***	−0.166***	−0.117***	−0.136**	−0.134**	−0.086**	−0.113*	−0.115*	−0.093**
BTL[b]			0.185***			0.196***			0.201***			0.137***
Balance at Orig. (Log)	0.123**	0.128**	0.144***	0.083	0.083	0.108***	0.091*	0.094*	0.108***	0.128**	0.129**	0.125***
LTV, Lag 3	−0.008***	−0.008***	−0.008***	−0.007***	−0.007***	−0.007***	−0.006***	−0.007***	−0.006***	−0.005***	−0.005***	−0.006***
IPI (YoY), Lag 3	1.286**	1.232*		1.658***	1.553**		1.286**	1.178*		−0.078	−0.172	
Unemp. Rate (D1)			−0.490***			−0.263			−0.049			0.211

***p < 0.01, **p < 0.05, *p < 0.1.
D1 = first difference; YoY = year-over-year growth rate.
[a]Marital Status: 1 = Single, 0 = Non-single.
[b]BTL: 1 = Buy-to-let, 0 = Owner-occupied.

Table 11
Stability of coefficient estimates, transition intensity 2 → 1.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
One Applicant[a]			−0.210*			−0.216*			−0.180*			−0.326***
BTL[b]	0.343*	0.415**		0.242	0.281*		0.278*	0.300*		−0.179	−0.174	
LTV, Lag 1	−0.007**	−0.007**		−0.004	−0.004		−0.001	−0.001		0.001	0.001	
LTV, Lag 3			−0.007**			−0.004**			−0.002			−0.002
Mortgage Rate (D1)	−0.905*	−0.938*		−1.023**	−1.113**		−1.062**	−1.081**		−0.986	−0.937	

***p < 0.01, **p < 0.05, *p < 0.1.
D1 = first difference.
[a]Number of applicants: 1 = One applicant, 0 = More than one applicant.
[b]BTL: 1 = Buy-to-let, 0 = Owner-occupied.

Table 12
Stability of coefficient estimates, transition intensity 2 → 0.

Covariate	Up to 2011			Up to 2012			Up to 2013			Random 70%		
	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT	AG	PWP-CP	PWP-GT
Strata	0.638*			0.538*			0.409*			−0.107		
Marital Status = Single[a]			−0.350**			−0.337**			−0.176			−0.066
Interest Rate at Orig.			−0.302***			−0.220**			−0.132*			−0.057
LTV, Lag 3			−0.007**			−0.006**			−0.006**			−0.006**
Property Value, Lag 3	0.135**	0.133**		0.085	0.090		0.108**	0.105**		0.023	0.027	
Int. Rate, Lag 3	−0.289***	−0.255*		−0.309***	−0.296***		−0.244***	−0.232***		−0.183***	−0.186***	
IPI (D1)	0.225**	0.228**	0.253**	0.177**	0.182**	0.186***	0.141*	0.152*	0.147**	0.174**	0.168**	0.190**

***p < 0.01, **p < 0.05, *p < 0.1.
D1 = first difference
[a]Marital Status: 1 = Single, 0 = Non-single.

Table 13
Exhaustive regression analysis for the Industrial Production Index (YoY).

Variable	Transformation	Lag	Expected sign
Real GDP	YoY	0, 1, 3, 6	Positive
Nominal GDP	YoY	0, 1, 3, 6	Positive
Disposable income	YoY	0, 1, 3, 6	Positive
Unemployment rate	none, YoY, 1st diff, 3rd diff, 6th diff, 12th diff	0, 1, 3, 6	Negative
FTSE-100	YoY	0, 1, 3, 6	Positive
HPI	YoY	0, 1, 3, 6	Positive

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A.1. Sources of macroeconomic data

Table 6 lists the sources of macroeconomic data used to estimate the transition intensities presented in Section 2.3.

Table 14
Exhaustive regression analysis for the five-year mortgage rate.

Variable	Unit	Lag	Expected sign
Monetary Policy Rate	Level	0, 1, 3, 6	Positive
3-month Yield	Level	0, 1, 3, 6	Positive
5-year Yield	Level	0, 1, 3, 6	Positive
10-year Yield	Level	0, 1, 3, 6	Positive

A.2. Training and test samples: Stability of coefficient estimates

The estimation of the transition intensities presented in Section 2.3 is constrained by the data availability. Given

Table 15
Industrial production index (YoY). Regression analysis.

Rank order	Explanatory variables	Coefficient estimates	*Adj.R*^{2}	RMSE	*Adj.R*^{2}/ *RMSE*
1	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (6th Diff) ; HPI (YoY) - Lag 6	0.13*** ; −0.01** ; 0.21***	0.8293	0.0342	24.265
2	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (3rd Diff) ; HPI (YoY) - Lag 6	0.13*** ; −0.02** ; 0.22***	0.8279	0.0342	24.243
3	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (3rd Diff) - Lag 1 ; HPI (YoY) - Lag 6	0.13*** ; −0.02** ; 0.22***	0.8275	0.0341	24.237
4	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (1st Diff) ; HPI (YoY) - Lag 6	0.13*** ; −0.04** ; 0.23***	0.8274	0.0341	24.235
5	FTSE-100 (YoY) - Lag 1 ; HPI (YoY) - Lag 6	0.14*** ; 0.25***	0.8196	0.0338	24.234
6	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (1st Diff) - Lag 1 ; HPI (YoY) - Lag 6	0.13*** ; −0.04** ; 0.23***	0.8269	0.0341	24.227
7	FTSE-100 (YoY) - Lag 1 ; Unemp. rate (1st Diff) - Lag 3 ; HPI (YoY) - Lag 6	0.13*** ; −0.04** ; 0.22***	0.8261	0.0341	24.215
8	FTSE-100 (YoY) ; HPI (YoY) - Lag 6	0.13*** ; 0.29***	0.8167	0.0338	24.188
9	FTSE-100 (YoY) - Lag 1 ; HPI (YoY) - Lag 3	0.09*** ; 0.3***	0.8124	0.0337	24.119
10	FTSE-100 (YoY) - Lag 3 ; Unemp. rate (3rd Diff) - Lag 1 ; HPI (YoY) - Lag 6	0.14*** ; −0.03*** ; 0.12**	0.8135	0.0339	24.012

** $p < 0.05$, *** $p < 0.01$. Newey–West HAC standard errors. Number of observations: 120.

Table 16
Five-year mortgage rate. Regression analysis.

Rank order	Explanatory variables	Coefficient estimates	*Adj.R*^{2}	RMSE	*Adj.R*^{2}/ *RMSE*
1	3-month Yield - Lag 1	0.497***	0.9858	0.1288	7.6550
2	Monetary Policy Rate	0.500***	0.9814	0.1474	6.6563
3	Monetary Policy Rate - Lag 1	0.501***	0.9803	0.1516	6.4651
4	3-month Yield	0.495***	0.9781	0.1597	6.1250
5	3-month Yield - Lag 3	0.495***	0.9688	0.1908	5.0786
6	Monetary Policy Rate - Lag 3	0.495***	0.9524	0.2356	4.0425
7	3-month Yield - Lag 6	0.479***	0.8974	0.3459	2.5941
8	5-year Yield - Lag 1	0.669***	0.8961	0.3480	2.5750
9	5-year Yield	0.668***	0.8930	0.3533	2.5279
10	5-year Yield - Lag 3	0.665***	0.8837	0.3683	2.3995

** $p < 0.05$, *** $p < 0.01$. Newey–West HAC standard errors. Number of observations: 144.

Table 17
Tests on residuals.

Estimated variable	Dickey–Fuller	Phillips–Perron	Shapiro–Wilk	Breusch–Pagan & Cook–Weisberg	Breusch–Godfrey
Ind. Prod. Index (YoY)	−5.642***	−5.629***	−0.254	1.42	39.683***
5-year Mort. Rate	−1.102	−0.740	2.897***	170.45***	126.486***

*** $p < 0.01$, ** $p < 0.05$. Newey-HAC standard errors were implemented to deal with the presence of autocorrelation and/or heteroskedasticity in the residuals.

that one of the objectives of the methodology presented in this paper is to predict transition probabilities between delinquency states, it is necessary to test the estimation outputs on an independent sample. The training sample was defined as all information observed from June 2006 to December 2011, while the test sample was defined as all accounts originated from January 2012 and observed until December 2015. Note that any cut-off date could have been selected without impacting the methodology, as the contribution is to develop an approach that enables us to estimate and predict transition probabilities between delinquency states and between any two points in time and under alternative macroeconomic scenarios for a mortgage portfolio.

To maximise the number of observations, the samples were determined such that each of them would roughly contain 50% of the accounts. This definition may have an impact on coefficient estimates because the training sample is observed during the financial crisis, and tested on a post-crisis period. Note that, in practice, re-calibrations or re-estimations of the transition intensities are expected once more data become available, and this might have an impact of the coefficient estimates. To investigate this further, we performed extra analyses by re-defining the training sample as 1) data observed until 2012, 2) data observed until 2013, and 3) a random sample containing 70% of the accounts observed at any point. Tables 7 to 12 compare the results obtained for each transition intensity. Although the parameters for each sample are in general quite stable, those estimated from the 70% random sample present the largest differences compared to those estimated for the period up to 2011, especially for movements out of states 1 and 2. This is expected, as the composition of the sample significantly differs from the development sample used in this research. Except for the transition 1 → 0, we observe that the estimates for the macroeconomic factors remain significant when extending the development sample to account for data observed up to 2012 and up to 2013. In terms of account-level data, most of the coefficient estimates also remain significant; however, we observed instability for some of them (e.g. Marital Status in transition 1 → 0, BTL in transition 2 → 1, and Strata in transition 2 → 0).

A.3. Estimated paths for industrial production index and five-year mortgage rate by scenario

We ran an exhaustive regression analysis on the year-on-year growth of the industrial production index and on the five-year mortgage rate considering a set of macroeconomic variables provided by the PRA as potential covariates. The aim was to produce scenario-conditional paths

(a) Industrial production index (YoY).

(b) Five-year mortgage rate.

Fig. 13. Observed values vs. estimated values.

Table 18
Estimated industrial production index (YoY). VIF.

Explanatory Variable	VIF
FTSE-100 (YoY) - Lag 1	1.91
Unemp. Rate (6th Diff.)	1.75
HPI (YoY) - Lag 6	1.33
Mean VIF	1.66

for these two variables that were consistent with the PRA's scenarios. These potential covariates were obtained from the 2016 stress testing scenarios and are listed in Tables 13 and 14.[14]

This exhaustive regression analysis implied analysing the combination of all potential covariates using up to three explanatory variables and considering both contemporaneous and lagged effects. Once all these regressions were obtained, we kept those with significant coefficient estimates at the 5% level that also showed the expected sign based on economic intuition. To avoid multicollinearity, combinations of covariates with a correlation above $|0.75|$ were disregarded. As several models satisfied these constraints, we then selected the specification with the highest $\frac{Adjusted\ R^2}{RMSE}$ ratio, as it represents the model with best in-sample fit.

Tables 15 and 16 present the top ten models based on in-sample fit. The specifications ranked first (highlighted in grey) were selected.

A.3.1. Regression analysis

See Tables 15 and 16.

A.3.2. In-sample fit

See Fig. 13.

A.3.3. Residual analysis

See Table 17.

A.3.4. Multicollinearity analysis

See Table 18.

References

Aalen, O., Borgan, Ø., & Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, *5*(3), 141–150.

Amorim, L. D., & Cai, J. (2015). Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*, *44*(1), 324–333.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*. Springer-Verlag New York, Inc..

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 1100–1120.

Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, *50*(12), 1185–1190.

Bellotti, T., & Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, *60*(12), 1699–1707.

Bellotti, T., & Crook, J. (2013a). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, *29*(4), 563–574.

Bellotti, T., & Crook, J. (2013b). Retail credit stress testing using a discrete hazard model with macroeconomic factors. *Journal of the Operational Research Society*, *65*(3), 340–350.

Beran, J., & Djaïdja, A.-Y. K. (2007). Credit risk modeling based on survival analysis with immunes. *Statistical Methodology*, *4*(3), 251–276.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists*. Cambridge University Press.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 89–99.

Calabrese, R., & Crook, J. (2020). Spatial contagion in mortgage defaults: A spatial dynamic survival model with time and space varying coefficients. *European Journal of Operational Research*, *287*(2), 749–761.

Cope, D., Hsu, C., Lively, C., Morgan, J., Schuermann, T., & Sekeris, E. (2022). Stress testing for commercial, investment, and custody banks. *Handbook of Financial Stress Testing*, 247.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, *34*(2), 187–220.

[14] Note that these series were not used to estimate the transition intensities, as we restricted the estimations presented in Section 2.3 to macroeconomic data available on a monthly frequency to avoid bias on the coefficient estimates driven by the usage of quarterly data.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data. 1984*. London: Chapman&Hall.
Crook, J., & Banasik, J. (2012). Forecasting and explaining aggregate consumer credit delinquency behaviour. *International Journal of Forecasting*, *28*(1), 145–160.
Dirick, L., Claeskens, G., & Baesens, B. (2017). Time to default in credit scoring using survival analysis: a benchmark study. *Journal of the Operational Research Society*, *68*(6), 652–665.
Djeundje, V. B., & Crook, J. (2018). Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards. *European Journal of Operational Research*, *271*(2), 697–709.
Djeundje, V. B., & Crook, J. (2019). Dynamic survival models with varying coefficients for credit risks. *European Journal of Operational Research*, *275*(1), 319–333.
Duffie, D. (2011). *Measuring corporate default risk*. Oxford University Press.
Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, *83*(3), 635–665.
Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, *72*(359), 557–565.
Guo, Z., Gill, T. M., & Allore, H. G. (2008). Modeling repeated time-to-event health conditions with discontinuous risk intervals: an example of a longitudinal study of functional disability among older persons. *Methods of Information in Medicine*, *47*(2), 107.
Hott, C. (2015). A model of mortgage losses and its applications for macroprudential instruments. *Journal of Financial Stability*, *16*, 183–194.
Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley & Sons.
Kau, J. B., Keenan, D. C., & Kim, T. (1994). Default probabilities for mortgages. *Journal of Urban Economics*, *35*(3), 278–296.
Kelly, P. J., & Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, *19*(1), 13–33.
Kelly, R., & O'Malley, T. (2016). The good, the bad and the impaired: A credit risk model of the Irish mortgage market. *Journal of Financial Stability*, *22*, 1–9.
Kiefer, N. M., & Larson, C. E. (2015). Counting processes for retail default modeling. *Journal of Credit Risk*, *11*(3), 45–72.
Kupiec, P. H. (2018). On the accuracy of alternative approaches for calibrating bank stress test models. *Journal of Financial Stability*, *38*, 132–146.
Lando, D., & Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, *26*(2), 423–444.
Lee, E. W., Wei, L., Amato, D. A., & Leurgans, S. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival analysis: State of the art* (pp. 237–247). Springer.
Leow, M., & Crook, J. (2014). Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, *236*(2), 685–694.
Leow, M., & Crook, J. (2016). The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. *European Journal of Operational Research*, *249*(2), 457–464.
Malik, M., & Thomas, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, *28*(1), 261–272.
McDonald, R. A., Matuszyk, A., & Thomas, L. C. (2010). Application of survival analysis to cash flow modelling for mortgage products. *Operational Research Society*, *23*(1), 1–14.
Ozenne, B., Sørensen, A. L., Scheike, T., Torp-Pedersen, C., & Gerds, T. A. (2017). Riskregression: predicting the risk of an event using Cox regression models. *The R Journal*, *9*(2), 440–460.
Park, K. A. (2016). FHA loan performance and adverse selection in mortgage insurance. *Journal of Housing Economics*, *34*, 82–97.
Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, *68*(2), 373–379.
Sagara, I., Giorgi, R., Doumbo, O. K., Piarroux, R., & Gaudart, J. (2014). Modelling recurrent events: comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. *Malaria Journal*, *13*(1), 293.
Schechtman, R. (2013). Default matrices: A complete measurement of banks' consumer credit delinquency. *Journal of Financial Stability*, *9*(4), 460–474.
Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting*, *30*(3), 717–728.
Schuermann, T., Jafry, Y., et al. (2003). Measurement and estimation of credit migration matrices. *Center for Financial Institutions Working Papers*, *3*(8).
Thenmozhi, M., Jeyaseelan, V., Jeyaseelan, L., Isaac, R., & Vedantam, R. (2019). Survival analysis in longitudinal studies for recurrent events: Applications and challenges. *Clinical Epidemiology and Global Health*, *7*(2), 253–260.
Therneau, T. M., & Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine*, *16*(18), 2029–2047.
Ullah, S., Gabbett, T. J., & Finch, C. F. (2014). Statistical modelling for recurrent events: an application to sports injuries. *British Journal of Sports Medicine*, *48*(17), 1287–1293.
Wei, L., & Glidden, D. V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine*, *16*(8), 833–839.
Wei, L.-J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, *84*(408), 1065–1073.
Wood, R., & Powell, D. (2017). Addressing probationary period within a competing risks survival model for retail mortgage loss given default. *Journal of Credit Risk*, *13*(3).
Zhang, J., & Thomas, L. C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, *28*(1), 204–215.