

Modelling Defaultable Loan Repayment Cashflows Using Neural Hawkes Processes

JUGRAJ S. DEOL

*Department of Mathematics, Brunel University of London, Kingston Lane, Uxbridge, UB8 3PH,
United Kingdom*

AND

PARESH DATE*

*Department of Mathematics, Brunel University of London, Kingston Lane, Uxbridge, UB8 3PH,
United Kingdom*

*Corresponding author: Paresh.Date@brunel.ac.uk

[]

We propose a novel framework for modelling defaultable loan repayment cashflows using Hawkes processes, where missed repayments are captured via a self-exciting point process and transition dynamics are learned through a recurrent neural network. Using a loan-level dataset of over 1,400 mortgages from 2014 to 2024, we demonstrate the model's ability to accurately classify and simulate repayment behaviour across performing, missed-payment, bankruptcy, and restructuring states. The model achieves 94.8% accuracy in forecasting repayment states and passes the Kupiec Value-at-Risk test at 90%, 95%, and 99% confidence levels with consistently high pass rates across borrower credit categories, indicating strong calibration and value for risk assessment.

Keywords: Finance; Forecasting; Risk analysis; Neural Hawkes process; Simulation

© The Author(s) 2026. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Accurately modelling and forecasting loan repayment behaviour remains one of the key challenges in quantitative finance and risk management. For institutions active in underwriting and management of mortgage-backed securities (MBS), anticipating repayment irregularities—missed instalments, loan modifications, or prolonged delinquency—is vital. Such forecasts affect pricing of collateralised securities, capital reserve allocations and cash flow planning. Traditional credit risk models concentrate on deriving the probability of terminal events such as default or prepayment. While useful, these approaches typically rely on restrictive assumptions of linear modelling, time invariance and temporal independence. In practice these assumptions are poorly suited to the empirical realities of individual loan trajectories which occur across prolonged time periods, exhibit clustering and reflect both endogenous borrower histories and exogenous macroeconomic shocks. Standard credit risk models therefore risk missing important features within loan level repayment data.

This paper addresses the following significant question:

How can we model and forecast mortgage ¹ repayments at the individual loan level to inform risk-sensitive decision-making in MBS management?

To address this, we propose a Neural Hawkes Process (NHP) framework that captures the dynamic, non-linear, and temporally dependent nature of mortgage repayments. Unlike classical parametric models that impose fixed functional forms, the NHP learns a data-driven conditional intensity function to govern state transitions. The model is parametrised by a recurrent neural network (RNN) that evolves in continuous time, allowing it to capture both short-term memory patterns (e.g., recent missed payments) and long-term dependencies (e.g., latent financial distress).

Our empirical study uses a granular dataset of over 1,400 residential mortgages observed between 2014 and 2024. The motivation for using a data-driven self-exciting process arises from several empirical observations in a granular dataset spanning a ten-year post credit crisis period (2014–2024). We observe:

- Repayment behaviour displays strong temporal clustering (e.g., consecutive missed payments or sustained current performance);
- Credit categories show clear heterogeneity, with repayment dynamics differing across borrower segments;
- Macroeconomic shocks do not trigger isolated outcomes but instead propagate through sequential waves of delinquency, restructuring, and recovery.

These features violate assumptions of Markovian dynamics and linearity, suggesting the need for a self-exciting point process framework that adjusts its forecasts due to evolving loan-level histories.

Our proposed Neural Hawkes approach offers several operational benefits.

- It forecasts entire repayment paths rather than single transitions;

¹ As our empirical study is based on a mortgage-backed securities, we will use the terms mortgage and loan interchangeably. We emphasise that the primary objective of this study is to model repayment dynamics and the associated cashflow trajectories, rather than to construct a complete valuation framework. We are interested in categorising missed payments and modelling the associated cashflow (as is relevant in MBS modelling). Consistent with standard reduced-form credit risk modelling, collateral and recovery effects can be incorporated downstream, conditional on bankruptcy or other default-like states. As the emphasis is on cashflow modelling and simulation, the fact that a mortgage has a collateral and hence a non-zero recovery rate is not explicitly modelled within the present framework and may be considered as a possible direction of future research.

- It captures non-Markovian and non-linear interactions between events;
- It yields probabilistic forecasts over future state paths, enabling robust simulation, value-at-risk estimation, and cash flow stress testing.

Beyond its methodological novelty, this framework provides direct utility for practitioners in mortgage servicing, securitization, and risk management. For example, servicers can use trajectory forecasts to pre-emptively offer loan modifications; underwriters can use model-generated risk trajectories to adjust expected loss estimates; and investors in MBS can assess tail risks under alternative macroeconomic scenarios. We emphasise that while the primary focus of this study is on modelling repayment dynamics and cashflow trajectories, the resulting simulated paths can be directly incorporated into downstream valuation frameworks, which we outline in Section 8.

Summary of Key Contributions.

- We develop an exponential-kernel Hawkes process as a benchmark model for mortgage repayments, providing a parametric self-exciting framework that captures clustering behaviour in repayment trajectories.
- We enhance this framework by relaxing the parametric structure and employing an RNN to model transition intensities, yielding a Neural Hawkes Process.
- We use a high-resolution, post-2014 mortgage dataset with granular monthly payment histories and borrower features.
- We propose a state space reduction framework that simplifies 11 repayment states into 4 interpretable classes while preserving the underlying dynamic structure.
- We provide empirical evidence of non-Markovian repayment dynamics, including dwell times, transition asymmetries, and behavioural clustering.
- We demonstrate the operational value of this approach in MBS underwriting, enabling scenario-based trajectory simulation and borrower-level credit risk assessment.

2. Related Work

The earliest studies of mortgage credit risk relied on classical statistical tools, in particular logistic regression and hazard rate models. These models linked observable borrower and loan features, such as credit scores, loan-to-value ratios and interest rates, in order to calculate the conditional probability of terminal events like default or prepayment (von Furstenberg and Green, 1974). Their widespread adoption at the time was largely due to the ease of implementation and being computationally light. However, the assumptions underpinning these models are restrictive, with models often assuming linear effects, additive covariates and memoryless transitions. In practice, repayment histories rarely look so tidy, instead exhibiting dynamic and non-linear repayment patterns involving regular payments and delinquency cycles. Standard exponential hazards, often used within Cox models, fail to account for periods of sustained delinquency followed by re-performance or for borrower histories with alternating periods of irregular and regular payments.

A parallel approach was the development of option-theoretic models, originally designed for valuing mortgage-backed securities (MBS). These models interpret prepayment and other borrower choices as rational responses to changes in interest rates and house prices (Kau et al., 1994). While the approach provides valuable theoretical insight, it quickly becomes analytically complex when multiple dimensions are considered. Moreover, it relies on assumptions of rationality and market completeness that seldom hold in practice. Both structural and reduced-form models also share a limitation: treating

repayments as isolated outcomes rather than sequences with temporal interdependence, limiting their capability to model evolving borrower trajectories.

More recently, structural credit risk models have been extended to account for finite maturity and endogenous default boundaries. For example, [Detemple and Kitapbayev \(2025\)](#) develop a valuation framework for equity and debt with finite-maturity obligations, formulating default boundaries via coupled Volterra integral equations. Their work highlights that the maturity structure of debt plays a crucial role in determining both firm value and default premia. While such approaches are analytically powerful, they remain tied to assumptions about firm-value dynamics. By contrast, our framework adopts a data-driven perspective, allowing default and repayment behaviour to emerge endogenously from observed loan-level histories.

Recent advances in machine learning have introduced more flexible methodologies for credit risk modelling by relaxing the rigid parametric forms used by classical models, with particular applications of deep learning used to move beyond rigid parametric functional forms. For instance, [Sadhvani et al. \(2021\)](#) used feed-forward neural networks to incorporate lagged repayment information, improving predictions of transitions between repayment, prepayment, and default. Building on this idea, [Manzo and Qiao \(2020\)](#) combined structural models with neural representations of firm balance sheet data and macroeconomic conditions. Other studies have looked at ensemble tree-based models ([Kündig and Sigrist, 2024](#)), spatio-temporal Gaussian processes, and transformer-based architectures ([Lei et al., 2024](#)). These methods highlight the potential of modern machine learning to capture heterogeneity and behavioural nuance. That said, most are geared toward static classification tasks or near-term forecasts. We found no models, to the best of our knowledge, which are designed to generate full repayment paths, including repeated delinquencies or restructuring, which are essential for long-horizon risk analysis.

Parallel strands of research have emphasised liquidity-sensitive cash management policies. [Salas-Molina et al. \(2020\)](#), for instance, propose a multi-criteria optimisation framework for cash management that incorporates liquidity delays, forecast uncertainty, and transaction costs. Their results underscore the importance of explicitly modelling liquidity terms when evaluating financial decision-making under uncertainty. Our work is complementary: while their focus is on liquidity-driven optimisation at the firm or portfolio level, we model the stochastic structure of loan-level cashflows, which can in turn provide inputs into such liquidity-aware optimisation frameworks.

A promising alternative to above approaches is offered by temporal point processes, which are designed to model sequences. In this view, the chance that a new event arrives at time t depends on the previous sequence of events up until time t through a conditional intensity function $\lambda(t)$. To fix ideas, consider the conditional intensity in the standard Hawkes process ([Hawkes, 1971](#)):

$$\lambda(t) = \mu + \sum_{t_i < t} \phi(t - t_i),$$

where $\mu > 0$ is a baseline rate and $\phi(\cdot)$ is a kernel that captures the excitation of past events. A common kernel choice uses an exponential decay,

$$\phi(t) = \alpha e^{-\beta t}, \quad \alpha, \beta > 0.$$

The exponential kernel yields self-excitation behaviour where the arrival of one event briefly lifts the likelihood of another arriving soon after, reinforcing the intuition of events occurring in clusters. The clustering feature has clear uses within finance, where Hawkes models have been applied to trade arrivals ([Bacry et al., 2015](#)), order-book dynamics ([Lu and Abergel, 2018](#)), and the transmission of systemic shocks ([Aït-Sahalia et al., 2015](#)).

Despite their strengths, classical Hawkes processes are limited by their reliance on fixed kernel forms and linear superposition, which cannot capture non-linear interactions or heterogeneous responses. [Mei and Eisner \(2017\)](#) address this by replacing the hand-crafted kernel with a recurrent neural network, giving the Neural Hawkes Process:

$$\lambda(t) = f(h(t)),$$

where $h(t)$ is a hidden state updated at event times evolving between each arrival, and $f(\cdot)$ is a non-negative mapping such as a softplus. This allows the model to learn from data how memory should decay, revive, or interact—rather than imposing a fixed functional form. This makes it possible to capture longer-range dependencies and more intricate excitation patterns while preserving the point-process interpretation.

Neural Hawkes models have surfaced in a wide range of settings. In healthcare, for example, they have been used to trace hospital admission histories by linking short-lived symptoms with longer-term conditions ([Mei and Eisner, 2017](#)). Related ideas have been used to describe activity on online platforms ([Du et al., 2016](#)) and to follow how rumours move through social networks ([Guo et al., 2018](#)). Finance has its own variants: work on market contagion ([Rambaldi et al., 2018](#)) and physics-informed versions for cryptocurrency markets ([Fabre and Muni Toke, 2025](#)).

In contrast, mortgage repayments present unique modelling challenges due to the quasi-periodic structure of repayments (such as missed payments, late repayments, arrears and changes in the interest rate), combined with complex regulatory interventions (e.g., loan modifications), and the interaction between micro-level borrower decisions and macroeconomic shocks. This blend of persistence of clustering (e.g. consecutive missed payments) and dormancy (e.g., prolonged current status) volatility across credit classes is hard to capture with standard hazard models and Hawkes processes that rely on a single exponential kernel.

Our work addresses the gap within literature by developing a history-sensitive NHP tailored to the state transitions in mortgage repayment. The aim is not only to predict the next status change, but to generate full repayment paths. This capability supports forecasting, risk quantification, and stress testing in the MBS context. We keep the point-process backbone for interpretability while allowing a recurrent network to learn how memory decays—or occasionally revives—over time. By integrating point process theory with deep sequence modelling, our framework provides a novel tool for credit risk assessment.

The benchmark models considered in this study—a first-order Markov chain and an exponential-kernel Hawkes process—are intentionally parsimonious. Their role is to isolate the incremental value of incorporating duration dependence, behavioural clustering, macro-level modifiers, and non-linear hidden-state dynamics within the Neural Hawkes framework. To the best of our knowledge, enriched parametric alternatives with comparable temporal structure have not yet been developed in the literature on stochastic loan cashflow modelling. In that sense, the exponential Hawkes specification serves not only as a benchmark, but also as a modest stand-alone contribution demonstrating the viability of self-exciting point-process methods for mortgage cashflow forecasting.

Fixed structure parametric models such as exponential Hawkes process offer far better interpretability than semi-parametric models such as the NHP, whereas NHP offers far more flexibility in capturing non-linear and history-dependent interactions. This trade-off, and especially the impact of using a data-driven model structure to capture the temporal dynamics for borrowers with low creditworthiness, is made clearer in section 7.

3. Methodology

3.1. Classical Hawkes Process

The classical Hawkes process is a self-exciting temporal point process which captures the probability of an event occurring through an intensity function, $\lambda(t)$, conditioned on past event arrivals. For event times $\{t_i\}_{i=1}^N$ on $[0, T]$, the univariate Hawkes process defines a conditional intensity function $\lambda(t)$, representing the instantaneous event rate given the history $\mathcal{H}_t = \{t_i < t\}$, as:

$$\lambda(t | \mathcal{H}_t) = \mu + \sum_{t_i < t} \phi(t - t_i), \quad (3.1)$$

with $\mu > 0$ representing the background (exogenous) rate and $\phi(\cdot)$ is the memory kernel capturing endogenous influences from past arrivals. A common parametric form is the exponential kernel:

$$\phi(s) = \alpha e^{-\beta s}, \quad \alpha, \beta > 0, \quad (3.2)$$

where each past event briefly increases the conditional intensity function, $\lambda(t)$, before decaying as time elapses. The process requires $\alpha/\beta < 1$ to remain stable.

In the multivariate setting, the model allows for multiple event types (e.g., different repayment states). Let K denote the number of event types, the intensity for type k becomes

$$\lambda_k(t) = \mu_k + \sum_{j=1}^K \sum_{t_i^j < t} \phi_{kj}(t - t_i^j), \quad (3.3)$$

where $\phi_{kj}(\cdot)$ captures the influence of type j events on type k intensities. In mortgage modeling, such formulations can describe transitions like missed payments triggering future delinquency or restructuring.

However, classical Hawkes models are limited by:

- fixed parametric kernels that impose linear decay and lack flexibility;
- additive structure that cannot model inhibition or interactions;
- Markovian assumptions that overlook long-range dependencies.

These limitations motivate a more expressive framework: the Neural Hawkes Process.

3.2. Neural Hawkes Process Formulation

The Neural Hawkes Process (NHP) extends the classical Hawkes framework by introducing non-linear, history-dependent intensity functions parameterized via recurrent neural networks (RNNs) (Mei and Eisner, 2017). Here, and throughout the sequel, the repayment states are drawn from a discrete state space $\mathcal{S}$, whose reduced four-state representation is defined formally in Section 4.2. Let $\{(t_i, x_i)\}_{i=1}^N$ denote a sequence of repayment events, with $x_i \in \mathcal{S}$ representing the discrete repayment state and t_i the corresponding time of occurrence. The conditional intensity for transition type k is parameterized as:

$$\lambda_k(t) = f_k(h(t)) = \text{Softplus}(w_k^\top h(t) + b_k), \quad (3.4)$$

where $h(t) \in \mathbb{R}^d$ is the continuous-time hidden state encoding the loans repayment history, w_k and b_k are transition-specific parameters, while the Softplus activation, $\text{Softplus}(z) = \log(1 + e^z)$, guarantees non-negativity.

The hidden state $h(t)$ should be interpreted as a compressed representation of the borrowers repayment history and temporal dependencies. It aggregates information about recent missed payments, duration in the current state, earlier episodes of delinquency or restructuring, and the timing between successive events. In this sense, $h(t)$ acts as the mechanism through which the model carries forward high-order dependence beyond a first-order Markov structure.

Between repayment events, the hidden state evolves according to a continuous-time decay–drift process:

$$h(t) = h_i \odot \exp(-\delta \cdot (t - t_i)) + \tilde{h} \odot (1 - \exp(-\delta \cdot (t - t_i))), \quad (3.5)$$

where h_i is the hidden state immediately after the i th event, $\delta \in \mathbb{R}_+^d$ is a learnable vector of decay rates, $\tilde{h}$ is a learned long-term memory attractor, and $\odot$ denotes element-wise multiplication.

At the event time t_i , the RNN updates its hidden representation via a Long Short-Term Memory (LSTM) unit:

$$(h_i, c_i) = \text{LSTM}(h(t_i^-), c(t_i^-), x_i, \Delta t_i), \quad (3.6)$$

where $h(t_i^-)$ and $c(t_i^-)$ are the hidden cell states immediately before the i th event, x_i is a one-hot encoding of the observed repayment state, and $\Delta t_i = t_i - t_{i-1}$ is the elapsed time since the previous event. For notational simplicity, subsequent formulations focus on the hidden state h_i , while the cell state c_i is retained internally by the LSTM to regulate information flow.

Although equations (3.4)–(3.5) contain parametric elements through the weights, biases, and decay terms, the model is flexible because the hidden-state dynamics are governed by the LSTM rather than by a fixed parametric structure. This means that the Neural Hawkes Process is substantially less restrictive than classical Hawkes models with a prescribed exponential decay structure.

This architecture provides three key modelling advantages:

- It captures non-linear excitation and inhibition effects in repayment behaviour;
- It accommodates high-order temporal dependencies through the LSTMs gating and memory mechanism;
- It generates context-sensitive repayment trajectories across heterogeneous borrower risk profiles.

The multivariate output across K repayment states yields event-wise probabilities and intensities that evolve continuously in time, thereby offering a flexible and data-driven extension of the traditional Hawkes process.

3.3. Model Assumptions and Design

To make the framework usable for mortgage repayment data, we adopt a few pragmatic choices that mirror how these datasets are recorded and analysed.

Discrete monthly observations Repayment information arrives on a monthly basis. The NHP is specified in continuous time, but for estimation we treat t_i as month-indexed event times. Between observations, the latent state and implied intensity evolve continuously; we interpolate within the month so that timing effects and decay are not lost simply because reporting is monthly.

State transition dynamics Each event x_i records a change of repayment state. The component intensity $\lambda_k(t)$ is interpreted as the instantaneous likelihood of moving into state k , conditional on the past. Borrower-specific memory enters through the hidden state $h(t)$, while shared or macro-level influences are captured by weights w_k that modulate the transition to k .

Temporal dependence Repayment behaviour is history-dependent. Rather than imposing a first-order Markov restriction, the NHP carries information forward through $h(t)$, allowing the model to reflect patterns such as strings of missed payments, subsequent re-performance, and occasional switching between regimes over extended horizons.

Non-stationarity Behaviour changes over time. We therefore allow decay and intensity terms to vary, learned directly from the data, so the model can pick up calendar effects or regime shifts—for example, seasonal risk upticks or policy-driven modification waves—without hard-coding time-invariant coefficients.

3.4. Training Procedure and Likelihood Optimisation

The training process is executed using individual loan repayment sequences. The full set of trainable parameters, denoted as Θ includes the following components: $\Theta = \{w_k, b_k, \delta, \text{LSTM parameters}\}$. Here w_k and b_k are the class-specific weights and biases which map the hidden state to the intensity for state k ; δ denotes an additional model hyperparameter (e.g., a decay or modulation term); and the remaining terms are the recurrent weights, biases, and gates of the LSTM layers.

We fit the model by maximising the standard point-process log-likelihood for an observed sequence $\{(t_i, x_i)\}_{i=1}^N$:

$$L(\Theta) = \sum_{i=1}^N \log \lambda_{x_i}(t_i) - \int_0^T \sum_{k=1}^K \lambda_k(t) dt.$$

The first part rewards placing high intensity on the events that actually occurred; the second is the compensator, penalizing excess predicted events to ensure valid probability mass.

Because $\lambda_k(t)$ is a non-linear function of the hidden state, the integral has no convenient closed form solution. We therefore approximate it on a discrete grid using a simple midpoint (or, equivalently here, uniform Riemann) rule:

$$\int_0^T \lambda_k(t) dt \approx \sum_{j=1}^J \lambda_k(t_j) \Delta t,$$

with t_j spaced on the monthly reporting grid and Δt the step size. This choice matches the data frequency, offers ease for implementation and yields stable gradients for stochastic optimisation.

3.4.1. Model Architecture and Complexity

The model uses a 2-layer LSTM architecture, where each loan has a sequence of N monthly observations, hidden width d , and $K = 4$ repayment states (C, M, B, R) (please refer to equation 6), the per-epoch computational cost scales on the order of $O(N \cdot d \cdot K)$. This reflects both the forward evaluation and the reverse-mode (backpropagation) pass. Hidden states are rolled one month at a time, so updates occur sequentially across the calendar.

In practice, mini-batching across loans amortises the cost over many sequences, and the memory footprint grows linearly with N and d . On our dataset, routine training runs converge within a few hundred epochs—typically under 400—indicating that the architecture is efficient enough for granular repayment histories while remaining computational feasible to optimise.

3.4.2. Model Calibration and Evaluation

The model is calibrated and evaluated using two out-of-sample metrics to assess its performance and suitability for operational credit risk:

- **Exact Match Score for Simulated Sequences:** The model's ability to simulate repayment sequences is evaluated by comparing simulated paths to observed repayment states. Multiple sequences are generated per loan, and the score measures the proportion of exact matches. The higher the average exact match score, the better the accuracy.
- **Value-at-Risk (VaR) Estimation:** VaR is estimated by simulating repayment sequences and computing the 1st, 5th, and 10th percentiles of the resulting loss distribution. A rolling window approach is applied, with VaR recalculated at each time step until loan maturity. Calibration is assessed using Kupiec's test, which evaluates whether observed violation frequencies align with predicted coverage.

These calibration metrics demonstrate that the model produces reliable, interpretable, and statistically consistent forecasts, supporting its suitability for real-world credit risk applications.

4. Data Description and Preprocessing

4.1. Dataset

Our empirical work uses a monthly, loan-level data panel sourced from Bloomberg. The collateral pool corresponds to the mortgage-backed security *AAM 2011-1 AG12* and covers origination cohorts issued between 2014 and 2024. Each loan is followed from first disbursement until one of four endpoints – full prepayment, charge-off, restructuring or the study cut-off in March 2024. In total, the raw panel contains over 1,400 loans and more than 120,000 monthly repayments observations across borrowers with varying credit characteristics.

For loan l , we observe the sequence

$$\{(t_i^l, x_i^l, \mathbf{z}_i^l)\}_{i=1}^{N_l},$$

where

- t_i^l is the i th observation (month) for loan l ;
- $x_i^l \in \mathcal{S}_{\text{raw}}$ is the recorded repayment status at t_i^l ;
- $\mathbf{z}_i^l \in \mathbb{R}^p$ is the associated vector of borrower and loan-level covariates.

The repayment status x_i^l is encoded using an internal 11-class schema reflecting granular outcomes such as full payment, partial payment, missed, modified, bought-out, and delinquent states.

4.2. State Space Reduction

In practice, repayment labels within loan-level datasets are granular and often noisy. To enhance learnability and improve the statistical reliability of transition density estimation under sparse state-switching regimes, we introduce a reduced state space defined as

$$S = \{C, M, B, R\},$$

where the elements correspond to the dominant repayment regimes of practical and theoretical interest:

- **C (Current):** full, on-time payment;
- **M (Missed):** missed, delinquent, or partial payment;
- **B (Bankruptcy):** default, foreclosure, or charge-off;
- **R (Restructured):** modification, deferment, or workout aimed at avoiding default.

The mapping function

$$\psi : S_{\text{raw}} \rightarrow S$$

is applied to all repayment histories prior to sequence construction, where S_{raw} denotes the original 11-state coding. Table 1 lists the correspondence, consolidating idiosyncratic or short-lived administrative labels into coherent behavioural classes.

Original State	Reduced Class
C — Current	C
3 — 30-day Delinquent	M
6 — 60-day Delinquent	M
9 — 90-day Delinquent	M
B — Bankruptcy	B
F — Foreclosure	B
M — Modified	R
L — Liquidated	R
S — Subsequent Loss / Recovery	R
R — REO	R
P — Paid off	Dropped

TABLE 1 Mapping from the original 11-state system (S_{raw}) to the reduced four-state representation (S).

This reduction serves two aims. First, it preserves the main behavioural modes while smoothing volatile or administrative label changes (e.g., short transitions between delinquency buckets). Second, it yields a tractable and interpretable state space for intensity-based modelling—particularly for the Neural Hawkes Process, where sparse transitions and long-range dependence are central—thereby supporting long-horizon forecasting and scenario-driven risk simulation at the portfolio level.

4.3. Feature Construction and Summary

We construct a set of borrower loan-level features $\mathbf{z}_i^l$ capturing creditworthiness, contract terms, and payment obligations. Table 2 summarizes all the model features used as RNN inputs.

TABLE 2 Summary of Features Used in the Neural Hawkes Model

Feature	Description
Credit Category	Categorical: Good, Fair, Poor
Original Balance	Initial loan amount
Gross Coupon	Annualised interest rate
Payment Due	Monthly scheduled payment amount
Current State	One-hot vector: C, M, R, B
Time Since Start	Number of months since loan origination

All features are standardised within the training set and remain fixed at each time step. The current repayment state is encoded using a one-hot vector and passed into the RNN at each update. Time intervals between events are critical for computing decay in the hidden state dynamics.

4.4. Preprocessing and Filtering

We apply a simple, transparent pipeline to clean and structure the repayment histories before modelling:

- (a) **Censoring.** Loans that finish before March 2024 (e.g., full repayment or buyout) are right-censored at their last observed month, preserving the observed history without extrapolation.
- (b) **Exclusion threshold.** Loans with fewer than 24 months of observations are removed. Such short histories tend to destabilise training and provide insufficient context for the models memory. In our data, these series typically conclude with state P (full payoff), often driven by factors outside the recorded repayment path.
- (c) **Invalid sequences.** We remove sequences with consecutive missing entries or raw labels that cannot be mapped to the reduced state space. Fewer than 0.2% of loans are excluded.
- (d) **Event construction.** For each loan, a discrete-time event sequence $\{(t_i, x_i)\}_{i=1}^N$ is constructed, where $x_i \in \mathcal{S}$ denotes the repayment state at month t_i . Only state transitions or payment status changes are retained to reduce redundancy.
- (e) **Train/test split.** The dataset is split into an 80/20 calibration/validation partition by loan, ensuring no data leakage. All preprocessing is applied prior to splitting.

We stratify by borrower credit quality using FICO scores to form three groups: *Good*, *Fair*, and *Poor*. From each group, 150 loans are randomly selected, yielding a total sample of 450 loans from the full dataset of 1,441 loans (see Section 6.1). This ensures a balanced design across credit categories while respecting data availability: after filtering, the smallest segment contains fewer than 200 loans, constraining the maximum feasible balanced sample size.

This balanced sampling design ensures comparable representation across credit categories while retaining a sufficient number of observations within each segment to support statistically meaningful estimation and out-of-sample comparison. Within each stratum, data is split using the 80/20 (120 training, 30 validation) training/test ratio. We then fit a separate Neural Hawkes Process on the training loans in each stratum, yielding group-specific parameters. These fitted models are evaluated on their corresponding validation sets, allowing performance comparisons across credit segments.

5. Empirical Patterns and Motivation

Before formalising our model, we first document empirical patterns in mortgage repayment behaviour that demonstrate the limitations of conventional approaches. The evidence highlights loan level repayments exhibit behaviours which are history-dependent, uneven over time, and distinctly non-linear—features that standard modelling techniques struggle to capture. We found both borrower-level persistence and system-wide variation have a meaningful impact for forecasting.

5.1. Transition Asymmetries and Non-Markovian Effects

Classical approaches—Markov transition models or hazard-type specifications—typically assume that next months state depends only on the current state and contemporaneous covariates. However, empirical evidence from our dataset reveals strong deviations from these assumptions of first order Markov dynamics with time homogenous transition probabilities.

Figure 1 reports the empirical transition matrix over the loan life cycle for the $K = 11$ raw states $\mathcal{S}_{\text{raw}} = \{C, 3, 6, 9, B, F, M, L, S, P, R\}$. The visualisation reveals several patterns:

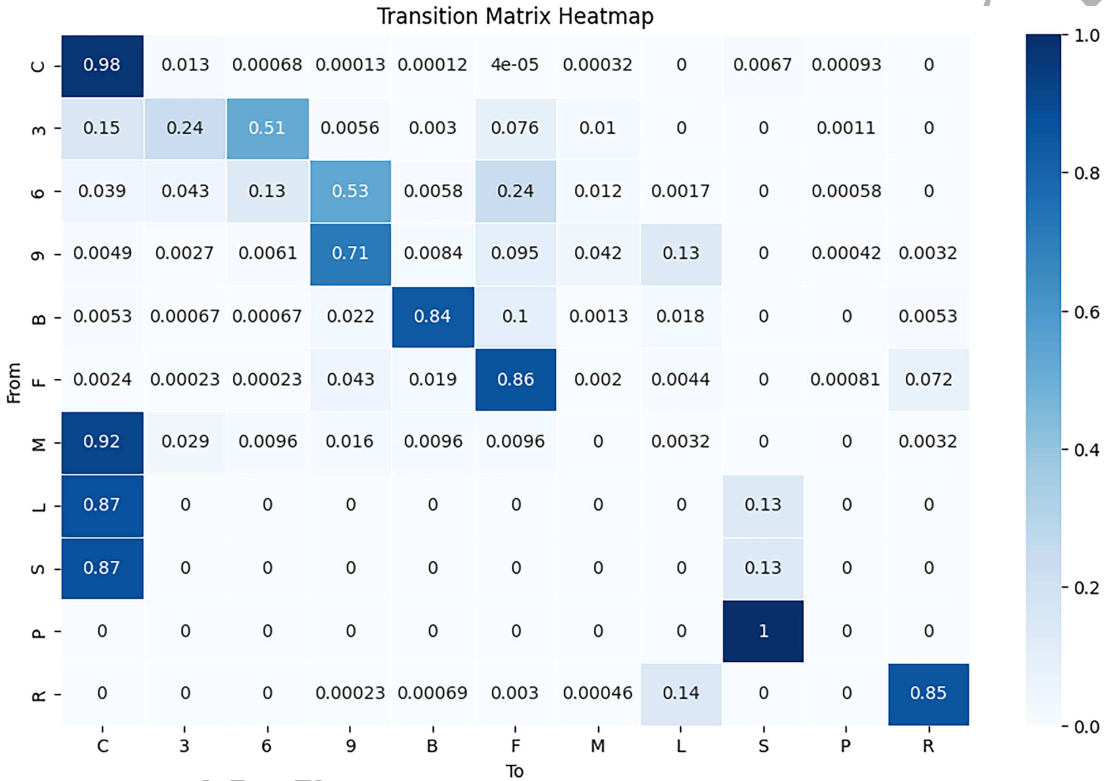


FIG. 1. Empirical repayment transition probabilities (monthly, 2014–2024). (Single-column fitting image).

- **Strong persistence.** The largest transition probabilities lie on the main diagonal: loans tend to remain in their current state for prolonged periods. This persistence is observed across all repayment states.
- **Sequential progression.** A secondary band of elevated transition probabilities immediately to the right of the diagonal suggests a common deterioration pathway: $C \rightarrow 3 \rightarrow 6 \rightarrow 9$, followed by resolution or exit states (B, F, M, L, S, P, R).
- **Acceleration in mid-delinquency.** From state 6, transitions to 9 are comparatively likely, indicating that once a loan reaches mid-stage delinquency, reversion to a better state becomes unlikely before further deterioration.

- **Near-absorbing endpoints.** States such as B (bankruptcy) and P (paid off) behave almost as absorbing, with transitions back into delinquency being rare without intervention or restructuring.
- **Cycles after resolution.** Returns from restructuring to current ($R \rightarrow C$) are observed, with re-performance more likely when time spent in R is short, suggesting a window in which remediation is most effective.

Taken together—high within-state persistence, staged progression toward distress, acceleration at intermediate delinquency, quasi-absorbing endpoints, and post-resolution cycling—the evidence conflicts to the use of a time-homogeneous, first-order Markov model. A better suited modelling approach would be one that is conditioned upon the full history of a loan repayments and allows the intensity to adapt over time. The Neural Hawkes framework, which is explicitly history-sensitive, is designed for precisely this kind of temporal dependence.

5.2. Behavioural Cluster Persistence and Post-Cluster Transition Dynamics

Our analysis reveals that mortgage repayment trajectories exhibit long-form regimes, such as sustained current repayments, persistent delinquency or extended periods in restructuring, contradicting the assumption of purely memoryless dynamics in most traditional credit risk models. To define this persistent dynamic, we call a *state cluster* any run of two or more consecutive months in the same state. This lets us measure typical durations and see how often prolonged regimes occur. Table 3 reports the mean cluster length for the four states within our reduced state space: Current (C), Missed (M), Bankruptcy (B), and Restructured (R).

State	Average Length
C	41.23
M	4.65
B	9.44
R	7.59

TABLE 3 Mean length (months) of contiguous state clusters.

The results highlight that, C is highly persistent: once the borrowers establish a pattern of on-time payment, they tend to keep it. Secondly, M clusters are shorter, whereas B and R last longer on average—consistent with the drawn-out nature of distress and the time needed for restructuring to take effect.

The empirical distributions of cluster lengths are markedly right-skewed, with short-duration clusters occurring far more frequently than long ones, resulting in mean values that understate the maximum observed persistence. This skew is a reminder that models should allow for widely varying durations rather than impose a single “typical” persistence.

Not all state occurrences form clusters. We define a *non-clustering arrival* as an isolated month in a given state with different states on both sides. These are uncommon, accounting for only 2.12% of all state observations. The prevalence of clustering is exactly the kind of pattern a self-exciting specification is designed to accommodate. We then ask what happens *after* a cluster ends. Figure 2 summarises the most frequent next state for clusters of C , M , B , and R .

Our findings show that the transitions are asymmetric. After C , the next state is most often M . Clusters of B and R frequently lead to each other, reflecting movement between bankruptcy-related

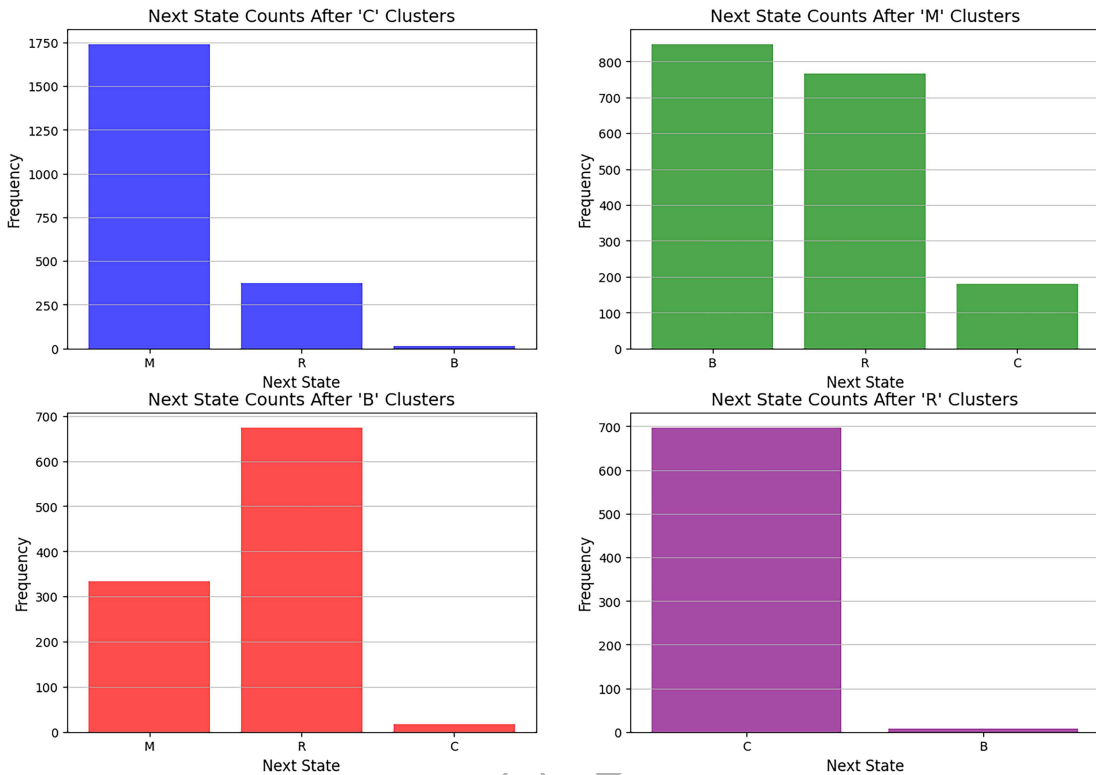


FIG. 2. Most frequent next state after a cluster terminates. (Single-column fitting image).

outcomes and restructuring. By contrast, *R* commonly returns to *C*, consistent with the idea that successful restructuring restores regular payment. The *M* state displays a bifurcated pattern, with roughly balanced transitions to *B* and *R*, suggesting that missed payments can lead either to default or to lender intervention.

We further test the hypothesis that cluster length influences the subsequent transition probabilities. The evidence indicates longer spells in prolonged delinquency (*M*) are associated with a higher chance of bankruptcy (*B*) next and a lower chance of returning to *C*. Extended time in *C* eventually raises the probability of slipping to *M*, which fits with life-cycle effects in long mortgages (income shocks, interest rate resets, and so on). Long *B* clusters tend to be followed by *M*, signalling ongoing instability, whereas *R* clusters revert to *C* at high rates almost regardless of length, suggesting restructuring is relatively effective once in place.

These findings confirm that repayment behaviour exhibits both strong temporal persistence and path dependent post-cluster dynamics, necessitating modelling inter-temporal approaches — such as the NHP — that can incorporate full repayment histories and learn complex inter-temporal dependencies beyond the reach of conventional Markovian models.

5.3. Systemic Behavioural Variation over Time

The temporal dynamics of repayment behaviour reveal pronounced variation over the sample period, closely linked to macroeconomic conditions and policy interventions. Figure 3 presents the monthly arrival frequencies of each repayment state (*C*, *M*, *B*, and *R*) from 2014 to 2024. Several systemic patterns emerge:

- **Sustained growth in *C* arrivals:** The frequency of loans in the Current (*C*) state shows an overall increasing trend, stabilising at high levels from 2018 onwards, indicating sustained repayment stability.
- **Decline in *M* arrivals:** Missed payments (*M*) are more frequent in the early years (2014–2015) and then recede, consistent with improved borrower capacity or tighter screening after 2016.
- **Early peaks in *B* and *R* arrivals:** Bankruptcy (*B*) and Restructured/Remortgage (*R*) exhibit elevated activity in 2014–2016 before tapering, suggesting that defaults and restructuring waves were concentrated early in the sample.
- **Long-run suppression of distress states:** From 2018 onwards, both *M* and *B* frequencies remain persistently low while *C* dominates, reflecting a systemic shift towards repayment stability.

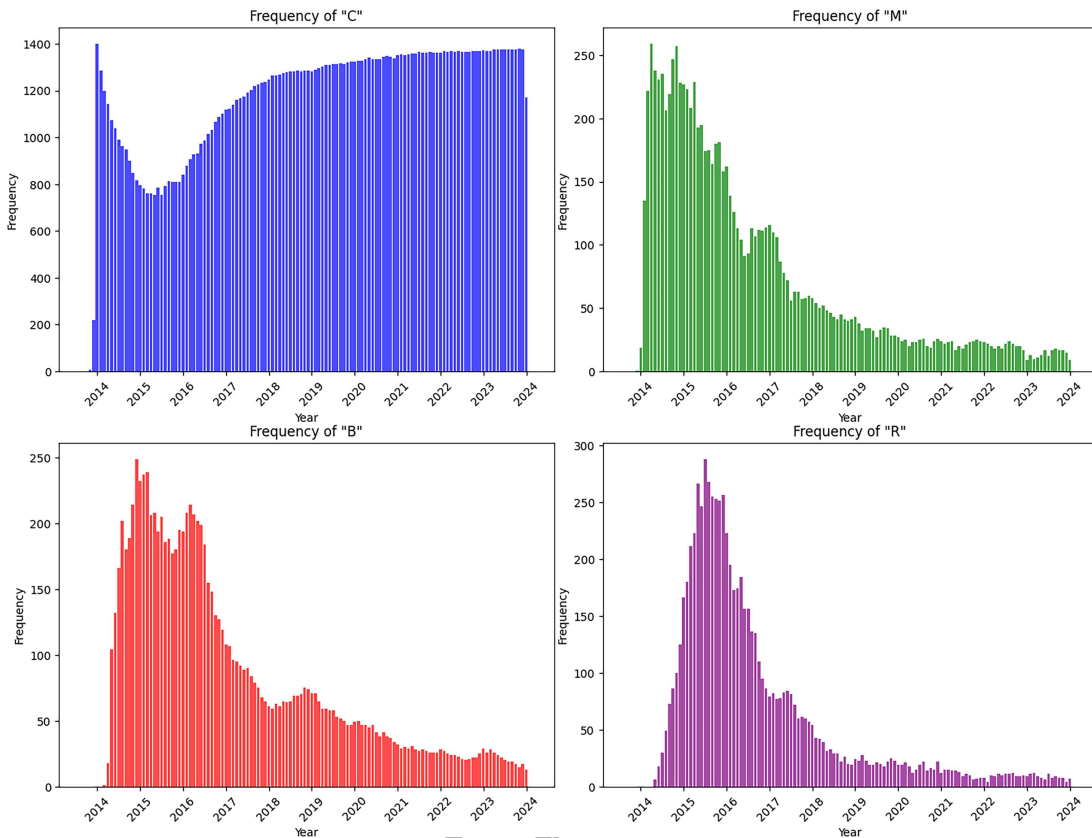


FIG. 3. Monthly arrivals by repayment state (C, M, B, R), 2014–2024. (Single-column fitting image).

These findings highlight that repayment state dynamics are not stationary over time. Periods with elevated M , B , and R give way to years dominated by C . Models that treat transitions as stationary in time—or that rely on a single cross-section—will fail to model these regime changes. A time-dependent specification is therefore warranted, one that allows intensity levels to adjust with macro conditions, borrower composition, and policy interventions as they evolve.

5.4. Implication for Model Design

The empirical evidence above motivates three key design criteria:

- (i) The model must incorporate full trajectory history and not assume Markovian sufficiency.
- (ii) Temporal patterns such as clustering, decay, and recurrence must be embedded within the modelling structure.
- (iii) The model must remain flexible to accommodate system-wide shifts in borrower behaviour induced by exogenous shocks.

These criteria motivate the use of a Neural Hawkes Process framework, which incorporates long-memory encoding, time decay, non-linear intensities, and dynamic simulation—all required for robust credit risk modelling under real-world repayment dynamics.

5.5. Incorporating Systemic and Behavioural Modifiers

The empirical analyses presented above points to two dominant features of mortgage repayment behaviour: (i) repayment states tend to persist in clusters (behavioural inertia), and (ii) the mix of states shifts over calendar time with the macro environment. To ensure that these dynamics are explicitly embedded within the Neural Hawkes specification, we add two auxiliary inputs to the intensity: a *systemic temporal modifier* $f(t)$ and a *behavioural decay modifier* $g(t)$ — which serve as additional exogenous inputs to the intensity process.

Systemic temporal modifier $f(t)$ To ensure adaptedness, $f(t)$ is constructed using lagged information only. In particular, for each calendar month t , we define $f(t)$ from the portfolio-wide distribution of the four reduced states observed up to time $t - 1$, so that $f(t)$ is measurable with respect to the filtration $\mathcal{F}_{t-1}$. Operationally, this means the model only uses information that would have been available at the time the forecast is formed, thereby avoiding look-ahead bias.

Let $f(t) \in \mathbb{R}^4$ record, for each calendar month t , the portfolio share in the four reduced states (C, M, B, R) . Each loans history is aligned to the same calendar index, so $f(t)$ provides a time-varying signal of system-wide credit conditions, including shifts associated with economic downturns, institutional interventions, or policy driven refinancing waves. To ensure interpretability, the state distributions are normalised relative to their long-run mean (yielding a centred representation). Per loan n , the resulting input $f_n(t) \in \mathbb{R}^{4 \times T}$ is fed to the neural network alongside the event history, allowing intensities to move with calendar-time regimes rather than relying solely on individual trajectories.

Behavioural decay modifier $g(t)$ The behavioural modifier $g(t)$ is constructed independently of the training and validation samples. Specifically, it is estimated using an external dataset with highly similar repayment and borrower characteristics, which serves as a proxy for duration-dependent transition behaviour. In our empirical implementation, this external dataset is constructed from a set of comparable mortgage-backed securities ABMT 2014-1, AP 2014-3, AABST 2014-1 B1, and AAM 2014-3 AG12 which exhibit similar borrower profiles and repayment dynamics, thereby ensuring a representative and economically meaningful estimation of duration-dependent transition behaviour. This separation preserves statistical validity and ensures that $g(t)$ does not leak future information from the evaluation sample into the forecasting mechanism. The availability of such proxy datasets—for example, comparable mortgage-backed securities—is a realistic assumption in developed and liquid markets such as the UK and US. The function $g(t) \in \mathbb{R}^{4 \times T}$ encodes borrower-specific persistence effects by quantifying how the probability of state transitions varies with the duration of time spent in a given state. Concretely, for each $s \in \{C, M, B, R\}$ we estimate, from the data, the distribution of next-state outcomes conditional on the current *cluster length* (the number of consecutive months in s). We then normalise by the states unconditional average to obtain a multiplicative scaling: values above one indicate heightened propensity relative to typical, while values below one indicate suppression. For example, extended delinquency clusters tend to carry lower recovery probabilities—an inertia effect which scaling makes explicit. Supplying $g(t)$ to the model lets the intensity adjust with “time-in-state” rather than treating all months in s as equivalent.

Together, $f(t)$ and $g(t)$ give the network two simple levers: one for *when* the system displays macro-economic changes (calendar-time regimes), and one for *how long* a borrower has been where they are

(duration dependence). In effect, the NHP can respond to macro cycles and borrower persistence at the same time, matching the patterns we observe without imposing stationarity or memoryless transitions.

6. Experimental Setup and Implementation

This section outlines the design of the empirical study, including the Neural Hawkes Process (NHP) model implementation, input feature construction, training/test split methodology, computational setup, and baseline comparisons. The goal is to assess how effectively the proposed model can learn repayment dynamics and forecast future borrower behaviour in a realistic mortgage underwriting environment.

6.1. Feature Construction and Target Definition

The source panel contains 1,441 anonymised residential mortgages. For the experiments, we draw a simple random sample of 450 loans originated between 2014 and 2024. Each loan is represented as a monthly sequence of repayment states

$$x_i \in \mathcal{S} = \{C, M, R, B\},$$

together with borrower- and loan-level covariates.

Each month, we augment the observation with the following:

- **Static features.** Original balance, interest rate, gross coupon, FICO-equivalent credit score (300–850), and loan term. These do not change over time but help anchor cross-loan differences.
- **Temporal features.** Months since origination; time spent in the current state (duration); the lagged repayment state; and the number of active state clusters at time t , captured by the function $f(t)$.
- **Exogenous macro variables.** A normalised temporal signal $g(t)$ that tracks, across the full set of 1,441 loans, deviations in repayment states from their long-run means, providing a centred indicator of system-wide shifts beyond individual loan histories.

The supervised learning framework uses as its target the sequence of repayment states from x_{i+1} to x_T , conditioned on the historical sequence $\{(x_j, t_j)\}_{j=1}^i$. At each time step t , the model simulates all remaining periods, producing a complete projected repayment trajectory for each loan.

6.2. Train/Test Split and Evaluation Procedure

To mimic real deployment and avoid leakage across sequences, we split *by loan*: 80% for training (360 loans) and 20% for validation (90 loans). No loan appears in both sets. Model training is performed on the training set, with performance monitored on a held-out validation set. Hyperparameters are selected via cross-validation within the training folds. We report two out-of-sample checks on the validation set:

- **Exact-match simulation accuracy.** The proportion of correctly predicted next repayment states.
- **Kupiec test pass rate.** Empirical coverage of simulated Value-at-Risk (VaR) paths relative to the nominal level.

Forecasts extend to the terminal month T using a stochastic simulator driven by the learned conditional intensities. At each step, the model computes the conditional intensity vector $\lambda_k(t)$ for all states $k \in \mathcal{S}$ using the trained LSTM. These intensities are then normalised via a softmax transformation to yield a

categorical distribution over the next state:

$$p_k(t) = \frac{\exp(\lambda_k(t))}{\sum_{k' \in \mathcal{S}} \exp(\lambda_{k'}(t))}, \quad k \in \mathcal{S}.$$

The next state is then drawn from this distribution using multinomial sampling:

$$x_{t+1} \sim \text{Categorical}(p_1(t), p_2(t), \dots, p_{|\mathcal{S}|}(t)),$$

thereby introducing the stochastic nature of the forecast trajectory. This sampling procedure is iteratively applied until the terminal time step T , ensuring that each simulated path reflects both the deterministic influence of the learned intensity dynamics and the inherent uncertainty of repayment behaviour. By repeating this process N times per loan, an ensemble of repayment trajectories is obtained:

$$X_{\text{sim}} = \left\{ (x_t^{(n)})_{t=1}^T \mid n = 1, \dots, N \right\},$$

which facilitates downstream risk estimation, distributional analysis, and scenario-based stress testing. The combination of next-step accuracy and VaR coverage provides both pointwise and risk-sensitive views of out-of-sample performance.

6.3. Model Implementation and Runtime Environment

The Neural Hawkes Process model is implemented in PyTorch version 2.1, using a two-layer long short-term memory (LSTM) recurrent architecture with hidden dimension, $H = 64$, to capture long-range temporal dependencies in repayment sequences. The hidden state $h(t) \in \mathbb{R}^H$ is mapped to an unnormalised intensity vector $\tilde{\lambda}(t) \in \mathbb{R}^{|\mathcal{S}|}$ through a fully connected layer. A Softplus activation ensures positivity:

$$\lambda_k(t) = \log\left(1 + \exp\left(w_k^\top h(t) + b_k\right)\right), \quad \forall k \in \mathcal{S}.$$

During simulation, a Softmax transformation converts the vector of intensities into a categorical distribution from which the next repayment state is sampled.

To reiterate, the empirical contribution of the Neural Hawkes Process is twofold. First, the model incorporates complex temporal features—including duration dependence, behavioural clustering, and macro-level modifiers—directly into the hidden-state dynamics. Second, it avoids the need to specify a restrictive parametric excitation kernel in advance, allowing the data to determine how repayment histories should influence future transition risk.

The model is trained with the Adam optimiser ($\eta = 10^{-3}$) using a fixed number of epochs. Model selection is performed by monitoring the validation log-likelihood at the end of each epoch. The training objective is minimising the negative log-likelihood of the observed repayment sequences under the model. The integral term in the likelihood, which accounts for periods with no observed events, is approximated numerically using the trapezoidal rule on a discretised monthly grid.

Hardware Configuration. Training was conducted on an NVIDIA RTX A6000 GPU (48GB) with 64GB RAM and a 24-core CPU (AMD Ryzen Threadripper PRO 3965WX). Total training time for the final model was approximately 38 minutes, with per-epoch time of 84 seconds. Inference (forecasting till the end of each loan, time step T) required less than 40 seconds.

6.4. Baseline Models for Comparison

To provide a comparative benchmark for the proposed Neural Hawkes Process, we consider two modelling approaches:

1. **Markov Chain Model:** A discrete-time, first-order Markov chain estimated from the training set, in which transition probabilities are conditioned solely on the current repayment state. Formally, the probability of transitioning from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$ is given by:

$$p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i),$$

where $\mathcal{S}$ denotes the set of repayment states. This specification assumes time-homogeneity and history-independence beyond the most recent state, and thus cannot capture higher-order temporal dependencies or clustering effects.

2. **Exponential Hawkes Process Model:** A parametric point process specification in which the conditional intensity for repayment state $k \in \mathcal{S}$ at time t is defined as

$$\lambda_k(t) = f_k(t)\mu_k + \sum_{t_i < t} g_k(t)\alpha_{k,x_i} \exp(-\beta_{k,x_i}(t-t_i)),$$

where $\mu_k > 0$ denotes the baseline intensity scaled by the function $f_k(t)$, $\alpha_{k,x_i} \geq 0$ captures the excitation effect of a past event of type x_i on state k , and $\beta_{k,x_i} > 0$ governs the exponential rate at which such effects decay. The exponential kernel

$$\alpha_{k,x_i} \exp(-\beta_{k,x_i}\tau), \quad \tau > 0,$$

induces a history-dependent but linearly decaying memory structure, in which the influence of past repayment events diminishes at a constant exponential rate. This idiosyncratic element of the Hawkes process is scaled by the function $g_k(t)$ to introduced the affects of cluster length into the model. Parameters are estimated via maximum likelihood from the repayment sequences in the training set. While serving as a benchmark against the Neural Hawkes Process, it is worth noting that, to the best of our knowledge, the application of a parametric Hawkes process to mortgage cashflow modelling has not been reported in the literature and represents a novel contribution in itself.

Each model is trained using identical train/test splits and evaluated using the same forecast accuracy and calibration metrics.

7. Results and Analysis

We assess the performance of the proposed Neural Hawkes Process (NHP) as well as the benchmark models using two evaluation tools: Exact match simulation accuracy and Value-at-Risk forecast evaluated by means of the Kupiec Proportion of Failures (POF) test. The following results demonstrate that the NHP provides a substantial advance in credit risk forecasting for mortgage repayments.

7.1. Exact Match Simulation Accuracy

To evaluate the predictive capability of the proposed model in replicating observed repayment dynamics, we introduce a bespoke performance metric, Exact Match Simulation Accuracy (EMSA). Starting from

time t , we simulate the future states to T and compare each month to the realised state. The score is the share of exact matches across the horizon. Formally, for loan i ,

$$\text{EMSA}_i(t) = \frac{\sum_{\tau=t}^T \mathbf{1}[\hat{x}_{i,\tau} = x_{i,\tau}]}{T - t},$$

where $\hat{x}_{i,\tau}$ and $x_{i,\tau}$ denote the simulated and observed states at time τ , respectively, and $\mathbf{1}[\cdot]$ is the indicator function. By construction, this measure directly quantifies the proportion of repayment states that are predicted exactly over the simulation horizon, thereby providing an interpretable and stringent accuracy benchmark.

Across 90 test loans (spanning *Good*, *Fair*, and *Poor* credit tiers) each with over 60 monthly observations each, the Neural Hawkes Process (NHP) achieves an overall EMSA of 94.80%, averaged across the three forecast horizons. Table 4 presents the average accuracy by horizon and credit class for the NHP, while Tables 5 & 6 report the corresponding results for the benchmark Exponential Hawkes Process (EHP) with a fixed exponential kernel and the first-order Markov chain model. Table 7 reports the overall averages.

TABLE 4 *Exact-Match Simulation Accuracy by Forecast Horizon and Credit Category (Neural Hawkes Process)*

Horizon Category	Good	Fair	Poor
Short (≤ 12 months)	0.9814	0.9736	0.9627
Medium (13–36 months)	0.9954	0.9811	0.9531
Long (≥ 36 months)	0.9418	0.9211	0.8217

TABLE 5 *Exact-Match Simulation Accuracy by Forecast Horizon and Credit Category (Exponential Kernel Hawkes Process)*

Horizon Category	Good	Fair	Poor
Short (≤ 12 months)	0.8867	0.8671	0.8544
Medium (13–36 months)	0.8799	0.8627	0.8151
Long (≥ 36 months)	0.8891	0.8638	0.8112

Two patterns are worth emphasising:

- The Neural Hawkes Process (NHP) exhibits strong predictive performance in the short to medium term, with accuracy exceeding 95% at short and medium horizons across all credit categories. With the exception of long-horizon performance in the *Poor* credit class, accuracy remains above 92%, demonstrating robustness across both time and borrower segments.
- The benchmark models perform consistently worse than the NHP across all horizons and credit categories, with a substantial margin in most cases. Notably, both parametric benchmarks exhibit

TABLE 6 *Exact-Match Simulation Accuracy by Forecast Horizon and Credit Category (Markov Chain Model)*

Horizon Category	Good	Fair	Poor
Short (≤ 12 months)	0.8009	0.7601	0.7354
Medium (13–36 months)	0.7873	0.7699	0.7044
Long (≥ 36 months)	0.7869	0.7544	0.7111

TABLE 7 *Average Forecast Accuracy Comparison on Test Set*

Model	Forecast Accuracy
Neural Hawkes Process	0.9480
Exponential-kernel Hawkes Process	0.8589
Markov Chain (First Order)	0.7567

weaker performance when modelling the *Poor* credit class relative to the *Good* and *Fair* classes. This highlights the limitations of fixed-structure models in capturing the complex and highly stochastic dynamics of higher-risk borrowers.

Looking at averages across all test loans (Table 7), the NHP achieves 94.8% accuracy, compared with 85.9% for the exponential Hawkes model and approximately 75.7% for the Markov chain. Improvements of 8–9 percentage points over the exponential Hawkes and nearly 19 percentage points over the Markov model underscore the importance of flexible memory and non-linear interactions for accurate mortgage forecasting.

7.2. Evaluation of Predictive Calibration via the Kupiec Test

A crucial requirement of a credit risk model is for its probabilistic forecasts to be statistically consistent with real world observations. In the context of mortgage repayments, this means the forecasted chances of default, delinquency, or other irregular repayment should be consistent with realised outcomes. We assess this calibration for the Neural Hawkes Process (NHP) and, for comparison, the exponential-kernel Hawkes model (EHP) and first-order Markov chain model, using the Kupiec Proportion of Failures (POF) test (Kupiec, 1995)—a likelihood ratio method widely applied in Value-at-Risk (VaR) backtesting.

For each loan i , we generate 100 forward repayment simulations beyond a fixed terminal date, in all cases this is the last repayment recorded from the original dataset. From each simulation we compute rolling cumulative losses over horizons of 3, 6, and 12 months. At each horizon, we estimate Value-at-Risk thresholds at confidence levels $\alpha \in \{0.90, 0.95, 0.99\}$ from the simulated loss distribution:

$$\text{VaR}_\alpha(i, k) = \inf\{x \in \mathbb{R} : \mathbb{P}(L(i, k) \leq x) \geq \alpha\}.$$

We then transform the realised history into the same rolling-loss form and flag a violation whenever observed losses exceed the corresponding VaR.

Let x be the number of violations across n evaluation points. With $\hat{p} = x/n$ and $\pi = 1 - \alpha$, the Kupiec unconditional-coverage statistic is given as:

$$LR_{uc} = 2 \left[x \ln \left(\frac{\hat{p}}{\pi} \right) + (n - x) \ln \left(\frac{1 - \hat{p}}{1 - \pi} \right) \right],$$

which is asymptotically χ_1^2 under the null hypothesis of correct unconditional convergence. We treat forecasts as well-calibrated when the null is not rejected at the 5% level.

7.2.1. Results Across Horizons and Confidence Levels

Tables 8, 9 and 10 report the percentage of simulations that pass the Kupiec test, grouped by credit tier and forecast horizon.

TABLE 8 *Kupiec test pass rates (%) for the Neural Hawkes Process (NHP).*

Horizon	Poor			Fair			Good		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
3 Months	0.9152	0.9082	0.8001	0.9743	0.9480	0.9087	0.9796	0.9415	0.9136
6 Months	0.9030	0.8400	0.8097	0.9598	0.9175	0.9077	0.9603	0.9374	0.9174
12 Months	0.8540	0.8348	0.8100	0.9699	0.9390	0.9017	0.9497	0.9245	0.9177

TABLE 9 *Kupiec test pass rates (%) for the Exponential Hawkes Process (EHP).*

Horizon	Poor			Fair			Good		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
3 Months	0.8777	0.8487	0.8348	0.9011	0.8754	0.843	0.9075	0.8842	0.8678
6 Months	0.8673	0.8489	0.8124	0.8703	0.8489	0.8411	0.9020	0.8812	0.8537
12 Months	0.8277	0.8046	0.7911	0.8611	0.8558	0.8311	0.8889	0.8531	0.8122

TABLE 10 *Kupiec test pass rates (%) for the Markov Chain Model.*

Horizon	Poor			Fair			Good		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
3 Months	0.2786	0.2489	0.1876	0.3078	0.2741	0.2107	0.4722	0.4452	0.4310
6 Months	0.2487	0.2047	0.1674	0.2400	0.1984	0.1776	0.4020	0.3501	0.2147
12 Months	0.2044	0.1713	0.1323	0.2149	0.1994	0.1638	0.3911	0.3000	0.1633

Comparative Analysis A comparative examination of the Kupiec test results across the three modelling frameworks Neural Hawkes Process (NHP, Table 8), Exponential Hawkes Process (EHP, Table 9), and the Markov Chain benchmark (Table 10) reveals three key findings.

First, the NHP consistently achieves higher pass rates than both benchmarks across all horizons, VaR levels, and credit tiers (Table 8). Pass rates remain high overall typically above 0.90 for *Fair* and

Good loans while remaining robust, though lower, in the *Poor* credit tier. In comparison, the EHP (Table 9) performs moderately well, whereas the Markov model (Table 10) exhibits substantially weaker calibration, often below 0.50.

Second, while all models show declining performance as the VaR level increases from 90% to 99%, the NHP deteriorates more gradually. For example, in the *Poor* tier at 12 months, the NHP declines from 0.8540 to 0.8100 (Table 8), compared to 0.8277 to 0.7911 for the EHP (Table 9) and 0.2044 to 0.1323 for the Markov model (Table 10), indicating greater robustness in capturing tail risk.

Third, longer forecast horizons adversely affect the benchmark models more strongly. Both the EHP (Table 9) and Markov model (Table 10) exhibit clear deterioration as the horizon extends, particularly for riskier credit tiers. In contrast, the NHP (Table 8) remains relatively stable and retains strong calibration for *Fair* and *Good* loans even at 12 months (around 0.930.94 at 95% VaR). Overall, incorporating temporal dependence and non-linear dynamics leads to more reliable alignment between simulated and realised losses.

7.3. Summary of Findings

Our empirical results show: mortgage repayment paths are not well described by memoryless dynamics. They exhibit clustering, long spells in the same state, and transitions that depend on the full history rather than only the most recent status. Against this backdrop, the Neural Hawkes Process (NHP) performs best across all the chosen measures. In sequence-level accuracy (Exact Match Simulation Accuracy) and in risk calibration (Kupiec tests for VaR), the NHP consistently exceeds the two benchmarks—the exponential-kernel Hawkes model and the discrete-time Markov chain.

In practical terms, the NHP keeps accuracy high at short and medium horizons and gives up relatively little at longer spans, suggesting it captures sustained repayment patterns and borrower-specific persistence. Its Kupiec pass rates are consistently higher across credit tiers and confidence levels relative to the benchmarks, indicating improved alignment with realised loss distributions and a more reliable capture of tail risk, particularly in higher-quality credit segments. The advantage is most visible in the higher-risk segment, where behaviour is more volatile and simple parametric or one-step models are least reliable.

Although the Neural Hawkes Process uses a 2 layer LSTM (and hence a large number of parameters) as compared to more parsimonious benchmark models, its behaviour can still be interpreted in economically meaningful ways. The learned state intensities indicate which repayment transitions are currently most likely, while the hidden-state representation summarises the borrowers temporal repayment profile. In addition, sensitivity of simulated trajectories to input features and repayment histories can be examined through perturbation analysis or gradient-based saliency, thereby linking the model outputs back to observable repayment dynamics. This interpretability is particularly important in model-risk governance settings, where predictive performance must be complemented by an intelligible explanation of the underlying transition mechanism.

8. Valuation Framework, Discounting, and Collateral Effects

While the primary contribution of this study is the modelling of repayment dynamics via a NHP, the resulting simulated repayment trajectories naturally serve as inputs into a broader valuation and risk management framework. In particular, the simulated state paths allow for the construction of loan-level cashflows, which can subsequently be discounted under appropriate risk-adjusted measures.

8.1. From Repayment Dynamics to Cashflow Valuation

Let $\{x_t\}_{t=1}^T$ denote a simulated repayment path for a given loan, where $x_t \in \mathcal{S} = \{C, M, R, B\}$ represents the repayment state at time t . We define the associated cashflow function $CF(x_t)$ as

$$CF(x_t) = \begin{cases} \text{full scheduled payment,} & \text{if } x_t = C, \\ \text{zero or partial payment,} & \text{if } x_t = M, \\ \text{modified payment,} & \text{if } x_t = R, \\ \text{recovery payment,} & \text{if } x_t = B. \end{cases}$$

Given a simulated trajectory, the total discounted value of the loan is given by

$$V = \mathbb{E} \left[\sum_{t=1}^T D(t) CF(x_t) \right],$$

where $D(t)$ is a stochastic discount factor and the expectation is taken over simulated repayment paths generated by the Neural Hawkes Process.

8.2. State-Dependent Discounting and Risk Premia

A key limitation of standard valuation approaches is the use of a single discount rate across all repayment states. In practice, the risk profile of a loan varies significantly depending on its repayment condition. To address this, we introduce state-dependent discount rates of the form:

$$r_k(\tau) = r_f(\tau) + \pi_k(\tau),$$

where $r_f(\tau)$ denotes the risk-free rate at horizon τ , and $\pi_k(\tau)$ represents a state-dependent risk premium associated with repayment state $k \in \mathcal{S}$.

This formulation captures the intuition that delinquent or distressed states (such as M or B) carry higher risk and therefore require higher discount rates, while performing states (C) are discounted at lower spreads.

The discount factor is then given by:

$$D(t) = \exp \left(- \sum_{\tau=1}^t r_{x_\tau}(\tau) \right),$$

which allows the discounting mechanism to adapt dynamically to the realised repayment path.

Furthermore, both the transition dynamics (through the Neural Hawkes Process) and the risk premia $\pi_k(\tau)$ may depend on macroeconomic conditions, ensuring consistency between the simulated repayment behaviour and the valuation framework.

8.3. Incorporating Collateral and Loss Given Default

Collateral plays a central role in mortgage valuation through its impact on recovery rates and loss given default (LGD). While the Neural Hawkes Process models the timing and evolution of repayment states, we treat recovery as a downstream component conditional on the occurrence of a bankruptcy event.

Specifically, upon entering the bankruptcy state B , the cashflow is augmented by a recovery component:

$$CF(B) = \text{Recovery} = (1 - \text{LGD}) \times \text{Exposure at Default}.$$

The LGD may be modelled as a function of collateral characteristics, such as loan-to-value (LTV) ratios, house price dynamics, and liquidation costs. In particular, one may write:

$$\text{LGD} = f(\text{LTV}, \text{House Price Index}, \text{Time to Liquidation}),$$

allowing recovery rates to vary across loans and over time in response to market conditions.

Importantly, this separation between repayment dynamics (modelled by the NHP) and recovery modelling is consistent with standard reduced-form credit risk frameworks, where default timing and loss severity are treated as distinct components.

8.4. *Integration with MBS Pricing and Risk Management*

At the portfolio level, the simulated loan cashflows can be aggregated to produce tranche-level cashflows for mortgage-backed securities (MBS). These cashflows can then be used to compute key risk metrics, including expected loss, Value-at-Risk (VaR), and stress-test scenarios under alternative macroeconomic conditions.

The proposed framework therefore provides a flexible and modular architecture: the Neural Hawkes Process generates realistic, history-dependent repayment trajectories, while the valuation layer incorporates discounting, risk premia, and collateral effects. This separation allows practitioners to adapt the valuation component to specific institutional or regulatory requirements without altering the underlying behavioural model.

9. **Conclusion and Future Work**

This study introduces a Neural Hawkes Process (NHP) framework tailored to the modelling of irregular mortgage repayment trajectories at the individual loan level. By embedding non-linear, history-dependent transition dynamics into a discrete-time temporal point process, the model steps beyond Markov or static specifications and captures features we empirically observe in repayment data—sequential dependence, clustering, behavioural switching, and differences across credit segments.

On a ten-year panel of more than 1,400 residential mortgages, the NHP delivers strong out-of-sample performance. Next-step repayment classification averages 90.90% across test loans, and Value-at-Risk backtests (Kupiec POF) show high pass rates, 98.7% in aggregate in our experiments. These results suggest not only accurate point forecasts but also credible probabilistic calibration. Importantly, the improvements are most visible where behaviour is volatile—loans in weaker credit tiers and horizons beyond a few months—precisely the settings where simple parametric or memoryless models tend to falter.

Since the framework allows full path simulation, it allows lenders and servicers to price dynamically, run scenario-based cash-flow projections, and flag accounts for early intervention when the model anticipates a drift into delinquency or restructuring. For risk and finance teams, the same engine supports IFRS 9/CCAR-style exercises—portfolio aggregation of losses, VaR/CVaR estimation, and stress testing under macro alternatives. Although the model is non-linear, it remains interpretable in practice through state intensities and input saliency, which helps with auditability and model risk governance.

An important next step is to integrate collateral dynamics more directly into the framework. This includes modelling loan-to-value ratios, exogenous or endogenous house-price processes, and recovery rates conditional on default-related repayment states. Such an extension would move the model

from a repayment-dynamics engine toward a more unified valuation framework for mortgage-backed securities.

Future work will pursue three principal directions. First, a multivariate NHP that brings exogenous signals (interest rates, unemployment, house-price indices) into the intensity could capture borrower–macro feedback more directly. Second, allowing covariate-dependent intensities—e.g., functions of loan-to-value, age of loan, or region—would add richer conditioning beyond the event history. Third, for portfolio applications, jointly simulating correlated loans would enable asset–liability analyses and tranche-level attribution for MBS investors and balance-sheet managers.

Collectively, this work establishes a mathematically principled and operationally grounded foundation for discrete-time credit risk modelling, with broad applicability across securitization, capital planning, and risk-sensitive pricing frameworks.

Data Availability Statement

The data that support the findings of this study is available from the corresponding author for non-commercial use. Meta-data provided (including dates for training/validation data, security identification numbers for MBS and the details of RNN architecture used) is adequate to reproduce the results, if a commercial database such as Bloomberg is accessible.

REFERENCES

- Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. A. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015. doi: 10.1016/j.jfineco.2015.03.002.
- E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):1550005, 2015.
- J. Detemple and Y. Kitapbayev. The valuation of corporate securities with finite maturity debt. *IMA Journal of Management Mathematics*, page dpaf028, 2025. doi: 10.1093/imaman/dpaf028.
- N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1555–1564, 2016.
- T. Fabre and I. Muni Toke. Neural hawkes: Non-parametric estimation in high dimension and causality analysis in cryptocurrency markets. *Quantitative Finance*, pages 1–28, 2025.
- H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li. Rumor detection with hierarchical social attention network. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 943–951, 2018.
- A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. doi: 10.1093/biomet/58.1.83.
- J. B. Kau, D. C. Keenan, and T. Kim. Default probabilities for mortgages. *Journal of Urban Economics*, 35(3): 278–296, 1994.
- P. Kündig and F. Sigrist. A spatio-temporal machine learning model for mortgage credit risk: Default probabilities and loan portfolios, 2024.
- P. H. Kupiec. Techniques for verifying the accuracy of risk measurement models. Technical report, Board of Governors of the Federal Reserve System, Washington, DC, 1995.
- Y. Lei, Z. Wang, C. Liu, T. Wang, and D. Lee. Finlangnet: A novel deep learning framework for credit risk prediction using linguistic analogy in financial data, 2024.
- X. Lu and F. Abergel. High-dimensional hawkes processes for limit order books: Modelling, empirical analysis and numerical calibration. *Quantitative Finance*, 18(2):249–264, 2018.

- G. Manzo and X. Qiao. Deep learning credit risk modeling. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn.3651379. Subsequently published in *The Journal of Fixed Income*.
- H. Mei and J. Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- M. Rambaldi, V. Filimonov, and F. Lillo. Detection of intensity bursts using hawkes processes: An application to high-frequency financial data. *Physical Review E*, 97(3):032318, 2018.
- A. Sadhwani, K. Giesecke, and J. Sirignano. Deep learning for mortgage risk. *Journal of Financial Econometrics*, 19(2):313–368, 2021.
- F. Salas-Molina, A. García-Bernabeu, D. Pla-Santamaria, and J. A. Rodríguez-Aguilar. Multiple-criteria cash-management policies with particular liquidity terms. *IMA Journal of Management Mathematics*, 31(2):177–196, 2020.
- G. M. von Furstenberg and R. J. Green. Home mortgage delinquencies: A cohort analysis. *The Journal of Finance*, 29(5):1545–1548, 1974.