The University of Durham

Modelling the Effects of Varying Affordability on the Probability of Default

James Stuart

Submitted for: Mathematics and Statistics Bsc

October 6, 2025

Contents

Aknowledgments

I would like to thank to my supervisor, Prof. Camila Caiado, for their guidance and encouragement throughout the course of this project. Their insights and feedback have been instrumental in shaping the direction of this work. A special acknowledgement is reserved for Atom Bank for the invaluable information and experience they have provided. Finally, I would to express my gratitude to my parents for their continued support throughout my studies.

Plagiarism Declaration

I hereby declare that this report is my own original work and has not been submitted elsewhere in any form. All sources used have been properly acknowledged. No part of this work has been plagiarized from any other source.

Abstract

Following the introduction of the internal ratings based system (IRB) outlined in Basel II, there has been a significant rise in interest for models which can accurately estimate probability of default (PD). This interest stems from PD's central role in assessing credit risk and constructing affordability metrics. This report explores the use of Markov chains in application to estimating PD. We begin by introducing Markov chains in the context of discrete time, before highlighting the limitations of this framework that naturally lead to its extension into continuous-time models. We then review methodologies for assessing the validity of our model's core assumptions; specifically, the properties of time-homogeneity and Markovian behaviour. We perform our data analysis on representative IRB data provided by Atom Bank.

Chapter 1

Introduction

1.0.1 Introduction to Credit Risk

During the early 1970s, there was growing speculation within the financial lending market that the lack of supervision and regulation over institutions was resulting in poor-quality capital held by banks to cover their risk-weighted assets. These concerns were confirmed in 1974 when the international currency markets experienced significant turmoil, with the most notable event being the collapse of Herstatt Bank [3]. This event is widely seen as the key catalyst behind the creation of the Basel Committee. In response to the crisis, the Basel Committee was established in 1975 by a collection of G10 countries, with the goal of developing international banking regulations to ensure that banks were being adequately supervised. Over the next decade, quarterly meetings laid the foundation for a supervisory framework aimed at preventing such crises in the future. However, the focus eventually shifted from supervision to the issue of capital adequacy, leading to the introduction of the Basel I Accord in 1988.

Basel I imposed the requirement for lending institutions to withhold an 8% capital ratio against risk weighted assets whilst maintaining an emphasis on supervision. This was a significant first step in ensuring that banks had a financial cushion to absorb losses in times of crisis, but it was far from perfect. The main limitation was its one-size-fits-all approach to risk assessment, which failed to account for the nuances of market risk (such as currency fluctuations) and operational risk (such as fraud or internal breakdowns). This rigidity left banks exposed to risks that couldn't be captured by a single capital ratio, which led to the creation of Basel II in 2004.

Basel II represented a fundamental shift in how banks were required to assess and maintain capital. The introduction of the Internal Ratings-Based (IRB) approach was a key development, as it allowed banks to use their own internal data to estimate the amount of capital needed to cover risk-weighted assets. This was a far more flexible and accurate approach, as it moved away from the blanket 8% capital ratio of Basel I and allowed banks to consider their specific risk profiles [21]. In essence, Basel II acknowledged that one bank's risks might be vastly different from another's, and thus, capital requirements should reflect these differences.

The IRB approach also introduced a more sophisticated method of risk assessment by incorporating not only credit risk but also market and operational risks. One of the most important factors in the IRB framework is the Probability of Default (PD). We will provide a more rigorous definition in Section 1.0.2 but for now we define this to be the likelihood that a borrower will fail to repay their loan. PD is a critical component when constructing affordability metrics, as the capital needed to cover risk-weighted assets is heavily dependent on expected losses; which, in turn are directly influenced by the PD. In addition to PD, Basel II also incorporated factors like loss given default (LGD) and exposure at default (EAD) to further refine risk assessments. Van Gestel (2008) [26] define LGD to be the expected proportion of exposure which will be lost at the point of a default occurring and EAD to be the outstanding debt pending payment at the time of default. Equation 1.0.1 provides an example calculation of expected loss during the event of a default.

$$\text{Expected Loss given default} = PD \cdot EAD \cdot LGD \tag{1.0.1}$$

The introduction of the IRB approach under Basel II was a crucial step in making banking regulations more risk-sensitive and reflective of the actual conditions banks face. It marked a move from a rigid, one-size-fits-all approach to a more tailored, data-driven system that better reflected the complexity of modern financial institutions. Following the introduction of the IRB approach, there has been a surge demand for advanced statistical models which can accurately estimate expected losses through estimation of PD. A range of techniques have been employed for this task, with the most popular being variations of a logistic regression and, more recently, several banks have applied modern machine learning techniques such as Random Forest. This report will explore a modelling technique which has grown in popularity in recent years due to its straightforwardness and scalability: Markov chains.

1.0.2 Definition of Default

Previously, we classified defaulted accounts as those which are unable to make payments on a loan. While this definition holds true in a general sense, banks must adopt a more rigorous approach to defining default to ensure compliance with the regulations set by the Financial Conduct Authorities (FCA), with particular reference to the IFRS-9 regulations [11]. To further emphasize the importance of a precise definition of default, consider that the Internal Ratings-Based (IRB) approach, which we will be working under, is designed to ensure banks have sufficient capital to cover their risk-weighted assets in events such as default. Thus, the definition of default plays a critical role in determining the capital that Atom must hold to meet these regulatory requirements. It is important to note that this definition may evolve over time, as Atom regularly reviews and updates its policies.

Atom Bank's Default Write-off Policy defines default using the measure of days past due (DPD). If an account is overdue by 90 days or more, it is classified as having defaulted. However, due to the monthly resolution of the IRB data, Atom does not have access to real-time DPD values. As a result, Months in Arrears (MIA) is used as a proxy for DPD. The calculation for MIA is straightforward: the total sum of missed repayments is divided by the amount the account pays in each period to estimate the MIA. If the MIA value is greater than or equal to three, the account is classified as defaulted. Although this is the primary method Atom uses to define default, other circumstances must also be considered, where an account might still meet the default classification. We will briefly discuss some of these circumstances, but it is worth noting that Atom's policy rigorously outlines over 23 different "types" of default in their write-off policy.

One of the most common ways an account enters default, aside from DPD considerations, is through Unlikeliness to Pay (U2P) indicators. These are internal markers that suggest an account or small-to-medium enterprise (SME) may be experiencing financial difficulties and could potentially miss payments as a result. The most common U2P indicator is when the client or SME has defaulted on another account, either with Atom or another bank. Additionally, other circumstances might include when the sector in which the SME operates is declared to be in crisis. For example, if a period of political unrest disrupts key internationally-based resources that the SME relies on, this may flag to Atom that default is imminent.

Finally, we will touch upon default status achieved through forbearance or repossession. While Atom aims to support customers experiencing financial hardship, it is important to note that the bank cannot waive financial obligations. However, Atom can offer payment deferrals or extensions on payments and interest. This leads to another classification of default; that is, accounts for which Atom provides forbearance due to mitigating circumstances. A notable example of this occurred in March 2020, when, as part of the UK Government's economic response to the COVID-19 pandemic, lenders were advised to offer payment holidays of up to three months to borrowers [9].

1.0.3 The data

Following the introduction of the Internal Ratings-Based (IRB) approach, it becomes evident that a key limitation of this framework lies in the quality of a given bank's internal data. As a result, since the implementation of the Basel II Accord, there has been a strong emphasis on the careful maintenance and interpretability of such data. Given the maturity of Atom's portfolio, we will perform our analysis using representative external data provided by TransUnion as a proxy for internal data within an IRB framework. The dataset consists of 114 variables spanning over 12,000 accounts. It includes account information, bureau affordability fields and account performance data. This dataset has been updated monthly since its inception in March 2019, and we will focus our analysis on data collected up until November 2024. All analysis performed within this report was conducted through the use of the programming language R [22].

Foreshadowing to the Markov Chain analysis we will soon conduct, we now introduce the key variable of interest: the Risk Grade classification. Atom classifies each account into one of eleven risk grades, with Risk Grade 1 representing a "perfect" customer with the lowest Probability of Default (PD) and PD increases as the customer ascends through the risk grades. These risk grades will henceforth be referenced by the prefix "RG". To motivate our interest in this variable, consider that this variable will serve as the singular input for the model we will soon introduce, exemplifying the concept of "straightforwardness" described earlier. Atom define accounts in risk grades 9, 10 and 11 as "non-performing" accounts" - accounts which fall into this classification are considered to be of significant risk, for example in some circumstances Atom may consider support for such customers. Risk Grade 11 (RG11) is of particular interest as Atom defines an account in this grade as having defaulted. Therefore, in our analysis, the probability of default is equivalent to the probability of an account being classified as RG11.

In terms of Markov chain theory, which will be elaborated upon shortly, it is important to note that within this dataset RG11 is not an absorbing state. In short, this means that an account will not be stuck in the default position or, in other words, there is a non-zero probability of transitioning out of RG11 (this is rigorously defined in Section 2.2). This convention is in stark contrast to the majority of studies in this sector where Risk Grade 11 is typically modelled as an absorbing state (e.g. Trueck et al. (2009) [25], Jarrow et al. (1997) [17]). In the Markovian model we will soon implement, we will be particularly focused on the RG value of an account in the past and so another variable in the data we will pay close attention to is the account number. Since we are considering longitudinal data, this variable will serve as the means by which we will uniquely identify accounts at different points in time. However, looking into the data we see that there is a significant proportion of accounts recorded as having the following account number: "99999997". Instances of this value in the data correspond to a range of degenerate cases such as missing information. In this study, we choose to remove such accounts to ensure that each account, and in particular their history, are traceable. Figure 1.1 highlights the significant proportion of instances of this value, affirming the importance of data-maintenance in an IRB approach as in some months we have been forced to discarded over 15% of the data. To avoid the discarding of such data, one could perform imputation through an unsupervised machine learning technique such as hierarchical clustering to identify each unidentified account based on the remaining 112 variables.

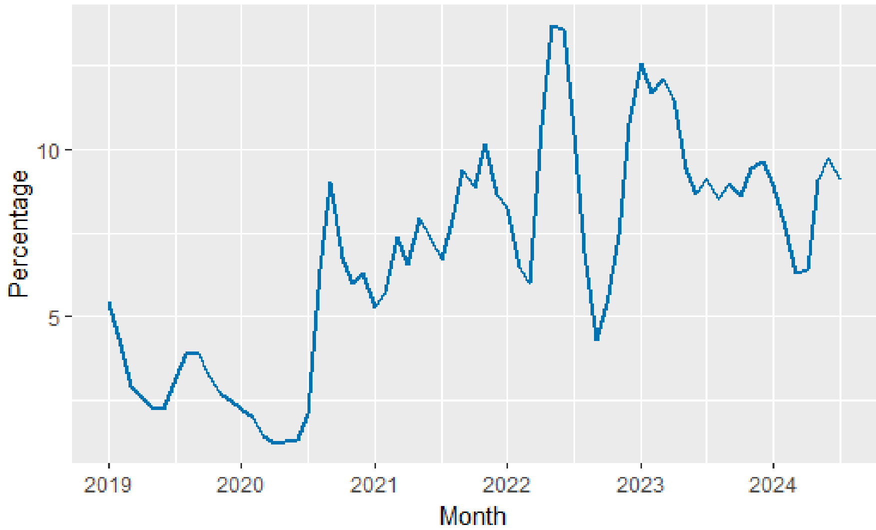

Figure 1.1: Percentage of accounts which have account number “99999997” for each monthly instalment of Atom’s IRB data

Since we still retain an average of 13680 accounts each month after discarding degenerate entries, we will omit such techniques here. For an overview of imputation techniques we suggest Wang (2024) [28] as further reading. Figure 1.2 displays the percentage of accounts which are classified as having defaulted at each point in time. Clearly, there has been a significant percentage increase in the number of defaulted accounts with the proportion in 2024 being over 4 times that of 2019. The financial market has been proven to undergo periods of contraction and expansion, for example during a recession (see Section 4.2.3 for a discussion on this). For this reason, one would never expect this graph to be constant however an increase of the magnitude we have seen here warrants in investigation. It should be noted that this increase does not necessarily imply a general increase in default rates. To see this suppose an account defaults through bankruptcy. They then begin an extensive legal procedure to have their debts discharged, during which time they remain marked as RG11 in the IRB data despite their default potentially occurring months ago. Such scenarios perhaps explain the trend we are seeing here; that being, its not necessarily that default is becoming more common but rather that we are seeing the cumulative frequency of previously defaulted accounts. To determine if default is becoming a more frequent event, we must investigate the proportion of accounts entering into RG11 over a given time period. It happens that these proportions will be explored in much detail later as they will be used as the sole parameter of the Markov model we will soon introduce (see transition probabilities in Section 2.2).

Figure 1.3 displays the proportion of accounts in RG01 and RG02 at each point in time. Once again, we would never expect these to be constant due to fluxations in the financial market however, the proportion of accounts in RG02 is somewhat constant over a suitably short interval. RG01 displays interesting behaviour opposite to what we observed in Figure 1.2 for RG11. We see a general decrease in the proportion of accounts in RG01 with the proportion approximately being half of what it was in 2021 by 2024. After presenting this result to Atom, we concluded that this is most likely to be due to a changes in the definitions of the risk grades that occurred post-pandemic. In short, these changes over time have led to the classification as RG01 or RG02 being harder to achieve which explains the observed decline from 2021 onwards.

The final consideration of this section is the following: Given that we primarily define an account to have defaulted if they are 90 or more days past due (DPD) on a payment - how do we address the lag by which an account cannot plausibly default during their first 90 days? One solution to this issue is to remove the lag entirely by at each point in time only considering accounts which have a history of at least 90 days in the data, therefore ensuring that default by DPD is a plausible event for all remaining accounts. To demonstrate the issue with this, consider the first 90 days of Atoms portfolio from March 2019 to May 2019. According to our default definition established in Section 1.0.2, we would expect that it is impossible to default through being 90 or more DPD over this period. Interestingly we see that there were a total of 37 defaults which, given our default definition established in Section 1.0.2, we assume to have occurred by means other than DPD considerations (e.g U2P indicators). This highlights the issue with removing the DPD-lag entirely; that is, that DPD considerations are not the only means by which an account can default. For example not including an account in subsequent analysis due to it not having a maturity that exceeds 90 days overlooks the possibility that they it is possible for them to attain RG11 by other means during this period, for example by fraudulent activity. It is for this reason that we choose to omit such considerations in this report, as although it may be true that an account cannot default through DPD during their first 90 days, Atom's definition of default outlined in Section 1.0.2 considers several means of defaulting, not just DPD.

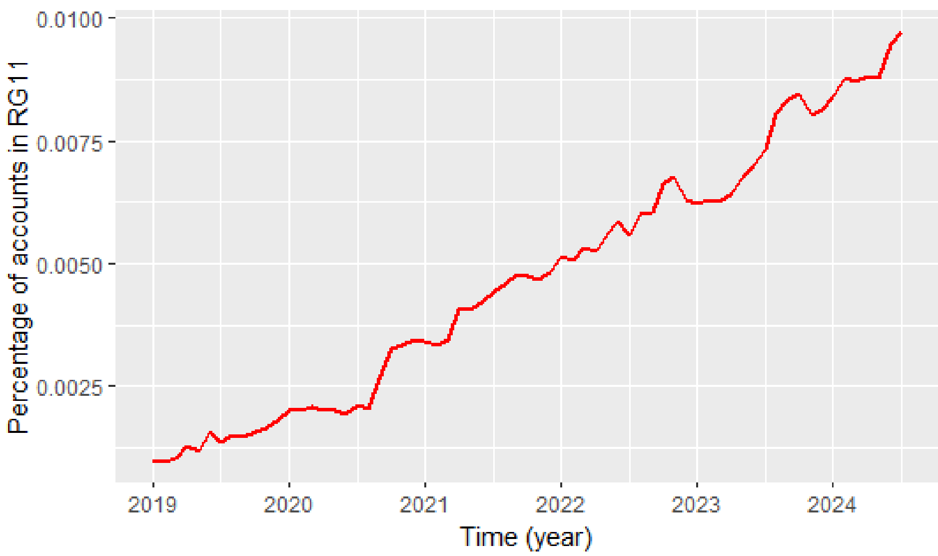

Figure 1.2: Percentage of accounts classified as RG11 at each point in time.

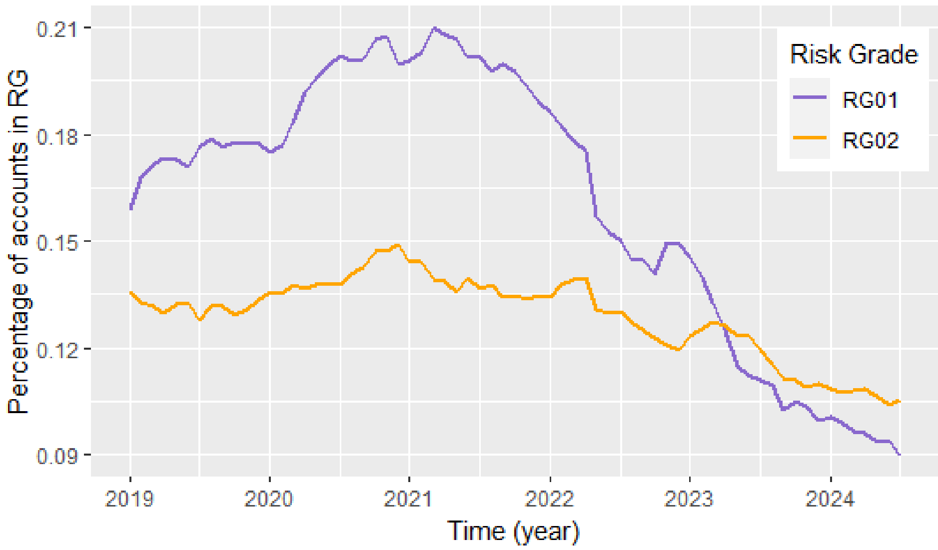

Figure 1.3: Percentage of accounts classified as RG01 and RG02 at each point in time

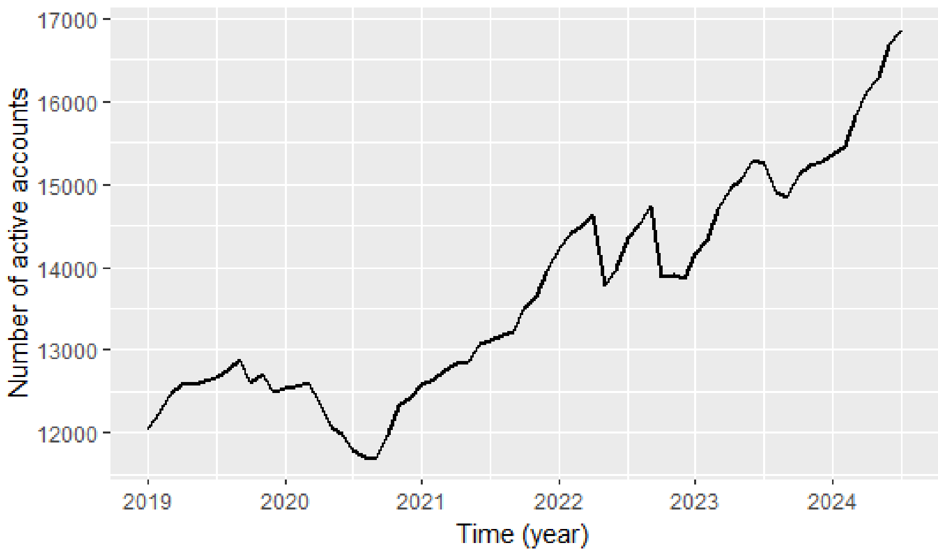

Figure 1.4: Total number of active accounts at each point in time

1.0.4 Markov chains introduction (A Faulty Machine)

In short, this report is focused on whether the movement of an account through the risk grades can be modelled by a certain type of process called a Markov Chain (MC). In the following section a more formal definition will be given but for now lets introduce this special type of process through an example. Imagine a temperamental machine in a workshop which breaks frequently. Suppose that at the beginning of each day the machine is either in a “Broken state (B)”” or in a “Working state (W)”. Assuming it starts the day as broken, there is a 0.6 probability that it is not fixed and therefore ends the day still broken. There is also a 0.4 probability it is fixed and ends the day as working. Conversely, if it starts the day in the working state then there is a 0.7 probability it is still working at the end of the day and hence a 0.3 probability that it brakes during the day. Note that we are assuming that only one transition can occur per day, for example it cannot break and be fixed on the same day. Here we have process which transitions between two states {W,B} according to some probability distribution. Figure 1.5 displays the process with the branches joining the states representing the probability of transitioning between them. Processes of this kind are described as Markov if they adhere to an additional property: the evolution of the process does not depend on where it has been in the past, only where it is presently. In the context of our example, this means that the probability of the machine breaking depends only on whether the machine is currently broken or working. If we assume this property to hold, the Figure 1.5 displays a simple 2-state Markov chain.

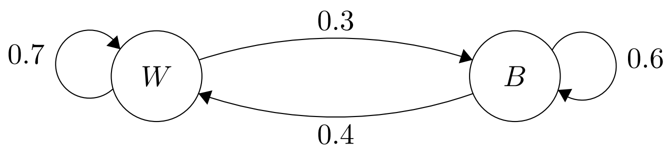

Figure 1.5: Markov Chain example: A Faulty Machine

Now suppose that the probability of the machine breaking was unknown. An observant worker makes note of the status of the machine for every day in the past. They seek to estimate the probability that tomorrow the machine will break based on their observations, or in other words, can the machines past be used to estimate probabilities regarding its future behaviour? Suppose the worker observes the machine over the last 3 days and notices the following pattern: Day1 = W, Day2 = W, Day3 = B. They then construct the following probability that the machine is working on day 4 (tomorrow):

$$\mathbb{P}(X_4 = W \mid X_3 = B, X_2 = W, X_1 = W) \overset{!}{=} \mathbb{P}(X_4 = W \mid X_3 = B)$$

By modelling this process as a Markov chain, we can “ignore” the past and say that the future only depends on where it is currently (as shown by $\overset{!}{=}$). This ‘ignoring” of the past is often characterised by the process being memoryless. This report explores if Markov chains are a suitable tool for answering the worker’s question, except in our case we will be interested in the analogous question of whether the history of an account banking with Atom can be used in estimation of probabilities regarding their future behaviour.

More formally, a Markov chain is a Stochastic process $(X_n)_{n\geq 0}$ which transitions from one state to another with probabilities dependent on only the current state and the time elapsed. In the case of time-homogeneity (a concept explored in Section 4.2.1), one can further relax this definition to only depending on the current state. The Markov property says that the next state is independent of where the chain might have been previously, or in short: the process does not retain memory. We shall consider markov chains in both discrete time:

$$n \in \mathbb{Z}^+ = \{1, 2, 3, 4 \dots\} \tag{1.0.2}$$

and in continuous time:

$$n \in \mathbb{R}^+ = [0, \infty) \tag{1.0.3}$$

Chapter 2

Discrete Markov chains

2.1 Discrete-time Markov chains

2.1.1 Chapter outline

Section 2.1.2 will begin by introducing discrete-time Markov chains in a more rigorous mathematical framework than in the case of the faulty machine example given previously following a similar construction to J.R Norris (1997) [20]. Section 2.1.3 will then explore how, under certain distributional assumptions, we can find maximum likelihood estimates for the transition probabilities between states. Section 2.2.1 then applies these estimates to the context of estimating PD in Atom's IRB data by replicating the methodology outlined in Lando et al. (2002) [17]. Finally, Section 2.2.2 will explore the results of this methodology and how its weaknesses motivate more complex models, in particular continuous MC models. Results in Section 2.2.2 imply similar conclusions to that of several papers which utilise this methodology such Bangia et al (2002) [4] and Trueck et al (2009) [25] regarding the key limitation being the modelling of rare events. Furthermore, note that all MCs in this chapter are assumed to be time-homogeneous. In essence this means that the transition probabilities are independent of the time at which the transition occurs (see Section 2.1.2).

2.1.2 Discrete-time Markov chains

Let $\mathcal{I}$ be a countable set called the state-space, where each $i \in \mathcal{I}$ is a state. Continuing the faulty machine example given previously in Section 1.0.4 we would take this to be {Broken,Working}. The processes we define here transition between these states according to some probability distribution, therefore making them Stochastic processes. Markov chains are a type of stochastic processes which are defined by an additional property of being memoryless. In other words, a Markov chain is a stochastic process which does not retain memory of the states it has visited in the past. To formalise this first declare that we will work over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a collection of outcomes, $\mathcal{F}$ is the set of subsets of Ω and for $A \in \mathcal{F}$, $\mathbb{P}(A)$ is the probability of A. We then consider a random variable $(X_n)_{n\geq 0}$ which takes values in $\mathcal{I}$ and is therefore a $\mathcal{F}$-measurable function which maps $X : \Omega \to \mathcal{I}$. We seek to construct a joint distribution of the sequence $X_0, X_1, ..., X_n$ in an attempt to understand its behaviour. We will start by constructing a distribution for where the process begins and then extend to subsequent transitions.

We say that a row vector $\{\lambda_i : i \in \mathcal{I}\}$ is a measure on $\mathcal{I}$ if $0 \leq \lambda_i < \infty \ \forall \ i \in \mathcal{I}$. Furthermore if $\sum_{i\in I} \lambda_i = 1$ then we say that λ is a distribution. Equipped with this, we define the starting position of the process to be given by the initial distribution $\{\lambda_i : i \in \mathcal{I}\}$. To extend to subsequent transitions given that we now have defined where it started, we define the one-step transition matrix

$P := (p_{i,j} : i, j \in \mathcal{I})$ where $p_{i,j}$ represents the probability of transitioning from state i to j over a unit interval. We define P to have the property of being a Stochastic matrix, meaning that each entry is non-negative and the row sums are zero. Note that this is equivalent to saying at each row of P is a distribution. Continuing the faulty-machine example in Section 1.0.4, a discrete-time MC would have the following transition matrix:

$$\mathrm{P} = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

Definition 2.1 (Markov chain) *We say that $(X_n)_{n\geq 0}$ is a **Markov chain** with initial distribution λ and transition matrix P if for all $n \geq 0$ and $i_0, \dots, i_{n+1} \in \mathcal{I}$,*

(i) $P(X_0 = i_0) = \lambda_{i_0}$;

(ii) $P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n i_{n+1}}$.

*For short, we say $(X_n)_{n\geq 0}$ is **Markov**$(\boldsymbol{\lambda}, P)$.*

In this definition, (ii) displays the memoryless property we have spoke of so far, once again we see that the future (X_{n+1}) is independent of the past $(X_{n-1}, X_{n-2}, ..., X_0)$. It turns out that this is a special case of a broader property of Markov chains, a property that completely encapsulates the notion of memory we have referred to so far:

Theorem 2.1 (The Markov property) *Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then conditional on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, X_1, \dots, X_m$.. Where:*

$$\delta_i = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

Throughout this Chapter, we also make the assumption the assumption of time-homogeneity. In short, this means that the transition probabilities are independent of the time at which the transition occurs. For a discussion on the validity of this assumption see Section 4.2 .Formally, this can be expressed by the following:

Definition 2.2 (Discrete Time-homogeneity) *A discrete-time Markov chain with state space $\mathcal{I}$ is said to be time-homogeneous if the transition probabilities are independent of time, i.e., for all $t \geq 0$ and $i, j \in \mathcal{I}$:*

$$P(X_{t+1} = j \mid X_t = i) = P(X_{t+1} = j \mid X_0 = i)$$

It should now be evident that the transition probabilities completely define the Discrete Markov chain. Suppose the parameters p_{ij} are unknown, we will now explore how to estimate the transition probabilities in such a senario.

2.1.3 Estimating transition probabilities

We will now explore how through maximum likelihood estimation we can attain estimates the unknown parameters p_{ij} in our model. Before we can derive such estimates, we must first define the Multinomial distribution. The Multinomial distribution is a generalization of the binomial distribution to more than two categories. It describes the probability of obtaining a specific combination of outcomes in a fixed number of independent trials, where each trial results in one of $k \in \mathbb{N}$ possible outcomes.

Definition 2.3 (Multinomial distribution) *The multinomial distribution is parametrised by the number of trials n and the probabilities $\pi_1, \pi_2, \dots, \pi_k$ of each outcome, where the vector $\pi := (\pi_i : i \in 1, ..k)$ defines a distribution such that $\sum_i \pi_i = 1$. A variable following a multinomial distribution is denoted $X \sim Mult(n, \pi)$ and has the following PDF for a realisation of X such that $x = (x_i : i \in 1, ...k)$:*

$$\mathbb{P}(X = x | n, \pi_1, .., \pi_k) = \binom{n!}{x_1! \cdot x_2! \cdots x_k!} \cdot \prod_{i=1}^{k} \pi_i^{x_i} \tag{2.1.1}$$

For simplicity, suppose we have a state-space consisting of just 4 unique positions, $\mathcal{I} := \{i_0, i_1, i_2, i_3\}$. We then wish to model the process given by $(X_n)_{n \geq 0}$ which takes values within this state space. Now suppose that $P := (p_{i,j} : i, j \in \mathcal{I})$ is unknown. We wish to model this process as a Markov chain but since the tranistion matrix defines such a process we must estimate some $\hat{P} := (\hat{p}_{i,j} : i, j \in \mathcal{I})$ such that $(X_n)_{n \geq 0}$ is $Markov(\lambda, \hat{P})$. To find such an estimate, we will employ Maximum Likelihood Estimation.

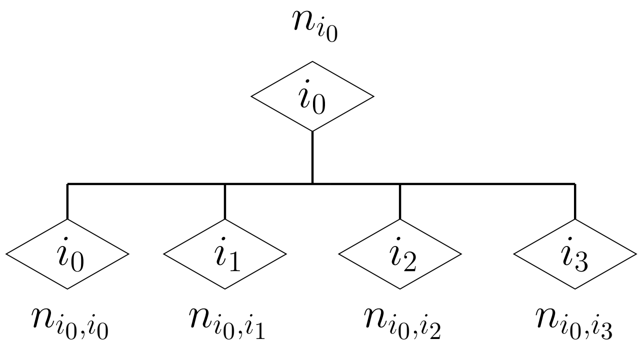

Figure 2.1: Tree showing possible subsequent transitions i_0, i_1, i_2, and i_3.

Suppose that we are observing many processes at once, all of which can all be identically and independently modelled by $Markov(\lambda, P)$ where P is some unknown transition probability matrix. Note that all of the processes have the same transition matrix P. The consideration of multiple processes at once is motivated by the fact that we will soon be modelling multiple mortgage accounts evolving simultaneously. We observe the processes at a fixed point in time and note the number which are currently in a given state, say $i_0 \in \mathcal{I}$. From this initial cohort in i_0, each process will either transition to another state or remain in i_0 over the following period. We define the following variables to represent the initial cohort and the frequencies of each subsequent transition:

n_{i_0} := The number of processes starting in state i_0

n_{i_0,i_j} := The number of processes which started in state i_0 and ended the period in state i_j

Figure 2.1 demonstrates this scenario. We now make the assumption that the proportion of the initial cohort n_{i_0} which will split off into each of the 4 possible states can be modelled by a multinomial distribution. Note that this is a very natural assumption to make as clearly we have a fixed number of trials (processes), each trial is independent (all processes have the same P) and subsequent transitions are mutually exclusive (a process cannot end the period in two states).

Figure 2.1 demonstrates the application of the Multinomial distribution in the context of Markov chains: The fixed number of trials corresponds to the inital cohort n_{i_0} in state i_0 and the probabilities of each of the processes splitting off into one of the 4 states over the next period are given by the probabilities encoded in some unknown probability distribution vector $(p_i : i \in I)$ which parametrises a multinomial distribution. To further display this, we can reconstruct equation 2.1.1 in the context of our Markov chain notation:

Let X be a vector containing the number of observations in each of the 4 states at the end of the period. Let x be a realisation of this vector such that $x := (n_{i_0,i_0}, n_{i_0,i_1}, n_{i_0,i_2}, n_{i_0,i_3})$. Also we define a distribution p such that $(p_i : i \in i_0, i_1, i_2, i_3)$ where p_i represents the probability of transitioning to the i^{th} state over the following period. Then assuming $X \sim \text{Mult}(n_{i_0}, p)$:

$$\mathbb{P}(X = x | n_{i_0}, p_{i_1}, .., p_{i_4}) = \left(\frac{n_{i_0}!}{n_{i_0,i_0}! \cdot n_{i_0,i_1}! \cdots n_{i_0,i_4}!} \right) \cdot \prod_{j=1}^{4} p_j^{n_{i_0,i_j}} \tag{2.1.2}$$

Of course, the choice of selecting the starting position to be the state i_0 is arbitrary. In fact, we could have chosen any of the 4 states and modelled the counts for subsequent transitions as a Multinomial distribution. This leads to the main conclusion of this section: each row of our transition probability matrix $P := (p_{i,j} : i, j \in \mathcal{I})$ corresponds to the probability distribution parameter in a multinomial distribution of the kind we have just seen. Therefore, we can estimate the probability parameter in each of the multinomial distributions for each starting state (row of P) and attain an estimate for the transition probability matrix, denoted $\hat{P}$. Estimation of the parameters of a Multinomial distribution through maximum likelihood estimation is a well-established procedure we will now outline:

2.1.4 Maximum likelihood for the Multinomial distribution

Upon inspection of the PDF given in the definition of the Multinomial distribution 2.1.1, it is clear we will have the following Log-Likelihood function:

$$log(L(\pi)) = log(n!) + \sum_{i=1}^{m} x_i log(\pi_i) - \sum_{i=1}^{m} log(x_i!) \tag{2.1.3}$$

We seek to estimate $\hat{\pi}$ by maximizing Equation 2.1.3 subject to the following constraint:

$$\sum_{i=1}^{m} \pi_i = 1 \tag{2.1.4}$$

And so we must maximise the Langarian:

$$\mathcal{L}(\pi, \lambda) = log(L(\pi)) + \lambda \left(1 - \sum_{i=1}^{m} \pi_i \right) \Rightarrow \tag{2.1.5}$$

$$\frac{\partial \mathcal{L}(\pi, \lambda)}{\partial \pi_i} = \frac{\partial}{\partial \pi_i} \left(log(n!) + \sum_{i=1}^{m} x_i log(\pi_i) - \sum_{i=1}^{m} log(x_i!) \right) - \lambda \Rightarrow \tag{2.1.6}$$

$$\frac{\partial \mathcal{L}(\pi, \lambda)}{\partial \pi_i} = \frac{x_i}{\pi_i} - \lambda \tag{2.1.7}$$

Computing the extremum as:

$$\frac{x_i}{\hat{\pi_i}} = \lambda \tag{2.1.8}$$

Then summing both sides and employing the constraint $\sum_{i=1}^{m} \pi_i = 1$ yeilds $\lambda = n$. Therefore the Maximum Likelihood Estimator for π is given by:

$$\hat{\pi_i} = \frac{x_i}{n}, i \in \{1, \ldots, n\} \tag{2.1.9}$$

Returning to context of Markov chains this now gives us a method for estimating an unknown transition probability matrix P. Lets demonstrate this by once again using the scenario outlined in Figure 2.1. Recall that we had an initial cohort n_{i_0} in state i_0 and that over the next period next period n_{i_0,i_1} accounts will transition to state i_1. The MLE above says that probability for transitioning between these state is best estimated by the proportion of observations which transition to state i_1. Therefore

$$\hat{p}_{i_0,i_1} = \frac{n_{i_0,i_1}}{n_{i_0}} \tag{2.1.10}$$

We then simply repeat this calculation for all pairs $(p_{i,j} : i, j \in I)$ and we have an estimate $\hat{P}$ for the transition probability matrix of the Markov process.

2.2 Application to estimating probability of default

Now that we have formally established our definition of the Markov chain and how to estimate its parameters, we shift our focus to applying this to the financial sector and in particular to analysis of probability of default. The use of Markov chains in application to credit migrations was first introduced in by Jarrow, Lando and Turnball [17] in 1997 leading to the method of applying such models being commonly referred to as JLT. The joint paper introduced the framework of using discrete and continuous Markov chains as a means for estimating probability of default. Let us begin to reformulate the Markov chain theory seen so far in the context of Atom banks IRB data with the objective to estimate probability of default. Recall that in section 1.0.3 we foreshadowed our interest in the Risk grade variable, in particular, this will become our state-space. Formally, we are now considering a random variable $(X_n)_{n\geq 0}$ which takes values in the finite set {RG01,RG02,...RG11}. We then construct a probability matrix P:= $(p_{i,j} : i, j \in I)$ wherein the entry $p_{i,j}$ denotes the probability of transitioning between any pair of risk grades $i, j \in \{RG01, RG02, ...RG11\}$. We now outline the primary objective of this report, to estimate P. Our motivation for this objective is the following. Recall that an account in risk grade 11 is classified as having defaulted, therefore PD is equivalent to the probability of entering risk grade 11. Suppose we are able to model the process $(X_n)_{n\geq 0}$ as $Markov(\lambda, \hat{P})$ where $\hat{P}$ is the maximum likelihood estimate for P given element-wise by 2.1.10. We then simply need to to take the column of $\hat{P}$ corresponding to transitioning into RG11 to attain PD estimates. Note that we henceforth assume this to be the 11^{th} column without loss of generality.

Lets introduce this through an intuitive example. Suppose we have an account with the risk grade history displayed in Figure 2.2. We seek to estimate the probability that next month they will default (X_4 = RG11). We begin by constructing the following probability:

Figure 2.2: Example risk grade migration pattern

$$\begin{aligned}\mathbb{P}(X_4 = RG11 \mid X_3 = RG08, X_2 = RG06, X_1 = RG04, X_0 = RG02)\\ \stackrel{!}{=} \mathbb{P}(X_4 = RG11 \mid X_3 = RG08) = p_{8,11}\end{aligned}$$

The first equality ($\stackrel{!}{=}$) displays the Markov property in action, often characterised by the process being memoryless. In essence, this greatly simplifies the calculations as we are able to “ignore” the past through conditional independence and exclusively condition on where the process is currently. The validity of this step (or equivalently the validity of the Markov assumption), is a concept we explore extensively in Section 4.1. For now however, we assume ($\stackrel{!}{=}$) is a valid step exemplifying the straightforwardness of the Markov chain model. The second equality shows how the probability of observing such a migration pattern, and any pattern for that matter, simplifies to an entry in the transition matrix. In this case the probability of default for an account currently in RG08 can be interpreted as $p_{8,11}$, regardless of their history. To clarify, changing the values of X_0, X_1 and X_2 in this scenario will result in the same default probability as above.

In section 1.0.3 we defined RG11 to not be an absorbing state. Let us now formally define what was meant by this. In short, an absorbing state is a state that once entered, cannot be left. Two states are said to communicate if the probability of transitioning between them is non-zero [20]. For our model, we assume that the sate-space uniquely partitions into one communicating class. In other words, the probability of transitioning between any pair of risk grades is non-zero. Note that this condition will be important later as we will observe a clear contradiction of this in section 2.2.1. One should also note that this is in stark contrast to most studies conducted in this area which define risk grade 11 to be an absorbing state [4][18][25].

Assuming an account's risk grade migrations to behave as a Markov chain, we have now shown that the task of estimating probability of default is equivalent to finding the transition probability matrix and taking the $(i, 11)^{th}$ for an account currently in risk grade $i \in \{RG01, ...RG11\}$. Since we have found an estimate for such probability in Section 2.1.4, we are now equipped to estimate PD using Atom's IRB data. First we estimate $\hat{P}$ through the maximum likelihood result (2.1.10) we derived earlier, then we simply take the column corresponding to RG11. This procedure is commonly referred to as the Cohort Method or JLT method due to it being first applied in the context of credit migrations by Jarrow, Lando and Turnball (1997) [17].

2.2.1 The Cohort Method

As outlined previously, if we make the assumption that an accounts migration pattern through the risk grades can be modelled by the discrete time Markov chain then have a simplistic method for estimating PD. Let $n_{i,j}$ denote the number of accounts transitioning between risk grades $i, j \in \{RG01, ..., RG11\}$ and n_i denote the number of accounts initially in risk grade i. Through application of equation (2.1.10) We can then estimate the probability of transitioning between these risk grades by:

$$\hat{p}_{i,j} = \frac{n_{i,j}}{n_i} \tag{2.2.1}$$

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	24438	2273.8	304.2	129.6	35.00	60.6	21.0	14.2	0.4000	0.0000	0.8000
RG02	2334	16551	2016	393.4	97.20	70.0	64.4	38.0	0.2000	0.0000	0.6000
RG03	396.0	2094	26991	2773	466.2	190	119	99.0	1.800	0.0000	0.4000
RG04	160.4	496	3061	23410	2184	462.2	244.8	244.6	2.000	0.0000	3.200
RG05	38.40	115.2	504.2	2553	14335	1325	429.4	220.0	3.400	0.0000	3.600
RG06	42.80	69.60	139	429.8	1698	8659.6	1032	223	2.800	0.2000	2.400
RG07	5.400	10.0	54.2	148.4	440.2	1393.8	9084	871	4.400	0.2000	4.400
RG08	14.2	11.0	43.0	82.8	106	164.4	1121	8933	23.0	0.6000	21.20
RG09	0.2000	0.0000	0.2000	0.4000	0.4000	0.0000	0.6000	17.2	10.2	10.2	7.000
RG10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.800	1.000	3.800	8.000
RG11	0.0000	0.2000	0.8000	0.2000	0.6000	0.8000	1.400	10.6	0.6000	0.6000	763

Table 2.1: Average one year transition frequencies using Atom's IRB data from 2019 to 2024. Entries correspond to $n_{i,j}$ in equation 2.1.1. $(i, j)^{th}$ entry denotes the expected number of transitions from risk grade i to j over a year.

10^{-2}	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	89.590	8.336	1.115	0.475	0.128	0.222	0.077	0.052	0.0010	0.00000	0.0030
RG02	10.823	76.749	9.349	1.824	0.451	0.325	0.299	0.176	0.0010	0.00000	0.0030
RG03	1.195	6.320	81.468	8.370	1.407	0.573	0.360	0.299	0.0050	0.00000	0.0010
RG04	0.530	1.639	10.112	77.342	7.217	1.527	0.809	0.808	0.0070	0.00000	0.0110
RG05	0.197	0.590	2.582	13.074	73.408	6.788	2.199	1.127	0.0170	0.00000	0.0180
RG06	0.348	0.566	1.130	3.495	13.805	70.410	8.389	1.813	0.0230	0.0020	0.0200
RG07	0.045	0.083	0.451	1.235	3.663	11.600	75.599	7.249	0.0370	0.0020	0.0370
RG08	0.135	0.105	0.409	0.787	1.008	1.563	10.653	84.915	0.2190	0.0060	0.2020
RG09	0.431	0.000	0.431	0.862	0.862	0.000	1.293	37.069	21.983	21.983	15.086
RG10	0.000	0.000	0.000	0.000	0.000	0.000	1.293	37.069	21.983	21.983	54.086
RG11	0.000	0.026	0.103	0.026	0.077	0.103	0.180	1.361	0.077	0.077	97.971

Table 2.2: Average one year transition probabilities, estimated using Cohort method applied to frequencies shown in Table 2.1. $(i, j)^{th}$ entry denotes the probability of transitioning between risk grade i and j over a unit interval, scaled by 10^{-2}.

Table 2.1 displays the average number of yearly transitions between each pair of risk grades. We observe that the row frequencies are heavily weighted around the diagonal entry and in general we see a decrease in frequency moving away from the diagonal. This supports the anticipated behaviour that most accounts will remain in their current grade with larger changes such as jumps of multiple risk grades appearing as rare events. Note that the fractional frequencies arise from the fact that we are averaging over each year. We then divide each of the entries by their row sums in accordance with equation 2.2.1 to achieve our estimated transition probability matrix, shown in Table 2.2. We will focus on the column corresponding to risk grade 11 as these are the default probabilities (highlighted). We observe reasonable estimates with a general monotonic increase in PD ascending through the grades. Excluding RG11, We observe that on average it is most probable to default from RG10, this is to be expected since RG10 contains the accounts which have the worst performance. Interestingly, we observe that PD for RG02 is higher than that of RG03, a clear contradiction of the risk grade definition we gave earlier where we defined PD to increase strictly monotonically as RG increases. To investigate this further we will consider Table 2.3 which displays the estimated PD rates using Atom's IRB data over April 2021 to 2022. This timespan has been carefully selected to emphasise the issue we have just observed with RG03. This time, the contradiction is even more clear as PD for RG03 is far less than PD for RG02, in fact, its zero! Therefore according to this model, defaulting from RG03 is impossible. This contradicts the assumption of the state space being a single communicating class which implied that all transitions should have a non-zero probability. This leads on to the main limitation of the cohort method: rare events are assigned as impossible events.

$\hat{P}_{i,11}$	RG11
RG01	0.000329
RG02	0.000435
RG03	**0**
RG04	0.000318
RG05	0.000651
RG06	0.000861
RG07	0.001109
RG08	0.003293
RG09	0.250811
RG10	0.583029
RG11	0.832577

Table 2.3: Default probabilities estimated using Cohort method applied to Atom's IRB data from April 2021 to April 2022.

2.2.2 Limitations of the Cohort method

In the previous section we saw that the model introduced so far is capable of producing results which suggest that defaulting from a given risk grade is an impossible event. This highlights the main limitation of the Cohort method: if there are not any observations of a given transition, then it is automatically assigned a zero probability, as was the case for time-period considered in Table 2.3. Figure 2.3 explains the results of Table 2.3 where we observed a zero probability of defaulting from RG03 as we observe that in general defaulting from RG03 is a very rare event, with it only occurring four times over the entire history of Atom Bank's portfolio. Due to it being such a rare event, we can carefully select an interval such that there is not a single observation of a default from RG03. As a result, the numerator in equation 2.2.1 will be zero and therefore the model assigns a probability of zero from the considered risk grade. To summarise, it is possible for rare events to be assigned zero probabilities despite our model assuming all transition probabilities to be non-zero. This motivates the next topic of study in which we extend our discrete time Markov chains to its continuous time analogue in order to resolve this issue.

On the other hand, Figure 2.3 displays that RG08 is the most common risk grade to default from. Interestingly we are seeing that it is more common to default from RG08 than RG09 or RG10. We will explore the cause of this result in greater depth in Section 4.1 but for now we may speculate that this is due to the following effect that is often referred to as "momentum". In order to have reached RG08 the customer must be experiencing financial crisis, it is then likely that by the next month they have already defaulted. This draws upon another weakness of the cohort method, the model is confined to the monthly discrete resolution of the data and there is no indication of which point

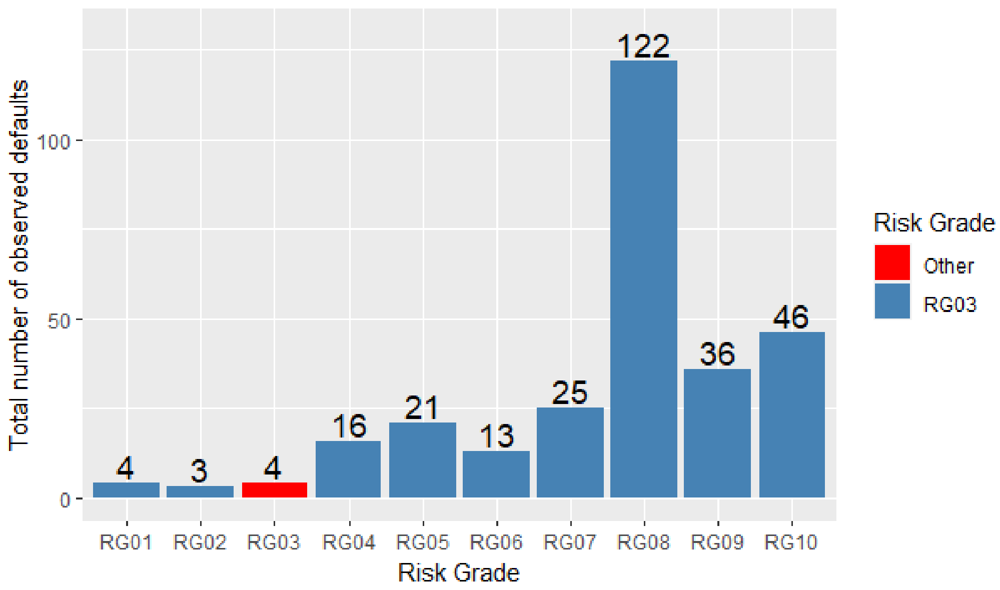

Figure 2.3: Total number of default occurrences from each of the risk grades over the entire history of Atom Bank's IRB data.

within the month a default occurs. If you reach RG08 then it is unlikely you will be able to "hang-on" another month until the process is observed again, hence the high frequency of defaults over a month starting in RG08. Suppose we are able to scale our time horizon to a week. We might expect to see an increase in the number accounts defaulting from RG09 and RG10 as this shorter time period would allow us to observe jumps occurring less than a month apart. With this new time-resolution, we might observe that many of the accounts which were thought to have defaulted from RG08 were in fact instead performing a sequence of rapid jumps ,say from RG08 to RG10, followed by a default but due to the monthly resolution we could not detect these rapid jumps. The assumption that the jumps after RG08 are "rapid" is predicated on the idea of financial momentum. This is the commonly observed phenomenon that entering a current state from a downgrade increases the probability of a subsequent downgraded, compared to say entering into the current state by upgrade (see Section 4.1.3).

To further demonstrate this weakness of the discrete-time MC, consider the following example. Suppose an account is constantly observed to be in RG01 except for the third week of every month when they spend this week in RG10. At the end of the third week they return to RG01. Under the model constructed in this section, we observe their position at the start and end of each month to be in RG01 and hence and assign them a very low PD. We are effectively "blind" to the spikes in risk grade that occur in the 3rd week - subsequently assigning them as a low-risk customer despite their volatile behaviour. In essence, the third week can be viewed as a "bind-spot" of the model. This motivates the following chapter in which we will extend our model to continuous-time with the goal of filling in these blind-spots (see motivating question, Section 3.0.2)

To summarise, discrete time Markov chains provide a "quick and dirty" approach banks can use as a baseline estimate for PD values. However there are two key limitations: modelling rare events and an inability to scale to arbitrary time horizons. We will next see that both of these issues can be addressed through extending our model to continuous time.

Chapter 3

Continuous Markov chains

3.0.1 Chapter outline

This chapter is concerned with the goal of extending the discrete-time Markov chain theory to continuous time, motivated by the discussion at the end of the previous chapter. Section 3.0.2 begins by introducing the extension to the continuous case through viewing it as an embedding problem. For consistency with the discrete-time theory provided earlier, we construct the Continuous Markov chain (CTMC) once again in a similar fashion to Norris JR. (1997) [20]. This section then introduces a certain type of matrix known as a Generator matrix in Definition 3.3 and justifies its significance for our application. The main result of Section 3.0.2 is Theorem 3.1 which provides a means for recovering transition probabilities (and hence default rates) given a Generator matrix. The framework, which we refer to as the Duration method, for the use of generator matrices in estimating credit migration rates was introduced by Lando et al. (2002) [18] and we replicate their methodology in Section 3.1. Our results presented in Section 3.1.2 regarding the advantages of the continuous-time extension compared to the discrete "Cohort Method" agree with those found in Christisen (2002) [6] and Trueck (2009) [25]. Finally, a comparison of these two approaches can be found in section 3.1.3.

3.0.2 Continuous-time Markov chains

Following the previous section in which we outlined limitations of discrete-time Markov chains when used to estimate PD, we now shift our focus to extending our theory to the continuous-time case. Whereas in the Discrete case we consider the process $(X_n)_{n\geq 0}$ where $n \in \mathbb{Z}^+$, we are now interested in the state of the process at time t given by $(X_t)_{t\geq 0}$ where $t \in \mathbb{R}^+$. The initial obstacle to this extension is that in the discrete time we spoke of a "one-step" transition matrix; in the continuous case however, the notion of a smallest step does not hold. we are now concerned with the task of viewing the process in discrete time as being embedded in its continuous time analogue. Such a task can be interpreted mathematically as the following embedding problem:

Motivating question: Consider a transition probability matrix $\mathrm{P} = (p_{i,j} : i, j \in \mathcal{I})$ on a finite set $\mathcal{I}$, is there a natural way to fill the gaps in the sequence $(\mathrm{P}^n : n = 0, 1, 2, ...)$? Recall at the end of the previous chapter the key limitation of the discrete MC model was that incorrect inferences could be easily made due to inherent "blind-spots" in our model (see Section 2.2.2). This gives rise to our motivating question as by extending to continuous time we effectively fill in the blind-spots of the discrete time MC. To demonstrate how we extend the discrete-time theory and answer such a question, suppose an account has the following risk grade migration pattern:

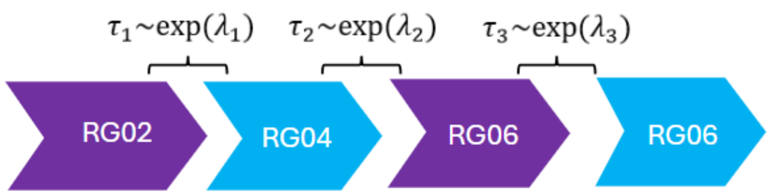

Figure 3.1: Migration pattern for an account in Atom's IRB data illustrating the holding time construction

Suppose the account enters into RG02 as shown in Figure 3.1. In discrete-time, jumps can occur at times belonging to a known finite set. We simply wait until the next jump and then observe where the process is. In continuous time, the amount of time the process spends in a state before transitioning is a random variable. Formally, this means that the trajectory of the process to be the right continuous step function $t \mapsto X_t$ with jumps at the times for which transitions occur. The amount of time spent in a state before transitioning is a random variable τ known as a holding time. What then can we say about the distribution of this random variable? Before we can establish the distribution of τ, we must first define time-homogeneity in continuous time. Note that, as in the discrete case, we restrict our interest to time-homogenous processes for now.

Definition 3.1 *A Markov process is **time-homogeneous** if, for every $0 \leq s < t$,*

$$P(X_t = j \mid X_s = i) = P(X_{t-s} = j \mid X_0 = i). \tag{3.0.1}$$

Suppose X_t is a time-homogeneous Markov process that starts at state $X_0 = i$. Let τ be the time it waits until transitioning to a new state $j \neq i$. We want to determine the distribution of τ.

Notice that:

$$P(\tau \geq t+s \mid \tau \geq s) = P(X_u = i \text{ for } u \leq t+s \mid X_u = i \text{ for } u \leq s).$$

By the Markov property, this is equivalent to:

$$P(X_u = i \text{ for } u \leq t+s \mid X_s = i).$$

By time homogeneity, this simplifies to:

$$P(X_u = i \text{ for } u \leq t \mid X_0 = i) = P(\tau \geq t).$$

Therefore, τ has the memoryless property:

$$P(\tau \geq t+s \mid \tau \geq s) = P(\tau \geq t).$$

This implies that the holding times τ have the defining properties of being independent and stationary, therefore making them a Poisson process and hence exponentially distributed with some rate λ_i (see Ross (2019) [23]). These rates λ_i represent the rate at which the process leaves state i. After entering a state x_i, the process stays there for a random length of time $\tau_i \sim exp(\lambda_i)$ before transitioning. This is illustrated in Figure 3.1. To recap, the answer to the question of <u>when</u> will the transitions occur is encoded within the transition rates λ_i. In essence, these $\lambda' s$ encapsulate the temporal structure of the process. We now turn to the question of where will the process go next? As in the discrete–time case, this is answered by the transition probabilities $p_{i,j}(t)$. So together both $(\lambda_i : i \in I)$ and $(p_{i,j}(t) : i,j \in I)$ completely define the Continuous-Time Markov chain (CTMC). The former contains the information of when the transitions will occur, and the latter defines the embedded Markov structure determining where the process will go next. Notice that the transition probabilities are now a function of a time parameter, this is the result of there no longer being the notion of a "step" as used in the discrete case where we speak of a one-step transition matrix. We are now equipped to formally define the Markov chain in continuous time.

Definition 3.2 (Continuous time Markov chain (CTMC)) *A stochastic process $(X_t)_{t\geq 0}$ on the finite state-space I is Markov if given a finite duration $0 \leq s < t$ and a set of prior times denoting the times in which the process migrates from one state to another $0 \leq s_0 < s_1 < \cdots < s_n = s$, one has the defining Markov property that $\forall$ choice of $\{x_1, x_2, ..., x_n\}$:*

$$\mathbb{P}(X_t = j \mid X_{s_1} = x_1, X_{s_2} = x_2, \dots X_{s_n} = x_n) = \mathbb{P}(X_t = j \mid X_s = x_n)$$

Recall that λ_i represents the rate at which the process leaves state i, so by taking its product with $p_{i,j}$ we can define $q_{i,j} := \lambda_i p_{i,j}$ which can be thought of as the rate of transitioning from state i to j. The concept of this being a rate is critical in what follows as we will soon express it as a derivative of $p_{i,j}(t)$. Since $q_{i,j}$ contains both structural elements of the CTMC (Temporal and the embedded discrete Markov chain) we can say that $Q := (q_{i,j} : i,j \in I)$ completely determines the CTMC. In fact, under certain conditions given below, we say that it generates the CTMC.

Definition 3.3 (Generator/Q Matrix) *Let $\mathcal{I}$ be a countable set. A Q-matrix on $\mathcal{I}$ is a matrix $Q = (q_{ij} : i,j \in I)$ satisfying the following conditions:*

(i) $0 \leq -q_{ii} \leq \infty$ *for all* $i \in \mathcal{I}$

(ii) $q_{ij} \geq 0$ *for all* $i \neq j$

(iii) $\sum_{j\in J} q_{ij} = 0$ *for all* $i \in \mathcal{I}$

For a matrix defined as $q_{i,j} := \lambda_i p_{i,j}$ which satisfies the three properties above, we say that Q generates the CTMC. Moreover, it is the key tool in answering our motivating question of this chapter given earlier. To see this consider the following theorem:

Theorem 3.1 *For a Markov process $(X_t)_{t\geq 0}$ admitting a generator Q such that the following limit exists for all $i,j \in I$:*

$$q_{i,j} \triangleq \lim_{h\to 0} \frac{p_{i,j}(h) - p_{i,j}(0)}{h}$$

Then we have that:

$$\frac{d}{dt}P(t) = QP(t)$$

and therefore:

$$P(t) = e^{Qt}$$

[Proof of Theorem 3.1]

First, observe that:

$$p_{ij}(0) = I = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Now, we use the Chapman-Kolmogorov equations (see JR. norris (1997) [20]), which state that for any $s, t \geq 0$:

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s).$$

Therefore:

$$\begin{aligned} \frac{p_{ij}(t+s) - p_{ij}(t)}{h} &= \frac{p_{ij}(t) p_{ij}(h) - p_{ij}(t)}{h} \\ &= \frac{p_{ij}(t)\left(p_{ij}(h) - I\right)}{h} \\ &= p_{ij}(t) \lim_{h \to 0} \frac{p_{ij}(h) - I}{h}. \\ &= P(t) \lim_{h \to 0} \frac{P(h) - P(0)}{h}. \\ &= P(t) Q \end{aligned}$$

Together with the initial condition that P(0)=I, this forms what is known as the Kolomorgorov forward equation. It has a unique solution which we will now see. To solve this equation, first notice that the matrix exponential can be reformulated as an infinite sum:

$$e^{Qt} = \sum_{n=1}^{\infty} \frac{(Qt)^n}{n!} \tag{3.0.2}$$

Taking the derivative term-by-term:

$$\begin{aligned}
\frac{d}{dt}e^{Qt} &= \frac{d}{dt}\sum_{n=1}^{\infty}\frac{(Qt)^n}{n!} \\
&= \sum_{n=1}^{\infty}\frac{d}{dt}\frac{(Qt)^n}{n!} \\
&= \sum_{n=1}^{\infty}\frac{(Qt)^{n-1}}{(n-1)!}Q \\
&\overset{!}{=} \sum_{m=0}^{\infty}\frac{(Qt)^m}{(m)!}Q \\
&= Qe^{Qt}
\end{aligned}$$

Therefore $P(t) = e^{Qt}$ satisfies the Kolmogorov forward equation. An identical method follows for the Kolmogorov backwards equation, which we omit here. It is worth noting that the assumption ($\triangleq$) of the theorem is very natural as we defined $q_{i,j}$ to be a rate of transitioning from i to j and therefore its expression as the limit of the derivative of $p_{i,j}$ follows. Theorem 3.1 is important as, given a Q-matrix, we are equipped with means for recovering P; that is, we simply take the matrix exponential of Q scaled by an arbitrary time horizon t.

Let us now return to our motivating question of recognising CTMCs as embedded in their discrete-time analogue. Recall that our aim was to fill the gaps in the sequence ($\mathrm{P}^n : n = 0, 1, 2, ...$). Suppose that our CTMC admits a matrix Q, through Theorem 3.1 we then can find a P such that:

$$P = e^{Q} \tag{3.0.3}$$

Then, by taking powers:

$$e^{nQ} = (e^{Q})^n = \mathrm{P}^n \tag{3.0.4}$$

And therefore ($e^{tQ} : t \geq 0$) fills the gaps in the discrete sequence.

To recap, the matrix Q completely defines the CTMC. Assuming we can find such a matrix, we can recover our transition probabilities (and later default probabilities) through Theorem 3.1 by taking the matrix exponential. We now turn our attention to the problem of finding such a matrix Q, we present this as a step in a broader method for estimating transition probabilities called the Duration method.

3.1 The Duration Method

The following method for estimating transition probabilities is an extension of the previously established Cohort method to continuous time. It was first introduced in the context of credit migrations by Lando et al (2002) [18] and here we replicate the methodology in a similar fashion to Trueck et al. (2009) [25]. We choose to introduce this method through the context of estimating PD in Atoms IRB data however the methods discussed here apply to all continuous Markov processes admitting a valid generator. Note that we omit discussion regarding the existence and uniqueness of a valid generator here however Appendix A provides the necessary conditions for a valid generator provided by Norris (1997) [20].

Let us begin by reformulating the CTMC theory seen thus far in the context of Atom's IRB data with the objective of estimating PD. Similar to the discrete case, we take our state-space to be the different risk-grades $\{RG01, ..RG11\}$. We define a sequence of random variables $(X_t)_{t\geq 0}$ where $t \in \mathbb{R}^+$, which will again be the focus of our study. We model the evolution of this sequence by a CTMC, in essence meaning that the probability of future migrations between risk grade occurring is conditionally independent of where the process has been previously. Recall that we define a CTMC through its generator matrix. In the discrete case, we were faced with the challenge of estimating an unknown transition matrix $\hat{P}$. In the continuous case, we instead focus on estimating the generator matrix $\hat{Q}$. The following maximum likelihood estimate for such a matrix was provided by Küchler and Sorensen in their 1997 paper "Exponential Families of Stochastic Processes" [16]:

$$\hat{q}_{ij} = \frac{n_{ij}}{\int_0^T Y_i(s)\, ds}, \quad \text{for all } i \neq j, \tag{3.1.1}$$

$$\hat{q}_{ii} = -\sum_{i \neq j} \hat{q}_{ij}, \quad \text{for all } i \in I. \tag{3.1.2}$$

Where:

$n_{i,j}$:= Total number of transitions from i to j over the period [0,T]

$\hat{q}_{i,j}$:= maximum likelihood estimate for the rate of transitioning from i to j.

Y_i := Number of accounts in state i at time $s \in [0, T]$ of accounts in the state i at time s.

We will provide a sketch of the derivation. Begin by noting the similarity to the Cohort (multinomial) from the previous chapter (2.2.1). In fact, the numerator in equation 3.1.1 is identical to that of the discrete counterpart 2.2.1 and the denominator has just been scaled to the arbitrary time horizon. The estimate for the off-diagonal elements $\hat{q_{ii}}$ is an application of (iii) in definition 3.3 by enforcing a zero row-sum. We are now fully equipped to define the CTMC for the process $(X_t)_{t\geq 0}$ and hence estimate its transition probabilities. The Duration Method can be summarised as follows: first we estimate the generator matrix $\hat{Q}$ using the maximum likelihood estimates provided in equations 3.1.1 and 3.1.2. Through theorem 3.1 we are then able to recover our estimate for $\hat{P}$ by the following:

$$\hat{P(t)} = e^{\hat{Q}t} \tag{3.1.3}$$

We then simply take the column corresponding to risk grade 11 to achieve estimated PD values. One of the key advantages of this method now becomes clear; that is, we can scale to any arbitrary time horizon through variation of t. To clarify, the transition probabilities for any time horizon (and therefore default rates) are found using the same mechanism of $\hat{Q}$. This means that for two different time horizons we do not need to "start over" and find a new $\hat{Q}$, we simply scale t accordingly. However, this does raise interpolation issues regarding the use of time horizons beyond that of data used to estimate Q. Before we apply this method to Atom's IRB data in an attempt to estimate PD, we will first perform a simple example to demonstrate the calculations in equations 3.1.1 and 3.1.2.

3.1.1 Worked Example of Duration Method

In similar fashion to Trueck et al. (2007) [25], we now present an example of utilising the Duration Method to recover transition probabilities. For simplicity, suppose there are three risk grade classifications: RG01, RG02, RG11 (Default). At the beginning of the year, there are 10 accounts in RG01, 10 accounts in RG02 and zero in RG11. Now suppose that over the following year we observe the following behaviour:

- one account translations from RG01 to RG02 after one month and stays there for the remainder of the year.
- one account tranistions from RG02 to RG11 after 3 months and stays there for the remainder of the year.
- one account tranistions from RG11 to RG01 after 6 months and stays there for the remainder of the year.

Modelling this with a discrete-time Markov chain defined in Section 2.1.1 we can utilise the Cohort method outlined in Section 2.2.1 to estimate tranistion probabilites given in $\hat{P}_{cohort}$. Consider the element $\hat{p}_{1,2}$ which denotes the probability of transitioning from RG01 to RG02. Through equation 2.2.1 we can estimate $\hat{p}_{1,2}$ by:

$$\hat{p}_{1,2} = \frac{1}{10}$$

Since there was 10 accounts beginning the year in RG01 and only one transition observed from RG01 to RG02. We repeat this calucation for the remaining elements to arrive at the following transition matrix:

$$\hat{P}_{cohort} = \begin{pmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

We will now alternatively use the Duration method to estimate the transition probability matrix $\hat{P}_{duration}$ for the same scenario. The following demonstrates the use of the MLE equations 3.1.1 and 3.1.2 in estimating a generator matrix given the above scenario. We first define the following:

$n_{1,2}$:= number of transitions from RG01 to RG02 over the entire year.

$\hat{q}_{1,2}$:= maximum liklihood esitmate for the rate of transitioning from RG01 to RG02.

$Y_1(s)$:= number of accounts in state RG01 at time s.

We then apply 3.1.1 to estimate the non-diagonal elements of the underlying generator matrix. If we once again seek to estimate $p_{1,2}$ then the numerator is clearly equal to 1 as there was only once instance of transitioning from RG01 to RG02 over the year. The denominator requires more consideration: 9 of the accounts spent their entire year in RG01; one account spent their first month in RG01 and finally one account spent half of the year in RG01. We sum these proportions to arrive at the following calculation:

$$\hat{q}_{1,2} = \frac{n_{1,2}}{\int_0^1 Y_1(s)\,ds} = \frac{1}{9 + \frac{1}{12} + \frac{6}{12}} \approx 0.1043$$

Next we find the diagonal elements through enforcing the row-sum condition (iii) in Definition 3.3. In this senario this results in:

$$\hat{q}_{1,1} = -\hat{q}_{1,2} \approx -0.1043$$

Repeating this procedure for the remaining entries, we arrive at the following esimtate for the generator:

$$\hat{Q} = \begin{pmatrix} 0.1043 & -0.1043 & 0 \\ 0 & -0.0983 & 0.0983 \\ 0.8 & 0 & -0.8 \end{pmatrix}$$

Then by taking the matrix exponential in equation 3.1.3 we find the following estimate for our transition matrix:

$$\hat{P}_{Duration} = \begin{pmatrix} 0.9020 & 0.0942 & 0.0037 \\ 0.0285 & 0.9074 & 0.0641 \\ 0.5195 & 0.0326 & 0.4502 \end{pmatrix}$$

We delay the full comparison of these methods to Section 3.1.3 however here we highlight the fact that in $\hat{P}_{duration}$, all of the elements are non-zero. This a result of the basic property that the exponential function is strictly non-zero. Note that this agrees with the assumption of our model made in Section 2.2 that the state space is a single communicating class and hence all transition probabilities are non-zero. Note that $\hat{P}_{cohort}$ fails to meet this requirement as there are several zero entries. Furthermore, the entries in $\hat{P}_{cohort}$ are non-zero despite there being not a single observation of many of transition pairings, we give a full description of this observation in Section 3.1.3.

3.1.2 An empirical study: The duration method

We will now apply the Duration method to Atom's IRB data with the aim of achieving PD estimates. Once again, we take our state-space to be the different risk-grades $\{RG01, ..RG11\}$. We define a sequence of random variables $(X_t)_{t\geq 0}$ taking values in the state-space which we assume can be modelled by a CTMC. Following Lando (2002) [18], our methodology is the following. The duration method first estimates the generator matrix underlying this process and then recovers the transition probability matrix using equation 3.1.3. Since we are interested in Default probabilities we take the column of this matrix corresponding to transitioning into RG11. An example of results from this methodology are show in Table 3.1 (left).

$\hat{P}_{i,11}$	RG11
RG01	0.000238
RG02	0.000246
RG03	**0.000170**
RG04	0.000318
RG05	0.000149
RG06	0.000242
RG07	0.000181
RG08	0.001076
RG09	0.252134
RG10	0.753031
RG11	0.964426

$\hat{P}_{i,11}$	RG11
RG01	0.000329
RG02	0.000435
RG03	**0**
RG04	0.000318
RG05	0.000651
RG06	0.000861
RG07	0.001109
RG08	0.003293
RG09	0.250811
RG10	0.583029
RG11	0.832577

Table 3.1: Estimated monthly PD values using the Duration Method (Left) and Cohort Method (Right) using data taken from Atom Bank's IRB dataset over April 2021 to April 2022.

Looking the results of the duration method over April 2021 to April 2022 (Table 3.1) we see reasonable estimates for default rates with a somewhat monotonic increase in default rates ascending through the risk grades which agrees with our definitions given in section 1.0.4. Recall from the previous chapter that the main limitation of the Cohort method was its tendency to characterise rare events as impossible. At the end of the previous chapter we foreshadowed that by extending to continuous-time we would overcome this limitation. To see this consider that both of these methods used the same interval of IRB data, and yet one models defaulting from RG03 as impossible whereas the Duration method provides strictly positive probabilities. We have already established why the Cohort method assigns this probability of zero; that being if there are no observation instances of a transition then the numerator in equation 2.2.1 is zero. We will now focus on demonstrating how the Duration method overcomes this limitation. Consider Figure 3.2 which shows the total frequency of defaults from each of the eleven risk grades over the same period used in the results from Table 3.1. There are zero observations of default during this period for RG03 and so the Cohort method proves a PD of zero, as seen in Table 3.1 (right).

On the other hand, due to the basic properties of the exponential in equation 3.1.3, every transition probability resulting from the Duration method will be non-zero. To clarify, the Duration Method assigns non-zero probabilities even in the event that zero observations of the given transition occurred. Note that this aligns with the assumption we made earlier that the state-space uniquely partitions into one communicating class, or in other words, every transition is possible.

3.1.3 Comparison of Cohort and Duration methods

Here we summarise the discussion above and outline the advantages of each of the methods. Recall that we introduced continuous Markov chains with the aim of overcoming the limitation that the Cohort method assigns zero probabilities to rare events. Above we outlined that the Duration method overcomes this limitation and can in fact provide non-zero estimates for transition probabilities even in the instance that there is not a single observation of such a transition. The second advantage of the Duration method is that it is easily scalable to different time horizon; simply scale the value of t in equation 2.2.1. Lando (2002) [18], Trueck (2009) [25] and Christisen (2002) [6] report similar findings on the inability for the Cohort method to capture rare events.

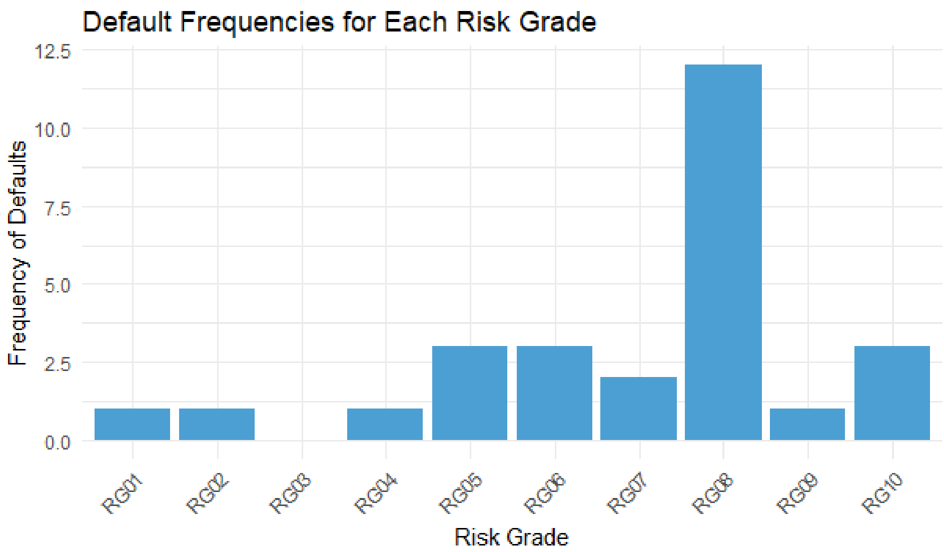

Figure 3.2: Default frequencies for each risk grade over April 2021 to April 2022 in Atom's IRB data

On the other hand, there are several advantages to the Cohort Method. Firstly, it is worth noting the calculations required are far simpler than in the Duration since the computing the matrix exponential can become computationally expensive for a large state-space. Next consider that for internal ratings systems it is common for observations to be reported quarterly or even yearly. In such scenarios, the Duration method is not applicable as we often do not know what time the transitions occurred within the considered period and so the denominator in equation 3.1.1 is not computable. Finally consider that the Duration method makes the assumption that there exists a unique and valid generator. Up until now, we have assumed that the Markov process admits a unique generating matrix. Is Israel et al. (2000) [12] portrays issue with this assumption in two parts. Firstly the authors discuss the existence of a generator for a Markov process (embedability) and secondly they discuss the task of finding such a generator (identifiability) once its existence is proven. In addition, there is the issue of deciding which generator to use in the case that there exits more than one valid generator for a given process. We omit such discussion here but draw upon the fact that the assumptions of Cohort Method are much simpler. In fact we have already discussed in 2.1.3 that the sole assumption of the cohort method is that the risk grade migration frequencies follow a multinomial distribution and this is natural for our application. For results on confirming the existence of a valid generator and a proposal of how to chose which generator to use in the case that there exists more than one, see Appendix A.

Chapter 4

Stability of credit migrations

This chapter focuses on accessing the validity of the two core assumptions we have made in the previous chapters; in particular these being the assumptions of the Markov property and time-homogeneity. Recall that the Markov assumption states that proceeding transition probabilities are only dependent on the current state and the time at which the transition occurs. The time-homogeneity assumption allows us to relax the latter condition and so our transition probabilities become exclusively conditional only on the current state. Several studies have collected evidence against the validity of the Markov assumption, for instance Bangia et al. (2002) reports non-Markovian "path-dependence" in S&P issuer ratings [4]. Trueck et al. (2009) [25] reports evidence of rejecting a first-order Markov model in favour of a second-order counterpart for an internal ratings based model in addition to further evidence of path-dependence through analysis of conditional momentum matrices. It is worth noting that despite there being substantial evidence against the assumption of time-homogeneity, this does not implicitly contradict the Markovian assumption and so this motivates us investigating them separately.

4.1 Markov assumption investigation

Recall from Chapter 2 that we define a stochastic process to be Markov if it possesses the "memoryless" property. In short, means that the probability distribution of future transitions is conditionally independent of previous states. Or perhaps more intuitively, it does not collect any "memory" of where it has been in the past. This section aims to outline a statistical test for the validity of this assumption by estimating migration matrices conditional on where the process has been previously. For instance, we seek to answer questions such as: does the probability of a downgrade increase if the account entered into its current state by a downgrade? This section will replicate the methodology outlined in Bangia et al. (2002) [4] and Trueck et al. (2009) [25] for answering this question. We will begin by conducting a likelihood ratio test provided by Goodman (1957) [8].

Before we can identify the hypotheses we wish to test, we must first outline what kind of behaviour would provide evidence against the Markov assumption. Up until now, the term "Markov Chain" has in fact referred to a first-order MC. We omitted the order previously to aid the clarity of explanations but since we are now going to consider the history of the process beyond the previous state and so we provide the following definition:

Definition 4.1 ($k^{\mathbf{th}}$ order Markov Chain) *The random variable $(X_n)_{n\geq 0}$ exhibits k^{th} order Markov behaviour if the conditional distribution of X_n on past states depends only on the k most recent states, $\{X_{n-1}, X_{n-2}, \ldots, X_{n-k}\}$, and does not depend on earlier states, such as $\{X_{n-k-1}, X_{n-k-2}, \ldots\}$. Formally, this is expressed as:*

$$\mathbb{P}\left(X_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0\right) = \mathbb{P}\left(X_n \mid X_{n-1} = x_{n-1}, \ldots, X_{n-k} = x_{n-k}\right)$$

To aid interpretation, one might consider this as saying that the process only retains "memory" of where it has been over the last k transitions. We will now refer to a Markov chain of first order as MK1 with higher orders denoted analogously.

4.1.1 Likelihood ratio test

Goodman (1958) [8] provides a Likelihood ratio (LR) test for testing the Markov assumption by testing whether, given a MC of k^{th} order, there is evidence to suggest that a MC of higher order is better suited. If there is evidence, then this is a clear violation of the k^{th} order Markov property as it suggests that the process retains memory beyond the k^{th} previous transitions. Mathematically, this is formulated by the following hypotheses. Let $m > k$, then define:

$$\begin{aligned}
H_0 &: \mathbb{P}(X_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_0 = x_0) = \mathbb{P}(X_n \mid X_{n-1} = x_{n-1}, ..., X_{n-k} = x_{n-k}) \\
&\equiv \text{"Transition probabilites are indepependent of tranistions occuring before time n-k"} \\
H_1 &: \mathbb{P}(X_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_0 = x_0) = \mathbb{P}(X_n \mid X_{n-1} = x_{n-1}, ..., X_{n-m} = x_{n-m}) \\
&\equiv \text{"Transition probabilites are indepependent of tranistions occuring before time n-m"}
\end{aligned}$$

Let L_0 and L_1 be the likelihood functions under the assumptions of the null and alternative hypotheses respectively. For example, if we take the null assumption to be the naïve independent identical distribution (idd) which assumes that transition probabilities are independent of any previous states, including the current. In addition, if we take the alternative hypotheses to be a first order MC model then we have the following likelihood functions L_0 and L_1 respectivly:

$$L_0 := \prod_{j=1}^{n} p_{ij}^{n_j}$$

$$L_1 := \prod_{i=1}^{n} \prod_{j=1}^{n} p_{ij}^{n_i j}$$

As illustrated in Goodman (1958) [8], the following test-statistic is Chi square distributed through a simple application of Wilks theorem [27]:

$$LR = 2 \ln \left(\frac{L_1}{L_0} \right) \sim \chi_{\Delta m} \tag{4.1.1}$$

Where Δm is the difference in the number of estimated parameters in the null and alternative models. Continuing the example in the previous paragraph (iid vs MK1), we have the following test statistic:

$$LR = \ln\left(\frac{\prod_{i=1}^{n}\prod_{j=1}^{n} p_{ij}^{n_{ij}}}{\prod_{j=1}^{n} p_{ij}^{n_j}}\right) = 2\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\ln p_{ij} - \sum_{j=1}^{n} n_j \ln p_j\right) \sim \chi_{\Delta m}$$

Where Δm is the difference in the number estimated parameters in the naive iid (H_0) and first order MC (H_1) models respectively. When calculating Δm, it is important to impose the row sum constraint of a stochastic transition matrix. In addition, we only estimate parameters for which non-zero transition frequencies where observed. An example of such a calculation will be given in the following subsection. p_{ij} can simply be estimated using the Cohort and Duration methods outlined in Chapters 2 and 3. For simplicity we will use the Cohort estimates over a suitably large interval as to ensure that rare events are observed.

$$\hat{p}_{i,j} = \frac{n_{i,j}}{N_i} \tag{4.1.2}$$

However if we wish to test hypotheses of higher order Markov chains we must now define such an estimate for Markov chains of order greater than one. Consider a second order Markov chain. Let p_{ijk} denote the probability of transitioning from j to i given that the previous transition was from k to j:

$$p_{ijk} := \mathbb{P}(X_n = i \mid X_{n-1} = j, X_{n-2} = k)$$

In order to calculate LR for a test involving a second order MC we will need to estimate this value. For this reason, we extend the Cohort/multinomial estimate as follows:

$$\hat{p}_{ijk} = \frac{n_{ijk}}{N_{jk}} \tag{4.1.3}$$

Where n_{ijk} is the number of transitions transitioned from j to i in the current period and from k to j in the previous period. N_{jk} is the number of transitions from k to j that occurred in the previous period. Transition probabilities for higher order Markov chains are defined analogously.

4.1.2 Empirical Study: Markov assumption LR-test

We will now apply the theory discussed above to Atom's IRB dataset. As discussed previously, we take our state space to be the risk grade spectrum ranging from grade RG01 to RG11 (default). To calculate test statistics given by equation (4.1.1) we estimate our transition probabilities using the multinomial maximum likelihood estimates given by equation (4.1.2)

Table 4.1 displays the average one-year transition frequencies for a first-order Markov chain, this provides us with the n_{ij} values needed to calculate LR. Table 4.2 shows the estimated one-year transition probabilities, $\hat{p}_{ij}$, calculated using the frequencies provided in Table 4.1. Table 4.3 displays the average one-year transition frequencies (n_j) for a iid model, with the corresponding estimates for the average one-year transition probabilities provided also ($\hat{p}_j$).

Substituting the values provided by these tables, the LR test statistic outlined in equation (4.1.1) can be calculated as 434466.71. To find Δm, one can first consider the iid model: Clearly there are 11 probabilities to calculate as there are 11 states but since we impose the condition that they sum to 1, we conclude that the iid model has 11-1=10 degrees of freedom. As for the first-order MC

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	24438.20	2273.80	304.20	129.60	35.00	60.60	21.00	14.20	0.40	0.00	0.80
RG02	2334.00	16551.00	2016.20	393.40	97.20	70.00	64.40	38.00	0.20	0.00	0.60
RG03	396.00	2094.00	26991.40	2773.20	466.20	190.00	119.40	99.00	1.80	0.00	0.40
RG04	160.40	496.00	3060.60	23409.80	2184.40	462.20	244.80	244.60	2.00	0.00	3.20
RG05	38.40	115.20	504.20	2553.20	14335.20	1325.60	429.40	220.00	3.40	0.00	3.60
RG06	42.80	69.60	139.00	429.80	1697.80	8659.60	1031.80	223.00	2.80	0.20	2.40
RG07	5.40	10.00	54.20	148.40	440.20	1393.80	9084.00	871.00	4.40	0.20	4.40
RG08	14.20	11.00	43.00	82.80	106.00	164.40	1120.60	8932.60	23.00	0.60	21.20
RG09	0.20	0.00	0.20	0.40	0.40	0.00	0.60	17.20	10.20	10.20	7.00
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	1.00	3.80	8.00
RG11	0.00	0.20	0.80	0.20	0.60	0.80	1.40	10.60	0.60	0.60	763.00

Table 4.1: Average one-year transition frequencies, observed over 2019 to 2024. The $(i,j)^{th}$ entry denotes the average number of transitions from Risk Grade (RG) i to j over a year.

10^{-2}	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	89.59	8.34	1.12	0.48	0.13	0.22	0.08	0.05	0.00	0.00	0.03
RG02	10.82	76.75	9.35	1.82	0.45	0.33	0.30	0.18	0.00	0.00	0.03
RG03	1.20	6.32	81.47	8.37	1.41	0.57	0.36	0.30	0.01	0.00	0.00
RG04	0.53	1.64	10.11	77.34	7.22	1.53	0.81	0.81	0.01	0.00	0.01
RG05	0.20	0.59	2.58	13.07	73.41	6.79	2.20	1.13	0.02	0.00	0.02
RG06	0.35	0.57	1.13	3.50	13.81	70.41	8.39	1.81	0.02	0.00	0.02
RG07	0.05	0.08	0.45	1.24	3.66	11.60	75.60	7.25	0.04	0.00	0.04
RG08	0.14	0.11	0.41	0.79	1.01	1.56	10.65	84.92	0.22	0.01	0.20
RG09	0.43	0.00	0.43	0.86	0.86	0.00	1.29	37.07	21.98	21.98	15.09
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.33	6.85	26.03	54.80
RG11	0.00	0.03	0.10	0.03	0.08	0.10	0.18	1.36	0.08	0.08	97.97

Table 4.2: Average one-year transition probabilities, estimated over 2019 to 2024. The $(i,j)^{th}$ entry denotes the maximum likelihood estimate of transitioning from Risk Grade (RG) i to j over a year. Scaled by 10^{-2}.

model, there are a total of 11^2 parameters in the transition probability matrix so there are $11^2 - 11$ degrees of freedom remaining once we impose the Stochastic row-sum constraint. Finally we can only consider non-zero parameters, meaning we remove the 15 parameters corresponding to zero entries in Table 4.2. This results in $11^2 - 11 - 15 = 95$ degrees of freedom for the first-order MC. Therefore $\Delta m = 95 - 10 = 85$ which corresponds to a Chi Square statistic of $\chi^2_{\alpha=0.05}(85) = 64.75$. Since $434466.71 \gg 64.75$ we conclude there is a substantial amount of evidence at the 5% significance level supporting the alternative hypotheses of a first-order MC model, as expected. These results, alongside higher order tests, are summarised in Table 4.4.

Table 4.4 displays the results of testing a null hypotheses of a first-order MC against an alternative hypotheses of a second-order MC. The test successfully rejects H_0 at the 5% level and so we conclude that a MC of order-two is a more suitable model for the process, and therefore in the interest of Atom it will yield better forecasting capabilities. Similar results are observed for a testing of a second-order MC against an alternative third-order MC at the 5% level. We conclude that the third order MC improves upon the forecasting ability of the second order MC. Finally, we test a null hypotheses of a third-order MC against a fourth-order MC and conclude the test to be insignificant at the 5% level, as shown by the highlighted entry in Table 4.4. Therefore we conclude that for Atoms internal ratings based system, the distribution of a customer's risk grade is best modelled by a third-order MC. Altman and Kao (1992) [1], Lando and Skødeberg (2002) [18] and Bangia et al. (2002) [4] all reject a null hypothesis of a first-order Markov chain either through a LR test similar to here or through detection of path dependence, a topic we will soon explore. However it is worth noting that all of these studies explored data from major credit rating agencies such as Moody's S&P whereas

$\hat{p}_j$	n_j
1.63 $\times 10^{-2}$	27429.60
1.29 $\times 10^{-1}$	21620.80
1.97 $\times 10^{-1}$	33113.80
1.78 $\times 10^{-1}$	29920.80
1.15 $\times 10^{-1}$	19363.00
7.36 $\times 10^{-2}$	12327.00
7.23 $\times 10^{-2}$	12117.40
6.37 $\times 10^{-2}$	10672.00
2.97 $\times 10^{-4}$	49.80
9.31 $\times 10^{-5}$	15.60
4.86 $\times 10^{-3}$	814.60

Table 4.3: I.I.D model estimated average yearly tranistion probabilities p_j and observed tranistion frequencies n_j, calculated over 2019 to 2024

	H_0	H_1	LR	df	$\chi^2_{\alpha=0.05}(df)$
1	iid	MK1	424466.71	85	64.75
2	MK1	MK2	34196.01	487	436.83
3	MK2	MK3	13798.77	1804	1706.35
4	MK3	MK4	3113.05	5075	**4910.43**

Table 4.4: Results from applying Goodman's 1957 LR test in search of higher order Markov behaviour

here we consider an internal rating system based on credit scores only. Trueck et all (2009) applies Goodman's (1957) Likelihood ratio test to an internal rating system and also rejects a first-order MC model but concludes a Second order MC is most suitable, in comparison to our study which concludes a third-order is most suitable. It is worth noting that for each of the four tests conducted, the number of accounts considered decreases as we test higher order hypotheses. To see this, consider that in the instance of testing a third-order MC, we can only consider account which have a history of three periods available.

4.1.3 Rating drift

Rating drift is a term used in many studies (e.g Lando et al (2002) [18]) to describe the theorised non-Markovian tendency for accounts which have previously been downgraded/upgraded to be more frequently further downgraded/upgraded in the following period. Assuming this tendency to be true, it would be expected that monotonic "UpUp" transitions (accounts which are upgraded two periods in a row) to be more frequently observed than "UpDown" transitions. Analagously, we would expect that "DownDown" transitions are more probable than "DownUp" transitions. Mathematically, the two cases of rating drift can be expressed by the following:

$$\begin{aligned}\mathbb{P}(\text{UpUp} > \text{UpDown}) &\equiv \mathbb{P}(X_{t+1} > X_t \mid X_t > X_{t-1}) > \mathbb{P}(X_{t+1} < X_t \mid X_t > X_{t-1}) \\ \mathbb{P}(\text{DownDown} > \text{DownUp }) &\equiv \mathbb{P}(X_{t+1} < X_t \mid X_t < X_{t-1}) > \mathbb{P}(X_{t+1} > X_t \mid X_t < X_{t-1})\end{aligned}$$

Both of these cases are a clear violation of the Markov property as in such a case the process evolves with probabilities conditional on its previous state. This phenomenon is a particular case of a broader characterisation known as path dependency (see Bangia et al (2002) [4]). Path dependency is the general condition that the future transitions of the process are dependent on the "path" which the process took into the current state. In order to detect such Behaviour, we will consider a first order Markov chain with transition frequency matrix $\{M(t)\}_{ij}$ such that the $(i,j)^{th}$ entry denotes the number of transitions from state i to j over the time period t to t+1. We will then partition this matrix into three components *Up-Momentum-Matrix*, *Down-Momentum-Matrix* and *Neutral-Momentum-Matrix* such that:

$$M(t) = M_{Up}(t) + M_{down}(t) + M_{Neutral}(t) \tag{4.1.4}$$

where:

$\{M_{Up}(t)\}_{ij}$:= The number of transitions from i to j over the period t to t+1 such that the account as upgraded during the period t-1 to t

$\{M_{Down}(t)\}_{ij}$:= The number of transitions from i to j over the period t to t+1 such that the account as Downgraded during the period t-1 to t

$\{M_{Neutral}(t)\}_{ij}$:= The number of transitions from i to j over the period t to t+1 such that the account was neither upgraded or downgraded during the period t-1 to t

4.1.4 Empirical study: Path dependence detection

For this study we will recall the results in Table 4.1 which provide us with observed transition frequencies and maximum likelihood estimates for transition probabilities, both averaged yearly over 2019 to 2024. The objective of this section will be to find evidence path-dependency through analysis of momentum matrices. We take the entries of Table 4.1 to be $\{M(t)\}_{ij}$ as defined in the previous section. We can then partition the frequencies in $\{M(t)\}_{ij}$ by conditioning on weather they entered state i by upgrade or downgrade to provide us with $\{M_{Down}(t)\}_{ij}$, $\{M_{Up}(t)\}_{ij}$ and $\{M_{Neutral}(t)\}_{ij}$ as presented in the following Tables:

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	24438.20	2273.80	304.20	129.60	35.00	60.60	21.00	14.20	0.40	0.00	0.80
RG02	2334.00	16551.00	2016.20	393.40	97.20	70.00	64.40	38.00	0.20	0.00	0.60
RG03	396.00	2094.00	26991.40	2773.20	466.20	190.00	119.40	99.00	1.80	0.00	0.40
RG04	160.40	496.00	3060.60	23409.80	2184.40	462.20	244.80	244.60	2.00	0.00	3.20
RG05	38.40	115.20	504.20	2553.20	14335.20	1325.60	429.40	220.00	3.40	0.00	3.60
RG06	42.80	69.60	139.00	429.80	1697.80	8659.60	1031.80	223.00	2.80	0.20	2.40
RG07	5.40	10.00	54.20	148.40	440.20	1393.80	9084.00	871.00	4.40	0.20	4.40
RG08	14.20	11.00	43.00	82.80	106.00	164.40	1120.60	8932.60	23.00	0.60	21.20
RG09	0.20	0.00	0.20	0.40	0.40	0.00	0.60	17.20	10.20	10.20	7.00
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	1.00	3.80	8.00
RG11	0.00	0.20	0.80	0.20	0.60	0.80	1.40	10.60	0.60	0.60	763.00

Table 4.5: $\{M(t)\}_{ij}$:Average one-year transition frequencies, observed over 2019 to 2024. $(i,j)^{th}$ entry denotes the average number of transitions from Risk Grade (RG) i to j over a year.

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RG02	504.00	1440.19	148.19	36.38	2.48	1.71	4.19	1.52	0.00	0.00	0.00
RG03	120.19	512.76	1418.48	114.86	32.76	3.43	6.29	2.10	0.19	0.00	0.19
RG04	56.57	145.52	694.48	2035.05	159.43	26.67	10.67	13.14	0.00	0.00	0.57
RG05	7.05	34.86	170.86	635.24	1634.29	131.05	40.00	18.67	0.57	0.00	0.19
RG06	16.00	26.86	48.19	153.52	438.48	1170.10	133.33	31.05	0.76	0.00	0.19
RG07	1.14	2.48	10.67	44.19	146.10	368.57	1157.71	120.57	0.76	0.19	0.76
RG08	2.67	3.05	11.43	24.76	40.76	59.05	335.05	1194.48	1.52	0.00	3.62
RG09	0.19	0.00	0.19	0.38	0.38	0.19	0.57	14.10	7.62	9.90	4.57
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.33	0.76	2.10	7.24
RG11	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.19	52.00

Table 4.6: $\{M_{Down}(t)\}_{ij}$: Average one-year transition frequencies for accounts which maintained their risk grade over the previous period, observed over 2019 to 2024. $(i,j)^{th}$ entry denotes the average number of transitions that occurred over a year from i to j in which the account maintained state j over the previous period.

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	2206.67	506.67	66.29	29.52	4.76	6.29	3.81	1.14	0.00	0.00	0.19
RG02	220.57	1858.86	449.33	89.71	17.14	3.05	9.90	1.90	0.00	0.00	0.00
RG03	21.90	169.71	2635.62	657.14	100.38	20.00	11.81	7.62	0.38	0.00	0.00
RG04	5.90	19.24	236.76	2181.33	493.52	84.19	35.24	19.24	0.19	0.00	0.38
RG05	3.43	5.33	27.81	215.81	1472.38	301.14	96.76	24.19	1.33	0.00	0.76
RG06	0.76	0.57	5.90	27.62	186.67	976.19	260.38	40.19	0.19	0.00	0.19
RG07	0.19	0.38	4.19	6.10	21.90	87.62	759.43	212.19	0.76	0.00	0.57
RG08	0.00	0.00	0.00	0.19	0.38	0.00	0.57	24.38	2.48	0.00	1.14
RG09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.57	0.57	0.00	0.38
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.19	0.00
RG11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.7: $\{M_{Up}(t)\}_{ij}$: Average one-year transition frequencies for accounts which were downgraded in the previous period, observed over 2019 to 2024. $(i,j)^{th}$ entry denotes the average number of transitions that occurred over a year from i to j in which the account entered i by an downgrade.

	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	21118.67	1661.52	202.29	69.33	13.90	39.81	12.76	4.95	0.38	0.00	0.57
RG02	1471.81	12471.43	1353.90	231.24	57.14	35.62	43.05	17.52	0.19	0.00	0.57
RG03	230.48	1324.95	21700.95	1911.05	299.81	105.14	76.00	57.33	1.33	0.00	0.57
RG04	91.81	309.33	2010.29	17871.24	1488.95	281.52	178.29	150.67	1.71	0.00	2.10
RG05	25.71	71.62	294.29	1632.95	10448.76	858.67	291.43	152.76	1.71	0.00	2.86
RG06	22.86	37.14	83.05	238.86	1038.29	6157.90	633.71	159.62	1.90	0.19	1.90
RG07	3.81	7.24	38.48	96.19	268.00	924.57	6937.33	557.33	2.86	0.00	3.43
RG08	10.48	7.81	29.14	54.86	64.00	102.86	787.81	7642.29	21.90	0.57	18.48
RG09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.43	4.38	1.14	1.90
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.19	2.10	1.33
RG11	0.00	0.38	0.76	0.19	0.57	0.76	0.95	10.10	0.57	0.38	744.00

Table 4.8: $\{M_{Neutral}(t)\}_{ij}$: Average one-year transition frequencies for accounts which were neither upgraded or downgraded in the previous period, observed over 2019 to 2024. $(i,j)^{th}$ entry denotes the average number of transitions that occurred over a year from i to j in which the account remained in state i during the previous period.

Analysing the four Tables above, it is clear that our partition of $\{M(t)\}_{ij}$ approximately holds as it is close to the sum of the latter three matrices. It is worth noting that here the equality in equation (4.1.4) is not exact as $\{M(t)\}_{ij}$ was calculated using a larger selection of accounts than the momentum matrices as the latter could only be calculated using accounts which had a history of two periods, therefore excluding accounts during their first two months banking with Atom. However for the purposes of our study, an approximate equality is sufficient. In addition, notice that the first row of $\{M_{Down}(t)\}_{ij}$ are all zeros, this is due to the fact that one cannot be downgraded into RG01. Similarly, one cannot be upgraded into RG11 which explains why the final row of $\{M_{Up}(t)\}_{ij}$ are zeros. It is worth considering the similarity between $\{M(t)\}_{ij}$ and $\{M_{Neutral}(t)\}_{ij}$. The reason for this is simple, there were a total of 852867 observed transitions used in the calculation of $\{M(t)\}_{ij}$, of these 852867, 678162 belong to the "no momentum" category. Therefore we have an $\approx 80\%$ overlap in the universe used to calculate $\{M(t)\}_{ij}$ and $\{M_{Neutral}(t)\}_{ij}$, hence their similarity. Consequently, the following analysis must be viewed in the context that $\{M_{Up}(t)\}_{ij}$ and $\{M_{Down}(t)\}_{ij}$ make up a combined $\approx 20\%$ of the data, and so a large enough sample size must be taken to ensure that the 20% is in fact representative of the data. To aid interpretation when comparing $M_{Neutral}(t)$ to the smaller momentum matrices we scale the results by calculating the corresponding maximum likelihood estimates of transition probabilities produced from the momentum matrices. We denote these estimates by $P_{Up}(t)$, $P_{Down}(t)$ and $P_{Neutral}(t)$:

$\times 10^{-5}$	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	14876.74	3415.81	446.88	199.04	32.10	42.38	25.68	7.70	0.00	0.00	1.28
RG02	1585.13	13358.66	3229.13	644.73	123.20	21.90	71.18	13.69	0.00	0.00	0.00
RG03	115.11	891.87	13850.54	3453.38	527.52	105.10	62.06	40.04	2.00	0.00	0.00
RG04	36.56	119.13	1466.11	**13507.54**	3056.06	521.34	218.21	119.13	1.18	0.00	2.36
RG05	30.39	47.27	246.49	1912.87	13050.71	2669.23	857.67	214.42	11.82	0.00	6.75
RG06	9.68	7.26	75.05	351.03	2372.48	12407.10	3309.37	510.81	2.42	0.00	2.42
RG07	3.32	6.64	73.00	106.19	381.62	1526.46	13230.46	3696.70	13.27	0.00	9.96
RG08	0.00	0.00	0.00	124.49	248.99	0.00	373.48	15935.26	1618.43	0.00	746.97
RG09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7142.86	7142.86	0.00	4761.90
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9523.81	0.00	9523.81	0.00
RG11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.9: $P_{Up}(t)$: Cohort estimates for transition probabilities based on observed frequencies of accounts which were upgraded in the previous period. $(i,j)^{th}$ entry denotes the probability of transitioning from risk grade i to j conditional on a previous upgrade.

$\times 10^{-5}$	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG010	RG011
RG01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RG02	4488.78	12826.78	1319.83	324.02	22.05	15.27	37.32	13.57	0.00	0.00	0.00
RG03	1035.32	4416.93	12218.76	989.38	282.21	29.53	54.15	18.05	1.64	0.00	1.64
RG04	342.94	882.18	4209.97	**12336.61**	966.47	161.66	64.66	79.67	0.00	0.00	3.46
RG05	50.23	248.41	1217.63	4527.07	11646.85	933.92	285.06	133.03	4.07	0.00	1.36
RG06	150.99	253.44	454.76	1448.75	4137.74	11041.76	1258.22	292.98	7.19	0.00	1.80
RG07	11.75	25.45	109.64	454.21	1501.65	3788.38	11899.62	1239.30	7.83	1.96	7.83
RG08	30.30	34.63	129.86	281.35	463.15	670.92	3806.93	13572.05	17.31	0.00	41.12
RG09	95.24	0.00	95.24	190.48	190.48	95.24	285.71	7047.62	3809.52	4952.38	2285.71
RG10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2222.22	1269.84	3492.06	12063.49
RG11	0.00	0.00	0.00	0.00	0.00	0.00	138.03	0.00	0.00	69.01	18840.58

Table 4.10: $P_{Down}(t)$: Cohort estimates for transition probabilities based on observed frequencies of accounts which were downgraded in the previous period. $(i,j)^{th}$ entry denotes the probability of transitioning from risk grade i to j conditional on a previous downgrade.

In order to check for non-Markovian path-dependancy, we analyse how the probabilities are distributed either side of the diagonal. For instance, higher probabilities left of the diagonal in $\{P_{Up}(t)\}_{ij}$ would provide evidence that the probability of further upgrade being higher. Likewise, higher probabilities right of the diagonal in $\{P_{Down}(t)\}_{ij}$ would suggest that the probability of further downgrade is greater. These statements are equivalent to conditions (4.1.3) established earlier. To illustrate this, consider the $(RG04, RG04)$ entry in $\{P_{Up}(t)\}_{ij}$ highlighted in red, if we look to the left of the diagonal, we see that the probability of a further unit upgrade in the next period is 1.466×10^{-2}. Now if we consider the $(RG04, RG04)$ entry in $P_{Down}(t)$ highlighted in red and look to the left we see that the probability of a subsequent upgrade is 4.209×10^{-2}. To conclude, if the account enters into RG04 by a downgrade, then they are ≈ 3 times as likely to be upgraded to RG03 in the following period than if they entered into RG04 by an upgrade. This illustrates what is meant by path-dependency, as clearly the probability distribution of the future transition is not independent of the path the process took into the current state, as seen in this example in which a given path approximately trebled the probability of a future upgrade. Under the markov assumption, we would expect the highlighted entries to be equal.

	Diagonal	Left	Right
$M_{Neutral}(t)$	82.49 %	8.87 %	8.64 %
$M_{Up}(t)$	71.49 %	7.50 %	21.02 %
$M_{Down}(t)$	63.94 %	29.30 %	6.76 %

Table 4.11: Momentum matrices divided into percentage contributions from diagonal entries (trace), strictly lower triangular and strictly upper triangular components

Table 4.1.4 displays the distribution of frequencies either side of the diagonal for each of the three "momentum matrices". For example the "Right" column for $M_{Up}(t)$ displays the proportion of strictly upper triangular contributions. As expected, $M_{Neutral}(t)$ is most heavily weighted along the diagonal with 82.49% of its contributions being diagonal entries. This suggests that, on average, accounts which maintained their risk grade over the previous period will continue to maintain their risk grade over the next. Looking at the non diagonal contributions, we see an approximately equal proportion of 8% of customers upgrading and downgrading. Looking at $M_{Up}(t)$, we see once again the majority of customers maintaining the risk grade reached in the previous period. Interestingly, we do not observe similar non-diagonal contributions as in this case there are many, in fact three times as many, accounts being downgraded compared to upgraded. This suggests that for accounts which entered into their current state via upgrade, they are three times as likely to be downgraded in the next period compared to further upgraded. If we analyse $M_{Down}(t)$, the opposite phenomenon is observed: for accounts which were previously downgraded, there are approximately five times as many accounts being subsequently upgraded compared to further downgraded, as indicated by the 29.30 % contribution from the strictly lower diagonal entries. To illustrate the clear path-dependency here, consider the percentage of accounts being upgraded (Right of diagonal) for $M_{Up}(t)$ and $M_{Down}(t)$, the two percentages are 21.02 % and 6.67 % respectively. These proportions suggest that the probability of a subsequent upgrade is over three times as likely for an account which was previously upgraded than an account that was previously downgraded. Under a first-order Markov model, we would expect these proportions to be equal as under such a models future transitions are independent of the manner in which the process attained its current state.

To conclude, many studies find significant evidence for general path-dependency in credit rating migrations [4][18][25]. Most of these studies, also found evidence of a certain type of path-dependency characterised by "rating-drift" in which they observed a greater frequency of "DownDown" transitions compared to "DownUp" transitions. In contrast, we have found evidence for path-dependancy, but not the type of dependency characterised by these monotonic downgrades or upgrades. The type of dependency we have observed is such that accounts which downgraded in the previous period are more likely to be upgraded in the next period compared to further downgraded. In short this can be expressed as: "DownUp" is more common than "DownDown". An analogous statement follows for accounts which previously upgraded. Interestingly, the only study aforementioned that found similar results to this is Trueck et al. (2009) [25]. This is interesting as this is the only study mentioned that considered an internal ratings based system similar to Atom's, the others conducted research on major rating systems such as Moody's S&P.

$\hat{P}_{i,11}$	RG11
RG01	0.000329
RG02	0.000435
RG03	**0**
RG04	0.000318
RG05	0.000651
RG06	0.000861
RG07	0.001109
RG08	0.003293
RG09	0.250811
RG10	0.583029
RG11	0.832577

$\hat{P}_{i,11}$	RG11
RG01	0.000238
RG02	0.000246
RG03	**0.000170**
RG04	0.000318
RG05	0.000149
RG06	0.000242
RG07	0.000181
RG08	0.001076
RG09	0.252134
RG10	0.753031
RG11	0.964426

Table 4.12: Estimated transition probabilities over a single month, $\hat{P}_{i,11}$. Estiamted using data taken from Atom Bank's IRB dataset over April 2021 to April 2022

4.2 Time-homogeneity investigation

This section is concerned with testing the validity of the time homogeneity assumption we have made throughout this report. Recall that in the context of Markov chains, the process being time-homogenous means that the transition probabilities are independent of the time at which the transition is occurring. To motivate our interest in testing this assumption consider that up until this point we have been assigning probabilities of defaulting based on a singular estimated transition matrix that we assumed to be constant over time. There is significant evidence within the literature that credit migrations are not time-homogenous. For instance, Nickell et al (2002) [19] and Bangia et al. (2002) [4] attribute observed time-heterogeneity to economic cycle affects. In particular, Nickell et al. (2002) [4] produces evidence that default rates vary conditional on whether the economy is believed to be in a period of economic expansion or contraction. Declaring a given period to be a period of contraction or expansion is a difficult task and it for this reason we consider the results of Wei et al. (2003) [29] and Trueck et al (2009) [25]. Both of these studies overcome the identifiability issues in economic cycle analysis by attributing the cause of the observed heterogeneity to be exogenous macroeconomic variables instead; for instance, Trueck et al (2009) [25] compares the evolution of PD rates alongside GDP fluctuations. In light of these studies, we consider the reasoning behind observed heterogeneity to be a well researched area and therefore in this report we instead choose to focus on the task of formalising a statistical test for detecting time-heterogeneity. Section 4.2.2 outlines such a test by replicating the methodology of Trueck et al. (2009) [25] in which we use a Chi-square statistic provided in Goodman (1958) [8]. We conclude that there is significant evidence for time-heterogeneity in the models outlined during preceding chapters and introduce regime-conditioned migration matrices as a means to recover the time-homogenous assumption following suggestions in Kim (1999) [14].

4.2.1 Detecting Time-homogeneity

Intuitively, we know that stability is scarcely found within financial markets. For instance consider currency fluxuations; stock market crashes ; interest rate changes or commodity price increases. This section instead aims to formalise this notion of instability within the context of default rates with the aim of developing a statistical test for its detection. To see this consider Figures 4.2 and 4.1. Figure 4.2 displays how default rates vary for accounts in RG09 for years 2021 to 2024. Clearly, PD is not constant over this period with default rates varying from 3% at the end of 2021 to 30% in 2023. Trueck et al. (2009) [25] reports a similar degree of variation for Moody's S&P default frequencies.

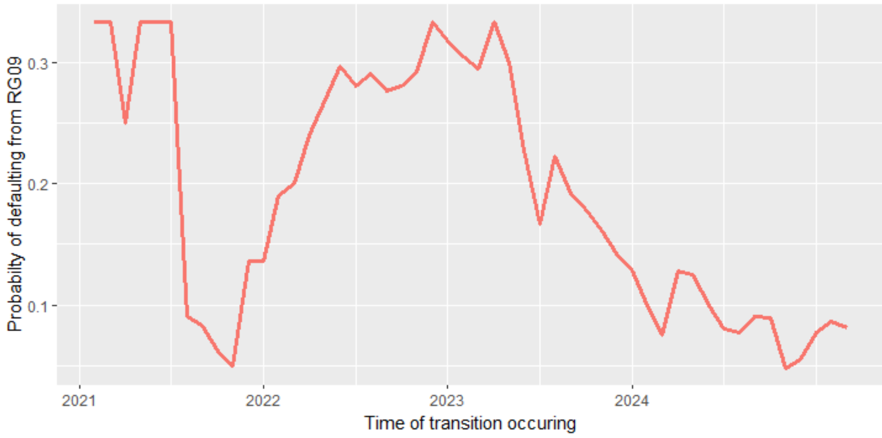

Figure 4.1: Estimated default rates using Atom's IRB data for accounts in RG09 and time horizon 2019 to 2024.

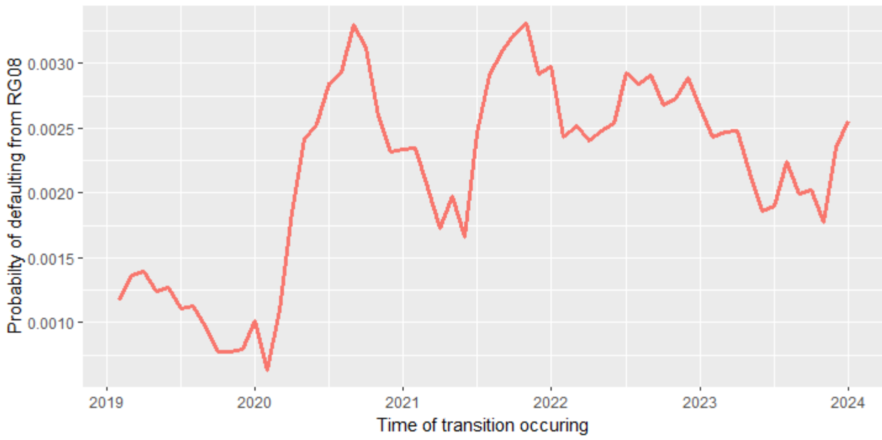

Figure 4.2: Estimated default rates using Atom's IRB data for accounts in RG08 and time horizon 2019 to 2024.

Figure 4.1 displays similar deviations from the mean default rate for accounts in RG08 over years 2019 to 2024. Note that here we do not consider variation of default rates for accounts in performing risk grades (i.e RG01 to RG07) as due to the low observed default frequencies from these lower risk grades, variation in default rates is most likely due to a lack of data rather than the theorised macroeconomic effects. This is once again a result of the Cohort methods inability to accurately capture rare events (see discussion in Section 2.2.2). Following these results we conclude that assuming default probabilities to be constant in a time-homogenous Markov chain model may yield incorrect forecasting results for future credit rating migrations. However through some idealization, we can use average transition matrices over a suitably long time period as a starting point for estimating credit risk, as suggested in Jafry et al. (2004) [13].

We now provide a statistical test for detecting time-inhomogeneity. The following test statistic was developed by Andersen and Goodman (1958) [2] and Billingsley (1961) [5]. We replicate its application to credit risk outlined in Trueck et al. (2009) [25] (Chapter 6). They begin by partitioning the data into at least two independent mutually exclusive samples of observations. For example we might select one sample of observations that we believe to belong to a period of economic expansion and one that we believe to belong to a period of economic contraction. The goal is to then see if the resulting transition probability matrices from these samples are significantly different. Note that in this section we will use the Cohort method to estimate the transition matrices, this decision was made to avoid issues regarding existence of a valid generator in the Duration method (see Appendix A). There are many means by which we can compare the migration matrices resulting from the two samples, Trueck et al. (2009) (Chapter 7) [25] provides a comprehensive analysis of different distance measures such as using classical matrix norms and eigenvalue analysis. Here we conduct our analysis using a Chi-square goodness of fit test. Before we conduct this test, we must first declare our hypotheses. Since for reasons aforementioned we will be using the discrete-time Cohort method to estimate our migration matrices and so we will also construct our hypotheses from a discrete-time perspective in alignment with Definition 2.2. Let X_u and X_s denote the positions of a discrete-time Markov chains at arbitrary times u and s respectively.

$$
\begin{aligned}
H_0 :=\ & \text{Transition probabilities are independent of the time at which the transition occurs.} \\
\equiv\ & \mathbb{P}(X_u = x_u \mid X_s = x_s) = \mathbb{P}(X_{u-1} = x_{u-1} \mid X_{s-1} = x_{s-1}) \quad \forall s, u
\end{aligned}
$$

$$
\begin{aligned}
H_1 :=\ & \text{Transition probabilities are dependent on the time at which the transition occurs.} \\
\equiv\ & \exists\, u, s \ \text{ s.t } \quad \mathbb{P}(X_u = x_u \mid X_s = x_s) \neq \mathbb{P}(X_{u-1} = x_{u-1} \mid X_{s-1} = x_{s-1})
\end{aligned}
$$

4.2.2 Time-homogenity test using Chi-square Distance

Following the procedure outlined in Trueck et al. (2009) [25], Goodman (1958) [8] provides the following test. First suppose that we divide the entire sample into T sub-samples. We will test to see if the migration matrices resulting from each of the T sub-samples differs significantly from the migration matrix resulting from using the entire sample. To measure this difference we use the following test statistic which we assume to be asymptotically chi square distributed with degrees of freedom Δ (see Goodman (1958) [8]):

$$
Q_t := \sum_{i}^{T} \sum_{j} \sum_{j \in V_i} n_i(t) \frac{(\hat{p}_{i,j} - p_{i,j})^2}{p_{i,j}} \sim \chi^2(\Delta) \tag{4.2.1}
$$

where:

$p_{i,j}$:= probability of transitioning from state i to j, estimated from the entire sample.

$\hat{p}_{i,j}(t)$:= probability of transitioning from state i to j, estimated from sub-sample $t \in \{1, ..., T\}$.

$n_i(t)$:= number of accounts in state i at the begging of the observation period in subsample t.

$V_i := \{j : \hat{p}_{i,j} > 0\}$. This ensures that we do not consider transitions for which no observations are available.

$\Delta :=$ Degrees of freedom given by the number of summands minus the number of estimated transition probabilities $\hat{p}_{i,j}$, corrected for the number of restrictions imposed by $\sum_j p_{i,j} = 1$ and $\sum_j \hat{p}_{i,j} = 1$.

We would then reject the null hypotheses, H_0, of time-homogeneity in favour of H_1 at significance level α if $\sum_{t=1}^{T} Q_t > \chi^2_{1-\alpha}(\Delta)$.

4.2.3 Empirical Study: Detecting Time-homogeneity in Atom's IRB data

As in previous Chapters, we declare that we take our state-space to be the risk grade classifications {RG01,..RG11} and assume that the migration pattern of an account be modelled by a discrete time-homogenous Markov chain. Let $(X_n)_{n\geq 0}$ represent this process and we say that this process is $Markov(\lambda, P)$ where λ is the initial distribution of the process and P represents the transition probability we wish to investigate. In alignment with the hypotheses above we wish to test if such a matrix P is dependent on a time parameter. We partition the observations $\{X_1, ...X_n\}$ into T independent sub-sets. For simplicity, we partition the five years of IRB data into individual years meaning T=5. For each of these years we then estimate a transition probability matrix using the data from that year. Let P_i denote the probability transition matrix resulting from year i. Our objective is to test whether the matrices $\{P_i\}_{i=1}^{5}$ are significantly difference through the test statistic 4.2.1. The calculation of the summations in equation 4.2.1 is trivial for a computer and so we instead give an example calculation of Δ which requires more caution. Table 4.13 represents the matrix P in this scenario as it was estimated using the entire five years of Atom's IRB data. In the first row we observe that there are 10 non zero entries. We then subtract one from this to enforce the stochastic matrix constraint that rows must sum to one. We repeat this for each row and sum the resulting degrees of freedom. In this case we have that $\Delta = 5 \cdot 9 + 6 \cdot 11 = 111$. We find that $\sum_{t=1}^{5} Q_t = 2441.32 \gg \chi^2_{0.95}(\Delta = 111) = 136.5911$ and therefore conclude that there is significant evidence to reject the null hypothesis of time-homogeneity within Atom's IRB data for a discrete Markov chain model at the 5% level.

10^{-2}	RG01	RG02	RG03	RG04	RG05	RG06	RG07	RG08	RG09	RG10	RG11
RG01	89.590	8.336	1.115	0.475	0.128	0.222	0.077	0.052	0.0010	0.00000	0.0030
RG02	10.823	76.749	9.349	1.824	0.451	0.325	0.299	0.176	0.0010	0.00000	0.0030
RG03	1.195	6.320	81.468	8.370	1.407	0.573	0.360	0.299	0.0050	0.00000	0.0010
RG04	0.530	1.639	10.112	77.342	7.217	1.527	0.809	0.808	0.0070	0.00000	0.0110
RG05	0.197	0.590	2.582	13.074	73.408	6.788	2.199	1.127	0.0170	0.00000	0.0180
RG06	0.348	0.566	1.130	3.495	13.805	70.410	8.389	1.813	0.0230	0.0020	0.0200
RG07	0.045	0.083	0.451	1.235	3.663	11.600	75.599	7.249	0.0370	0.0020	0.0370
RG08	0.135	0.105	0.409	0.787	1.008	1.563	10.653	84.915	0.2190	0.0060	0.2020
RG09	0.431	0.000	0.431	0.862	0.862	0.000	1.293	37.069	21.983	21.983	15.086
RG10	0.000	0.000	0.000	0.000	0.000	0.000	1.293	37.069	21.983	21.983	54.086
RG11	0.000	0.026	0.103	0.026	0.077	0.103	0.180	1.361	0.077	0.077	97.971

Table 4.13: Average one year transition probabilities, estimated using Cohort method applied to the entire 5 year duration of Atom's IRB data. $(i,j)^{th}$ entry denotes the probability of transitioning between risk grade i and j over a unit interval, scaled by 10^{-2}.

4.2.4 Conditional migration matrices

The above results confirm the existence of the time-inhomogeneity that we speculated to exist from Figures 4.2 and 4.1. This was anticipated considering that a large number of studies have concluded that default rates vary over time (e.g Nickell et al. (2000) [19]). The motivating question now is how will an institution such as Atom decide which default rate to use for forecasting future risk grade migrations? One solution to this is to simply calculate the average transition probability matrix over a large time interval. This subsection will explore an alternative solution of conditioning migration matrices based on external macroeconomic variables which yields better forecasting abilities compared to the average matrix. As a starting point, consider Baniga et al. (2002) [4] which links the changes in default rates to business cycle effects by developing a regime-switching model. In short, this study produces separate matrices for periods of economic expansion and contraction and shows that by modelling the two regimes independently an improved forecasting ability can be achieved, compared to simply averaging. Taking this idea of conditioning on external factors further, several papers have attempted to condition migration matrices on external macroeconomic variables such as GDP, consumer price indexes (CPI), and unemployment rates (e.g Wilson (1997) [30] Wei (2003) [29]). Conditioning on such variables is a technique introduced first in Wilson (1997). Kim (1999) extends the methodology by proposing the following two step method. The first step is to build a credit cycle index. This is a variable Z_t which represents the state of market at time t. It should take positive values when the given sector is performing well (e.g expansion) and negative values when the sector experiences turmoil (e.g recession). Kim (1997) suggests the following credit cycle index:

$$Z_t = \frac{\Phi^{-1}(SDP_t) - \mu_t}{\sigma_t} \tag{4.2.2}$$

Where SDP_t is the speculative grade default probability of period t and μ and σ are the mean and standard deviation of the inverse Gaussian transformation of the speculative grade default probability. Furthermore due to the restriction that SDP_t lies between zero and one, the following probit model is suggested:

$$SDP_t = \Phi(X_{t-1}\beta + \epsilon) \tag{4.2.3}$$

Where X_{t-1} is a set of macroeconomic variables chosen by the practitioner (e.g change in CPI) and ϵ_t such that $\mathbb{E}_{t-1}(\epsilon_t) = 0$. The second step is to then condition migration matrices upon this variable Z_t. We will omit demonstration of how to condition upon such a variable here as the main point of this discussion is to draw attention to the fact that the main obstacle in using conditional matrices to account for a lack of time-homogeneity is arises from the difficulty in selecting a suitable set of variables X_{t-1}. For instance, consider that an element of X_{t-1} is an indicator variable which partitions observations into periods of expansion or contraction. As suggested previously, declaring a given period to be during an expansion it is difficult task and can introduce significant bias. After considering many variables, Kim (1997) [14] finds that conditioning on quarterly GDP growth and CPI inflation yielded the largest improvement in forecasting ability compared to the average transition matrix.

To summarise, time-homogeneity is a difficult assumption to maintain for Markovian credit migration models leading to potentially incorrect inferences despite averaging over large time periods. Conditioning migration matrices on external macroeconomic variables provides an alternative to simply averaging, offering more nuanced representation of credit dynamics.

Chapter 5

Concluding Remarks and Suggestions of Further Study

5.1 Concluding Remarks and Further Study

The aim of this report was to outline the use of Markov chains in estimating probability of default. This interest was motivated by PD's central role in assessing credit risk and constructing affordability metrics. We began this report by providing an overview of the history of credit scoring. In particular, we emphasised the importance of PD in forecasting expected loss and hence its critical role in developing affordability metrics. Chapter 2 introduced the discrete time-homogenous Markov chain model as seen in Jarrow et al. (1997) [17]; commonly referred to within the literature as the JLT approach. This paper is considered by most researchers within the field as the first to establish a Markovian based model for credit migrations and inspired over a decade of research in which we explore during subsequent chapters. We concluded this chapter by discussing the limitations of the discrete model; the main observation being that the Cohort method cannot accurately capture probabilities regarding rare events. This pit-fall inspired the extension to a continuous-time model through consideration of transition intensities in a generating matrix, as suggest in Lando et al. (2002) [18] and Bangia (2002) [4]. Chapter 3 explores the continuous models suggested in these papers and paints the picture of viewing the extension to continuous time as an embedding problem. We found that the continuous Duration method provided a significant improvement in ability to capture rare events compared to the discrete counterpart, in addition to the advantage of being easily scalable to arbitrary time horizons. Chapter 3 focused on testing the validity of the assumptions made within the previous two chapters. Since time-homogeneity does not imply Markovian behaviour and vice-versa, we conducted tests for detecting such properties independently. Section 4.1.4 provided a test for the existence of non-Markovian path dependency such as rating drift. We concluded that there was sufficient evidence to reject a null hypotheses of first-order Markovian behaviour in favour of a higher order process. This conclusion aligns with the findings of Trueck et al. (2009) [25] and Bangia et al. (2002) [4], who reached similar results. Perhaps the most interesting result of this report can be found in Section 4.1.4 where we attempted to reject the Markov assumption through detection of rating drift. We set out attempting to find evidence that downgrading in the previous period increases the probability of further downgrade in the following period, as found in Bangia et al. (2002) [4]. What we found was in fact the opposite; that is, if an account was downgraded in the previous period then they are more likely to be upgraded and recover their position compared to a further downgrade. What makes this result interesting is that Trueck et al. (2009) [25] is the only paper within the literature who find identical behaviour to us. The reasoning behind this similarity becomes apparent when we analyse the datasets which were used. Trueck et al. (2009) [25] conducts their study on empirical results originally published in Krüger (2005) [15]. This paper analyses the IRB data of a German bank, similar to Atom's IRB dataset considered here. On the other hand Baniga (2002) [4], which detects rating drift, uses data taken from a major agency such as Moody's S&P. The authors of Trueck et al. (2009) [25] suggest that the difference in results arises from the fact that major agencies implicitly input "soft-factors" such as personal judgements into their models (see page 108 Chapter 6, Trueck et al. (2009) [25]). Finally we concluded the Chapter by testing the time-homogeneity assumption. We suggested the use of a Chi-square test for the detection of inhomogeneous behaviour and conclude that there is significant evidence to reject a null hypothesis of time homogeneity. This result was anticipated due to the large number of studies finding similar results (e.g Nickell et al. [19] Kim (1997)[14]). This in turn motivated the discussion of conditional migration matrices as a means to account for the time-inhomogeneity.

5.1.1 Suggestions of Further study

To conclude we will provide an brief discussion on areas of further study. The methodologies outlined within this report encapsulate the use of Markov chains in estimating PD. However, there are many alternative approaches to modelling credit migrations and estimating the probability of default. A variety of predictive models have been developed and are gaining popularity, including logistic regression, Naive Bayes, decision trees, support vector machines, and artificial neural networks (e.g see Haung (2007) [10] Chaung (2011) [7]). Atom bank model credit risk through the construction of a scorecard wherein the variables considered are chosen based on their predictive power in a logistic regression. The use of logistic regression in application to estimating probability of default has become popular in recent years due to its strong interpretability of results. This advantage becomes apparent when we consider that many machine learning models are considered "black boxes" due to their complex structure, making their evaluation processes difficult to interpret and justify to regulatory bodies. It is for this reason that we suggest the application of such methods as a topic of further study.

Chapter 6

Appendix: A

Appendix A

The following section provides some conditions for the existence of a valid generator matrix. We will present this as a problem with two parts; that is, we will begin by providing conditions for the existence of such a matrix and then secondly give a criteria for the the matrix being unique. In the case that the matrix is not unique, we will suggest a procedure for selecting which valid generator to choose. Firstly, recall that we seek a generator matrix Q such that for a transition matrix P we have that:

$$P = e^{Q} = \sum_{k=0}^{\infty} \frac{Q^k}{k!} = I + Q + \frac{Q^2}{2!} + \frac{Q^3}{3!} + \cdots \tag{6.0.1}$$

We begin by addressing the problem of existence. Israel et al.(2000) [12] provides the following proposition regarding the non-existence of a generator matrix:

Proposition 1 *Let P be a transition matrix. If one of the following conditions is satisfied then there does not exist a valid generator Q such that $P = e^{Q}$:*

- $\det(P) \leq 0$
- $\det(P) > \prod_i p_{ii}$
- *There are states i and j such that j is accessible from i, but $p_{ij} = 0$*

Singer and Spilerman (1976) [24] provide a stronger result regarding the existence and uniqueness:

Proposition 2 *Let P be a transition matrix with real, distinct eigenvalues.*

- *If all eigenvalues of P are positive, then the matrix obtained by 6.0.1 is the only real matrix Λ such that $P = e^{Q}$.*
- *If P has any negative eigenvalues, then there exists no real matrix Λ such that $\exp(\Lambda) = P$.*

Secondly, we provide a means for determining if there is more than one generating matrix Q such that equation 6.0.1 holds. Trueck et al. (2009) [25] provides the following result for confirming the uniqueness of such a matrix:

Theorem 6.1 (5.5) *Let P be a transition matrix.*

- *If* $\det(P) > 0.5$*, then P has at most one generator.*
- *If* $\det(P) > 0.5$ *and* $\|P - J\| < 0.5$*, then the only possible generator for P is the one obtained by (6.0.1).*
- *If P has distinct eigenvalues and* $\det(P) > e^{-\pi}$*, then the only possible generator is the one obtained by (6.0.1).*

In the case that there exists more than one valid generator, how shall we choose which one to model credit migrations with? Israel et al. (2002) propose that the choice should be made such that it minimises the following:

$$J = \sum_{i,j} |j - i| \, |\lambda_{ij}| \tag{6.0.2}$$

The rational behind this suggestion is that a transition to a risk grade which is "far away" from the current grade is less likely than a transition to a risk grade which is "close" to the current grade. Choosing a Q such that it minimises J in equation 6.0.2 ensures that this phenomenon is preserved as transition intensities are weighted according to how "far away" they are by $|j - i|$.

Bibliography

[1] Edward Altman and David Kao. "The implications of corporate bond rating drift". In: *Financial Analysts Journal* 48.3 (1992). a, pp. 64–75.

[2] T. W. Anderson and Leo A. Goodman. "Statistical Inference About Markov Chains". In: *Annals of Mathematical Statistics* 28.1 (1957), pp. 89–110.

[3] James S. Angell, Robert E. Litan, and Richard Herring. *The Failure of the Franklin National Bank: Challenge to the International Banking System*. New Haven, CT: Yale University Press, 1981. ISBN: 978-1893122346.

[4] A. Bangia et al. "Ratings migration and the business cycle, with application to credit portfolio stress testing". In: *Journal of Banking and Finance* 26 (2002), pp. 445–474.

[5] Patrick Billingsley. "Statistical Methods in Markov Chains". In: *Annals of Mathematical Statistics* 32.1 (1961), pp. 12–40.

[6] Jens Christensen, Esben Hansen, and David Lando. "Confidence Sets for Continuous Time Rating Transition Probabilities". In: *Journal of Banking and Finance* 28 (2004), pp. 2575–2602.

[7] Chia-Ling Chuang and Shih-Tang Huang. "A Hybrid Neural Network Approach for Credit Scoring". In: *Expert Systems* 28.2 (2011), pp. 185–196. DOI: `10.1111/j.1468-0394.2010.00565.x`. URL: `https://doi.org/10.1111/j.1468-0394.2010.00565.x`.

[8] L. Goodman. "Simplified runs tests and likelihood ratio tests for Markov chains". In: *Biometrika* 45.1 (1958), pp. 181–197.

[9] HM Treasury. *Help with mortgages to continue for homeowners affected by Coronavirus*. [Accessed: 01-May-2025]. 2020. URL: `https://www.gov.uk/government/news/help-with-mortgages-to-continue-for-homeowners-affected-by-coronavirus`.

[10] Chih-Long Huang, Mu-Chen Chen, and Chieh-Jen Wang. "Credit Scoring with a Data Mining Approach Based on Support Vector Machines". In: *Expert Systems with Applications* 33.4 (2007), pp. 847–856. DOI: `10.1016/j.eswa.2006.07.007`. URL: `https://doi.org/10.1016/j.eswa.2006.07.007`.

[11] IFRS Foundation. *IFRS 9 Financial Instruments*. Accessed: 2025-03-11. 2025. URL: `https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/`.

[12] Robert Israel, Jeffrey S. Rosenthal, and Jason Z. Wei. "Finding Generators for Markov Chains via Empirical Transition Matrices, with Application to Credit Ratings". In: *Mathematical Finance* 11 (2000), pp. 245–265.

[13] Yaser Jafry and Til Schuermann. "Measurement, Estimation and Comparison of Credit Migration Matrices". In: *Journal of Banking and Finance* 28 (2004), pp. 2603–2639.

[14] Jin-Chuan Kim. "Conditioning the Transition Matrix". In: *Risk Credit Risk Special Report* (1999), pp. 37–40.

[15] U. Krüger, M. Stötzel, and S. Trueck. "Time series properties of a rating system based on financial ratios". In: *Deutsche Bundesbank Discussion Paper Series: Banking and Financial Studies* 14 (2005), pp. 1–60.

[16] U. Küchler and M. Sorensen. *Exponential Families of Stochastic Processes*. New York: Springer, 1997.

[17] David Lando. "A Markov Model for the Term Structure of Credit Risk Spreads". In: *Journal of Financial Economics* 47.1 (1998), pp. 115–148. DOI: 10.1016/S0304-405X(97)00039-5. URL: https://mx.nthu.edu.tw/~jtyang/Teaching/Risk_management/Papers/Models/A%20Markov%20Model%20for%20the%20Term%20Structure%20of%20Credit%20Risk%20Spreads.pdf.

[18] David Lando and Thomas Skødeberg. "Analyzing rating transitions and rating drift with continuous observations". In: *Journal of Banking and Finance* 26.2 (2002), pp. 423–444.

[19] Patrick Nickell and Stefano Varotto. "Stability of Rating Transitions". In: *Journal of Banking and Finance* 1–2 (2000), pp. 203–227.

[20] J.R. Norris. *Markov Chains*. Cambridge: Cambridge University Press, 1997.

[21] Michael K. Ong. *The Basel Handbook: A Guide for Financial Practitioners*. London, UK: Risk Books, 2007. ISBN: 978-1904339395.

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. Accessed: 2025-04-30. R Foundation for Statistical Computing. 2024. URL: https://www.r-project.org/.

[23] Sheldon M. Ross. *Introduction to Probability Models*. 12th ed. Academic Press, 2019. ISBN: 9780128143469.

[24] Burton Singer and Seymour Spilerman. "The Representation of Social Processes by Markov Models. Fonctions de Réparation à n Dimensions et Leurs Marges". In: *American Journal of Sociology* 82 (1976), pp. 1–54.

[25] S. Trueck et al. *Ratings-Based Credit Migrations*. English. Amsterdam: Academic Press, 2009. ISBN: 978-0-12-373683-3. DOI: https://doi.org/10.1016/B978-0-12-373683-3.X0001-2.

[26] Theo Van Gestel and Bart Baesens. *Credit Risk Management: Basic Concepts*. Oxford: Oxford University Press, 2008.

[27] Dennis D. Wackerly, William Mendenhall, and Richard L. Scheaffer. *Mathematical Statistics with Applications*. 7th. Belmont, CA: Brooks/Cole, 2008, Chapter 9, Section 9.6 (Likelihood Ratio Tests).

[28] Y. Wang et al. "There Can Be Diverse Effects of Using Synthetic Data on the Performance of Statistical Analyses: An Empirical Study". In: *BMC Medical Research Methodology* 24.1 (2024), p. 73. DOI: 10.1186/s12874-024-02173-x. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-024-02173-x.

[29] Jian Wei. "A Multi-Factor, Credit Migration Model for Sovereign and Corporate Debts". In: *Journal of International Money and Finance* 22 (2003), pp. 709–735.

[30] Thomas C. Wilson. "Portfolio Credit Risk". In: *Risk Magazine* 10 (1997).