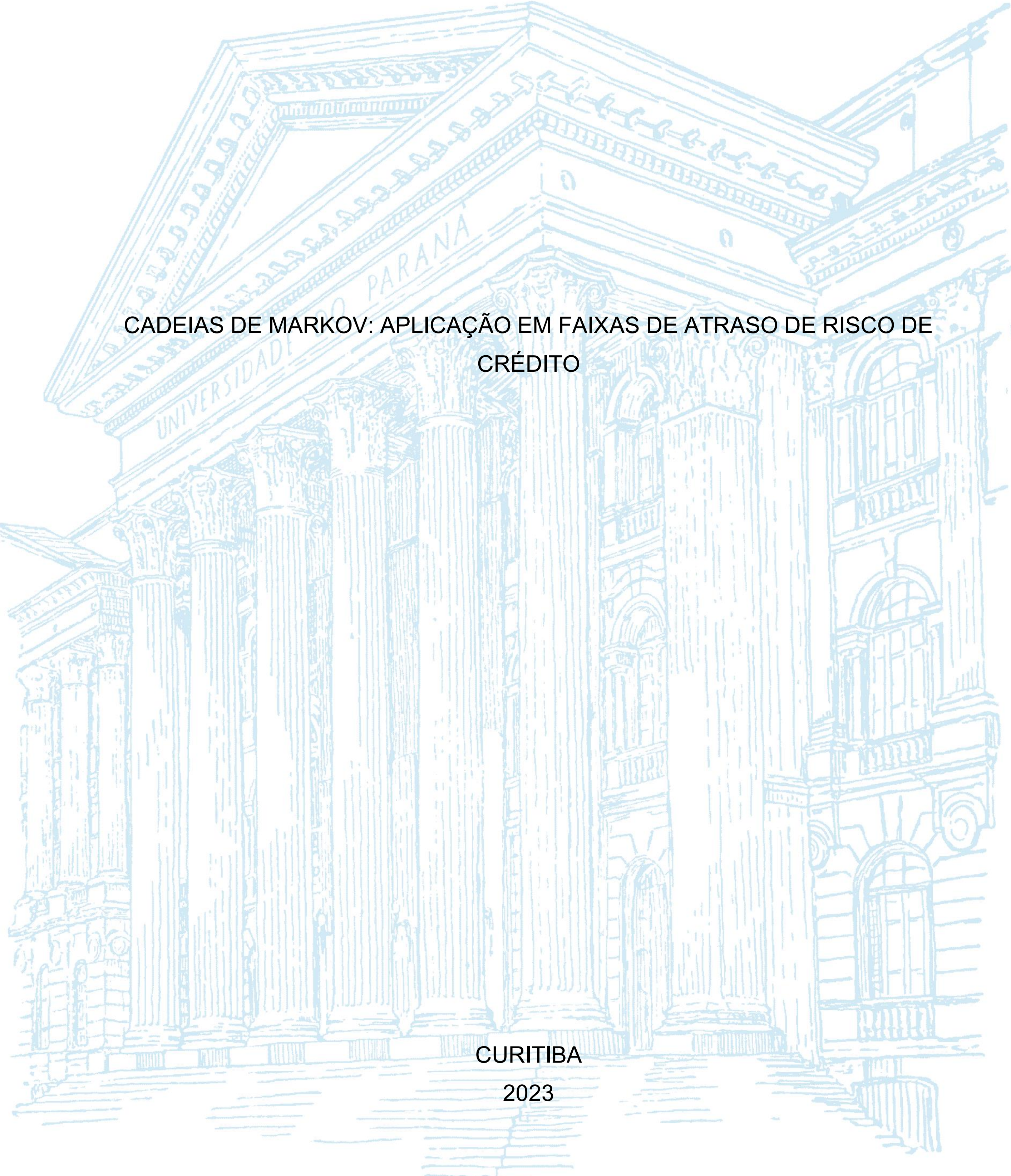
UNIVERSIDADE FEDERAL DO PARANÁ

EMERSON BERLANDA



CADEIAS DE MARKOV: APLICAÇÃO EM FAIXAS DE ATRASO DE RISCO DE CRÉDITO

CURITIBA

2023

EMERSON BERLANDA

CADEIAS DE MARKOV: APLICAÇÃO EM FAIXAS DE ATRASO DE RISCO DE CRÉDITO

Monografia apresentada ao curso de Graduação em Ciências Econômicas, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Economia.

Orientador: Prof. Dr. José Guilherme Silva Vieira

CURITIBA

2023

TERMO DE APROVAÇÃO

EMERSON BERLANDA

CADEIAS DE MARKOV: APLICAÇÃO EM FAIXAS DE ATRASO DE RISCO DE CRÉDITO

Monografia apresentada ao curso de Graduação em Ciências Econômicas, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Economia.

Orientador - Prof. Dr. José Guilherme Silva Vieira
Departamento de Economia, UFPR

Prof. Dr. Fabiano Abranches Silva Dalto
Departamento de Economia, UFPR

Prof. Me. Françoise Iatski de Lima
Departamento de Economia, UFPR

Curitiba 28 de Novembro de 2023.

RESUMO

O presente trabalho faz uma aplicação de Cadeias de Markov discretas para estimar as matrizes de transição entre as faixas de atraso de uma base de dados de uma carteira de crédito de uma instituição financeira brasileira (anônima) a fim de melhor entender o comportamento histórico da carteira em questão. Uma matriz de transição é uma matriz com as probabilidades de cada uma das faixas de atraso rolar para uma outra faixa de atraso de um mês para o outro e como esta matriz pode ser utilizada para aprofundar as técnicas de análise de Risco de Crédito de uma instituição financeira, bem como indicar as possibilidades de estudos que podem ser desenvolvidos com a obtenção desses dados. Os resultados obtidos demonstram uma melhoria na interpretação das rolagens de crédito entre faixas de atraso.

Palavras-chave: Cadeias de Markov, Matriz de Transição, Análise de Risco de Crédito, Faixas de atraso.

ABSTRACT

The present work makes na application of discrete-time Markov Chains to estimate transition matrices between delinquency rates of a data base obtained of a loan portfolio of a real financial institution (remained anonymous) so that we can better understand the historical behaviour of the said portfolio. A transition matrix is a matrix with the probabilities of each delinquency rate migrating to another delinquency rate inbetween months, and how that matrix can be used to deepen the Credit Risk Analysis of a financial institution, aswell as indicate further study possibilities that can be developed with the new obtained data. The results found demonstrate a better understanding of roll rates between delinquency buckets.

Key words: Markov Chains, Transition Matrix, Credit Risk Analysis, Delinquency Rates.

1 INTRODUÇÃO

A Análise de Risco de Crédito é fundamental para a gestão financeira, especialmente em instituições que concedem empréstimos e detêm carteiras de crédito. A capacidade de prever e compreender o comportamento dos devedores é crucial para a tomada de decisões e a mitigação de riscos financeiros.

Esta monografia explora a aplicação das Cadeias de Markov na análise de risco de crédito, com foco em uma abordagem detalhada das transições entre faixas de atraso dos devedores. O objetivo é fornecer uma compreensão aprofundada do comportamento dos devedores e das probabilidades associadas a essas transições, abrindo caminho para uma gestão de risco de carteira mais eficaz.

Ao longo deste trabalho, serão examinados os resultados obtidos por meio da aplicação das Cadeias de Markov em uma carteira de crédito. Esses resultados incluem a análise das probabilidades de transição entre faixas de atraso e suas possíveis aplicações e derivações. Esses elementos demonstram como as probabilidades de transição entre as faixas de atraso se estabilizam ao longo do tempo, bem como o equilíbrio de longo prazo da carteira de crédito e fornece uma base sólida para aprimorar a gestão de risco financeiro, com implicações que vão além do campo financeiro e se estendem a diversas áreas de tomada de decisões informadas.

As probabilidades de transição também podem demonstrar a qualidade de uma safra, isto é, créditos gerados dentro de um período específico de tempo, podendo também ser comparado com outras safras, com outras políticas de crédito ou em cenários econômicos diferentes e averiguar quais são as mais adequadas e/ou mais eficazes, melhorando assim a tomada de decisão futura e a qualidade da carteira de crédito como um todo.

2 REFERENCIAL TEÓRICO E METODOLÓGICO

As Cadeias de Markov, nomeadas em homenagem ao matemático russo Andrei Markov, têm se destacado como uma poderosa ferramenta na modelagem de sistemas dinâmicos e na análise de processos estocásticos. A teoria das Cadeias de Markov, frequentemente referida como uma área fundamental na probabilidade e na teoria dos processos estocásticos, tem uma ampla gama de aplicações em campos diversos como finanças, engenharia, biologia, e muito mais. No geral, Cadeias de Markov são separadas entre Discretas e Contínuas. Nos limitaremos às Discretas por conta do tipo de dados que será trabalhado posteriormente.

2.1 CADEIA DE MARKOV DISCRETA

Cadeias de Markov, são um processo fundamental nos processos estocásticos, isto é, uma sequência de observações de um fenômeno em função do tempo, que possuem uma propriedade de que o estado (ou classe, para referenciar métodos de classificação) futuro depende apenas do estado-presente, e não de estados passados, ou, conforme (SIMON, 2004) "[...] se a probabilidade com que o sistema está no estado i no período $n+1$ depende somente do estado em que o sistema esteve no período n, em outras palavras, a mudança entre estados é constante no tempo. Matematicamente podemos dizer que

$$P(X_{n+1} = x | X_1 = x_1 + X_2 = x_2 + \cdots + X_n = x_n) = P(X_{n+1} | X_n = x_n)$$

onde x é algum estado do processo. Essa identidade é conhecida como o Teorema de Propriedade de Markov (SERICOLA, 2013) e, ainda segundo SERICOLA (2013, p.13), "uma cadeia de Markov é dita irredutível se todos os seus estados se comunicam, isto é, se possuem uma única classe equivalente ou se seu gráfico admite um único componente conectado", que determina que os estados de transição e de recorrência são propriedades de classe e que se pode partir de qualquer estado para chegar em qualquer outro estado.

Desta maneira, podemos descrever uma matriz de transição como uma matriz de probabilidades de estados de uma determinada observação, onde a matriz

Imagem 1 – Matriz de Transição

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots & t_{1m} \\ t_{21} & t_{22} & t_{23} & \cdots & t_{2m} \\ t_{31} & t_{32} & t_{33} & \cdots & t_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & t_{n3} & t_{n4} & t_{nm} \end{bmatrix}$$

Fonte: Elaboração Própria.

em que t_{ij} é a probabilidade de mudança do estado j para o estado i em uma unidade de tempo. Ou seja, aproximadamente, temos que probabilidade histórica de um estado de um certo tipo de dados é dada pela razão entre quantidade de aparições naquele estado pelo número de observações total, enquanto a matriz

Imagem 2 - Distribuição Inicial

$$P_0 = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_n \end{bmatrix}$$

Fonte: Elaboração própria.

é a distribuição inicial entre os estados no tempo 0. E, portanto, a matriz de estado P, após uma unidade de tempo, se torna o produto das matrizes $P = TP_0$ onde, após k unidades de tempo, se torna $P_k = TP_{(k-1)} = T^2P_{(k-2)} = \cdots = T^kP_0$, que nada mais é do que uma matriz diagonalizável e, dado que se trata de uma cadeia irredutível, contínua no tempo, com estados recorrentes positivos e também sabemos que a soma das probabilidades é igual a 1, isso significa que existe apenas um único vetor de probabilidades chamado de distribuição estacionária, ou de estado estável (SERICOLA, 2013). Em outras palavras a distribuição das probabilidades de transição tende a uma distribuição única de probabilidades permanentes.

Portanto, podemos usar duas maneiras de chegar ao estado estacionário desta matriz: calculá-la com o maior número de observações possíveis (aproximação), ou utilizando autovetores (SANTOS, 2003).

2.2 EM MERCADOS FINANCEIROS

Cadeias de Markov são frequentemente utilizadas para modelagem de mercados financeiros devido a sua relativa simplicidade de estabelecimento inicial, e também muitas vezes são utilizados em comparação com modelos de regressão dentro deste meio, como por exemplo uma aplicação em investimentos na IBOVESPA (SCARIOTE; CORSO, 2021) utilizando o dólar como influenciador dentro dos cenários de incerteza do mercado financeiro, bem como para a previsão de oscilações no índice IBOVESPA, NASDAQ e NYSE (SALVADOR; CORSO, 2022) através da transformação das variáveis numéricas dos índices em variáveis de classe, estabelecendo assim um conjunto de variáveis no tempo que se encaixam nos requisitos de classificação do processo de Markov e, portanto, definindo-os como processos finitos e de estados contáveis, tais dados podem ser caracterizados como uma cadeia de Markov.

2.2.1 Simulações de Monte Carlo

Outra grande aplicação de Cadeias de Markov são em modelos de Monte Carlo. Uma simulação de Monte Carlo pode ser compreendida como uma simulação de amostragem randomizada (Sobol, 1975) e, para se obter uma amostragem estatisticamente significativa, ela depende primariamente de dois fatores: viés e tamanho.

O viés é um fator importante porque ele pode fazer com que a sua amostragem não represente de forma efetiva a sua população. Qualquer viés na geração de números aleatórios ou na estrutura do modelo pode levar a resultados distorcidos e imprecisos. Portanto, é fundamental minimizar ou, preferencialmente, eliminar o viés na simulação de Monte Carlo para obter estimativas confiáveis.

O tamanho da amostragem também desempenha um papel crítico na precisão dos resultados da simulação de Monte Carlo. Conforme estabelecido na Lei dos Números Grandes (Chibisov, 2016), à medida que o tamanho da amostragem aumenta, a média das estimativas tende a convergir para o valor esperado p. Isso significa que, para obter resultados mais confiáveis e estáveis, é necessário realizar simulações com um grande número de amostras. Um tamanho de amostragem inadequado pode levar a estimativas imprecisas e não confiáveis.

Portanto, a combinação de uma amostragem livre de viés e um tamanho de amostragem adequado são fundamentais para garantir que uma simulação de Monte Carlo forneça resultados estatisticamente significativos e úteis para a análise de sistemas complexos e estocásticos.

2.2.2 Métodos de Monte Carlo baseados em Cadeias de Markov

Métodos de Monte Carlo baseados em cadeias de Markov são frequentemente chamados de métodos MCMC, sendo a sigla "MCMC" uma abreviação de "Markov Chain Monte Carlo". Se Monte Carlo é uma técnica de amostragem e Cadeias de Markov são uma sequência de observações no tempo sem memória de seu passado, então o MCMC é um método que envolve a geração de sequências de amostrar que formam uma Cadeia de Markov, onde cada amostra obtida é condicionada à amostra anterior, de modo que sua distribuição estacionária π seja garantida em uma cadeia de tempo n (Silva, 2017).

Conforme demonstrado anteriormente, uma distribuição são probabilidades de transição quando o período de tempo n tende ao infinito, o que seria aproximadamente o cálculo de probabilidades de transição de uma base histórica muito grande e, portanto, podemos utilizar MCMC quando os processos se tornam impossíveis, ou muito lentos, devido a demanda computacional necessária, então através do MCMC podemos obter uma amostra cujo resultado final π seja igual ao de uma distribuição de uma população de dados.

> Em particular, o MCMC é muito útil para a estimativa de modelos não lineares com integrais de alta dimensão na verossimilhança (como modelos com muitas variáveis latentes) ou uma estrutura hierárquica. Isso inclui, mas não se limita a, modelos de escolha discreta, modelos de correspondência e outros modelos de auto seleção, modelos de duração, dados em painel e modelos estruturais, abrangendo uma ampla gama de tópicos em finanças corporativas, como estrutura de capital, emissão de títulos, intermediação financeira, governança corporativa, falência e modelos estruturais da empresa.[1] (Korteweg, 2011, p.2, tradução própria)

---

1 In particular, MCMC is very useful for estimating non-linear models with high-dimensional integrals in the likelihood (such as models with many latent variables), or a hierarchical structure. This includes, but is not limited to, discrete-choice, matching and other self-selection models, duration, panel data and structural models, encompassing a large collection of topics in corporate finance such as capital

2.3 RISCO DE CRÉDITO E GERENCIAMENTO DE PORTFÓLIO

Segundo Siu (Siu, et al., 2005), há 2 métodos que eram predominantemente utilizados dentro das modelagens de risco de crédito: de cópulas e de Monte Carlo, porém modelos de Markov multivariados demonstram uma maneira natural e conveniente para descrever as dependências entres os *ratings* de ativos individuais dentro de um portfólio. Essas dependências entre os ratings então podem ser utilizadas em modelos de precificação e cobertura de dívida e também em cobertura de derivativos de crédito, conforme demonstrado por Jarrow, Lando e Turnbull (1977).

2.4 METODOLOGIA

Para fazer uma aplicação de cadeias de Markov em faixas de atraso de uma carteira de crédito, faz-se necessário uma base de dados que contenham as faixas de atraso necessárias. Uma das dificuldades de se trabalhar com bases públicas do Sistema Financeiro Nacional (SFN) é a de que não há informações detalhadas o suficiente (como a de acompanhamento de um cliente ao longo do tempo) tampouco a de uma abertura maior do que a quantidade em atraso acima de 90 dias (em inadimplência).

2.5 BASE DE DADOS E AMOSTRAGEM

A base de dados utilizada neste trabalho teve medidas tomadas para garantir que a privacidade e a confidencialidade das informações contidas nas observações fossem mantidas. As informações obtidas contém as faixas de atraso de 500 mil clientes de uma carteira de crédito não definida entre os anos de 2016 até 2022, selecionados por amostragem aleatória simples, que contém seu contrato, que estão mascarados sob uma coluna chamada de "ID", a data de origem do contrato do cliente sob uma coluna chamada de "DT_BASE" e a evolução da faixa de atraso do contrato deste cliente desde o seu primeiro mês (FAIXA_ATRASO_M0) ao longo de 83 meses

---

structure and security issuance, financial intermediation, corporate governance, bankruptcy, and structural models of the firm.

(FAIXA_ATRASO_M83) que é o número máximo de meses que um dos clientes da base esteve na carteira até a conclusão de seu contrato. Dentro das colunas de faixas de atraso, foram agrupadas todas as possíveis faixas de atraso em 4:

TABELA 1 – TRATAMENTO DE FAIXAS DE ATRASO

Faixa Original	Faixa Tratada
000 - 000	0
001 - 030	1
031 - 060	2
061 - 090	3
>90	4

FONTE: Elaboração própria.

Estes dados foram coletados de uma instituição financeira brasileira como parte de suas atividades comerciais.

2.6 TRATAMENTOS ADICIONAIS E BASE FINAL

Além dos tratamentos já mencionados anteriormente, a base de dados também passou por um tratamento de transposição de colunas que estará definido no anexo técnico A.

Desta maneira, a base final está composta por 3 colunas: ID, DT_BASE e FAIXA_ATRASO. Onde cada observação repetida do mesmo ID e DT_BASE demonstra a evolução em meses desde o estabelecimento de seu contrato. Por exemplo, dadas as seguintes informações:

TABELA 2 – EXEMPLO DA BASE DE DADOS

n	ID	DT_BASE	FAIXA_ATRASO
0	254	31/01/2016	0
1	254	31/01/2016	0
2	254	31/01/2016	0
3	254	31/01/2016	1
4	254	31/01/2016	1
5	254	31/01/2016	0

FONTE: Elaboração própria.

Portanto, quando $n = 4$ o contrato 254 estava na faixa de atraso 1 na data de maio/2016, dado que sua originação é em janeiro/2016. Desde abril/2016 ele se encontra na faixa de atraso 1 e em junho/2016 ele voltou para a faixa 0 .

2.7 ALGORITMO

O código desenvolvido para a obtenção da matriz de transição, diferentemente de uma matriz de transição normal, conforme descrita no capítulo de revisão bibliográfica, aqui estamos interessados na quantidade de vezes que um contrato passou de um estado para o outro, ou continuou no mesmo estado, e não quantas vezes aquele estado apareceu na base de dados. Isto é, estamos contando a transição entre os meses, portanto a contagem de transição de um ID para outro não é contabilizada. Utilizando a tabela 2, temos apenas uma transição da faixa 0 para a faixa 1 , duas transições da faixa 0 para a faixa 0 (a originação não é contada), uma transição da faixa 1 para a faixa 1 e, finalmente, uma transição da faixa 1 para a faixa 0 isto é:

TABELA 3 – Exemplo de cálculo de probabilidades de transição I.

FAIXA_ATRASO	**0**	**1**
0	2	1
1	1	1

FONTE: Elaboração própria.

Portanto a probabilidade de transição das faixas que partem do estado 0 é dada pela razão da quantidade de vezes que uma faixa em estado 0 “rolou” para outros estados ou continuou no mesmo estado e pela quantidade total de rolagens que partiram do estado 0 . Neste mesmo exemplo temos que:

TABELA 4 – Exemplo de cálculo de probabilidades de transição II.

FAIXA_ATRASO	**0**	**1**
0	$\frac{2}{3} = 67\%$	$\frac{1}{3} = 33\%$
1	$\frac{1}{2} = 50\%$	$\frac{1}{2} = 50\%$

FONTE: Elaboração própria.

Importante ressaltar que a soma das probabilidades das transições que partem da mesma faixa de atraso devem ser 100% e que as contagens são agrupadas por ID, portanto a transição final de um contrato para a inicial de um contrato não é contabilizada. Mais informações sobre o código desenvolvido podem ser obtidas pelo anexo técnico B.

3 RESULTADOS

Nesta seção serão apresentados os resultados obtidos através da aplicação de cadeias de Markov em faixas de atraso, bem como suas possíveis aplicações dentro de uma carteira de crédito.

3.1 MATRIZ DE TRANSIÇÃO E ANÁLISE

Os resultados finais do algoritmo desenvolvido aplicado a esta base foi a seguinte:

TABELA 5 – Resultados

n	DE_FAIXA_ATRASO	PARA_FAIXA_ATRASO	CONT_TRANSICAO	TRANSICAO_TOTAL	PROB
0	0	0	40.415.232	40.577.950	99,60%
1	0	1	159.797	40.577.950	0,39%
2	0	2	2.874	40.577.950	0,01%
3	0	3	14	40.577.950	0,00%
4	0	4	33	40.577.950	0,00%
5	1	0	98.892	222.531	44,44%
6	1	1	55.097	222.531	24,76%
7	1	2	67.515	222.531	30,34%
8	1	3	1.008	222.531	0,45%
9	1	4	19	222.531	0,01%
10	2	0	16.925	77.893	21,73%
11	2	1	4.231	77.893	5,43%
12	2	2	6.465	77.893	8,30%
13	2	3	49.314	77.893	63,31%
14	2	4	958	77.893	1,23%
15	3	0	8.979	53.377	16,82%
16	3	1	788	53.377	1,48%
17	3	2	845	53.377	1,58%
18	3	3	2.630	53.377	4,93%
19	3	4	40.135	53.377	75,19%
20	4	0	40.039	568.249	7,05%
21	4	1	396	568.249	0,07%
22	4	2	194	568.249	0,03%
23	4	3	411	568.249	0,07%
24	4	4	527.209	568.249	92,78%

FONTE: Elaboração própria.

Onde a coluna DE_FAIXA_ATRASO representa a origem, PARA_FAIXA_ATRASO representa a chegada, CONT_TRANSICAO é a quantidade de vezes que a transição daquela origem para aquela chegada ocorreu, a TRANSICAO_TOTAL é a quantidade total de transições com a mesma origem e a PROB é a razão entre CONT_TRANSICAO e a TRANSICAO_TOTAL, contendo o resultado da probabilidade obtida.

Reconvertendo os valores das faixas para os valores originais e transpondo a tabela para melhor visualização, temos a seguinte tabela:

TABELA 6 – Matriz de Transição

FAIXA ATRASO	000-000	001-030	031-060	061-090	>90
000-000	99,60%	0,39%	0,01%	0,00%	0,00%
001-030	44,44%	24,76%	30,34%	0,45%	0,01%
031-060	21,73%	5,43%	8,30%	63,31%	1,23%
061-090	16,82%	1,48%	1,58%	4,93%	75,19%
>90	7,05%	0,07%	0,03%	0,07%	92,78%

FONTE: Elaboração própria.

Nesta tabela temos as probabilidades de transição históricas de 500 mil contratos de uma carteira de crédito, com todas as transições possíveis entre elas e cada número nas faixas de atraso representam a quantidade de dias em atraso. É preciso destacar que as probabilidades de um contrato sair da faixa 000-000 dias para a faixa 061-090 dias ou da 000-000 dias para a >90 dias de um mês para o outro não são absolutamente 0. Não é possível identificar ao certo o porquê esses contratos tiveram esses saltos nas faixas de atraso por uma limitação de base, mas de qualquer maneira, eles são estatisticamente irrelevantes. O mesmo pode ser atribuído para todos os saltos maiores de 1 mês de intervalo, obviamente com exceção dos contratos que voltam em faixas de atraso, dado que é possível o pagamento de mais de um mês de parcelas em atraso em um intervalo de um mês.

Ademais, observamos uma concentração maior nas probabilidades de transição entre as faixas subsequentes, como era esperado, com exceção das faixas de transição de 000-000 dias para 000-000 dias e de >90 dias para >90 dias. É possível argumentar que as atribuições de riscos desta carteira específica têm sido eficazes e que a qualidade do crédito desta carteira é alta, dado que a probabilidade de um cliente entrar em faixas de atraso é baixa, bem como a recuperação (voltar uma faixa de atraso) de um cliente que está na faixa de 001-030 dias é mais alta do que a probabilidade deste cliente avançar para a faixa de atraso subsequente. No caso das faixas subsequentes é possível observar uma notável perda do controle das recuperações desses clientes, pois eles passam a avançar nas faixas de atraso com altas probabilidades, sobretudo quando chegam a um estado de inadimplência (>90 dias), dado que a probabilidade de um cliente continuar inadimplente é de $92{,}78\%$.

3.1.1 Safras

Adicionalmente, fazendo um pequeno filtro de datas no algoritmo, é possível fazer uma análise de safras com a base utilizada.

Considerando a safra do ano de 2016, isto é, contratos (ID) que foram gerados em todos os meses do ano de 2016 (DT_BASE = 2016) obtêm-se a seguinte matriz de transição:

TABELA 7 – Matriz de Transição da Safra de 2016

FAIXA ATRASO	000-000	001-030	031-060	061-090	>90
000-000	99,64%	0,36%	0,01%	0,00%	0,00%
001-030	45,41%	27,02%	27,17%	0,38%	0,02%
031-060	20,51%	6,46%	10,16%	61,84%	1,03%
061-090	15,12%	1,90%	1,83%	5,56%	75,59%
>90	5,73%	0,07%	0,03%	0,07%	94,10%

FONTE: Elaboração própria.

Comparando com a matriz de transição da safra de 2021:

TABELA 8 – Matriz de Transição da Safra de 2021

FAIXA ATRASO	000-000	001-030	031-060	061-090	>90
000-000	99,56%	0,43%	0,01%	0,00%	0,00%
001-030	40,66%	22,71%	35,97%	0,65%	0,01%
031-060	22,25%	3,83%	5,92%	66,50%	1,50%
061-090	18,33%	1,22%	1,04%	3,51%	75,89%
>90	14,09%	0,11%	0,05%	0,10%	85,65%

FONTE: Elaboração própria.

É possível observar que a maior diferença entre as safras é na recuperação do crédito em inadimplência (de >90 dias para 000-000 dias), que é significativamente maior na safra de 2021. A isso é possível atribuir uma melhor política de recuperação de crédito sendo utilizada em 2021, ou, por se tratar do período acometido pela pandemia de COVID-19, essa recuperação de crédito pode ser por conta das políticas de prorrogação e melhores condição de renegociação de dívida tomadas durante o período de pandemia (Pupo, 2022).

3.2 DISCUSSÃO

Uma das maneiras frequentemente utilizadas de se analisar o risco de crédito em rolagem de uma carteira de crédito é através do cálculo das rolagens (*roll rate*). Matematicamente podemos dizer que

$$Rolagem_{(i,j)} = \frac{FaixaAtraso_{(i,j)}}{FaixaAtraso_{(i-1,j-1)}},\ i = (1, \dots, n), j = (1, \dots, n)$$

onde i são faixas de atraso e j são meses.

Desta maneira podemos obter aproximadamente quantidade de ativos que rolaram de uma faixa para a outra no mês subsequente. Uma média móvel de, por exemplo, 12 meses do saldo dessas rolagens nos permite fazer uma projeção de saldos em atraso que, apesar de simples, já é capaz de auxiliar na tomada de decisões futuras. É dita simples pois esta é uma metodologia com diversas limitações de interpretação, mas a mais aparente, e talvez mais importante, é a de que nesta metodologia um ativo só pode rolar adiante (para a faixa subsequente no mês subsequente).

No entanto, como é possível observar com os resultados das matrizes de transição, há muitas rolagens que permanecem numa mesma faixa de atraso, significando que o cliente pode ter pagado o novo vencimento e ficando assim na mesma faixa de atraso, bem como há também a possibilidade de o cliente pagar uma ou mais parcelas em atraso, fazendo com que ele salte de faixas de atraso mais avançadas para a faixa de 000-000 dias. Em outras palavras, a aplicação de cadeias de Markov em faixas de atraso permite observar não só o saldo que rola adiante, mas também o saldo que continua na mesma faixa e o saldo que volta para as faixas anteriores também.

Conforme estabelecido pelo Teorema da Convergência ao Equilíbrio, em que uma cadeia de Markov irredutível e de tempo homogêneo, sabemos que o vetor π de probabilidades é uma distribuição estacionária, isto é, converge para uma probabilidade única. Isso é facilmente observável com uma simulação simples com os mesmos dados utilizados neste trabalho. Ao cortarmos o número de linhas pela metade, sem nem precisar fazer uma outra amostragem aleatória simples (mais informações no anexo C), e subtrairmos pelos valores da tabela 6, obtemos diferenças percentuais muito baixas, conforme demonstrado abaixo:

TABELA 9 - Diferença entre as matrizes de transição

FAIXA ATRASO	000-000	001-030	031-060	061-090	>90
000-000	0,02%	-0,02%	0,00%	0,00%	0,00%
001-030	0,93%	-1,90%	0,58%	0,38%	0,01%
031-060	1,03%	0,02%	-4,66%	2,38%	1,23%
061-090	1,71%	0,31%	-2,20%	-1,47%	1,65%
>90	1,97%	0,03%	-0,01%	-0,05%	-1,94%

FONTE: Elaboração própria.

Desta maneira temos um modelo, ainda simples, porém muito mais pertinente que o cálculo de rolagem simples, cujos valores podem ser utilizados como base para os mais variados modelos de projeção de indicadores de risco de crédito, como de inadimplência, modelos de Perda Esperada (IFRS9), testes de estresse e afins.

3.2.1 Projeções

Uma outra interessante possível aplicação de cadeias de Markov seria a utilização das probabilidades como uma variável explicativa em modelos lineares de regressão. No geral, para esse tipo de dados, não estamos interessados na distribuição estável da matriz para projeções pois sabemos que o futuro não pode ser previso apenas com dados históricos. O mercado é um conjunto de crenças determinados tanto por movimentações históricas como pelas movimentações de variáveis macroeconômicas que, por sua vez, são influenciadas por uma infinitude de questões econômicas, sociais, poder, políticas e etc.

Desta maneira, poderíamos reduzir nossas probabilidades históricas utilizadas nesse trabalho e transformá-los em probabilidades de transição mensais e obtendo assim uma variável que pode ser utilizada em séries temporais. Assim é possível fazer testes de correlação com variáveis macroeconômicas, identificar sazonalidades, correlações e etc., a fim de identificar comportamentos e variáveis que possam explicar as oscilações nas probabilidades de correlação e, a partir disso, fazer projeções do comportamento desta carteira de crédito no futuro.

Uma outra possibilidade seria a utilização de Cadeias de Markov multivariadas para este tipo de trabalho, pois espera-se que haja uma interdependência entre as rolagens de uma carteira de crédito e as variáveis macroeconômicas.

3.2.2 Gerenciamento de Portfólio

Outra aplicação dos resultados obtidos seria uma análise dos *ratings* dos clientes, individuais, por grupos, por região ou simplesmente para verificar se os *ratings* atribuídos aos clientes estão sendo eficazes nesta carteira em relação ao seu comportamento de atrasos. Talvez exista a possibilidade de que vários clientes com um rating com certa probabilidade de inadimplência esteja se comportando melhor, ou pior, ao visualizar seu comportamento histórico de probabilidades de transição entre as faixas de atraso. Em outras palavras, ao invés de trabalhar com a carteira toda, é possível identificar as probabilidades de transição por recortes mais específicos dentro de uma carteira de crédito, como a de *rating*, ou por um certo grupo de empresas ou setores e etc. Desta maneira podemos verificar a qualidade do crédito de um portfólio. Desta mesma maneira ainda pode ser feita a mesma manipulação nos dados conforme o capítulo anterior, a fim de tentar prever o futuro comportamento daquele grupo, *rating*, setor, região e etc.

4 CONCLUSÃO

Nesta pesquisa foram exploradas algumas das aplicações das cadeias de Markov na análise de risco de crédito de uma dada carteira de crédito. Através da análise das transições entre diferentes faixas de atraso nos pagamentos, é possível entender sobre o comportamento dos devedores, bem como suas probabilidades de avançar em atrasos. Esse tipo de informação tem muitas aplicações na gestão de risco de crédito e projeções de inadimplência.

Na análise presente dos resultados das matrizes de transição, identificamos padrões de transição entre as faixas de atraso que vão além da simples rolagem de contratos. Podemos observar que os devedores não só avançam para faixas maior atraso, mas também permanecem na mesma faixa de atraso ou retornam para faixas de atraso anteriores ao longo do tempo. Essa dinâmica é uma aproximação da realidade muito mais significativa do que as metodologias mais simples, sem trazer muitas dificuldades metodológicas de aplicação.

Com o Teorema da Convergência ao Equilíbrio podemos obter uma transição única para toda uma carteira de crédito. É claro que como estamos tratando de questões econômicas, a resposta também não poderia ser tão simples, já que o comportamento futuro é influenciado por uma infinidade de fatores complexos, mas de qualquer maneira, saber que uma carteira de crédito tem um comportamento que, em média, converge para uma probabilidade única, já é bastante útil para a análise de Risco de Crédito.

Além disso, foram identificadas algumas opções de continuidade desde tipo de análise com a transformação das probabilidades em variáveis de série temporal para serem utilizadas tanto como variáveis explicativas ou como variáveis dependentes de modelos de regressão, abrindo assim um caminho para entender como os fatores externos estão influenciando as probabilidades de transição entre as faixas de atraso. Esse tipo de análise também pode enriquecer as projeções de os indicadores de Risco de Crédito.

Em suma, as Cadeias de Markov demonstraram ser uma ferramenta poderosa na análise de risco de crédito. As probabilidades de transição e a distribuição estacionária são elementos-chave para entender o comportamento da carteira de crédito. Essas descobertas podem ser aplicadas em uma variedade de contextos, incluindo modelos de inadimplência, Perda Esperada (IFRS9), testes de estresse e muito mais. Esta pesquisa oferece uma base sólida para aprimorar a gestão de risco de crédito e tomar decisões informadas no complexo mundo das finanças.

5 REFERÊNCIAS

CHIBISOV, D. M. (2016). **Bernoulli's Law of Large Numbers and the Strong Law of Large Numbers. Theory of Probability & Its Applications,** 60(2), 318–319. doi:10.1137/s0040585x97t987696

JARROW, R. A.; Lando, D.; Turnbull, S. M. (1997). **A Markov Model for the Term Structure of Credit Risk Spreads**. Review of Financial Studies, 10(2), 481–523. doi:10.1093/rfs/10.2.481

KORTEWEG, A. G. (2011). **Markov Chain Monte Carlo Methods in Corporate Finance**. SSRN Electronic Journal. doi:10.2139/ssrn.1964923

PUPO, Fabio. **Governo negocia novas regras contra inadimplência em crédito emergencial.** Folha de São Paulo, 20, fevereiro de 2022. Disponível em: Governo negocia novas regras contra inadimplência em crédito emergencial - 20/02/2022 - Mercado - Folha (uol.com.br) acesso em: 17, novembro de 2023.

SALVADOR, D. H., & Corso, L. L. (2022). **APLICAÇÃO DE CADEIAS DE MARKOV E REGRESSÃO MÚLTIPLA LINEAR PARA ANÁLISE DOS ÍNDICES IBOVESPA, NASDAQ E NYSE.** Revista CIATEC-UPF, 14(3), 1-13. https://doi.org/10.5335/ciatec.v14i3.14025

SANTOS, Reginaldo. **Um Curso de Geometria Analítica e Álgebra Linear. Imprensa Universitária UFMG**, Belo Horizonte, 2003. P. 383-406

SCARIOTE, R.; CORSO, L. L. **APLICAÇÃO DE CADEIAS DE MARKOV MULTIVARIADA E REGRESSÃO LINEAR MÚLTIPLA PARA INVESTIMENTO NO IBOVESPA COM O DÓLAR COMO INFLUENCIADOR.** ConTexto - Contabilidade em Texto, Porto Alegre, v. 21, n. 49, p. 17–29, 2021. Disponível em: https://seer.ufrgs.br/index.php/ConTexto/article/view/117552.

SERICOLA, Bruno. **Markov Chains: Theory, Algorithms and Applications. London**: ISTE Ltd, 2013.

SILVA, M. Larissa da, **Cadeias de Markov e Aplicações**, TCC – Faculdade de Matemática da Universidade Federal Fluminense. Rio de Janeiro.

SIMON, Carl P.; BLUME, Lawrence; DOERING, Claus Ivo. **Matemática para economistas.** Bookman, 2004.

SIU T.-K., Ching, W.-K., Fung, S. E., & Ng, M. K. (2005). **On a multivariate Markov chain model for credit risk measurement. Quantitative Finance**, 5(6), 543–556. doi:10.1080/14697680500383714

SOBOL, I. M. **The Monte Carlo Method**, Moscow: Mir Publishers, 1975.

6 APÊNDICE A - TRANSPOSIÇÃO DAS COLUNAS VIA SQL

```
SELECT [ID], [DT_BASE], [FAIXA_ATRASO]
FROM(
  SELECT
[ID],[DT_BASE],[FAIXA_ATRASO_M0],[FAIXA_ATRASO_M1],[FAIXA_ATRASO_M2],[FAIXA_ATRASO_M3],[FAIXA_ATRASO_M4],[FAIXA_ATRASO_M5],[FAIXA_ATRASO_M6],[FAIXA_ATRASO_M7],[FAIXA_ATRASO_M8],[FAIXA_ATRASO_M9],[FAIXA_ATRASO_M10],[FAIXA_ATRASO_M11],[FAIXA_ATRASO_M12],[FAIXA_ATRASO_M13],[FAIXA_ATRASO_M14],[FAIXA_ATRASO_M15],[FAIXA_ATRASO_M16],[FAIXA_ATRASO_M17],[FAIXA_ATRASO_M18],[FAIXA_ATRASO_M19],[FAIXA_ATRASO_M20],[FAIXA_ATRASO_M21],[FAIXA_ATRASO_M22],[FAIXA_ATRASO_M23],[FAIXA_ATRASO_M24],[FAIXA_ATRASO_M25],[FAIXA_ATRASO_M26],[FAIXA_ATRASO_M27],[FAIXA_ATRASO_M28],[FAIXA_ATRASO_M29],[FAIXA_ATRASO_M30],[FAIXA_ATRASO_M31],[FAIXA_ATRASO_M32],[FAIXA_ATRASO_M33],[FAIXA_ATRASO_M34],[FAIXA_ATRASO_M35],[FAIXA_ATRASO_M36],[FAIXA_ATRASO_M37],[FAIXA_ATRASO_M38],[FAIXA_ATRASO_M39],[FAIXA_ATRASO_M40],[FAIXA_ATRASO_M41],[FAIXA_ATRASO_M42],[FAIXA_ATRASO_M43],[FAIXA_ATRASO_M44],[FAIXA_ATRASO_M45],[FAIXA_ATRASO_M46],[FAIXA_ATRASO_M47],[FAIXA_ATRASO_M48],[FAIXA_ATRASO_M49],[FAIXA_ATRASO_M50],[FAIXA_ATRASO_M51],[FAIXA_ATRASO_M52],[FAIXA_ATRASO_M53],[FAIXA_ATRASO_M54],[FAIXA_ATRASO_M55],[FAIXA_ATRASO_M56],[FAIXA_ATRASO_M57],[FAIXA_ATRASO_M58],[FAIXA_ATRASO_M59],[FAIXA_ATRASO_M60],[FAIXA_ATRASO_M61],[FAIXA_ATRASO_M62],[FAIXA_ATRASO_M63],[FAIXA_ATRASO_M64],[FAIXA_ATRASO_M65],[FAIXA_ATRASO_M66],[FAIXA_ATRASO_M67],[FAIXA_ATRASO_M68],[FAIXA_ATRASO_M69],[FAIXA_ATRASO_M70],[FAIXA_ATRASO_M71],[FAIXA_ATRASO_M72],[FAIXA_ATRASO_M73],[FAIXA_ATRASO_M74],[FAIXA_ATRASO_M75],[FAIXA_ATRASO_M76],[FAIXA_ATRASO_M77],[FAIXA_ATRASO_M78],[FAIXA_ATRASO_M79],[FAIXA_ATRASO_M80],[FAIXA_ATRASO_M81],[FAIXA_ATRASO_M82],[FAIXA_ATRASO_M83]
  FROM [Monografia].[dbo].[Base_Monografia]
) as [TESTE]
UNPIVOT (
FAIXA_ATRASO FOR FAIXA_COLUMN IN (
[FAIXA_ATRASO_M0],[FAIXA_ATRASO_M1],[FAIXA_ATRASO_M2],[FAIXA_ATRASO_M3],[FAIXA_ATRASO_M4],[FAIXA_ATRASO_M5],[FAIXA_ATRASO_M6],[FAIXA_ATRASO_M7],[FAIXA_ATRASO_M8],[FAIXA_ATRASO_M9],[FAIXA_ATRASO_M10],[FAIXA_ATRASO_M11],[FAIXA_ATRASO_M12],[FAIXA_ATRASO_M13],[FAIXA_ATRASO_M14],[FAIXA_ATRASO_M15],[FAIXA_ATRASO_M16],[FAIXA_ATRASO_M17],[FAIXA_ATRASO_M18],[FAIXA_ATRASO_M19],[FAIXA_ATRASO_M20],[FAIXA_ATRASO_M21],[FAIXA_ATRASO_M22],[FAIXA_ATRASO_M23],[FAIXA_ATRASO_M24],[FAIXA_ATRASO_M25],[FAIXA_ATRASO_M26],[FAIXA_ATRASO_M27],[FAIXA_ATRASO_M28],[FAIXA_ATRASO_M29],[FAIXA_ATRASO_M30],[FAIXA_ATRASO_M31],[FAIXA_ATRASO_M32],[FAIXA_ATRASO_M33],[FAIXA_ATRASO_M34],[FAIXA_ATRASO_M35],[FAIXA_ATRASO_M36],[FAIXA_ATRASO_M37],[FAIXA_ATRASO_M38],[FAIXA_ATRASO_M39],[FAIXA_ATRASO_M40],[FAIXA_ATRASO_M41],[FAIXA_ATRASO_M42],[FAIXA_ATRASO_M43],[FAIXA_ATRASO_M44],[FAIXA_ATRASO_M45],[FAIXA_ATRASO_M46],[FAIXA_ATRASO_M47],[FAIXA_ATRASO_M48],[FAIXA_ATRASO_M49],[FAIXA_ATRASO_M50],[FAIXA_ATRASO_M51],[FAIXA_ATRASO_M52],[FAIXA_ATRASO_M53],[FAIXA_ATRASO_M54],[FAIXA_ATRASO_M55],[FAIXA_ATRASO_M56],[FAIXA_ATRASO_M57],[FAIXA_ATRASO_M58],[FAIXA_ATRASO_M59],[FAIXA_ATRASO_M60],[FAIXA_ATRASO_M61],[FAIXA_ATRASO_M62],[FAIXA_ATRASO_M63],[FAIXA_ATRASO_M64],[FAIXA_ATRASO_M65],[FAIXA_ATRASO_M66],[FAIXA_ATRASO_M67],[FAIXA_ATRASO_M68],[FAIXA_ATRASO_M69],[FAIXA_ATRASO_M70],[FAIXA_ATRASO_M71],[FAIXA_ATRASO_M72],[FAIXA_ATRASO_M73],[FAIXA_ATRASO_M74],[FAIXA_ATRASO_M75],[FAIXA_ATRASO_M76],[FAIXA_ATRASO_M77],[FAIXA_ATRASO_M78],[FAIXA_ATRASO_M79],[FAIXA_ATRASO_M80],[FAIXA_ATRASO_M81],[FAIXA_ATRASO_M82],[FAIXA_ATRASO_M83]
)
  ) AS TESTE2
ORDER BY ID, DT_BASE;
```

7 APÊNDICE B – MATRIZ DE TRANSIÇÃO VIA PYTHON

```
import pandas as pd
import numpy as np
atraso = pd.read_csv("monografia_final.csv")
atraso.head()
atraso = atraso.sort_values(by=['ID', 'DT_BASE'])
atraso['DE_FAIXA_ATRASO'] = atraso.groupby('ID')['FAIXA_ATRASO'].shift(1)
atraso = atraso[atraso['ID'] == atraso['ID'].shift(1)]
atraso = atraso.rename(columns={'FAIXA_ATRASO': 'PARA_FAIXA_ATRASO'})
atraso['DE_FAIXA_ATRASO'] = atraso['DE_FAIXA_ATRASO'].astype(int)
atraso['PARA_FAIXA_ATRASO'] = atraso['PARA_FAIXA_ATRASO'].astype(int)
transicoes = atraso.groupby(['DE_FAIXA_ATRASO',
'PARA_FAIXA_ATRASO']).size().reset_index(name='CONT_TRANSICAO')
transicao_total=atraso.groupby('DE_FAIXA_ATRASO').size().reset_index(name='TRANSICAO_TOTA
L')
transicoes = pd.merge(transicoes, transicao_total, on='DE_FAIXA_ATRASO')
transicoes['PROB'] = transicoes['CONT_TRANSICAO'] / transicoes['TRANSICAO_TOTAL']
print('TRANSICAO:\n', transicoes)
```

8 APÊNDICE C – TESTE DE DISTRIBUIÇÃO ESTACIONÁRIA VIA PYTHON

```
atraso = pd.read_csv("monografia_final.csv")
atraso2 = atraso[:250000]
atraso.shape[:]
atraso2 = atraso.sort_values(by=['ID', 'DT_BASE'])
atraso['DE_FAIXA_ATRASO'] = atraso.groupby('ID')['FAIXA_ATRASO'].shift(1)
atraso = atraso[atraso['ID'] == atraso['ID'].shift(1)]
atraso = atraso.rename(columns={'FAIXA_ATRASO': 'PARA_FAIXA_ATRASO'})
atraso['DE_FAIXA_ATRASO'] = atraso['DE_FAIXA_ATRASO'].astype(int)
atraso['PARA_FAIXA_ATRASO'] = atraso['PARA_FAIXA_ATRASO'].astype(int)
transicoes = atraso.groupby(['DE_FAIXA_ATRASO',
'PARA_FAIXA_ATRASO']).size().reset_index(name='CONT_TRANSICAO')
transicao_total=atraso.groupby('DE_FAIXA_ATRASO').size().reset_index(name='TRANSICAO_TOTA
L')
transicoes = pd.merge(transicoes, transicao_total, on='DE_FAIXA_ATRASO')
transicoes['PROB'] = transicoes['CONT_TRANSICAO'] / transicoes['TRANSICAO_TOTAL']
print('TRANSICAO:\n', transicoes)
```