

UNIVERSIDADE DE SÃO PAULO  
ESCOLA POLITÉCNICA

MATHEUS FERREIRA FIGUEIREDO

**Uso de séries temporais para modelagem de inadimplência em recebíveis imobiliários e  
seu monitoramento por gráficos de controle**

São Paulo  
2020



MATHEUS FERREIRA FIGUEIREDO

**Uso de séries temporais para modelagem de inadimplência em recebíveis imobiliários e seu monitoramento por gráficos de controle**

Trabalho de Formatura apresentado à  
Escola Politécnica da Universidade de  
São Paulo para obtenção do Diploma de  
Engenheiro de Produção.

Orientador: Profa. Dra. Linda Lee Ho

São Paulo  
2020



MATHEUS FERREIRA FIGUEIREDO

**Uso de séries temporais para modelagem de inadimplência em recebíveis imobiliários e seu monitoramento por gráficos de controle**

Trabalho de Formatura apresentado à  
Escola Politécnica da Universidade de  
São Paulo para obtenção do Diploma de  
Engenheiro de Produção.

Orientador: Profa. Dra. Linda Lee Ho

São Paulo  
2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte

#### Catálogo-na-publicação

Figueiredo, Matheus Ferreira

Uso de séries temporais para modelagem de inadimplência em recebíveis imobiliários e seu monitoramento por gráficos de controle / M. F. Figueiredo – São Paulo, 2020.

104 p.

Trabalho de Formatura – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1. Séries temporais 2. Modelos paramétricos de previsão 3. Monitoramento por gráficos de controle 4. Séries de inadimplência I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

Nome: FIGUEIREDO, Matheus Ferreira

Título: Uso de séries temporais para modelagem de inadimplência em recebíveis imobiliários e seu monitoramento por gráficos de controle

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do Diploma de Engenheiro de Produção. São Paulo, 2020.

Aprovado em:

Banca Examinadora

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_





## **AGRADECIMENTOS**

À professora Linda Lee Ho, por sua dedicação, compreensão e apoio durante todo este processo. Sua orientação e ensinamentos foram essenciais ao desenvolvimento deste trabalho.

À minha família, por todo o apoio e motivação que sempre me deram, e por seu esforço em sempre me oferecer as melhores oportunidades. Foi graças a vocês que eu pude chegar onde estou hoje.

Aos meus amigos da Poli e do Porto, que me acompanharam por toda essa jornada, por me ajudar nos momentos de dificuldade e partilhar dos momentos de alegria.



## RESUMO

Este trabalho de formatura propõe o desenvolvimento de uma metodologia para previsão e monitoramento de índices de inadimplência em carteiras de recebíveis imobiliários cedidas para operações de securitização, para fornecer informações quantitativas para embasar a tomada de decisão e diminuir os riscos de não pagamento durante a operação. Primeiro foi analisada e tratada uma série temporal de inadimplência para o ajuste de um modelo paramétrico de classe ARIMA. Em seguida, foi selecionado o gráfico de média móvel exponencialmente ponderada para o monitoramento desta série, com o objetivo de identificar um processo fora de controle. Para tal, foram desenvolvidos algoritmos usando a linguagem de programação R para definir seus parâmetros, de modo a atingir o  $ARL_0$  pré-estabelecido de 36 meses. Então foi simulado um processo fora de controle para estudar a sensibilidade do gráfico em relação a mudanças na média da série. A análise dos resultados consistiu de uma avaliação da capacidade de previsão do modelo ajustado, a partir da comparação dos valores previstos com as observações reais. Quanto ao gráfico de controle, este foi avaliado pela sua capacidade de identificar processos fora de controle e perceber pequenas variações na média da variável monitorada. Ao final do desenvolvimento, foi constatado que a solução proposta apresentou uma capacidade de previsão adequada para o curto prazo e uma sensibilidade satisfatória a variações na média da inadimplência.

**Palavras-chave:** Inadimplência. Séries temporais. Modelos ARIMA. Controle estatístico de processos. Gráficos de controle.



## ABSTRACT

This paper proposes the development of a methodology for the forecasting and monitoring of default rates in real estate credit portfolios granted for a securitization operation, in order to provide quantitative information to support the decision-making process and mitigate the risks of payment default throughout the operation. First, the default rate time series was analyzed and treated for the adjustment of an ARIMA-class parametric model. Then the exponentially weighted moving average chart was chosen for the series monitoring, for the purpose of identifying an out-of-control process. To this end, several algorithms were developed using the R programming language to define the chart's parameters, in order to achieve the pre-established  $ARL_0$  of 36 months. The next step consisted of simulating an out-of-control process to determine the chart's sensibility to changes in the process average. The result analysis began with the evaluation of the adjusted model forecasting capacity, by comparing the forecasts with the actual values observed in the process. Regarding the control chart, it was evaluated on its ability to identify an out-of-control process and to detect small variations in the monitored variable. Having concluded the result analysis, it was verified that the proposed solution presented an appropriate forecasting ability in the short-term and a satisfactory sensitivity in its monitoring capacity.

**Keywords:** Default ratios. Time series. ARIMA models. Statistical process control. Control charts.



## LISTA DE FIGURAS

Figura 01 - Evolução de emissões de CRI.....	20
Figura 02 - Exemplo de distribuição normal .....	30
Figura 03 - Exemplo de gráfico Q-Q.....	31
Figura 04 - Gráfico de Pareto .....	37
Figura 05 - Diagram de causa e efeito .....	38
Figura 06 - Histograma.....	38
Figura 07 - Folha de verificação.....	39
Figura 08 - Gráfico de controle .....	39
Figura 09 - Diagrama de dispersão.....	40
Figura 10 - Diagrama de concentração de defeito.....	40
Figura 11 - Fluxograma de seleção de gráfico de controle .....	56
Figura 12 - Fluxograma do algoritmo de definição do ARL em processo sob controle .....	61
Figura 13 - Fluxograma do algoritmo de definição do ARL em processo fora de controle.....	63
Figura 14 - Série de índices de inadimplência.....	66
Figura 15 - Série transformada .....	66
Figura 16 - Série estacionária .....	67
Figura 17 - Q-Q Plot da série estacionária .....	68
Figura 18 - Valores de autocorrelação.....	68
Figura 19 - Valores de autocorrelação parcial.....	69
Figura 20 - Q-Q Plot dos resíduos .....	73
Figura 21 - Valores de autocorrelação nos resíduos.....	73
Figura 22 - Valores das autocorrelações parciais nos resíduos .....	74
Figura 23 - Valores do ARL em função de L .....	76
Figura 24 - Valores de ARL em função de c .....	77
Figura 25 - Série Estacionária versus projeção .....	78
Figura 26 - Comparação entre previsões e observações reais para série estacionária.....	79
Figura 27 - Série original versus projeção.....	80
Figura 28 - Comparação entre previsões e observações reais para série original .....	81
Figura 29 - Gráfico MMEP aplicado à série estacionária .....	82





## LISTA DE TABELAS

Tabela 01 - Valores de AIC para os modelos testados .....	70
Tabela 02 - Estimativa dos coeficientes do modelo ARMA(4,5).....	71
Tabela 03 - Estimativa dos coeficientes do modelo ARMA(4,5) reduzido.....	71
Tabela 04 - Valores de MSE e RMSE versus horizonte de previsão .....	81



## LISTA DE ABREVIATURAS E SIGLAS

<b>AIC</b>	Critério de Informação de Akaike
<b>AR</b>	Modelo Autorregressivo
<b>ARIMA</b>	Modelo Autorregressivo Integrado de Médias Móveis
<b>ARL</b>	<i>Average Run Length</i>
<b>ARMA</b>	Modelo Autorregressivo de Médias Móveis
<b>CEP</b>	Controle Estatístico de Processos
<b>CRI</b>	Certificado de Recebíveis Imobiliários
<b>CUSUM</b>	Gráfico de Controle de Soma Cumulativa
<b>CVM</b>	Comissão de Valores Imobiliários
<b>FAC</b>	Função de Autocorrelação
<b>FACP</b>	Função de Autocorrelação Parcial
<b>FACV</b>	Função de Autocovariância
<b>LC</b>	Linha Central
<b>LIC</b>	Limite Inferior de Controle
<b>LSC</b>	Limite Superior de Controle
<b>MA</b>	Modelo de Médias Móveis
<b>MMEP</b>	Gráfico de Controle de Média Móvel Exponencialmente Ponderada
<b>MSE</b>	<i>Mean Square Error</i>
<b>RMSE</b>	<i>Root Mean Square Error</i>



## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO.....</b>	<b>19</b>
<b>1.1.</b>	<b>Contexto.....</b>	<b>19</b>
<b>1.2.</b>	<b>A empresa .....</b>	<b>21</b>
<b>1.3.</b>	<b>Problema a ser abordado .....</b>	<b>22</b>
<b>1.4.</b>	<b>Estrutura do trabalho.....</b>	<b>23</b>
<b>2.</b>	<b>REVISÃO BIBLIOGRÁFICA – SÉRIES TEMPORAIS.....</b>	<b>25</b>
<b>2.1.</b>	<b>Processos estocásticos .....</b>	<b>26</b>
<b>2.2.</b>	<b>Estacionariedade.....</b>	<b>26</b>
<b>2.3.</b>	<b>Normalidade.....</b>	<b>29</b>
<b>2.4.</b>	<b>Independência .....</b>	<b>32</b>
2.4.1.	Função de autocovariância .....	32
2.4.2.	Função de autocorrelação .....	33
2.4.3.	Função de autocorrelação parcial .....	33
<b>2.5.</b>	<b>Modelos Paramétricos .....</b>	<b>34</b>
2.5.1.	Modelos Autorregressivos .....	34
2.5.2.	Modelos de Medias Móveis.....	35
2.5.3.	Modelos Autorregressivos de Médias Móveis .....	35
2.5.4.	Modelos Autorregressivos Integrados de Médias Móveis.....	36
<b>3.</b>	<b>REVISÃO BIBLIOGRÁFICA – CONTROLE ESTATÍSTICO DE PROCESSOS .....</b>	<b>37</b>
<b>3.1.</b>	<b>Gráficos de controle.....</b>	<b>41</b>
3.1.1.	Gráficos de Controle Shewhart.....	43
3.1.2.	Gráficos de Controle de Soma Cumulativa (CUSUM) .....	45
3.1.3.	Gráficos de Controle de Média Móvel Exponencialmente Ponderada (MMEP) .....	46
<b>4.</b>	<b>METODOLOGIA.....</b>	<b>49</b>
<b>4.1.</b>	<b>Obtenção dos dados .....</b>	<b>49</b>

<b>4.2.</b>	<b>Análise e ajuste da série .....</b>	<b>50</b>
<b>4.3.</b>	<b>Identificação do modelo .....</b>	<b>51</b>
<b>4.4.</b>	<b>Diagnóstico do modelo .....</b>	<b>54</b>
<b>4.5.</b>	<b>Escolha do gráfico de controle .....</b>	<b>55</b>
<b>4.6.</b>	<b>Definição dos limites de controle.....</b>	<b>56</b>
4.6.1.	Cálculo do Fator de Suavização $\lambda$ .....	57
4.6.2.	Definição ARL sob Controle.....	58
4.6.3.	Determinação ARL Fora de Controle .....	61
<b>4.7.</b>	<b>Avaliação dos Resultados .....</b>	<b>63</b>
<b>5.</b>	<b>APLICAÇÃO SOBRE O PROBLEMA PROPOSTO.....</b>	<b>65</b>
<b>5.1.</b>	<b>Coleta dos dados .....</b>	<b>65</b>
<b>5.2.</b>	<b>Análise e ajustes na série .....</b>	<b>65</b>
<b>5.3.</b>	<b>Identificação do modelo .....</b>	<b>69</b>
<b>5.4.</b>	<b>Diagnóstico do modelo .....</b>	<b>72</b>
<b>5.5.</b>	<b>Escolha do gráfico de controle .....</b>	<b>74</b>
<b>5.6.</b>	<b>Definição dos limites de controle.....</b>	<b>75</b>
5.6.1.	Definição ARL sob controle .....	75
5.6.2.	Definição ARL fora de controle.....	76
<b>5.7.</b>	<b>Avaliação dos resultados.....</b>	<b>77</b>
<b>6.</b>	<b>CONCLUSÃO .....</b>	<b>87</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>91</b>
	<b>APÊNDICE A – Código para análise da série temporal e ajuste do modelo ARMA .....</b>	<b>93</b>
	<b>APÊNDICE B – Código para determinação do fator de suavização do MMEP .....</b>	<b>95</b>
	<b>APÊNDICE C – Código para determinação do ARL para processo sob controle .....</b>	<b>97</b>
	<b>APÊNDICE D – Código para determinação do ARL para processo fora de controle ..</b>	<b>101</b>
	<b>ANEXO A – Série de índices de inadimplência .....</b>	<b>103</b>

## **1. INTRODUÇÃO**

Este trabalho de formatura foi desenvolvido junto à Forte Securitizadora S.A. (Fortesec), instituição financeira na qual o autor está realizando estágio supervisionado na área comercial. A empresa foi escolhida para servir de alvo para o trabalho pela proximidade do autor com ela, que permitiu a identificação do problema e uma grande disponibilidade de informações. Neste primeiro capítulo serão apresentados o contexto em que se elaborou o trabalho, o problema levantado e sua relevância e, finalmente, a abordagem proposta para solução deste problema.

### **1.1. Contexto**

A securitização é uma operação de estruturação e emissão de títulos de dívida lastreados em direitos creditórios, com o intuito de captar recursos para empreendedores via investidores do mercado de capitais. Apenas empresas especializadas, chamadas de securitizadoras, tem a permissão da Comissão de Valores Mobiliários (CVM) para emitir esses títulos. Os direitos creditórios envolvidos podem ser de diversas naturezas, mas neste trabalho será abordado o caso específico de créditos oriundos de empreendimentos imobiliários.

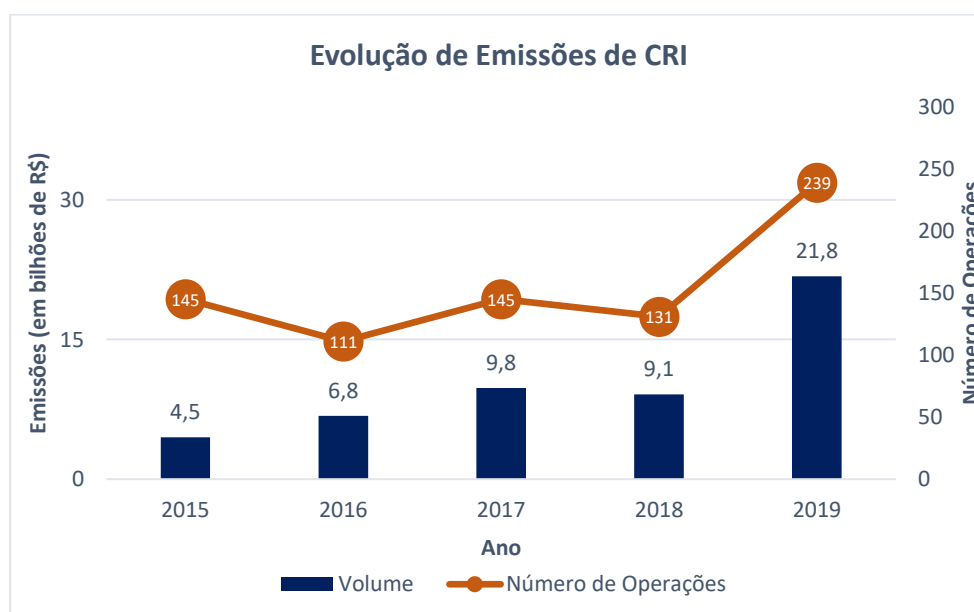
Normalmente, durante a venda de unidades de um empreendimento imobiliário, os compradores assinam contratos com pagamento parcelado por um longo período. O conjunto de todas as parcelas a receber de todos os contratos de um empreendimento constitui o que é chamado de carteira de recebíveis. No entanto, em diversas situações, é interessante para o empreendedor antecipar o recebimento dessas parcelas para poder usá-lo no presente, para financiar obras, aplicar em outros projetos, ou diversas outras finalidades. E é essa necessidade que a securitização busca satisfazer. Empresas com interesse em antecipar recebíveis contatam a securitizadora, e esta analisa o valor da carteira de recebíveis oferecida. Uma vez feita a precificação, a securitizadora contata potenciais investidores, e oferece a eles o direito aos recebíveis da carteira por um prazo determinado. Caso haja interesse, é feito um acordo e os investidores oferecem um certo valor em troca dos direitos aos recebíveis. A securitizadora atua então como intermediária, que repassa o recurso dos investidores ao empreendedor, e faz a gestão da carteira por toda duração da operação, acompanha e garante os pagamentos aos investidores. A transferência dos recursos pelos direitos creditórios é formalizada pela emissão de um título de dívida chamado de Certificado de Recebíveis Imobiliários (CRI).

Esse processo ganhou espaço no mercado como uma alternativa às linhas mais tradicionais de financiamento, como o financiamento bancário, devido a seu formato mais flexível e customizável de acordo com as necessidades dos clientes. Algumas das vantagens da securitização que podem ser citadas em relação às linhas tradicionais são um maior prazo de financiamento, transferência de risco para os investidores, menores impactos no balanço patrimonial e desintermediação bancária.

O mercado de securitização brasileiro ainda é pouco desenvolvido, especialmente se comparado aos mercados em países europeus ou nos Estados Unidos, o que sugere que ainda há um enorme potencial a ser aproveitado pelas empresas do setor. Nos últimos anos já foi registrado um crescimento relevante, sinal de que existem cada vez mais empresas buscando captar recursos por esta modalidade.

A Fortesec faz um mapeamento constante das novas operações de CRI para monitorar as condições de mercado e acompanhar novas tendências. Os dados consolidados em uma base anual estão apresentados na Figura 01. Em 2019, foi calculado um volume recorde de recursos captados via CRI, um total de R\$ 21,8 bilhões. Fica clara a tendência de crescimento desse mercado observando os últimos 5 anos, tanto em relação ao volume quanto ao número de operações de securitização. Importante ressaltar que a Figura 01 apresenta dados relativos apenas às operações nas quais as cedentes dos créditos não são os próprios investidores da operação e não é usada a Taxa Referencial como indexador.

Figura 01 - Evolução de emissões de CRI



Fonte: Elaborado pelo autor baseado em dados passados pela Fortesec



## **1.2. A empresa**

A Fortesec foi fundada em 2015 com o objetivo de conectar empreendedores que buscam recursos com investidores que procuram investimentos rentáveis. Atua na securitização de recebíveis dos setores de agronegócio e imobiliário, sendo este último seu foco e especialidade, que corresponde a 98,7% de suas operações. Atende empresas de diversos portes em todas as regiões do país. Tem como diferenciais a modelagem customizada de cada estrutura de dívida, para atender as necessidades específicas de cada cliente, e um foco maior na gestão do CRI, ao manter um contato próximo com empresas e investidores, o que proporciona maior êxito e segurança a todas suas operações.

Segue uma breve explicação do processo interno de securitização. Inicia-se na área comercial, na qual o autor deste trabalho estagia. A área comercial faz um primeiro contato com a empresa interessada em securitizar recebíveis e solicita todos documentos necessários à análise. É feita uma avaliação detalhada das características do empreendimento e do valor da carteira de recebíveis, através de critérios como número de contratos, desempenho das vendas, valor presente dos recebíveis e comportamento de pagamentos da carteira, que é analisado através de índices de inadimplência. Em seguida, realiza-se uma modelagem financeira inicial da operação, que determina características como o volume de captação, a taxa de juros, o prazo, a forma de pagamento, entre outros diversos aspectos. Essa primeira estrutura é então apresentada a potenciais investidores, e é negociada uma estrutura final para a operação que seja financeiramente interessante para eles. Essas condições são então apresentadas em forma de proposta de emissão de CRI para a empresa.

Uma vez que a proposta é assinada, a área de estruturação assume a operação. Ela irá elaborar todos os documentos necessários, registrar a emissão junto às instituições reguladoras do mercado, supervisionar o processo de auditoria da carteira de recebíveis e do empreendimento e finalmente contratar todos os prestadores de serviço que irão atuar junto à securitizadora. Assim que todos requisitos legais e operacionais forem cumpridos, ocorre a liquidação, ou seja, recebimento de capital dos investidores que é repassado para a empresa cedente dos créditos.

A partir da liquidação até o vencimento da operação, a área de gestão assume. Ela fará um acompanhamento e monitoramento do empreendimento junto a todos os prestadores de serviços, que avalia o progresso de obras, o recebimento dos créditos e os índices de

inadimplência. E a partir desse acompanhamento irá prover tanto os empreendedores quanto os investidores de todas informações que necessitam para garantir transparência por todo processo.

### **1.3. Problema a ser abordado**

Este trabalho tem como foco a análise dos índices de inadimplência dos recebíveis imobiliários, que como explicado anteriormente, impactam tanto a etapa de modelagem financeira quanto de gestão da operação. Atualmente, em ambas as etapas, analisa-se o histórico de pagamentos da carteira e calculam-se as quantidades de pagamentos em dia, atrasados, antecipados e inadimplentes, e então toma-se as decisões com base no comportamento passado da carteira.

No entanto, a empresa carece de projeções desses índices para períodos futuros para apoiá-la na tomada de decisões. Essas projeções serão de grande auxílio, principalmente ao levar em consideração a longa duração dessas operações, que necessariamente exige um planejamento de longo prazo desde as primeiras etapas da securitização.

O índice a ser estudado é a inadimplência mensal bruta. Este pode ser calculado através de dados fornecidos por um prestador de serviços contratado para auditar mensalmente a carteira de recebíveis.

Para resolver o problema será empregado um modelo de séries temporais e estatística para se desenvolver projeções do índice de inadimplência e uso de métodos de controle de processos para se elaborar critérios de decisões através de limites de controle, que possibilitem a tomada de decisões com um maior embasamento quantitativo.

A relevância do problema abordado se justifica pela utilidade da solução proposta para as áreas comercial e de gestão. A começar pela área comercial, a possibilidade de prever um comportamento futuro dos pagamentos de uma carteira dará uma visão muito mais precisa sobre os riscos da securitização, e esses riscos serão levados em conta na modelagem, onde serão aplicados mecanismos para mitigá-los, como ajustes no volume de emissão ou estabelecimento de garantias adicionais. Além disso, levar essas informações para a negociação com os investidores disponibilizará uma base mais rica para suportar uma decisão de investimento, o que aumentará a atratividade dos produtos ofertados pela Fortesec.

Em relação ao departamento de gestão, atualmente ao notar um aumento na inadimplência, eles alertam a empresa dona do empreendimento e auxiliam no planejamento e implantação de medidas para corrigir a situação, como, por exemplo, a contratação de agentes de cobrança. Mas dado que é analisado o comportamento atual e passado, esses sinais surgem apenas quando o problema se apresenta, o que possibilita apenas a tomada de medidas corretivas. Projeções bem fundamentadas possibilitariam medidas preventivas, fornecendo uma maior segurança aos investidores e agregando um maior valor para seus clientes, através de informações que permitem uma administração mais eficiente.

#### **1.4. Estrutura do trabalho**

Este trabalho foi iniciado através do presente capítulo “Introdução”, que fornece ao leitor todo o contexto necessário para se entender o problema que será abordado e qual a sua relevância para a empresa alvo deste trabalho.

Em seguida, nos capítulos “Revisão Bibliográfica- Séries Temporais” e “Revisão Bibliográfica – Controle Estatístico de Processos” será feito um relato de todo conhecimento teórico necessário para abordagem do problema proposto. Serão apresentados conceitos sobre modelos paramétricos para séries temporais e ferramentas de controle estatísticos de processo, com foco em gráficos de controle.

No capítulo “Metodologia” será estruturado um processo para desenvolvimento de um modelo paramétrico de previsão e monitoramento por gráficos de controle, utilizando dos conceitos abordados na revisão bibliográfica.

Tendo sido apresentada toda base conceitual e teórica, será feita a “Aplicação sobre o problema proposto”, no qual os métodos levantados são utilizados para uma base de dados específica fornecida pela Fortesec. Após o término do desenvolvimento da solução será feita uma análise dos resultados.

Por último, o capítulo “Conclusão” trará um resumo deste trabalho de formatura, uma avaliação geral da solução encontrada, suas limitações e o quanto ela pode conseguir solucionar o problema levantado, além de prover alternativas de estudos futuros para aprimorar essa solução.

Para se realizar todas as análises necessárias e desenvolver o modelo proposto, serão usados os softwares Microsoft Excel, para tratamento de dados e elaboração de gráficos para apresentar os resultados de forma mais visual, e o RStudio, ambiente de desenvolvimento em linguagem R, usado para a construção do modelo de previsão e determinação dos limites do gráfico de controle.

## 2. REVISÃO BIBLIOGRÁFICA – SÉRIES TEMPORAIS

Neste capítulo serão apresentados os principais conceitos e metodologias empregados no estudo da série de dados de inadimplência, e que servirão de base para a construção do modelo de projeção de valores futuros.

De acordo com Morettin e Toloi (2006), séries temporais podem ser definidas como “qualquer conjunto de observações ordenadas no tempo”. Segundo esta definição, pode-se afirmar que uma série de observações mensais de índices de inadimplência configura uma série temporal. Morettin e Toloi (2006) também enumeram alguns dos principais objetivos associados à aplicação deste conceito:

- a) investigar o mecanismo gerador da série temporal;
- b) fazer previsões de valores futuros da série;
- c) descrever apenas o comportamento da série;
- d) procurar periodicidades relevantes nos dados.

A partir desses possíveis usos é possível justificar o uso de séries temporais em diversas áreas do conhecimento. Apesar de suas aplicações mais tradicionais estarem relacionadas aos campos da economia ou física, por exemplo, seu uso também é bem justificado neste trabalho, uma vez que o objetivo é obter projeções de valores futuros de índices de inadimplência a partir de comportamento passado da série.

A análise por séries temporais pode ser abordada por duas perspectivas diferentes. Uma delas é a análise no domínio temporal, que seria a observação dos diversos valores de  $Z(t)$  ao longo de um período de observação. Neste caso, são construídos os chamados modelos paramétricos, que são aqueles definidos por um número finito de parâmetros. Esta é a perspectiva que será adotada neste trabalho. A segunda perspectiva é a análise da série no domínio das frequências, onde é feita uma análise espectral que resulta em um modelo não paramétrico.

Nas seções seguintes serão explicados alguns conceitos fundamentais para a aplicação de séries temporais.

## 2.1. Processos estocásticos

Morettin e Toloi (2006) definiram processos estocásticos como:

**Definição:** Seja  $\mathbb{T}$  um conjunto arbitrário. Um processo estocástico é uma família  $Z = \{Z(t), t \in \mathbb{T}\}$ , tal que, para cada  $t \in \mathbb{T}$ ,  $Z(t)$  é uma variável aleatória.

Assim, definido um espaço de probabilidades  $(\Omega, A, \mathbb{P})$ , pode-se considerar processos estocásticos como uma família de variáveis aleatórias definidas neste espaço. Esses processos são regularmente usados como forma de descrever séries temporais.

Ainda segundo Morettin e Toloi (2006), é possível classificar os processos estocásticos de acordo com três categorias:

- a) processos estacionários ou não estacionários, de acordo com a independência ou não relativamente à origem dos tempos;
- b) processos normais (Gaussianos) ou não normais, de acordo com as funções densidade de probabilidade que caracterizam os processos;
- c) processos Markovianos ou não Markovianos, de acordo com a independência dos valores do processo, em dados instante, de seus valores de instantes precedentes.

## 2.2. Estacionariedade

A estacionariedade é a existência de um comportamento constante no desenvolvimento da série ao longo do tempo, sempre mantendo sua média e variância para todas as observações. Dessa forma, caso o processo seja estacionário, a origem dos tempos não tem impacto na análise, uma vez que as mesmas características seriam encontradas independente do ponto de partida. Segundo Morettin e Toloi (2006), as séries podem ter estacionariedade fraca ou forte, de acordo com as definições apresentadas a seguir:

**Definição:** Um processo estocástico  $Z = \{Z(t), t \in \mathbb{T}\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n)$$

para quaisquer  $t_1, \dots, t_n, \tau \in \mathbb{T}$ . Isso significa que, para qualquer  $t \in \mathbb{T}$  tem-se que:

a)  $\mu(t) = \mu$ ;

b)  $V(t) = \sigma^2$ .

**Definição:** Um processo estocástico  $Z = \{Z(t), t \in \mathbb{T}\}$  diz-se fracamente estacionário ou estacionário de segunda ordem (ou em sentido amplo) se e somente se:

a)  $E\{Z(t)\} = \mu(t) = \mu$ , constante para todo  $t \in \mathbb{T}$ ;

b)  $E\{Z(t)\} < \infty$ , para todo  $t \in \mathbb{T}$ ;

c)  $\gamma(t_1, t_2) = Cov\{Z(t_1), Z(t_2)\}$  é uma função de  $|t_1 - t_2|$ , onde  $\gamma(t_1, t_2)$  é a função de autocovariância, que é apresentada em detalhes na subseção 2.4.1. deste trabalho.

Existem diversas alternativas de testes estatísticos disponíveis para verificar se uma série é ou não estacionária. Aqui será abordado o caso do teste desenvolvido por Dickey e Fuller (1979), que leva o nome dos autores. Este método busca determinar se o polinômio autorregressivo de uma série temporal possui raízes sobre o círculo unitário. Caso possua, conclui-se que o processo estocástico não apresenta estacionariedade, e é necessário o tratamento da série original através das suas diferenças (esse conceito é apresentado em detalhe na subseção 2.5.4).

Para exemplificar a aplicação do teste, é considerado o seguinte modelo de média zero:

$$Z_t = \phi Z_{t-1} + a_t, \quad a_t \sim RBN(0, \sigma^2)$$

sendo  $a_t$  um ruído branco de média zero e desvio variância igual a  $\sigma^2$ , com distribuição aproximada pela distribuição normal. Ruído branco foi definido por Morettin e Toloi (2006) como:

**Definição:** Dizemos que  $\{\varepsilon_t, t \in \mathbb{Z}\}$  é um ruído branco discreto se as variáveis aleatórias  $\varepsilon_t$  não são correlacionadas, isto é  $Cov\{\varepsilon_t, \varepsilon_s\} = 0, t \neq s$ .

Já  $\phi$  é o operador translação para o futuro, definido por Morettin e Toloi (2006) como  $\phi Z_t = Z_{t+1}$ .

Para ajustar o modelo proposto, é necessário subtrair  $Z_{t-1}$  de ambos lados da equação, que resulta em:

$$\Delta Z_t = \phi^* Z_{t-1} + a_t,$$

onde  $\phi^* = \phi - 1$ . A partir do resultado desta equação, será aplicado o seguinte teste de hipótese:

$$H_0: \phi^* = 0$$

$$H_1: \phi^* < 0$$

A hipótese  $H_0$  representa o caso em que existe ao menos uma raiz unitária para o polinômio autorregressivo. Caso seja verificado como verdadeira, implica na não estacionaridade da série. Por outro lado, caso  $H_0$  seja falsa, e então  $H_1$  considerada verdadeira, assume-se que não existem raízes unitárias e que a série é estacionária.

Para realizar o teste, é utilizada a seguinte estatística:

$$\hat{\tau}_{DF} = \frac{N^{-1} \sum Z_{t-1} a_t}{S(N^{-2} \sum Z_{t-1}^2)^{1/2}}$$

na qual  $N$  corresponde tamanho = da amostra e  $S^2$  representa o estimador da variância  $\sigma^2$ , dado por:

$$S^2 = \frac{1}{N-2} \sum_{t=2}^N (\Delta Z_t - \hat{\phi}_{MQ}^* Z_{t-1})^2$$

no qual  $\hat{\phi}_{MQ}^*$  é o estimador considerado de mínimos quadrados de  $\phi^*$  por meio da regressão de  $\Delta Z_t$  sobre  $Z_{t-1}$ .

Uma limitação do teste Dickey-Fuller é que ele opera sobre a premissa de que o processo estocástico analisado tenha dependência apenas em relação ao valor imediatamente anterior, ou seja, a  $Z_{t-1}$ . Dessa forma, um método mais robusto criado a partir do teste Dickey-Fuller é conhecido como Dickey-Fuller aumentado, que conta com as mesmas hipóteses, mas com uma nova forma de cálculo da estatística:



$$\hat{\tau}_{ADF} = \frac{N^{-1} \sum Z_{t-1} a_t}{S(N^{-2} \sum (Z_{t-1} - \bar{Z})^2)^{1/2}}$$

Serão definidos valores críticos para cada tabulação da estatística em função de  $N$ . Caso a estatística para um determinado  $N$  seja menor que o valor crítico, toma-se  $H_0$  como falso, e considera-se a série como estacionária.

Um problema comum ao se analisar séries temporais é a não apresentação de um comportamento estacionário em sua variância devido a grandes variações no decorrer do tempo. Dessa forma, para que se obtenha estacionariedade é necessário manipular esta série para estabilizar sua variância. Existem diversas categorias de transformações para este fim, como as exponenciais e as logarítmicas. Neste caso, será considerada a transformação proposta por Box e Cox (1964), da forma:

$$Z_t^{(\gamma)} = \begin{cases} \frac{Z_t^\gamma - c}{\gamma}, & \text{se } \gamma \neq 0 \\ \log Z_t, & \text{se } \gamma = 0 \end{cases}$$

sendo  $\gamma$  e  $c$  parâmetros a serem estimados de forma a se obter uma variância constante e uma série com distribuição aproximadamente normal.

### 2.3. Normalidade

Segundo Morettin e Toloi (2006) um processo é definido como gaussiano se:

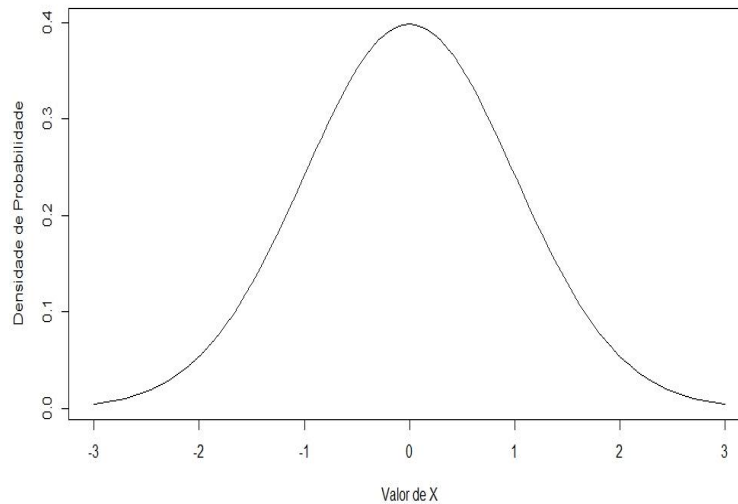
**Definição:** um processo estocástico  $Z = \{Z(t), t \in \mathbb{T}\}$  diz-se Gaussiano se, para qualquer conjunto  $t_1, t_2, \dots, t_n$  de  $\mathbb{T}$ , as variáveis aleatórias  $Z(t_1), \dots, Z(t_n)$  tem distribuição normal  $n$ -variada.

Sendo o processo estocástico identificado como normal, ou gaussiano, ele pode ser determinado pelas suas médias e covariâncias. Um caso particular é aquele no qual a série seja gaussiana e estacionária de segunda ordem, que implica necessariamente que ele seja estritamente estacionário.

Para processos gaussianos, a função de densidade de probabilidade a ser usada é aquela da distribuição normal, tal como mostrado na Figura 02, dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \text{para } -\infty \leq x \leq \infty$$

Figura 02- Exemplo de distribuição normal



Fonte: Elaborado pelo autor

Devido à complexidade dos cálculos, muitas vezes não estão disponíveis valores da esperança  $\mu$  e desvio padrão  $\sigma$ . Assim, uma aproximação aceitável são os estimadores  $\bar{X}$  e  $S$  para esperança e desvio padrão, respectivamente, que são dados por:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

sendo  $n$  o tamanho da amostra e  $X_i$  os valores de cada uma das observações da amostra.

Para determinar se uma série apresenta distribuição próxima à normal, pode-se aplicar um teste não paramétrico. Uma opção popularmente utilizada nesses casos é aquele proposto por Shapiro e Wilk (1965), que leva o nome dos autores. Ele se baseia em um teste de hipótese para verificar se uma série pode ser considerada Gaussiana, a partir das seguintes hipóteses:

$H_0$ : os dados apresentam distribuição normal  $N(\mu, \sigma^2)$

$H_1$ : os dados não apresentam distribuição normal

A reprovação, ou não, da hipótese nula é feita a partir do cálculo da estatística  $W$ , calculada da seguinte forma:

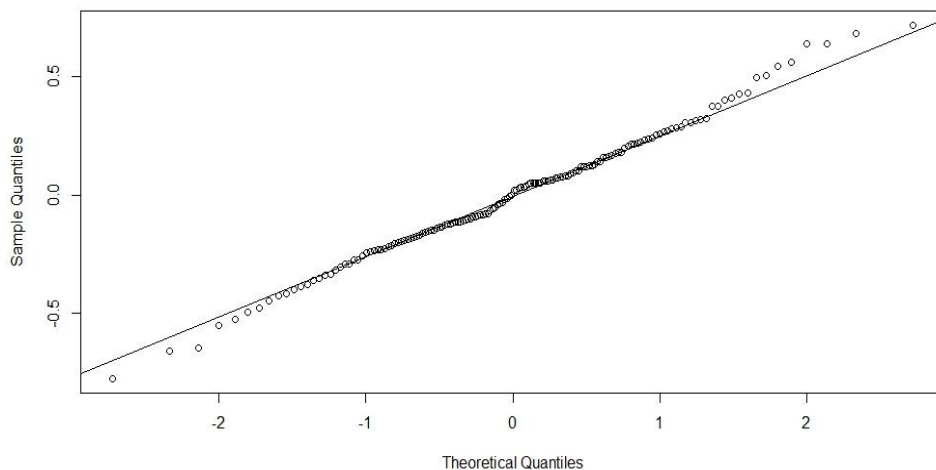
$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

sendo  $x_i$  os valores das observações,  $\bar{x}$  a média desses valores,  $n$  o tamanho da amostra e  $b$  uma constante obtida em função de  $n$ . Como critério de avaliação da hipótese nula, a estatística  $W$  é comparada com valores críticos de estatística  $W_\alpha$ , obtidos através de  $n$  e do nível de significância  $\alpha$  adotado. Caso  $W < W_\alpha$ , a rejeita-se  $H_0$  e conclui-se que a série não possui distribuição normal.

Importante citar que este teste é indicado para amostras com um número relativamente baixo de observações. Dado que a série de percentual de inadimplência a ser analisada posteriormente possui menos de 200 amostras, esta metodologia se mostra apropriada para o estudo proposto.

Para analisar visualmente a distribuição de uma série, um dispositivo gráfico muito usado é o Q-Q Plot, que se baseia na comparação entre os quantis de uma distribuição teórica e da série estudada, na qual se deseja verificar se há ou não normalidade. Conforme mostrado na Figura 03, a distribuição normal é representada pela reta teórica, e os pontos distribuídos ao seu redor simbolizam a distribuição que se deseja testar. Quando mais aderência houver, ou seja, mais próximos da reta os pontos estiverem, mais próximo de uma distribuição normal a distribuição testada está.

Figura 03 – Exemplo de gráfico Q-Q



Fonte: Elaborado pelo autor

## 2.4. Independência

Aqui serão apresentadas as funções de autocovariância, autocorrelação e autocorrelação parcial, que são utilizadas na análise de séries temporais com o intuito de classificá-las como Markovianas ou não.

### 2.4.1. Função de autocovariância

A função de autocovariância (facv) mede o grau de variação de segundo momento, ou segunda ordem, entre dois elementos situados em diferentes espaços de tempo. Morettin e Toloi (2006) definem esta função como:

**Definição:** Seja  $\{X_t, t \in \mathbb{Z}\}$  um processo estacionário real com tempo discreto, de média zero e  $facv \gamma_t = E \{X_t, X_t + \tau\}$ .

A facv  $\gamma_t$  satisfaz as seguintes propriedades:

- a)  $\gamma_0 > 0$ ;
- b)  $\gamma_{t-T} = \gamma_t$ ;
- c)  $|\gamma_t| \leq \gamma_0$ ;
- d)  $\gamma_t$  é não negativa definida, no sentido que:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0,$$

para quaisquer números reais  $a_1, \dots, a_n$  e  $\tau_1, \dots, \tau_n$  de  $\mathbb{Z}$ .

Como seu cálculo pode se provar muito complexo, muitas vezes a função é desconhecida. Assim, pode ser usada a estimativa  $c_j$  dada por:

$$c_j = \frac{1}{N} \sum_{t=1}^{N-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], j = 0, 1, \dots, N-1,$$

sendo  $\bar{X} = \sum_{t=1}^N X_t$  a média amostral, e com  $c_j = c_{-j}$ .

### 2.4.2. Função de autocorrelação

A função de autocorrelação (fac) mede o efeito de uma observação  $X_t$  sobre as demais observações da série, levando em consideração um intervalo de tempo  $\tau$  entre as duas. Morettin e Tolo (2006) definem esta função como:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \tau \in \mathbb{Z}$$

São mantidas as propriedades de  $\gamma_t$ , exceto que agora  $\rho_0 = 1$ . Quando  $\rho_j = 1$ , temos uma correlação perfeita e  $\rho_j = -1$  indica uma correlação perfeita em sentidos opostos entre duas variáveis.  $\rho_j = 0$  indica a ausência de correlação elas. A partir da definição da estimativa da facv, é possível obter uma estimativa da função de autocorrelação  $r_j$  dada por:

$$r_j = \frac{c_j}{c_0}, j = 0, 1, \dots, n-1$$

A fac é utilizada para avaliar se o processo de identificação de um modelo está adequado à série proposta.

### 2.4.3. Função de autocorrelação parcial

Box, Jenkins e Reinsel (1994) sugerem o conceito de função de autocorrelação parcial (facp) como ferramenta adicional no processo de identificação de modelos para simplificá-lo. A facp mede a correlação pura entre duas variáveis  $X_t$  e  $X_{t+j}$ , excluindo efeitos de quaisquer outras variáveis sobre elas. Morettin e Tolo (2006) definem a facp da seguinte forma:

**Definição:** Vamos denotar por  $\phi_{kj}$  o  $j$ -ésimo coeficiente de um modelo AR(k) (o modelo AR(k) será definido posteriormente na subseção 2.5.1), de tal modo que  $\phi_{kk}$  seja o último coeficiente. Sabemos que:

$$\rho_j = \phi_{k1}\rho_1 + \phi_{k2}\rho_2 + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k,$$

A partir das quais obtemos as equações de Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \vdots & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \dots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_k \end{bmatrix}$$

Resolvendo estas equações sucessivamente para  $k = 1, 2, 3, \dots$  obtemos:

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

Em geral,

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}$$

onde  $P_k$  é a matriz de autocorrelações e  $P_k^*$  é a matriz  $P_k$  com a última coluna substituída pelo vetor de autocorrelações.

## 2.5. Modelos Paramétricos

Conforme dito anteriormente, a série alvo de estudo deste trabalho será analisada sobre o escopo do domínio temporal, e para isso são construídos modelos paramétricos. Nesta subseção serão estudadas quatro variedades de modelos: autorregressivos, de médias móveis, autorregressivos de médias móveis e, finalmente, autorregressivos integrados de médias móveis.

### 2.5.1. Modelos Autorregressivos

Os modelos autorregressivos são representados pela notação AR(p), no qual p representa a ordem. Modelos desta natureza assumem que o valor de um elemento  $Z_t$  de uma série temporal é definido pelos valores das p observações passadas, acrescido de um erro, representado por um ruído branco.

A forma genérica de um modelo AR(p) é definida por Morettin e Toloi (2006) como:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_p + a_t$$

onde  $\tilde{Z}_t = Z_t - \mu$ , sendo  $\mu$  a média das observações,  $\phi_p$  o peso atribuído a observação passada  $\tilde{Z}_{t-i}, i = 1, \dots, p$  e  $a_t$  um ruído branco  $RB(0, \sigma^2)$ . Segundo Morettin e Toloi (2006) outra forma de representar modelos  $AR(p)$  é através de um operador autorregressivo de ordem  $p$ , definido por eles como:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

sendo  $B$  o operador translação para o passado, definido por Morettin e Toloi (2006) como:

$$BZ_t = Z_{t-1}, \quad B^m Z_t = Z_{t-m}$$

Assim, podemos representar modelos  $AR(p)$  por:

$$\phi(B)\tilde{Z}_t = a_t$$

### 2.5.2. Modelos de Medias Móveis

Os modelos de médias móveis são representados pela notação  $MA(q)$ , sendo  $q$  a ordem. Diferente dos modelos  $AR(p)$ , neste caso o modelo assume que os valores de  $Z_t$  são obtidos através de um sistema linear, nos quais as entradas são os ruídos brancos verificados nas  $q$  observações passadas. Morettin e Toloi (2006) apresentaram a seguinte equação genérica para representar modelos  $MA(q)$ :

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

sendo  $\theta_j$  o peso atribuído ao ruído  $a_{t-j}, j = 1, \dots, q$  e  $a_t, \dots, a_{t-q}$  ruídos brancos  $RB(0, \sigma^2)$ . Analogamente, é possível representar essa equação através do operador de médias móveis apresentado por Morettin e Toloi (2006):

$$\theta(B) = 1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_q B^q$$

Dessa forma, o modelo  $MA(q)$  é representado por:

$$\tilde{Z}_t = \theta(B)a_t$$

### 2.5.3. Modelos Autorregressivos de Médias Móveis

Modelos autorregressivos de média móvel, denotados ARMA(p,q), são resultados da aplicação simultânea dos modelos AR(p) e MA(q), para se analisar a progressão das observações  $Z_t$  sob um escopo mais amplo. Na prática, modelos ARMA(p,q) são uma alternativa eficiente a modelos puramente de autorregressão ou de médias móveis, por possibilitar uma análise com um número menor de parâmetros. Segundo Morettin e Toloi (2006), modelos ARMA(p,q) podem ser descritos através da seguinte equação:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_p + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

sendo  $a_t, \dots, a_{t-q}$  ruído branco  $RB(0, \sigma^2)$ . Usando o operador autorregressivo de ordem p e o operador de médias móveis de ordem q, é possível reescrever a equação para modelos ARMA(p,q) da seguinte forma:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t$$

#### 2.5.4. Modelos Autorregressivos Integrados de Médias Móveis

É comum ao se analisar uma série temporal que ela não apresente o comportamento estacionário necessário à aplicação dos modelos ARMA(p,q). Uma ferramenta útil para usar na série de forma que o resultado seja estacionário é o operador diferença, definido por Morettin e Toloi (2006) como:

$$\Delta Z_t = Z_t - Z_{t-1}$$

Ou, considerando a n-ésima diferença, tem-se:

$$\Delta^n Z_t = \Delta[\Delta^{n-1} Z_t]$$

Na maior parte dos casos, uma ou duas diferenças são suficientes para se obter a estacionariedade. Este operador é combinado aos modelos ARMA(p,q) para se obter modelos ARIMA(p,d,q), sendo d o número de diferenças aplicadas sobre a série original. Morettin e Toloi (2006) apresentam a seguinte equação genérica para modelos ARIMA(p,d,q):

$$\phi(B)\Delta^d \tilde{Z}_t = \theta(B)a_t$$



### 3. REVISÃO BIBLIOGRÁFICA – CONTROLE ESTATÍSTICO DE PROCESSOS

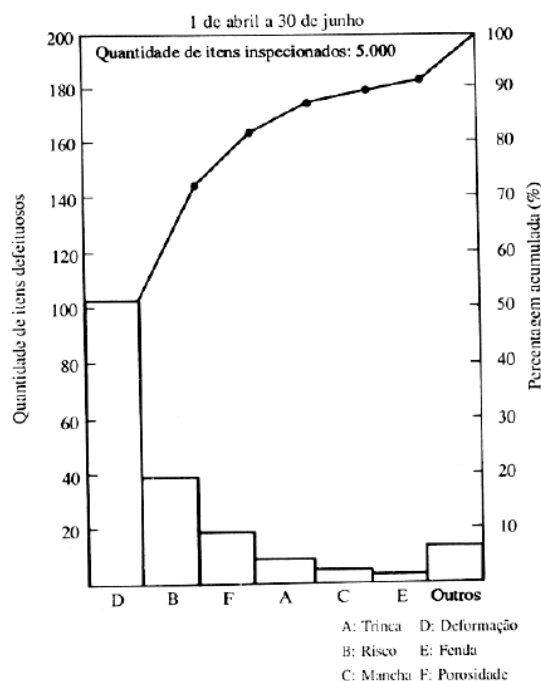
Neste capítulo serão apresentados os conceitos e ferramentas de controle estatístico de processos (CEP) necessários para a concepção de um modelo de monitoramento da série de índices de inadimplência para auxiliar a tomada de decisões no cotidiano da empresa.

CEP é um conjunto de ferramentas estatísticas de controle de qualidade, aplicadas para se avaliar um determinado processo visando aumentar sua estabilidade e previsibilidade, e assim otimizá-lo através da redução da variabilidade.

Esse conjunto é composto por sete ferramentas diferentes. Elas serão brevemente apresentadas a seguir, e ao final será determinada qual é a mais apropriada para os fins propostos neste trabalho.

- a) gráfico de Pareto: é usado para se analisar a distribuição de defeitos em um processo, que possibilita a identificação dos problemas mais recorrentes e uma abordagem mais focada nas causas de maior impacto (vide Figura 04);

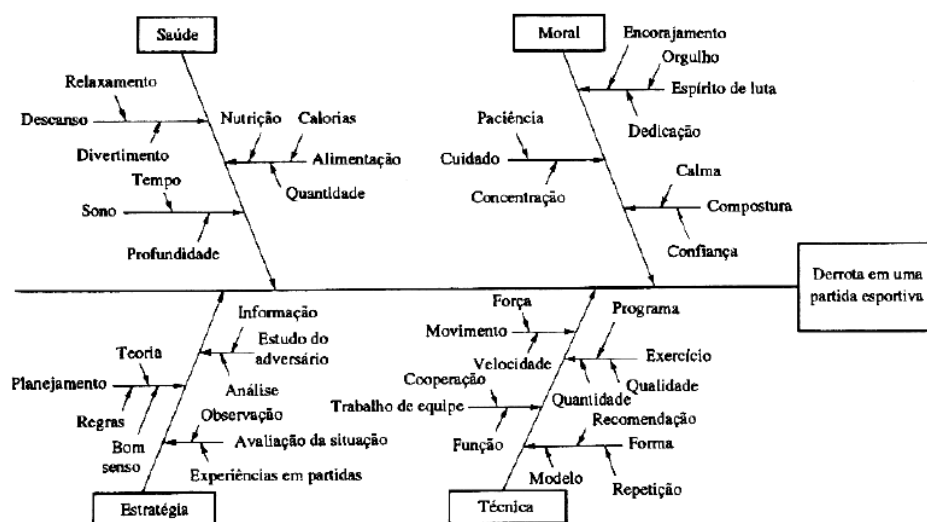
Figura 04 – Gráfico de Pareto



Fonte: Kume (1993)

- b) diagrama de causa-e-efeito: desenvolvido em 1953 por Kaoru Ishikawa, consiste em um diagrama que determina diversas causas, primárias e secundárias, que levam à ocorrência de um determinado efeito ou evento (vide Figura 05);

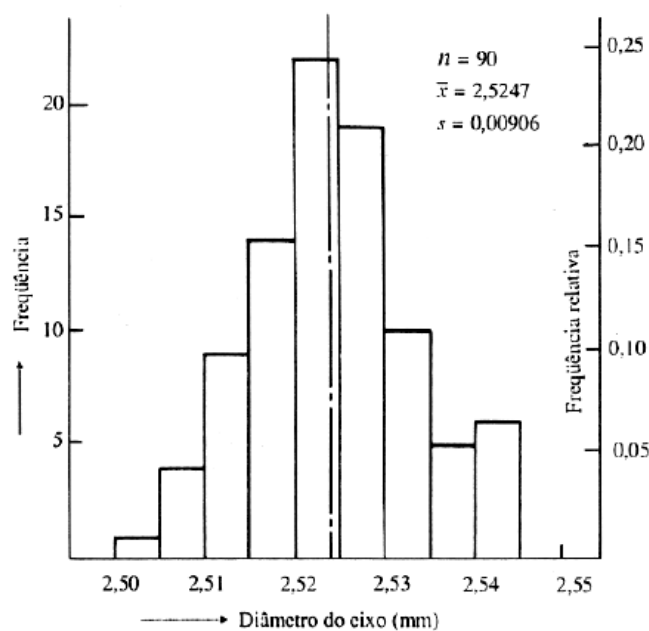
Figura 05 – Diagram de causa e efeito



Fonte: Kume (1993)

- c) histograma: uma forma gráfica de apresentar a distribuição de frequências em uma amostra. Permite uma análise visual rápida para avaliar a média e dispersão dos dados (vide Figura 06);

Figura 06 - Histograma



Fonte: Kume (1993)

- d) folha de verificação: um formulário impresso elaborado de forma a facilitar a coleta de dados durante a observação de um processo, e permitir sua organização de maneira simples (vide Figura 07);

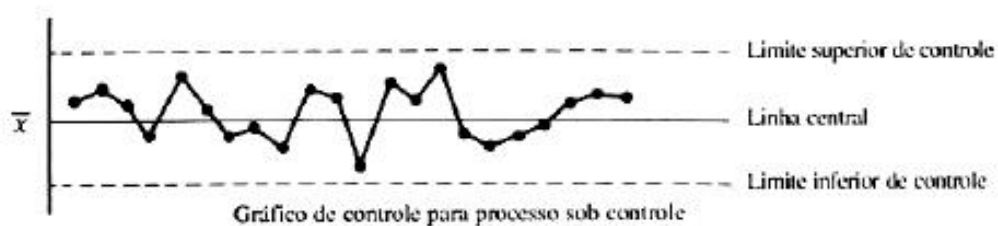
Figura 07 – Folha de verificação

Folha de Verificação		
Produto:	Data:	
Estágio de fabricação: inspeção final	Seção:	
Tipo de defeito: marca, peça incompleta, trinca, deformação	Inspetor:	
Total inspecionado: 1525	Lote nº:	
Observações: todos os itens inspecionados	Pedido nº:	
Defeito	Marca	Sub-Total
Marcas na superfície		17
Trincas		11
Peça Incompleta		26
Deformação		3
Outros		5
	Total:	62
Total Rejeitado	 	42

Fonte: Kume (1993)

- e) gráfico de controle: esta ferramenta busca classificar um processo como dentro ou fora de controle com base em seu comportamento ao longo do tempo e o estabelecimento de limites de controle (vide Figura 08);

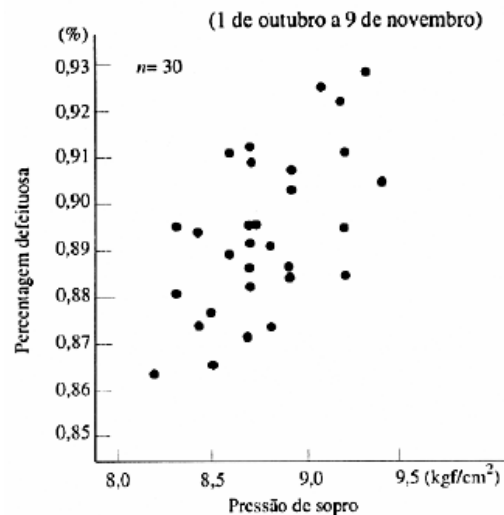
Figura 08 - Gráfico de controle



Fonte: Montgomery (2004)

- f) diagrama de dispersão: ideal para estudar a relação entre duas variáveis, que podem ser características de qualidade ou fatores que as afetam. Seu objetivo é constatar se existe correlação entre os dados, e se houver, classificar esta correlação (vide Figura 09);

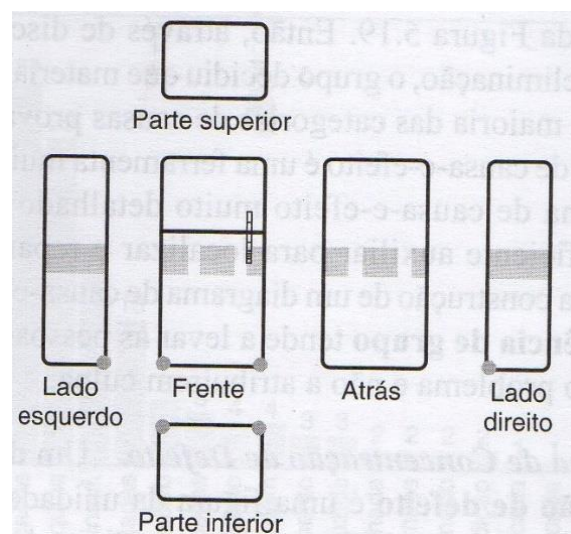
Figura 09 - Diagrama de dispersão



Fonte: Kume (1993)

- g) diagrama de concentração de defeito: representação plana de um determinado objeto, sobre a qual são marcados os locais onde há incidência de defeitos. A partir de sua análise, é possível reconhecer pontos de concentração levantar informações sobre possíveis causas (vide Figura 10).

Figura 10 - Diagrama de concentração de defeito



Fonte: Kume (1993)

Essas ferramentas normalmente são citadas em contexto de análise de processos industriais. Porém, segundo Montgomery (2004), “aplicações não industriais não diferem

substancialmente das aplicações industriais, mais usuais”. Portanto, neste trabalho de formatura, a série de inadimplência será considerada com um processo, a ser avaliado a partir da aplicação de ferramentas de controle da qualidade.

Dentre todas as ferramentas apresentadas, o gráfico de controle é aquele que apresenta a maior utilidade na abordagem do problema proposto, por ser a mais adequada para analisar a variação no decorrer do tempo. Esta ferramenta será apresentada em detalhe nas subseções a seguir.

### 3.1. Gráficos de controle

Kume (1993) apresenta a seguinte definição para gráficos de controle:

**Definição:** Um gráfico de controle consiste em uma linha central, um par de limites de controle, um dos quais localiza-se abaixo e outro acima da linha central, e valores característicos marcados no gráfico representando o estado do processo.

O limite abaixo da linha central (LC) é conhecido como limite inferior de controle (LIC), e o limite acima é chamado de limite superior de controle (LSC). Um processo pode ser considerado sob controle se os valores observados para o processo estiverem contidos entre os dois limites, sem apresentar tendências. Caso contrário, o processo pode ser considerado fora de controle, e é, portanto, passível de ações corretivas para melhorar seus resultados.

É inerente a qualquer processo a existência de variações nas características observadas. Segundo Kume (1993), existem dois tipos de causas para tais variações, as aleatórias, que são aquelas que inevitavelmente acontecerão e são de difícil prevenção, e as assinaláveis, que sinalizam a existência de fatores relevantes que podem ser analisados para buscar a otimização de um processo. Classificar um processo como fora de controle através da aplicação do gráfico de controle é equivalente a dizer que existem uma ou mais causas assinaláveis fora de controle.

Montgomery (2004) apresenta um modelo geral para gráficos de controle. Sendo  $u_w$  uma estatística,  $\mu_w$  sua média e  $\sigma_w$  seu desvio padrão, os limites de controle e a linha central são dados pelas seguintes equações:

$$LSC = \mu_w + L \sigma_w$$

$$LC = \mu_w$$

$$LIC = \mu_w - L \sigma_w$$

nas quais  $L$  representa a distância entre a linha central e os limites de controle, expressa em unidades de desvios padrões. Normalmente seu valor é encontrado a partir de processos iterativos que buscam atender um critério de desempenho pré-determinado de acordo com as especificidades do processo analisado.

Montgomery (2004) afirma que “há uma relação muito próxima entre gráficos de controle e testes de hipótese”. De fato, analisar um processo a partir dos limites de controle do gráfico pode ser análogo a aplicar um teste de hipótese repetidamente para cada período de tempo, com as seguintes hipóteses:

$$H_0: O \text{ processo está sob controle}$$

$$H_1: O \text{ processo está fora de controle}$$

Dessa forma, a hipótese nula é rejeitada caso a estatística calculada esteja fora dos limites de controle, ou apresente algum tipo de tendência em seu comportamento. No entanto, pode-se notar que o tamanho do intervalo entre os limites de controle depende do  $L$  usado no modelo. Assim, o valor de  $L$  é fundamental na decisão do processo como dentro ou fora de controle. Mas como todo processo de decisão, pode ocorrer dois tipos de erros. São eles:

- a) erro tipo I: rejeitar a hipótese nula, enquanto esta é verdadeira. Significa que o modelo aponta a existência de um processo fora de controle, sem que de fato esteja. É conhecido como alarme falso. Montgomery (2004) apresenta a probabilidade de ocorrência  $\alpha$  dada por:

$$\alpha = P\{\text{erro tipo I}\} = P\{\text{rejeitar } H_0 | H_0 \text{ é verdadeira}\};$$

- b) erro tipo II: não rejeitar a hipótese nula, sendo que esta é falsa. O modelo deixa de detectar um processo fora de controle. De acordo com Montgomery (2004), a probabilidade de ocorrência  $\beta$  é dada por:

$$\beta = P\{\text{erro tipo II}\} = P\{\text{deixar de rejeitar } H_0 | H_0 \text{ é falsa}\}.$$

Uma métrica muito usada para avaliar a eficiência um gráfico de controle quanto a ocorrência de erros é o comprimento médio de sequência (ARL, do termo inglês “Average Run Length”). Segundo Montgomery (2004), o ARL é a quantidade média de observações dentro dos limites de controle até a ocorrência de uma observação fora dos limites de controle.

Existem duas medidas de ARL, o  $ARL_0$ , usado para processos sob controle, e o  $ARL_1$ , para processos fora de controle, que são dados por:

$$ARL_0 = \frac{1}{\alpha}$$

$$ARL_1 = \frac{1}{1 - \beta}$$

Como o objetivo é obter o máximo de observações dentro do limite de controle (para um processo sob controle), a situação ideal é aquela em que o  $ARL_0$  é maximizado e o  $ARL_1$  minimizado. No entanto, para se aumentar  $ARL_0$  e obter mais pontos dentro dos limites, é necessário aumentar a distância entre os limites de controle e a linha central. Ao fazer isso, o modelo perde uma parte de sua capacidade de detectar pontos fora de controle, ou seja, aumenta a incidência de erro tipo II. Dessa forma, ao se tentar aumentar o  $ARL_0$ , conseqüentemente há um aumento do  $ARL_1$  também. É preciso então encontrar valores satisfatórios para ambos.

Existem diferentes categorias de gráficos de controle que podem ser utilizadas, cada uma com características próprias que a tornam adequada a um certo tipo de processo. Nas subseções a seguir serão exploradas algumas variedades que podem ser úteis ao monitoramento de índices de inadimplência em carteiras de recebíveis.

### 3.1.1. Gráficos de Controle Shewhart

Os gráficos de Shewhart são amplamente utilizados para controle estatístico de processos, principalmente por ser uma ferramenta de fácil implantação e interpretação, e buscar avaliar a estabilidade de uma determinada variável com a evolução do tempo.

Os gráficos são classificados de acordo com a estatística que está sendo monitorada. A seguir serão apresentados os gráficos de controle mais usados para monitoramento de média e variabilidade do processo.

#### I. Gráfico de controle $\bar{x}$

Este tipo de gráfico de controle é usado para o monitoramento da média dos valores observados em determinada amostra  $m$ , denotada por  $\bar{x}$  e dada por:

$$\bar{x}_m = \frac{x_{m,1} + x_{m,2} + \dots + x_{m,i}}{t}$$

com  $i = 1, \dots, t$ , sendo  $t$  o número de observações na amostra  $m$ .

Montgomery (2004) apresenta os limites de controle como:

$$LSC = \bar{\bar{x}} + A_2 \bar{R}$$

$$LC = \bar{\bar{x}}$$

$$LIC = \bar{\bar{x}} - A_2 \bar{R}$$

onde  $A_2$  é uma constante tabulada para diversos tamanhos de amostra,  $\bar{\bar{x}}$  é a média das observações das  $m$  amostras e  $\bar{R}$  a média das amplitudes das  $m$  amostras, que são dadas por, respectivamente:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m}$$

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m}$$

onde  $R_j, j = 1, \dots, m$  amplitude calculada para a amostra  $j$ , dada por:

$$R_j = \max(x_{j,i}) - \min(x_{j,i})$$

## II. Gráfico da amplitude (R)

Este gráfico usa a amplitude de um processo como estatística a ser controlada, com o intuito de monitorar, e controlar, a variabilidade de um processo. Neste caso, os parâmetros para o gráfico R apresentados por Montgomery (2004) são:

$$LSC = D_4 \bar{R}$$

$$LC = \bar{R}$$

$$LIC = D_3 \bar{R}$$

onde  $D_4$  e  $D_3$  são constantes tabuladas de acordo com o número de observações da amostra.



### III. Gráfico para desvio-padrão

Também conhecido como gráfico S, o gráfico para desvio padrão é utilizado para controlar a variabilidade do processo, assim como os gráficos R, mas a partir do monitoramento do desvio padrão, que pode ser calculado da seguinte forma:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Para elaborar os gráficos de controle, são usados os seguintes limites definidos por Montgomery (2004):

$$LSC = B_4 \bar{S}$$

$$LC = \bar{S}$$

$$LIC = B_3 \bar{S}$$

onde  $B_3$  e  $B_4$  são constantes tabuladas para diversos tamanhos de amostras e  $\bar{S}$  é o desvio padrão médio das  $m$  amostras.

#### 3.1.2. Gráficos de Controle de Soma Cumulativa (CUSUM)

Os gráficos CUSUM foram introduzidos por Page (1954), e são classificados como gráficos com memória. Isso significa que, diferente dos gráficos de Shewhart, ele considera informações fornecidas por toda a sequência de observações, não apenas a mais recente. Esta característica permite que gráficos com memória sejam mais sensíveis a pequenas variações.

O processo de soma cumulativa pode ser aplicado a diversas variáveis, mas nesta subseção será abordado apenas o caso do CUSUM tabular para monitoramento de média de processos. Basicamente, considerando a média do processo  $\mu_0$  como valor alvo, é calculado a soma total de todos os desvios acima deste valor, o que resulta na estatística CUSUM unilateral superior  $C^+$ . Analogamente, para os desvios abaixo do valor alvo é calculada a estatística CUSUM unilateral inferior  $C^-$ . Montgomery (2004) apresenta a seguinte forma de cálculo destas estatísticas:

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+]$$

$$C_i^- = \max[0, (\mu_o + K) - x_i + C_{i-1}^-]$$

onde  $C_i^+$  e  $C_i^-$  são os valores da estatística acumulados até a observação  $i$ , sendo que  $C_o^+ = C_o^- = 0$ , e  $K$  é o valor de referência, ou valor de folga. Seu valor é calculado como a metade da distância entre  $\mu_o$  e o valor da média fora de controle  $\mu_1$ :

$$K = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_o|}{2}$$

Para que seja classificado como fora de controle, as estatísticas  $C_i^+$  e  $C_i^-$  devem exceder o intervalo de decisão  $H$ . Este intervalo é definido a partir do valor de ARL que se deseja obter, através da testagem de diversos cenários.

### 3.1.3. Gráficos de Controle de Média Móvel Exponencialmente Ponderada (MMEP)

O gráfico MMEP apresentado por Roberts (1959), assim como o modelo CUSUM, é classificado como gráfico com memória. É normalmente aplicado aos casos em que as observações feitas são individuais e é regularmente lembrado por ser de fácil aplicação. Segundo Montgomery, Jennings e Kulachi (2008), este tipo de gráfico é muito utilizado para previsões e modelos de séries temporais.

O primeiro passo da construção da estatística  $z_i$ , para a qual Samohyl (2009) apresenta a seguinte equação:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}$$

onde  $i$  representa o período de tempo,  $x_i$  é o valor da série observado em  $i$  e  $\lambda$  é uma constante, sendo que  $0 < \lambda \leq 1$ . Para  $i = 0$ , é considerado  $z_0 = \mu_o$ , a média dos valores observados.

Samohyl (2009) aponta que como  $z_i$  é calculado em função de  $x_i$  e  $z_{i-1}$ , a partir de uma série de substituições sequenciais para valores passados é possível observar que o MMEP calcula o valor atual da estatística com base em todos valores passados, com as observações mais recentes tendo um peso maior no cálculo da média.

A partir do cálculo da estatística  $z_i$  para todos valores de  $i$  do processo analisado, Montgomery (2004) apresenta as seguintes formas de cálculo dos parâmetros LC, LSC e LIC do gráfico de controle:

$$LSC = \mu_o + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$

$$LC = \mu_o$$

$$LIC = \mu_o - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$

Sendo  $\sigma$  o desvio padrão dos valores observados e  $\mu_o$  o valor alvo do processo. Tanto os valores de  $\lambda$  quanto  $L$  devem ser selecionados de acordo com o ARL que se deseja alcançar. Vale ressaltar aqui, que a partir da equação da estatística  $z_i$  percebe-se que um  $\lambda$  menor significa um maior peso atribuído aos dados históricos do processo, o que possibilita uma detecção mais eficaz de pequenas variações. Se for usado  $\lambda = 1$ , o gráfico MMEP é equivalente a um gráfico de Shewhart.



## **4. METODOLOGIA**

Neste capítulo será explicada detalhadamente a metodologia aplicada para se obter um modelo de previsão da série e como será feito seu monitoramento a partir de controle estatístico de processos e gráficos de controle.

### **4.1. Obtenção dos dados**

O primeiro passo do trabalho será a coleta de dados para os quais se busca construir o modelo de monitoramento. Um ponto importante a ser considerado é a confiabilidade das informações. Para o modelo se provar útil e cumprir seu propósito, é necessário que ele seja alimentado com informações que realmente reflitam a situação real. Esse ponto se torna ainda mais importante considerando a complexidade de se realizar o controle de uma carteira de recebíveis. Para se calcular a inadimplência, é necessário possuir, para todas as parcelas de todos os devedores, as seguintes informações:

- a) data de Vencimento;
- b) valor da parcela;
- c) data de pagamento;
- d) valor pago.

Uma operação de securitização pode envolver diversos empreendimentos, que podem totalizar milhares de devedores, cada um com dezenas de parcelas a pagar. Para se calcular o índice de inadimplência, é necessário apurar todos pagamentos recebidos, associar cada um desses pagamentos a um devedor específico e comparar o valor pago com o valor da parcela, e a data de pagamento com a data de vencimento.

Para realizar esse controle, normalmente a securitizadora contrata um prestador de serviço denominado Servicer, que realiza um monitoramento mensal dos pagamentos recebidos. O Servicer então condensa todas essas informações em uma planilha de uma forma apropriada para tratamento dos dados e análise. Dessa forma, ao se utilizar este relatório pode-se contar com informações confiáveis e de fácil uso.

No entanto, este relatório apresenta apenas uma relação dos pagamentos devidos e recebidos, não contemplando o índice de inadimplência em si, que deverá ser calculado pela seguinte fórmula:

$$\text{Inadimplência no Mês} = 1 - \frac{\text{Pagamentos Recebidos em Dia}}{\text{Pagamentos Esperados} - \text{Pagamentos Antecipados}}$$

sendo os pagamentos expressos em seu valor em reais.

Assim, ao final desta etapa, deve-se estar em posse de uma planilha com o cálculo deste índice para cada mês desde o início das vendas do empreendimento, para poder contemplar a evolução dessa métrica ao longo do tempo, obtendo assim uma série temporal.

## 4.2. Análise e ajuste da série

O objetivo desta seção é aplicar métodos e conceitos para que a série analisada possua as características necessárias ao uso de modelos paramétricos lineares. Os códigos desenvolvidos para esta subseção, bem como para as seções 4.3 e 4.4, estão apresentados no Apêndice A. Foram considerados os seguintes passos:

Passo 1: uma análise visual da série. Caso seja observada instabilidade na variância ao longo do tempo, será aplicada a transformação proposta por Box e Cox (1964), introduzida na seção 2.2 deste trabalho, para que se estabilize a variância. Conforme explicado, para aplicar esse método é necessário estimar os parâmetros  $\gamma$  e  $c$ . Estes serão definidos a partir da função `boxcox.lambda()` do RStudio, que fornece os valores que minimizam a variação da variância.

Passo 2: avaliar a estacionariedade da série. Para isso, será aplicado o teste Dickey-Fuller aumentado. Como output, o programa retorna um valor  $p - value$ , e caso ele seja menor que 0,05, pode-se rejeitar a hipótese nula e assumir estacionariedade. Caso seja verificada a não estacionariedade da série, serão tomadas quantas diferenças forem necessárias até ela se torne estacionária. No entanto, Morettin e Tolo (2006) defendem que na maioria dos casos 1 ou 2 diferenças costumam ser suficientes para obter uma série estacionária.

Passo 3: é avaliado se a série possui uma distribuição aproximadamente igual a normal com a aplicação do teste de Shapiro-Wilk, através da função `shapiro.test()` do RStudio, que já calcula a variável  $b$  adequada. Esta função retorna o valor  $p - value$ , e caso este seja superior a 0,05, pode-se aceitar a hipótese nula e concluir que a série é gaussiana. Ainda para analisar a

normalidade, será elaborado o gráfico Q-Q Plot com os pontos da série, e será feita uma análise visual deste. Se os pontos estiverem próximos da reta que representa a distribuição normal, pode-se assumir normalidade. Caso não haja normalidade, deverão ser aplicadas transformações adicionais para corrigir a série.

Passo 4: finalmente será avaliada a independência da série, através da análise das autocorrelações e autocorrelações parciais, usando as funções *fac()* e *facp()* do RStudio, respectivamente. Essas funções serão aplicadas até o lag 50, para se avaliar um intervalo suficientemente longo e perceber a existência de todas autocorrelações relevantes. O intervalo limite que divide as autocorrelações relevantes das demais já é calculado pelo programa, e mostrado visualmente junto aos valores encontrados para facilitar a análise.

### 4.3. Identificação do modelo

Uma vez analisadas as características da série e aplicados os devidos ajustes, parte-se então para a etapa de identificação do modelo paramétrico que se adequa melhor a ela. Uma metodologia que serve de referência neste campo é a abordagem iterativa em 4 estágios de Box e Jenkins (1976):

- a) especificação: uma classe geral de modelos é considerada para análise;
- b) identificação: com base nas autocorrelações, nas autocorrelações parciais e outros critérios, identificar um candidato de modelo;
- c) estimação: determinar os parâmetros para o modelo candidato;
- d) diagnóstico/verificação: análise de resíduos para determinar se o modelo ajustado é adequado para os fins que se procura atingir.

Durante as etapas de estimação ou diagnóstico, caso os resultados obtidos não sejam satisfatórios, volta-se para a etapa de identificação, seguindo desta forma até se achar um modelo adequado. Nesta subseção serão abordados os três primeiros passos.

Em relação à especificação, aqui serão utilizados modelos autorregressivos integrados de média móveis ARIMA(p,d,q), por ser uma classe mais ampla que engloba tanto os casos de modelos AR(p), MA(q) e ARMA(p,q).

A etapa de identificação é uma das mais trabalhosas e importantes de todo o processo. Morettin e Toloi (2006) citam as seguintes características da autocorrelação de uma série que auxiliam na identificação de um modelo adequado:

- a) um processo AR(p) tem fac que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- b) um processo MA(q) tem fac finita, no sentido em que ela apresenta um corte após o “lag” q;
- c) um processo ARMA(p,q) tem fac infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o “lag” p-q.

Da mesma forma, Morettin e Toloi (2006) apresentam as características das autocorrelações parciais para cada classe de modelos:

- a) Um processo AR(p) tem facp  $\phi_{kk} \neq 0$ , para  $k \leq p$  e  $\phi_{kk} = 0$  para  $k > p$ ;
- b) Um processo MA(q) tem facp que se comporta de maneira similar à fac de um processo AR(p): é denominada por exponenciais e/ou senóides amortecidas;
- c) Um processo ARMA(p,q) tem facp que se comporta como a facp de um processo MA puro.

Devido à sua maior complexidade, a identificação de modelos ARMA pode ser difícil através da análise de sua fac e facp. Assim, neste caso, é recomendado o teste de uma variedade de modelos, preferencialmente com número reduzido de parâmetros, e comparação deles através de métodos pré-estabelecidos. Akaike (1973) indica como mais adequado o modelo ARIMA(k,i,l), onde  $k$ ,  $i$  e  $l$  são as ordens com as quais o seguinte critério apresenta seu valor mínimo:

$$AIC(k, d, l) = N \ln \hat{\sigma}_a^2 + \frac{N}{N-d} 2(k + l + 1 + \delta_{d0})$$

em que:

$$\delta_{d0} = \begin{cases} 1, & d = 0 \\ 0; & d \neq 0 \end{cases}$$



sendo  $N$  o número de observações na amostra e  $\hat{\sigma}_a^2$  o estimador da máxima verossimilhança de  $\sigma_a^2$ , definido por:

$$\hat{\sigma}_a^2 = \frac{\sum_{i=1}^n (x_i - z_i)^2}{n}$$

sendo  $x_i$  os valores das observações da série e  $z_i$  os valores obtidos através do modelo. Assim, o melhor modelo deverá ter o menor valor do AIC.

Importante ressaltar que até agora foi abordado apenas o caso dos modelos cheios, que são os modelos de ordem  $p$  e  $q$  que possuem  $p + q + 1$  parâmetros diferentes de zero. O caso ideal é aquele em que se acha um modelo com um número mínimo de parâmetros. Assim, é interessante estudar também os chamados modelos reduzidos, que são aqueles de ordem  $p$  e  $q$ , mas em que apenas uma fração dos  $p + q + 1$  parâmetros são diferentes de zero. Para se validar o uso do modelo reduzido ao invés do completo, utiliza-se o seguinte teste de hipótese:

$H_0$ : o modelo reduzido é equivalente ao completo

$H_1$ : o modelo reduzido não é equivalente ao completo

Para tal, são comparados o valor  $\Delta AIC$ , diferença entre o critério de Akaike dos modelos completo e reduzido, e o valor de  $\chi^2$ , que é o valor atribuído para a distribuição qui-quadrado para um grau de liberdade igual a diferença entre o número de parâmetros nos modelos cheio e reduzido, e um  $\alpha = 0,05$ . Caso  $\Delta AIC < \chi^2$ , a hipótese nula é tomada como verdadeira e segue-se a análise com o modelo reduzido.

Outro critério importante para avaliação dos modelos é a precisão de suas projeções. Para tal avaliação, é usada apenas uma porção da série de dados para definição do modelo paramétrico. Por exemplo, em uma série de 1.000 observações, a modelagem seria feita a partir dos 900 primeiros. Uma vez que o modelo é encontrado e todos seus parâmetros são calculados, é feita uma projeção dos valores futuros da série. Estas são então comparadas aos valores reais para determinar a sua precisão.

Sendo  $\gamma_t$  a observação medida em  $t$  e  $F_t$  a previsão para  $\gamma_t$ , o erro da previsão é dado por  $e_t = \gamma_t - F_t$ . Hyndman e Koehler (2005) apresentam critérios de avaliação amplamente utilizados, calculados a partir dessas variáveis: a média do erro ao quadrado MSE (do inglês Mean Square Error) e a raiz da média do erro ao quadrado RMSE (do inglês Root Mean Square Error), que são dados por:

$$MSE = \frac{e_1^2 + e_2^2 + \dots + e_f^2}{f}$$

$$RMSE = \sqrt{MSE}$$

sendo  $f$  o número de projeções a serem comparadas com a série original. Os modelos mais apropriados serão aqueles que minimizem os resultados de MSE e RMSE, dada a série de inadimplência utilizada.

Uma vez descobertas a classe e as ordens do modelo, o próximo passo é estimar seus parâmetros. Para isso será utilizada a função *arima()* do RStudio. Eles são obtidos através do método dos mínimos quadrados. Em suma, este método consiste em buscar os parâmetros que registrem os menores erros entre os valores encontrados pelo modelo e os valores experimentais.

#### 4.4. Diagnóstico do modelo

Neste momento será aplicado o quarto passo do método de Box e Jenkins (1976), o diagnóstico do modelo. Aqui serão realizados uma série de testes para verificar se ele é adequado para prever os valores futuros da série. Mais especificamente, são analisados os resíduos da série, que é a diferença entre os valores experimentais e aqueles encontrados pelo modelo. Sendo  $Z_t$  uma série temporal, e  $W_t = \Delta Z_t$ , suponha o modelo:

$$\phi(B)W_t = \theta(B)a_t$$

Morettin e Toloi (2006) definem o erro verdadeiro desse modelo como:

$$a_t = \theta^{-1}(B)\phi(B)W_t$$

Segundo os autores, caso o modelo seja adequado, esses erros irão se comportar como ruídos brancos. Isso portanto servirá de critério para sua aprovação. Caso este se prove inadequado, volta-se para a etapa de identificação do modelo.

O primeiro teste a ser aplicado é o de Ljung e Box (1978). Esse método se baseia em um teste de hipótese, onde:

$H_0$ : os resíduos são independentes e identicamente distribuídos

$H_1$ : os resíduos não são independentes e identicamente distribuídos

Para isso, é calculada a seguinte estatística:

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)}$$

na qual  $K$  é o valor do lag analisado,  $n$  é igual o número de observações menos o número de diferenças tomadas, e  $\hat{r}_j$  é a autocorrelação do resíduo  $\hat{a}_t$ , dada por:

$$\hat{r}_j = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2}$$

Considerando um nível de confiança  $\alpha$ , a hipótese nula  $H_0$  é rejeitada se o valor de  $Q(K)$  for superior ao valor atribuído a distribuição qui-quadrado para  $\alpha$  e grau de liberdade igual a  $K - p - q$ . Assim, o modelo será aprovado no diagnóstico caso a hipótese nula não seja rejeitada.

Mesmo que haja resultados positivos na aplicação do teste de Ljung e Box (1978), serão aplicados testes adicionais para se garantir que o modelo é apropriado com maior rigor. O segundo teste então será o de Shapiro-Wilk, para verificar se os resíduos apresentam distribuição próxima da normal. Novamente, será assumida normalidade caso o programa retorne um  $p - value$  superior a 0,05.

Por último serão avaliadas a  $fac$  e  $facp$  dos resíduos. Uma característica de ruídos brancos é que as observações em diferentes momentos não estão correlacionadas. Assim, para aprovar o modelo, é preciso que a análise das duas funções não mostre autocorrelações fora dos intervalos de confiança estabelecidos. Caso mostre, significa que há informação relevante contida nos resíduos que não está sendo captada pelo modelo.

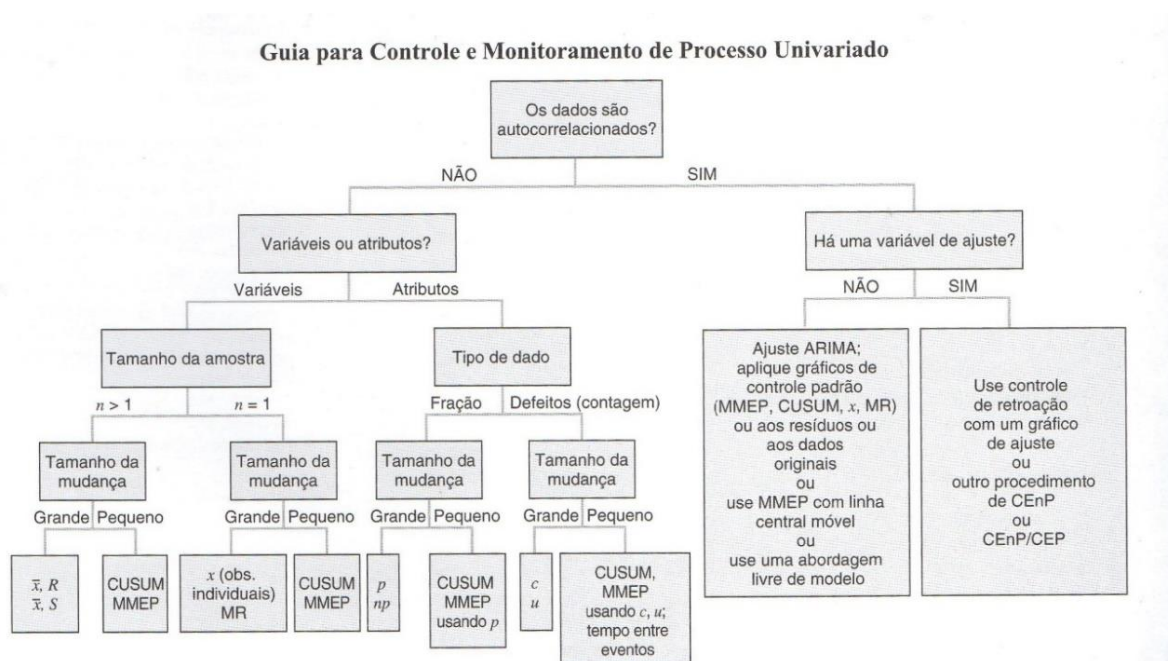
Tendo resultados favoráveis nos três testes, pode-se assumir que o modelo encontrado é adequado e avançar para a próxima etapa. Caso qualquer um dos três tenha resultado negativo, não se pode assumir que os resíduos se comportam como ruídos brancos e deve-se selecionar outro modelo encontrado na fase de identificação, e submetê-lo ao diagnóstico aqui explicado.

#### 4.5. Escolha do gráfico de controle

Nesta etapa, será selecionada a melhor opção de gráfico de controle dentre as alternativas apresentadas na subseção 3.1, tendo em mente o propósito deste trabalho e o modelo paramétrico encontrado durante a análise da série de inadimplência.

Para esta escolha, será utilizado o fluxograma apresentado na Figura 11, proposto por Montgomery (2004).

Figura 11 - Fluxograma de seleção de gráfico de controle



Fonte: Montgomery (2004)

Logo, o primeiro passo será avaliar a função de autocorrelação, conforme indicado na seção 4.2, e a partir dela serão avaliados os demais ramos do fluxograma até se encontrar a alternativa mais adequada.

#### 4.6. Definição dos limites de controle

Como será relatado em detalhes na seção 5.5, o gráfico de controle escolhido através de fluxograma foi o de média móvel exponencialmente ponderada. Portanto, nesta seção serão apresentados os procedimentos a serem adotados para definição de todos parâmetro necessários à aplicação do MMEP.

O ponto de partida para definição dos limites de controle é definir um valor alvo para o  $ARL_0$ . Esta variável indica o intervalo médio de observações entre alarmes falsos em um

processo sob controle, e, portanto, deve ser adequada ao contexto em que se busca aplicar o gráfico de controle.

Para servir de referência, será apresentado o caso do gráfico  $\bar{x}$  com limite de três sigmas (desvio padrão da estatística), apresentado por Montgomery (2004). Neste caso, o  $ARL_0$  é calculado como o inverso da probabilidade  $p$  de um ponto estar fora dos limites de controle. Para este modelo de limite de três sigmas, tem-se que  $p = 0,0027$ , logo:

$$ARL_0 = \frac{1}{p} \cong 370$$

Esse é o valor considerado padrão e serve como ponto de partida para outras análises. No entanto, considerando que os dados de inadimplência são obtidos em uma frequência mensal, um  $ARL_0$  igual a 370 significaria um alarme falso a cada 30 anos, um período muito longo considerando o horizonte de observação da amostra usada. E como foi explorado na seção 3.1., ao se aumentar demais o  $ARL_0$  há uma redução do  $ARL_1$ , o que prejudica a capacidade do gráfico de detectar observações fora de controle.

Assim, conclui-se que devido às características da série é necessário adotar um  $ARL_0$  significativamente menor. Será adotado então  $ARL_0$  igual a 36 meses, ou 3 anos. Dado o contexto em que aumentos na inadimplência podem gerar perdas para investidores, e como elas podem se manifestar em curtos espaços de tempo, é preferível que seja considerado um intervalo menor entre alarmes falsos em prol de uma maior capacidade de detecção de pontos fora de controle.

Os próximos passos então consistem na determinação do fator de suavização  $\lambda$  e, finalmente, a definição dos limites de controle, através de um processo iterativo de simulações que será explicado em detalhes nas próximas subseções.

#### 4.6.1. Cálculo do Fator de Suavização $\lambda$

Conforme explicado na subseção 3.1.3, os gráficos da classe MMEP são definidos por:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}$$

Os valores de  $x_i$  são as observações da série de inadimplência, faltando assim definir apenas o parâmetro  $\lambda$  para que se possa calcular as estatísticas  $z_i$ .

Hunter (1986) sugere o método de mínima soma dos quadrados dos erros para determinar  $\lambda$ . Primeiramente, é adotado  $\hat{x}_{i+1}$  como a predição feita no final do período  $i$  para a observação que será verificada no período  $i + 1$ . Define-se então:

$$\hat{x}_{i+1} = z_i$$

Ou seja, a estatística  $z_i$  é equivalente à predição feita no período  $i$  para a observação que ocorrerá em  $i + 1$ . Pode-se então calcular o erro do modelo através da diferença entre a predição da observação para o período  $i$  e o valor de fato observado:

$$e_i = x_i - \hat{x}_i$$

que também pode ser descrita como:

$$e_i = x_i - z_{i-1}$$

Portanto, através da ferramenta RStudio, será desenvolvido um processo iterativo para definir um  $\lambda$  ótimo, cujo código está apresentado no Apêndice B. Inicialmente serão calculados os valores de  $z_i$  da série e seus respectivos erros para um determinado valor de  $\lambda$ , e então será obtida a soma dos quadrados dos erros. Esse processo será repetido para diversos valores de  $\lambda$ , e o valor da soma dos quadrados dos erros anotado. Aquele  $\lambda$  que minimizar esta soma será o valor ótimo adotado para o MMEP. Para testar uma variedade significativa de valores, será usado um valor inicial de  $\lambda$  de 0,01 e um passo de 0,005, que resulta em 198 iterações.

#### 4.6.2. Definição ARL sob Controle

Após a definição do parâmetro  $\lambda$  e do valor de  $ARL_0$ , o último passo para implementação do gráfico é definir seus limites de controle. Como apresentado na subseção 3.1.3., os limites de controle e a linha central do MMEP são calculados por:

$$LSC = \mu_o + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$

$$LC = \mu_o$$

$$LIC = \mu_o - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$

Em relação a linha central, o valor alvo do processo  $\mu_o$  é específico para cada operação de securitização. Durante a análise, são verificados os níveis atuais de inadimplência e a modelagem é feita a partir deles. Dessa forma, o valor de  $\mu_o$  normalmente corresponde a média do processo observada até então.

A partir da análise da série temporal será determinado seu desvio padrão, de modo que a última variável a ser definida é o  $L$ . Para tal será aplicado um algoritmo desenvolvido com a ferramenta RStudio, apresentado no Apêndice C.

O primeiro passo consiste em obter valores da série simulada  $x_t$ . Para isso será usado o modelo paramétrico a ser encontrado pelo procedimento descrito na seção 4.3. A partir deste modelo, serão introduzidos valores de resíduos  $a_t$  para se obter uma simulação de  $x_t$ . Segundo a bibliografia apresentada,  $a_t$  precisa se comportar como ruído branco em processo sob controle. Para adaptar o modelo à série de inadimplência estudada, os valores de  $a_t$  usados serão gerados a partir da variância amostral dos resíduos  $\sigma_R^2$  do modelo paramétrico ajustado, ou seja,  $a_t \sim N(0, \sqrt{\sigma_R^2})$ . Em seguida, com os valores da série, serão calculados os valores da estatística  $z_t$  do MMEP.

Tendo definido os valores de  $z_t$ , serão calculados os limites de controle para cada período  $t$  analisado. O programa irá seguir comparando os valores de  $z_t$  com os limites de controle, com o intuito de obter o intervalo de observações entre a primeira observação e a primeira ocorrência de uma observação fora dos limites. Quando esse valor para o comprimento de sequência (o número de observações até que haja um ponto fora de controle) for obtido, ele é anotado e se inicia uma nova simulação com os mesmos parâmetros. Serão obtidos  $N = 2.000$  comprimentos de sequência para cada iteração. Ao fim das simulações, são somados todos os comprimentos de sequência, e essa soma é dividida pelo número de simulações, resultando no valor de ARL médio para estas condições.

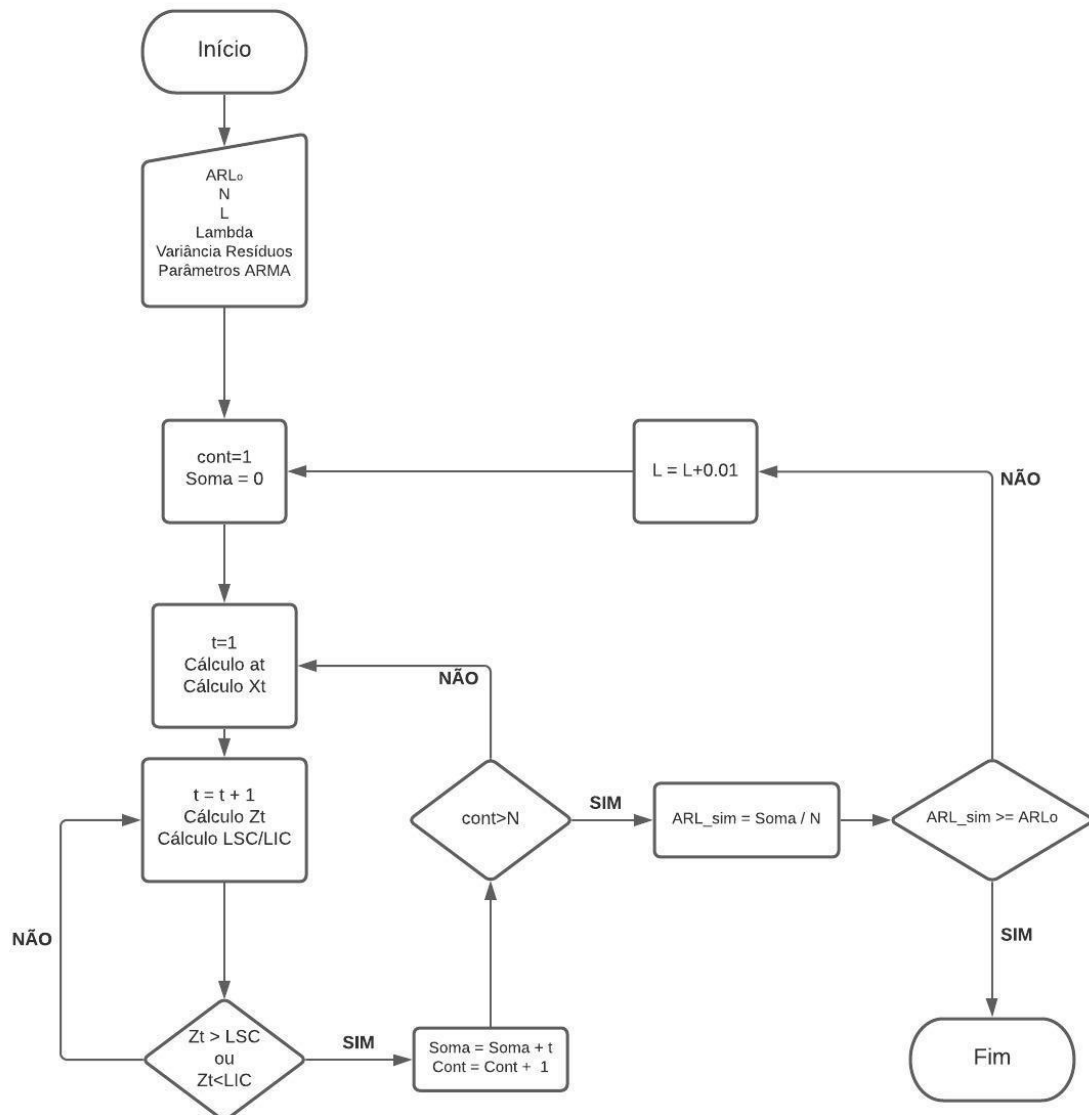
O processo citado acima é então iterado para diferentes valores de  $L$ , até que o ARL encontrado seja menor que o  $ARL_0$  pré-definido. Para apresentar o processo de uma forma mais estruturada, foi elaborado o fluxograma apresentado na Figura 12. Segue um breve resumo das variáveis envolvidas para orientar a leitura do fluxograma:

- a)  $ARL_0$ : valor alvo do ARL, é usado como referência para definir o fim das simulações;

- b)  $N$ : número de simulações a serem feitas para cada valor de  $L$ ;
- c)  $X_t$ : lista contendo os valores simulados da série, obtida através do modelo ARIMA ajustado e dos valores de  $a_t \sim N(0, \sqrt{\sigma_R^2})$ ;
- d)  $Z_t$ : valor da estatística para cálculo do MMEP. Utiliza o parâmetro  $\lambda$  calculado na subseção 4.6.1. Será usado  $z_0$  igual a média dos valores de  $X_t$ ;
- e)  $LSC$ : valor para o limite superior de controle, calculado a partir da equação apresentada no início desta subseção;
- f)  $LIC$ : valor para o limite inferior de controle, calculado a partir da equação apresentada no início desta subseção;
- g)  $L$ : parâmetro usado para cálculo dos limites de controle e valor que se deseja encontrar. Começa como 1 e vai sendo incrementado em 0,01 por iteração;
- h)  $Cont$ : variável para acompanhar o número de simulações feitas para cada valor de  $L$ , varia de 1 a 2.000;
- i)  $t$ : variável que acompanha o número de observações analisadas;
- j)  $Soma$ : soma dos valores de até uma ocorrência fora dos limites de controle nas  $N$  simulações;
- k)  $ARL_{sim}$ : ARL encontrado para a simulação atual pela divisão de Soma por  $N$ .



Figura 12 - Fluxograma do algoritmo de definição do ARL em processo sob controle



Fonte: Elaborado pelo autor

#### 4.6.3. Determinação ARL Fora de Controle

Com os limites de controle definidos, será calculado o  $ARL_1$  do gráfico de controle elaborado, através da simulação de um processo fora de controle. Retomando a bibliografia, o  $ARL_1$  representa o número médio de observações até que seja verificada uma observação fora dos limites, para processos fora de controle.

Para isso, foi novamente desenvolvido um algoritmo iterativo, muito semelhante ao apresentado na subseção 4.6.2. Em suma, o algoritmo irá simular uma série temporal a partir

do modelo paramétrico ajustado durante a etapa de identificação do modelo. Serão simulados diversos valores de  $x_t$  a partir de resíduos  $a_t$ , e nesse ponto há uma diferença relevante em relação ao cálculo do  $ARL_0$ . Para simular um processo fora de controle, a média dos resíduos será incrementada por 10% do desvio padrão por vez, isto é,  $a_t \sim N(c \cdot \sqrt{\sigma_R^2}, \sqrt{\sigma_R^2})$ , com  $c = 0, \dots, 1,5$ .

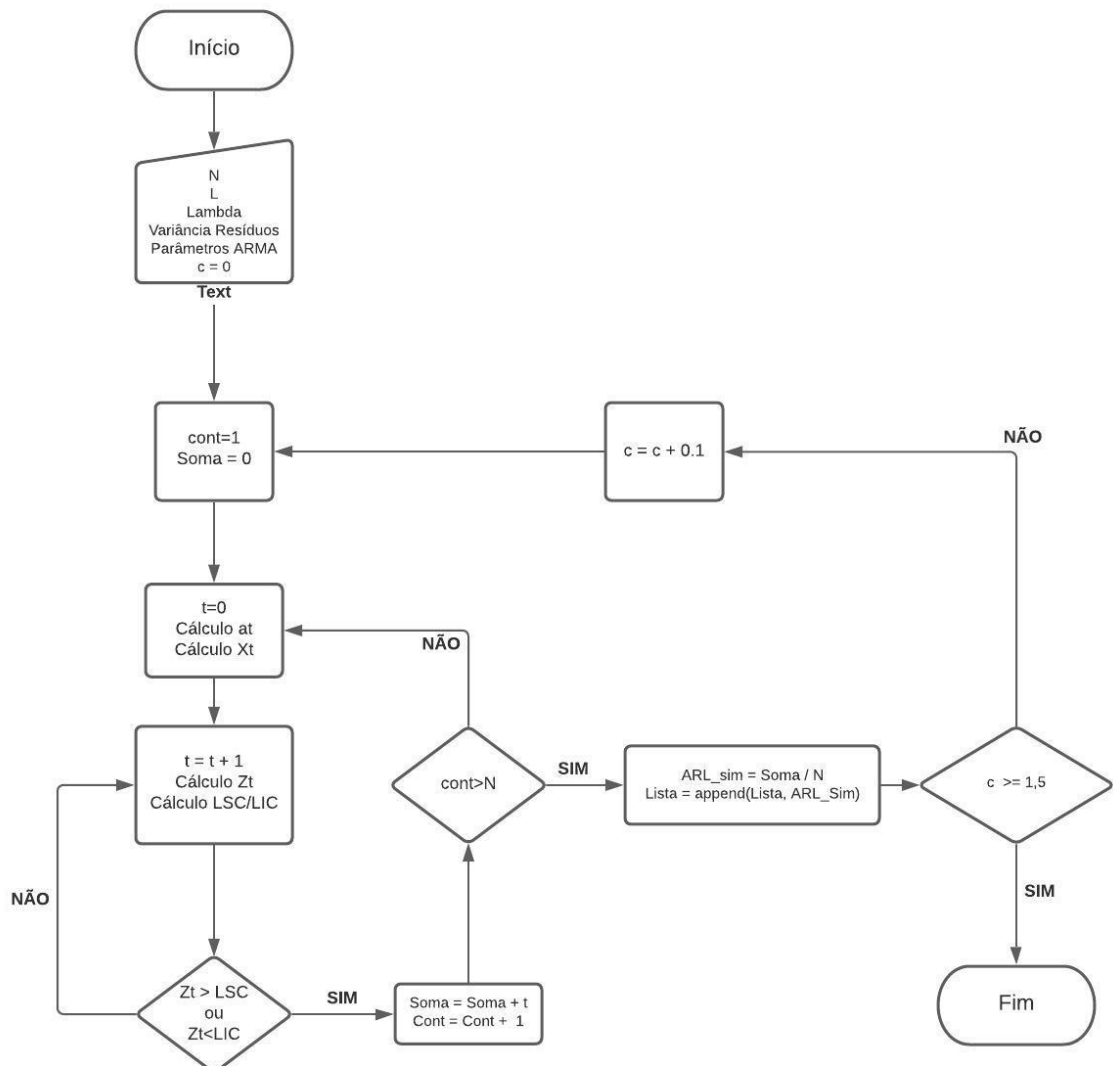
Com esses dados, serão calculadas as estatísticas  $z_t$  (com o fator de suavização ótimo encontrado) e os limites de controle para cada período  $t$ . Os dois serão comparados para cada período  $t$  até que um ponto seja detectado fora do limite de controle. Será então anotado o número de ocorrências até que haja um alarme, e o processo será repetido para estes parâmetros  $N = 2.000$  vezes. Ao fim das simulações, o  $ARL_1$  é calculado como a média desses valores. Esse processo então é iterado até que  $c = 1.5$ , a partir de incrementos de 0,1 por iteração.

Diferente do algoritmo para definição do  $ARL_0$ , o objetivo deste não é encontrar um valor fixo para a variável a ser iterada, neste caso o fator  $c$ , mas sim realizar um estudo de sensibilidade sobre o efeito de um crescimento na média dos resíduos sobre a habilidade do gráfico de detectar pontos fora de controle em processos que não estão sob controle.

A Figura 13 mostra o fluxograma que explica a lógica do algoritmo, cujo código está apresentado no Apêndice D. Em adição as variáveis apresentadas no algoritmo do  $ARL_0$ , serão usadas neste caso:

- a)  $c$ : fator de incremento da média dos resíduos. Irá variar entre 0 e 1,5;
- b) *Lista*: lista que registra o  $ARL$  encontrado para cada valor de  $c$ .

Figura 13 - Fluxograma do algoritmo de definição do ARL em processo fora de controle



Fonte: Elaborado pelo autor

#### 4.7. Avaliação dos Resultados

O objetivo final desse trabalho é o desenvolvimento de um modelo paramétrico que possa projetar valores de séries de inadimplência com base em seu histórico, e a aplicação de um gráfico de controle para que se avalie o comportamento dessa série, que aqui será vista como um processo, e aponte a necessidade de intervenção quando estiver fora de controle. Com esses objetivos em mente, a avaliação dos resultados obtidos será dividida em três etapas:

- a) **Modelo Paramétrico:** a série de inadimplência que será utilizada neste trabalho tem um total de 175 observações, correspondente a um período de 15,4 anos, lembrando que a frequência de observações é mensal. No entanto, para o desenvolvimento do

modelo, serão usadas apenas as 155 primeiras observações. Portanto, haverá 20 observações que não foram introduzidas no modelo, mas deverão ser previstas por ele. A primeira avaliação será feita pelo cálculo do erro médio entre as projeções e os valores reais. Com essas 20 observações será possível avaliar a capacidade de projeção em um horizonte de 1 anos e 8 meses, que já seria de grande utilidade à empresa;

- b) Gráfico de controle: a eficiência do gráfico desenvolvido será julgada através da capacidade de detectar um processo fora de controle, com base no histórico conhecido dessa série em particular. É importante que a partir dele seja possível diferenciar variações aleatórias das assinaláveis, que são aquelas passíveis de medidas preventivas ou corretivas. Como os recebíveis de carteira imobiliária geralmente representam quantias elevadas de capital, é interessante que o gráfico seja capaz de detectar pequenas variações na inadimplência, pois estas podem representar variações relevantes no montante de dinheiro a ser recebido. Serão analisadas as observações fora dos limites de controle e comparado com o histórico de inadimplência para constatar se correspondem a períodos em que a série apresentou um comportamento atípico;
- c) Após ser constatado que ambas as etapas acima atendem o seu propósito, será feita uma comparação da aplicação da metodologia desenvolvida em relação à situação atual da empresa. Esta última etapa, de caráter mais qualitativo, visa esclarecer quais benefícios a Fortesec obterá com a uso de séries temporais e gráficos de controle em suas operações e validar a adoção da solução proposta.

## 5. APLICAÇÃO SOBRE O PROBLEMA PROPOSTO

Neste capítulo será apresentada a aplicação da metodologia desenvolvida no capítulo anterior para o caso da Fortesec. A partir de uma série temporal de dados de inadimplência, será ajustado um modelo paramétrico para previsão dos valores futuros, e na sequência serão definidos os parâmetros necessários para desenvolvimento do MMEP que será usado para o monitoramento desta série. Por último, os resultados serão avaliados para determinar se a solução foi capaz de atender o propósito proposto neste trabalho.

### 5.1. Coleta dos dados

Os dados foram extraídos de um relatório fornecido por um agente de monitoramento contratado pela Fortesec para acompanhar os pagamentos em uma de suas operações de securitização. O empreendimento em questão se trata de um loteamento no estado de São Paulo, voltado ao público de classe B e C.

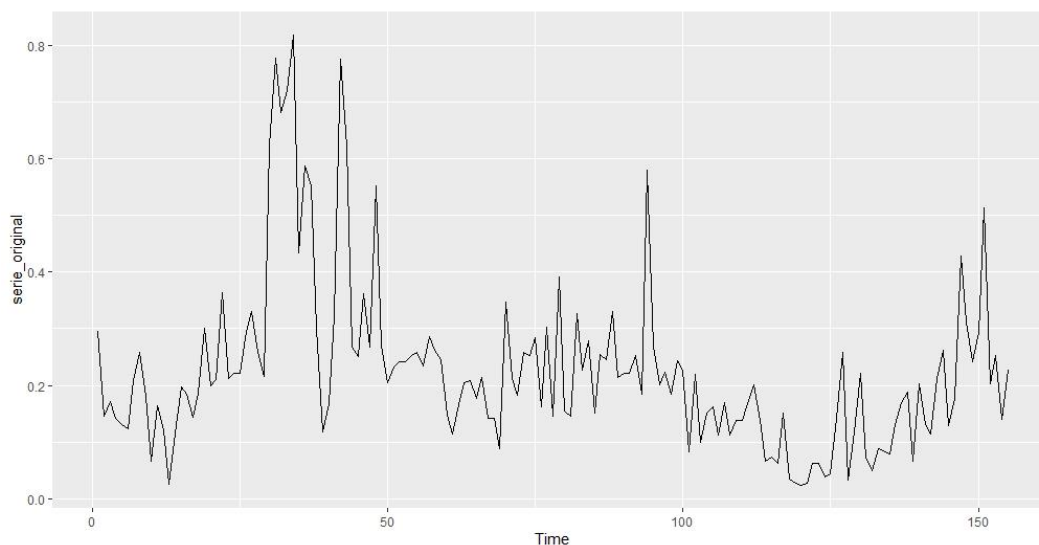
Foi escolhida uma série com um histórico grande de pagamentos para obter uma amostra com uma quantidade significativa de observações que permitisse a aplicação de modelos paramétricos. Neste caso, a amostra é composta por 175 observações, equivalente a quase 15 anos de monitoramento. Mas para fins de validação posterior do modelo, apenas os 155 primeiros meses serão usados nas próximas etapas.

Uma vez obtidos os dados, eles serviram de base para o cálculo da inadimplência mensal, a partir do uso do Excel. Os valores são apresentados no Anexo A. Como se pode notar, a série é composta por números positivos, entre 0 e 1.

### 5.2. Análise e ajustes na série

A série original  $x_t$  então foi introduzida ao RStudio, no qual foi gerado sua representação gráfica, mostrada na Figura 14.

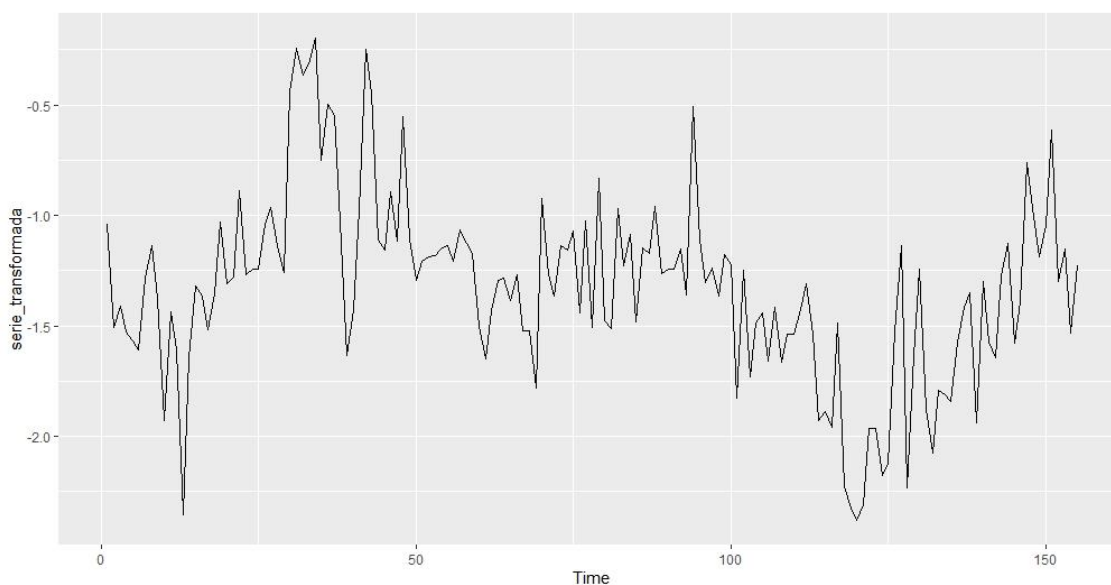
Figura 14 - Série de índices de inadimplência



Fonte: Elaborado pelo autor

Fica claro ao observar a imagem de que é necessário estabilizar a variância da série. Principalmente no início nota-se uma variação alta de um mês para o outro. Para corrigir a série, foi aplicada a transformação de Box e Cox (1964). O parâmetro  $\gamma$  utilizado foi definido pelo RStudio, com valor igual a 0,26487. A série  $x_t^{(\gamma)}$  após a aplicação da transformação é mostrada na Figura 15.

Figura 15 - Série transformada

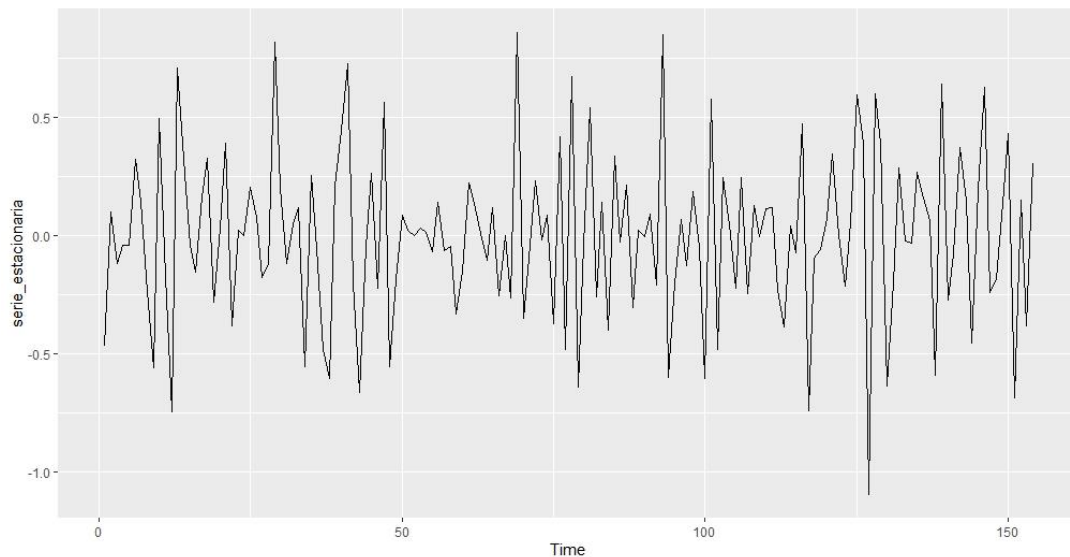


Fonte: Elaborado pelo autor

Visualmente, nota-se que há uma estabilização da variância, apesar de sugerir que ainda é necessário estabilizar a média da série. Foi então aplicado o teste de Dickey-Fuller aumentado.

O programa retornou um  $p - value$  igual a 0,3102. Assim, não se pode rejeitar a hipótese nula de não estacionariedade. Para tentar estabilizar a média, foi utilizada a primeira diferença da série  $x_t^{(\gamma)}$ . A Figura 16 mostra o resultado desta operação.

Figura 16 - Série estacionária

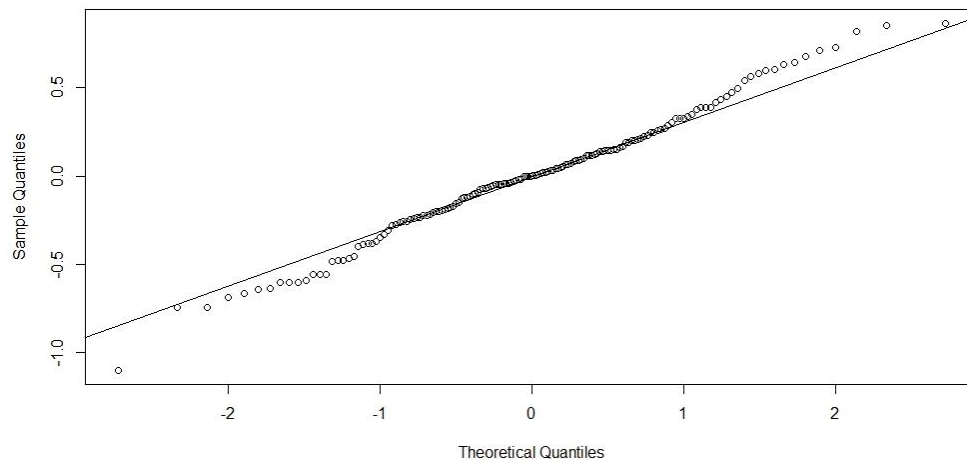


Fonte: Elaborado pelo autor

Realizando novamente o teste de Dickey-Fuller aumentado, obteve-se um  $p - value$  menor que 0,01. Dessa forma, a hipótese nula é rejeitada e pode-se assumir a estacionariedade da série  $\Delta x_t^{(\gamma)}$ .

Para se analisar  $\Delta x_t^{(\gamma)}$  quanto a sua distribuição, foi elaborado o gráfico Q-Q plot segundo uma distribuição normal, conforme mostrado na figura 17.

Figura 17 - Q-Q Plot da série estacionária

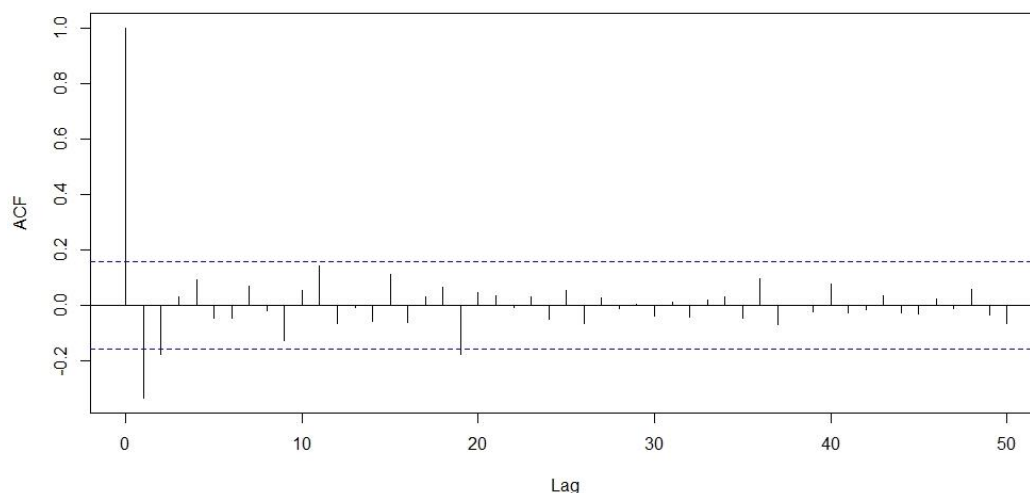


Fonte: Elaborado pelo autor

Apesar de nas extremidades os pontos da série se afastarem da reta normal, levando em consideração o comportamento médio da série, o teste sugere que há normalidade. Para confirmar, foi aplicado o teste não paramétrico de Shapiro-Wilk. Para a série estacionária, o programa retornou um  $p - value$  igual a 0,4311. Como é maior do que 0,05, pode-se tomar a hipótese nula como verdadeira e assumir que a série segue uma distribuição gaussiana.

Por último, a série será analisada sob o ponto de vista da independência. Os valores encontrados para as autocorrelações e autocorrelações parciais estão apresentados nas Figuras 18 e 19, respectivamente.

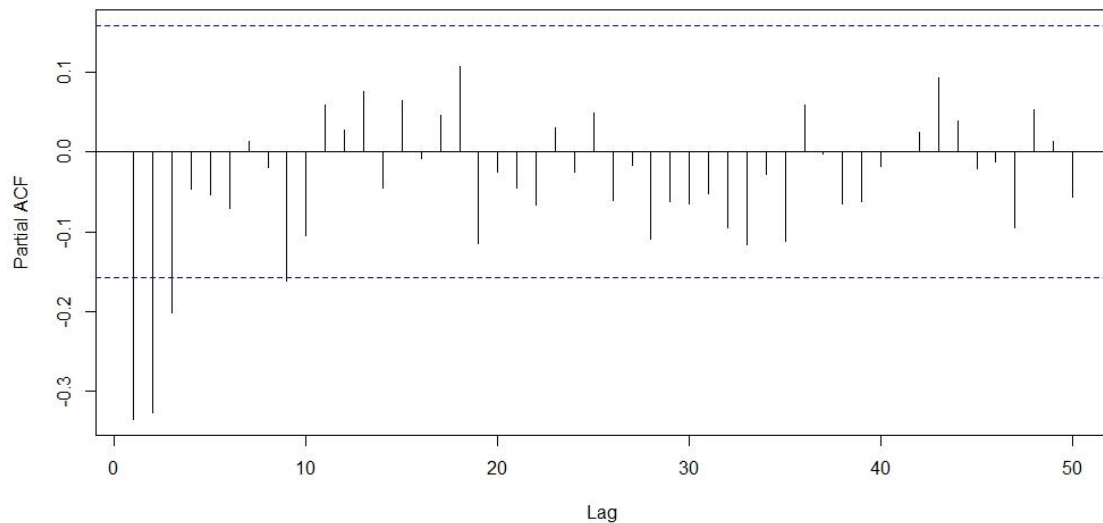
Figura 18 - Valores de autocorrelação



Fonte: Elaborado pelo autor



Figura 19 - Valores de autocorrelação parcial



Fonte: Elaborado pelo autor

Como se pode observar, ambas as funções apresentam um comportamento que lembra uma senóide amortecida, o que sugere que o modelo adequado seja da classe  $ARMA(p,q)$ . Ao analisar as autocorrelações, nota-se que valores significativos ocorrem apenas para os lags 1, 2 e 19. Em relação à autocorrelação parcial, os lags relevantes são o 1, 2, 3 e 9.

### 5.3. Identificação do modelo

A partir da análise da função de autocorrelação e autocorrelação parcial, a etapa de identificação começou com o teste de diversos modelos  $ARMA(p,q)$ . Foi considerada esta classe de modelos, pois já está sendo usada a primeira diferença da série transformada, e foi constatado que esta é estacionária, o que dispensa a necessidade de diferenças adicionais. Portanto serão avaliados modelos  $ARMA(p,q)$ , equivalentes ao  $ARIMA(p,0,q)$ .

Pela análise da autocorrelação, não é possível definir  $p$  e  $q$ . Assim, para se encontrar o modelo apropriado, foram testadas alternativas para diferentes combinações de  $p$  e  $q$ , ambos variando de 1 a 12. A partir de dados empíricos, foi adotada a premissa de que apenas as observações ocorridas no último ano podem ter efeito sobre a inadimplência em um dado período, portanto o limite para as ordens  $p$  e  $q$  escolhido foi 12.

Para cada modelo analisado, foi calculado o valor de seu AIC como critério de seleção. Por se tratar de um número muito grande de alternativas, foi aplicado um teste preliminar de diagnóstico do modelo, através da geração um gráfico de autocorrelação dos resíduos. Aqueles

que possuíam autocorrelações relevantes foram eliminados. Dessa forma pôde-se realizar uma seleção preliminar mais rápida que o diagnóstico completo, e assim avançar apenas com os modelos mais promissores. A Tabela 01 mostra os 10 modelos com os menores valores de AIC encontrados.

Tabela 01 - Valores de AIC para os modelos testados

Modelo	AIC
ARMA (4,5)	85,107
ARMA (5,5)	86,915
ARMA (6,4)	87,069
ARMA (3,6)	87,664
ARMA (6,6)	87,817
ARMA (6,5)	87,919
ARMA (7,4)	88,600
ARMA (5,8)	88,887
ARMA (8,2)	88,916
ARMA (7,4)	89,150

Fonte: Elaborado pelo autor

Dentre as possibilidades, a melhor escolha é o modelo ARMA(4,5) pelo critério de Akaike. Como pode ser notado ao visualizar a Tabela 01, a avaliação pelo AIC privilegia os modelos de menor ordem.

Em seguida foi analisada a possibilidade de se utilizar um modelo reduzido, com o intuito de simplificar o modelo através de uma menor quantidade de parâmetros. A Tabela 02 mostra as estimativas dos elementos do modelo completo, bem como o  $p - value$  associado à hipótese nula: se o  $p - value$  associado ao componente do modelo apresentar um valor inferior a 0,05, ele é considerado relevante. O modelo completo encontrado é então dado por:

$$Z_t = -0,5033Z_{t-1} - 0,3322Z_{t-2} + 0,42211Z_{t-3} + 0,7202Z_{t-4} + a_t + 0,0205a_{t-1} \\ + 0,0598a_{t-2} + 0,6751a_{t-3} + 0,4858a_{t-4} - 0,5106a_{t-5} + 0,0034$$

Tabela 02 - Estimativa dos coeficientes do modelo ARMA(4,5)

Parâmetro	Estimativa	P-value
AR1	-0,5033	0,0001*
AR2	-0,3322	0,0001
AR3	0,4221	0,0001*
AR4	0,7202	0,0001*
MA1	-0,0205	0,8201
MA2	-0,0598	0,4933
MA3	-0,6751	0,0001*
MA4	-0,4858	0,0001*
MA5	0,5106	0,0001*
Intercept	0,0034	0,7262

\*valores menores do que 0,0001

Fonte: Elaborado pelo autor

. Como pode ser observado, os parâmetros considerados significativos pela análise do  $p - value$  foram  $AR_1, AR_2, AR_3, AR_4, MA_3, MA_4$  e  $MA_5$ . Foi então ajustado um modelo reduzido considerando apenas estes parâmetros, conforme Tabela 03.

Tabela 03 - Estimativa dos coeficientes do modelo ARMA(4,5) reduzido

Parâmetro	Estimativa	P-value
AR1	-0,5152	0,0001*
AR2	-0,4340	0,0001*
AR3	0,3771	0,0001*
AR4	0,4638	0,0001*
MA3	-0,6933	0,0001*
MA4	-0,2930	0,0001*
MA5	0,3532	0,0001*

\*valores menores do que 0,0001

Fonte: Elaborado pelo autor

Por último, foi verificado se o modelo reduzido é equivalente ao cheio. A variação entre o AIC do modelo cheio e do reduzido é igual a 1,997. O valor encontrado na tabela qui-quadrado foi 7,815, para  $\alpha = 5\%$  e um número de graus de liberdade igual a diferença entre o número de parâmetros nos dois modelos, igual a 3. Assim, conclui-se que não se pode rejeitar a hipótese nula, então o modelo reduzido foi considerado equivalente ao cheio.

Dessa forma, a versão reduzida é dada por:

$$Z_t = -0,5152Z_{t-1} - 0,4340Z_{t-2} + 0,3771Z_{t-3} + 0,4638Z_{t-4} + a_t + 0,6933a_{t-3} \\ + 0,930a_{t-4} - 0,3532a_{t-5}$$

Ao analisar os parâmetros autorregressivos encontrados, é possível notar que o valor observado nos dois meses anteriores afeta inversamente  $Z_t$ , apesar do mês  $t - 1$  ter uma influência um pouco maior. Assim, um aumento em  $Z_{t-1}$  e  $Z_{t-2}$  resulta em uma redução de  $Z_t$ . Dados que os coeficientes  $AR_1$  e  $AR_2$  são próximos de 0,5, pode-se dizer que o valor de  $Z_t$  varia em função de uma aproximação da média aritmética de  $Z_{t-1}$  e  $Z_{t-2}$ .

Por outro lado, o modelo leva à conclusão de que aumentos em  $Z_{t-3}$  e  $Z_{t-4}$  geram aumentos no valor observado no período atual.

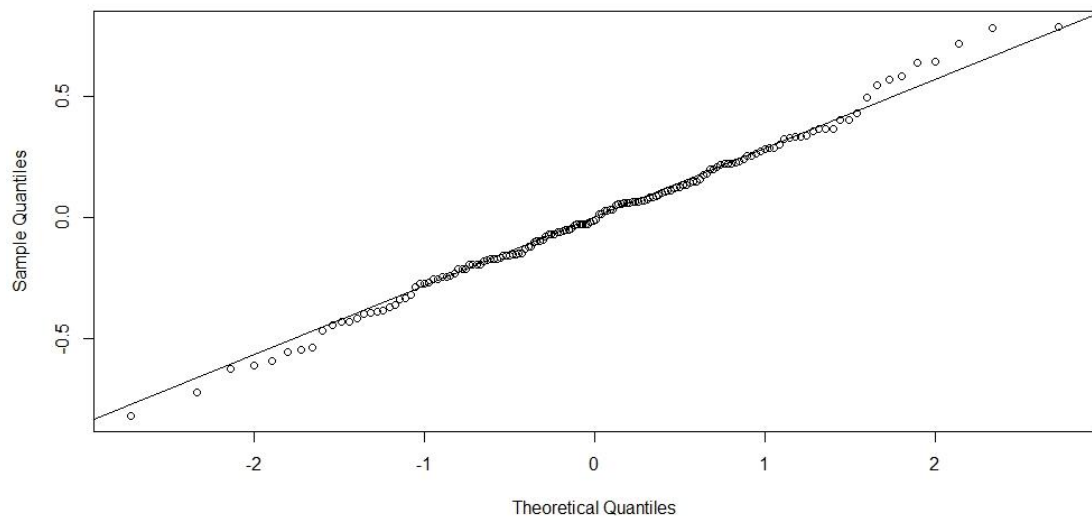
Interessante notar que de acordo com o modelo identificado, o valor do índice só é afetado por eventos em um horizonte de 5 meses. Por se tratar de um estudo sobre percentual de pagamentos não cumpridos em uma base mensal, faz sentido o modelo depender principalmente de observações no curto prazo.

#### 5.4. Diagnóstico do modelo

Para validar o modelo, o primeiro teste aplicado foi o de Ljung e Box (1978). Para o modelo reduzido, o programa retornou um valor  $p - value$  igual a 0,9641. Dessa forma, não se pode rejeitar a hipótese nula e se considera que os resíduos são independentes e identicamente distribuídos.

O segundo teste aplicado sobre os resíduos do modelo foi o de Shapiro-Wilk. O valor de  $p - value$  encontrado foi 0,825. Assim, não se pode rejeitar a hipótese nula e conclui-se que os resíduos apresentam distribuição normal. Para apresentar esse resultado de forma visual, foi gerado o gráfico Q-Q Plot, mostrado na Figura 20. A proximidade dos valores retirados da série com a reta normal corrobora o resultado do teste de Shapiro-Wilk.

Figura 20 - Q-Q Plot dos resíduos

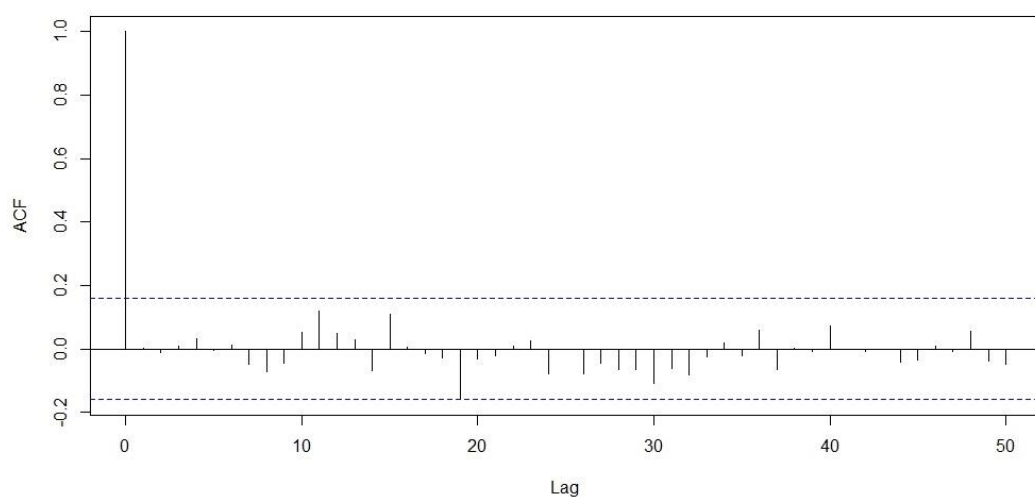


Fonte: Elaborado pelo autor

Por último, foi analisada a existência de autocorrelações e autocorrelações parciais significantes dentre os resíduos. O resultado obtido está demonstrado nas Figuras 21 e 22. Como pode ser observado, não existem autocorrelações fora do critério utilizado, o que leva à conclusão de que não há autocorrelação relevante entre os resíduos.

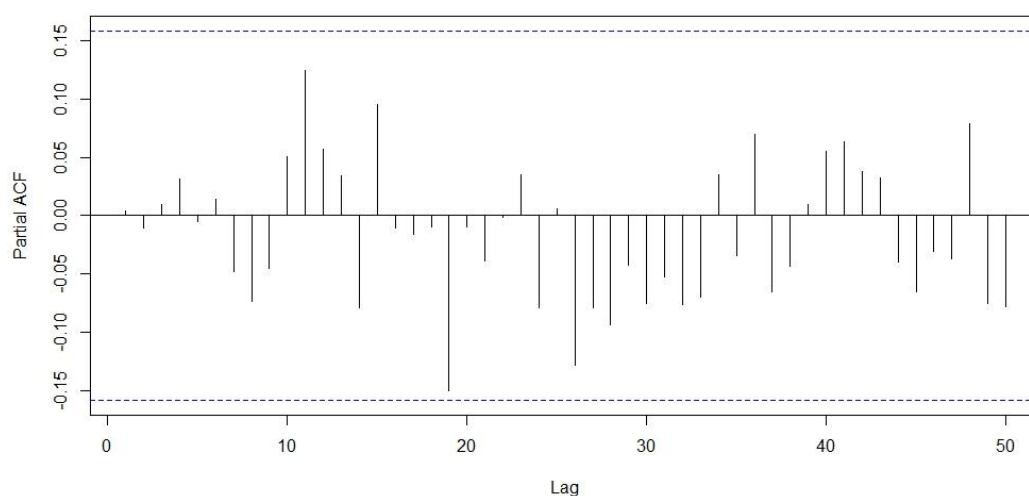
A partir dos resultados favoráveis obtidos em todos os estágios da etapa de diagnóstico, conclui-se que o modelo obtido é adequado para a previsão da série de inadimplência.

Figura 21 – Valores de autocorrelação nos resíduos



Fonte: Elaborado pelo autor

Figura 22 - Valores das autocorrelações parciais nos resíduos



Fonte: Elaborado pelo autor

### 5.5. Escolha do gráfico de controle

Conforme visto durante a etapa de análise da série temporal, existe uma correlação entre os dados analisados. Dessa forma, segundo a metodologia representada na Figura 11, o próximo passo para a definição do tipo de gráfico de controle mais adequado é verificar a existência de uma variável de ajuste, que não faz parte da série analisada.

Portanto, de acordo com o fluxograma de Montgomery (2004), para este caso são escolhas adequadas tanto os modelos MMEP, CUSUM,  $\bar{x}$  e MR (aplicados ou aos resíduos ou à série original), quanto um MMEP com linha central móvel ou um modelo livre de abordagem.

Ao se considerar que a área de gestão da empresa que fará uso deste modelo carece de pessoas especializadas no assunto de monitoramento estatístico de processos, é desejável que seja empregada a opção de mais fácil visualização e interpretação dos dados. Assim, foi decidido a aplicação de um gráfico MMEP sobre a média da série estacionária obtida ao final do tratamento da série de índices de inadimplência.

Optou-se pela série estacionária pela praticidade em calcular seus limites de controle. Caso fosse escolhido o monitoramento da série original, seria necessário aplicar o inverso da diferenciação e da transformação de Box-Cox aos limites de controle. Devido à natureza da operação de diferenciação, realizar a operação inversa poderia distorcer os valores finais, comprometendo a eficiência e confiabilidade do gráfico de controle.

## 5.6. Definição dos limites de controle

Através da aplicação do método da mínima soma dos quadrados dos erros, foi definido que o valor ótimo para o fator de suavização é  $\lambda = 0,7615$ . Assim, o cálculo da estatística  $z_t$  do MMEP é dada por:

$$z_t = 0,7615x_t + 0,2385z_{t-1}$$

Portanto, a estatística calculada no período anterior possui um peso menor que valor observado da série para o cálculo da estatística do período atual. De acordo com Hunter (1986), na maioria dos casos há uma certa subjetividade na escolha de  $\lambda$ , sendo que dados econométricos testados levam a sugestão de  $\lambda = 0,2 \pm 0,1$ . No entanto, considerando que valores maiores de  $\lambda$  diminuem a influência dos dados históricos, um valor mais alto de  $\lambda$  está coerente com o modelo ARMA(4,5) identificado. Dessa forma, tanto o modelo quanto a fórmula da estatística são influenciados em maior parte pelas observações mais recentes.

### 5.6.1. Definição ARL sob Controle

Com o algoritmo apresentado na subseção 4.6.2. será encontrado o valor da variável  $L$  necessária para determinar os limites de controle. Para tal foram geradas séries de observações a partir do modelo paramétrico ARIMA (4,1,5) reduzido encontrado nas etapas anteriores, a partir de valores de  $a_t$  simulados com a variância amostral dos resíduos  $\sigma_R^2 = 0,08900$ .

Tendo como alvo o  $ARL_0$  pré-determinado igual a 36, o algoritmo retornou o valor de  $L$  a ser usado como 1,82. Com esse valor, o ARL médio encontrado nas 5.000 simulações foi igual a 36,4. Ou seja, é esperado que o haja um alarme falso aproximadamente a cada 36 meses, próximo suficiente ao valor proposto inicialmente.

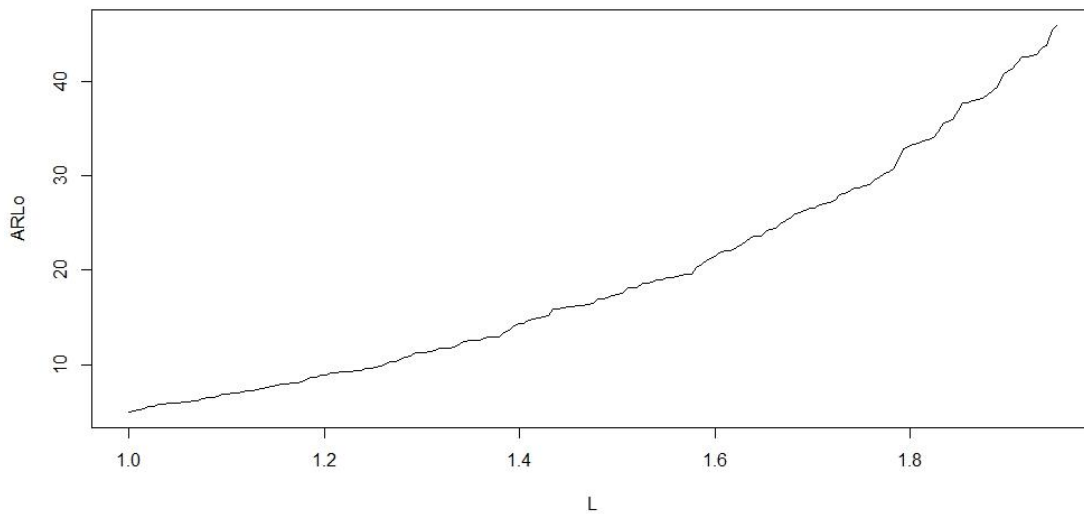
O valor de  $L$  encontrado é significativamente inferior ao valor padrão usado em gráficos de Shewhart, conhecido como três sigma ( $L=3$ ). Mas como os dados de inadimplência apresentam uma frequência mensal, significativamente mais longa que as aplicações mais tradicionais de gráficos de controle, foi necessário adotar um  $ARL_0$  mais baixo, e então faz sentido que o valor de  $L$  seja menor.

Esse valor baixo também é resultado do fator de suavização aplicado ao cálculo dos limites de controle. Conforme estudo conduzido por Lucas e Saccucci (1990), uma redução no

parâmetro  $\lambda$  provoca uma redução do  $ARL_0$ , mantidas as demais condições. Ainda que o fator de suavização encontrado de 0,7615 possa ser considerado alto, é menor que o  $\lambda = 1$  equivalente a um gráfico de Shewhart.

A Figura 23 mostra os valores do  $ARL_0$  encontrados pelo algoritmo para cada valor de  $L$  testado. Conforme esperado, há uma correlação positiva entre os valores de  $ARL_0$  e  $L$ .

Figura 23 - Valores do ARL em função de L



Fonte: Elaborado pelo autor

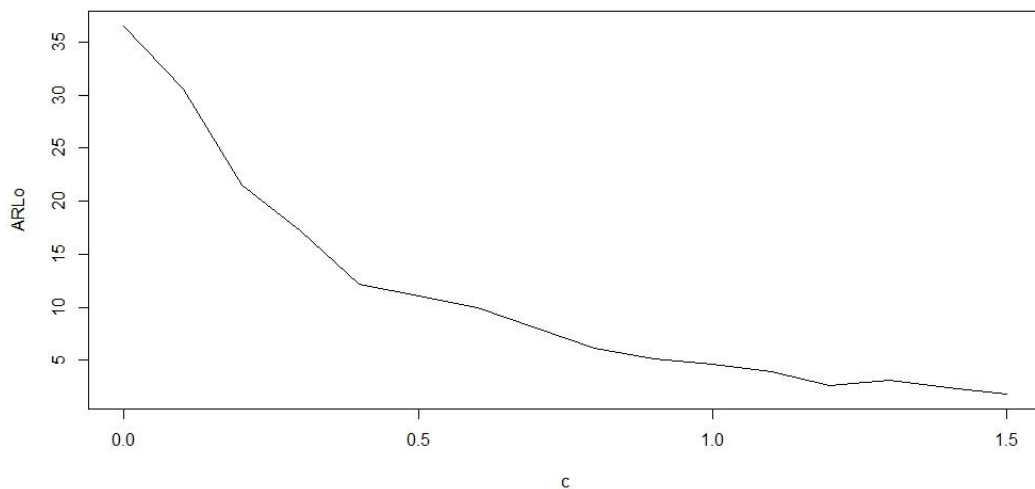
#### 5.6.2. Definição ARL fora de Controle

Com o parâmetro  $L$  definido, o passo seguinte consiste em um estudo da variação da média dos resíduos da série temporal sobre o valor do ARL encontrado. Para isso, foram feitas 5.000 simulações para cada valor de  $c$ , que variou de 0 a 1,5 em incrementos de 0,1, de modo que os resíduos introduzidos no modelo foram calculados como  $a_t \sim N(c * \sqrt{\sigma_R^2}, \sqrt{\sigma_R^2})$ .

Os resultados obtidos estão apresentados na Figura 24. Como se pode notar, o gráfico se mostrou bastante sensível a mudanças na média para valores de  $c$  entre 0 e 1. De acordo com os cálculos realizados, se a média aumentar em um desvio padrão, o ARL apresenta uma queda de 36,5 para 4,66, redução de 87,2%. Para variações maiores ainda, de 1,5 vezes o desvio padrão, o ARL passa a ser 1,92, um curto tempo de resposta. Portanto o gráfico MMEP elaborado aparenta ser capaz de detectar variações no processo com eficácia.



Figura 24 - Valores de ARL em função de c



Fonte: Elaborado pelo autor

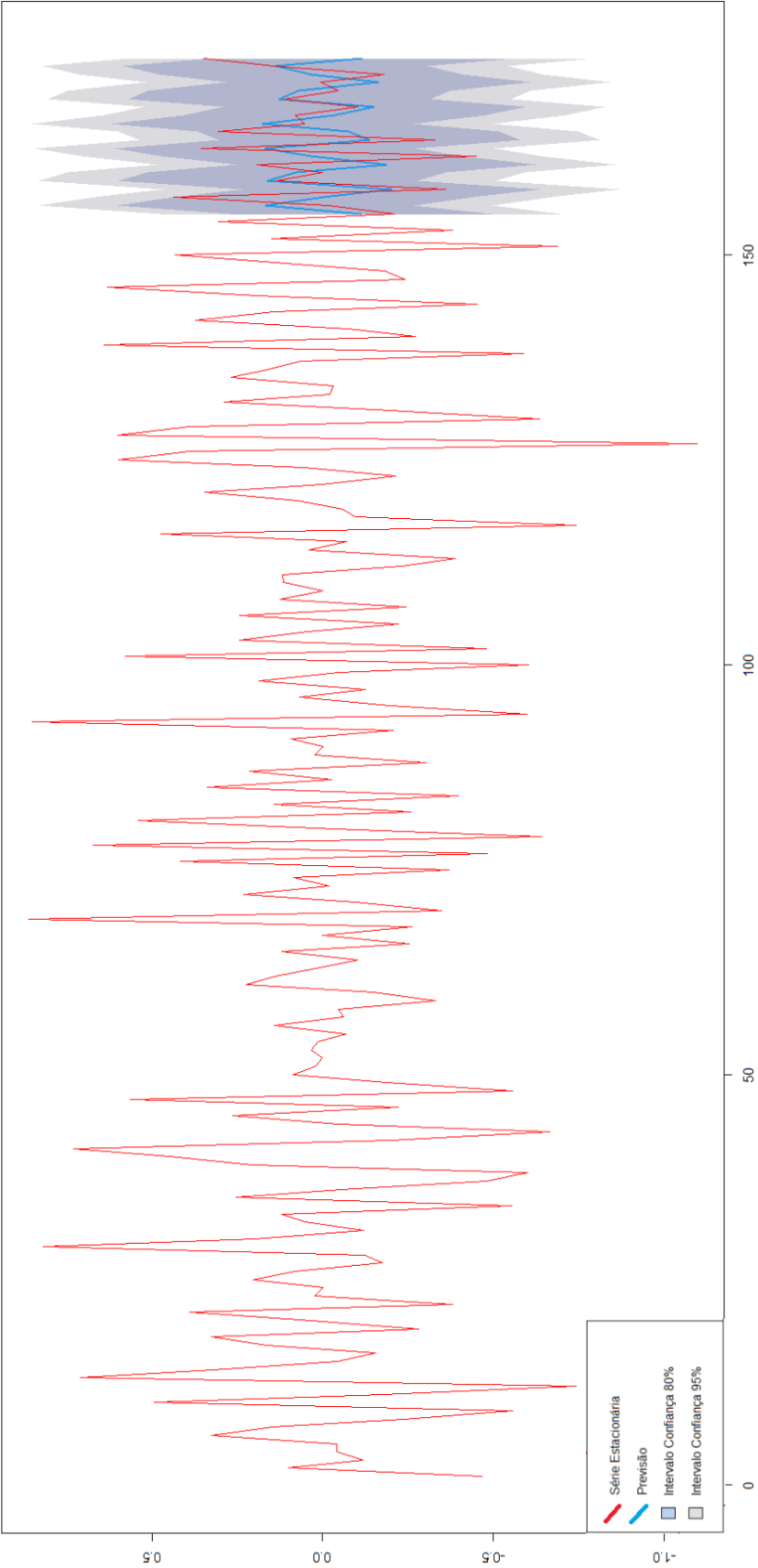
### 5.7. Avaliação dos resultados

A validação do resultado alcançado começa pela avaliação das previsões elaboradas pelo modelo paramétrico ARMA(4,5) reduzido. Para tal, foi usada a função *forecast()* do RStudio para projetar as observações dos períodos entre 155 e 174 (devido à diferenciação, a série estacionária possui um elemento a menos que a original). O resultado está apresentado na Figura 25. A linha vermelha da Figura 25 representa a série  $\Delta x_t^{(v)}$  completa com todas as 174 observações. A linha azul à direita representa a média das projeções. A área colorida com cinza escuro representa o intervalo de confiança de 80% para as projeções, e a área cinza claro o intervalo de confiança de 90% para as mesmas.

Como pode ser visto, as médias das projeções acompanham bem os valores da série estacionária, o que indica uma boa capacidade de previsão do modelo. Ainda, ao se observar os limites de confiança, entre os vinte valores previstos apenas 2 estão fora do intervalo de confiança de 80%, e nenhum valor ficou fora do intervalo de 95%. Na Figura 26 são mostrados os valores previstos e observados apenas para o intervalo de  $t$  entre 155 e 174 para facilitar a visualização.

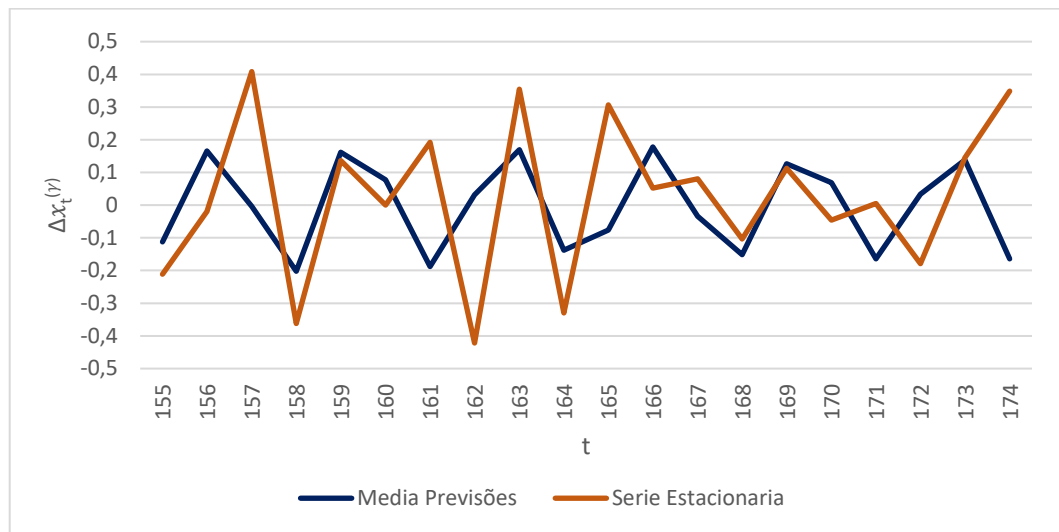
Figura 25 - Série Estacionária versus projeção

Forecasts from ARIMA(4,0,5) with zero mean



Fonte: Elaborado pelo autor

Figura 26 - Comparação entre previsões e observações reais para série estacionária

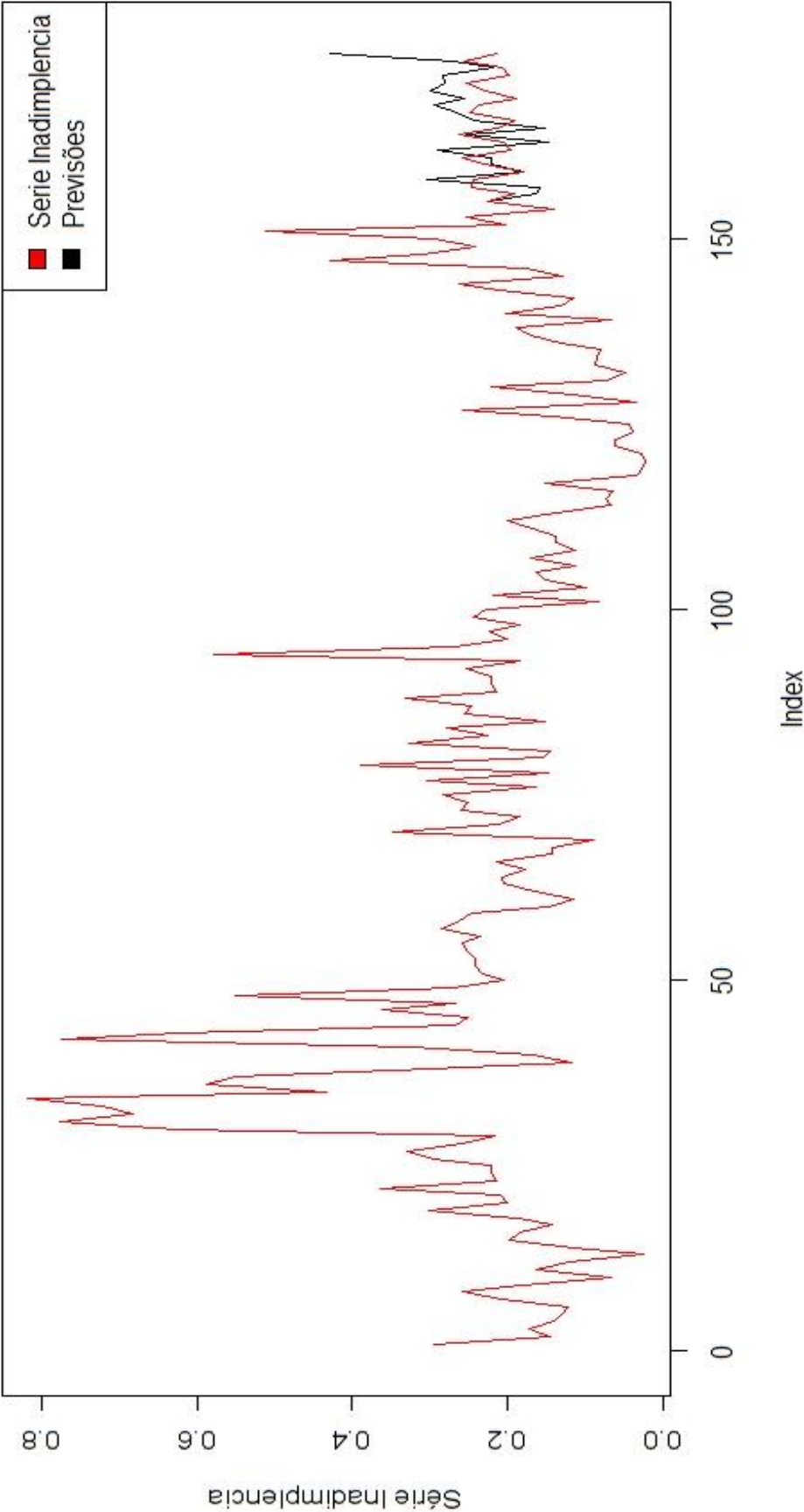


Fonte: Elaborado pelo autor

Em seguida foram aplicados sobre as médias das previsões os processos inversos à primeira diferenciação e à transformação de Box-Cox (1964) para se obter os valores de inadimplência previstos. A Figura 27 mostra o resultado final, no qual a linha vermelha representa a série original e a linha preta os valores previstos. Na Figura 28 pode-se ver os valores previstos e observados apenas para o intervalo de  $t$  entre 155 e 174. Fica claro ao comparar os valores previstos e observados que há uma aderência maior nas projeções de curto prazo.

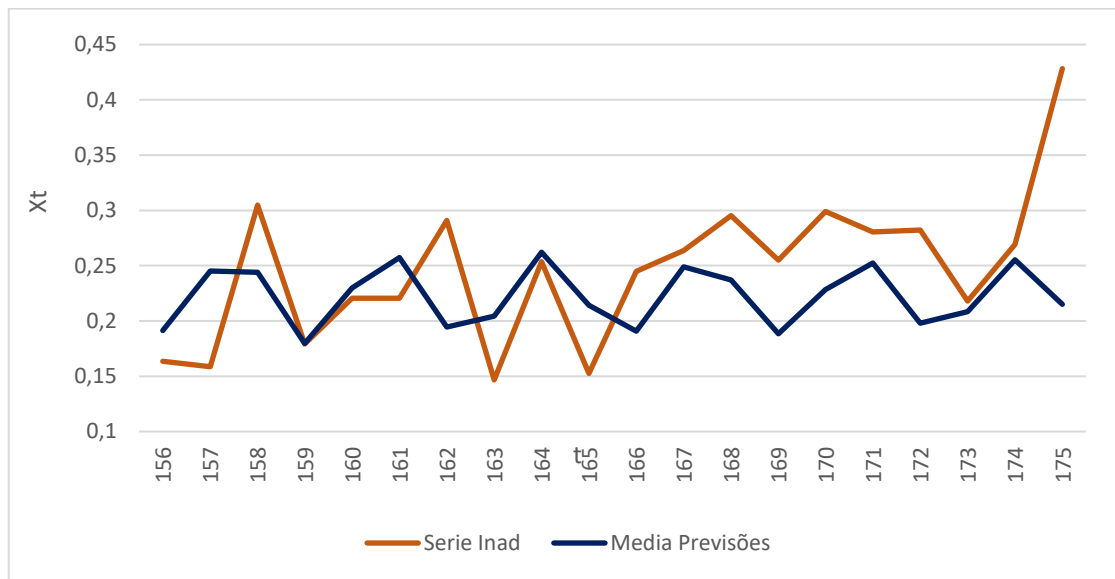
Novamente, foi notada uma aderência satisfatória entre os valores encontrados. Para finalizar a análise de previsão, foram calculados os índices MSE e RMSE para diferentes horizontes de projeção, conforme Tabela 4, com o objetivo de confirmar se há uma diferença relevante entre previsões de curto e médio prazo. Como pode ser observado pelos resultados da Tabela 04, o aumento do horizonte de projeção causa um aumento tanto no MSE quanto no RMSE, o que mostra que o modelo é mais confiável para projeções de curto prazo.

Figura 27 - Série original versus projeção



Fonte: Elaborado pelo autor

Figura 28 - Comparação entre previsões e observações reais para série original



Fonte: Elaborado pelo autor

Tabela 04 - Valores de MSE e RMSE versus horizonte de previsão

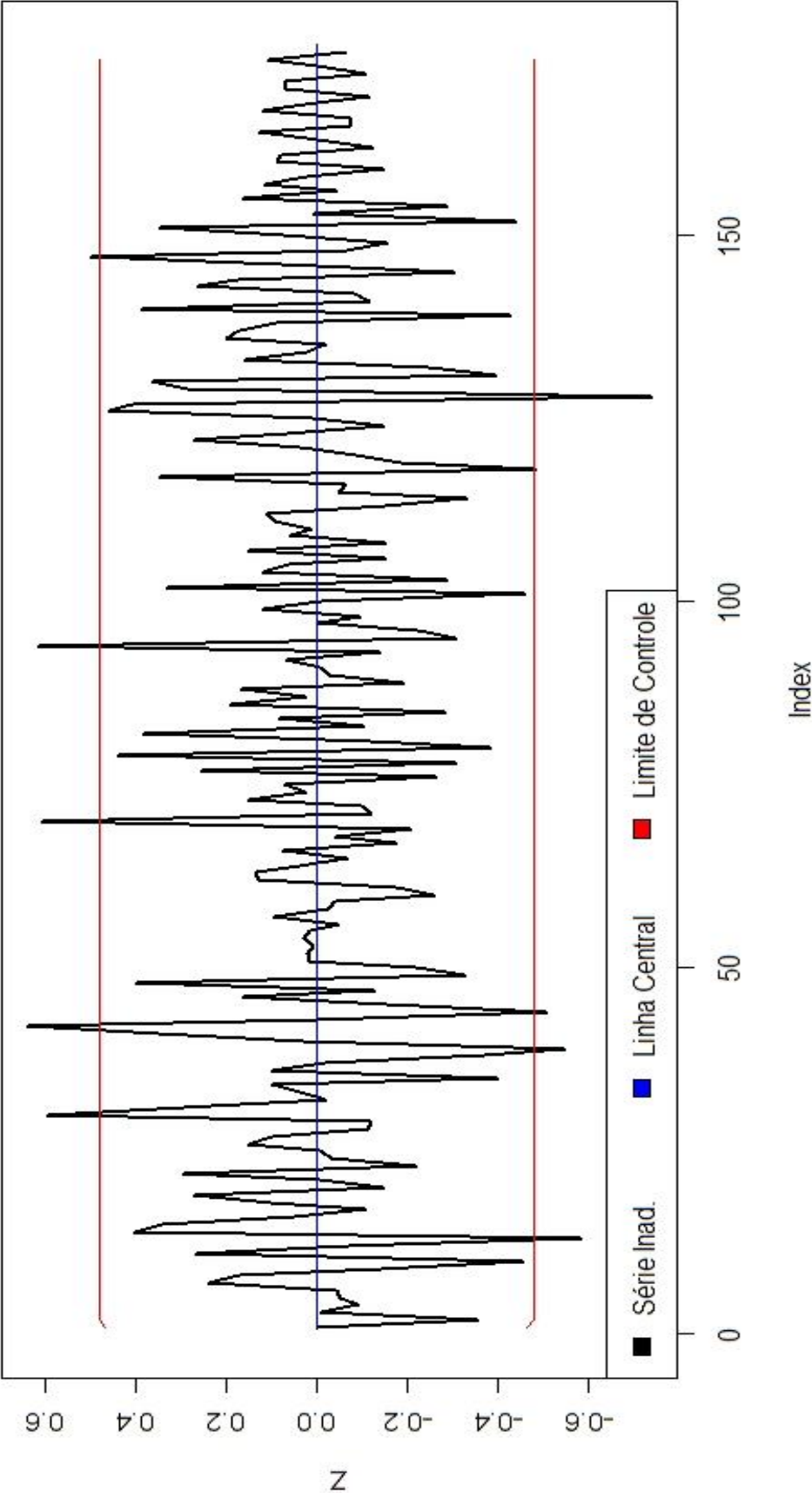
Horizonte de Previsão	MSE	RMSE
5 meses	0.002414	0.04914
10 meses	0.002989	0.05467
15 meses	0.003054	0.05527
20 meses	0,004972	0.07051

Fonte: Elaborado pelo autor

O próximo passo é verificar a eficiência do gráfico de controle desenvolvido. Primeiramente foi calculada a série de estatísticas  $z_t$ , usando os valores da série diferenciada para  $t = 1, \dots, 154$  e os valores das previsões do modelo para  $t = 155, \dots, 174$ . O objetivo aqui é introduzir as previsões no gráfico de controle para comparar seus resultados com o que foi verificado na realidade. Por último, foram calculados os limites de controle para  $t = 1, \dots, 174$ , usando os parâmetros calculados na seção 5.6. O gráfico final é apresentado na Figura 29.

A análise do gráfico mostra que houve 10 períodos em que foram verificados pontos fora dos limites de controle. Observando a série mostrada na Figura 27, existem três ocasiões de inadimplência significativamente acima da média do período, que serão comentados a seguir:

Figura 29 - Gráfico MMEP aplicado à série estacionária



Fonte: Elaborado pelo autor

- a) meses 30 a 43: o período mais duradouro de alta na inadimplência. Analisando o MMEP, percebe-se que há um alerta de processo fora de controle logo no mês 30. Como foi visto, o gráfico possui uma alta sensibilidade para variações na média do processo, então era esperado que uma variação dessa magnitude provocasse um alerta. Neste período, houve um total de 4 pontos fora do limite de controle. É importante citar aqui a primeira vantagem percebida da solução proposta. Na operação atual esse primeiro pico poderia ser interpretado como um aumento pontual, mas o MMEP já aponta como uma instabilidade no processo desde o primeiro mês de aumento de não-pagamentos;
- b) meses 147 e 151. Mais uma vez o MMEP apontou o processo como fora de controle já no primeiro mês deste período, acusando um crescimento na média de inadimplência e a necessidade de correção;
- c) mês 94: também foi identificado no próprio mês 94 pelo MMEP, mas pode ser considerado um *outlier*, pois é possível perceber que já no período seguinte a série retoma a média que vinha apresentando, se tratando, portanto, de um evento pontual.

Uma característica importante do gráfico é identificada a partir de seu desempenho durante períodos de crescimento acelerado da inadimplência, nos quais pode-se ver que não houve ocasiões em que o início desses picos não foi detectado. Assim, conforme mostrado na subseção 5.6.2., o para casos de aumento na média maior ou igual a 1,5 desvios padrões o MMEP possui um tempo de resposta de aproximadamente 1 mês.

Percebe-se também que o gráfico possui uma incidência erro tipo II muito baixa. Essa conclusão é coerente com o que foi proposto durante seu desenvolvimento, uma vez que dentre os erros, o pior para o investidor da operação é uma queda inesperada nos valores a receber.

Importante também verificar que na passagem entre os meses 69 e 70 houve a detecção de ponto fora do limite de controle. Diferente dos primeiros casos comentados, aqui não há um pico chamativo na inadimplência, mas a média passa de 16,0% entre os meses 60 e 69 para 24,0% entre os meses 70 e 79. Novamente esse aumento foi identificado logo no início do período. Aqui, por exemplo, o analista da Fortesec poderia ter antecipado essa tendência de crescimento que duraria 3 trimestres e se preparado de acordo já a partir do mês 69.

Todos resultados acima corroboram a sensibilidade do gráfico para variações consideráveis na média, e reforçam esta como um requisito fundamental do gráfico para atender seu propósito. Apesar do gráfico apresentar um tempo de resposta não tão rápido para pequenas variações, a série apresentada mostrou que em períodos de instabilidade sua inadimplência apresenta variações significativamente superiores a um desvio padrão, e, portanto, o uso do MMEP teria sido benéfico à empresa.

Naturalmente, ao se discutir a inadimplência há uma preocupação maior em monitorar aumentos nela, porém também existe benefícios em se identificar períodos de potencial redução. É uma prática comum a renegociação de dívidas, e se for possível perceber que a inadimplência está tendendo para abaixo dos limites estipulados, e suportar isso com base no modelo estatístico, é possível detectar os períodos ideais para clientes tentarem uma renegociação com os investidores. Nesse sentido, podem ser ressaltados dois períodos em que o gráfico de controle apontou para uma redução na média da operação:

- a) mês 10: neste período o gráfico indicou uma variação negativa na média das observações a partir dos valores dos meses anteriores. No entanto, foi observado que na ocasião do alarme já havia sido verificado um novo crescimento na média;
- b) mês 118: assim que houve uma redução da inadimplência de 15,1% para 3,4% entre os meses 117 e 118, o MMEP já identificou isso como uma alteração na média do processo, que pode ser comprovada pelo comportamento nos próximos 26 meses onde a máxima da inadimplência foi 25,8%, perto da média global. É interessante notar que a série neste período adota um comportamento semelhante ao de uma parábola, e que o MMEP apontou outra observação fora dos limites de controle perto do centro da parábola, análogo a um ponto de inflexão, quando a inadimplência volta a apresentar um crescimento.

Analisando as previsões introduzidas no MMEP, nota-se que se trata de um período sem grandes variações no processo, e o MMEP se comportou de acordo na maior parte dele, sem acusar nenhuma necessidade de correção. No entanto, um aumento na inadimplência no mês 175 não foi detectado pelo gráfico. Isso pode ser justificado pela característica do modelo paramétrico de ser mais preciso para horizontes menores de tempo. Por isso, sugere-se que o modelo seja usado para projeções em um horizonte de 12 meses.



Em suma, em diversas ocasiões o MMEP demonstrou um tempo de resposta rápido para detecção de processo fora de controle, o que providenciará a empresa com um maior tempo de resposta para lidar com crescimentos da inadimplência.



## 6. CONCLUSÃO

Neste capítulo será retomado tudo aquilo que foi abordado até agora neste trabalho e seus principais resultados, bem como uma avaliação geral da solução encontrada, suas limitações e sugestões de estudos futuros que possam aprimorar a solução elaborada.

A motivação deste trabalho de formatura foi a percepção de um problema na empresa em que o autor realiza seu estágio. O acompanhamento de índices de inadimplência nas carteiras de recebíveis imobiliários que possuem em gestão é uma atividade vital para as operações de securitização. Apesar disso, a empresa não conta com nenhum método de projeção desses índices. Na configuração atual, em caso de problemas de instabilidade da inadimplência, há margem apenas para ações corretivas, que podem ser tomadas tarde demais. Foi proposto então a criação de uma metodologia para desenvolver ferramentas de previsão e monitoramento.

O trabalho se inicia com o levantamento da bibliografia relevante ao tema proposto. Decidiu-se pelo desenvolvimento de um modelo de previsão a partir de séries temporais e modelos paramétricos. Para o monitoramento, foram pesquisadas ferramentas de controle estatístico de processos, dentre as quais optou-se pelo uso de gráficos de controle para o monitoramento da série estudada. A partir desta base teórica, foi estruturada uma metodologia para o desenvolvimento da solução final.

O desenvolvimento prático começou pelo estudo da série temporal em questão, sob os aspectos de estacionariedade, normalidade e dependência, bem como a realização de ajustes para torná-la compatível com modelos paramétricos. Conhecendo suas características, foram estudados diversos modelos para selecionar o mais adequado, que acabou sendo uma versão reduzida de um ARMA (4,5).

Para o monitoramento da série, foi escolhido o gráfico de média móvel exponencialmente ponderada (MMEP), por sua facilidade de desenvolvimento e aplicação e por ser ideal para detectar pequenas mudanças no processo. Colocou-se como meta um gráfico com um  $ARL_0$  igual a 36 meses. Para o seu uso foi necessário definir seus parâmetros. O fator de suavização  $\lambda$  foi encontrado a partir do método do mínimo erro dos quadrados, enquanto os limites de controle foram definidos a partir do uso de algoritmos iterativos que testaram diversas possibilidades até encontrar valores correspondentes ao  $ARL_0$  pré-definido.

A solução desenvolvida foi então submetida a uma avaliação para se determinar se cumpre seu propósito com eficácia. Em relação às projeções, foram deixadas de fora 20 elementos da série original para servir de base de comparação para as previsões do modelo. Analisando o modelo ARMA(4,5) foi possível verificar que há uma aderência satisfatória entre os valores reais e suas projeções. Conforme abordado na análise, ficou claro que quanto maior o horizonte de previsão maior o erro de previsão registrado, então o modelo é melhor utilizado com o foco em previsões de curto prazo, apesar de ter apresentado uma margem adequada para previsões de até 15 meses no futuro.

Quanto ao gráfico de controle desenvolvido, foi observado que ele apresenta uma sensibilidade adequada a variações na média dos resíduos, com um tempo de resposta muito baixo para variações na média maiores que 1 desvio-padrão. Cabe citar aqui uma limitação do MMEP, que serviria melhor seu propósito caso demonstrasse um  $ARL_1$  menor para pequenas variações. Porém, como foi discutido durante a avaliação dos resultados, as grandes variações na inadimplência monitorada fizeram com que na prática os resultados fossem favoráveis.

A partir de um paralelo com aquilo que foi de fato registrado ao decorrer da evolução da série, pode-se concluir que caso o modelo fosse adotado no mês 155 e servisse de base para a tomada de decisão nos próximos 20 meses o resultado teria sido positivo uma vez que a inadimplência ficou muito próxima ao previsto. Ainda, com base no histórico, foi mostrado que o MMEP foi capaz de detectar variações importantes em diferentes períodos, o que corrobora sua utilidade. Em luz desses resultados, conclui-se que a ferramenta desenvolvida cumpre seu objetivo e pode ser de grande valor para a Fortesec.

Este trabalho serviu para mostrar a utilidade e os benefícios do uso de séries temporais e gráficos de controle no monitoramento em índices de inadimplência, mas naturalmente ainda é passiva de aprofundamento e melhorias. Uma proposta para aprimoramento desta solução seria a análise da adoção de modelos Beta ARMA. Modelos Beta ARMA são utilizados nos casos específicos em que as observações da série estão entre 0 e 1, como é o caso de um índice de inadimplência. Este se trata de uma classe de modelos menos conhecida que o ARIMA, possivelmente por ser mais recente e não ter sido tão amplamente estudada quanto o ARIMA. Seu uso é mais complexo que o dos modelos apresentados neste trabalho e exigiria um grande volume de estudo, porém se tratando de uma ferramenta mais direcionada às características da série de inadimplência, é possível sua utilização possibilite obter ganhos interessantes em termos de capacidade de previsão.

Outro possível estudo futuro para aprofundamento e refinamento da solução seria a abordagem do uso de séries temporais não apenas para fins de previsão. Como foi explorado na bibliografia, estas séries podem ser aplicadas para diversos propósitos, dentre eles a investigação do mecanismo responsável por sua geração. Esta análise seria de grande utilidade, uma vez que conhecer os mecanismos por trás da inadimplência criaria a possibilidade de tomar ações mais direcionadas para mitigá-la, assim como pode fornecer informações que poderiam ser incorporadas ao modelo de projeção. Outra sugestão seria o estudo detalhado de um processo de implantação dessa solução na empresa, uma vez que são dezenas de operações ocorrendo simultaneamente e teria um impacto muito alto no operacional da Fortesec.

Por último, é sugerido um estudo de uma metodologia que possibilite inverter a transformação de Box-Cox e a diferenciação sobre os limites de controle do MMEP, para possibilitar o monitoramento a partir da série original, que facilitaria muito sua interpretação.

Quanto às dificuldades enfrentadas, para este trabalho de formatura foi necessário o aprendizado de séries temporais e modelos paramétricos, visto que o autor não havia tido contato com esses conceitos anteriormente, e que, portanto, exigiu diversas horas de estudo. A etapa de identificação do modelo se provou especificamente desafiadora, uma vez que foi necessário o teste de dezenas de modelos diferentes, tanto versões cheias quanto reduzidas, para que se obtivesse uma versão final que fosse aprovada tanto na etapa de diagnóstico quanto na avaliação dos resultados da previsão.



## REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. **Maximum likelihood identification of Gaussian autoregressive moving average models.** *Biometrika*, Oxford University Press, v.60, n. 2, pp. 255-265, 1973.
- BOX G. E. P.; COX D. R. **An analysis of transformations.** *Journal of the Royal Statistical Society, Series B (methodological)*, Wiley Online Library, v. 26, n. 2, pp. 211-243, 1964.
- BOX G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control.** *San Francisco: Holden-Day*, revisited edition, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. **Time Series Analysis: Forecasting and Control.** *Englewood Cliffs: Prentice hall*, 3ª ed. 1994.
- DICKEY, D. A.; FULLER, W.A. **Distribution of the estimators for autoregressive time series with a unit root.** *Journal of the American Statistical Association*, v. 74, pp. 427-431, 1979.
- HUNTER, J. S. **The Exponentially Weighted Moving Average.** *Journal of Quality Technology*, v.18, n. 4, pp. 203-210, 1986.
- HYNDMAN, R. J.; KOEHLER, A. B. **Another look at measures of forecast accuracy.** *International Journal of Forecasting*, v. 22, n. 4, pp. 679-688, 2006.
- KUME, H. **Métodos Estatísticos para Melhoria da Qualidade.** São Paulo: Editora Gente, 1993.
- LJUNG, G. M.; BOX, G. E. P. **On a measure of lack of fit in time series models.** *Biometrika*, Oxford University Press, v. 65, pp. 297 – 303, 1978.
- LUCAS, J. M.; SACUCCI, M. S. **Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements.** *Technometrics*, v. 32, n. 1, pp. 1 – 12, 1990.
- MONTGOMERY, D. C. **Introdução ao Controle Estatístico da Qualidade.** 4ª ed. Rio de Janeiro: Editora LTC, 2004.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULACHI, M. **Introduction to Time Series Analysis and Forecasting.** Nova Jersey: John Wiley e Sons, 2008.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2ª ed. São Paulo: Editora Blucher, 2006.

PAGE, E.S. **Continuous Inspection Schemes**. *Biometrics*, v. 41, n. 1, pp. 100-115, 1954.

ROBERTS, S.W. **Control Chart Tests Based on Geometric Moving Averages**. *Technometrics*, v. 42, n. 1, pp. 97-102, 1959.

SAMOHYL, R. W. **Controle Estatístico de Qualidade**. 1ª ed. Rio de Janeiro: Elsevier, 2009.

SHAPIRO, S. S.; WILK, M. B. **An analysis of variance test for normality (complete samples)**. *Biometrika*, v. 52, pp. 591-611, 1965.



## APÊNDICE A – Código para análise da série temporal e ajuste do modelo ARMA

```
library(ggplot2)
library(ggfortify)
library(forecast)
library(tseries)

setwd("C:/Users/mathe/Rstudio")
base <- read.csv2("Inad_G100_v2.csv", sep = ";")

serie_completa <- ts(base$Inad)
serie_original <- ts(serie_completa[1:155])

#Transformação Estabilizar a Variância
lbd <- BoxCox.lambda(serie_original)
serie_transformada <- BoxCox(serie_original, lbd)

#Teste estacionariedade
adf.test(serie_transformada, alternative="stationary")
serie_estacionaria <- ts(diff(serie_transformada))
adf.test(serie_estacionaria, alternative="stationary")

#Teste Normalidade
qqnorm(serie_estacionaria)
qqline(serie_estacionaria)
shapiro.test(serie_estacionaria)
```

```
#Teste Independência
```

```
estimador_acf <- acf(serie_estacionaria, lag=50, plot=TRUE)
```

```
estimador_pacf <- pacf(serie_estacionaria, lag=50, plot = TRUE)
```

```
#Estimação Modelo ARIMA
```

```
modelo_cheio <- arima(serie_estacionaria, order=c(4,0,5))
```

```
acf(modelo_cheio$residuals, lag=50)
```

```
coeftest(modelo_cheio)
```

```
#Estimação Modelo ARIMA Reduzido
```

```
modelo_reduzido <- arima(serie_estacionaria, ,order=c(4,0,5),transform.pars =  
F,fixed=c(NA,NA,NA,NA,0,0,NA,NA,NA),include.mean = F)
```

```
modelo_reduzido$aic
```

```
coeftest(modelo_reduzido)
```

```
#Teste Resíduos
```

```
acf(modelo_reduzido$residuals, lag=50)
```

```
pacf(modelo_reduzido$residuals, lag=50)
```

```
shapiro.test(modelo_reduzido$residuals)
```

```
Box.test(modelo_reduzido$residuals, type="Ljung-Box")
```

```
qqnorm(modelo_reduzido$residuals)
```

```
qqline(modelo_reduzido$residuals)
```

## APÊNDICE B – Código para determinação do fator de suavização do MMEP

```
#Importar dados

Zo <- mean(serie_original)
Xt <- serie_original[1:155]

lambda <- 0.01
lambda_otimo <- 0
erro_min <- 50

while (lambda <= 1) {
  Z <- c(Zo)
  i <- 1
  j <- 1
  E <- c()

  # cálculo da estatística Z
  while (i <= length(Xt)) {
    zt <- lambda*Xt[i] + (1-lambda)*Z[i]
    Z <- append(Z, zt)
    i <- i + 1
  }

  #Cálculo Erro
  while (j <= length(Xt)) {
    erro <- Xt[j] - Z[j]
    E <- append(E, erro)
    j <- j + 1
  }
}
```

```
}
```

```
#Cálculo Erro Quadrado
```

```
erro_quadrado <- E^2
```

```
#Soma dos Erros
```

```
soma_erros <- sum(erro_quadrado)
```

```
if (soma_erros < erro_min){
```

```
  erro_min <- soma_erros
```

```
  lambda_otimo <- lambda
```

```
}
```

```
lambda <- lambda + 0.0005
```

```
}
```

### APÊNDICE C – Código para determinação ARL para processo sob controle

```
ARLo <- 36
```

```
L <- 1
```

```
N <- 2000
```

```
ar1 <-modelo_reduzido$coef[["ar1"]]
```

```
ar2 <-modelo_reduzido$coef[["ar2"]]
```

```
ar3 <-modelo_reduzido$coef[["ar3"]]
```

```
ar4 <-modelo_reduzido$coef[["ar4"]]
```

```
ma3 <-modelo_reduzido$coef[["ma3"]]
```

```
ma4 <-modelo_reduzido$coef[["ma4"]]
```

```
ma5 <-modelo_reduzido$coef[["ma5"]]
```

```
dp_residuos <- sd(modelo_reduzido$residuals)
```

```
check_ARL <- F
```

```
Lista_ARL <- c()
```

```
while (check_ARL==F){
```

```
  cont <- 1
```

```
  soma <- 0
```

```
  while (cont<N){
```

```
    t <- 1
```

```
    fora_limite <- F
```

```
    sim_Xt <- arima.sim(n = 1005, list(order=c(4,1,5), ar=c(ar1,ar2,ar3,ar4),  
    ma=c(0,0,ma3,ma4,ma5)), sd=dp_residuos)
```

```

sim_Zo <- mean(sim_Xt)

sim_Zt <- c(sim_Zo)

valor_alvo <- mean(sim_Xt)

#Comparar limites de controle

while (fora_limite==F){

  zt <- lambda_otimo*sim_Xt[t]+(1-lambda_otimo)*sim_Zt[t]

  sim_Zt <- append(sim_Zt,zt)

  LSC <- valor_alvo + L*sd(sim_Xt)*sqrt(lambda_otimo*(1/(2-Lambda_otimo))*(1-(1-
lambda_otimo)^(2*t)))

  LIC <- valor_alvo - L*sd(sim_Xt)*sqrt(lambda_otimo*(1/(2-lambda_otimo))*(1-(1-
lambda_otimo)^(2*t)))

  if (zt>LSC | zt<LIC) {

    soma <- soma + t

    cont <- cont + 1

    fora_limite <- T

  } else {

    t <- t + 1

  }

}

cont <- cont + 1

}

ARL_sim <- soma/N

Lista_ARL <- append(Lista_ARL,ARL_sim)

if (ARL_sim >= ARLo){

```

```
    check_ARL <- T
  } else{
    L <- L + 0.01
  }
}
```





## APÊNDICE D – Código para determinação ARL para processo fora de controle

```
c <- 0
```

```
N <- 2000
```

```
ar1 <-modelo_reduzido$coef[["ar1"]]
```

```
ar2 <-modelo_reduzido$coef[["ar2"]]
```

```
ar3 <-modelo_reduzido$coef[["ar3"]]
```

```
ar4 <-modelo_reduzido$coef[["ar4"]]
```

```
ma3 <-modelo_reduzido$coef[["ma3"]]
```

```
ma4 <-modelo_reduzido$coef[["ma4"]]
```

```
ma5 <-modelo_reduzido$coef[["ma5"]]
```

```
dp_residuos <- sd(modelo_reduzido$residuals)
```

```
check_ARL <- F
```

```
Lista_ARL <- c()
```

```
while (c<=1.5){
```

```
  cont <- 1
```

```
  soma <- 0
```

```
  while (cont<N){
```

```
    t <- 1
```

```
    fora_limite <- F
```

```
      sim_Xt <- arima.sim(n = 1005, list(order=c(4,1,5), ar=c(ar1,ar2,ar3,ar4),  
ma=c(0,0,ma3,ma4,ma5)),mean=(c*dp_residuos), sd=dp_residuos)
```

```

sim_Zo <- mean(sim_Xt)

sim_Zt <- c(sim_Zo)

valor_alvo <- mean(sim_Xt)

#Comparar limites de controle

while (fora_limite==F){

  zt <- lambda_otimo*sim_Xt[t]+(1-lambda_otimo)*sim_Zt[t]

  sim_Zt <- append(sim_Zt,zt)

  LSC <- valor_alvo + L*sd(sim_Xt)*sqrt(lambda_otimo*(1/(2-lambda_otimo))*(1-(1-
lambda_otimo)^(2*t)))

  LIC <- valor_alvo - L*sd(sim_Xt)*sqrt(lambda_otimo*(1/(2-lambda_otimo))*(1-(1-
lambda_otimo)^(2*t)))

  if (zt>LSC | zt<LIC) {

    soma <- soma + t

    cont <- cont + 1

    fora_limite <- T

  } else {

    t <- t + 1

  }

}

cont <- cont + 1

}

ARL_sim <- soma/N

Lista_ARL <- append(Lista_ARL,ARL_sim)

c <- c + 0.1

}

```

### ANEXO A – Séries de índices de inadimplência

Data	Inadimplência	Data	Inadimplência	Data	Inadimplência
Mês 1	29,57%	Mês 41	32,31%	Mês 81	14,52%
Mês 2	14,58%	Mês 42	77,43%	Mês 82	32,64%
Mês 3	17,14%	Mês 43	62,15%	Mês 83	22,66%
Mês 4	14,12%	Mês 44	26,78%	Mês 84	27,82%
Mês 5	13,16%	Mês 45	25,10%	Mês 85	15,18%
Mês 6	12,28%	Mês 46	36,11%	Mês 86	25,55%
Mês 7	20,81%	Mês 47	26,67%	Mês 87	24,57%
Mês 8	25,86%	Mês 48	55,13%	Mês 88	33,07%
Mês 9	18,15%	Mês 49	26,98%	Mês 89	21,44%
Mês 10	6,70%	Mês 50	20,51%	Mês 90	22,16%
Mês 11	16,42%	Mês 51	23,34%	Mês 91	22,06%
Mês 12	12,26%	Mês 52	24,08%	Mês 92	25,23%
Mês 13	2,50%	Mês 53	24,15%	Mês 93	18,49%
Mês 14	11,54%	Mês 54	25,33%	Mês 94	57,86%
Mês 15	19,68%	Mês 55	25,81%	Mês 95	26,90%
Mês 16	18,42%	Mês 56	23,39%	Mês 96	20,11%
Mês 17	14,31%	Mês 57	28,65%	Mês 97	22,25%
Mês 18	18,66%	Mês 58	26,23%	Mês 98	18,39%
Mês 19	30,07%	Mês 59	24,58%	Mês 99	24,38%
Mês 20	19,99%	Mês 60	14,72%	Mês 100	22,74%
Mês 21	21,02%	Mês 61	11,39%	Mês 101	8,21%
Mês 22	36,33%	Mês 62	16,60%	Mês 102	21,92%
Mês 23	21,31%	Mês 63	20,56%	Mês 103	9,89%
Mês 24	22,09%	Mês 64	20,82%	Mês 104	15,15%
Mês 25	22,07%	Mês 65	17,77%	Mês 105	16,31%
Mês 26	29,55%	Mês 66	21,36%	Mês 106	11,19%
Mês 27	32,94%	Mês 67	14,26%	Mês 107	16,92%
Mês 28	25,80%	Mês 68	14,25%	Mês 108	11,17%
Mês 29	21,55%	Mês 69	8,94%	Mês 109	13,89%
Mês 30	62,42%	Mês 70	34,73%	Mês 110	13,83%
Mês 31	77,64%	Mês 71	21,24%	Mês 111	16,71%
Mês 32	68,17%	Mês 72	18,36%	Mês 112	20,07%
Mês 33	72,05%	Mês 73	25,91%	Mês 113	13,82%
Mês 34	81,83%	Mês 74	25,19%	Mês 114	6,73%
Mês 35	43,32%	Mês 75	28,42%	Mês 115	7,29%
Mês 36	58,78%	Mês 76	16,30%	Mês 116	6,31%
Mês 37	55,19%	Mês 77	30,32%	Mês 117	15,15%
Mês 38	30,05%	Mês 78	14,64%	Mês 118	3,45%
Mês 39	11,80%	Mês 79	39,03%	Mês 119	2,71%
Mês 40	16,72%	Mês 80	15,41%	Mês 120	2,31%

Data	Inadimplência	Data	Inadimplência
Mês 121	2,80%	Mês 165	15,26%
Mês 122	6,23%	Mês 166	24,49%
Mês 123	6,26%	Mês 167	26,39%
Mês 124	3,88%	Mês 168	29,52%
Mês 125	4,41%	Mês 169	25,50%
Mês 126	14,16%	Mês 170	29,89%
Mês 127	25,87%	Mês 171	28,05%
Mês 128	3,40%	Mês 172	28,24%
Mês 129	11,79%	Mês 173	21,79%
Mês 130	22,16%	Mês 174	26,92%
Mês 131	7,44%	Mês 175	42,82%
Mês 132	4,89%		
Mês 133	8,82%		
Mês 134	8,48%		
Mês 135	7,96%		
Mês 136	13,04%		
Mês 137	17,01%		
Mês 138	18,82%		
Mês 139	6,58%		
Mês 140	20,35%		
Mês 141	13,11%		
Mês 142	11,49%		
Mês 143	21,17%		
Mês 144	26,19%		
Mês 145	12,90%		
Mês 146	17,68%		
Mês 147	42,83%		
Mês 148	31,20%		
Mês 149	24,12%		
Mês 150	29,38%		
Mês 151	51,26%		
Mês 152	20,30%		
Mês 153	25,34%		
Mês 154	14,01%		
Mês 155	22,67%		
Mês 156	16,36%		
Mês 157	15,85%		
Mês 158	30,48%		
Mês 159	17,91%		
Mês 160	22,06%		