



UNIVERSIDADE
E D U A R D O
M O N D L A N E

Faculdade de Economia

Mestrado em Ciências Actuarias

**Aplicação da Metodologia de Box-Jenkins para a Previsão de Custos com
Sinistros no Seguro Automóvel em Moçambique**

Joaquim Matias Caetano Matias

Maputo, Outubro de 2024

Faculdade de Economia
Mestrado em Ciências Actuarias

**Aplicação da Metodologia de Box-Jenkins para a Previsão de Custos com
Sinistros no Seguro Automóvel em Moçambique**

Dissertação apresentada como pré-requisito para
a Culminação do Curso de Mestrado em Ciências
Actuarias na Faculdade de Economia da
Universidade Eduardo Mondlane.

Supervisor: Professor Doutor Nuno Ricardo Martins Sobreira

Co-Supervisora: Mestre Gulshan Cadir

Maputo, Outubro de 2024

DECLARAÇÃO

Eu, Joaquim Matias Caetano Matias, declaro que esta Dissertação é da minha autoria e resulta da minha investigação, o seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e nas referências bibliográficas. É apresentada em cumprimento parcial para a obtenção do grau de Mestre em Ciências Actuarias na Universidade Eduardo Mondlane.

Maputo, aos _____ de _____ de 2024

O declarante

Joaquim Matias Caetano Matias

APROVAÇÃO DO JÚRI

Esta Dissertação de Mestrado foi aprovada no dia ____ de _____ de 2024 por nós, membros do júri examinador nomeados pela Faculdade de Economia da Universidade Eduardo Mondlane.

O Presidente do Júri

O Arguente

O Supervisor

DEDICATÓRIA

Dedico este trabalho aos meus pais Bete Caetano Matias e Luísa Margarida Mulungo (em memória); a minha Esposa Aurora Eusébio Caetano Matias; aos meus filhos Milena Isabel, Tedes, Hélivio e Wamna Aurora; aos meus netos Luanga, Telvin e Aylane, pela motivação e apoio incondicional que deram durante essa longa caminhada.

AGRADECIMENTOS

A minha gratidão em primeiro lugar vai para Deus por me ter dado saúde e força para superar as dificuldades.

Agradeço, especialmente a minha família pelo apoio incondicional para que concretizasse a minha formação.

Ao meu Supervisor Professor Doutor Nuno Ricardo Martins Sobreira e a minha Co-Supervisora Mestre Gulshan Cadir, pelo suporte, pelas correcções e incentivos.

Ao Instituto de Supervisão de Seguros de Moçambique (ISSM), pela abertura na recolha de dados.

Agradeço igualmente aos Docentes do Mestrado em Ciências Actuarias e aos meus colegas de turma pelo suporte, troca e partilha de conhecimento.

Agradeço a todos que directa ou indirectamente fizeram parte da minha formação.

ÍNDICE

| | |
|--|------|
| DECLARAÇÃO..... | i |
| APROVAÇÃO DO JÚRI..... | ii |
| DEDICATÓRIA..... | iii |
| AGRADECIMENTOS..... | iv |
| ÍNDICE..... | v |
| RESUMO..... | viii |
| ABSTRACT..... | ix |
| LISTA DE SIGLAS E ABREVIATURAS..... | x |
| LISTA DE TABELAS..... | xi |
| LISTA DE GRÁFICOS..... | xii |
| CAPÍTULO I: INTRODUÇÃO..... | 1 |
| 1.1.Contextualização..... | 1 |
| 1.2.Motivação..... | 3 |
| 1.3.Contribuição..... | 3 |
| 1.4.Problematização..... | 4 |
| 1.5.Objectivos..... | 5 |
| 1.6.Estrutura da Dissertação..... | 5 |
| CAPÍTULO II: REVISÃO DE LITERATURA..... | 7 |
| 2.1. Definição de Conceitos..... | 7 |
| 2.1.1.Seguro..... | 7 |
| 2.1.2.Seguro Automóvel..... | 8 |
| 2.1.3.Sinistro..... | 9 |
| 2.1.4.Provisões Técnicas – Seguro Não Vida..... | 9 |
| 2.1.4.1.Provisão para Sinistros..... | 10 |
| 2.1.5.Série Temporal..... | 11 |
| 2.1.5.1.Diferenciação e Ordem de Integração de uma Série Temporal..... | 15 |
| 2.1.6.Modelos de Séries Temporais..... | 16 |
| 2.1.6.1.Modelos Auto-Regressivos e de Médias Móveis..... | 16 |
| 2.1.6.2.Modelos Auto-Regressivos..... | 17 |

| | |
|---|----|
| 2.1.6.3. Modelo de Médias Móveis | 17 |
| 2.1.6.4. Modelos Auto-Regressivos e de Médias Móveis | 17 |
| 2.1.6.5. Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA) | 18 |
| 2.1.7. Regressão..... | 19 |
| 2.1.8. Estacionaridade..... | 20 |
| 2.1.8.1. Testes de Estacionaridade..... | 20 |
| 2.1.7.2. Correlograma | 21 |
| 2.1.7.3. Teste Geral de <i>Dickey-Fuller</i> Aumentado (ADF) | 22 |
| 2.2. Métodos de Previsão Utilizados no Sector de Seguros | 23 |
| 2.2.1. Método de <i>Chain Ladder</i> | 24 |
| 2.2.2. Método de <i>Thomas Mack</i> | 28 |
| 2.2.3. Método de <i>Bootstrap</i> | 35 |
| 2.2.4. Método de <i>Box-Jenkins</i> | 39 |
| 2.2.4.1. Identificação | 39 |
| 2.2.4.2. Estimação | 42 |
| 2.2.4.3. Diagnóstico | 43 |
| 2.2.4.4. Previsão | 44 |
| 2.2.4.5. Critério para a Selecção de Modelos | 45 |
| CAPÍTULO III: METODOLOGIA..... | 46 |
| 3.1. Classificação da Pesquisa | 46 |
| 3.1.1. Quanto a Natureza | 46 |
| 3.1.2. Quanto a Abordagem..... | 46 |
| 3.1.3. Quanto aos Objectivos..... | 47 |
| 3.2. Método de Procedimento..... | 47 |
| 3.2.1. Técnicas e Instrumento de Recolha de Dados | 47 |
| 3.2.1.1. Pesquisa Bibliográfica | 47 |
| 3.2.1.2. Pesquisa Documental..... | 48 |
| 3.3. Fonte de Dados | 48 |
| 3.4. Análise de Dados | 48 |
| CAPÍTULO IV: ANÁLISE E DISCUSSÃO DE RESULTADOS | 50 |

| | |
|--|----|
| 4.1. Descrição da Série de Custos com Sinistros no Seguro Automóvel e Verificação do seu Comportamento entre os Anos de 2010 a 2023..... | 50 |
| 4.2. Identificação e Estimação do Modelo ARIMA de <i>Box-Jenkins</i> que Melhor se Ajusta aos Dados da Série | 57 |
| 4.3. Utilização do Modelo Estimado para Fazer Previsões dos Custos com Sinistros no Seguro Automóvel para o Período de Janeiro de 2024 a Dezembro de 2028..... | 59 |
| CAPÍTULO V: CONSIDERAÇÕES FINAIS | 61 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 63 |
| ANEXOS | 66 |

RESUMO

A presente pesquisa tem como objectivo estimar através da metodologia de Box-Jenkins um modelo ARIMA adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 a 2028. O estudo parte da seguinte questão: qual é o modelo ARIMA mais adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 à 2028? Trata-se de uma pesquisa aplicada baseada na abordagem quantitativa e exploratória a partir de consulta de material documental sobre os custos com sinistros no seguro automóvel em Moçambique no período de 2010 a 2023 e a análise de dados efectuada foi essencialmente econométrica. O estudo concluiu que dos modelos estimados o ARIMA (0,1,2) é o mais adequado para os dados referentes aos custos com sinistros no seguro automóvel em Moçambique. A previsão de custos com sinistros no seguro automóvel no período de 2024 a 2028 demonstra um comportamento crescente, o que sugere que a procura pelo seguro automóvel e o pagamento dos benefícios pelas seguradoras serão constantes e esse comportamento pode ser explicado pelo contínuo aumento do parque automóvel no país.

Palavras-Chave: Séries Temporais, Previsão, Sinistros, Seguro Automóvel.

ABSTRACT

The objective of this research is to estimate, using the Box-Jenkins methodology, an ARIMA model suitable for forecasting the costs of car insurance claims in Mozambique for the period from 2024 to 2028. The study starts from the following question: what is the ARIMA model that represents the best fit for forecasting the costs of car insurance claims in Mozambique for the period from 2024 to 2028? This is an applied research based on a quantitative and exploratory approach based on consultation of documentary material on the costs of car insurance claims in Mozambique from 2010 to 2023 and the data analysis carried out was essentially econometric. The study concluded that of the estimated models, ARIMA (0,1,2) is the most suitable for data relating to the costs of car insurance claims in Mozambique. The forecasts of car insurance claims costs in the period from 2024 to 2028 show an increasing behavior, which suggests that the demand for car insurance and the payment of benefits by insurers will be constant and this behavior can be explained by the continuous increase in the car park in the country.

Keywords: Time Series, Forecasting, Claims, Car Insurance.

LISTA DE SIGLAS E ABREVIATURAS

a.c – Antes de Cristo.

AR - Auto-Regressivo.

ARMA - Auto-Regressivo e de Média Móvel.

ARIMA - Auto-Regressivo, Integrado e de Média Móvel.

DF – Dickey Fuller.

EMOSE - Empresa Moçambicana de Seguros.

IGS – Inspeção Geral de Seguros.

ISSM - Instituto de Supervisão de Seguros de Moçambique.

SIM - Seguradora Internacional de Moçambique.

UEM – Universidade Eduardo Mondlane.

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1. Triângulo <i>Run-off</i> incremental..... | 24 |
| Tabela 2. Principais características das FAC e PFAC e Teorias dos processos estacionários não sazonais..... | 40 |
| Tabela 3. Teste de Dickey-Fuller com intercepto dos custos com sinistros no seguro automóvel em Moçambique em níveis..... | 55 |
| Tabela 4. Teste de Dickey-Fuller com intercepto e tendência dos custos com sinistros no seguro automóvel em Moçambique em níveis | 55 |
| Tabela 5. Teste de Dickey-Fuller com intercepto dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças..... | 56 |
| Tabela 6. Teste de Dickey Fuller com intercepto e tendência dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças | 56 |
| Tabela 7. Selecção do modelo ARIMA..... | 58 |
| Tabela 8. Previsão e intervalos de confiança dos custos com sinistros no seguro automóvel em Moçambique (2024:1 - 2028:12) – valores em milhões de meticais..... | 60 |

LISTA DE GRÁFICOS

| | |
|---|----|
| Gráfico 1. Histograma e estatística descritiva dos custos com sinistros no seguro automóvel em Moçambique | 51 |
| Gráfico 2. Evolução em níveis dos custos com sinistros no seguro automóvel (2010 - 2023) | 52 |
| Gráfico 3. Evolução em primeiras diferenças dos custos com sinistros no seguro automóvel (2010 - 2023) | 52 |
| Gráfico 4. Correlograma dos custos com sinistros no seguro automóvel em Moçambique em níveis..... | 53 |
| Gráfico 5. Correlograma dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças..... | 54 |
| Gráfico 6. Correlograma Residual..... | 59 |

CAPÍTULO I: INTRODUÇÃO

Este trabalho pretende, a partir dos dados agregados dos montantes pagos com sinistros pelas Seguradoras a operar no país no período de 2010 à 2023 compilados e publicados pelo ISSM¹, efectuar a previsão dos custos com sinistros para o período de 2024 à 2028, utilizando a metodologia de Box-Jenkins.

Nas secções que se seguem traz-se uma contextualização sobre a actividade seguradora (sua origem e evolução) a nível mundial e em Moçambique, convergindo para o seguro automóvel que é o objecto da presente pesquisa. A posterior, apresenta-se a motivação para a elaboração da pesquisa, bem como a contribuição que esta pretende trazer para o país. A parte introdutória continua com a apresentação do problema, os objectivos e finaliza com a estrutura da dissertação.

1.1. Contextualização

A actividade seguradora começou como um meio de redução de riscos dos comerciantes, a 500 a.c na China e 4500 a.c na Babilónia. A Lloyds foi a primeira associação de seguros que existiu na história. Nos Estados Unidos da América a primeira companhia de seguros foi fundada em Charleston, Sul da Califórnia em 1731, mas só cobria seguros de incêndio (Banze, 2007).

Banze (2007), acrescenta que em Moçambique a actividade seguradora teve o seu início no século XX (no ano de 1943) e era desenvolvida por empresas comerciais a praticarem paralelamente com as suas actividades comerciais e de navegação, mas antes, tal como ocorreu nas restantes partes do mundo, a solidariedade humana era prática comum em casos de desgraças ou infortúnios.

Após a Independência Nacional e no âmbito das Nacionalizações, a actividade seguradora foi uma das primeiras a ser abrangidas pelo reconhecimento da sua importância. Assim, em 13 de Janeiro de 1977, ao abrigo do Decreto-Lei nº 03 / 77 foi criada a EMOSE E. E. (Empresa

¹ ISSM – Instituto de Supervisão de Seguros de Moçambique (entidade reguladora da actividade seguradora e de fundos de pensões no país).

Moçambicana de Seguros Empresa Estatal, actualmente Empresa Pública), acto justificado pelos seguintes argumentos: (1) Por gerir seguros sociais tais como os de acidente de trabalho, com reflexos sociais; (2) Para garantir a formação técnico-profissional acelerada aos quadros nacionais aptos, num sector de serviços complexos que na altura não existiam; (3) Como forma de melhor servir as necessidades da economia, assegurando-se que sirva a plenitude dos interesses nacionais.

A Lei nº 24/91 de 31 de Dezembro liberalizou a actividade seguradora e resseguradora. Esta abertura, só veio a materializar-se em 1992 após o governo autorizar, ao abrigo deste decreto, a constituição de seguradoras privadas no país, nomeadamente: Companhia de Seguros IMPAR, Companhia Geral de Seguros de Moçambique (actual Global Alliance), Seguradora Internacional de Moçambique (SIM que detinha as acções da IMPAR) e a Hollard Moçambique Companhia de Seguros.

Com o crescimento do tráfego rodoviário, mercê do desenvolvimento económico e social que Moçambique vem conhecendo nos últimos anos, por forma a garantir uma protecção cada vez mais eficaz, o estado moçambicano decretou a obrigatoriedade do seguro automóvel de responsabilidade civil através da Lei nº 2/2003, de 21 de Janeiro, que viu o seu regulamento aprovado pelo Decreto nº 47/2005, de 22 de Novembro.

O seguro de responsabilidade civil automóvel tem como finalidade cobrir o risco que consiste na ameaça do património do segurado em razão de acontecimento futuro, incerto e danoso, independente da sua vontade – acidente de trânsito - que causará prejuízos nos bens materiais ou morais de terceiros ou pessoas transportadas no veículo (Vasques, 2005, p. 36).

Em Moçambique a actividade seguradora ainda está em franco crescimento, criando oportunidades para o aparecimento de novas Companhias Seguradoras, actualmente estão a operar no mercado 19 seguradoras (14 no ramo não vida, 2 no ramo vida e 3 mistas – vida e não vida).

O desafio reside no facto de na ocorrência de sinistros haver demoras na atribuição dos benefícios aos segurados, devido a questões meramente burocráticas o que diminui a confiança dos cidadãos em relação aos seguros. Ademais, muitos só subscrevem aqueles que são obrigatórios, como é o caso do Seguro de Responsabilidade Civil Automóvel. Para a

melhoria deste sector, deve haver maior publicidade e facilidade de adesão aos seguros e flexibilização da concessão dos benefícios.

1.2. Motivação

A razão da escolha do tema está ligada ao facto de o mesmo possibilitar a aplicação dos conhecimentos adquiridos ao longo do curso e à sua importância na planificação das actividades das seguradoras nacionais, pois a previsão de custos com sinistros ajudarão no cálculo dos montantes para as provisões. A provisão para sinistros corresponde ao custo total estimado que a Seguradora suportará para regularizar todos os sinistros que tenham ocorrido até ao final do exercício, quer tenham sido comunicados ou não, após dedução dos montantes já pagos respeitantes a esses sinistros. Esta provisão representa uma grande parte do passivo de uma seguradora (de ramo não vida).

Outro factor para a escolha deste tema é a existência hoje de *softwares* avançados para a análise econométrica como é o caso do EViews e do R que por serem de distribuição gratuita e mais acessíveis e ao mesmo tempo tem um leque bastante vasto de bibliotecas que permitem desenvolver uma programação bastante completa.

O último factor tem que ver com a vontade de desenvolvimento na área profissional do pesquisador, pois como analista de dados, entende que deve ter a habilidade de fazer uma avaliação dos riscos a que incorre a actividade seguradora para que estas não caiam em insolvência.

1.3. Contribuição

No ramo não vida, as provisões para sinistros são as que apresentam maior relevância, logo exigem uma análise actuarial detalhada e apropriada, para caucionar o pagamento de qualquer sinistro que venha a ocorrer, que seja da responsabilidade da seguradora.

O ramo automóvel é, de todos os ramos não vida, o que apresenta maior dimensão, a nível de produção. Este facto realça a contribuição deste estudo. A previsão dos custos com sinistros

no seguro automóvel utilizando a metodologia de Box-Jenkins trará uma alternativa mais económica às seguradoras para efectuarem as suas provisões anuais.

As técnicas de previsão auxiliam na tomada de decisões em actividades que necessitam de planeamento e da redução de incerteza, tornando assim os possíveis e futuros riscos mais visíveis e, conseqüentemente, mais controláveis.

Assim, os resultados deste estudo visam trazer melhorias para as empresas Seguradoras na prossecução dos seus objectivos de geração de lucros pois já existe um tratado de solvência II que exige às seguradoras um maior enfoque no controlo interno e governação assim como os sistemas de gestão de riscos, sendo a política de reservas das companhias muito mais exigente do que era no passado. Isto faz com que as companhias de seguro tenham uma menor margem de manobra nos valores incorridos de sinistros mas por outro lado protege muito mais os segurados e tomadores de seguros.

Todavia, esta situação leva a que as seguradoras tenham como objectivo estimar o custo final de cada sinistro ocorrido no menor espaço de tempo para que não haja surpresas no futuro. A previsão de custos em sinistros no seguro automóvel é sempre desafiante dado que não se trata apenas de danos materiais, pois em muitos casos incluem vidas humanas.

1.4. Problematização

No sector de seguros em Moçambique, as provisões técnicas são deveras controladas pelo regulador (Instituto de Supervisão de Seguros de Moçambique), principalmente no seguro automóvel por ser aquele com maior número de apólices detidas pelas seguradoras.

Para cumprir com os requisitos estabelecidos pelo regulador, as seguradoras contratam o serviço de actuários na África do Sul, Portugal e Reino Unido para efectuarem os cálculos dos montantes que estas devem reter para as provisões.

Do levantamento efectuado nas seguradoras nacionais, constatou-se que o modelo utilizado para o cálculo das provisões é o de *Chain Ladeer*, por ser um método de fácil utilização e aplicabilidade. Mas também apresenta algumas limitações, tal como o pressuposto de que existe independência entre os diversos anos de ocorrência e assumir que o factor de desenvolvimento é o mesmo ao longo dos anos de ocorrência dos sinistros. Essas limitações

do modelo *Chain Ladder* levam com que os cálculos das provisões sejam efectuados anualmente o que acarreta custos elevados.

Um modelo alternativo para o cálculo das provisões poderia ser facultado pela metodologia ARIMA por acarretar menores custos e permitir efectuar previsões com fiabilidade até um período de cinco anos. Assim, a presente pesquisa levanta a seguinte pergunta:

- *Qual é o modelo ARIMA que melhor se ajusta aos dados e que previsões oferecem para os custos com sinistros no seguro automóvel em Moçambique para o período de 2024 à 2028?*

1.5. Objectivos

1.5.1. Objectivo Geral

- Estimar através da metodologia de Box-Jenkins um modelo ARIMA adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 à 2028.

1.5.2. Objectivos Específicos

- Descrever a série de custos com sinistros no seguro automóvel para verificar o seu comportamento ao longo dos anos de 2010 a 2023;
- Identificar e estimar o modelo ARIMA de Box-Jenkins que melhor se ajusta a série;
- Utilizar o modelo estimado para fazer previsões dos custos com sinistros no seguro automóvel para o período de Janeiro de 2024 a Dezembro de 2028.

1.6. Estrutura da Dissertação

A pesquisa está organizada nos seguintes capítulos, sendo o primeiro a Introdução, onde apresenta-se o tema, expõe-se a justificativa para o desenvolvimento do tema, formulam-se os objectivos e o problema de estudo.

O segundo capítulo refere-se a Revisão de Literatura, que aborda os principais conceitos da pesquisa e apresenta os modelos de previsão mais utilizados no sector de seguros. O terceiro

capítulo aborda a Metodologia, onde faz-se a descrição dos procedimentos metodológicos que delinearão a pesquisa, como as técnicas, instrumentos de recolha de dados, fonte de dados, as técnicas de análise de dados.

No capítulo referente a Análise e Discussão de Resultados, são apresentados os resultados e é feita a discussão dos mesmos. Por fim, seguem as Considerações Finais e as Referências bibliográficas que serviram de base para a construção do argumento teórico.

CAPÍTULO II: REVISÃO DE LITERATURA

Este capítulo visa apresentar informações sobre o debate teórico no contexto dos métodos de previsão, focalizado para a ligação entre estes e o sector de seguros em Moçambique. Assim, primeiramente serão introduzidos conceitos teóricos gerais, nomeadamente: seguros, seguro automóvel, sinistros, provisões técnicas – seguro não vida, série temporal e estacionariedade; a *posteriori* são apresentados os métodos de previsão utilizados no sector de seguros.

2.1. Definição de Conceitos

2.1.1. Seguro

O Instituto de Supervisão de Seguros de Moçambique (ISSM, 2021), o regulador da actividade seguradora em Moçambique define o seguro como o contrato pelo qual a seguradora, mediante retribuição pelo tomador do seguro, se obriga, a favor do segurado ou de terceiro, à indemnização de prejuízos resultantes, ou ao pagamento de valor pré-definido, no caso de se realizar um determinado evento futuro e incerto.

Para Vasques (2005), o seguro consiste numa operação pela qual uma das partes (o segurador) obtém, mediante certa remuneração (prémio) paga à outra parte (segurado), a promessa de uma indemnização para si ou para terceiro, no caso de se realizar um risco.

Na visão de Carvalho (2010), o seguro pode ser classificado como sendo um contrato oneroso, visto que dele resulta uma atribuição patrimonial e um correspondente sacrifício patrimonial às partes nele envolvido, consubstanciando-se no pagamento do prémio pelo tomador do seguro e na prestação indemnizatória ou convencional a carga do segurador.

A mais importante distinção ao nível dos ramos de seguro cabe aos seguros do ramo vida e não vida. Segundo o ISSM (2021), são seguros do ramo vida: seguros de vida (individual e grupo); seguros financeiros: seguro de capitalização e planos de poupança reforma. Acrescenta ainda que os seguros do ramo não vida são: seguro Automóvel – responsabilidade civil e outros danos; responsabilidade civil geral; Crédito; seguro de Doença; Marítimo e Transportes; Aéreo; Incêndio e Outros Danos; Acidentes de Trabalho; Acidentes Pessoais; Diversos.

ISSM (2021) explica que fica ao critério de cada companhia a prática das várias opções de seguros, mediante a licença de comercialização dos mesmos, assim como a designação e as especificações de cada seguro (produto) podem variar entre seguradoras.

2.1.2. Seguro Automóvel

O seguro de responsabilidade civil automóvel tem como finalidade cobrir o risco que consiste na ameaça do património do segurado em razão de acontecimento futuro, incerto e danoso, independente da sua vontade – acidente de trânsito - que causará prejuízos nos bens materiais ou morais de terceiros ou pessoas transportadas no veículo (Instituto de Supervisão de Seguros - ISSM, 2021).

Para Carvalho (2010), o seguro de responsabilidade civil automóvel, é definido como sendo um acordo de vontades entre o tomador do seguro e o segurador, mediante o qual o primeiro se obriga a pagar uma quantia em dinheiro, designada por prémio, e o último se obriga a manter indemnização ao segurado dos prejuízos eventualmente decorrentes de sinistros por si causados no exercício de condução de veículos terrestres a motor, isto é, obriga-se a suportar o risco de circulação daqueles veículos, para o que a lei exige determinado capital e cuja celebração é legalmente imposta.

Martins (2011) considera que o contrato de seguro automóvel como o contrato celebrado entre o proprietário de um veículo (e uma entidade) devidamente autorizada a explorar a indústria de seguros (segurador), por meio do qual aquele, mediante o pagamento de uma importância previamente estabelecida (prémio) transfere para este o encargo de reparação pelos prejuízos ou danos causados a terceiros por aquele veículo, na sua integridade física ou património.

Percebe-se que todos os autores convergem na ideia de que o contrato de seguro obrigatório na responsabilidade civil automóvel é um contrato nominado e típico, visto que resulta da lei, e o seu regime tem uma natureza maioritariamente imperativa. Como já vimos anteriormente, o contrato de seguro obrigatório é bilateral, pois na sua celebração encontram-se duas partes: o tomador do seguro e o segurador. Acrescenta-se que o contrato de seguro obrigatório é um contrato de natureza aleatório, visto o seu resultado depender de factos futuros e incertos. Ocorrendo ou não esse facto, os contraentes sabem que, com a celebração do contrato têm

uma possibilidade de ganhar ou perder e também não sabem quanto vão beneficiar ou perder com a aleatoriedade do contrato em causa.

2.1.3. Sinistro

Segundo Carvalho (2010), sinistro é o nome dado a ocorrência de todo evento que tem cobertura no seguro contratado e que esteja especificado na apólice ou contrato.

Na visão de Martins (2011), sinistro é a ocorrência de um invento involuntário no qual existe um contrato de seguro para quando o mesmo vier a acontecer e no qual há pagamento de uma indemnização ou benefício.

Borginho (2003) explica que existem duas maneiras de receber a indemnização pela seguradora: de forma parcial ou de forma integral. O critério para classificar um sinistro como perda parcial ou total é o seguinte: se o valor para a reposição for inferior a 75% então está caracterizado como perda parcial e se o valor for igual ou superior a 75%, então caracteriza-se por perda total.

2.1.4. Provisões Técnicas – Seguro Não Vida

Segundo Borginho (2003), as provisões técnicas correspondem aos montantes que a Seguradora necessita constituir e manter, para que consiga cumprir todas as responsabilidades assumidas. Esse montante deve ser suficiente para permitir à Seguradora cumprir os compromissos resultantes dos contratos de seguro que realizou.

Carvalho (2010) explica que as provisões técnicas do ramo Não Vida a serem constituídas e mantidas são as seguintes:

- 1. Provisão para prémios não adquiridos** – Esta provisão é constituída com parte dos prémios brutos emitidos, relativamente a cada um dos contratos de seguro em vigor, com excepção dos respeitantes ao ramo vida, a imputar a um ou vários dos exercícios seguintes. Na concretização de um contrato de seguro, nem sempre o mesmo inicia a 1 de Janeiro de um determinado ano, assim como existem diferentes modalidades de pagamento e fraccionamento (mensal, trimestral, semestral e anual). Por isso a

seguradora, no fim do ano em exercício nem sempre recebe na totalidade o prémio correspondente do contrato de seguro. Resumindo, trata-se de uma reserva do valor de prémio que ainda está por receber por parte da seguradora;

2. **Provisão para riscos em curso** - corresponde ao montante necessário para fazer face a prováveis indemnizações e encargos a suportar após o termo do exercício e que excedem o valor dos prémios não adquiridos e dos prémios exigíveis relativos aos contractos em vigor;
3. **Provisão para sinistros** - corresponde ao custo total estimado que a Seguradora suportará para regularizar todos os sinistros que tenham ocorrido até ao final do exercício, quer tenham sido comunicados ou não, após dedução dos montantes já pagos respeitantes a esses sinistros. Esta provisão representa uma grande parte do passivo de uma seguradora (de ramo não vida);
4. **Provisão para desvios de sinistralidade** - destina-se a fazer face a sinistralidade excepcionalmente elevada em ramos de seguros que, pela sua natureza, se preveja que tenha maiores oscilações, como por exemplo seguros associados a risco de fenómenos da natureza (sismos, vulcões, tempestades e outros).

Uma vez que a presente pesquisa versa sobre a previsão de custos com sinistros, no ponto seguinte falaremos sobre a provisão com sinistros.

2.1.4.1. Provisão para Sinistros

Segundo Borginho (2001), a provisão para sinistros pode ser dividida nas seguintes quatro categorias:

1. Provisões para sinistros já declarados, para estimar o custo final;
2. Provisões IBNR (*“Incurred But Not Reported”*), para sinistros ocorridos mas ainda não declarados;
3. Provisões para sinistros ocorridos mas insuficientemente provisionados (IBNER);

4. Provisão para despesas de regularização de sinistros, que podem ser do tipo administrativo.

Borginho (2003) explica que a estimativa das provisões pode ser realizada de duas formas:

- a) Caso a caso, analisando todos os processos um a um para constituir a provisão. Esta situação não é viável dada a dimensão do número de sinistros (processos), mas é a mais adequada quando se trata de sinistros atípicos ou de sinistros grandes (sinistros que ultrapassam determinados limites de valor, previamente estabelecidos pela companhia). Para além de que não permite estimar os IBNR, dado que se tratam de sinistros que ainda não foram participados;
- b) Na globalidade, por meio de técnicas estatísticas, que permitem estimar a evolução dos pagamentos futuros dos sinistros.

Na presente pesquisa, a estimativa das provisões a realizar é na globalidade, pois pretende-se estimar a evolução dos custos com sinistros no seguro automóvel em Moçambique no período de 2024 a 2028.

2.1.5. Série Temporal

Uma série temporal é uma sequência de realizações (observações) de uma variável ao longo do tempo. Dito de outra forma, é uma sequência de pontos (dados numéricos) em ordem sucessiva, geralmente ocorrendo em intervalos uniformes. Portanto, uma série de tempo é simplesmente uma sequência de números colectados em intervalos regulares durante um período de tempo (Gujarati, 2006).

Uma série temporal é uma realização de um processo estocástico que consiste em uma sequência de variáveis aleatórias (Morretin e Tolo, 2006). Seja $\{Z_{(t)}\}, t = 0, 1, 2, \dots, n$, uma série temporal de interesse.

Morettin e Tolo (2006) afirmam que uma série temporal é um conjunto de informações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e fica-se interessado em analisar e modelar esta

dependência. Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante para a análise de séries temporais a ordem dos dados é crucial. Referindo-se ao parâmetro t como sendo o tempo, a série $Z_{(t)}$, poderá ser função de algum outro parâmetro físico como espaço ou volume.

Uma série temporal ou cronológica é composta por um conjunto de observações contínuas, ordenadas em intervalos temporais igualmente espaçados (Gujarati, 2006).

A acção mais importante a tomar quando se analisam dados é a sua visualização através de gráficos. As características básicas dos dados, tais como comportamento e valores estranhos são mais facilmente detectáveis em gráficos. Por vezes, os gráficos também sugerem explicações possíveis para algumas variações nos dados. Por exemplo, disputas laborais geralmente afectam as séries temporais da produção; mudanças governamentais afectam séries temporais económicas. Os gráficos são a maneira mais eficaz de identificar o efeito de tais eventos nos dados. O tipo de dados determina o tipo de gráfico mais apropriado.

Para as séries temporais, a representação gráfica mais usual é o gráfico temporal, no qual os dados são representados em função do tempo. Um gráfico temporal revela qualquer tendência ao longo do tempo, comportamento sazonal regular e outras características sistemáticas dos dados. Estas características necessitam de ser identificadas para que possam ser incorporadas num modelo estatístico.

Murteira e Black (1983), afirmam serem muitos os objectivos com que se procede ao estudo das séries temporais. Porém, podem agrupar-se num conjunto de cinco grupos fundamentais: caracterização, previsão, modelagem e comparações. Podendo serem sintetizados da seguinte maneira:

- Caracterização - determinação das propriedades da série (padrões de tendência, existência de variação sazonal, regimes não-lineares, auto correlação, leis de potência, etc);
- Previsão - prever variações passadas e futuras em relação a uma sequência de medidas;
- Modelagem - aplicar ferramentas para modelagem matemática e ou fenomenológica da série;

- Comparação - controle estatístico de processos e qualidade dos dados e usar a variação numa série para explicar a variação em outra.

Como o objectivo da pesquisa é de estimar através da metodologia de Box-Jenkins um modelo ARIMA adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 à 2028, os objectivos propostos por Murteira e Black (1983) são aplicáveis.

Segundo Makridakis; Wheelwright e Hyndman (1998), os modelos de séries temporais podem classificar baseando-se no número de séries temporais envolvidas no modelo e se classificam em:

- Modelo univariados - grupo de dados que incluem todos os métodos que se baseiam em uma única série histórica, ou seja série temporal é explicado (prevista) apenas por seus valores passados (modelos auto regressivos).
- Modelos multivariados ou casuais - grupo de dados que modela simultaneamente duas ou mais séries temporais permitindo, assim, relações de interdependência e casualidade, mas sem qualquer exigência com relação á direcção da casualidade entre elas. A série temporal é explicada (prevista) pelos seus valores passados e também pelos valores passados de outras séries.

Murteira e Black (1983) afirmam que qualquer estudo que envolve séries temporais, o primeiro e fundamental passo é a representação gráfica da série, visto que, revela padrões de comportamento importantes, como a tendência de crescimento (ou decrescimento), padrões cíclicos, alterações estruturais, observações aberrantes, etc. Para Makridakis et al (1998) é muito importante, uma representação gráfica para um estudo prévio da série temporal. Estes autores acrescentam que, a representação gráfica mostra a visão geral do comportamento da série mas porém, é necessário efectuar análises estatísticas adequadas para uma conclusão confiável do real comportamento da série.

A representação gráfica da série temporal envolve uma variável Y ilustrada por meio da construção de um gráfico de Y em função do tempo (t), a partir do sistema de eixos cartesiano (Murteira e Black, 1983).

Segundo Morettin e Toloí (2006), uma série temporal pode adoptar basicamente dois tipos de modelos: o modelo aditivo e o modelo multiplicativo. O modelo aditivo pode exprimir-se como:

$$Y_t = T_t + S_t + C_t + I_t \quad (1)$$

Onde:

- Y_t representa o valor da série cronológica num dado período de tempo.
- Tendência (T_t): é caracterizada como um movimento regular e contínuo de longo prazo que pode ser crescente, decrescente ou constante. Também conhecida como ciclo-tendencial, consiste no comportamento a longo prazo de um dado fenómeno;
- Sazonalidade (S_t): corresponde às oscilações crescentes ou decrescentes que sempre ocorrem em um determinado período do dia, da semana, do mês ou do ano. Descreve as variações em relação à tendência em geral que se repete com periodicidade anual. Estas variações ocorrem todos os anos seguindo um padrão e podem ter causas naturais (ex: as estações do ano têm efeitos sobre actividades como agricultura, o turismo, etc.) ou sociais (ex: disposições físicas, costumes e tradições sociais etc.).
- Movimentos oscilatórios ou Ciclo (C_t): Corresponde às oscilações periódicas de longo prazo em torno da tendência. Estão associados as fases de expansão e de depressão que afectam o sistema económico, não apresentando qualquer periodicidade definida e verifica-se em séries de produção, preços, etc. Quando os ciclos são longos são difícil separa-los da tendência.
- A variação Irregular (I_t): Esta componente das séries temporais engloba as variações remanescentes após identificação da tendência e dos movimentos cíclicos e sazonais. Representa dois tipos de variações:
 - a) Alterações esporádicas
 - b) Movimentos aleatórios que agrupam a influência do conjunto de factores que sendo explicativos não são explicitados devido ao seu diminuto peso individual e ao desconhecimento do sentido da sua influência.

Existe ainda uma componente habitualmente designada por resíduo ou ruído (ϵ) que tem uma natureza aleatória, originada por diferentes causas, cuja importância é habitualmente reduzida e de curta duração temporal.

Esta decomposição conceptual das séries temporais nas suas diferentes componentes fornece-nos uma orientação útil para descrever o seu comportamento. O enfoque mais moderno na análise de séries temporais, envolve a construção de um modelo formal no qual os diferentes componentes estão explícita ou implicitamente presentes, para descrever o comportamento dos dados da série. De acordo com Gujarati (2006), podemos apresentar a série consistindo da soma das duas componentes (Modelo Aditivo).

$$Y_t = T_t + S_t + C_t + \varepsilon_t \quad (2)$$

Este modelo pressupõe a independência entre as quatro componentes das séries temporais, ou seja, pressupõe que a importância de cada uma delas é independente do valor das outras.

Em razão desta limitação, prefere-se frequentemente o modelo multiplicativo.

O modelo multiplicativo pode exprimir-se como:

$$Y_t = T_t \cdot S_t \cdot C_t \cdot I_t \quad (3)$$

Onde os símbolos têm o mesmo significado que no modelo aditivo.

A hipótese implicitamente assumida no modelo multiplicativo é a da interdependência entre os quatro factores.

De notar que o modelo multiplicativo pode transformar-se num modelo linear aditivo através da utilização de logaritmos, esta será dada por:

$$\log Y_t = \log T_t + \log S_t + \log C_t + \log I_t \quad (4)$$

2.1.5.1. Diferenciação e Ordem de Integração de uma Série Temporal

Gujarati (2006) afirma que a maior parte das séries temporais na área de economia em geral tornam estacionárias somente depois que atingirem as suas primeiras diferenças. Morettin (2004) explica que esse procedimento pode suscitar a perda de informações de longo prazo.

Segundo England e Verall (2002), um filtro especial é muito útil, para remover uma componente de tendência, consiste em diferenciar a série até que ela se torne estacionária.

Morettin (2004) afirma que, a maioria dos procedimentos de análise estatística de séries temporais supõem que estas sejam estacionárias. Quando elas não apresentam esta característica, se deve transformá-las para obter a estacionaridade. A transformação mais

comum é a operação de diferenciação que consiste em tornar diferenças de valores sucessivos da série original. A série constituída pelas primeiras diferenças de uma série Y_t indicada por ΔY_t é definida como:

$$\Delta Y_t = Y_t - Y_{t-1} \quad (5)$$

Segundo Gujarati (2006), se a série temporal X_t ou de Y_t for diferenciada uma vez e a série resultante for estacionária, diz-se que a série X_t ou de Y_t é integrada de ordem 1. Analogamente, se forem necessárias duas diferenciações seguidas para atingir a estacionaridade, isto é, se for preciso tornar a primeira diferença da série diferenciada, diz-se que a série original é integrada de ordem 2, e assim sucessivamente.

Exemplo se Y é I (2) então $\Delta \Delta Y_t = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$ se tornara estacionária.

No geral:

- Se uma série Y_t é não-estacionárias tem de ser diferenciada d vezes para se tornar estacionária, diz-se que é integrada de ordem d e denota-se $Y_t \sim I(d)$.
- Se uma série Y_t é estacionária desde o inicio, isto é, estacionária em nível, diz-se que ela é integrada de ordem zero, e denota-se $Y_t \sim I(0)$.

2.1.6. Modelos de Séries Temporais

2.1.6.1. Modelos Auto-Regressivos e de Médias Móveis

Wheelwright (1985) descreve três procedimentos capazes de representar as observações de uma série temporal estacionária: modelo Auto-regressivo (AR), modelo de Médias Móveis (MA) e o modelo Auto-regressivo e de Medias Móveis (ARMA).

2.1.6.2. Modelos Auto-Regressivos

Um modelo auto-regressivo de ordem p , representa-se por AR (p) e define-se como:

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + \varepsilon_t \quad (6)$$

Onde Y_t serve como variável dependente, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ são valores desfasados de Y_t , ε_t é um ruído branco (erro com media nula, variância constante e covariância nula, e os b_j ($j = 0, \dots, p$) são parâmetros.

O processo é dito de ordem p porque o maior desfasamento presente é Y_{t-p} .

Caso as observações da série temporal possam ser representadas pela equação 6, a ordem do modelo poder ser determinada e os parâmetros estimados, é possível prever o valor futuro da série em análise.

2.1.6.3. Modelo de Médias Móveis

Um modelo a média móvel de ordem q , representa-se por MA (q) e define-se como:

$$Y_t = b_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} \quad (7)$$

Onde a notação adaptação adoptada tem um significado idêntico ao modelo AR (a utilização de sinais negativos é uma mera convenção).

A equação 6 é similar a equação 7, excepto pelo facto de o valor previsto para uma dada observação α depender dos valores dos erros observados em cada período passado, ao invés das observações propriamente ditas (Wheelwright, 1985).

2.1.6.4. Modelos Auto-Regressivos e de Médias Móveis

Wheelwright (1985) especifica o modelo misto Auto-regressivo e de Médias Móveis (ARMA) através da equação 8, como sendo a combinação dos modelos AR e MA.

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} \quad (8)$$

Analisando a equação 8 é possível verificar que os modelos ARMA relacionam os valores futuros com as observações passadas, assim como também com os erros passados apurados entre os valores reais e os previstos.

Em geral, em um processo ARMA (p, q), haverá p termos auto-regressivos e q termos media móvel.

2.1.6.5. Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA)

Segundo Gujarati (2006), se tivermos de diferenciar uma série temporal d vezes para torna-la estacionária e então aplicarmos a ela o modelo ARMA (p, q), dizemos que a série temporal original é ARIMA (p, d, q).

Se uma série não estacionária é adicionada a um modelo ARIMA, então a especificação geral do modelo torna-se ARIMA (p, d, q).

Portanto, se tivermos de diferenciar uma série temporal d vezes para torná-la estacionária e então aplicarmos a ela o modelo ARMA (p, q), dizemos que a série temporal original é ARIMA (p, d, q), ou seja, é uma série temporal auto-regressiva integrada de média móvel, em que p indica o número de termos auto regressivos, d, o número de vezes que a série deve ser diferenciada para se tornar estacionária e q, o número de termos de média móvel.

Naturalmente que se d = 0 (isto é uma série é antes de mais nada estacionária), então o modelo ARIMA (p, d=0,q) = ARMA (p, q). Note que um processo ARIMA (p, 0,0) significa um processo puramente estacionário AR (p); e um ARIMA (0,0,q) significa um processo puramente estacionário MA (q). Deste modo, dado p, d e q podem facilmente dizer que o processo esta sendo modelado.

Em geral os modelos ARIMA (p, d, q) podem gerar padrões diferentes das funções de ACF e PACF, pelo que é inútil definir regras para identificação dos modelos ARIMA.

Contudo, os modelos simples de AR (p) e MA (q) podem fornecer uma informação útil para a identificação destes modelos. É importante salientar que diferentes modelos podem gerar a mesma qualidade de resultados em termos de previsão. Na prática, é sempre recomendável trabalhar com valores de p, d e q que se situam no intervalo de 0,1 ou 2. Porém, parece surpreendente que tais valores por mais pequenos que possam resolver uma vasta gama de situações em termos de previsão.

2.1.7. Regressão

Segundo Gujarati (2006), a regressão envolve uma variável dependente e uma ou mais variáveis independentes. A regressão é simples quando envolve uma variável dependente e uma independente e pode ser considerada composta quando envolve mais do que uma variável independente.

Para Morettin (2004), a análise da regressão tem por objectivo permitir-nos estimar o valor médio da variável dependente com base nas variáveis independentes e testar as hipóteses acerca da validade de modelos (a natureza da dependência).

Segundo Gujarati (2006), são os seguintes os pressupostos da regressão:

- **Linearidade:** o modelo de regressão é linear em parâmetros;
- Os valores da variável independente são fixos em amostras repetidas casualmente;
- O valor médio do termo erro (ϵ_t) para cada intervalo da amostra independente é igual a zero (0);
- **Homocedasticidade:** para todos os valores da variável independente a variância do termo erro (ϵ_t) é igual em todas as observações;
- **Não-autocorrelação,** entre os valores da variável independente e o termo erro (ϵ_t);
- O número de observações deve ser superior ao número de parâmetros;
- O modelo deve ser correctamente especificado; deve-se incluir todas as variáveis relevantes e assegurar que a fórmula funcional é correcta;

- **Não a Multicolinearidade**, se esta existir dificilmente chegaremos a conclusão entre as variáveis independentes com as variáveis dependentes.

2.1.8. Estacionaridade

Segundo Morretin e Toloí (2006), uma das suposições mais frequentes que se faz a respeito de uma série temporal é que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, reflectindo alguma forma de equilíbrio.

Formalmente, estacionaridade significa que médias e variâncias são constantes ao longo do tempo em análise. Gujarati (2006), também apresenta formalização desse conceito afirmando que a série é estacionária se o valor da covariância entre os dois períodos depender apenas da distância ou defasagem entre os dois períodos.

Enquadram-se perfeitamente no significado do termo estacionaridade, todas as definições apresentadas acima, deste modo, mostrando a importância do termo para análise de séries temporais e para o presente trabalho, visto que, este é concentrado neste conceito.

2.1.8.1. Testes de Estacionaridade

Segundo Gujarati (2006), são séries estacionárias se a sua média e variância forem constantes ao longo do tempo e o valor da covariância entre os dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efectivo em que a covariância é calculada, com as seguintes propriedades:

$$\text{Média: } E(Y_t) = \mu$$

$$\text{Variância: } Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$\text{Covariância: } CoVar(K) = \gamma_K = E[(Y_t - \mu) \times (Y_{t+K} - \mu)]$$

Onde γ_k representa covariância ou auto covariância na defasagem k , entre os valores de Y_t e Y_{t+k} , ou seja entre dois valores Y separados por k períodos.

Se a série temporal não for estacionária, pode também ser considerada de fracamente estacionária e que se deve a mudança de média.

Para a modelação de toda a série temporal temos que nos preocupar não só com a estacionaridade, pois esta é condição necessária mas não suficiente. Neste processo é também fundamental que a série de resíduos apresente uma média e variância constantes no tempo, esta série é conhecida como Ruído Branco.

Segundo Murteira e Black (1983), Ruído Branco ou “White Noise” é um processo fracamente estacionário até segunda ordem, cuja importância não advém de si próprio mas do papel fundamental que representa na construção de outros processos estacionários particularmente relevantes na modelação de sucessões cronológicas.

Apesar de quase a totalidade dos métodos de previsão de séries temporais estar fundamentada apenas na análise das observações da série de interesse para a especificação de algum modelo que descreva essas observações, alguns procedimentos de previsão tentam explicar o comportamento de uma série temporal pela evolução dos fenómenos observacionais de outras séries. Desta forma, dependendo do número de séries temporais envolvidas na modelagem, classificam-se os métodos de previsão em univariados, funções de transferência e multivariados (Ljung e Box, 1978).

2.1.7.2. Correlograma

Uma forma simples para verificar se uma série é estacionária é através do coeficiente da autocorrelação da amostra, representado pela seguinte relação:

$$r_k = \frac{\sum (Y_t - \bar{Y}_t)(Y_{t-k} - \bar{Y}_{t-k})}{\sqrt{\sum (Y_t - \bar{Y}_t)^2 \sum (Y_{t-k} - \bar{Y}_{t-k})^2}} = \frac{Cov(Y_t, Y_{t-k})}{S_{Y_t} S_{Y_{t-k}}}, \quad -1 \leq r_k \leq 1 \quad (9)$$

O correlograma da amostra representa um gráfico dos valores da função de autocorrelação de r_k em relação aos desfasamentos k .

Uma indicação de estacionaridade da série Y_t seria quando os valores da função de autocorrelação são nulos a partir do primeiro desfasamento, ou quando os valores de r_k caem rapidamente para zero.

Existem testes utilizados para testar a significância estatística conjunta dos valores r_k :

(i) O teste de Box-Pierce, baseado na estatística Q definida como:

$$Q = n \sum_{k=1}^m r_k^2 \sim \chi_m^2 \quad (10)$$

Onde: n é o nº de observações e m = extensão do desfasamento.

A série é estacionária quando $Q < \chi_{crítico}^2$, uma vez que a hipótese nula implica que todos os r_k são iguais a zero.

2.1.7.3. Teste Geral de *Dickey-Fuller* Aumentado (ADF)

De acordo com Gujarati (2006), existem diversos testes que permitem analisar a questão da ordem de integração de uma série observada X_t , que são geralmente conhecidos como testes de não estacionaridade ou testes de raiz unitária. Os testes mais comuns são os de *Dickey-Fuller* ou DF e o teste Aumentado de *Dickey Fuller* ou ADF.

Segundo Gujarati (2006) os testes de DF e ADF também conhecidos por estatística ou teste de t-tau consistem em estimar o seguinte:

Se: $\delta = (\rho-1)$; ε_t são erros aleatórios independentes e identicamente distribuídos com média e variâncias constantes, e onde testa se a seguinte hipótese;

Ho: $\rho-1 = 0$ (a série tem raiz unitária)

Ha: $\rho-1 < 0$ (a série não tem raiz unitária)

O teste de estacionaridade pode ser aplicado ao caso geral, (assumindo um processo autoregressivo superior da primeira ordem) estimando a seguinte equação:

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} \quad (11)$$

Incluimos um número de variáveis desfasadas da primeira diferença de Y_t , até obtermos erros u_t não autocorrelacionados.

Estimando pelo método de mínimos quadrados (OLS) o seguinte critério pode ser considerado:

A série é estacionária quando $\tau_{estatístico} > \tau_{crítico}$

Alternativamente, podemos testar a hipótese da significância conjunta:

$H_0 : \beta = \delta^* = 0$ (não estacionariedade ou *trend* estocástico)²

$H_a : \beta, \delta^* < 0$ (estacionariedade)

Neste caso pode ser utilizado um *teste-F* para testar a hipótese de estacionariedade.

O critério de estacionariedade implica que $F_{estatístico} > F_{crítico}$ ($\tau_{crítico}$ e $F_{crítico}$ são dados pelas tabelas próprias de Dickey-Fuller).

2.2. Métodos de Previsão Utilizados no Sector de Seguros

Neste trabalho apresentam-se os métodos para estimar as provisões para sinistros – métodos determinísticos e métodos estocásticos. A característica comum dos métodos de *Chain Ladder*, *Thomas Mack* e de *Bootstrap* apresentados nos pontos seguintes é que se baseiam em triângulos *Run-off* e pressupõem que existe proporcionalidade entre colunas. Isto é, têm como fundamento principal que o comportamento das evoluções passadas, ou seja, os montantes dos sinistros já liquidados, se mantém estáveis ao longo dos anos.

Mais concretamente, neste estudo aplica-se o método de *Box-Jenkins*, devido à sua simplicidade e facilidade de pôr em prática e obtenção de dados reais.

Contudo, os modelos determinísticos no cálculo das provisões não envolvem uma medida de erro de estimação. Desta forma, torna-se impossível quantificar tanto o grau de incerteza como a variabilidade das estimativas obtidas, que conseqüentemente não nos permite quantificar o nível de prudência a aplicar no processo de estimação.

² Sob a hipótese nula a equação $\Delta Y_t = a + \beta T + \delta^* Y_{t-1} + u_t$ reduz-se para $\Delta Y_t = a + u_t$ indicando que a variável Y_t segue um padrão (*trend*) estocástico.

Com o desenvolvimento da área computacional e da área de Actuariado Não Vida, surgiram diversos modelos estocásticos que, além de produzirem o valor esperado dos montantes de indemnizações por liquidar, também produzem a variação sobre esse valor esperado, permitindo a obtenção de medidas de erro associadas às previsões.

Embora sejam de uma aplicação e interpretação mais difícil, os métodos estocásticos possuem a principal vantagem de ser possível a construção de intervalos de confiança, permitindo, desta forma, monitorizar a estabilidade dos pagamentos futuros e a detenção prévia de possíveis desvios dos pagamentos registados em exercícios anteriores.

Assim, a seguir, apresenta-se os métodos de previsão mais utilizados no sector de seguros para a provisão para sinistros, nomeadamente o método de *Chain Ladder*, o método de *Thomas Mack*, o método de *Bootstrap* e o método de *Box-Jenkins*.

2.2.1. Método de *Chain Ladder*

O clássico método *Chain Ladder* é um algoritmo determinístico de previsão e uma das técnicas actuariais mais usadas e conhecidas, principalmente por ser simples e prático, com resultados por vezes muito próximos da realidade. No entanto, não assume qualquer distribuição de probabilidade para os dados.

Este método parte do pressuposto base que o desenvolvimento passado, evidenciado na matriz de informação histórica, continuará a verificar-se no futuro.

Tabela 1. Triângulo *Run-off* incremental.

| $X_{1,J}$ | 1 | 2 | ... | j | ... | I |
|-----------|-----------|-----------|-----|-----------|-----|-----------|
| 1 | $X_{1,1}$ | $X_{1,2}$ | ... | $X_{1,j}$ | ... | $X_{1,J}$ |
| 2 | $X_{2,1}$ | $X_{2,2}$ | ... | $X_{2,j}$ | ... | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | |
| i | $X_{i,1}$ | $X_{i,2}$ | ... | $X_{i,j}$ | | |
| ⋮ | ⋮ | ⋮ | | | | |
| I | $X_{I,1}$ | | | | | |

Fonte: adaptado pelo Autor, a partir de Guambe (2022).

Em particular, o método *Chain Ladder* assume a existência de proporcionalidade entre colunas da matriz, ou seja, entre anos de desenvolvimento (England e Verrall, 2002). A técnica *Chain Ladder* faculta uma previsão pontual das reservas, para tal usa como base a matriz de *run-off* de montantes pagos acumulados.

Esta informação histórica dispõe-se em forma de uma matriz *Run-off* (matriz triangular superior). Trata-se de uma matriz incompleta, onde cada linha representa o ano de ocorrência do sinistro e cada coluna representa o ano de desenvolvimento, isto é, o ano em que é liquidado o sinistro.

Segundo Guambe (2022), a matriz *Run-off* pode apresentar dois tipos de estrutura:

- a) Incremental – apresenta os montantes pagos observados no ano de ocorrência i e de desenvolvimento j – representa-se por $x_{i,j}$;
- b) Acumulada – obtém-se através da soma dos pagamentos por ano de ocorrência i , e corresponde ao total de montantes pagos no ano de ocorrência i , até ao ano de desenvolvimento j :

O primeiro passo consiste em calcular os factores de desenvolvimento, de acordo com a seguinte fórmula: $C_{i,j} = \sum_{k=0}^j X_{i,k}$, $0 \leq i \leq n$, $0 \leq j \leq n-1$ (12)

Após a obtenção dos valores dos factores de desenvolvimento, é possível calcular os valores referentes ao triângulo inferior da matriz de pagamentos de sinistros, aplicando aos montantes da diagonal principal da matriz de montantes acumulados (com valores conhecidos) o respectivo factor de desenvolvimento. Com uma nova diagonal obtida, volta-se a multiplicar os valores desta pelo factor de desenvolvimento correspondente e assim sucessivamente até a parte inferior da matriz estar completa. Ou seja, cada novo elemento $\hat{r}_{i,i+j} > n+1$, do ano de ocorrência i e ano de desenvolvimento j é calculado recursivamente através de:

$$\hat{r}_{i,n} = \hat{c}_{i,n} - C_{i,n-i-1}, 0 \leq n \leq i \quad (13)$$

Por fim, para obter o montante total da reserva utiliza-se a seguinte expressão:

$$\hat{R} = \sum_{n=1}^i \hat{R} \quad (14)$$

O método clássico de *Chain Ladder*, segundo Zehnwirth (1989) é um algoritmo determinístico de previsão. É um dos métodos mais conhecidos e utilizados, devido principalmente à sua simplicidade e por não assumir qualquer distribuição de probabilidade.

Este método assume que os anos de desenvolvimento (colunas do triângulo) são directamente proporcionais e que os factores de desenvolvimento que se utilizam para a estimação dos pagamentos futuros se mantêm estáveis ao longo dos anos de ocorrência dos sinistros.

Considera também que os anos de ocorrência são independentes e que a sinistralidade, ao longo do tempo, pode variar (Guambe, 2022).

Consideramos um triângulo (de montantes de pagos acumulados) de sinistros com informação referente aos pagamentos acumulados $\{C_{i,j} = i = 0, \dots, I; j = 0, \dots, J\}$ em que i refere-se aos anos de origem e j refere-se aos anos de desenvolvimento:

Segundo England e Verrall (2002), pode-se descrever o método de *Chain Ladder* nos seguintes passos:

1. Verificação dos seguintes pressupostos:
 - Existem factores de desenvolvimento $f_1, \dots, f_{I-1} > 0$ em que:

$$E \left(\frac{C_{i,k+1}}{C_{i,k}} / C_{i,0}, \dots, C_{i,k} \right) = f_k, 0 \leq i \leq I, 0 \leq k \leq I \quad (15)$$

$$E (C_{i,k+1} / C_{i,0}, \dots, C_{i,k}) = C_{i,k} \times f_k, 0 \leq i \leq I, 0 \leq k \leq I \quad (16)$$

As variáveis $C_{i,j}$ de diferentes anos de ocorrência são independentes, isto é:

$$\{C_{i,1}, \dots, C_{i,I}\}, \{C_{j,1}, \dots, C_{j,J}\} \quad i \neq j \text{ são independentes} \quad (17)$$

- A variância das quantidades acumuladas de pagamentos $C_{i,j}$ expressa-se como:

$$Var (C_{j,k+1} / C_{j,1}, \dots, C_{j,k}) = C_{j,k} \times \sigma_k^2, 1 \leq i \leq n, 1 \leq j \leq n-1 \quad (18)$$

2. Determinação dos factores de desenvolvimento:

Os factores de desenvolvimento (f_j) recolhem uma estimação da variação da sinistralidade entre um ano de desenvolvimento e o seguinte. Calculam-se da seguinte forma:

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} ; j = 0, \dots, j-1 \quad (19)$$

3. Determinação dos factores de projecção:

Através do produto dos factores de desenvolvimento obtemos os factores de projecção:

$$\hat{F}_k = \prod_{j=k}^{J-1} f_j ; k = 0, \dots, J - 1 \quad (20)$$

4. Estimação das quantias acumuladas para o último ano de desenvolvimento:

O último ano de desenvolvimento, conhecido como “*ultimate*”, facilita o cálculo das provisões técnicas.

Os primeiros valores a serem estimados para o triângulo de sinistros são as quantias acumuladas de pagamentos ($\ddot{C}_{i,j}$):

$$\ddot{C}_{i,j} = C_{i,j-1} \times F_{j-1} ; i = 1, \dots, I, 0, \dots, J - 1 \quad (21)$$

5. Determinação das reservas por ano de origem:

Calculam-se as reservas por ano de origem (\ddot{R}_i) para cada ano de origem, realizando a diferença entre a estimação da quantidade acumulada do último ano de desenvolvimento e o último valor observado da mesma:

$$\ddot{R}_i = \ddot{C}_{i,j} - C_{i,j-1}, i = 1, \dots, I, 0, \dots, J - 1 \quad (22)$$

6. Determinação da reserva total:

A reserva total (\hat{R}), isto é, o valor total esperado das responsabilidades com sinistros ainda pendentes, calcula-se como o somatório das reservas de origem (\ddot{R}_i):

$$\hat{R} = \sum_{i=1}^I \ddot{R}_i \quad (23)$$

O método *Chain Ladder* é um método de fácil utilização e aplicabilidade. Mas também apresenta algumas limitações, tal como pressuposto de que existe independência entre os diversos anos de ocorrência e assumir que o factor de desenvolvimento é o mesmo ao longo dos anos de ocorrência dos sinistros.

Os modelos determinísticos apresentam apenas uma estimativa pontual da reserva, enquanto nos modelos estocásticos, em que as soluções possíveis dependem da distribuição de probabilidade usada, se podem construir intervalos de confiança. Isto não significa que os

métodos estocásticos são melhores que os métodos tradicionais, apenas fornecem mais informação sobre aspectos da reserva.

Este tipo de metodologia pode ser ajustada, isto é, não é necessariamente obrigatório a utilização de todos os dados históricos existentes na matriz de dados, até porque isso é variável de acordo com o estudo realizado. Na análise, pode ser mais adequado utilizar factores de desenvolvimento referentes a menos anos ou mesmo média de factores. Este tipo de critério e selecção é efetuado pelos actuários da empresa de seguros.

2.2.2. Método de *Thomas Mack*

Os métodos estocásticos e a sua aplicação no cálculo de provisão de sinistros, têm sido alvo de grande interesse nos últimos tempos, mesmo com a dificuldade inerente na utilização e compreensão dos mesmos sem o recurso a *softwares* adequados e tendo em conta, que é possível utilizar métodos determinísticos (*Chain Ladder*), como opção válida no cálculo de valores de provisão.

Os métodos estocásticos são como uma ampliação das técnicas tradicionais, que permitem obtenção de mais informação e avaliações.

Segundo Borginho (2001), este modelo é classificado como não paramétrico, dado que não assume que os dados seguem uma determinada distribuição de probabilidade.

Um dos objectivos com esta metodologia é manter as mesmas capacidades de previsão do *Chain Ladder*, mas proporcionar a produção de mais informação sobre os estimadores que permitam construir intervalos de confiança. Os intervalos de confiança possibilitam a avaliação da variabilidade de estimativas futuras e detectam eventuais desvios em relação aos valores registados nos anos considerados.

Caso os valores das provisões para sinistros estejam dentro do intervalo de confiança, com um determinado nível de confiança previamente seleccionado, podemos assumir que as provisões constituídas são adequadas para as responsabilidades assumidas. Se os valores estiverem abaixo do limite inferior, a seguradora deverá de reforçar as provisões. Se por outro lado, os

valores são superiores ao limite superior, então existe um excesso de provisionamento e será adequado a seguradora libertar alguma provisão.

Mack (1993) refere-se que antes da aplicação do modelo de *Thomas Mack*, é necessário testar os dados em estudo, ou seja, ver se o método é adequado para aplicar aos mesmos através da verificação de três pressupostos base. São os mesmos que o método de *Chain Ladder*, mas neste não são testados já que se assume que são implícitos.

Considera-se que $C_{i,k}$ corresponde ao montante de sinistros totais acumulados do ano de ocorrência i , $1 \leq i \leq I$ até ao ano de desenvolvimento k , $1 \leq k \leq I$.

Antes de aplicar o método *Thomas Mack* é necessário testar os dados em estudo através da verificação dos três pressupostos descritos anteriormente no método *Chain Ladder*. A verificação destes três pressupostos permite-nos avaliar a possibilidade do modelo ser adequado ao nosso conjunto de dados e, por conseguinte, assegura-nos uma maior fiabilidade do processo de estimação.

Após a verificação dos três pressupostos estaremos em condições de determinar uma estimação do Erro Padrão (EP) e construir os intervalos de confiança associados às reservas obtidas através da aplicação do Método de *Chain Ladder*.

2.2.2.1. Estimação das Reservas

Como foi dito anteriormente, através do método de *Chain Ladder* estimam-se os factores de desenvolvimento utilizando a seguinte expressão:

$$\hat{f}_k = \frac{\sum_{i=0}^{n-k} C_{i,j+1}}{\sum_{i=0}^{n-k} C_{i,j}}, \quad 0 \leq j \leq n-1 \quad (24)$$

Consequentemente, através da seguinte expressão, determinam-se as estimativas dos futuros pagamentos:

$$\hat{C}_{i,n} = C_{i,n-i} \times \hat{f}_{n-i} \times \dots \times \hat{f}_{n-1}, \quad 0 \leq i \leq n \quad (25)$$

De seguida, calculam-se as provisões a constituir para cada ano de ocorrência i através de:

$$\bar{R}_i = \bar{C}_{i,n} - C_{i,n-1}, 0 \leq i \leq n \quad (26)$$

Por fim, a estimativa do valor total das provisões é obtido através de:

$$\hat{R} = \sum_{i=0}^n \hat{R}_i \quad (27)$$

2.2.2.2. Variabilidade das Reservas

Para determinar uma medida de variabilidade das estimativas calcula-se o erro padrão (EP) dos estimadores, que corresponde à raiz quadrada do erro quadrático médio (EQM):

$$EP = \sqrt{EQM} \quad (28)$$

Como $E\hat{Q}M(\hat{R}_i) = E\hat{Q}M(\hat{C}_{i,n})$, segundo Mack (1993) pode estimar-se $E\hat{Q}M(\hat{C}_{i,n})$ através de:

$$EQM(\hat{C}_{i,n}) = \hat{C}_{i,n}^2 \times \sum_{k=n-1}^{n-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{C}_{i,k}} + i = \frac{1}{\sum_{i=0}^{n-k} C_{i,k}} \right) \quad (29)$$

Um estimador centrado para $\hat{\sigma}_k^2$ é dado por:

$$\hat{\sigma}_k^2 = \frac{1}{n-k-1} \times \sum_{i=0}^{n-k} C_{i,k} \times \left(\frac{C_{i,k+1}}{C_{i,k}} - \hat{f}_k \right)^2 \text{ para } 0 \leq k \leq n-2 \quad (30)$$

- Por sua vez, a equação não produz um estimador $\hat{\sigma}_{n-1}^2$. Portanto: se $f_{n-1} = 1$ e se espera que o desenvolvimento dos sinistros termine antes de $n-1$, então $\hat{\sigma}_{n-1}^2 = 0$;
- Caso contrário, pode realizar-se uma projecção da sequência $(\sigma_1, \sigma_2, \dots, \sigma_{T-3}, \sigma_{T-2})$, por exemplo, através da média ou de uma regressão loglinear; outra alternativa é definida por:

$$\hat{\sigma}_{n-1}^2 = \min\left(\frac{\hat{\sigma}_{n-3}^4}{\hat{\sigma}_{n-3}^2}, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2)\right)$$

(31)

Assumindo que se verifica $\frac{\hat{\sigma}_{n-3}^2}{\hat{\sigma}_{n-2}^2} = \frac{\hat{\sigma}_{n-2}^2}{\hat{\sigma}_{n-1}^2}$, pelo menos enquanto, $\hat{\sigma}_{n-3}^2 > \hat{\sigma}_{n-2}^2$.

Sendo $\hat{R} = \sum_{i=0}^n \hat{R}_i$, pode assumir-se que:

$$EQM(\hat{R}) = EQM\left(\sum_{i=0}^n \hat{R}_i\right) = EQM\left(\sum_{i=0}^n \hat{C}_{i,n}\right) \quad (32)$$

Por sua vez, $EQM(\sum_{i=0}^n \hat{C}_{i,n})$ pode ser estimado através da seguinte expressão:

$$EQM(\sum_{i=0}^n \hat{C}_{i,n}) \sum_{i=1}^n \left\{ EQM(\hat{C}_{i,n}) + C_{i,n}(\sum_{j=i+1}^n \hat{C}_{i,j}) \times \sum_{k=n+1-i}^{n-1} \frac{2\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\sum_{l=0}^{n-k} C_{l,k}} \right) \right\} \quad (33)$$

Assumindo que as estimativas anteriormente obtidas seguem uma distribuição Normal, determinam-se os intervalos de confiança para as estimativas \hat{R}_i e \hat{R} respectivamente, através de:

$$\left[\hat{R}_i - \Phi^{-1}[1 - \alpha/2] \times \sqrt{E\hat{Q}M(\hat{R}_i)}; \hat{R}_i + \Phi^{-1}[1 - \alpha/2] \times \sqrt{E\hat{Q}M(\hat{R}_i)} \right] \quad (34)$$

e

$$\left[\hat{R} - \Phi^{-1}[1 - \alpha/2] \times \sqrt{E\hat{Q}M(\hat{R})}; \hat{R} + \Phi^{-1}[1 - \alpha/2] \times \sqrt{E\hat{Q}M(\hat{R})} \right] \quad (35)$$

Em que $\Phi^{-1}[1 - \alpha/2]$ representa o quantil de probabilidade $1 - \alpha/2$ da distribuição Normal Padrão.

2.2.2.3. Verificação dos Pressupostos do Método *Thomas Mack*

1.º Pressuposto

O primeiro pressuposto traduz-se na proporcionalidade entre as colunas da matriz de desenvolvimento, isto é, na proporcionalidade dos anos de desenvolvimento.

A equação (3.1), para um qualquer k fixo e $i = 0, \dots, n$, traduz uma relação linear entre os montantes $C_{i,k+1}$ e $C_{i,k}$ com um declive igual a f_k .

Se considerarmos um modelo de regressão do tipo:

$$Y_i = c + x_i b + \varepsilon_i, 0 \leq i \leq n \quad (36)$$

Onde c e b representam os coeficientes de correlação e ε_i o erro.

Sendo $E(\varepsilon_i) = 0$, $c = 0$ e $b = f_k$, por substituição, temos $E(Y_i) = x_i f_k$.

Para verificar este pressuposto constrói-se um gráfico em que representamos os pares ordenados $(x_i, y_i) = (C_{i,k}, C_{i,k+1})$ e uma recta que passe na origem com declive f_k .

Se os pares ordenados estiverem próximos à recta na origem, considera-se que os estimadores se ajustam. Caso contrário, devem procurar-se novos estimadores que se adequem melhor aos dados ou, até mesmo, rejeitar aplicar o modelo.

Este pressuposto implica a inexistência de correlação entre todos os factores de desenvolvimento. Para verificar esta condição recorre-se ao teste de *Spearman*.

Partindo da matriz de factores de desenvolvimento individuais, constroem-se duas novas matrizes.

- Matriz com os elementos $r_{i,k}$ - constrói-se esta matriz tendo em conta cada ano de desenvolvimento k , ordenando em cada coluna por ordem crescente os correspondentes factores individuais $C_{i,k+1}/C_{i,k}$, $0 \leq i \leq I - K - 1$, e designa-se por $r_{i,k}$ o número de ordem atribuído a cada factor individual, $1 \leq r_{i,k} \leq I - k$.
- Matriz com elementos $s_{i,k}$ - constrói-se de forma semelhante para com os factores precedentes $C_{i,k}/C_{i,k-1}$, desprezando o último factor $C_{i-k,k}/C_{i-k,k-1}$, e designam-se por $s_{i,k}$ os números de ordem obtidos desta forma, $1 \leq s_{i,k} \leq I - k$.

De seguida, o coeficiente de correlação de *Spearman*, T_k , é determinado por:

$$T_k = 1 - 6 \times \sum_{i=0}^{n-k-1} \frac{(r_{i,k} - s_{i,k})^2}{((n-k)^2 - n + k)}, 1 \leq k \leq n - 2, \text{ onde } -1 \leq T_k \leq 1 \quad (37)$$

Sendo T_k um valor pertencente ao intervalo $] -1; 1[$, obter um valor próximo de zero indica a inexistência de correlação dos factores de desenvolvimento entre os anos de $k-1$ e k e os factores entre k e $k+1$. Qualquer outro valor indica uma correlação positiva ou negativa.

No caso de não existir correlação, tem-se de:

$$E(T_k) = 0 \text{ e } V(T_k) = \frac{1}{n-k-1} \quad (38)$$

Segundo Mack (1993), a estimativa T obtém-se através de:

$$T = \frac{\sum_{k=1}^{n-2} (n-k-1) T_k}{\sum_{k=1}^{n-2} (n-k-1)} \quad (39)$$

No caso de não existir correlação, tem-se que:

$$E(T) = 0 \text{ e } V(T) = \frac{1}{\frac{(n-1) \cdot ((n-2))}{2}} \quad (40)$$

Tendo em conta que a distribuição de T_k , com $n-k \geq 10$, se aproxima de forma razoável de uma distribuição Normal, e sendo que T é obtido através da soma ponderada de T_{k1s} não correlacionados, pode assumir-se que a distribuição T também se aproxima da distribuição Normal.

Como tal, caso a estimativa T, obtida através do teste de *Spearman*, se situar dentro do intervalo de confiança de 50% de probabilidade³, aceita-se a hipótese de inexistência de correlação entre os factores individuais de desenvolvimento.

$$\frac{-0,6745}{\sqrt{\frac{(I-1)(I-2)}{2}}} \leq T \leq \frac{0,6745}{\sqrt{\frac{(I-1)(I-2)}{2}}} \quad (41)$$

Caso contrário, é necessário analisar as correlações e talvez optar pela não aplicação do método.

2.º Pressuposto

Para verificar a hipótese de independência entre os diferentes anos de ocorrência, considera-se novamente a matriz de factores de desenvolvimento individuais e subdividem-se todos os factores de desenvolvimento em dois conjuntos, um com os mais elevados e o outro com os mais baixos de cada ano k :

- Conjunto *L* (*Large factors*) – contém metade dos factores de desenvolvimento de cada ano k , nomeadamente os maiores;

³ Utiliza-se um nível de significância de 50 %, em vez dos habituais 95%, uma vez que o teste apresentado é apenas uma aproximação. Além disso, pretende-se detectar correlações existentes numa parte significativa do triângulo de desenvolvimento (Borginho, 2003).

- Conjunto S (*Small factors*) – contém os restantes factores de desenvolvimento, ou seja, os menores.

De seguida, analisa-se se existe uma superioridade de elementos de um dos grupos nas diagonais na matriz *Run-off*. Tendo em conta que o número de elementos pertencentes ao conjunto L tem de ser igual ao número de elementos pertencentes ao conjunto S , nos casos em que os anos de desenvolvimento contêm um n.º ímpar de elementos, despreza-se um elemento, nomeadamente o de valor mediano. Por fim, cada um dos factores de desenvolvimento individuais ou pertence a um dos conjuntos, ou foi eliminado.

No passo seguinte, para cada diagonal $C_j, 0 \leq j \leq n - 2$ (j corresponde à soma dos índices de linha e coluna) de factores de desenvolvimento, contabiliza-se o número de elementos do conjunto L e S . Em caso de que exista independência, espera-se poder encontrar o mesmo número de elementos nos dois conjuntos.

Sendo L_j o número de elementos de L que pertencem a C_j , e S_j o número de elementos de S que pertencem a C_j . Caso o valor de $Z_j = \min(S_j, L_j)$ seja significativamente inferior ao valor de $(L_j + S_j)/2$ existe, então, um predomínio de factores elevados ou reduzidos na diagonal j de factores de desenvolvimento. Assim sendo, o pressuposto deve ser rejeitado nesse caso.

Ambos os valores L_k e S_k seguem a distribuição Binomial de parâmetros $n = L_j + S_j$ e $p = 0,5$. Assim, assumindo uma aproximação à distribuição Normal, deve rejeitar-se a hipótese de independência entre os anos de ocorrência, quando a variável $Z = \sum Z_j$ não se situar dentro do próximo intervalo de confiança:

$$E(Z) - \Phi^{-1}[1 - \alpha/2] \times \sqrt{V(Z)}; E(Z) + \Phi^{-1}[1 - \alpha/2] \times \sqrt{V(Z)} \quad (42)$$

Em que $\Phi^{-1}[1 - \alpha/2]$ – representa o quantil de probabilidade $1 - \alpha/2$ da distribuição Normal Padrão.

Onde os dois primeiros momentos da variável aleatória Z_j são determinados por:

$$E(Z_j) = \frac{n}{2} - \binom{n-1}{m} \times \frac{n}{2^n} \quad e \quad V(Z_j) = \frac{n(n-1)}{4} - \binom{n-1}{m} \times \frac{n(n-1)}{2^n} + E(Z_j) - [E(Z_j)]^2 \quad (43)$$

Em que m representa o maior número inteiro menor ou igual a $(n-1)/2$. Os momentos da variável aleatória Z sob a hipótese de teste são dados por:

$$E(Z) = \sum E(Z_i) \text{ e } V(Z) = \sum V(Z_i)$$

(44)

3.º Pressuposto

O terceiro pressuposto implica que:

$$\frac{C_{i,k+1} - C_{i,k} \times \bar{f}_k}{\sqrt{C_{i,k}}} \approx \sigma_k$$

(45)

Portanto, os valores obtidos segundo essa expressão não deverão apresentar qualquer relação com $C_{i,k}$.

De forma a verificar que o terceiro pressuposto é válido na matriz de desenvolvimento é necessário garantir que os estimadores de desenvolvimento obtidos correspondem aos que apresentam menor variância.

Para tal, representa-se graficamente os pares ordenados $(\frac{C_{i,k+1} - C_{i,k} \times \bar{f}_k}{\sqrt{C_{i,k}}}, C_{i,k})$ com o objectivo de analisar se o conjunto de pontos obtidos apresenta ou não algum tipo de tendência.

Caso apresentem um comportamento aleatório, é possível assumir que se verifica o pressuposto. Caso contrário, é necessário obter outros estimadores \bar{f}_k ou rejeitar aplicar o modelo.

2.2.3. Método de *Bootstrap*

O método de *Bootstrap* é uma técnica de reamostragem introduzida por Efron em 1979, que permite estimar a variabilidade de um determinado estimador de um parâmetro.

De acordo com England e Verrall (2002), o método de *Bootstrap* é, na sua essência, uma técnica de reamostragem, com o objectivo de estimar a variabilidade de um determinado

estimador de um parâmetro. Pode-se encontrar esta metodologia a ser aplicada em diferentes áreas de estudo, pois tem provado ser bastante útil e versátil.

É considerado como uma alternativa aos tradicionais métodos estatísticos e também pode ser usado como complemento na análise dos resultados obtidos.

Os procedimentos utilizados na geração de amostras *Bootstrap* podem ser paramétricos ou não-paramétricos. A utilização de métodos paramétricos exige um conhecimento prévio da distribuição de probabilidade dos dados em estudo, com um ou mais parâmetros desconhecidos e, neste caso, a geração das amostras com reposição é feita com base nas estimativas desses parâmetros que, geralmente, são obtidas através da máxima verosimilhança.

Quando se aplica tendo em conta métodos não-paramétricos, o processo de reamostragem tem como hipótese um conjunto de variáveis aleatórias, independentes e identicamente distribuídas, com distribuição de probabilidade desconhecida.

O método de *Bootstrap* é aplicado a um modelo determinístico de estimação de provisões com a finalidade de determinar medidas de variabilidade para as estimativas obtidas.

Hinkley (1997) explica que este método pode apresentar falhas quando aplicado a determinadas situações, tais como:

- a) Conjunto de dados dependentes: o método de reamostragem não paramétrica tem como hipótese um conjunto de observações independentes e identicamente distribuídas. Logo, em caso de dependência, não existe uma forma óbvia de estimar a densidade conjunta das observações, dada uma realização do processo de reamostragem;
- b) Conjunto de dados incompletos (ou incorrectos): se na amostra de dados em estudo faltar informação ou existirem dados com incorreções/aproximações, na aplicação do método esses erros serão propagados e ampliados pelo processo de *Bootstrap*;
- c) A presença de *outliers*, isto é, valores extremos: é fundamental uma análise cuidada dos dados que servirão de base ao processo de reamostragem. É recomendável a remoção ou correcção destas observações, com a finalidade de diminuir a sua

influência nos resultados obtidos. Caso sejam incluídos, é possível aplicar esta metodologia a valores negativos, isto é, reembolsos sob o ponto de vista da Seguradora. Tendo o cuidado que este tipo de informação terá influência sobre os resultados.

2.2.3.1. Fases do Método de *Bootstrap*

- a) Primeira fase do procedimento consiste na preparação dos dados:
 1. Obter os factores de desenvolvimento através da matriz de *Run-off* acumulada de montantes de pagamentos, aplicando o método *Chain Ladder* clássico.
 2. Obter uma nova matriz *Run-off* de valores ajustados acumulados.

Sendo $D_{i,j}$ cada entrada da nova matriz, que corresponde aos pagamentos acumulados ajustados para o ano de ocorrência i de desenvolvimento j . Partindo da última diagonal da matriz *Run-off* acumulada, calculam-se os valores dos anos anteriores por forma recorrente, dividindo o valor do ano t pelo factor de desenvolvimento do ano $t-1$:

$$\widehat{D}_{i,j} = \frac{\text{Montante pago ajustado}(i,j+1)}{\text{Factor de desenvolvimento}(j+1)} \quad (46)$$

3. Obter a matriz *Run-off* de valores ajustados incrementais a partir da matriz *Run-off* de valores ajustados acumulados no passo anterior.
4. Determinar os resíduos que permitem aferir a qualidade do ajustamento e que, por sua vez, constituirão parte da base do processo de reamostragem. Através da matriz *Run-off* incremental dos valores ajustados obtêm-se os resíduos de *Pearson*, $(r_{i,j})$

utilizando: $r_{\text{pearson}(i,j)} = \frac{x_{i,j} - D_{i,j}}{\sqrt{\widehat{D}_{i,j}}}$ (47)

5. Ajustar os resíduos de *Pearson* obtidos no passo anterior.

Sugere-se que, antes da implementação do procedimento, os resíduos sejam ajustados tendo em conta o número de graus de liberdade do conjunto de dados observados. Ou seja, substituir r_{pearson} por $r'_{\text{pearson}} \text{ ajustados}$:

$$r_{\text{pearson ajustados}(i,j)} = \frac{x_{i,j} - D_{i,j}}{\sqrt{D_{i,j}}} \times \sqrt{\frac{n}{n-p}} = r_{i,j} \times \sqrt{\frac{n}{n-p}} \quad (48)$$

Onde n corresponde ao número de observações; e p é igual 2 vezes ao número de anos de ocorrência menos 1.

- b) A segunda fase do procedimento consiste na aplicação da técnica de reamostragem *Bootstrap*, realizando N simulações:
6. Redistribuir os resíduos de *Pearson* ajustados de forma aleatória e com reposição, criando uma nova matriz *Run-off* de pseudo-dados.

Os valores da nova matriz de pseudo-dados, $D_{i,j}^{bs}$ calculam-se através de:

$$D_{i,j}^{bs} = r_{i,j}^{bs} \times \sqrt{D_{i,j}} + D_{i,j} \quad (49)$$

7. Aplicar de novo o método de *Chain Ladder* clássico à matriz de pseudo-dados acumulados e obter os respectivos factores de desenvolvimento, para, por sua vez, obter uma estimativa da provisão denominada de pseudo-reserva.

Repetir este procedimento N vezes, de forma a obter várias estimativas da pseudo-reserva. Armazena-se o conjunto de resultados obtidos, com o objectivo de criar uma distribuição preditiva e assim poder construir intervalos de confiança, e através do desvio padrão obter uma estimativa do erro padrão das provisões que se pretender constituir.

Para estimar o desvio padrão de cada estimador \bar{R}_i e \hat{R} calcula-se o desvio padrão da amostra das estatísticas simuladas, respectivamente $\hat{\sigma}_{bs}(\hat{R}_i)$ e $\hat{\sigma}_{bs}(\hat{R})$.

$$\hat{\sigma}_{bs}(\hat{R}_i) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{R}_{i(b)}^{bs} - \hat{R}_{i(\cdot)}^{bs})^2}, 0 \leq i \leq 1 \quad (50)$$

e

$$\hat{\sigma}_{bs}(\hat{R}) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{R}_{(b)}^{bs} - \hat{R}_{(\cdot)}^{bs})^2} \quad (51)$$

em que $\hat{R}_{i(\cdot)}^{bs}$ e $\hat{R}_{(\cdot)}^{bs}$ representam a média das B observações de $\hat{R}_{i(b)}^{bs}$ e $\hat{R}_{(b)}^{bs}$, respectivamente.

Considerando a proposta de England e Verrall (2002) calcula-se o erro padrão através de:

$$\widehat{EP}_{bs}(\hat{R}_i) = \sqrt{\hat{\phi} \times \hat{R}_i + \frac{n}{n-p} \hat{\sigma}_{bs}(\hat{R}_i)}, 0 \leq i \leq 1 \quad (52)$$

e

$$\widehat{EP}_{bs}(\widehat{R}) = \sqrt{\widehat{\phi} \times \widehat{R} + \frac{n}{n-p} \widehat{\sigma}_{bs}(\widehat{R})} \quad (53)$$

Em que:

$$\widehat{\phi} = \frac{1}{n-p} \sum_{i=0}^l \sum_{j=0}^{l-i} r_{i,j}^2;$$

ϕ – representa o parâmetro de dispersão;

n – número de valores observados;

$p = 2n-1$, número de parâmetros a estimar.

2.2.4. Método de *Box-Jenkins*

Segundo Gujarati (2006), o objectivo de *Box-Jenkins*, é identificar e estimar o modelo estatístico que possa ser interpretado como tendo gerado os dados amostrais. Se esse modelo estimado será usado para a previsão, devemos supor que as características desse modelo são constantes no tempo e particularmente, no período futuro. Obviamente, a grande questão é: observando uma série temporal, como sabemos se ela segue um processo AR (e, nesse caso, qual o valor de p) ou um processo puramente MA (e, nesse caso, qual o valor de q) ou um processo ARMA (e, nesse caso, quais são os valores de p e q). A metodologia *Box-Jenkins* vem a responder à questão colocada propondo quatro etapas:

2.2.4.1. Identificação

Determinar os valores de p , d e q através do uso do correlograma (ACF) e do correlograma parcial (PACF).

Segundo Murteira e Black (1983), a etapa de identificação está mais interessada em reduzir o número de possíveis modelos a escolher (isto é, o melhor). Neste contexto, trata-se de identificação pela FAC e pela respectiva FACP estimadas. A série por sua vez compreende as seguintes sub-etapas:

- Estacionarização da série temporal;

- Selecção do modelo ARMA.

Em geral, as séries não se encontram estacionarizadas e os modelos ARMA correspondem a processos estacionários, nesta sub-etapa deve ser feitas transformações com o fim de torná-la estacionárias.

Estacionarizada a série passa-se à sub-etapa seguinte que é realizada através da comparação das funções autocorrelação parcial, que é respeitante à selecção do modelo, sendo momento crucial neste processo de modelação é condição indispensável para a correcta identificação do seu conhecimento profundo.

Existem três classes de processos estacionários o AR, MA e ARMA com características diferentes em termos de funções autocorrelação e autocorrelação parcial, características essas que são ilustradas na tabela 2 com mais pormenores dos cinco processos estacionários mais comuns, os sazonais e não sazonais.

Tabela 2. Principais características das FAC e PFAC e Teorias dos processos estacionários não sazonais.

| Processo | FAC | PFAC |
|----------|---|---|
| AR(1) | Queda exponencial: com valores positivos se $\phi_1 > 0$ e com alternância do sinal começando do lado negativo se $\phi_1 < 0$. | Valor de “pico” (ou alto) na primeira defasagem, e depois diminuí drasticamente para zero: pico positivo se $\phi_1 > 0$, e negativo se $\phi_1 < 0$. |
| AR(p) | Queda exponencial ou “damped sine-wave”. O comportamento exacto da função depende do sinal e magnitude de ϕ_1, \dots, ϕ_p | Picos desde a defasagem 1 a p, e depois diminui drasticamente ou corta-se para zero. |
| MA(1) | Pico na defasagem 1 e depois corta-se para zero: pico positivo se $\phi_1 < 0$ e | Queda exponencial: do lado negativo se $\phi_1 > 0$ e alternado de sinal começando |

| | | |
|-------|---|---|
| | negativo se $\phi_1 > 0$. | do lado positivo se $\phi_1 > 0$. |
| MA(q) | Picos desde a defasagem 1 a q, e depois corta-se para zero. | Queda exponencial ou “damped sine-wave”. O padrão exacto depende dos sinais e das magnitudes de ϕ_1, \dots, ϕ_p |

Fonte: Murteira e Black (1983).

Morettin (2004) disse que a partir de 1970 foram propostos vários procedimentos para a identificação de modelos ARMA, que consistem na ideia de escolher k e l que minimizam a quantidade:

$$P_{(k,l)} = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(T)}{T} \quad (54)$$

Onde $\hat{\sigma}_{k,l}^2$ é uma estimativa residual obtida ajustando um modelo ARMA (k, l), T observações da série e C (T) é uma função do tamanho da série.

A quantidade $(k+l) \frac{C(T)}{T}$, é termo penalizador, que tem uma relação de proporcionalidade directa com o numero de parâmetros, isto é, aumenta com o numero de parâmetros, enquanto que $\hat{\sigma}_{k,l}^2$ diminui. Sendo assim, para minimizar, a equação (54) corresponde a identificação de parâmetros k e l que equilibrem esse comportamento.

Os critérios empregados com o objectivo de escolher um único modelo, sugerem que o modelo definitivo seja aquele que minimiza a variância dos ruídos e tenha a menor quantidade de parâmetros. Os principais critérios de desempate que serão aqui discutidos são: BIC e AIC.

a) AIC (*Akaike Information Criterion*)

Segundo este critério, o melhor modelo para uma mesma série é o que apresenta menor valor AIC. Quanto mais parâmetros, maior são o valor AIC.

$$AIC = \log \hat{\sigma}_{\varepsilon}^2 + \frac{2.k}{T} \quad (55)$$

Onde: k que é o número de parâmetros estimados, T é o número de observações da série e $\hat{\varepsilon}$ o erro amostral.

b) BIC (*Bayesian Information Criterion*)

Este critério é uma mutação do método AIC. O melhor modelo é do menor valor de BIC, ou seja, aquele modelo que está explicado mais da variância do erro com menor número de variáveis. Este critério foi sugerido por Shibata (1976).

$$BIC = \log \hat{\sigma}_{\varepsilon}^2 + \frac{2 \cdot \log(T)}{T} \tag{56}$$

Onde: T é o número de observações da série e $\hat{\varepsilon}$ o erro amostral.

Lembrando que estes critérios servem como desempate apenas quando os MAPE dos modelos se igualam. Então se os MAPE não se igualam, então, o melhor modelo será o aquele com menor erro.

2.2.4.2. Estimação

Depois de identificar os valores apropriados de p , d e q o próximo passo é estimar os parâmetros dos termos auto-regressivo e de média móvel incluída no modelo. Por vezes esse cálculo pode ser feito com o método dos mínimos quadrados simples.

Segundo Gujarati (2006), a problemática da estimação dos modelos ARIMA tem sido abordada por diversos autores a partir dos trabalhos pioneiros de *Box-Jenkins*. A etapa que se segue à da identificação é a estimação dos parâmetros desconhecidos, $(A_1, \dots, A_p), (\theta_1, \dots, \theta_q), \sigma_u^2 = v\{u_t\}$,

$$\text{Do modelo } ARIMA(p, q), X_t - A_1 X_{t-1}^* - \dots - A_p X_{t-p}^* = u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} \tag{57}$$

Onde $X_t^* = X_t - \mu$ e X_t ($t = 1, 2, \dots, N$) são observações correspondentes à sucessão cronológica por hipótese já estacionarizada.

A estimação dos parâmetros nos modelos auto-regressivos é feita sob a hipótese dos resíduos serem ruído branco, isto é, $u_t \sim iidn(0, \sigma_u^2)$, o processo $AR(p)$, $\{X_t\}$, é Gaussiano } onde Supõe-se sem perda de generalidade que $E(u_t)=0$.

Para a modelagem SARIMA o procedimento é idêntico ao mostrado para a modelagem ARIMA, com a superioridade da estimação por máxima verosimilhança (Gujarati, 2006).

2.2.4.3. Diagnóstico

Depois de escolher um modelo ARIMA em particular, e estimar seus parâmetros, seguidamente escolhe-se o modelo que se ajusta razoavelmente bem aos dados, pois é possível que um outro modelo ARIMA possa desempenhar o mesmo papel. Um teste simples do modelo consiste em verificar se os resíduos estimados desse modelo são “White Noise” (ou estacionários); se forem, então podemos aceitar o ajuste específico; se não deve-se começar todo o processo de novo. Assim, a metodologia de Box-Jenkins é um processo iterativo.

Tendo em conta o princípio de parcimónia, devem-se analisar as estimativas dos parâmetros obtidos com objectivo de eliminar aqueles que sejam desnecessários, ou seja, os parâmetros que não se possam considerar significativamente diferentes de zero. A aplicação deste princípio tem surtido grandes efeitos. Pois, a prática tem demonstrado que, em geral, os modelos parcimoniosos produzem melhores previsões.

Estimados os parâmetros α e β , as quantidades, que são chamadas resíduos estimados. Se o modelo for adequado os resíduos estimados, deverão estar próximos dos resíduos e , deverão ser aproximadamente não correlacionados. Se representarmos por as autocorrelações, os resíduos estimados, então deve-se ter em conta.

O desvio padrão pode ser consideravelmente menor especialmente para pequenos valores de k (Morettin, 2004)). Este autor provou que para um modelo puramente aleatório, pode ser bem menor que $1/n$. Morettin (2004) mostrou ainda que para um modelo AR (1) os valores das autocorrelações devem estar contidos dentro do intervalo de confiança de 95% com limites de $\pm 2/n$, para que não se quebre o comportamento do ruído branco.

Teste de Box-Pierce ou teste “*Portmanteau*”, estes autores formularam um teste global para os m “lags” considerados, $\rho_k(\hat{u}_t)$ $k = 1, \dots, m$, ou seja, testa a seguinte hipótese:

$$H_0: \rho_1(\hat{u}_t) = \rho_2(\hat{u}_t) = \dots = \rho_m(\hat{u}_t) = 0$$

Onde em geral se toma $m = N/4$. A rejeição da hipótese sugere que o modelo estimado é apropriado para descrever a série em causa.

Ljung & Box (1978), recomendaram um teste melhor na estimação de modelos ARIMA (p,d,q) , baseado na estatística,

$$\hat{Q} = n(n-2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}^2(\hat{u}) \quad (58)$$

$n=N-d$, que converge mais rapidamente, para uma distribuição do Qui-quadrado com $m-p-q$ graus de liberdade (por isso torna-se mais conveniente quando o tamanho da amostra não é muito grande). A hipótese de ruído branco para os resíduos é rejeitada para valores grandes, que ele sugere as 15 ou 20 primeiras. Para Murteira e Black (1983), deve-se ensaiar o H_0 e concluir pela má qualidade do ajustamento se o valor observado de Q exceder o limite de significância estabelecida, e as hipóteses dos testes são formuladas de acordo com Granger et al. (1986), são as seguintes:

H_0 : O modelo é adequado; os resíduos do modelo estimados são puramente aleatórios,

H_1 : O modelo não é adequado; os resíduos do modelo estimado não são puramente aleatórios.

2.2.4.4. Previsão

Uma das razões para a popularidade do método de *Box-Jenkins* é o seu sucesso em fazer previsão (principalmente dos modelos ARIMA). Em muitos casos as previsões obtidas com esse método são mais confiáveis do que as obtidas com a modelagem econométrica tradicional, especialmente para previsões de curto prazo.

2.2.4.5. Critério para a Selecção de Modelos

Na análise de uma série temporal podem existir vários modelos que podem descrever de forma satisfatória, *i.e.*, é possível, por vezes encontrar mais de um modelo que passa descrever a série.

No entanto, existem critérios de escolha de modelos, que geralmente são baseados em estatísticas construídas a partir da sucessão residual correspondente a um determinado modelo ajustado ou baseados no erro de previsão produzido por esse modelo. Os principais critérios sugeridos, são os critérios de informação de *Akaike* (AIC) e de *Schwarz* (SIC).

$$AIC = \log \hat{\sigma}_{\varepsilon}^2 + \frac{2 \cdot k}{T} \quad (59)$$

$$BIC = \log \hat{\sigma}_{\varepsilon}^2 + \frac{2 \cdot \log(T)}{T} \quad (60)$$

À luz desses critérios está na escolha do modelo para o qual o correspondente AIC e SIC tenha o menor valor. Pela forma como o modelo é definido verifica-se que são penalizados os modelos com maior número de parâmetros e com maior variância residual, o que significa que o modelo mais parcimonioso só é seleccionado se os seus resíduos tiverem menor variância significativamente menor (melhor ajustamento) do que os concorrentes mais parcimoniosos.

CAPÍTULO III: METODOLOGIA

Neste capítulo são apresentados os aspectos metodológicos que orientaram todo o processo de investigação. A metodologia da pesquisa abrange vários *itens*, pois responde a diversas questões relacionadas com a pesquisa, levando em conta a classificação da pesquisa, método de procedimento, técnicas e instrumentos de recolha de dados, fonte e análise de dados.

3.1. Classificação da Pesquisa

3.1.1. Quanto a Natureza

Esta pesquisa é do tipo aplicada, pois busca um resultado bastante específico. A definição deste tipo de pesquisa por Vergara (1997) refere que a pesquisa aplicada é fundamentalmente motivada pela necessidade de resolver problemas concretos, mais imediatos ou não. Tem, portanto, finalidade prática.

A escolha deste tipo de pesquisa prende-se no facto desta oferecer as ferramentas necessárias para a consumação do principal objectivo deste trabalho que consiste em estimar o modelo ARIMA de *Box-Jenkins* adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 a 2028.

3.1.2. Quanto a Abordagem

Para a concretização do estudo, adoptou-se pela abordagem quantitativa. Para Prodanov e Freitas (2013), a pesquisa quantitativa considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las. Requer o uso de recursos e de técnicas estatísticas.

Os montantes dos custos com sinistros no seguro automóvel em Moçambique entre o período de 2010 a 2023 são uma variável quantitativa numérica e portanto sofreu uma análise quantitativa.

Esta abordagem é fundamental para o estudo a partir do entendimento de que no contexto moçambicano existem poucos estudos que discutem o sector de seguro, o uso da abordagem

quantitativa mostra-se ideal para os estudos que se preocupam na descrição de problemas que foram menos explorados.

3.1.3. Quanto aos Objectivos

A presente pesquisa considera-se exploratória, na medida em que procurou obter informações da situação dos custos com sinistros no seguro automóvel em Moçambique, sendo este obrigatório no país.

3.2. Método de Procedimento

3.2.1. Técnicas e Instrumento de Recolha de Dados

Quanto às técnicas de recolha de dados a pesquisa privilegiou dois métodos: a documentação directa, que corresponde ao levantamento de dados no próprio local onde os fenómenos ocorrem e à documentação indirecta que se encontra dividida em pesquisa documental (fontes primárias) e em pesquisa bibliográfica (fontes secundárias).

3.2.1.1. Pesquisa Bibliográfica

Segundo Gil (2002) a pesquisa bibliográfica é desenvolvida com base em material já elaborado, constituído por livros e artigos científicos. Assim, assume-se a pesquisa como sendo bibliográfica porque parte da consulta de livros, relatórios, artigos electrónicos com o intuito de aprofundar o conhecimento referente ao tema em estudo, acompanhando-o em todas etapas de elaboração do trabalho.

A pesquisa bibliográfica foi relevante neste estudo ao permitir mapear os vários estudos que discutem o sector de seguros. Esta informação foi fundamental para o enriquecimento do debate teórico de estudo, bem como na fundamentação da informação contida nos diferentes capítulos que compõe a dissertação.

3.2.1.2. Pesquisa Documental

De acordo com Lakatos e Marconi (2009) a pesquisa documental tem como principal enfoque a análise de documentos institucionais. Assim com esta técnica foi possível aceder os relatórios Trimestrais sobre a Actividade Seguradora em Moçambique do Instituto de Supervisão de Seguros de Moçambique e outros documentos normativos. Esta informação foi fundamental para a formulação do problema e na componente da análise e discussão de resultados contidos no capítulo IV.

3.3. Fonte de Dados

Os dados relativos aos custos com sinistros no seguro automóvel em Moçambique de 2010 à 2023 foram obtidos no Instituto de Supervisão de Seguros de Moçambique (Entidade reguladora do sector de seguros e fundos de pensões no país).

3.4. Análise de Dados

A análise de resultados será feita com os pacotes estatísticos EViews 12.0 e Office Excel 2016, para o processamento e análise de dados, com vista a elaboração dos respectivos gráficos, quadros e tabelas, que ajudaram na interpretação e análise dos resultados da pesquisa.

A análise de dados efectuada foi essencialmente econométrica e a mesma cingiu-se nos passos que são descritos de seguida:

Inicialmente foi efectuada a representação gráfica da série de custos com seguros no seguro automóvel em Moçambique de 2010 a 2023 originais em níveis, posteriormente a observação de algumas estatísticas descritivas da mesma série como a moda, mediana, desvio padrão, mínimos e máximos.

Fez-se a seguir a análise do histograma para testar a normalidade da série pelo teste de *Jarque-Bera* sendo que rejeita-se a hipótese nula de existência de normalidade quando o p-

valor for inferior ao nível de significância definido (usualmente convencionou-se o nível de 5%). No histograma verifica-se também os coeficientes de curtose e assimetria.

Uma vez que as séries temporais devem ser estacionárias para a sua análise foi efectuado o teste *Dickey-Fuller* que consiste em testar a hipótese nula de a série ter raiz unitária (ou seja não estacionária), comparando o p-valor com os valores críticos de *Dickey Fuller* a 1%, 5% e 10%, sendo que sempre que o mesmo for maior em termos absolutos não se rejeita a hipótese nula de a série ter raiz unitária.

CAPÍTULO IV: ANÁLISE E DISCUSSÃO DE RESULTADOS

Este capítulo é reservado à análise e discussão de resultados, dos dados agregados dos montantes pagos com sinistros no seguro automóvel pelas Seguradoras a operar no país no período de 2010 a 2023. A análise e a discussão de resultados constituem dois processos distintos mas estreitamente relacionados.

De ponto de vista prático o capítulo apresenta os seguintes pontos: (i) descrever a série de custos com sinistros no seguro automóvel e verificar o seu comportamento entre os anos de 2010 e de 2023; (ii) identificar e estimar o modelo ARIMA de *Box-Jenkins* que melhor se ajusta aos dados da série; (iii) utilizar o modelo estimado para fazer previsões dos custos com sinistros no seguro automóvel para o período de Janeiro de 2024 a Dezembro de 2028.

4.1. Descrição da Série de Custos com Sinistros no Seguro Automóvel e Verificação do seu Comportamento entre os Anos de 2010 a 2023

No gráfico 1, apresenta-se o histograma e as estatísticas descritivas dos custos com sinistros no seguro automóvel em Moçambique no período de Janeiro de 2010 a Dezembro de 2023.

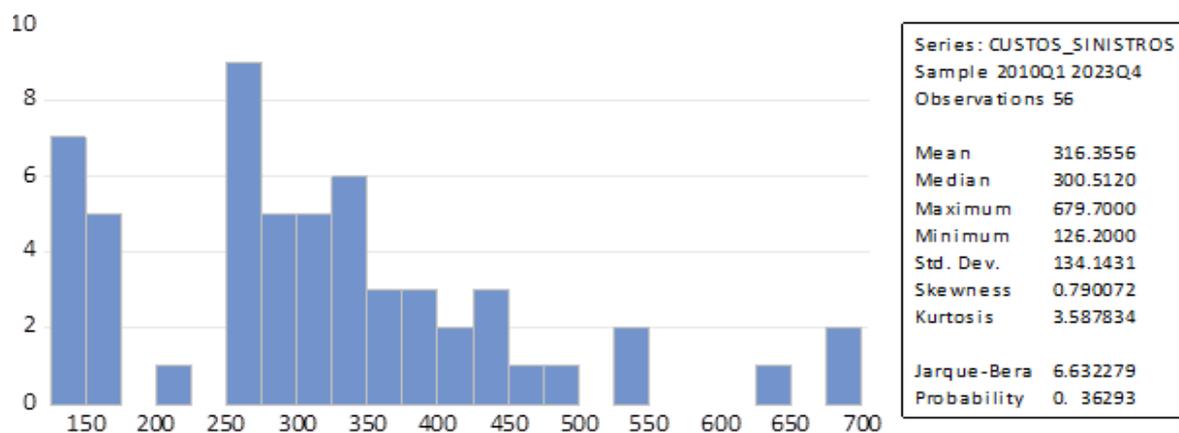
A média dos custos com sinistros no seguro automóvel observado é de 316,4 Milhões de Meticais, o custo mediano é de 300,5 Milhões de Meticais, o valor máximo dos custos observado no período em análise foi de cerca de 679,7 Milhões de Meticais, o mínimo de cerca de 126,2 Milhões de Meticais e o desvio padrão dos custos com sinistros é de 134,14 Milhões de Meticais.

O teste de normalidade de *Jarque-Bera* utiliza como parâmetros os coeficientes de curtose e assimetria (que na distribuição normal são de 3 e 0, respectivamente). Queremos saber se a nossa distribuição é aproximadamente normal porque desvios muito grandes como, por exemplo, uma curtose acima de 4 e assimetria acima de 1 invalidariam nossos erros-padrão e intervalos de confiança. Os valores da curtose são de 3,59 e a assimetria de 0,79.

O teste de *Jarque-Bera* tem como hipótese nula a normalidade. Assim, se o p-value for menor do que 5% ($p < 0,05$), então se rejeita a normalidade. Já se $p > 0,05$, não se rejeita a hipótese

de normalidade. O valor obtido de p-value é de 0,36, portanto superior a 0,05, logo não se rejeita a hipótese de normalidade dos dados.

Gráfico 1. Histograma e estatística descritiva dos custos com sinistros no seguro automóvel em Moçambique



Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Fazendo uma análise gráfica em relação a série em estudo em níveis (gráfico 2), podemos constatar que a série dos custos com sinistros no seguro automóvel em Moçambique, apresenta uma tendência crescente embora com flutuações ao longo de seu percurso, sugerindo nesse caso que durante o período em análise a média pode ter variado enquanto a variância não parece ser constante ao longo do tempo, dando-nos uma sensação de que esta série não é estacionária.

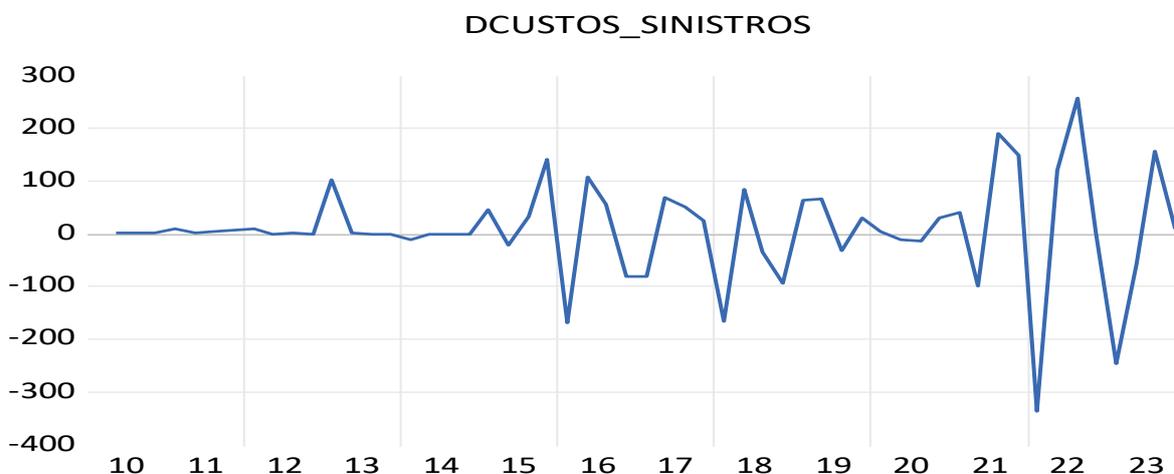
Gráfico 2. Evolução em níveis dos custos com sinistros no seguro automóvel (2010 - 2023)



Fonte: elaborado pelo Autor (extraído do EViews 12.0).

O gráfico 3 abaixo evidencia que a série dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças não apresenta nenhuma tendência, assemelhando-se ao comportamento de séries estacionárias, embora haja períodos em que apresentam grandes oscilações seguidas de períodos em que há relativa estabilidade, ela em primeiras diferenças oscila em torno da média zero, sugerindo-nos que é estacionária em primeiras diferenças.

Gráfico 3. Evolução em primeiras diferenças dos custos com sinistros no seguro automóvel (2010 - 2023)



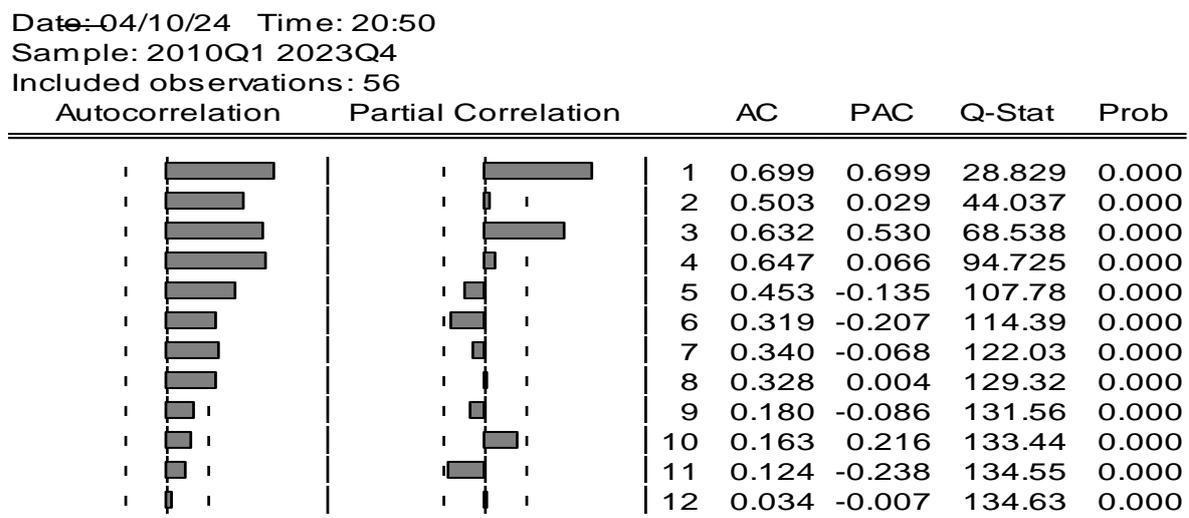
Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Uma forma simples de verificar se a série é estacionária é através do coeficiente de auto correlação da série. O correlograma da amostra representa um gráfico de valores da função em relação aos desfasamentos (Gujarati, 2006).

Uma indicação da estacionaridade da série é que quando os valores da função de autocorrelação estão próximos de zero a partir do primeiro desfasamento ou quando decaem rapidamente para valores próximos de zero.

Analisando o gráfico 4 (abaixo) do correlograma em níveis, através da função de autocorrelação (ACF) e dos intervalos de confiança, pode-se ter através de seu comportamento uma sinalização sobre a estacionaridade ou não da série. Como nesse caso a ACF apresenta um decréscimo suave em direcção a zero, tal sugere que a série não é estacionária e será necessário aplicar a primeira diferença.

Gráfico 4. Correlograma dos custos com sinistros no seguro automóvel em Moçambique em níveis

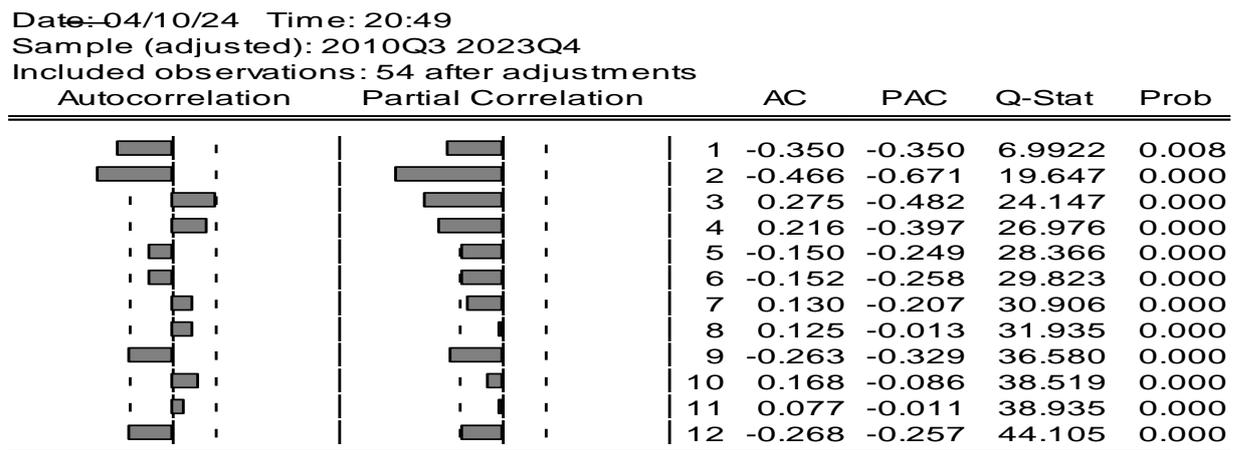


Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Olhando o comportamento do correlograma em primeiras diferenças (gráfico 5), tanto da ACF como da PACF podemos perspectivar tratar-se de uma série estacionária em primeiras diferenças.

A partir da leitura do correlograma, podemos ter uma indicação de que as séries são não estacionárias em níveis, mas sim em primeiras diferenças. Mas há a necessidade de se efectuar um teste formal de modo a verificar a estacionaridade da série em estudo.

Gráfico 5. Correlograma dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças



Fonte: elaborado pelo Autor (extraído do EViews 12.0).

O primeiro passo para o ajuste do modelo é a realização do teste de raiz unitária, o qual nos permitirá detectar ou não a presença de estacionaridade. Inicialmente foi feito o teste de raiz unitária em nível, com intercepto e também o teste em nível, com tendência e intercepto, conforme as tabelas a seguir:

Tabela 3. Teste de *Dickey-Fuller* com intercepto dos custos com sinistros no seguro automóvel em Moçambique em níveis

Null Hypothesis: CUSTOS_SINISTROS has a unit root
 Exogenous: Constant
 Lag Length: 2 (Automatic - based on SIC, maxlag=10)

| | t-Statistic | Prob.* |
|---|------------------|---------------|
| Augmented Dickey-Fuller test statistic | -1.254447 | 0.6440 |
| Test critical values: | | |
| 1% level | -3.560019 | |
| 5% level | -2.917650 | |
| 10% level | -2.596689 | |

*MacKinnon (1996) one-sided p-values.

Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Tabela 4. Teste de *Dickey-Fuller* com intercepto e tendência dos custos com sinistros no seguro automóvel em Moçambique em níveis

Null Hypothesis: CUSTOS_SINISTROS has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 2 (Automatic - based on SIC, maxlag=10)

| | t-Statistic | Prob.* |
|---|------------------|---------------|
| Augmented Dickey-Fuller test statistic | -2.646759 | 0.2623 |
| Test critical values: | | |
| 1% level | -4.140858 | |
| 5% level | -3.496960 | |
| 10% level | -3.177579 | |

*MacKinnon (1996) one-sided p-values.

Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Dados os resultados das tabelas acima, do teste de raiz unitária em nível com tendência e intercepto sendo o valor de *t-statistic* menor do que os valores críticos de *Dickey-Fuller* em termos absolutos a 1%, 5% e 10% de nível de significância, conclui-se pela presença de uma raiz unitária e, conseqüentemente, não estacionaridade da série dos custos com sinistros no seguro automóvel em Moçambique em níveis.

Pelo teste de *Dickey-Fuller* em níveis, verificou-se que as séries são não estacionárias, portanto, realizou-se o mesmo teste tendo as séries em primeiras diferenças. Os resultados obtidos são apresentados nas tabelas a seguir:

Tabela 5. Teste de *Dickey-Fuller* com intercepto dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças

Null Hypothesis: D(DCUSTOS_SINISTROS) has a unit root
 Exogenous: Constant
 Lag Length: 3 (Automatic - based on SIC, maxlag=10)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -9.105691 | 0.0000 |
| Test critical values: | | |
| 1% level | -3.568308 | |
| 5% level | -2.921175 | |
| 10% level | -2.598551 | |

*MacKinnon (1996) one-sided p-values.

Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Tabela 6. Teste de *Dickey Fuller* com intercepto e tendência dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças

Null Hypothesis: D(DCUSTOS_SINISTROS) has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 3 (Automatic - based on SIC, maxlag=10)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -9.042693 | 0.0000 |
| Test critical values: | | |
| 1% level | -4.152511 | |
| 5% level | -3.502373 | |
| 10% level | -3.180699 | |

*MacKinnon (1996) one-sided p-values.

Fonte: elaborado pelo Autor (extraído do EViews 12.0).

O resultado das tabelas acima, do teste de raiz unitária em primeiras diferenças com tendência e intercepto sendo o valor de *t-statistic* maior do que os valores críticos de *Dickey-Fuller* em termos absolutos até ao nível de 10% de significância, conclui-se pela estacionaridade da série dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças, ou seja, essa é a ordem que é preciso diferenciar a série para que ela se torne estacionária.

4.2. Identificação e Estimação do Modelo ARIMA de *Box-Jenkins* que Melhor se Ajusta aos Dados da Série

A estratégia de modelação de *Box-Jenkins* consiste em quatro etapas, sendo a primeira a de identificação de possíveis modelos que descrevem a série. Esta, com base na análise da função de auto correlação, nos possibilita seleccionar o modelo que melhor descreve a série dos custos com sinistros no seguro automóvel em Moçambique com base no resultado referente a estatísticas calculadas, escolhendo o modelo que apresenta menor valor nos critérios de informação.

Face aos resultados encontrados anteriormente, para determinar a ordem da parte auto regressiva (valor de p) utilizou-se o gráfico da função de auto correlação parcial, FACP, inicialmente, seleccionou-se o número de valores da FACP que estiveram acima do intervalo de confiança tendo um pico significativo. Para determinar a ordem da parte de média móvel (valor de q) utiliza-se o gráfico da função de auto correlação (FAC), assim como na parte auto regressiva, os números de valores da FAC que estiverem acima do seu intervalo de confiança representam o valor máximo que q poderá atingir.

Depois de identificados os valores de (p , d , q) o passo a seguir é de estimar os parâmetros de auto-regressão p e de média móvel, visto que a ordem de integração já foi identificada, pois conseguimos constatar que é necessário diferenciar uma vez a nossa série para que ela se torne estacionária.

A série apresentou *lags* significativos na função de autocorrelação parcial nas posições 1, 2, 3 e 4, bem como *lags* significativos na função de autocorrelação nos pontos 1 e 2 e foi integrado uma vez. Quando vários coeficientes FAC (*Autocorrelation*) e FACP (*Partial Correlation*)

são significativos é necessário testar vários modelos. Para tal foi necessário fazer várias combinações e considerar 5 modelos: ARIMA (0,1,2), ARIMA (0,1,3), ARIMA (0,1,4), ARIMA (3,1,0) e ARIMA (4,1,0). A estimação foi feita através de Mínimos Quadrados e o melhor modelo escolhido foi aquele que possui o menor critério *Akaike* e *Schwarz*:

Tabela 7. Selecção do modelo ARIMA

| Modelo ARIMA | Critério de <i>Akaike</i> | Critério de <i>Schwarz</i> |
|---------------------|----------------------------------|-----------------------------------|
| ARIMA (0,1,2) | 11.76389 | 11.87338 |
| ARIMA (0,1,3) | 12.01041 | 12.11990 |
| ARIMA (0,1,4) | 11.99919 | 12.10868 |
| ARIMA (3,1,0) | 12.03443 | 12.14392 |
| ARIMA (4,1,0) | 11.96942 | 12.07892 |

Fonte: elaborado pelo Autor.

Fazendo uma análise geral dos modelos estimados podemos constatar que o modelo ARIMA (0,1,2) é o mais adequado para os dados referentes aos custos com sinistros no seguro automóvel em Moçambique. Chegou-se a esta conclusão após comparados os resultados referentes as estatísticas calculadas, tendo sido constatado que são todas relativamente menores em relação aos outros modelos estimados.

Os resultados abaixo apresentam o modelo ARIMA (0,1,2), seleccionado com base nos critérios acima citados.

Gráfico 6. Correlograma Residual

Date: 04/10/24 Time: 21:49
 Sample (adjusted): 2010Q2 2023Q4
 Q-statistic probabilities adjusted for 2 ARMA terms

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.038 | -0.038 | 0.0821 | |
| | | 2 -0.007 | -0.008 | 0.0850 | |
| | | 3 0.114 | 0.113 | 0.8627 | 0.353 |
| | | 4 -0.005 | 0.003 | 0.8643 | 0.649 |
| | | 5 -0.034 | -0.033 | 0.9353 | 0.817 |
| | | 6 -0.083 | -0.100 | 1.3781 | 0.848 |
| | | 7 -0.070 | -0.079 | 1.7014 | 0.889 |
| | | 8 0.110 | 0.115 | 2.5088 | 0.867 |
| | | 9 -0.206 | -0.183 | 5.3944 | 0.612 |
| | | 10 0.059 | 0.069 | 5.6356 | 0.688 |
| | | 11 0.014 | -0.019 | 5.6504 | 0.774 |
| | | 12 -0.249 | -0.235 | 10.176 | 0.425 |

Fonte: elaborado pelo Autor (extraído do EViews 12.0).

Analisando o correlograma dos resíduos do modelo estimado (Gráfico 6), nota-se que não há valores significativos, tanto para a função de autocorrelação como para a função de autocorrelação parcial. O correlograma sugere que os resíduos do modelo são ruído branco. Estes resultados são corroborados pelos p-valores da última coluna que são todos superiores a 0,05 que nos permitem não rejeitar a hipótese de que os resíduos do modelo serem ruído branco.

4.3. Utilização do Modelo Estimado para Fazer Previsões dos Custos com Sinistros no Seguro Automóvel para o Período de Janeiro de 2024 a Dezembro de 2028

A tabela 8 (abaixo) corresponde a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de Janeiro de 2024 a Dezembro de 2028 (5 anos ou 20 trimestres).

A previsão de custos com sinistros no seguro automóvel no período de 2024 á 2028 demonstra um comportamento crescente, o que sugere que a procura pelo seguro automóvel e o pagamento dos benefícios pelas seguradoras serão constantes e esse comportamento pode ser explicado pelo contínuo aumento do parque automóvel no país.

Tabela 8. Previsão e intervalos de confiança dos custos com sinistros no seguro automóvel em Moçambique (2024:1 - 2028:12) – valores em milhões de meticais.

| | CUSTOS_SINF | LB | UB |
|--------|-------------|----------|----------|
| 2024Q1 | 523.3109 | 353.3623 | 693.2595 |
| 2024Q2 | 535.8542 | 293.4514 | 778.2571 |
| 2024Q3 | 543.1494 | 282.7929 | 803.5059 |
| 2024Q4 | 550.4446 | 272.7644 | 828.1247 |
| 2025Q1 | 557.7397 | 263.2547 | 852.2248 |
| 2025Q2 | 565.0349 | 254.1796 | 875.8901 |
| 2025Q3 | 572.3300 | 245.4739 | 899.1862 |
| 2025Q4 | 579.6252 | 237.0857 | 922.1647 |
| 2026Q1 | 586.9203 | 228.9734 | 944.8673 |
| 2026Q2 | 594.2155 | 221.1027 | 967.3283 |
| 2026Q3 | 601.5107 | 213.4453 | 989.5760 |
| 2026Q4 | 608.8058 | 205.9775 | 1011.634 |
| 2027Q1 | 616.1010 | 198.6792 | 1033.523 |
| 2027Q2 | 623.3961 | 191.5331 | 1055.259 |
| 2027Q3 | 630.6913 | 184.5245 | 1076.858 |
| 2027Q4 | 637.9864 | 177.6406 | 1098.332 |
| 2028Q1 | 645.2816 | 170.8702 | 1119.693 |
| 2028Q2 | 652.5768 | 164.2035 | 1140.950 |
| 2028Q3 | 659.8719 | 157.6318 | 1162.112 |
| 2028Q4 | 667.1671 | 151.1475 | 1183.187 |

Fonte: elaborado pelo Autor (extraído do Eviews 12.0).

Nota: CUSTOS_SINF – Valores previstos dos Custos com Sinistros; LB – Limite inferior da previsão e UB – Limite superior da previsão.

CAPÍTULO V: CONSIDERAÇÕES FINAIS

A pesquisa teve como objectivo estimar através da metodologia de Box-Jenkins um modelo ARIMA adequado para efectuar a previsão dos custos com sinistros no seguro automóvel em Moçambique para o período de 2024 à 2028. Com o desenvolvimento do trabalho, foi possível chegar as seguintes considerações finais:

Descrevendo o comportamento da série dos custos com sinistros no seguro automóvel em Moçambique, a análise gráfica da série em níveis, demonstrou que a série apresenta uma tendência crescente embora com flutuações ao longo de seu percurso, sugerindo nesse caso que durante o período em análise as médias podem ter variado, e a variância não ser constante ao longo do tempo, dando-nos uma sensação de que ela não é estacionária.

A série dos custos com sinistros no seguro automóvel em Moçambique em primeiras diferenças evidenciam que não apresenta nenhuma tendência, assemelhando-se ao comportamento de séries estacionárias, embora haja períodos em que apresentam grandes oscilações seguidas de períodos em que há relativa estabilidade, ela em primeiras diferenças oscilam em torno da média zero, sugerindo-nos que é estacionária em primeiras diferenças.

Testando a estacionaridade da série dos custos com sinistros no seguro automóvel em Moçambique, no correlograma em níveis a função de autocorrelação parcial, apresenta um decréscimo suave em direcção a zero, o que significa que não é estacionária, sendo que em primeiras diferenças o decréscimo foi rápido o que sugeriu que a mesma era estacionária.

Contudo, houve a necessidade de se realizar o teste formal (*Dickey Fuller* ou DF). Dos resultados do teste de raiz unitária concluiu-se que a série era estacionária em primeiras diferenças pois o valor de *t-statistic* era maior do que os valores críticos de *Dickey-Fuller* em termos absolutos até ao nível de 10% de significância.

Identificando e estimando o modelo ARIMA de *Box-Jenkins* que melhor se ajusta aos dados da série, na análise geral dos 4 modelos estimados podemos constatar que o modelo ARIMA (0,1,2) é o mais adequado para os dados referentes aos dos custos com sinistros no seguro automóvel em Moçambique.

Utilizando o modelo estimado, a previsão de custos com sinistros no seguro automóvel no período de 2024 á 2028 demonstra um comportamento crescente, o que sugere que a procura

pelo seguro automóvel e o pagamento dos benefícios pelas seguradoras serão constantes e esse comportamento pode ser explicado pelo contínuo aumento do parque automóvel no país.

REFERÊNCIAS BIBLIOGRÁFICAS

- Banze, L. F. (2007). *Análise da Contribuição da Actividade Seguradora no Crescimento Económico em Moçambique: Uma Exploração Empírica (1985 – 2005)*. Maputo: UEM.
- Borginho, H. M. M. (2001). *Metodologias Estocásticas de Estimação de Provisões para Sinistros: Relatório de Estágio Curricular*. Lisboa: Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- Borginho, H. M. M. (2003). *Provisões para Sinistros Não Vida - Metodologias de Estimação*. Lisboa: Departamento de Análise de Riscos e Solvência do ISP.
- Carvalho, A. (2010). *Modelos Estocásticos em Provisões para Sinistros*. Tese de Mestrado. Lisboa: Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.
- Decreto n° 3 / 1977. *Cria a Empresa Moçambicana de Seguros – EMOSE, E.E.* Maputo: Imprensa Nacional.
- Decreto n° 47 / 2005. *Regulamento da Lei n° 2 / 2003, de 21 de Janeiro, que introduz alterações ao Código de Estradas*. Maputo: Imprensa Nacional.
- England, P. D. e Verrall, R. J. (2002). *Stochastic Claims Reserving in General Insurance*. London: British Actuarial Journal.
- Gil, A. C. (2002). *Como Elaborar Projectos de Pesquisa*. 4ª Edição. São. Paulo: Atlas.
- Granger, C. W. J; Engle, R. F. and Newbold, P. (1986). *Forecasting Economic Time Series*. London: Academic Press.
- Guambe, C. J. (2022). *Manual de Teoria de Risco do Mestrado em Ciências Actuarias*. Maputo: UEM.
- Gujarati, D. N. (2006). *Econometria Básica*. 4ª Edição. Rio de Janeiro: Elsevier.
- Hinkley, D. (1997). *The Bootstrap Prediction Interval*. Cambridge: University Press.
- Instituto de Supervisão de Seguros de Moçambique – ISSM. (2021). *Relatório da Actividade Seguradora em Moçambique*. Maputo: ISSM.

Lakatos, E. M. & Marconi, M. A. (2009). *Metodologia de Trabalho Científico*. 7ª edição. São Paulo: Atlas.

Lei nº 24 / 91 de 31 de Dezembro. *Regula a Actividade Seguradora e Resseguradora em Moçambique*. Maputo: Imprensa Nacional.

Lei nº 2 / 2003 de 21 de Janeiro. *Aprova o Seguro Automóvel em Moçambique*. Maputo: Imprensa Nacional.

Ljung, G. M. e Box, G. E. P. (1978). *On a Measure of Lack of Fit in Time Series Models*. New York: Biometrika.

Mack, Th. (1993). *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates*. Cambridge: Astin.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. (1998). *Forecasting: Methods and Applications* (3rd ed.). New York: John Wiley.

Martins, J. V. (2011). *Notas Práticas Sobre o Contrato de Seguros*. Lisboa: Quid Juris.

Morretin, P. A. (2004). *Previsão de Séries Temporais*. São Paulo: Atual Editora.

Morretin, P. A. e Toloi, C. M. C. (2006). *Previsão de séries temporais*. 2ª Edição. São Paulo: Atual Editora.

Murteira, E. F. & Black, A. G. (1983). *Forecasting and Economic Series*. Cambridge: University Press.

Prodanov, C. C. e Freitas, E. C. (2013). *Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Académico*. 2ª Edição. Rio Grande do Sul: Universidade Feevale.

Shibata, R. (1976). *Selection of the Order of an Autoregressive Model by Akaike's Information Criterion*. New York: Biometrika.

Vasques, J. (2005). *Direito dos Seguros – Notas Para uma Teoria Geral*. Coimbra: Coimbra Editora.

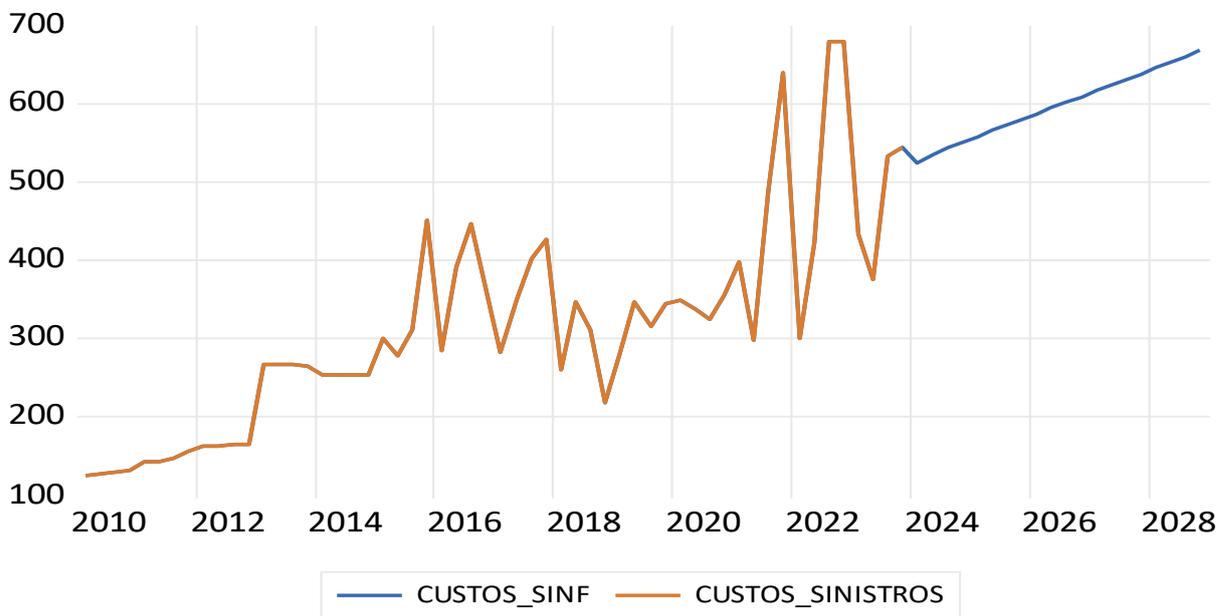
Vergara, S. C. (1997). *Projetos e Relatórios de Pesquisa em Administração*. São Paulo: Atlas.

Wheelwright, S. C. (1985). *Forecasting Methods for Management*. New York: John Wiley & Sons.

Zehnwirth, B. (1989). *The Chain Ladder Technique – a Stochastic Model*. New York: Kremer.

ANEXOS

Anexo 1: Valores Observados e Previstos dos Custos com Sinistros no Seguro Automóvel em Moçambique (2010:1 – 2028:12)



Nota: Laranja – Valores Observados; Azul – Valores Previstos.

Anexo 2: Modelos ARIMA estimados

Dependent Variable: D(CUSTOS_SINISTROS)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 08/12/24 Time: 10:21
 Sample: 2010Q2 2023Q4
 Included observations: 55
 Convergence achieved after 7 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 7.295157 | 6.066243 | 1.202582 | 0.2346 |
| MA(2) | -0.476735 | 0.107679 | -4.427391 | 0.0000 |
| SIGMASQ | 6684.517 | 1085.008 | 6.160802 | 0.0000 |
| R-squared | 0.272750 | Mean dependent var | | 7.585455 |
| Adjusted R-squared | 0.244779 | S.D. dependent var | | 96.75593 |
| S.E. of regression | 84.08425 | Akaike info criterion | | 11.76389 |
| Sum squared resid | 367648.4 | Schwarz criterion | | 11.87338 |
| Log likelihood | -320.5071 | Hannan-Quinn criter. | | 11.80623 |
| F-statistic | 9.751115 | Durbin-Watson stat | | 2.725358 |
| Prob(F-statistic) | 0.000253 | | | |
| Inverted MA Roots | .69 | -.69 | | |

Dependent Variable: D(CUSTOS_SINISTROS)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 08/12/24 Time: 10:24
 Sample: 2010Q2 2023Q4
 Included observations: 55
 Convergence achieved after 10 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 7.924991 | 19.38613 | 0.408797 | 0.6844 |
| MA(3) | 0.302964 | 0.144312 | 2.099375 | 0.0407 |
| SIGMASQ | 8588.592 | 1403.618 | 6.118897 | 0.0000 |
| R-squared | 0.065594 | Mean dependent var | | 7.585455 |
| Adjusted R-squared | 0.029655 | S.D. dependent var | | 96.75593 |
| S.E. of regression | 95.31048 | Akaike info criterion | | 12.01041 |
| Sum squared resid | 472372.6 | Schwarz criterion | | 12.11990 |
| Log likelihood | -327.2863 | Hannan-Quinn criter. | | 12.05275 |
| F-statistic | 1.825157 | Durbin-Watson stat | | 2.304079 |
| Prob(F-statistic) | 0.171368 | | | |
| Inverted MA Roots | .34+.58i | .34-.58i | -.67 | |

Dependent Variable: D(CUSTOS_SINISTROS)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 08/12/24 Time: 10:27
 Sample: 2010Q2 2023Q4
 Included observations: 55
 Convergence achieved after 8 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|-----------|
| C | 6.930358 | 16.72951 | 0.414260 | 0.6804 |
| MA(4) | 0.226625 | 0.098330 | 2.304748 | 0.0252 |
| SIGMASQ | 8504.829 | 1124.153 | 7.565543 | 0.0000 |
| R-squared | 0.074707 | Mean dependent var | | 7.585455 |
| Adjusted R-squared | 0.039119 | S.D. dependent var | | 96.75593 |
| S.E. of regression | 94.84457 | Akaike info criterion | | 11.99919 |
| Sum squared resid | 467765.6 | Schwarz criterion | | 12.10868 |
| Log likelihood | -326.9778 | Hannan-Quinn criter. | | 12.04153 |
| F-statistic | 2.099203 | Durbin-Watson stat | | 2.526002 |
| Prob(F-statistic) | 0.132820 | | | |
| Inverted MA Roots | .49+.49i | .49+.49i | -.49+.49i | -.49-.49i |

Dependent Variable: D(CUSTOS_SINISTROS)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 08/12/24 Time: 10:29
 Sample: 2010Q2 2023Q4
 Included observations: 55
 Convergence achieved after 8 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 7.891204 | 18.49546 | 0.426656 | 0.6714 |
| AR(3) | 0.194252 | 0.129488 | 1.500155 | 0.1396 |
| SIGMASQ | 8825.199 | 1309.112 | 6.741363 | 0.0000 |
| R-squared | 0.039852 | Mean dependent var | | 7.585455 |
| Adjusted R-squared | 0.002923 | S.D. dependent var | | 96.75593 |
| S.E. of regression | 96.61442 | Akaike info criterion | | 12.03443 |
| Sum squared resid | 485386.0 | Schwarz criterion | | 12.14392 |
| Log likelihood | -327.9469 | Hannan-Quinn criter. | | 12.07677 |
| F-statistic | 1.079152 | Durbin-Watson stat | | 2.331993 |
| Prob(F-statistic) | 0.347372 | | | |
| Inverted AR Roots | .58 | -.29-.50i | -.29+.50i | |

Anexo 3: Custos com Sinistros no Seguro Automóvel em Moçambique (2010:1 – 2023:12)

| Trimestre | Custo Com Sinistros (Milhões de Meticais) |
|------------------|--|
| 2010 – 1 | 126.2 |
| 2010 – 2 | 128.4 |
| 2010 – 3 | 129.3 |
| 2010 – 4 | 132.1 |
| 2011 – 1 | 142.6 |
| 2011 – 2 | 143.4 |
| 2011 – 3 | 147.8 |
| 2011 – 4 | 155.6 |
| 2012 – 1 | 164.2 |
| 2012 – 2 | 164 |
| 2012 – 3 | 165 |
| 2012 – 4 | 165.1 |
| 2013 – 1 | 266.5 |
| 2013 – 2 | 267 |
| 2013 – 3 | 266.5 |
| 2013 – 4 | 266 |
| 2014 – 1 | 254 |
| 2014 – 2 | 254.1 |
| 2014 – 3 | 254.2 |
| 2014 – 4 | 254.3 |
| 2015 – 1 | 300.524 |
| 2015 – 2 | 278.554 |
| 2015 – 3 | 310.705 |
| 2015 – 4 | 451.639 |
| 2016 – 1 | 283.994 |
| 2016 – 2 | 391 |
| 2016 – 3 | 445.7 |
| 2016 – 4 | 363.6 |
| 2017 – 1 | 282.8 |
| 2017 – 2 | 350.9 |
| 2017 – 3 | 401.6 |
| 2017 – 4 | 427.1 |
| 2018 – 1 | 260.3 |

| | |
|----------|-------|
| 2018 – 2 | 345.9 |
| 2018 – 3 | 312.1 |
| 2018 – 4 | 217.4 |
| 2019 – 1 | 280.3 |
| 2019 – 2 | 346.9 |
| 2019 – 3 | 315.6 |
| 2019 – 4 | 345.7 |
| 2020 – 1 | 349.5 |
| 2020 – 2 | 338.9 |
| 2020 – 3 | 325.8 |
| 2020 – 4 | 356.4 |
| 2021 – 1 | 397.8 |
| 2021 – 2 | 298.3 |
| 2021 – 3 | 488.4 |
| 2021 – 4 | 638.4 |
| 2022 – 1 | 300.5 |
| 2022 – 2 | 421.6 |
| 2022 – 3 | 677.9 |
| 2022 – 4 | 679.7 |
| 2023 – 1 | 432.2 |
| 2023 – 2 | 375.7 |
| 2023 – 3 | 532.8 |
| 2023 – 4 | 543.4 |

Anexo 4: Credencial



Faculdade de Economia

Exmos Senhores
Instituto Nacional de Supervisão de Seguros de Moçambique

MAPUTO

CREDECIAL

O Sr. Joaquim Caetano Matias é estudante do Curso de Mestrado Em Ciências Actuarias .

No âmbito do plano de estudos ora em curso na nossa Faculdade, este estudante é chamada nesta fase a realizar trabalho de pesquisa relacionado com a formação cujo tema se designa: "*Aplicação da Metodologia de BOX-JENKINS na Previsão dos Custos com Sinistros no Seguro Automóvel em Moçambique*" Junto da Instituição Supracitado.

Neste contexto a Faculdade de Economia aprecia o apoio de V. Excia. na disponibilização de toda informação relevante e possível para o sucesso do trabalho da estudante acima referido.

Com os nossos melhores cumprimentos.

Maputo, 08 de Setembro de 2023


Doutor Fernando Linduacha
(Assistente Universitário)
Faculdade de Economia



08.09.2023

sch