

Previsão da sinistralidade em seguros de vida utilizando modelos de séries temporais

Claims forecast in life insurance using time series models

Recebimento: 29/04/2023 - Aceite: 03/11/23 - Publicação: 01/04/2024

Processo de Avaliação: Double Blind Review- DOI: <https://doi.org/10.22567/rep.v13i1.961>

Amanda Santos Pandolfi

amandasantospandolfi@gmail.com

<https://orcid.org/0009-0002-5252-8045>

Universidade Federal de Minas Gerais (UFMG), Brasil

Jussiane Nader Gonçalves

jusianegoncalves@gmail.com

<https://orcid.org/0009-0009-1285-3928>

Universidade Federal de Minas Gerais (UFMG), Brasil

RESUMO

A condição de não estaticidade dos seguros em geral se dá por se tratar de uma ciência que depende de fatores sociais e econômicos para se sustentar. O Seguro de Vida, em especial, é impactado por diversos agentes, desde políticas econômicas e culturais, até o desempenho das seguradoras quanto a oferta e demanda do seu produto e aspectos sociodemográficos que influenciam no comportamento do próprio segurado. O conhecimento sobre o risco identificado é fundamental para que a avaliação seja realista e precisa. Para garantir a solvência dessas empresas e assegurar os compromissos futuros, diversas premissas atuariais estão envolvidas no gerenciamento e precificação de riscos, como análise de frequência de sinistros, severidade, risco biométrico, análises de sensibilidade e subscrição, acompanhamento de sinistralidade, entre outras, cabendo ao atuário responsável adequar tais medidas a realidade da instituição de risco. Portanto, este trabalho tem como objetivo avaliar modelos de previsão para uma das premissas atuariais mencionadas, a saber, a sinistralidade, por meio de métodos de regressão e Box-Jenkins. Foram analisados os dados mensais disponíveis no Sistema de Estatísticas (SES) da Superintendência de Seguros Privados (SUSEP), entre os anos de 2011 e 2021, das oito seguradoras com maior volume de prêmio. A capacidade de previsão dos modelos foi avaliada comparando os valores previstos com os dados reais do 1º semestre de 2022. O modelo de

regressão foi o mais adequado para a previsão da sinistralidade, embora ambos os métodos tenham apresentado desempenho satisfatório.

Palavras-chave: seguros de vida, sinistralidade, modelos de séries temporais.

ABSTRACT

The non-static nature of insurance, in general, is because it depends on many social and economic factors to sustain itself. Individual Life Insurance, in particular, is impacted by various factors that include economic and cultural policies, the performance of insurers in terms of supply and demand, and sociodemographic aspects that influence the behavior of the life insurance consumer. Analyzing each factor may facilitate the acceptance and pricing of insurance. Knowledge of the identified risk is essential for the evaluation to be realistic and accurate. To ensure the solvency of these companies and, thus, for them to fulfill their future commitments, the management and risk pricing involves several actuarial assumptions, such as analysis of claim frequency, severity, biometric risk, sensitivity and subscription analysis, and monitoring of claims experience. Consequently, it is up to the responsible actuary to manage such measures to the reality of the risk institution. Therefore, this paper aims to evaluate prediction models for one of the aforementioned actuarial assumptions, namely, loss ratio, through regression and Box-Jenkins models. This work analyzed monthly data available in the Statistics System (SES) of the Superintendence of Private Insurance (SUSEP), from the eight insurers with the highest premium volume, between 2011 and 2021. To perform model adequacy, we have compared the predicted values with the current data from the first half of 2022. Both regression and Box-Jenkins models showed satisfactory performance, but the regression model was more appropriate for predicting the loss ratio of the claims experience.

Keywords: life insurance, loss ratio, time series models.

1. INTRODUÇÃO

O estudo sobre seguros não é estático por se tratar de uma ciência que depende de muitos fatores sociais e econômicos para se sustentar. Segundo a Escola Nacional de Seguros (ENS,

antiga FUNENSEG), o Seguro de Vida Individual, em especial, é impactado por diversos agentes, como políticas econômicas e culturais em uma conjuntura macroeconômica, o desempenho das seguradoras quanto à oferta e demanda do seu produto e aspectos sociodemográficos que influenciam no comportamento do próprio consumidor do seguro de vida, por exemplo, no cenário microeconômico.

Conforme Huebner e Black (1976), existem certos fatores como idade, sexo, constituição física, estado físico, histórico pessoal, herança genética, risco moral, prática de atividades de risco, profissão, ocupação, entre outros, devem ser analisados pela seguradora para aceitação do risco. A análise de cada fator poderá facilitar a aceitação e a precificação do seguro. O conhecimento sobre o risco identificado é fundamental para que a avaliação seja realista e precisa. Tais fundamentos são estritamente importantes porque estão diretamente ligados ao volume de prêmio necessário pela instituição de risco para arcar com os possíveis montantes de sinistros futuros advindos dessas subscrições.

Os seguros de pessoas são, em geral, produtos de média ou longa duração, e sua gestão e precificação envolvem diversas premissas atuariais, tais como análise de frequência e severidade de sinistros, risco biométrico, análises de sensibilidade e subscrição, e acompanhamento de sinistralidade. É papel do atuário responsável adequar essas medidas à realidade da seguradora. Dessa forma, a avaliação da sinistralidade é um importante indicador nesse ramo, já que representa a relação entre o montante total de sinistros retidos e o volume de prêmios ganhos em determinado intervalo de tempo. Em termos gerais, a sinistralidade indica a representatividade dos custos em relação à receita de prêmios.

Logo, a análise da sinistralidade é um benchmark fundamental para gerenciar a aceitação do risco e avaliar a adequação da precificação do seguro. Além disso, é um indicador prudencial que ajuda a mensurar o nível de risco da instituição e identificar possíveis falhas no processo de subscrição de seguros. Com base nessa avaliação, a seguradora pode tomar medidas para minimizar prejuízos, como propor novos métodos de tarifação para ajustar os prêmios. Nesse contexto, o objetivo geral deste trabalho é avaliar modelos estatísticos para a previsão das taxas de sinistralidade. De modo específico, será realizado um estudo para o Seguro de Vida Individual por Cobertura de Morte, ramo 1391, definido pela SUSEP.

Uma vez que o objetivo deste trabalho consiste em prever o desenvolvimento futuro da sinistralidade através de uma investigação do seu comportamento no passado, bem como busca-se avaliar métodos estocásticos que permitem a quantificação da incerteza acerca dessa variável, propõe-se a utilização de modelos para séries temporais para tal fim.

Uma série temporal é formada por valores de variável avaliada em pontos distintos do tempo, e a suposição que norteia sua análise é de que há um sistema causal, aproximadamente constante, relacionado com o tempo, que exerceu influência sobre os dados no passado e pode continuar a fazê-lo no futuro (Chaves et al., 2014). Esse sistema costuma atuar criando padrões que podem ser detectados em análises estatísticas. Segundo Morettin e Toloí (2006), os principais objetivos para analisar uma série temporal são: i) investigar o mecanismo gerador da série temporal; ii) fazer previsões dos valores futuros da série; iii) descrever o comportamento da série, tais como sazonalidade e tendência; e iv) procurar periodicidades relevantes nestes dados.

Conforme explicam Gonçalves e Barreto-Souza (2020), a avaliação da adequação do modelo proposto é uma tarefa importante, uma vez que permite verificar desvios da distribuição de resposta assumida, bem como detectar especificações incorretas do modelo e ajuste inadequado. Adicionalmente, Machado (2012) discorre que o melhor critério para escolher um modelo de previsão é a sua capacidade preditiva, ou seja, quão perto estão as previsões dos valores posteriormente observados.

Diante dos fatos expostos, este trabalho tem como objetivos específicos: i) aplicar métodos estatísticos de previsão em dados de seguradoras que atuam no ramo 1391, disponibilizados no SES (Sistema de Estatísticas da SUSEP), a fim de estimar as sinistralidades para o 1º semestre de 2022; ii) avaliar o ajuste e a aplicabilidade dos métodos de previsão da série, comparando-os com os dados reais disponíveis no sistema da SUSEP. Os resultados dessa análise indicam que os modelos de regressão e de Box-Jenkins apresentaram um desempenho satisfatório na previsão da sinistralidade de seguros de vida, sendo que o modelo de regressão se mostrou mais adequado para esse propósito.

2. REFERENCIAL TEÓRICO

2.1. Contexto Histórico

A preocupação em resguardar e proteger a própria vida faz parte do instinto humano e está presente desde os primórdios da humanidade. Contudo, somente em meados da Revolução Industrial que foram estruturados os primeiros seguros de pessoas em moldes semelhantes aos que conhecemos hoje. Foram motivados pelas condições precárias de vida e trabalho, em que os trabalhadores da época eram obrigados a aceitar uma carga de trabalho excessivamente

elevada, e era perigoso, pois não havia nada que os resguardassem. Além disso, os inválidos da época eram apenas demitidos e substituídos, sendo tais casos ilustrações da importância de um seguro contra invalidez. Ainda no século XIX, Otto Von Bismarck montou o primeiro modelo de seguro social, que abrangia doença, invalidez e acidentes de trabalho, em resposta às greves e pressões dos trabalhadores da época.

No Brasil, o primeiro seguro registrado foi o marítimo. Até então, o “Código Comercial Brasileiro” (Lei N° 556, de 25 de junho de 1850), proibia o seguro de vida e permaneceu assim até 1855, quando a sua exploração foi autorizada. Segundo Lima (2018), dado grande interesse do mercado internacional no fluxo de seguros de vida no país, foi promulgada, em 5 de setembro de 1895, a Lei N° 294, dispondo exclusivamente sobre as companhias estrangeiras de seguros de vida, determinando que suas reservas técnicas fossem constituídas e tivessem seus recursos aplicados no Brasil, para fazer frente aos riscos aqui assumidos. Com a evolução do ramo segurador no país, houve a necessidade da criação de um órgão que fiscalizasse as operações desse mercado, de modo a garantir não apenas a livre concorrência, mas também a estabilidade e respeito aos segurados. Dessa forma, a Superintendência de Seguros Privados, SUSEP, criada em 1966, é responsável pela autorização, controle e fiscalização dos mercados de seguros, previdência complementar aberta, capitalização e resseguros no Brasil. Há também o órgão máximo, que estabelece as grandes diretrizes do setor de seguros, o Conselho Nacional de Seguros Privados (CNSP), cujo presidente é o Ministro da Fazenda. O CNSP é também o órgão que decide as pendências no campo de seguros em última instância, regulamenta os seguros obrigatórios e estabelece os limites das operações de seguros no país.

2.2. O Seguro Vida Individual

No Brasil, esse tipo de seguro pode ser dividido por cobertura de sobrevivência ou de risco - também chamado de cobertura por morte, e cada tipo possui suas próprias legislações específicas. A primeira, regida pela Resolução CNSP N° 348/2017, é estruturada sob o regime financeiro de capitalização e é definida como a cobertura que garante o pagamento do capital segurado, pela sobrevivência do segurado ao período de diferimento contratado, ou pela compra, mediante pagamento único, de renda imediata. Possui diversos tipos de planos e modalidades.

O presente trabalho analisará o Seguro de Pessoas por Cobertura de Risco, o qual é definido pela Resolução CNSP N° 117/2004 como cobertura do seguro de pessoas cujo evento

gerador não seja a sobrevivência do segurado a uma data pré-determinada. Em revés a sobrevivência, essa cobertura admite os regimes financeiros de Repartição Simples (RS), de Capitalização (RC) e de Capitais de Cobertura (RCC), e deve ser estruturada apenas na modalidade de benefício definido. Para mais, apesar de existir as formas de contratação individual ou coletiva, este relatório explora apenas o Seguro de Pessoas Individual.

Ademais, como exemplo de seguros de pessoas com coberturas de risco, temos seguro de vida, seguro funeral, seguro de acidentes pessoais, seguro educacional, seguro viagem, seguro prestamista, seguro de diária por internação hospitalar e seguro perda de renda, seguro de diária de incapacidade temporária. Este trabalho analisará em especial o Seguro Vida, indicado pela SUSEP como Ramo 1391. A depender do plano escolhido, há um leque de coberturas oferecidas pelas seguradoras que podem ser vendidas juntas ou separadamente. Segundo a Circular SUSEP N° 395/2009, que estabelece a codificação dos ramos de seguro, as contabilizadas no Ramo de Pessoas Individual são Morte por Qualquer Causa (MQC), Invalidez Ocasionada por Doença e Invalidez por Qualquer Causa (doença ou acidente), conforme o caso.

Segundo dados do SES, o Seguro de Pessoas segue crescendo expressivamente ao longo dos anos no Brasil e atualmente representa 16% das receitas acumuladas por segmento no país. Em especial, o Vida Individual apresentou aumento de aproximadamente 215% no volume de prêmios em 4,5 anos (jan/2017 a jun/2022), os quais estão na casa dos bilhões. Em relação aos sinistros, os valores absolutos são bem menores - estão nos milhões, como é o esperado, e o aumento não é proporcional aos prêmios. Nos mesmos 4,5 anos, esse crescimento foi de aproximadamente 58%.

A partir desses dados se calcula o índice de sinistralidade, observando a razão entre os sinistros pelos prêmios. Essa taxa é essencial para as seguradoras porque, por meio dela, conseguem analisar se o grau de aceitação de risco e a precificação estão adequadas. De acordo com os dados da SUSEP, pode-se observar que a sinistralidade não oscila muito, desconsiderando o aumento exorbitante durante a pandemia do Covid-19. Fora esse período, essa taxa variava entre 20% e 30%. No seu máximo, no ano de 2021, atingiu aproximadamente 44%, valor muito alto, mas condizente com o cenário.

2.3. Trabalhos Relacionados

Em um estudo proposto por Vianna (2018), foi analisado a sinistralidade no contexto da saúde, aplicado a operadoras de planos privados de assistência à saúde médico-hospitalar, e

objetivava a determinação dos fatores associados a esse índice e seus possíveis efeitos. Através desse trabalho, a autora concluiu que a sinistralidade se altera de acordo com a mudança de alguns fatores das Operadoras de Planos de Saúde (OPS), como a alteração do número de beneficiários, por exemplo. Além disso, apresentou medidas que podem ser tomadas visando a redução da sinistralidade, como a adoção de medidas preventivas com respeito à situação econômico-financeira, mantendo recursos próprios que auxiliam na capacidade da OPS em honrar com compromissos futuros e a importância da disponibilidade financeira, que auxilia a OPS a honrar com os compromissos a curto prazo.

Ainda no contexto de seguro saúde, Guimarães e Alves (2009) desenvolveram um modelo específico de previsão de insolvência para as OPS, a fim de antecipar a capacidade financeira dessas instituições em cumprir suas obrigações contratuais com clientes e prestadores de serviços de saúde. Para isso, os autores ajustaram um modelo de regressão logística a partir de 17 indicadores financeiros, levando em consideração cerca de 600 operadoras brasileiras de planos de saúde. Os autores concluíram que, ao comparar o modelo proposto com um modelo mais geral, o desempenho do modelo proposto foi superior em termos de precisão.

O trabalho de Carvalho e Gonçalves (2022) realiza a modelagem do Índice de Desempenho da Saúde Suplementar (IDSS), avaliando o efeito de fatores como região, quantidade de notificações de intermediação preliminar, modalidade e porte das OPS na previsão desse índice, o qual é composto de dimensões de qualidade em atenção à saúde, garantia de acesso, sustentabilidade no mercado e gestão de processos e regulação. As autoras realizaram uma validação e seleção dos modelos propostos, por meio da avaliação de medidas de adequação, chegando à conclusão de que o modelo de regressão beta apresentou o melhor ajuste para a previsão do IDSS.

Em seu trabalho de análise descritiva, Correa (2022) discorre sobre os impactos causados da pandemia de Covid-19, em particular o aumento da sinistralidade, sobre as seguradoras que atuam no ramo vida. Por meio dessa análise exploratória, a autora conclui que as seguradoras foram capazes de administrar e absorver a elevação desse índice de forma satisfatória. Como resultado geral, esse reflexo é observado em outros índices operacionais que foram avaliados para medir a situação das seguradoras. Já no contexto de séries temporais, Mori e Gonçalves (2016) analisam dados do sistema de saúde do Brasil e dos Estados Unidos, no período entre 2000 e 2012, utilizando modelos estatísticos para avaliação de séries temporais. Os modelos de séries temporais foram ajustados para a previsão do número de pessoas que possuem plano de saúde privado no Brasil, enquanto um modelo de regressão foi ajustado para

a estimação de novas adesões aos 14 planos de saúde avaliados dos Estados Unidos. Esses modelos foram desenvolvidos para auxiliar na tomada de decisões futuras, permitindo compreender a tendência dessas variáveis e orientar a gestão de forma mais efetiva.

No contexto do mercado de seguros brasileiro, há poucos estudos que aplicam modelos de séries temporais para avaliar índices que permitem uma análise econômico-financeira e atuarial das instituições de risco, especialmente para previsões de médio e longo prazo. Assim, esta pesquisa visa contribuir para a literatura sobre o tema.

3. DADOS E MÉTODOS

3.1. Dados

O mercado de seguros de vida individual não é bem distribuído, gerando uma falta de proporcionalidade no volume de prêmios entre as seguradoras que operam com esse tipo de serviço. Dentre as 43 entidades que operam com esse seguro disponíveis no SES, as duas maiores concentram aproximadamente 68,24% de todo o montante de prêmios. Para este trabalho, foram selecionadas empresas com receita de prêmios maior ou igual a 0,5 bilhão, uma vez que esse ponto de corte já garantiu a seleção das seguradoras que concentram mais de 90% do volume de prêmios. Dessa forma, por esse filtro passaram 8 empresas, as quais tiveram seus nomes omitidos para preservar suas imagens. Tais informações estão disponíveis no próprio painel de inteligência do mercado de seguros da SUSEP.

Além disso, as variáveis selecionadas para compor o banco foram Competência, Sinistros Retido e Ocorrido e Prêmio Ganho. Para as previsões, foram utilizadas a Competência e a Sinistralidade, sendo essa última calculada a partir do Prêmio Ganho e dos Sinistros. De acordo com a SUSEP, define-se:

- Prêmio Ganho: Prêmio Emitido + Variação das Provisões Técnicas.
- Montante de Sinistros: Sinistros Avisados + Despesas + Parcela de Cosseguro Aceito – Parcela de Cosseguro Cedido – Salvados e Ressarcimentos + Serviços de Assistência + Variação das Provisões Técnicas.

Ademais, foram considerados todos os meses de competências que possuíam dados completos disponíveis no SES. Assim, considerou-se a série a partir de janeiro de 2011, o que representa 132 observações mensais ou 11 anos completos, de 2011 a 2021. Para a avaliação

de adequação dos métodos propostos, foram coletadas as informações do 1º semestre de 2022 para fins de comparação com a previsão dos modelos.

3.2. Método de Previsão por Modelo de Regressão

Em um processo de regressão linear simples, supõe-se que $f(\cdot)$ qualquer pode ser aproximada por uma reta, tal qual

$$f(x) = \beta_0 + \beta_1 x + \varepsilon_i, \quad (1)$$

em que β_0 é o intercepto, β_1 é a inclinação da reta e ε_i é o termo de erro aleatório. O objetivo do modelo é ser uma aproximação simplificada da relação real entre as variáveis de interesse.

A regressão linear, no contexto de séries temporais, utiliza o método de mínimos quadrados para projetar uma equação linear que melhor se encaixe nos dados apresentados e assim fazer a previsão dos próximos meses. Assume Y_t como uma saída ou séries temporais dependentes, para $t = 1, \dots, n$. Essa variável está sendo influenciada por uma coleção de entradas possíveis ou séries independentes, tais como $X_{t1}, X_{t2}, \dots, X_{tq}$, consideradas como entradas fixas e conhecidas. Logo, essa relação pode ser expressa por

$$Y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_q X_{tq} + \varepsilon_t, \quad (2)$$

em que $\beta_1, \beta_2, \dots, \beta_q$ são coeficientes de regressão fixos desconhecidos e ε_t é um erro aleatório ou processo de ruído e $\varepsilon_i \sim N(0, \sigma^2)$ (lê-se os erros seguem distribuição normal com média zero e variância σ^2). Para mais detalhes sobre os modelos de regressão no contexto de séries temporais sugere-se a leitura de Wooldridge (2015).

3.3. Método de Previsão por Modelos de Séries Temporais

Shumway et al. (2011) descrevem uma série temporal como o conjunto de observações $\{Y(t), t \in T\}$ em que Y é a variável de interesse, a qual pode ser discreta, contínua ou multivariada, e T é o conjunto de índices. Segundo os autores, a correlação óbvia introduzida pela amostragem de pontos adjacentes no tempo pode restringir severamente a aplicabilidade de muitos métodos estatísticos convencionais, tradicionalmente dependentes da suposição de que essas observações são independentes e identicamente distribuídas. A abordagem sistemática a qual responde às questões estatísticas colocadas por essas correlações de tempo são referidas como análise de séries temporais.

A previsão de valores futuros pode ser feita por regressão linear, bem como por meio de métodos estatísticos mais complexos, como os modelos autoregressivos de médias móveis (Antunes et al., 2015). O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série. As séries temporais podem ser compostas dos seguintes fatores:

- tendência, que é o comportamento de longo prazo da série;
- sazonalidade, que são flutuações periódicas nos valores da variável;
- ciclos, são aquelas que provocam oscilações de subida e de queda nas séries, de forma suave e repetitiva, ao longo da componente de tendência;
- erro ou variação irregular, que ocorre devido a efeito de um dos outros fatores. Nas componentes de erro aparecem flutuações de período curto e com deslocamento inexplicável.

Os modelos utilizados para descrever séries temporais são processos estocásticos, controlados por leis probabilísticas. Denotando T um conjunto arbitrário, um processo estocástico é uma família $\{Y_t, t \in T\}$ tal que para cada $t \in T$, Y_t é uma variável aleatória. Um processo estocástico está determinado quando são conhecidas suas funções de distribuição de probabilidade conjuntas. Quando essas não são conhecidas e dispõe-se de apenas uma amostra do processo (a série temporal observada), assumem-se os pressupostos de estacionariedade e ergodicidade do processo estocástico. Segundo Morettin e Tolo (2006), esses fatores quanto à variabilidade das observações podem ser definidos como:

- Estacionariedade: as estatísticas não são afetadas por variações no tempo. Assim, uma série é estacionária se os dois primeiros momentos são constantes ao longo do tempo, ou seja, $E(Y_t) = \mu$ e $\text{Var}(Y_t) = \sigma^2$.
- Ergodicidade: se apenas uma realização do processo estocástico é suficiente para se obter todas as suas estatísticas. Um processo é dito ergódico em média se suas médias temporais e amostrais coincidem quando T tende a infinito.

Todo o processo ergódico também é estacionário, pois uma realização de um processo não estacionário não poderá conter todas as informações necessárias para a especificação do processo.

Diniz et al. (1998) afirmam que uma série temporal é estacionária se a média $E(Y_t) = \mu$, que mede o valor médio dos dados; a variância $\text{Var}(Y_t) = \sigma^2$, que mede o grau de dispersão dos dados em relação ao valor médio; e a covariância γ_t , que mensura o grau de dispersão entre um dado e seu subsequente, são constantes ao longo do tempo. Para verificar essa condição, será

realizado o teste de raiz unitária, no qual, a presença de uma ou mais raízes unitárias indica não estacionariedade no comportamento de uma série histórica, ou seja, os valores tendem a aumentar com o transcorrer do tempo.

Segundo Granger e Newbold (1974), ao se ajustar um modelo de regressão, mesmo que se constate a significância dos parâmetros, e que o coeficiente de determinação da regressão (R^2) seja elevado, pode-se obter uma regressão espúria quando a variável apresenta raiz unitária. Isso porque os pressupostos de que a média e a variância devem ser constantes ao longo do tempo são violados, comprometendo os resultados. Dessa forma, em uma série que não seja estacionária, a previsão para o futuro não será eficiente. Portanto, realizar-se-á o teste de razão unitária para verificação da condição de estacionariedade. Para maiores detalhes, recomenda-se a leitura das referências anteriormente mencionadas.

3.3.1. Funções de Autocorrelação

A autocovariância é a covariância entre Y_t e o seu valor Y_{t-k} separado por k unidades de tempo e denota-se por:

$$\gamma_k = Cov[Y_t, Y_{t-k}] = E([Y_t - \mu][Y_{t-k} - \mu]), \quad (3)$$

para $k = 0, \pm 1, \pm 2, \dots$

Autocorrelação, ou Autocorrelation Function (ACF), é a autocovariância padronizada, a qual mede o comprimento e a memória de um processo, ou seja, a extensão para a qual o valor tomado no tempo t depende daquele tomado no tempo $t-k$ ou $t+k$ e denota-se por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov[Y_t, Y_{t-k}]}{\sqrt{Var(Y_t)Var(Y_{t-k})}} \quad (4)$$

Essa ideia de autocorrelação pode ser estendida, isto é, pode-se medir a correlação entre duas observações, eliminando a dependência dos termos intermediários. A autocorrelação parcial, Partial Autocorrelation Function (PACF) é denotada por:

$$Cov[Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-(k+1)}]. \quad (5)$$

3.3.2. Modelo Autorregressivo Integrado e de Médias Móveis – ARIMA

De acordo com Shumway et al. (2011), uma regressão clássica, como proposta no tópico 3.2, muitas vezes não é suficiente para explicar todas as dinâmicas de uma série temporal. Em vez disso, a introdução da correlação como um fenômeno que pode ser gerado por meio de relações lineares defasadas leva a propor os modelos autorregressivo (AR) e autorregressivo de

média móvel (ARMA). Esses, apesar de complexos, ajustam-se aos fatores sazonais e de tendência, estimam parâmetros adequados de peso, testam o modelo e repetem o ciclo quando adequado.

O modelo ARIMA, proposto pelos estatísticos Box e Jenkins (1994), é uma difusão do modelo geral ARMA, aplicado a casos em que os dados são não estacionários. Essa questão da não estacionariedade pode ser resolvida a partir de diferenciações, o que corresponde a parte integrada (I) do modelo. Ambos os modelos podem ser usados para prever o futuro de uma série, dado seus pontos passados. Para esses espaços de tempo, Box-Jenkins indicam um número de no mínimo 40 a 50 pontos. Os dados utilizados neste trabalho atendem esse requisito, dado que o tamanho da amostra é 132. A previsão de Box-Jenkins é de maior utilidade quando se acredita que os fatores subjacentes que causam a demanda por produtos, serviços, receita e, neste caso, a taxa de sinistralidade se comportem no futuro da mesma maneira que no passado.

A parte autorregressiva dos modelos é estruturada por Y_t , que representa os valores atuais da série explicados em função de p valores passados. Logo, é expresso por AR(p) e denota-se por:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t, \quad (6)$$

em que Y_t é estacionária e os coeficientes $\varphi_1, \varphi_2, \dots, \varphi_p$ são constantes ($\varphi_p \neq 0$). Assume-se que $\varepsilon_i \sim N(0, \sigma^2)$ são independentes. Portanto, essa parte do modelo indica que a variável é regressada em seus próprios valores defasados. Por outro lado, assumindo que os ruídos brancos e_t estruturam a parte de Médias Móveis de ordem q , denota-se por MA(q) e se expressa por:

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

em que $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros e $\theta_q \neq 0$. Novamente, assume-se que $\varepsilon_i \sim N(0, \sigma^2)$, salva indicação contrária. Logo, indica que o erro de regressão é na verdade uma combinação linear dos termos de erro, cujos valores ocorreram contemporaneamente e em vários momentos no passado. Vale ressaltar que, diferente do AR(p), a parte MA(q) é estacionária para quaisquer valores dos parâmetros θ .

Ademais, a parte integrada (I) indica que os valores de dados foram substituídos com a diferença entre seus valores e os valores anteriores e este processo diferenciador pode ter sido realizado mais de uma vez, como explicado acima. Seguindo neste trabalho, na parte de resultados, será mostrado que não é este o caso, isto é, não foi preciso diferenciar a série pois ela já é estacionária. Finalmente, as partes se juntam, formando os modelos ARMA e ARIMA, a depender da estacionariedade. Portanto, o modelo pode ser representado por:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (8)$$

em que os parâmetros p e q são chamados ordens autorregressivas e de médias móveis, respectivamente. Neste trabalho, para calcular a previsão para os valores futuros a partir do modelo descrito, foi utilizada a função *arima*, do pacote *stats*, no software R.

4. RESULTADOS

4.1. Análise Descritiva

Os dados originais possuíam dois outliers negativos e dois valores de sinistralidade maiores do que 1, sendo, inclusive, consideravelmente maiores que o terceiro quartil. Essas taxas correspondem aos meses de jan/11, dez/11, mai/16 e dez/17. Dessa forma, a fim de gerar a melhor previsão possível com os dados disponíveis, esses outliers foram ajustados. Para os outliers maiores do que 1, foi utilizado a média dos 10 maiores valores de sinistralidade menores do que 1. Já para os outliers negativos, foi utilizado a média dos 10 menores valores de sinistralidade maiores do que zero. O resumo dos novos dados em comparação aos originais pode ser observado na Tabela 1.

Tabela 1: Estatísticas Descritivas dos Dados Originais e Ajustados

Estatísticas Descritivas	Sinistro		Sinistralidade	
	Originais	Ajustados	Originais	Ajustados
Mínimo	-4.046.445,00	2.335.441,00	-0,0800	0,0800
1º Quartil	9.465.237,00	9.465.237,00	0,2100	0,2100
Mediana	15.674.779,00	15.674.779,00	0,2600	0,2600
Média	20.821.577,00	20.900.212,00	0,2940	0,2852
3º Quartil	25.194.276,00	25.194.276,00	0,3225	0,3225
Máximo	95.432.888,00	95.432.888,00	1,4900	0,6400
Desvio Padrão	18.232.735,00	18.134.014,00	0,1750	0,1197

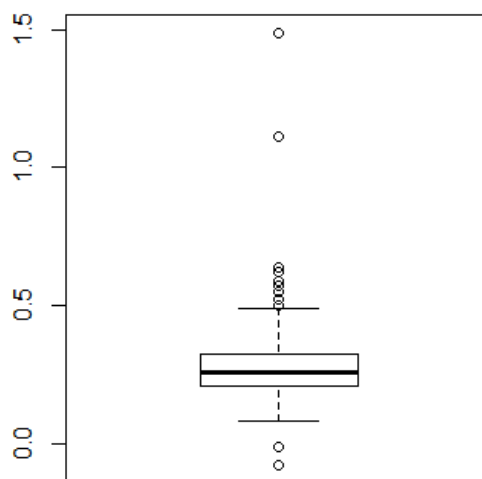
Fonte: Dados da pesquisa.

Observa-se que os valores mínimo e máximo de sinistralidade ficaram mais próximos da mediana e média sem outliers. Vale ressaltar que, mesmo com o ajuste proposto para os valores acima de 1, há ainda algumas taxas que se distanciam da realidade das demais, como pode ser analisado por meio do terceiro quartil, indicando que 75% dos dados apresentam

sinistralidade de até 0,32, em comparação com a sinistralidade máxima de 0,64. A efetividade desses ajustes pode ser melhor visualizada através dos boxplots da sinistralidade dos dados com e sem outliers, a seguir.

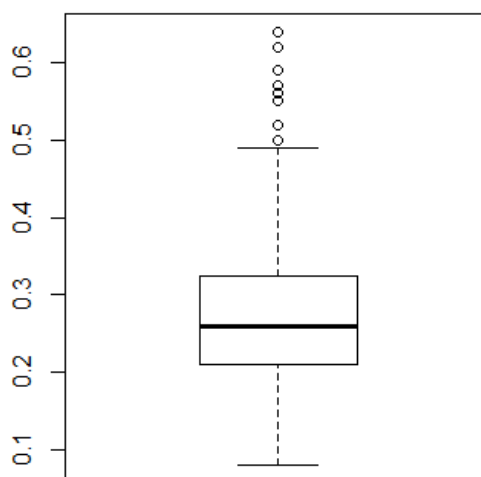
As Figuras 1 e 2 indicam limite superior da sinistralidade de aproximadamente 0,5. Embora valores acima desse limite sejam definidos como outliers, ele serão considerados na previsão por serem frequentes no conjunto de dados.

Figura 1: Boxplot da Sinistralidade (original)



Fonte: Elaborada pelas autoras.

Figura 2: Boxplot da Sinistralidade (ajustado)

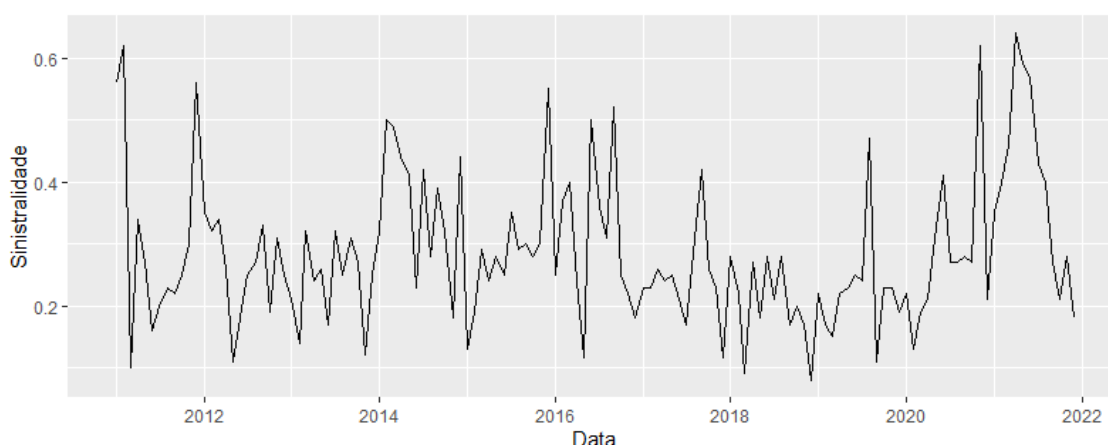


Fonte: Elaborada pelas autoras.

Por meio dos dados ajustados verifica-se que as taxas de sinistralidade observadas são, em maioria, menores do que 0,35. A reduzida proporção entre despesas com sinistros e o prêmio ganho tem um impacto direto na solidez financeira dessas seguradoras, influenciando positivamente a capacidade de retenção de lucros. Além disso, essa diminuição auxilia na comprovação da qualidade da subscrição e na eficácia dos cálculos para a precificação dos riscos. A partir deste ponto, foi utilizada a base com os dados ajustados para as demais análises.

A Figura 3 apresenta a série completa ao longo do tempo. Pelo gráfico é possível notar que o fator da estacionariedade, essencial para o cálculo das previsões por séries temporais, provavelmente será atendido. Essa hipótese será comprovada a partir dos testes de raiz unitária.

Figura 3: Sinistralidade no período de 2011 a 2021



Fonte: Dados da pesquisa.

4.2. Testes de Normalidade e de Raiz Unitária

Para testar a normalidade dos dados foi utilizado o teste de Shapiro-Wilk. A partir do p-valor do teste e considerando um nível de significância de 5%, concluiu-se que a hipótese nula dos dados serem normais é rejeitada. Diante disso, os dados da série foram transformados a partir do logaritmo (a transformação tende a suprimir flutuações maiores que ocorrem em partes da série) e o teste foi aplicado para a variável transformada, indicando, ao nível de significância de 5%, evidência estatística a favor da hipótese de normalidade.

Para a análise de raiz unitária, o primeiro teste adotado foi o Teste de Dickey Fuller aumentado (ADF). A hipótese nula desse teste é que a série tenha raiz unitária. Se o valor da estatística, calculada intrinsecamente no teste ADF, for maior que o valor absoluto tabulado por Dickey-Fuller, a hipótese nula é aceita e, portanto, a série é não estacionária. Foi utilizada a

função *adf.test* do pacote *tseries* no software R para realização do teste, o qual indicou evidência estatística a favor da hipótese de estacionariedade.

Proposto por Nusair (2003), o cruzamento dos testes ADF e o Kwiatkowski Philips Schmidt & Shin (KPSS) assegura uma conclusão mais precisa sobre a estacionariedade da série. Portanto, a fim de diminuir a incerteza do teste ADF, foi utilizado também o teste KPSS, a partir da função *PP.test* também do pacote *tseries*, o qual também indicou evidência estatística a favor da hipótese de estacionariedade.

4.3. Resultados dos Modelos de Regressão

Nesta seção, foram testados três modelos diferentes. No primeiro, a Sinistralidade foi ajustada pelo período de tempo t (correspondendo a jan/11 a dez/21) mais o quadrado de t , correspondendo a tendência. O segundo modelo adiciona ao Modelo 1 o fator de sazonalidade. Por fim, o terceiro adiciona ao Modelo 2 o termo autoregressivo de ordem 1 aos termos de erro e_t .

A seguir, serão apresentados os critérios de comparação, levando em consideração a normalidade dos resíduos pelo teste Shapiro-Wilk e a estatística Durbin-Watson. A estatística Durbin-Watson (DW) é um teste para autocorrelação de primeira ordem dos resíduos da regressão com valores entre 0 e 4. Valores entre 0 e 2 indicam autocorrelação positiva, enquanto valores entre 2 e 4 significam autocorrelação negativa. Já um valor igual a 2 indica que não há autocorrelação detectada na amostra. Para a seleção de modelo em caso de empate, será considerado o Critério de Informação de Akaike (AIC), o qual quanto menor melhor.

Tabela 2: Critérios de Seleção para os Modelos de Regressão

	AIC	Durbin-Watson Estatística	Durbin-Watson p-valor	Shapiro-Wilk p-valor
Modelo 1	147,90	1,5144	0,004	0,0394
Modelo 2	165,47	1,5232	0,004	0,1330
Modelo 3	160,47	2,0535	0,862	0,1246

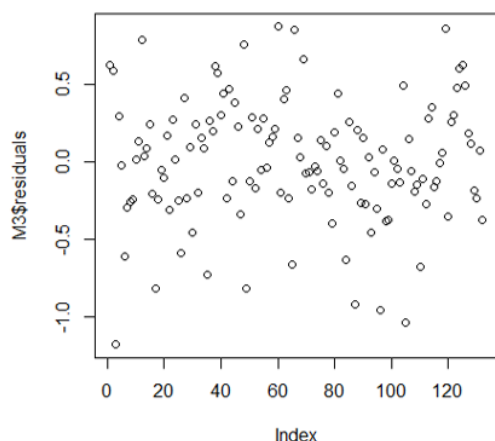
Fonte: Dados da pesquisa.

A Tabela 2 indica que o Modelo 1 foi o único a rejeitar o teste de normalidade dos resíduos (SW), a um nível de significância de 5%. Apesar do AIC do Modelo 3 ser maior do

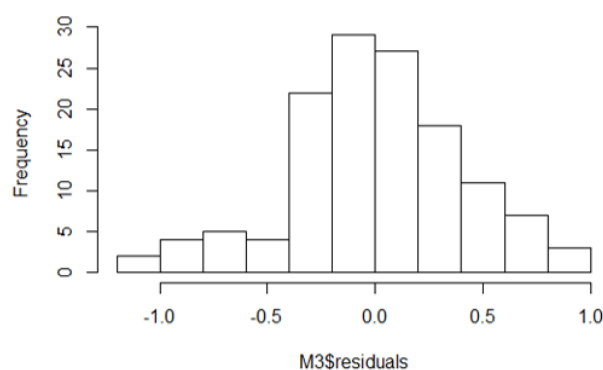
que do Modelo 1, é preciso primeiramente avaliar o teste de Durbin-Watson, o qual indica que o Modelo 3 é o mais adequado, já que a estatística DW é aproximadamente 2, enquanto os demais estão dentro do intervalo 0 a 2. Para uma análise mais completa do modelo selecionado, apresenta-se a seguir os gráficos de resíduos do Modelo 3.

A partir da Figura 4, é possível observar que os resíduos estão bem distribuídos, sem manifestar um padrão, ou seja, nuvem aleatória de pontos em torno de zero. A Figura 5 indica que os resíduos apresentam uma leve assimetria, indicando que cabe uma distribuição aproximadamente normal ao ajuste dos dados.

Figura 4: Dispersão dos Resíduos do Modelo 3 **Figura 5: Histograma dos Resíduos do Modelo 3**



Fonte: Dados da pesquisa.



Fonte: Dados da pesquisa.

Finalmente, a partir da metodologia descrita e do modelo escolhido, obtém-se as previsões para o 1º semestre de 2022, como pode ser observado na Tabela 3. Incluiu-se também os valores reais para fins de comparação entre o observado e o previsto.

Tabela 3: Sinistralidade Real versus Prevista – Modelo de Regressão

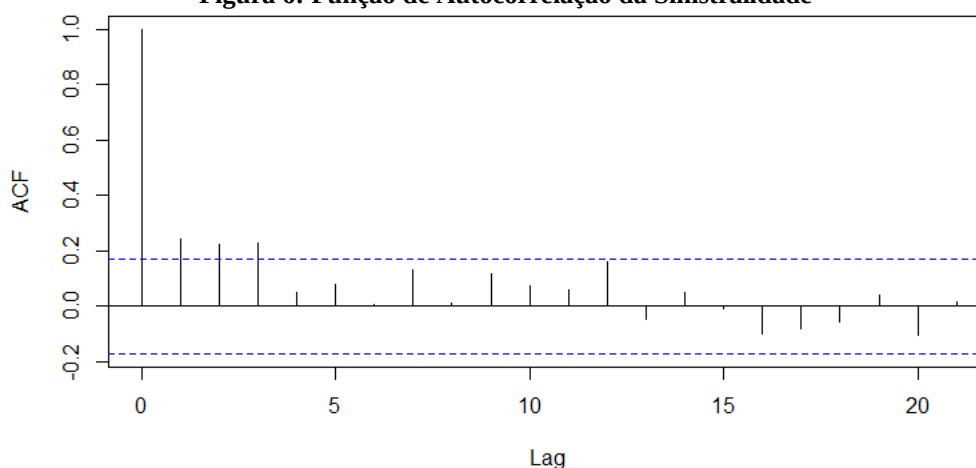
Mês	Real	Previsão	IC inferior	IC superior
jan/22	0,23	0,28	0,13	0,60
fev/22	0,29	0,30	0,13	0,66
mar/22	0,38	0,28	0,12	0,65
abr/22	0,34	0,33	0,14	0,81
mai/22	0,22	0,29	0,11	0,75
jun/22	0,29	0,31	0,11	0,89

Fonte: Dados da pesquisa.

A adequação da previsão é notável, especialmente quando comparada com os valores reais observados para sinistralidade do período, visto que todos os intervalos de confiança contêm os valores reais. Embora haja uma diferença de aproximadamente 30% nos meses de março e maio, é importante ressaltar que os meses de fevereiro e abril apresentaram diferenças inferiores a 3%.

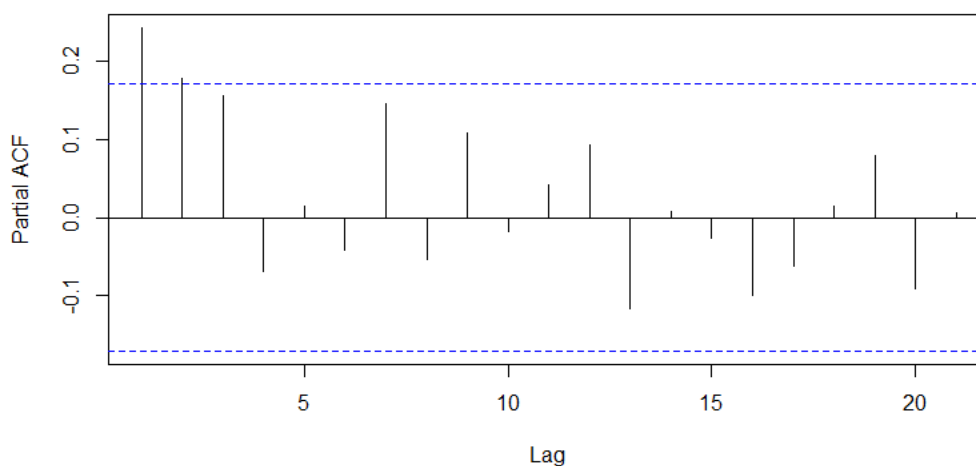
4.4. Resultados dos Modelos ARIMA

Em um processo estacionário, a Função de Autocorrelação (ACF) decai para zero rapidamente, indicando que a correlação em relação ao tempo passado diminui exponencialmente. Por outro lado, um decaimento lento da ACF sugere que a série não é estacionária e requer diferenciação. Portanto, a partir da Figura 6, é possível observar 3 lags além do ponto crítico.

Figura 6: Função de Autocorrelação da Sinistralidade

Fonte: Dados da pesquisa.

No processo autoregressivo, a Função de Autocorrelação (ACF) exibe um decaimento exponencial ou senoidal amortecido. Já a Função de Autocorrelação Parcial (PACF) mostra picos significativos nos lags 1 até p , seguido por uma queda abrupta para zero. Portanto, com base na Figura 7, é possível identificar 2 lags fora do ponto crítico.

Figura 7: Função de Autocorrelação Parcial da Sinistralidade

Fonte: Dados da pesquisa.

A combinação dos modelos $AR(p)$ e $MA(q)$ dá origem ao modelo $ARMA(p,q)$. O parâmetro p , da parte autoregressiva pode ser identificado a partir do gráfico PACF, e o q , das médias móveis, do gráfico ACF. A identificação se dá pelo número de lags fora dos pontos

críticos dos gráficos. Portanto, dado os 3 pontos observados no ACF e 2 no PACF, conclui-se que o modelo é um ARMA(2,3) ou ARIMA(2,0,3). Porém, para minimizar ao máximo o número de parâmetros do modelo, foi utilizada a função *coefstest* do pacote *lmtest* no software R. Através dela, é possível determinar quais coeficientes de fato são significativos, complementando a análise dos gráficos de autocorrelação. Nos testes foram considerados os coeficientes de maior p-valor até chegar no resultado final, um ARIMA(1,0,1).

Complementado esse processo inferencial, foi utilizada a função *auto.arima* do pacote *forecast*. Essa função usa uma variação do algoritmo desenvolvido por Hyndman e Khandakar (2008) que combina testes de raiz unitária para obter um modelo ARIMA. Por fim, dentre todas as opções analisadas pelos métodos acima, foram selecionados os modelos com menores AIC, na tabela seguir.

Tabela 4: Critério AIC para Seleção dos Modelos ARIMA

Modelo	AIC
ARIMA(1,0,1)(0,0,1)[12]	134,31
ARIMA(1,0,1)(1,0,0)[12]	134,64
ARIMA(1,0,1)	136,86
ARIMA(2,0,2)	137,40

Fonte: Dados da pesquisa.

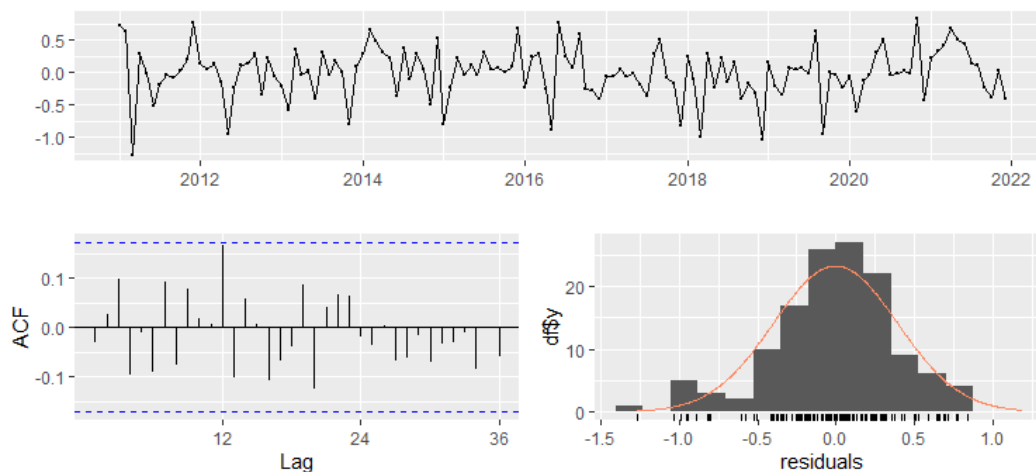
A Tabela 4 indica, portanto, que o melhor modelo para a sinistralidade é o ARIMA(1,0,1)(0,0,1)[12]. Esse resultado valida o resultado da análise para estimação do modelo realizada previamente, adicionando um termo que se refere ao desenvolvimento para a parte sazonal. O valor 12 entre colchetes é o número de períodos, o que vai de encontro aos dados, que são observações anuais subdivididas entre os meses de janeiro a dezembro.

4.4.1 Medidas de Diagnóstico

A seguir, serão considerados para o diagnóstico os modelos ARIMA(1,0,1) e ARIMA(1,0,1)(0,0,1)[12], a fim de analisar o efeito da possível sazonalidade na previsão. Será necessário verificar se os resíduos do modelo são autocorrelacionados e, para isso, será utilizado o teste de Ljung-Box. Sua hipótese nula é que os resíduos do modelo observado são conjuntamente não correlacionados ao longo do tempo. Por conseguinte, a não independência dos resíduos seria um indicador de falha na estrutura do modelo proposto. O p-valor do modelo

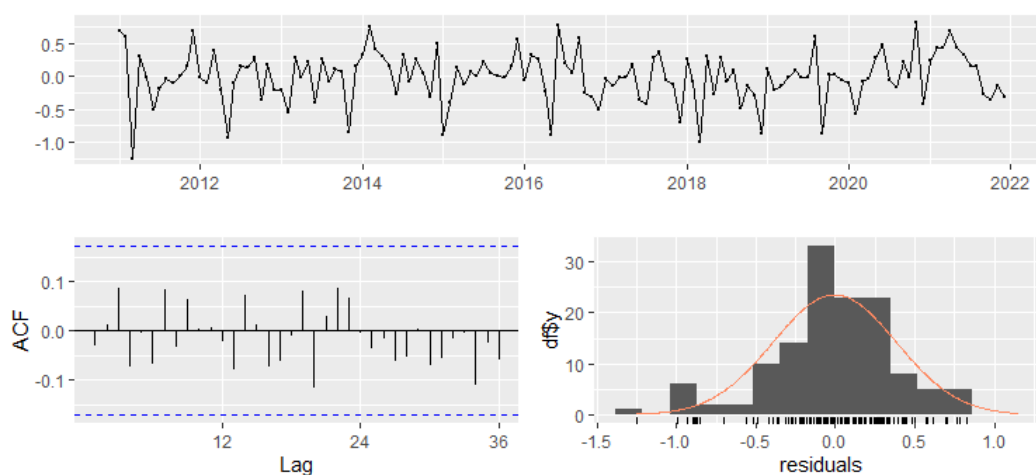
ARIMA(1,0,1) foi de 0,4557 e do modelo ARIMA(1,0,1)(0,0,1)[12] foi de 0,8859. Portanto, considerando um nível de significância de 5%, há evidências para concluir que os resíduos de ambos os modelos são conjuntamente não correlacionados.

Figura 8: Análise dos Resíduos para o Modelo ARIMA(1,0,1)



Fonte: Dados da pesquisa.

Figura 9: Análise dos Resíduos para o Modelo ARIMA(1,0,1)(0,0,1)[12]



Fonte: Dados da pesquisa.

Para complementar os testes, os correlogramas dos resíduos são expostos nas Figuras 8 e 9, acima apresentados. A partir deles, é possível observar que os resíduos de ambas as séries são bem distribuídos ao longo do tempo. Além disso, nos ACF, não há lags que ultrapassem as linhas de limite. Por fim, pelos histogramas, nota-se que ambos os modelos, sem e com sazonalidade, apresentam simetria em torno de zero.

4.4.2 Previsões

Seguindo a lógica da subseção anterior, foram realizadas duas sequências de previsões, cada uma utilizando um modelo ARIMA, para os dados do 1º semestre de 2022, como pode ser observado por meio da Tabela 5.

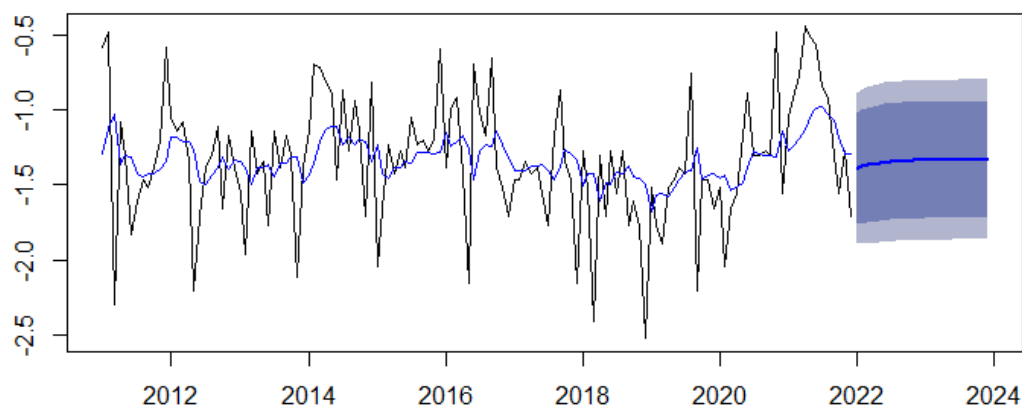
Tabela 5: Sinistralidade Real versus Prevista pelos Modelos ARIMA

Mês	Real	Previsão ARIMA(1,0,1)	Previsão ARIMA(1,0,1)(0,0,1)[12]
jan/22	0,23	0,39	0,39
fev/22	0,29	0,40	0,40
mar/22	0,38	0,41	0,41
abr/22	0,34	0,41	0,41
mai/22	0,22	0,41	0,41
jun/22	0,29	0,41	0,41

Fonte: Dados da pesquisa.

A projeção da série temporal, onde a linha azul representa os valores ajustados, será apresentada junto com intervalos de confiança, considerando níveis de 65% (representados em cinza claro) e 80% (em cinza escuro). Ao analisar a projeção do modelo ARIMA(1,0,1), é possível observar, a partir da Figura 10 e dos valores da Tabela 5, que este modelo projeta uma série de crescimento constante. A desvantagem desse modelo é que a projeção não leva em consideração os altos e baixos presentes na série observada.

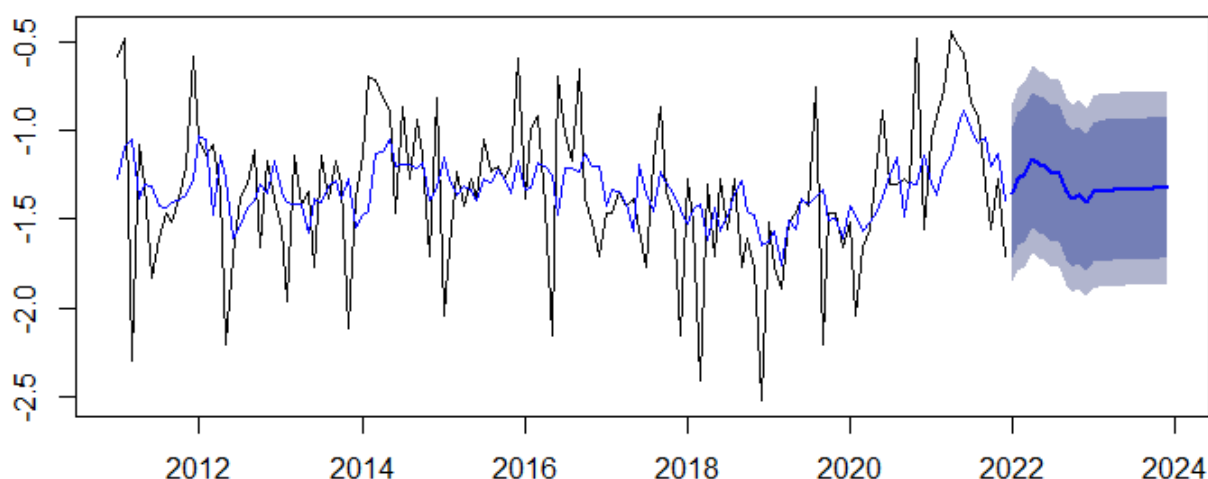
Figura 10: Projeção da Sinistralidade – Modelo ARIMA(1,0,1)



Fonte: Dados da pesquisa.

Na análise do modelo $ARIMA(1,0,1)(0,0,1)[12]$, percebe-se, com base na Figura 11, que a projeção desse modelo parece capturar de forma mais semelhante as variações presentes nos dados observados da série. No entanto, é importante ressaltar que, como somente os dados reais do primeiro semestre de 2022 estavam disponíveis para comparação com a previsão, verifica-se que não há diferença entre as previsões de ambos os modelos, conforme apresentado na Tabela 5.

Figura 11: Projeção da Sinistralidade – Modelo ARIMA(1,0,1)(0,0,1)[12]



Fonte: Dados da pesquisa.

Em relação às previsões, foi observado que a menor diferença entre o valor previsto e o observado da sinistralidade foi de 0,03, enquanto a maior diferença foi de 0,19. É importante

ressaltar que essa flutuação é esperada, uma vez que estamos lidando com uma variável não determinística. A variação não explicada pelo modelo pode ser atribuída a diversos fatores, muitos dos quais são de difícil mensuração (latentes), incluindo características específicas do período analisado, particularidades de cada seguradora, como seus processos de subscrição, entre outros. De maneira geral, as previsões são coerentes, especialmente após a crise da Covid-19 e o aumento das taxas em 2021.

4.5 Comparação entre os Modelos de Regressão e ARIMA

A métrica estatística dos Erros Quadráticos Médios (EQMs) é amplamente utilizada na literatura para comparar a performance de modelos em relação à capacidade preditiva. Assim, para a comparação e seleção entre os modelos propostos neste trabalho para previsão da sinistralidade, optou-se pela utilização do EQM. Quanto menor o EQM, melhor a qualidade de ajuste do modelo, resultando em uma previsão mais precisa.

Tabela 6: Erro Quadrático Médio para os Modelos de Regressão e ARIMA

Estatística	Regressão	ARIMA(1,0,1)	ARIMA(1,0,1)(0,0,1)[12]
EQM	0,00283	0,01629	0,01595

Fonte: Dados da pesquisa.

A partir das estatísticas apresentadas na Tabela 6, observa-se que o modelo de regressão possui o menor EQM. Conclui-se, portanto, que este modelo foi o que melhor previu a sinistralidade do 1º semestre de 2022, de acordo com o ajuste realizado com os dados de janeiro de 2011 a dezembro de 2021. Vale ressaltar a proximidade entre os EQMs do ARIMA(1,0,1) e do ARIMA(1,0,1)(0,0,1)[12], o que era esperado, dado que as projeções foram semelhantes, como observado na subseção 4.4.2 sobre as previsões.

5. CONSIDERAÇÕES FINAIS

Após uma vasta análise dos dados observados entre 2011 e 2021, este trabalho indicou três modelos diferentes para a previsão da sinistralidade em seguros de vida para o 1º semestre de 2022. No geral, os valores encontram-se dentro do intervalo de confiança, embora a análise comparativa dos modelos sugira que o modelo de regressão foi capaz de prever com mais precisão as taxas de sinistralidade para o período analisado, atingindo assim os objetivos desta

pesquisa. Espera-se que o trabalho contribua com a literatura sobre o tema, visto que há poucos estudos no contexto do mercado de seguros.

Para possíveis trabalhos futuros, sugere-se o estudo da aplicabilidade de outros modelos de previsão nos dados de seguro do ramo vida individual. Além disso, cabe o uso dos modelos propostos para outros tipos de seguro, como de vida em grupo, que possui outras especificidades. Além disso, sugere-se uma investigação mais aprofundada dos dados referentes ao período da pandemia de Covid-19, a fim de analisar o impacto da pandemia nas taxas de sinistralidade e verificar se as previsões seriam afetadas de alguma forma.

REFERÊNCIAS

- Antunes, J. L. F.; Cardoso, M. R. A. (2015). Uso da análise de séries temporais em estudos epidemiológicos. *Epidemiol. Serv. Saúde*, Brasília, 24(3), 565-576. <http://scielo.iec.gov.br/pdf/ess/v24n3/v24n3a24.pdf>
- Box, G. E.; Jenkins, G. M. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs.
- Carvalho, G. Q. F.; Gonçalves, J. N. (2022). Estimação do Índice de Desempenho da Saúde Suplementar (IDSS): Inferência Utilizando a Classe de Modelos GAMLSS. In: *Anais do Congresso de Ciências Contábeis e Atuariais da Paraíba - CONCICAT*, João Pessoa. <https://www.concicatufpb.com.br/anaisconcicat-2022>
- Chaves, M. E. D.; Mataveli, G. A. V.; Justino, R. C. (2014). Uso da modelagem estatística para monitoramento da vegetação no Parque Nacional da Serra da Canastra, Minas Gerais. *Caderno de Geografia*, 24(1), 120-132. <https://doi.org/10.5752/P.2318-2962.2014v24nespp120>
- CNSP. Resolução CNSP Nº 117, de 22 de dezembro de 2004; Resolução CNSP Nº 348, de 25 de setembro de 2017.
- Correa, B. (2022). Os impactos causados nas seguradoras do ramo vida pelo aumento da sinistralidade devido a pandemia de Covid-19. UFRGS.
- Diniz, H.; Andrade, L. C. M.; Carvalho, A. C. P.; Andrade, M. G. (1998). Previsão de séries temporais utilizando redes neurais artificiais e modelos de Box e Jenkins. In: *Anais do Simpósio Brasileiro de Redes Neurais*, 173-178.
- Gonçalves, J.N., Barreto-Souza, W. (2020). Flexible regression models for counts with high-inflation of zeros. *METRON* 78, 71-95. <https://doi.org/10.1007/s40300-020-00163-9>
- Granger, C. W. J.; Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2, 111-120. [https://doi.org/10.1016/0304-4076\(74\)90034-7](https://doi.org/10.1016/0304-4076(74)90034-7)
- Guimarães, A. L. S.; Alves, W. O. (2009). Prevendo a insolvência de operadoras de planos de saúde. *Revista de Administração de Empresas*, 49, 459-471. <https://doi.org/10.1590/S0034-75902009000400009>
- Huebner, S. S.; Black, K. J. (1976). *El seguro de vida*. Madrid: Editorial Mapfre. <https://documentacion.fundacionmapfre.org/documentacion/publico/es/bib/1110.do>



- Hyndman, R. J; Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 27(1), 1-22.
<<https://www.jstatsoft.org/article/view/v027i03>>
- Lima, C. A. (2018). A história do seguro. Revista Apólice.
<<https://revistaapolice.com.br/2018/10/a-historia-do-seguro/>>
- Machado, M. A. L. (2012). Modelos de previsão aplicados à otimização da gestão das actividades de um Call Center. Universidade de Lisboa, Lisboa.
<<http://hdl.handle.net/10451/9422>>
- Morettin, P. A.; Toloi, C. M. C. (2006). Análise de séries temporais. São Paulo: Edgard Blucher.
- Mori, F. T. M.; Gonçalves, L. R. (2016). Aplicação da metodologia de séries temporais ao sistema de saúde do Brasil e dos Estados Unidos (2000-2012). Revista Debate Econômico, 4(1).
<<https://publicacoes.unifal-mg.edu.br/revistas/index.php/revistadebateeconomico/article/view/288>>
- Nusair, S. (2003). Testing the validity of purchasing power parity for asian countries during the current float. Journal of Economic Development, 28(2), 129-147.
- Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). Time series analysis and its applications, 3, New York: Springer.
- SUSEP. Circular SUSEP N° 302, de 19 de setembro de 2005; Circular SUSEP N° 395, de 3 de dezembro de 2009.
- SUSEP. Sistema de Estatísticas da Superintendência de Seguros Privados.
- Vianna, F. G. (2018). Sinistralidade das operadoras de planos privados de assistências à saúde médico-hospitalar: determinação dos fatores associados a esse índice e seus efeitos. UFMG.
- Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.