

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas

Gustavo Pinheiro Alves

**MODELAGEM ESTOCÁSTICA PARA O TAP (TESTE DE ADEQUAÇÃO DO
PASSIVO) NO RAMO DA SAÚDE SUPLEMENTAR**

Belo Horizonte, MG
2023

GUSTAVO PINHEIRO ALVES

**MODELAGEM ESTOCÁSTICA PARA O TAP (TESTE DE ADEQUAÇÃO DO
PASSIVO) NO RAMO DA SAÚDE SUPLEMENTAR**

Trabalho de Conclusão de Curso apresentado ao Curso de Ciências Atuariais do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do grau de bacharel em Ciências Atuariais.

Orientador: Prof. Dr. Adrian Pablo Hinojosa Luna

Belo Horizonte, MG
Instituto de Ciências Exatas da UFMG
2023

2023, Gustavo Pinheiro Alves.
Todos os direitos reservados

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar saúde e sabedoria para que fosse possível terminar a possível terminar este Trabalho de Conclusão de Curso. Agradeço aos meus pais, pelo amor e incentivo que sempre me deram. Agradeço ao meu orientador, Adrian Hinojosa Luna, pela orientação e paciência que foram fundamentais para a realização deste trabalho. Agradeço a todos os professores, pelos ensinamentos que me proporcionaram e a todos que contribuíram para a conclusão deste TCC diretamente ou indiretamente.

RESUMO

O estudo apresenta uma análise estocástica do Teste de Adequação do Passivo no ramo da Saúde Suplementar. Utilizando-se de artigos e estudos de diferentes áreas da estatística e atuária com o intuito de explorar esse campo recente do TAP na saúde suplementar. Para tal, foram usados dados assistenciais disponíveis no portal de dados abertos disponível no site da Agência Nacional de Saúde (ANS) e no site do Banco Central do Brasil onde foram baixados os valores da taxa SELIC. Em seguida, após alguns trabalhos de filtração e transformação dos dados, além de pressupostos estatísticos, foram feitas previsões com os modelos ARIMA e modelo CIR a fim de mensurar a solvência da operadora de plano de saúde por meio do TAP. Além disso, foram analisados cenários de riscos envolvendo incertezas que afetam o setor de maneira significativa como aumento nas despesas assistenciais e elevação da taxa de juros. Com isso, traçar novas estratégias e métodos de estimar essa adequação das obrigações assumidas.

Palavras-chave: Teste de Adequação do Passivo, Medidas de Risco, Métodos de Monte Carlo, Equações Estocásticas, Saúde Suplementar, Série Temporais.

ABSTRACT

The study presents a stochastic analysis of the Liability Adequacy Test in the Supplementary Health sector. Using articles and studies from different areas of statistics and actuarial science with the aim of being a pioneer in this recent field of TAP in supplementary health. To this end, healthcare data available on the open data portal available on the National Health Agency (ANS) website and on the Central Bank of Brazil website where the SELIC rate values were downloaded were used. Then, after some filtering and data transformation work, in addition to statistical assumptions, predictions were made with the ARIMA models and CIR model in order to measure the solvency of the health plan operator through the TAP. In addition, risk scenarios were analyzed involving uncertainties that significantly affect the sector, such as an increase in assistance expenses and an increase in interest rates. Therefore, devise new strategies and methods to estimate the adequacy of the obligations assumed.

Key-words: Liability Adequacy Test, Risk Measures, Monte Carlo Methods, Stochastic Equations, Supplemental health insurance, Time Series.

LISTA DE ILUSTRAÇÕES E TABELAS

Figura 1: Exemplo de séries estimadas.....	23
utilizando o processo CIR com $a=1$, $b=1$ e $\sigma=1$	23
Figura 2: Exemplo de séries estimadas utilizando o processo CIR com $a=3$, $b=1$ e $\sigma=1$	23
Figura 3: Previsão Esperada do Ticket Médio de planos Coletivos Empresariais com os estimadores ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	28
Figura 5: Previsão Esperada do Ticket Médio de planos Coletivos por Adesão com os estimadores ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	29
Figura 6: Previsão Esperada do Custo per Capita dos plano Coletivos por Adesão com os estimadores ARIMA($p=1$, $d=0$, $q=1$)($P=1$, $D=2$, $Q=1$) ₁₂	29
Figura 7: Previsão Esperada do Ticket Médio de planos Individuais (Familiares) com os estimadores ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	30
Figura 8: Previsão Esperada do Custo per Capita de planos Individuais (Familiares) com os estimadores ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	30
Figura 9: Previsão Esperada de beneficiários nos planos Coletivos Empresariais com os estimadores ARIMA($p=1$, $d=1$, $q=0$)($P=2$, $D=1$, $Q=0$) ₁₂	31
Figura 10: Previsão Esperada de beneficiários nos planos Coletivos por Adesão com os estimadores ARIMA($p=1$, $d=1$, $q=0$)($P=1$, $D=0$, $Q=0$) ₁₂	31
Figura 11: Previsão Esperada de beneficiários nos planos Individuais (Familiares) com os estimadores ARIMA($p=0$, $d=1$, $q=2$)($P=0$, $D=1$, $Q=1$) ₁₂	32
Tabela 1: Parâmetros estimados.....	33
Figura 13: Taxa SELIC observada e previsão média da Taxa estimada ao longo do tempo.....	34
Figura 14: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Coletivos Empresariais ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	34
Figura 15: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Coletivos por Adesão ARIMA($p=1$, $d=0$, $q=1$)($P=1$, $D=2$, $Q=1$) ₁₂	35
Figura 16: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Individuais ARIMA($p=1$, $d=0$, $q=0$)($P=1$, $D=2$, $Q=1$) ₁₂	35
Figura 17: Taxa SELIC observada e previsão máxima da Taxa estimada ao longo do tempo.....	35
Figura 18: Histograma do TAP.....	36
Tabela 2: Intervalo de confiança de 95% para o VaR e TVaR com $\alpha=0,05$	36
Figura 19: Histograma do TAP com custo per capita agravado.....	37
Tabela 3: Intervalo de confiança de 95% para o VaR e TVaR do TAP com $\alpha=0,05$ e com custo per capita agravado.....	37
Figura 20: Histograma do TAP com com a taxa SELIC agravada.....	38

SUMÁRIO

1. INTRODUÇÃO.....	10
2. MODELO DE SOLVÊNCIA VIA TAP.....	16
2.1. TESTE DE ADEQUAÇÃO DO PASSIVO.....	16
2.2. MÉTODOS DE SIMULAÇÃO DE MONTE CARLO.....	18
3. MODELO ESTOCÁSTICO DE JUROS.....	20
3.1. MODELOS ESTOCÁSTICOS: EQUAÇÕES ESTOCÁSTICAS.....	20
3.2. PACOTE YUIMA: INFERÊNCIA PARAMÉTRICA E ESTIMAÇÃO DE QUASE-MÁXIMA VEROSSIMILHANÇA.....	20
3.3. INFERÊNCIA SOBRE EQUAÇÃO ESTOCÁSTICA: MODELO COX-INGERSOLL-ROSS.....	21
4. MODELOS PARA O NÚMERO DE BENEFICIÁRIOS, TICKET MÉDIO E CUSTO PER CAPITA.....	24
4.1. SÉRIES TEMPORAIS E MODELOS ARIMA.....	24
4.2. ANÁLISE E ESTUDOS DA BASE DE BENEFICIÁRIOS.....	27
4.3. ANÁLISE E ESTUDOS DA BASE DE DESPESAS E RECEITAS ASSISTENCIAIS.....	27
5. MODELAGEM DO TAP PARA OPERADORAS DE SAÚDE.....	28
5.1. DESCRIÇÕES DOS MODELOS.....	28
5.2. PACOTE YUIMA: MODELANDO A TAXA DE JUROS.....	32
5.3. EXPLORANDO CENÁRIOS.....	34
5.4. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA PELO TAP.....	36
5.5. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA AGRAVANDO O CUSTO.....	37
5.6. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA AGRAVANDO O JUROS.....	38
6. CONCLUSÕES.....	40
7. BIBLIOGRAFIA.....	41
8. ANEXOS.....	43
8.1. TRABALHANDO AS BASES DE BENEFICIÁRIOS:.....	43
8.2. TRABALHANDO AS BASES DE DESPESAS E RECEITAS ASSISTENCIAIS:.....	46
8.3. TRABALHANDO OS DADOS DA TAXA SELIC:.....	49
8.4. TRABALHANDO AS PREVISÕES DOS PARÂMETROS:.....	51
8.5. TRABALHANDO AS PREVISÕES DOS PARÂMETROS COM AGRAVO NOS CUSTOS PER CAPITA:.....	58
8.6. TRABALHANDO AS PREVISÕES DOS PARÂMETROS COM AGRAVO NA TAXA SELIC:.....	66
8.7. TRABALHANDO A BASE COM PREVISÕES PARA A ESTIMAÇÃO DO TAP:.....	73
8.8. TRABALHANDO OS INTERVALOS DE CONFIANÇA DO VAR E TVAR:.....	75

1. INTRODUÇÃO

O conceito de saúde suplementar começou a partir da década de 1960, momento em que as empresas começaram a oferecer planos de assistência médica aos empregados. Foi uma época marcada pelo crescimento econômico do Brasil e pelo avanço do trabalho formal.

Já a regulação da área teve início somente em 1999, com a entrada em vigor do CONSU por meio da lei 9.656/98 que dispõe sobre os planos de saúde. No ano seguinte, a Agência Nacional de Saúde Suplementar (ANS) foi criada para atuar na regulamentação, criação e implementação de normas, controle e fiscalização das atividades do segmento.

Uma mudança que é percebida quando analisa a saúde suplementar no país é que os procedimentos que antes eram para uma população abrangente estão mudando para um público específico com a adição de medicamentos e procedimento para pessoas com doenças raras, por esse motivo é necessário que as operadoras busquem por análises e estudos para assegurar de riscos decorrentes dessas novas obrigações assumidas.

O Teste de Adequação do Passivo (TAP) foi estabelecido à princípio pela International Financial Reporting Standards (IFRS) em 2005. Porém, só em 2010 foi instituída a obrigação do TAP no Brasil pela SUSEP (Superintendência de Seguros Privados). Por fim, a ANS (Agência Nacional de Saúde Suplementar) em 2020, passou a exigir o TAP das Operadoras com mais de 100.000 (cem mil) beneficiários.

O TAP é um teste que possui o objetivo de analisar, por meio de estimativas correntes de fluxo de caixa futuro, se o passivo está adequado, sendo uma ferramenta importante para avaliar a saúde financeira de empresas que assumem riscos.

A Resolução Normativa 435 publicada em 2018 formulou os parâmetros mínimos deverão ser observados na elaboração do TAP, tais como: (a) para a realização do TAP, os contratos deverão ser segregados, no mínimo, entre as modalidades: (i) individual, (ii) coletiva empresarial, (iii) coletiva por adesão e (iv) corresponsabilidade assumida; (b) as estimativas correntes dos fluxos de caixa deverão ser apuradas considerando as vigências dos contratos, limitadas ao horizonte máximo de 8 (oito) anos; (c) para o cálculo das estimativas de sobrevivência e de morte deverão ser utilizadas as tábuas BR-EMS vigentes no momento da realização do TAP, ajustadas, quando for o caso, por critério de desenvolvimento de longevidade; (e) as estimativas correntes dos fluxos de caixa deverão ser descontadas a valor presente com base nas estruturas a termo da

taxa de juros (ETTJ) livre de risco pré-fixada definidas pela ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais).

Assim, para que uma operadora de Saúde Suplementar consiga administrar melhor a sua carteira de beneficiários é necessário ter um equilíbrio atuarialmente justo entre os preços dos planos de saúde e as despesas assistenciais, administrativas e comerciais. Isso significa que os preços dos planos devem ser suficientes para cobrir todas as despesas da operadora, mas também devem ser justos para os beneficiários. Logo, a operadora deve monitorar constantemente suas despesas e receitas para garantir que o equilíbrio seja mantido ao longo do tempo. Contudo, estimar esse equilíbrio pode ser um desafio, pois muitos fatores devem ser levados em consideração.

Os riscos e incertezas que envolvem eventos e circunstâncias, como mudanças na legislação ou na economia, devem ser considerados para chegar à melhor estimativa. Então, uma análise atuarial é fundamental para que haja um equilíbrio econômico-financeiro, uma gestão eficiente e uma análise cuidadosa dos dados.

A inflação é um dos eventos que pode mudar ao longo do tempo devido a vários fatores, como mudanças na política monetária ou na economia. Quando a curva de juros muda, isso afeta o valor das despesas assistenciais, aumentando os custos de medicamentos, insumos, etc., e também as despesas administrativas e comerciais, consequentemente, levando a uma variedade de preços futuros.

Além disso, a dinâmica demográfica também tem um impacto significativo nos preços das mensalidades. Pois, à medida que o processo de envelhecimento da população brasileira intensifica, a demanda por utilização do plano de saúde tende a aumentar significativamente, segundo o boletim Panorama – Saúde Suplementar divulgado pela ANS o mês de julho de 2023 registrou um recorde de consumidores em planos de assistência médica, alcançando 50,7 milhões de beneficiários, maior número desde novembro de 2014. Portanto, é importante levar em consideração tanto a curva de juros quanto a dinâmica demográfica ao tomar decisões de investimento. A Agência Nacional de Saúde Suplementar (ANS) é responsável por regular o setor de saúde suplementar no Brasil.

Ademais, o Teste de Adequação do Passivo (TAP) é uma ferramenta importante para avaliar a saúde financeira das operadoras de saúde suplementar. Assim, com o objetivo de estimar o TAP através de uma modelagem estocástica, que utiliza uma abordagem estatística para calcular essa provisão, verificar se essa abordagem é adequada para estimar o TAP no campo da saúde suplementar. Então, ajudar as

operadoras a gerenciar melhor seus riscos e assim garantir um equilíbrio atuarialmente justo.

Para estimar o TAP de maneira consistente, primeiramente, necessita-se de alguns pressupostos, pois a variação dessa estimativa abre caminho para uma quantificação das provisões de contingência. Como a estimativa de uma reserva de contingência adequada, sendo esta ligada ao custo adicional previsto do pior caso, num nível de confiança específico, projetado ao longo de um período contábilístico fixo. Também tem a complexidade numérica que será abordada através de uma metodologia de simulação, apropriada também no caso de um grande número de fatores de risco.

Assim, utilizaremos o Método de Monte Carlo para obter aproximações numéricas de do TAP, pela complexidade de obter uma solução analítica ou, pelo menos, determinística. Em princípio, métodos de Monte Carlo podem ser usados para resolver quaisquer problemas com um interpretação probabilística, pois, pela Lei dos Grandes Números, integrais descritas pelo valor esperado de alguma variável aleatória podem ser aproximadas obtendo a média empírica de amostras independentes de variáveis.

Além disso, para conseguir uma estimativa melhor sobre as receitas e despesas futuras é necessário estudar o comportamento da série temporal até o presente a fim de descobrir quais são suas principais características e realizar a modelagem mais adequada para os próximos anos. Para isso serão utilizados os modelos ARIMA (autoregressive integrated moving average), pois fornecem métodos mais sofisticados e versáteis para modelagem de tendência e componentes sazonais. Neste trabalho usaremos, para a estimação e previsão sobre as receitas e despesas futuras, o pacote do R forecast, desenvolvido por (Hyndman, R. J., & Khandakar, Y., 2008).

Procedimentos concernentes aos testes de adequação têm sido muito utilizados na literatura atuarial e para uma formulação mais formal do problema em (Cocoza, R. et.al., 2008). Em particular, como parte das metodologias implementadas pela Agência Nacional de Saúde (ANS), uma discussão das premissas e da metodologia pode ser encontrada em (Ribeiro, A., 2012). O estudo que pretendemos fazer corresponde à implementação via métodos Monte Carlo do TAP, onde a estrutura do termo da taxa de juros é estocástica.

A implementação via métodos de Monte Carlo é mais recente na literatura brasileira, tem alguns exemplos sobre TAP e sobre o uso de técnicas estocásticas para a determinação das provisões técnicas no site do (Nazareno Jr, 2019), outras técnicas Monte Carlo são discutidas no (Nazareno Jr, 2015). Como dito por (Korn R., et al., 2010),

os métodos de Monte Carlo são fundamentais em aplicações no setor financeiro e de seguros. Frequentemente, eles são a única ferramenta acessível para engenheiros financeiros e atuários quando se trata de cálculos complicados de preços ou riscos. Contudo, como eles tendem a ser lentos, é muito importante buscar estratégias que possam acelerá-los e também aumentar sua precisão.

Segundo a Wikipédia, no Brasil, a taxa Selic é a taxa básica de juros da economia. É o principal instrumento de política monetária utilizado pelo Banco Central do Brasil (BC) para controle da inflação. Essa taxa é referência para todas as demais taxas de juros do país, tais como aquelas incidentes sobre empréstimos, financiamentos e aplicações financeiras, por isso, será a taxa usada para se estimar o TAP neste estudo. A taxa Selic diária corresponde à taxa média ponderada e ajustada das operações de financiamento por um dia lastradas em títulos públicos federais, custodiados no Sistema Especial de Liquidação e de Custódia.

A base da taxa SELIC utilizada possui medições diárias e foram utilizados os dados de janeiro de 2011 a 2021 para fazer previsões, pois nos períodos anteriores a este a SELIC apresentava um comportamento distinto com variações grandes, além de uma taxa alta em comparação ao observado atualmente. Em seguida, os valores diários foram acumulados para ter a taxa mensal.

O uso de métodos estocásticos para a determinação estrutural do termo da taxa de juros usada na determinação do valor presente no TAP pode ser visto no trabalho de (Duarte et. al. 2015). Além disso, como observado em (Iacus, S. M., & Yoshida, N. 2018), equações diferenciais estocásticas são frequentemente usadas para modelar evoluções aleatórias ao longo do tempo contínuo ou praticamente contínuo, como os movimentos aleatórios de preços das ações e a dinâmica populacional. A teoria da inferência estatística para equações diferenciais estocásticas tem uma história bastante longa, a mais de três décadas atrás, e ainda hoje novas metodologias estão em desenvolvimento e expansão. Neste estudo, a referência principal no tratamento dos dados, estimação dos modelos e previsão para futuro das séries para juros é o trabalho de (Iacus, S. M., & Yoshida, N., 2018) e a implementação dele no pacote YUIMA do R.

Nesta monografia usaremos alguns métodos estatísticos para a estimação de um modelo estocástico para juros, destes o mais modelo mas popular é o modelo CIR (Cox–Ingersoll–Ross). O modelo CIR foi introduzido por Feller como um modelo para o crescimento populacional e tornou-se bastante popular nas finanças depois que Cox, Ingersoll e Ross propuseram para modelar taxas de juros. Para uma revisão deste

modelo ver (Cox, J. C., Ingersoll, J. E., & Ross, S. A., 1978), e para a parte inferencial deste modelo ver (Overbeck, L., & Ryden, T., 1997) onde tem uma revisão dos estimadores paramétricos. A implementação correspondente ao ambiente das livrarias do R e que utilizaremos neste trabalho é parte dos estimadores de quase-verosimilhança, cuja implementação faz parte do pacote YUIMA do R, as referências e guias de uso estão em (Iacus, S. M., & Yoshida, N. 2018).

Após essa introdução, no capítulo 2 será introduzido o cálculo do Teste de Adequação de Passivo (TAP) utilizando um modelo de solvência, além de medidas de riscos como o Valor em Risco (VaR) e a Cauda do Valor de Risco (TVaR), utilizando o modelo para o valor presente estocástico do fluxo de caixa o artigo de (Cocoza, R. et.al., 2008). As medidas de risco Var e TVaR são populares na determinação dos riscos nos estudos e na prática atuarial, para uma revisão delas ler o artigo de (Peng, J., 2009). Neste capítulo introduziremos o método de Monte Carlo que serve para achar os valores esperados de funções de variáveis aleatórias, assim como, o erro de aproximação produzido ao usar este método de aproximação, para um estudo desses métodos ver (A.H. Luna, 2019). Mostramos no caso do TAP, Var e TvaR o procedimento para aproximar estes valores usando o método de Monte Carlo.

No capítulo 3, será feita uma introdução às equações estocásticas, e como é o procedimento para estimar os parâmetros destes modelos, como aplicação dos modelos estocásticos que são utilizados para modelar a taxa de juros. Além disso, será mostrada a função de verosimilhança usado na inferência dos parâmetros do modelo CIR (Cox-Ingersoll-Ross) que foi o modelo de equação diferencial estocástica escolhido para esse estudo.

No capítulo 4 descreveremos os modelos de séries temporais tipo ARIMA que usaremos para fazer as previsões dos dados tratados no capítulo 2. Por meio dos modelos ARIMA, obtemos estimativas paramétricas e previsões para o número de beneficiários, os ticket médios e os custos per capita em cada tipo de contratação: coletivo empresarial, coletivo por adesão e Individual (familiar). Também, será abordado quais as bases de dados utilizadas, as manipulações que foram feitas e alguns pressupostos usados para seja possível estimar o TAP, ainda, será debatido sobre a saúde suplementar no Brasil, como surgiu, qual o órgão responsável por reger esse sistema e sobre a regulamentação prudencial do setor.

No capítulo 5 detalharemos também a modelagem e previsão da taxa SELIC por meio do modelo CIR e com isto a análise do TAP com o VaR e o TVaR introduzidos no

Capítulo 2. Por fim, analisaremos cenários de incertezas de maior impacto que podem acontecer no mercado de saúde suplementar e como isso é refletido no teste de adequação do passivo.

Finalmente no capítulo 6 foram colocadas as conclusões deste trabalho.

2. MODELO DE SOLVÊNCIA VIA TAP

2.1. TESTE DE ADEQUAÇÃO DO PASSIVO

Como é conhecido, um risco é definido como uma variável aleatória X de valor real não negativa definida em algum espaço de probabilidade e neste caso a solvência das operadoras é o evento (risco) que será estudado. Já a medida de risco é definida como um mapeamento do conjunto de variáveis aleatórias, ou seja, perdas ou pagamentos, ao conjunto dos números reais. Nas ciências atuariais, as medidas de risco são utilizadas para determinar provisões e requisitos de capital de uma instituição, sendo neste estudo usado para calcular as estimativas correntes de fluxo de caixa do TAP e conseqüentemente analisar a solvência da operadora de plano de saúde.

Para a realização do TAP serão estimados o fluxo de caixa futuro tipo de contratação, sendo elas: Coletivo empresarial, Coletivo por Adesão e Individual (ou Familiar). Os planos coletivos empresariais são aqueles que só é possível contratar com a intermediação de empresa empregadora, já os planos coletivos por adesão são contratados com a intermediação de associação ou sindicato e os planos individuais podem ser contratados por qualquer pessoa física.

Também serão analisados cenários de incertezas que possuem um impacto significativo no equilíbrio de um plano de saúde, como a elevação da taxa de juros e o aumento das despesas, em consequência dos custos com medicamentos, novas tecnologias hospitalares e incorporação de novas doenças ao rol de procedimentos.

Em cada um desses cenários, são projetados os tickets médios e os custos *per capita* de todos os tipos de contratação e um balanço é desenvolvido no final do horizonte finito. Dado que muitas linhas de produtos envolvem garantias e opções de investimento, a avaliação de passivos consistente com o mercado exige frequentemente a utilização de avaliações neutras ao risco. A receita necessária para garantir a solvência é determinada pelas despesas para esse cenário, descontado as contraprestações na avaliação da data, trazendo-os ao valor presente ao longo do horizonte finito.

Um das medidas de Risco mais populares são o Valor em Risco ou Value at Risk VaR e o Valor em Risco na Cauda ou Tail Value at Risk (TVaR), e assim como em Coccozza, R. et.al.,2008, vamos usar essas medidas para avaliar a solvência. Além disso, será estudado a aplicação do Valor de Risco da provisão matemática num contexto de avaliação justa para cada um dos eventos estudados. Já que, através do cálculo do VaR, é possível

conhecer a estimativa de uma reserva de contingência adequada que está ligada ao custo adicional previsto do pior caso e também a um nível de confiança específico, projetado durante um período contábil fixo. Cujo a fórmula é a seguinte:

$$VaR_{\alpha}(X) = \text{Inf} \left\{ x \in \frac{\mathbb{R}}{F_x}(x) \geq \alpha \right\}, \text{ com } \alpha \in (0, 1). \quad (1)$$

Onde X é uma variável aleatória da reserva e estamos interessados no quantil α , que neste trabalho será igual a 5%. Também será utilizado o Valor de Risco na Cauda (TVaR) pela importância na ciência atuarial e nas finanças como medida da cauda superior de uma função de distribuição. Sendo essa uma medida de risco associada ao valor em risco mais geral, mensurando o valor esperado da perda dado que ocorreu um evento fora de um determinado nível de probabilidade. Sua fórmula está informada a seguir:

$$TVaR_{\alpha}(X) = E \left[\frac{X}{X} \leq VaR_{\alpha}(X) \right] \quad (2)$$

Assim, dividindo a carteira de beneficiários em m subcarteiras homogêneas caracterizadas por aspectos comuns, isto é, por tipo de contratação, por os pagamentos, etc. Conforme visto em (Cocozza, R. et.al.,2008), se encontra a seguinte função:

$$V = E \left[\sum_{r=t+1}^n \sum_{i=1}^m [S_{i,r}(L_i - P_{i,r})] v(t, r) \right] \quad (3)$$

$S_{i,r}$ = número de beneficiários no momento r de cada i-ésimo tipo de contratação,

L_i = benefício mensal para cada i-ésimo tipo de contratação,

$P_{i,r}$ = prêmio pago no momento r para cada i-ésimo tipo de contratação,

$v(t, r)$ = taxa de desconto no momento t da taxa SELIC no momento r

Esta equação é enquadrada numa perspectiva futura, dentro de uma avaliação atual fornecida na tempo inicial 0. Percebe-se ainda que esta equação é semelhante às fórmulas de reservas matemáticas.

2.2. MÉTODOS DE SIMULAÇÃO DE MONTE CARLO

A aproximação de Monte Carlo pode ser descrita do seguinte modo. Suponha que seja fácil simular uma amostra de tamanho n , denotada por X_1, \dots, X_n , da variável aleatória X . Então, para uma função real h , a média amostral de $h(X)$ é

$$\hat{h}(X) = \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (4)$$

$\hat{h}(X)$ pela lei dos grandes números ela converge para $E[h(X)]$ quando $n \rightarrow \infty$. Isto significa que podemos aproximar o valor esperado por :

$$E[h(X)] \approx \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (5)$$

Quando $h(X)$ tem uma variância finita $Var(h(X))$, pelo teorema do limite central, o erro desta aproximação pode ser caracterizado por:

$$\frac{\hat{h}(X) - E[h(X)]}{\sqrt{\frac{Var(h(X))}{n}}} \approx N(0, 1) \quad (6)$$

O termo de variância $Var(h(X))$ pode ser aproximado pela variância da amostra S_{n-1} :

$$s_{n-1} = \frac{1}{n-1} \sum_{i=1}^n (h(X_i) - \hat{h}(X))^2 \quad (7)$$

deste modo é possível provar que o erro de aproximação segue uma distribuição t-Student com $n-1$ graus de liberdade, $t_{(n-1)}$ e a aproximação do erro fica:

$$\frac{\hat{h}(X) - E[h(X)]}{\sqrt{\frac{s_{n-1}}{n}}} \approx t_{(n-1)} \quad (8)$$

Juntando ao resultado da equação (5) e da equação (8) podemos resumir o método de Monte Carlo para aproximar o valor esperado de uma função h da variável aleatória X , $E[h(X)]$ nestes passos:

Passo (1): Simulamos uma amostra de n cópias da variável aleatória X : X_1, \dots, X_n

Passo (2): Aproximamos $E[h(X)] \approx \frac{1}{n} \sum_{i=1}^n h(X_i)$ para assim obter um estimador de $E[h(X)]$

Passo (3): Obtemos um intervalo de confiança Monte Carlo usando a equação (8):

$$IC(E[h(X)], (1 - \alpha) \times 100\%) = h^{\wedge}(X) \pm t_{(n-1), \frac{\alpha}{2}} \cdot \frac{s_{n-1}}{\sqrt{n}} \quad (9)$$

Agora, para aplicar este poderoso método ao nosso problema vamos a especificar os valores da função h para o $TVaR$ e V , vamos a fixar o valor do VaR e depois explicamos como encontramos ele a partir das simulações.

1) Simulamos N cópias da variável aleatória X definida abaixo:

$$X = \sum_{r=t+1}^n \sum_{i=1}^m [S_{i,r}(L_i - P_{i,r})]v(t, r) \quad (10)$$

Abaixo explicamos como vamos a simular cada termo desta expressão.

2) Observe que $TVaR_{\alpha}(X) = E\left[\frac{X}{X} \leq VaR_{\alpha}(X)\right]$, assim é suficiente considerar as simulações de X menores que o VaR e aplicar a elas o método de Monte Carlo com a função $h(x)=x$.

3) Para obter V , que é nosso principal objetivo, precisamos simplesmente fazer $h(x)=x$ e aplicar o método de Monte Carlo a todas as simulações.

4) O VaR é o $\alpha - quantil$ da distribuição da variável X , assim um estimador dele é simplesmente a média empírica dos $\alpha - quantis$ de subamostras simuladas da variável aleatória X , e aplicamos os método de Monte Carlo à esta média empírica para obter um intervalo de confiança para o $\alpha - quantil$ estimado.

A continuação vamos explicar como simular cada componente da variável aleatória X da equação (10).

3. MODELO ESTOCÁSTICO DE JUROS

3.1. MODELOS ESTOCÁSTICOS: EQUAÇÕES ESTOCÁSTICAS

As equações diferenciais estocásticas são processos estocásticos a tempo contínuo. Elas são usadas para modelar diversos fenômenos, como os sistemas físicos sujeitos a flutuações térmicas, nas áreas biomédicas devido à necessidade natural da teoria do cálculo estocástico e também para modelar os preços das ações na bolsa de valores.

Um processo estocástico também é definido como uma coleção de variáveis aleatórias que é indexada por algum conjunto matemático, o que significa que cada variável aleatória do processo estocástico está exclusivamente associada a um elemento no conjunto, para uma introdução ao assunto ver (Milanés, A ,2004).

Além disso, os processos estocásticos são classificados da seguinte forma quanto ao espaço de estados: se o espaço de estado são os inteiros ou números naturais, então o processo estocástico é chamado de processo estocástico discreto ou com valor inteiro. Se o espaço de estado é a linha real, o processo estocástico é referido como um processo estocástico de valor real ou um processo com espaço de estado contínuo. Processos estocásticos de tempo discreto são considerados mais fáceis de estudar porque processos de tempo contínuo requerem técnicas e conhecimentos matemáticos mais avançados, particularmente devido ao conjunto de índices ser incontável.

Uma equação diferencial estocástica (stochastic differential equation (SDE)) é uma equação diferencial na qual um ou mais dos termos é um processo estocástico, resultando em uma solução que também é um processo estocástico. Como dito anteriormente, os SDEs têm muitas aplicações no análise financeiro. Os SDEs têm uma evolução que que no caso mais básico da literatura é uma equação diferencial ordinária e que além disso tem um termo aditivo extra que é dependente de uma variável aleatória do tipo ruído branco chamada processo de Wiener, ver (Milanés, A ,2004).

3.2. PACOTE YUIMA: INFERÊNCIA PARAMÉTRICA E ESTIMAÇÃO DE QUASE-MÁXIMA VEROSSIMILHANÇA

O pacote yuima implementa várias técnicas para a estimação paramétrica, semi e não paramétrica de equações diferenciais estocásticas (multidimensionais), ver (Iacus, S. M., & Yoshida, N., 2018). A maioria das técnicas apresentadas abaixo se aplica a dados de

alta frequência, ou seja, quando Δ_n , o intervalo de tempo máximo entre duas observações consecutivas do processo, converge para zero à medida que o número n de observações aumenta.

Assim, considerando um processo de difusão ou equação diferencial estocástica geral:

$$dX_t = a(X_t, \theta_2)dt + b(X_t, \theta_1), \text{ com } X_0 = x_0 \quad (11)$$

Onde W_t é um processo de Wiener padrão (também chamado de processo de movimento browniano padrão ou movimento browniano) independente da variável inicial x_0 . Ainda, $\theta_1 \in \Theta_1 \subset \mathbb{R}$ e $\theta_2 \in \Theta_2 \subset \mathbb{R}$ e as funções $a: \mathbb{R} \times \Theta_2 \rightarrow \mathbb{R}$ e $b: \mathbb{R} \times \Theta_1 \rightarrow \mathbb{R} \times \mathbb{R}$. Para mais detalhes recomendo a leitura do artigo de (Milanés, A, 2004).

O estimador de quase-verossimilhança usado neste trabalho para estimar os parâmetros do modelo CIR utilizado para modelar a taxa selic é uma aproximação ao estimador de máxima verossimilhança para os parâmetros θ_1 e θ_2 do modelo CIR. Vamos descrever a função de quase-verossimilhança usada para achar estas estimativas, para mais detalhes veja (Iacus, S. M., & Yoshida, N., 2018). Considere os dados observados $\mathbf{X}_n = (X_{t_i})_{i=0, \dots, n}$, com $t_i = i\Delta_n$, $\Delta_n \rightarrow 0$ quando $n \rightarrow \infty$, o estimador de quase-máxima verossimilhança (QMLE):

$$l_n(X_n, \theta_1, \theta_2) = -\frac{1}{2} \sum_{i=1}^n \log(\det(\Sigma_{i-1}(\theta_1))) + \frac{1}{\Delta_n} \Sigma_{i-1}^{-1}(\theta_1) [(\Delta X_i - \Delta_n a_{i-1}(\theta_2))^2] \quad (12)$$

Onde $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $\Sigma_i(\theta_1) = \Sigma(\theta_1, X_{t_i})$, $a_i(\theta_2) = a(X_{t_i}, \theta_2)$ e $\Sigma = b \times b^t$

Então o QMLE de $\theta = (\theta_1, \theta_2)$ é um estimador que satisfaz $\hat{\theta} = \text{argmax}_{\theta} (l_n, \theta)$ (13)

aproximadamente.

3.3. INFERÊNCIA SOBRE EQUAÇÃO ESTOCÁSTICA: MODELO COX-INGERSOLL-ROSS

Em finanças matemáticas, o modelo de Cox-Ingersoll-Ross (CIR) descreve a evolução das taxas de juros. É um tipo de "modelo de um fator", pois descreve os movimentos das taxas de juros como impulsionados por apenas uma fonte de risco de mercado. Assim, o modelo pode ser utilizado na avaliação de derivativos de taxa de juros.

O modelo CIR descreve a taxa de juros instantânea r_t , cuja equação diferencial estocástica é $dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}$

onde:

r_t = Taxa de juros instantânea no instante t,

a = Taxa de reversão média,

b = Média da taxa de juros,

W_t = Processo Wiener (variável aleatória modelando o fator de risco do mercado)

σ = Desvio padrão da taxa de juros (medida de volatilidade)

O fator de desvio padrão, $\sigma\sqrt{r_t}$, com a raiz quadrada do processo evita a possibilidade de taxas de juros negativas para todos os valores positivos de a e b. e quando a taxa r_t é próxima de zero, o desvio padrão $\sigma\sqrt{r_t}$ também torna-se muito pequeno, o que diminui o efeito do choque aleatório sobre a taxa. Consequentemente, quando a taxa se aproxima de zero, sua evolução torna-se influenciada pelo fator de deslocamento, que empurra a taxa para cima (em direção ao equilíbrio).

A seguir será apresentado alguns gráficos simulando trajetórias do processo CIR para entender melhor o papel dos parâmetros no processo CIR:

Figura 1: Exemplo de séries estimadas utilizando o processo CIR com $a=1$, $b=1$ e $\sigma=1$

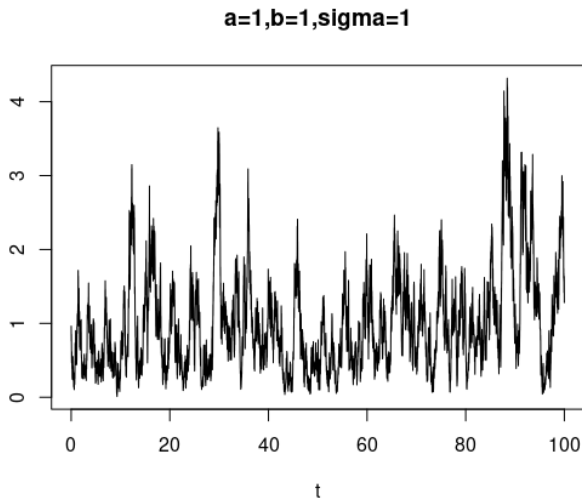
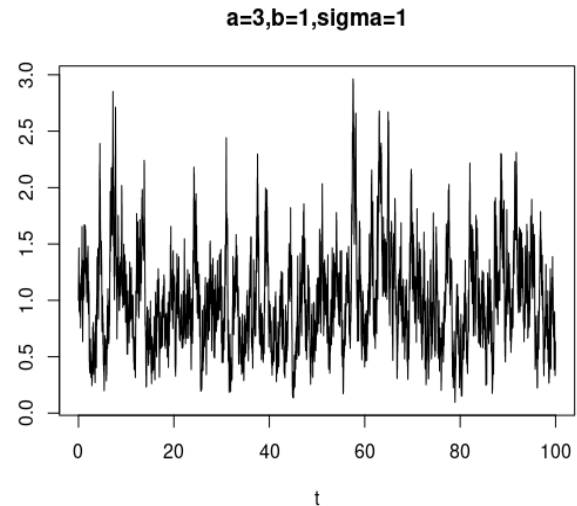


Figura 2: Exemplo de séries estimadas utilizando o processo CIR com $a=3$, $b=1$ e $\sigma=1$



Nestes exemplos temos que $b = 1$ é a média global do processo. O processo acompanha esta média b com uma força representada pelo parâmetro a assim na figura da esquerda $a=1$ e mais fraca do que $a=3$, neste último caso a força a faz com que o processo fique próximo da média global. Em geral, o processo CIR acompanha a média local e tende a não afastar-se da média global b . O parâmetro σ é o desvio global, ele é modificado pela escala $\sqrt{r_t}$, assim, quando r_t fica próximo de zero isso faz com que a variação do processo se aproxime de zero e quando r_t se afasta do zero, então as variações também se incrementarão.

No tópico anterior, foi verificado que todos parâmetros do modelo CIR que iremos usar foram estimados a partir dos dados da série temporal da taxa Selic, para isto usamos o estimador de máxima quase-verossimilhança para equações estocásticas que foi apresentado acima. Vejamos isso:

Vamos a trabalhar com a seguinte parametrização do CIR,

$dr_t = (\theta_1 - \theta_2 \cdot r_t)dt + \theta_3 \sqrt{r_t} dW_t$, assim $a = \theta_2$, $b = \frac{\theta_1}{\theta_2}$ e $\sigma = \theta_3$, para colocar esta expressão como uma difusão do tipo da equação (11) fazemos $\theta_2 = (\theta_1, \theta_2)$ e $\theta_1 = \theta_3$, assim:

$$1) a(r_{t_i}, \Theta_2) = \theta_1 - \theta_2 \cdot r_{t_i} e,$$

$$2) b(r_{t_i}, \theta_1) = \theta_3 \cdot \sqrt{r_{t_i}}$$

logo,

$$3) \Sigma_i = b \cdot b^t = \theta_3^2 \cdot r_{t_i}.$$

Finalmente, a verosimilhança (12) pode ser escrita como:

$$l_n(r_t, \theta_1, \theta_2, \theta_3) = -\frac{1}{2} \sum_{i=1}^n \log(\theta_3^2 \cdot r_{t_i}) + \frac{1}{t_i - t_{i-1}} (\theta_3^2 \cdot r_{t_{i-1}})^{-1} \left[\left((r_{t_i} - r_{t_{i-1}}) - (t_i - t_{i-1})(\theta_1 - \theta_2) \cdot r_{t_i} \right)^2 \right] \quad (14)$$

Poderíamos obter numericamente os valores de $\theta_1, \theta_2, \theta_3$ que maximizam a equação (14) mas o que vamos fazer é usar o pacote YUIMA que faz isso e também simula um modelo CIR com os parâmetros estimados.

4. MODELOS PARA O NÚMERO DE BENEFICIÁRIOS, TICKET MÉDIO E CUSTO PER CAPITA

4.1. SÉRIES TEMPORAIS E MODELOS ARIMA

De acordo com a Wikipédia, uma série temporal é uma série de pontos de dados indexados (ou listados ou grafados) em ordem temporal. Mais comumente, uma série temporal é uma sequência tomada em pontos sucessivos igualmente espaçados no tempo. Trata-se, portanto, de uma sequência de dados de tempo discreto. Além disso, existem duas formas de estudar séries temporais. Uma análise da série temporal é um método para tentar entender a série temporal, de forma a entender a estrutura que gerou a série. Uma previsão a partir da série temporal procura construir um modelo matemático a partir do qual seja possível prever valores futuros da série. Os modelos para estudar as séries temporais são muito conhecidos por seus acrônimos em inglês, montados a partir de AR (modelos auto-regressivos), 'I' (modelos integrados) e MA (modelos de média móvel). Por exemplo, o modelo ARIMA é um modelo auto-regressivo, integrado e de média móvel.

Os modelos ARIMA (autoregressive integrated moving average) são modelos ajustados aos valores da série temporal para entender melhor os dados ou para prever pontos futuros na série. Modelos ARIMA são aplicados em alguns casos em que os dados mostram evidências de não estacionariedade.

A parte auto-regressiva (AR) do modelo ARIMA indica que a variável de interesse é

regressada em seus próprios valores anteriores. Todo modelo AR é invertível e isto indica que os modelos autorregressivos são notavelmente flexíveis para lidar com uma ampla gama de diferentes padrões de séries temporais.

A parte de média móvel (MA) indica que o erro de regressão é na verdade uma combinação linear dos termos de erro, cujos valores ocorreram em vários momentos no passado. Ainda, toda série MA é estacionária e em geral, uma série temporal estacionária não terá padrões previsíveis no longo prazo. Os gráficos de tempo irão mostrar que uma série Média Móvel é aproximadamente horizontal (embora seja possível algum comportamento cíclico), com variância constante.

Modelos ARIMA não sazonais são geralmente denotados como ARIMA(p, d, q) em que os parâmetros p, d, q são números inteiros não negativos, p é a ordem (número de defasagens) do modelo auto-regressivo, d é o grau de diferenciação (o número de vezes em que os dados tiveram valores passados subtraídos) e q é a ordem do modelo de média móvel. Modelos ARIMA sazonais são geralmente denotados como ARIMA(p,d, q)(P, D, Q)_m em que m se refere ao número de períodos em cada temporada e P, D e Q se referem aos termos de auto-regressão, diferenciação e média móvel para a parte sazonal do modelo ARIMA.

Ou seja, Dada uma série temporal de dados X_t , em que t é um índice representado por um número inteiro e X_t são números reais, um modelo ARMA(p,q) é dado por:

$$X_t = \alpha_1 X_{t-1} - \dots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

ou equivalentemente por $(1 - \sum_{i=1}^{p'} \alpha_i L^i) X_t = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t$, em que L é operador de defasagem, α_i são os parâmetros ε_t são os termos de erro. Os termos de erro ε_t são geralmente assumidos como variáveis independentes e identicamente distribuídas amostradas a partir de uma distribuição normal com média zero.

Assumindo que o polinômio $(1 - \sum_{i=1}^{p'} \alpha_i L^i)$ tem uma raiz unitária (um fator $(1 - L)$) de multiplicidade d. Então um processo ARIMA(p,q,d) expressa a propriedade de fatoração deste polinômio com $p = p' - d$ pode ser escrito como:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t \quad (15)$$

Em relação ao AIC para determinar a ordem de um modelo ARIMA não sazonal, um critério útil é o critério de informação de Akaike (AIC) é escrito como:

$$AIC = -2 \log(L) + 2(p + q + k + 1),$$

em que L é verossimilhança dos dados, p é a ordem da parte auto-regressiva e q é a ordem da parte de média móvel. O parâmetro k neste critério é definido como o número de parâmetros no modelo sendo ajustado aos dados.

O objetivo é o de minimizar os valores de AIC para um bom modelo, pois o AIC tenta aproximar modelos da realidade da situação. Quanto menor o valor de um destes critérios para uma gama de modelos investigados, melhor o modelo se adequará aos dados.

Como dito anteriormente, o modelo ARIMA pode ser visto como uma ligação de dois modelos. O primeiro é não estacionário: $Y_t = (1 - L)^d X_t$, enquanto o segundo é estacionário em sentido amplo: $(1 - \sum_{i=1}^p \phi_i L^i) = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t$.

Os intervalos de previsão (intervalos de confiança para previsões) para modelos ARIMA possuem a seguinte fórmula:

$$Y_{T+h|T}^{\wedge} \pm 1.96 \sqrt{v_{T+h|T}}, \text{ em que } v_{T+h|T} \text{ é a variância de } y_{T+h} | y_1, \dots, y_T.$$

Para $h = 1$, $v_{T+h|T} = \sigma^2$ para todos os modelos ARIMA, independentemente dos parâmetros e das ordens.

Caso seja um modelo Média Móvel, ou seja, um modelo ARIMA(0,0,q),

$$y_t = e_t + \sum_{i=1}^q \theta_i e_{t-i}, \text{ temos então: } v_{T+h|T} = \sigma^2 \left[1 + \sum_{i=1}^{h-1} \theta_i^2 \right], \text{ para } h = 2, 3, \dots$$

Em geral, intervalos de previsão de modelos ARIMA aumentarão conforme o horizonte de previsão aumenta. Para automatizar o cálculo de previsões dos modelos ARIMA vamos usar o pacote forecast do R, para mais detalhes sobre o modelo ARIMA e o pacote forecast veja (Hyndman, R. J., & Khandakar, Y., 2008).

4.2. ANÁLISE E ESTUDOS DA BASE DE BENEFICIÁRIOS

Utilizando a base de beneficiários de uma operadora de plano de saúde A disponível no ANS TabNet foi obtido o número de pessoas por cada tipo de contratação, sendo eles, Individual ou Familiar, Coletivo Empresarial e Coletivo por adesão. Tais informações são divulgadas de forma trimestral, assim foi necessário fazer uma interpolação uniforme para conseguir os número de beneficiários por mês.

Para ter uma quantidade de dados consistentes foram utilizados os dados de junho de 2011 até o último levantamento, que é o mês de junho de 2023. Assim, a previsão dos beneficiários para os próximos sete anos foi feita usando os modelos ARIMA (autoregressive integrated moving average), sendo escolhido como melhor previsão aqueles em que todos os parâmetros do modelo foram significativos ao nível de 95% de confiança e tiveram o menor AIC (critério de informação de Akaike), que é uma métrica que mensura a qualidade de um modelo estatístico visando também a sua simplicidade. Os modelos ARIMA foram usados, pois, fornece uma estimativa melhor dos parâmetros, além de ser possível desenvolver diferentes cenários apenas alterando os valores do modelo criado.

4.3. ANÁLISE E ESTUDOS DA BASE DE DESPESAS E RECEITAS ASSISTENCIAIS

A base de receitas e despesas utilizada foi baixada pelo portal de dados abertos e são divulgadas de forma anual, para ter a estimativa mensal dos valores considerou-se que as despesas e receitas são normalmente distribuídos, logo o valor mensal médio seria o valor anual dividido pelos 12 meses com uma variação de 20% em torno deste valor médio. Além disso, para o estudo foram usados os dados do ano de 2012 até o ano de 2022.

Para calcular o ticket médio usou-se as receitas assistenciais mensais informadas anteriormente e em seguida o valor foi dividido pela quantidade de beneficiários no respectivo mês. O mesmo processo foi aplicado para calcular o custo per capita mensal, sendo o valor da despesa mensal dividido pelo número de beneficiários em cada mês.

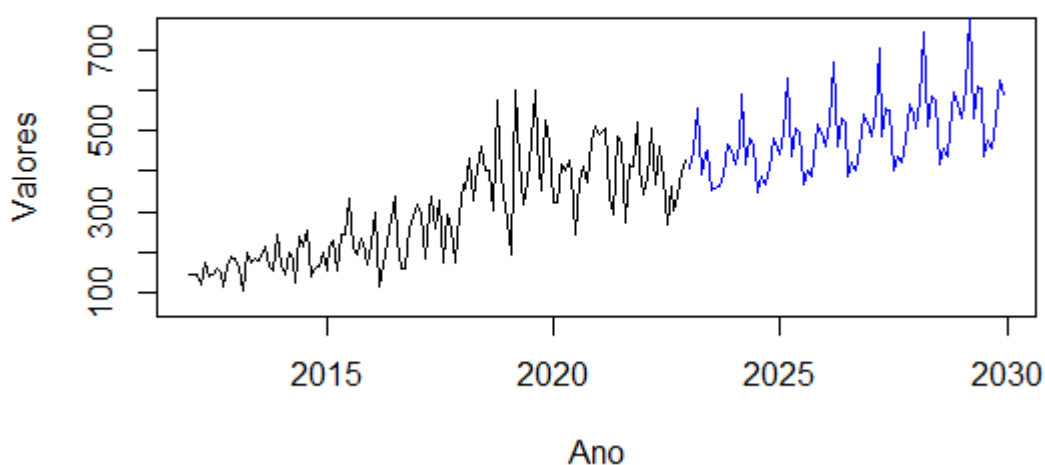
A previsão dos ticket médio e do custo per capita para os próximos sete anos foram feitas usando os modelos ARIMA (autoregressive integrated moving average), sendo escolhido como melhor previsão aqueles os que tinham uma variação menor do intervalo de 95% de confiança. Os modelos ARIMA foram usados pois fornece uma estimativa melhor dos parâmetros, além de ser possível desenvolver diferentes cenários apenas alterando os valores do modelo criado.

5. MODELAGEM DO TAP PARA OPERADORAS DE SAÚDE

5.1. DESCRIÇÕES DOS MODELOS

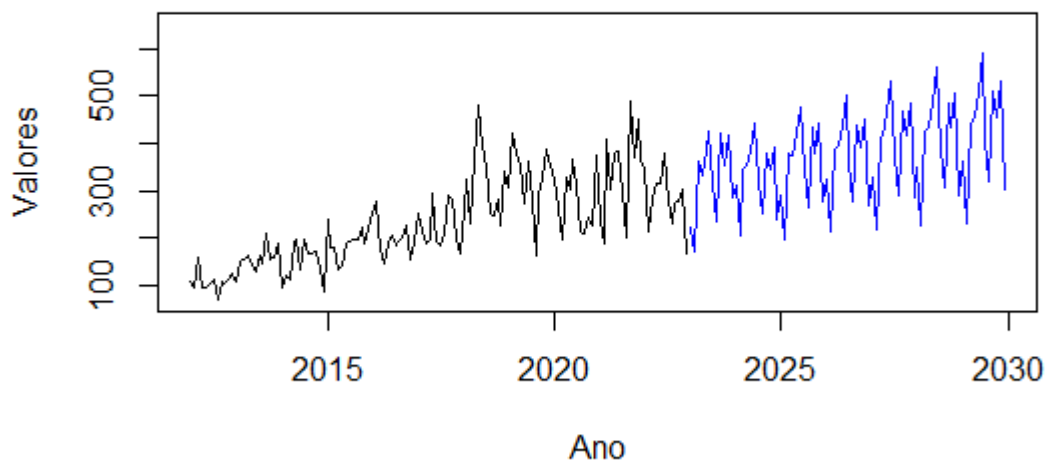
Como informado nos capítulos anteriores, para fazer as previsões dos tickets médios, custos per capitas e do número de beneficiários foram utilizados os modelos ARIMA. Abaixo estão os gráficos dos modelos de previsões com as melhores previsões encontradas:

Figura 3: Previsão Esperada do Ticket Médio de planos Coletivos Empresariais com os estimadores $ARIMA(p=1, d=0, q=0)(P=1, D=2, Q=1)_{12}$



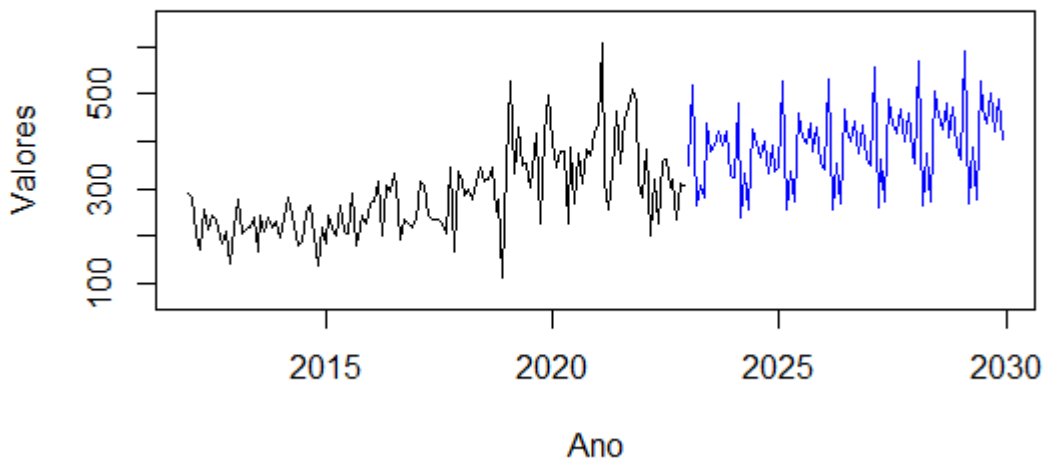
Na Figura 3 percebe-se além da tendência de crescimento, picos sazonais nas previsões estimadas com um pequeno aumento na amplitude da sazonalidade ao longo do tempo.

Figura 4: Previsão Esperada do Custo per Capita de planos Coletivos Empresariais com os estimadores $ARIMA(p=1, d=0, q=0)(P=1, D=2, Q=1)_{12}$



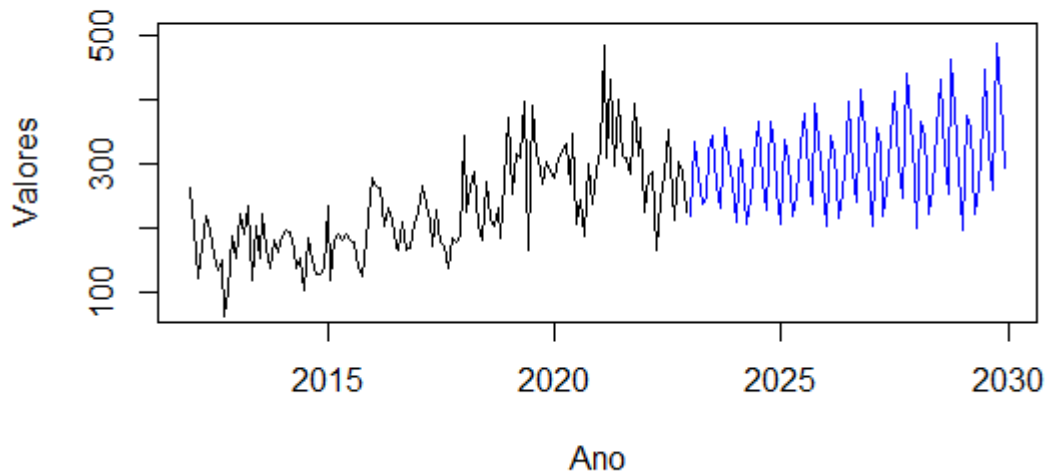
Na Figura 4 observa-se também uma tendência de crescimento, e picos sazonais nas previsões estimadas, além disso, assim como na Figura 1 ocorre um aumento no ruído da sazonalidade ao longo do tempo.

Figura 5: Previsão Esperada do Ticket Médio de planos Coletivos por Adesão com os estimadores $ARIMA(p=1, d=0, q=0)(P=1, D=2, Q=1)_{12}$



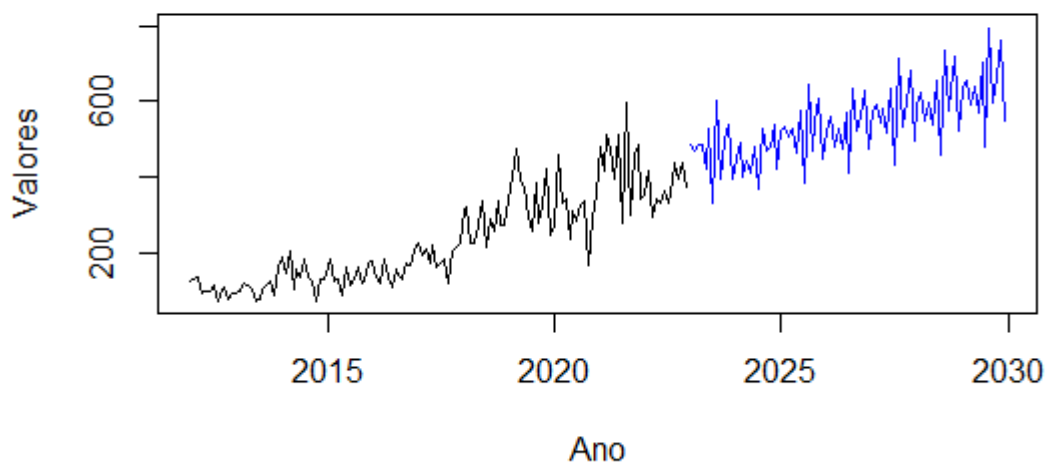
Na Figura 5 observa-se que aproximadamente entre 2019 e 2022 que a série apresenta um comportamento heterogêneo com quedas e aumento significativos nos valores. Na previsão estimada possui picos sazonais e que apresentam sazonalidade multiplicativa de aumento e também uma tendência crescente ao longo do tempo.

Figura 6: Previsão Esperada do Custo per Capita dos plano Coletivos por Adesão com os estimadores $ARIMA(p=1, d=0, q=1)(P=1, D=2, Q=1)_{12}$



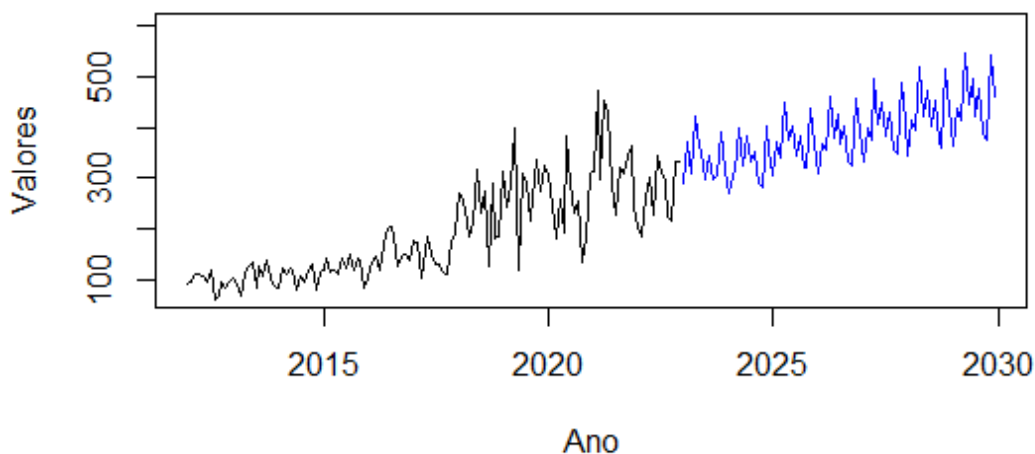
Na figura acima a série não aparenta ter sazonalidade e possui uma tendência crescente, mesmo que nos últimos meses observados apresenta um decréscimo rápido na tendência. Estimando a previsão, o comportamento de série mostrou sazonalidade crescente ao longo do tempo e a tendência de crescimento.

Figura 7: Previsão Esperada do Ticket Médio de planos Individuais (Familiares) com os estimadores $ARIMA(p=1, d=0, q=0)(P=1, D=2, Q=1)_{12}$



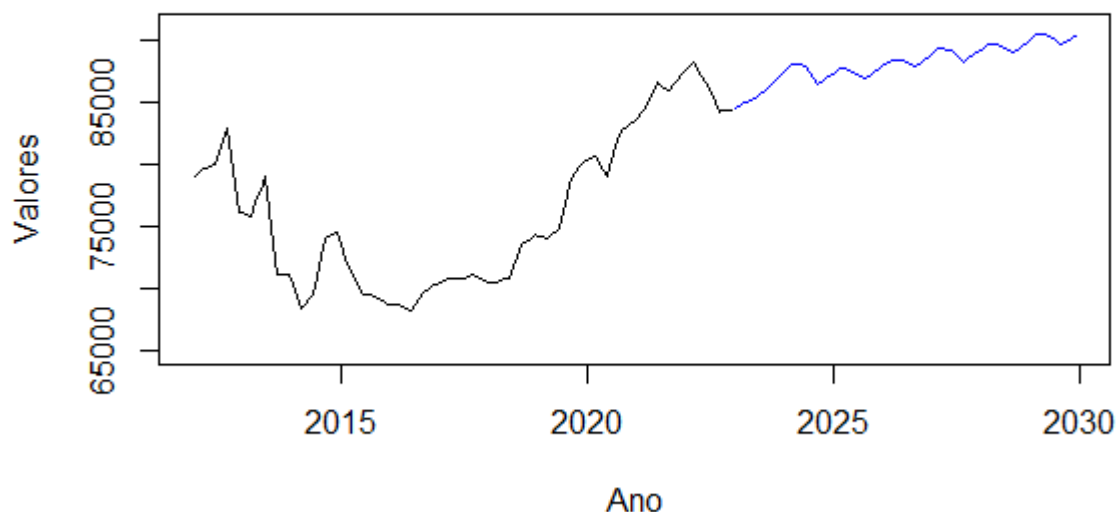
No gráfico acima percebe-se uma variação menor em comparação com os demais gráficos e além da tendência de crescimento, tem a ocorrência de sazonalidade nas previsões estimadas com um pequeno aumento na magnitude da sazonalidade ao longo do tempo.

Figura 8: Previsão Esperada do Custo per Capita de planos Individuais (Familiares) com os estimadores $ARIMA(p=1, d=0, q=0)(P=1, D=2, Q=1)_{12}$



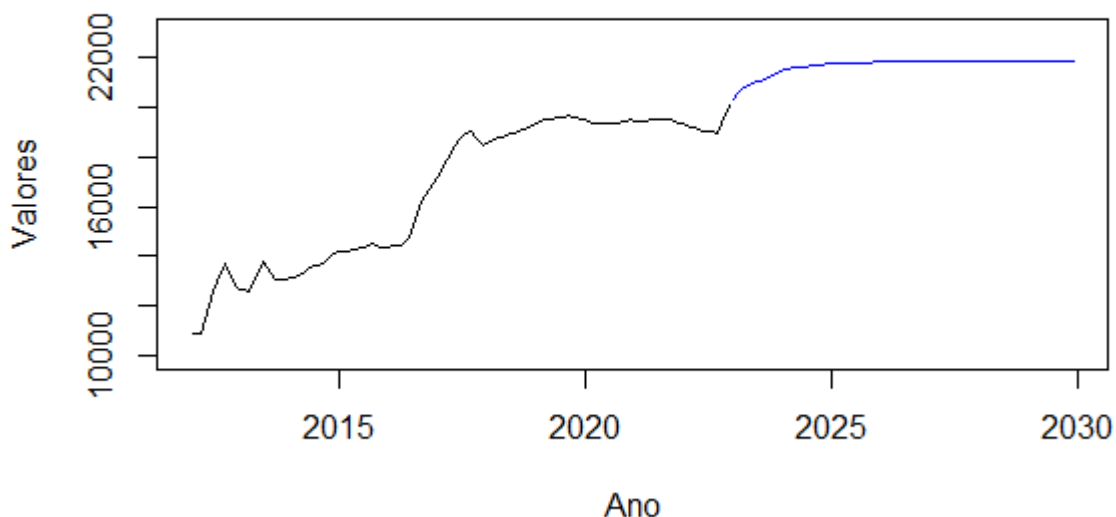
Na Figura 8 observa-se também uma tendência de crescimento, e picos sazonais nas previsões estimadas e como mostrado nos gráficos anteriores, ocorre um aumento na magnitude da sazonalidade ao longo do tempo, ainda que nesse gráfico seja menor em comparação aos demais.

Figura 9: Previsão Esperada de beneficiários nos planos Coletivos Empresariais com os estimadores $ARIMA(p=1, d=1, q=0)(P=2, D=1, Q=0)_{12}$



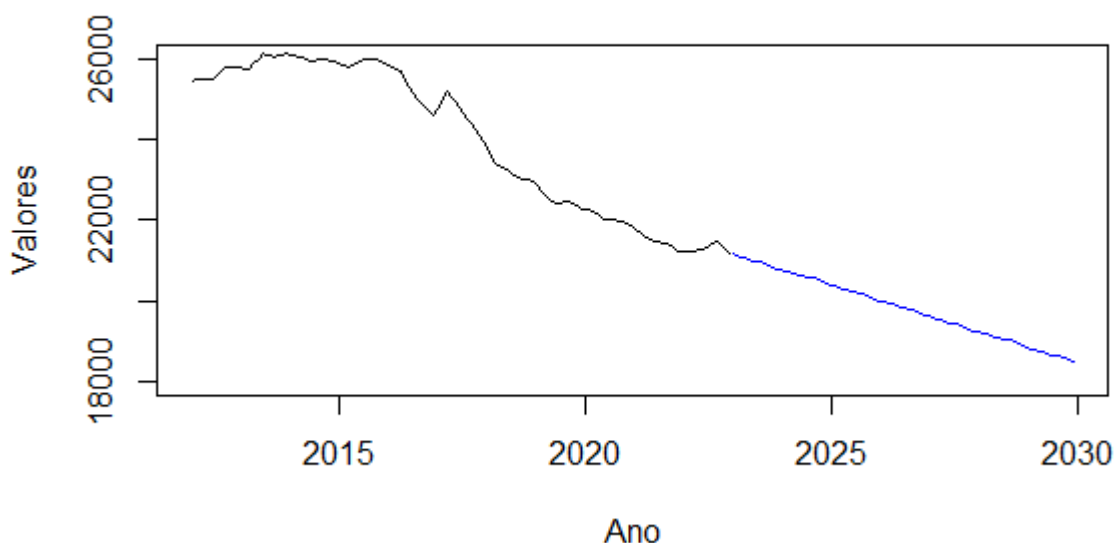
O número de beneficiários da operadora escolhida para o estudo apresenta no geral uma decrescimento de 2011 ao final de 2015, depois ocorre um aumento entre o período de 2016 a 2022 tendo uma queda em seguida no começo de 2023. Assim, na previsão que foi obtida não tem um cenário de grande crescimento.

Figura 10: Previsão Esperada de beneficiários nos planos Coletivos por Adesão com os estimadores $ARIMA(p=1, d=1, q=0)(P=1, D=0, Q=0)_{12}$



Nos planos coletivos por adesão ocorreu um aumento significativo em 2011 e 2018, o número de beneficiários quase que dobrando durante esse período. Em seguida, ocorreu uma certa estabilidade na carteira, voltando a crescer novamente no começo de 2023. Na previsão estimada percebemos um pequeno crescimento até 2025 e em seguida a série temporal prevista mantém uma estabilidade sem crescimento ou decrescimento.

Figura 11: Previsão Esperada de beneficiários nos planos Individuais (Familiares) com os estimadores $ARIMA(p=0, d=1, q=2)(P=0, D=1, Q=1)$ ¹²



Ao contrário dos demais tipos de contratação, os planos individuais apresentam queda no número de beneficiários. Essa situação também está ocorrendo em outras operadoras de plano de saúde, conforme mostrado em (Agência Nacional de Saúde Suplementar, 2023). Na série prevista, a tendência é decrescente e não possui sazonalidade.

5.2. PACOTE YUIMA: MODELANDO A TAXA DE JUROS

Com relação à SELIC, a previsão foi realizada utilizando o pacote `yuimaGUI` do R. O Projeto YUIMA foi inicialmente iniciado por Nakahiro Yoshida como investigador principal e depois Stefano M. Iacus juntou-se como coordenador da implementação do R, para uma pesquisa mais abrangente do pacote ver (Iacus, S. M., & Yoshida, N. (2018)). O Projeto YUIMA foi desenvolvido principalmente por estatísticos que publicam ativamente na área de inferência e simulação para equações diferenciais estocásticas. O pacote `yuimaGUI` é uma

interface para yuima que simplifica tarefas como estimativa e simulação de processos estocásticos, incluindo ferramentas adicionais relacionadas a finanças quantitativas, como recuperação de dados de preços de ações e indicadores econômicos, agrupamento de séries temporais, análise de pontos de mudança, etc.

Para modelar os dados da taxa SELIC utilizando o processo CIR introduzido anteriormente lembrando que vamos a trabalhar com a seguinte parametrização do modelo CIR:

$$(7) \quad dr_t = (\theta_1 - \theta_2 \cdot r_t)dt + \theta_3 \sqrt{r_t} dW_t$$

assim $a = \theta_1$, $b = \frac{\theta_1}{\theta_2}$ e $\sigma = \theta_3$, usando o pacote yuima do R obteve-se os seguintes estimadores para o processo CIR quando usamos a série da taxa Selic:

Tabela 1: Parâmetros estimados

	θ_1	θ_2	θ_3
Estimativas	0	0.007	0.00793
Erro Padrão	2e-04	0.008	9e-05

Em seguida, esses parâmetros estimados foram utilizados para simular a previsão nos próximos 7 anos (84 meses). Assim, fizemos N=20 destas simulações, um exemplo destas simulações está informado no gráfico abaixo:

Figura 12: Simulações prevista da taxa SELIC feitas no pacote YuimaGUI

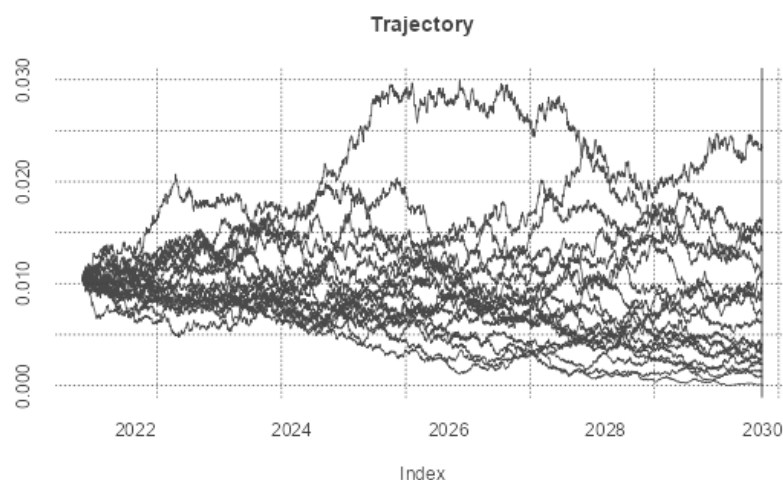
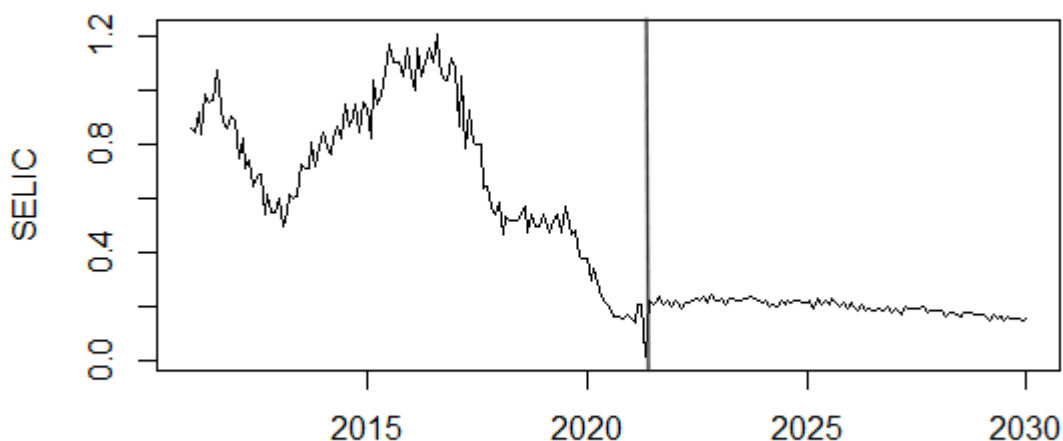


Figura 13: Taxa SELIC observada e previsão média da Taxa estimada ao longo do tempo



5.3. EXPLORANDO CENÁRIOS

Neste tópico serão analisados cenários de incertezas que possuem um impacto significativo no equilíbrio de um plano de saúde. Assim, foi estudado a elevação do custo caso a despesa for igual ao limite superior de 90% de confiança para os diferentes tipos de contratações e a também a elevação da taxa de juros.

Em ambos os três próximos gráficos percebe-se que a distância entre a previsão média estimada (série em azul) e a série do limite de 90% de confiança (série em vermelho) aumenta ao longo do tempo de previsão. O que demonstra que quanto maior o intervalo de tempo de previsão, maior também será o intervalo de confiança.

Figura 14: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Coletivos Empresariais ARIMA($p=1, d=0, q=0$)($P=1, D=2, Q=1$)¹²

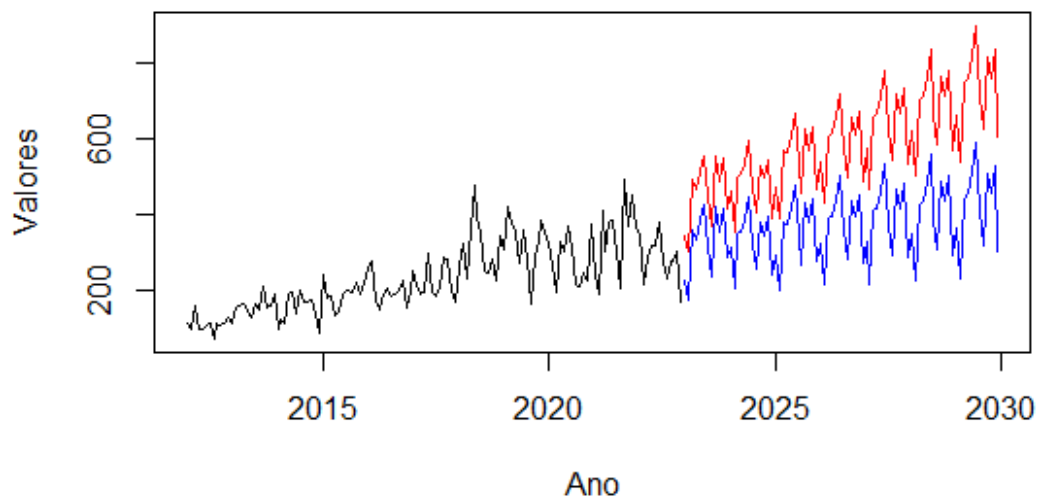


Figura 15: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Coletivos por Adesão ARIMA($p=1, d=0, q=1$)($P=1, D=2, Q=1$)¹²

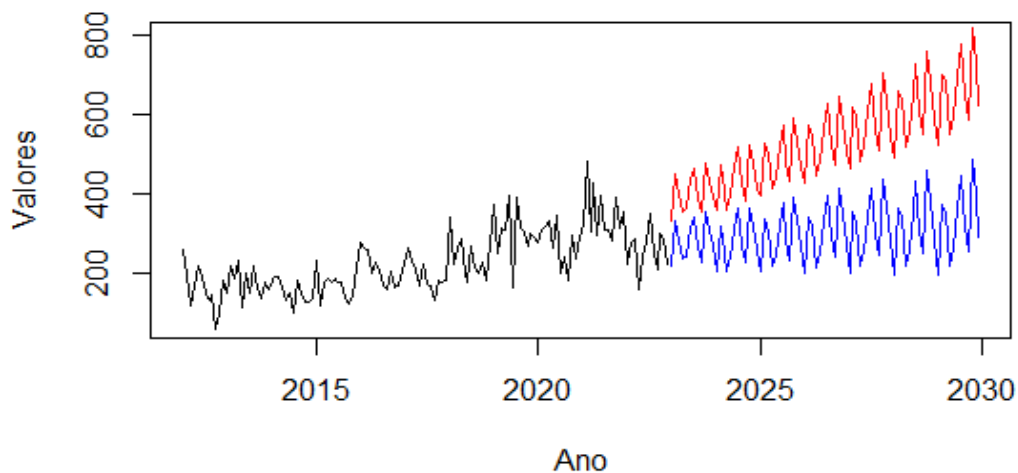


Figura 16: Limite superior de previsão e previsão média estimada do Custo per Capita de planos Individuais ARIMA($p=1, d=0, q=0$)($P=1, D=2, Q=1$)¹²

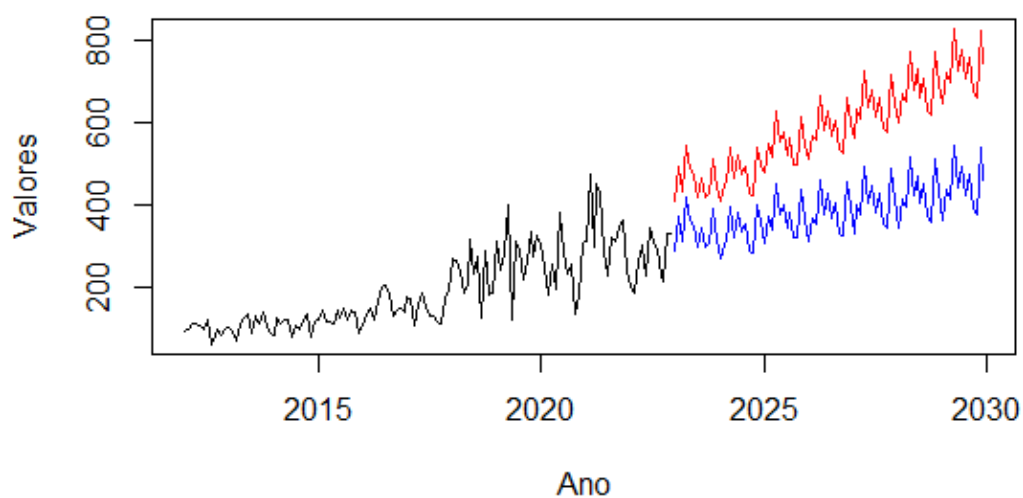
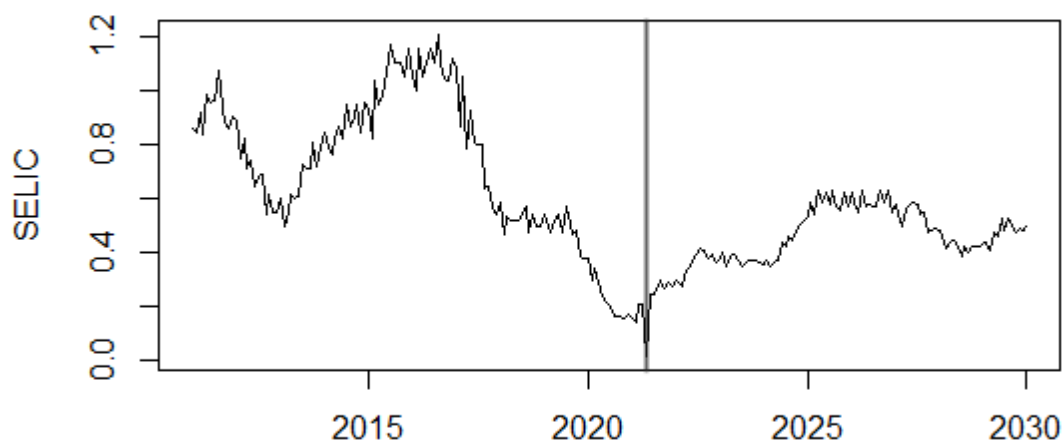


Figura 17: Taxa SELIC observada e previsão máxima da Taxa estimada ao longo do tempo



5.4. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA PELO TAP

Para verificar a solvência da operadora escolhida foi utilizada a fórmula (3) introduzida no capítulo 2. Assim, numa perspectiva de avaliação de solvência, os modelos que envolvem a chamada reserva de quantil desempenham um papel fundamental devido às suas ligações específicas ao Value at Risk. Logo, para obter os quantis foram feitas 5000 simulações usando o Método de monte Carlo sob a suposição que os tickets médios, custos per capita, total de beneficiários para cada tipo de contratação e o juros ao longo do período de 84 meses (7 anos) são uniformemente distribuídos. Desse modo, o valor do TAP (indicado pela letra V na fórmula) pode ser tratado como uma amostra de uma distribuição normal segundo o Teorema Central do Limite, sendo usada para estimar o VaR e o TVaR por meio do pacote “actuar” no R.

Figura 18: Histograma do TAP

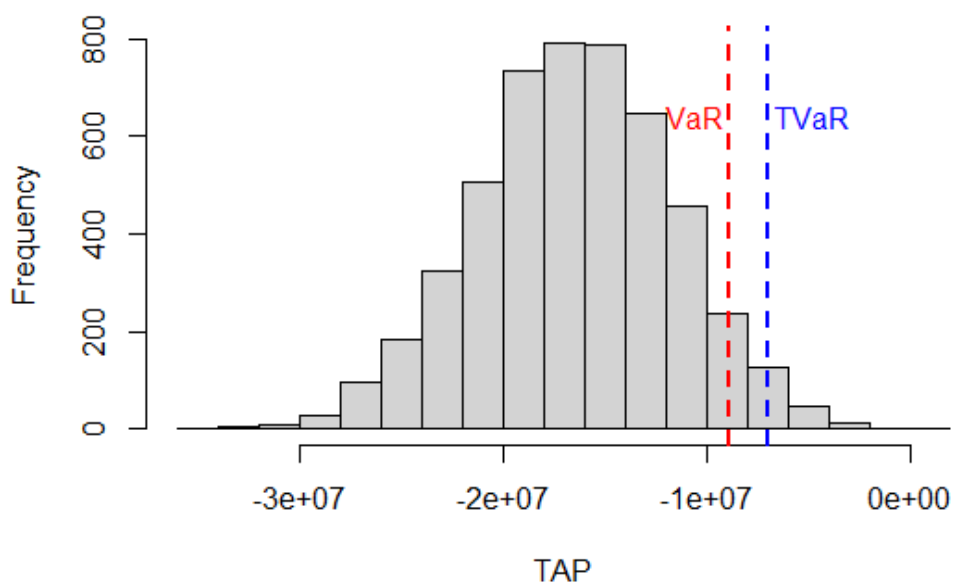


Tabela 2: Intervalo de confiança de 95% para o VaR_{α} e $TVaR_{\alpha}$ com $\alpha = 0,05$.

Medidas de Risco	Média	Limite Inferior	Limite Superior
VaR	-8.876.713,60	-9.124.685,96	-8.628.741,25
TVaR	-6.958.856,22	-7.200.482,24	-6.717.230,20

Percebe-se que a operadora escolhida apresenta solvência quando se estima o TAP de maneira estocástica, pois toda a reserva de quantil está abaixo de zero, demonstrando que não ocorreu nenhum cenário de ruína observado.

5.5. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA AGRAVANDO O CUSTO

Em uma situação onde ocorra o pior cenário possível ao nível de 95% de confiança, ou seja, igual ao limite superior de probabilidade foi obtido a reserva quantílica a seguir:

Figura 19: Histograma do TAP com custo per capita agravado

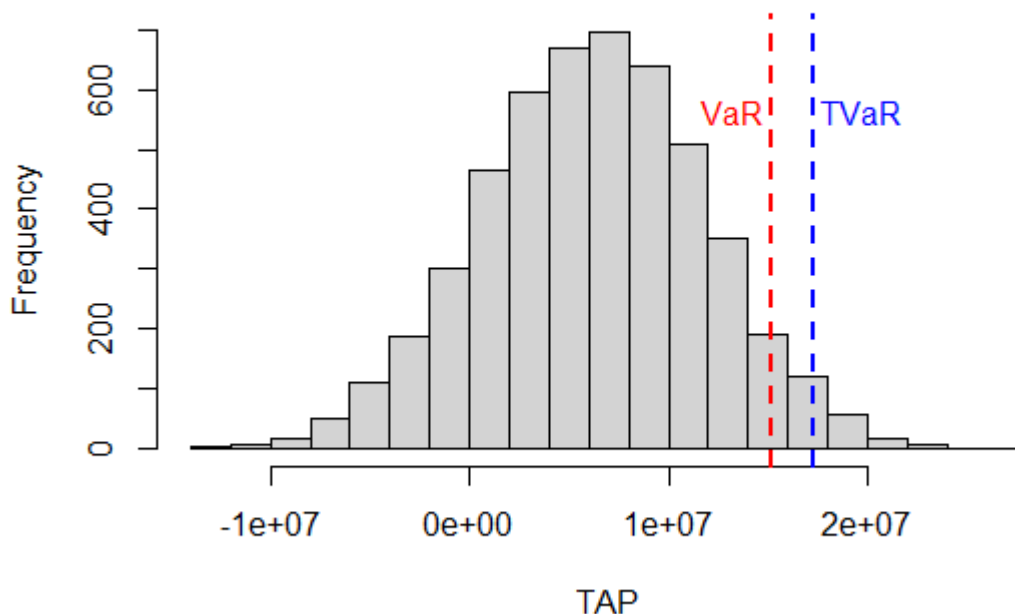


Tabela 3: Intervalo de confiança de 95% para o VaR e TVaR do TAP com $\alpha = 0,05$ e com custo per capita agravado

Medidas de Risco	Média	Limite Inferior	Limite Superior
VaR	15.131.083,41	14.843.807,54	15.418.359,27
TVaR	17.279.671,95	16.978.903,75	17.580.440,15

Nota-se nessa situação que a operadora apresentou insolvência, pois utilizando o VaR e o TVaR como medida de risco, ambos possuem valores distantes de zero demonstrando que ruína ocorreu.

5.6. OBSERVAÇÃO DA SOLVÊNCIA DA OPERADORA AGRAVANDO O JUROS

Em uma situação que a taxa de juros apresente valores com uma tendência crescente ao longo do tempo (conforme a Figura 13), foi obtido a seguinte reserva quantílica:

Figura 20: Histograma do TAP com com a taxa SELIC agravada

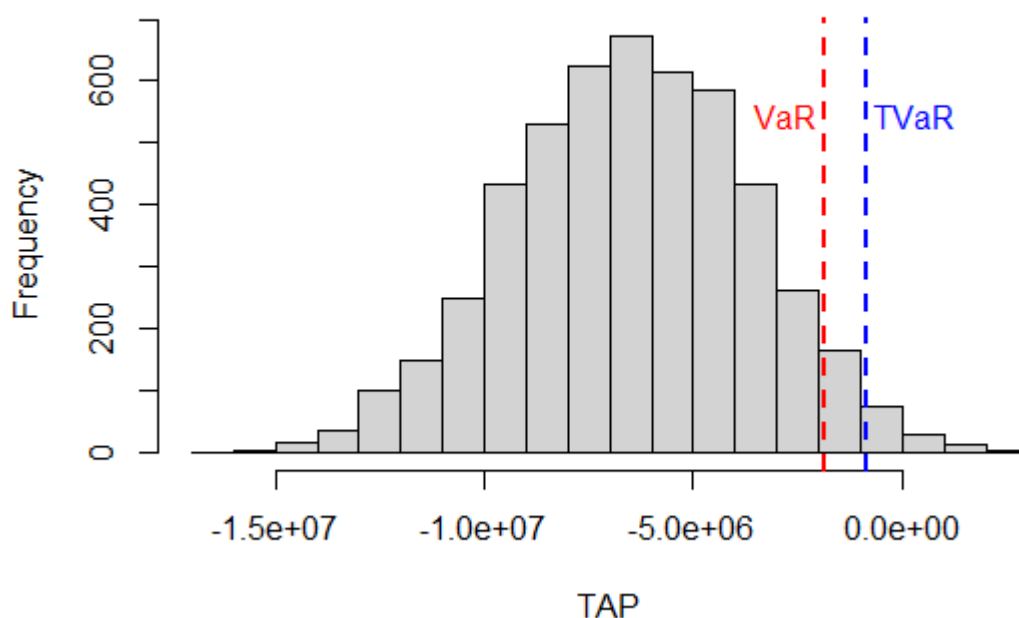


Tabela 4: Intervalo de confiança de 95% para o VaR e TVaR do TAP com $\alpha = 0,05$ e com a taxa SELIC agravada.

Medidas de Risco	Média	Limite Inferior	Limite Superior
VaR	-1.880.540,92	-2.034.950,36	-1.726.131,48
TVaR	-837.464,075	-976.747,39	-698.180,76

Observa-se nessa situação que a operadora não apresentou insolvência, pois utilizando o VaR e o TVaR como medida de risco, ambos possuem valores abaixo de zero demonstrando que ruína não ocorreu, ainda que em algumas simulações o valor do TAP ficou positivo.

6. CONCLUSÕES

Construímos um modelo estocástico para avaliação do Teste de Adequação do Passivo, pelo método de aproximação Monte Carlo, onde os valores da taxa de juros é estocástica e os dados assistenciais estudados são modelados como séries temporais, para isso usamos modelos estocásticos como o processo CIR para estimar e fazer a previsão da taxa de juros e os modelos ARIMA para modelar e fazer a previsão esperada dos dados assistenciais como número de beneficiários, valor do ticket médio e custo per capita por contratação.

Como dito anteriormente, foram utilizados os modelos ARIMA para estimar os dados assistenciais e como resultado dessa análise foi possível entender o comportamento dessas séries temporais e perceber tendências que aconteceram ao longo dos anos de estudo observados. Além disso, também foram gerados agravamentos no custo per capita usando os modelos ARIMA para obter cenários de incertezas que podem ocorrer no setor.

Usamos o modelo de equação diferencial estocástica CIR para modelar a taxa SELIC, que foi utilizada como um fator de desconto no cálculo do TAP. Também foi gerado um agravamento nos juros para estudar o impacto que a taxa de desconto tem sobre o valor do TAP.

A partir do uso das simulações pelo Método de Monte Carlo foi possível avaliar os riscos no valor do TAP, este risco foi avaliado por meio das medidas de risco VaR e TvaR. Avaliações que geraram resultados condizentes com a realidade do setor e ainda que existam técnicas de estimação mais sofisticadas, nossa esperança é que esse estudo servirá como uma sugestão introdutória para que novas análises possam ser desenvolvidas para o ramo atuarial da saúde suplementar que ainda está em crescimento no país.

7. BIBLIOGRAFIA

Agência Nacional de Saúde Suplementar (2º trimestre de 2023. Rio de Janeiro). Boletim panorama: saúde suplementar.

ARIMA – Wikipédia, a enciclopédia livre (wikipedia.org)
<https://pt.wikipedia.org/wiki/ARIMA> (acessado em 12/12/2023).

Cocozza, R., Di Lorenzo, E., Orlando, A., & Sibillo, M. (2008). A liability adequacy test for mathematical provision. In *Mathematical and Statistical Methods in Insurance and Finance* (pp. 75-81). Springer Milan.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1978). A Theory of the Term Structure of Interest Rates. Research Paper, Graduate School of Business, Stanford Univ.

Duarte, A. A., Silva, A. F. D., Oliveira, L. V., Weffort, E. F. J., & Chan, B. L. (2015). A Estrutura a Termo da Taxa de Juros e seu Impacto no Teste de Adequação de Passivo para Seguradoras no Brasil. *Revista Contabilidade & Finanças*, 26, 223-236.

Gonçalves, R. S., Ohayon, P., Rodrigues, A., & da Silva Macedo, M. A. (2018). ANÁLISE CONTÁBIL DO TESTE DE ADEQUAÇÃO DE PASSIVOS REALIZADO POR SOCIEDADES REGULADAS PELA SUSEP NO PERÍODO ENTRE 2011 A 2014. *Pensar Acadêmico*, 16(2), 279-297.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, 27, 1-22.

Iacus, S. M., & Yoshida, N. (2018). Simulation and inference for stochastic processes with YUIMA. A comprehensive R framework for SDEs and other stochastic processes. Use R.

Korn R., et al. (2010). [Chapman & Hall_CRC Financial Mathematics Series] - Monte Carlo methods and models in finance and insurance.

Milanés, A (2004) Uma introdução ao cálculo estocástico e às equações diferenciais estocásticas Departamento de Estatística UFMG. Disponível em: <https://www.est.ufmg.br/portal/wp-content/uploads/2023/01/RTE-03-2004.pdf>. (Acessado em 12/12/2023).

Nazareno Junior, José (2019) Teste de Adequação de Passivo (TAP) – aspectos técnicos, cenários, desafios e um exemplo hipotético., <https://pt.linkedin.com/pulse/teste-de-adequa%C3%A7%C3%A3o-passivo-tap-aspectos-t%C3%A9cnicos-e-um-maci-el-junior> (acessado em 07/08/2023).

Nazareno Junior, José (2016) Uma sugestão introdutória de modelagem estocástica do CBR de subscrição para operadoras médico-hospitalares., <https://pt.linkedin.com/pulse/uma-sugest%C3%A3o-introdu%C3%A7%C3%A3o-de-modelagem-es-toc%C3%A1stica-do-maci-el-junior> (acessado em 07/08/2023).

Overbeck, L., & Ryden, T. (1997). Estimation in the cox-ingersoll-ross model. *Econometric Theory*, 13(3), 430-461.

Peng, J. (2009, July). Value at risk and tail value at risk in uncertain environment. In Proceedings of the 8th international conference on information and management sciences (pp. 787-793).

Ribeiro, A et.al., (2012), Nota Técnica sobre o Teste de Adequação de Passivos (TAP). ESTUDOS FUNENSEG 26.

Taxa Selic – Wikipédia, a enciclopédia livre (wikipedia.org)
https://pt.wikipedia.org/wiki/Taxa_Selic (acessado em 12/12/2023).

Vu, K. M. (2007). *The ARIMA and VARIMA time series: Their modelings, analyses and applications*. AuLac Technologies Inc..

8. ANEXOS

Códigos em R usados no trabalho:

8.1. TRABALHANDO AS BASES DE BENEFICIÁRIOS:

```

library(stringr)

readLines("Base de Beneficiários.csv", n=8)

Operadora <- str_sub( readLines("Base de Beneficiários.csv", n=3)[3], start = 15, end = 20)

Historico_benef <- read.table("Base de Beneficiários.csv", header = FALSE, skip = 5, sep = ";")

if(length(Historico_benef) == 7 ){
Historico_benef <- read.table("Base de Beneficiários.csv",
                             col.names = c("Competência", "Individual ou Familiar",
                                             "Coletivo Empresarial", "Coletivo por Adesão",
                                             "Coletivo não identificado", "Não Informado", "Total"),
                             header = FALSE, skip = 5, sep = ";")
} else {
  if(length(Historico_benef) == 6 ){
    Historico_benef <- read.table("Base de Beneficiários.csv",
                                  col.names = c("Competência", "Individual ou Familiar",
                                                  "Coletivo Empresarial", "Coletivo por Adesão",
                                                  "Não Informado", "Total"), header = FALSE, skip = 5, sep = ";")
  } else {Historico_benef <- read.table("Base de Beneficiários.csv",
                                        col.names = c("Competência", "Individual ou Familiar",
                                                      "Coletivo Empresarial", "Coletivo por Adesão",
                                                      "Total"),
                                        header = FALSE, skip = 5, sep = ";")
  }}

require(lubridate)
require(dplyr)

Historico_benef$Competência <- sub("Dez", "Dec", Historico_benef$Competência)
Historico_benef$Competência <- sub("Set", "Sep", Historico_benef$Competência)

Historico_benef$Competência <- my(Historico_benef$Competência)

Novo_mes <- Historico_benef %>%
  mutate(mês_anterior = floor_date(Competência-1, unit = "month")) %>%
  mutate(meses_anterior = floor_date(mês_anterior-1, unit = "month"))

Competência <- as.Date(rbind(Novo_mes$Competência, Novo_mes$mês_anterior,
                             Novo_mes$meses_anterior))
Competência <- as.data.frame(Competência); Competência <- as.vector(t(Competência))

```

```

Ind_Fam <- matrix(0,length(Historico_benef$Competência),1)
Col_Ade <- matrix(0,length(Historico_benef$Competência),1)
Col_Emp <- matrix(0,length(Historico_benef$Competência),1)

Individual.ou.Familiar <- for (i in 1:length(Historico_benef$Competência)){
  Ind_Fam[i,1] = Historico_benef$Individual.ou.Familiar[i+1] -
  Historico_benef$Individual.ou.Familiar[i]
}

Coletivo.Empresarial <- for (i in 1:length(Historico_benef$Competência)){
  Col_Emp[i,1] = Historico_benef$Coletivo.Empresarial[i+1] -
  Historico_benef$Coletivo.Empresarial[i]
}

Coletivo.por.Adesão <- for (i in 1:length(Historico_benef$Competência)){
  Col_Ade[i,1] = Historico_benef$Coletivo.por.Adesão[i+1] -
  Historico_benef$Coletivo.por.Adesão[i]
}

Benef <- as.data.frame(cbind(Historico_benef$Competência,(-1)*Ind_Fam, (-1)*Col_Emp,
(-1)*Col_Ade))
Benef$V1 <- as.Date(Benef$V1)

names(Benef)[1]<-paste("Competência")
names(Benef)[2]<-paste("Individual.ou.Familiar")
names(Benef)[3]<-paste("Coletivo.Empresarial")
names(Benef)[4]<-paste("Coletivo.por.Adesão")

Ind_Fam <- matrix(0,length(Historico_benef$Competência),3)
Col_Ade <- matrix(0,length(Historico_benef$Competência),3)
Col_Emp <- matrix(0,length(Historico_benef$Competência),3)

Individual.ou.Familiar <- for (i in 1:length(Historico_benef$Competência)){
  Ind_Fam[i,1] = Historico_benef$Individual.ou.Familiar[i+1] -
  Historico_benef$Individual.ou.Familiar[i]
}

Coletivo.Empresarial <- for (i in 1:length(Historico_benef$Competência)){
  Col_Emp[i,1] = Historico_benef$Coletivo.Empresarial[i+1] -
  Historico_benef$Coletivo.Empresarial[i]
}

Coletivo.por.Adesão <- for (i in 1:length(Historico_benef$Competência)){
  Col_Ade[i,1] = Historico_benef$Coletivo.por.Adesão[i+1] -
  Historico_benef$Coletivo.por.Adesão[i]
}

```

```

Individual.ou.Familiar <- for (i in 1:length(Benef$Competência)){
  Ind_Fam[i,1] = Historico_benef$Individual.ou.Familiar[i]
  Ind_Fam[i,2] = Historico_benef$Individual.ou.Familiar[i] -
round(Benef$Individual.ou.Familiar[i]/3)
  Ind_Fam[i,3] = Historico_benef$Individual.ou.Familiar[i] -
round(2*Benef$Individual.ou.Familiar[i]/3)
}

Coletivo.Empresarial <- for (i in 1:length(Benef$Competência)){
  Col_Emp[i,1] = Historico_benef$Coletivo.Empresarial[i]
  Col_Emp[i,2] = Historico_benef$Coletivo.Empresarial[i] -
round(Benef$Coletivo.Empresarial[i]/3)
  Col_Emp[i,3] = Historico_benef$Coletivo.Empresarial[i] -
round(2*Benef$Coletivo.Empresarial[i]/3)
}

Coletivo.por.Adesão <- for (i in 1:length(Benef$Competência)){
  Col_Ade[i,1] = Historico_benef$Coletivo.por.Adesão[i]
  Col_Ade[i,2] = Historico_benef$Coletivo.por.Adesão[i] - round(Benef$Coletivo.por.Adesão[i]/3)
  Col_Ade[i,3] = Historico_benef$Coletivo.por.Adesão[i] -
round(2*Benef$Coletivo.por.Adesão[i]/3)
}

rm(Individual.ou.Familiar, Coletivo.Empresarial, Coletivo.por.Adesão)

Ind_Fam <- as.data.frame(Ind_Fam); Individual.ou.Familiar <- as.vector(t(Ind_Fam))
Col_Emp <- as.data.frame(Col_Emp); Coletivo.Empresarial <- as.vector(t(Col_Emp))
Col_Ade <- as.data.frame(Col_Ade); Coletivo.por.Adesão <- as.vector(t(Col_Ade))

Base_Completa <- cbind.data.frame(Competência, Individual.ou.Familiar,
                                Coletivo.Empresarial, Coletivo.por.Adesão)

Base_Completa <- Base_Completa %>%
  mutate(Total = rowSums(select_if(., is.numeric), na.rm = TRUE))

Base_Completa <- Base_Completa[-nrow(Base_Completa), ]
Base_Completa <- Base_Completa[-nrow(Base_Completa), ]

Base_Completa <- Base_Completa %>%
  filter(Competência<="2022-12-01")

manual <- Base_Completa[81:82,1:4]

dif <- round((manual[1,2:4] - manual[2,2:4])/4)
dif_1 <- manual[1,2:4]- dif
dif_2 <- dif_1 - dif
dif_3 <- dif_2 - dif
Competência <- rbind("2016-03-01", "2016-02-01", "2016-01-01")
Total <- rbind(sum(dif_1), sum(dif_2), sum(dif_3))
manual <- rbind(dif_1, dif_2, dif_3)

```

```

manual <- cbind(Competência, manual, Total)

Base_Completa <- rbind(Base_Completa, manual)

Base_Completa <- Base_Completa %>%
  arrange(desc(Competência)) %>%
  filter(Competência >="2012-01-01")

rm(Benef,Col_Ade,Col_Emp,Ind_Fam,Novo_mes,i,Coletivo.Empresarial,Coletivo.por.Adesão,
  Competência, Individual.ou.Familiar, Historico_benef, dif, dif_1, dif_2, dif_3,
  manual, Total)

```

8.2. TRABALHANDO AS BASES DE DESPESAS E RECEITAS ASSISTENCIAIS:

```

require(dplyr)

{
base_2001 <- read.csv("Bases de Receitas e Despesas/tb_rc_2001.csv")
base_2002 <- read.csv("Bases de Receitas e Despesas/tb_rc_2002.csv")
base_2003 <- read.csv("Bases de Receitas e Despesas/tb_rc_2003.csv")
base_2004 <- read.csv("Bases de Receitas e Despesas/tb_rc_2004.csv")
base_2005 <- read.csv("Bases de Receitas e Despesas/tb_rc_2005.csv")
base_2006 <- read.csv("Bases de Receitas e Despesas/tb_rc_2006.csv")
base_2007 <- read.csv("Bases de Receitas e Despesas/tb_rc_2007.csv")
base_2008 <- read.csv("Bases de Receitas e Despesas/tb_rc_2008.csv")
base_2009 <- read.csv("Bases de Receitas e Despesas/tb_rc_2009.csv")
base_2010 <- read.csv("Bases de Receitas e Despesas/tb_rc_2010.csv")
base_2011 <- read.csv("Bases de Receitas e Despesas/tb_rc_2011.csv")
base_2012 <- read.csv("Bases de Receitas e Despesas/tb_rc_2012.csv")
base_2013 <- read.csv("Bases de Receitas e Despesas/tb_rc_2013.csv")
base_2014 <- read.csv("Bases de Receitas e Despesas/tb_rc_2014.csv")
base_2015 <- read.csv("Bases de Receitas e Despesas/tb_rc_2015.csv")
base_2016 <- read.csv("Bases de Receitas e Despesas/tb_rc_2016.csv")
base_2017 <- read.csv("Bases de Receitas e Despesas/tb_rc_2017.csv")
base_2018 <- read.csv("Bases de Receitas e Despesas/tb_rc_2018.csv")
base_2019 <- read.csv("Bases de Receitas e Despesas/tb_rc_2019.csv")
base_2020 <- read.csv("Bases de Receitas e Despesas/tb_rc_2020.csv")
base_2021 <- read.csv("Bases de Receitas e Despesas/tb_rc_2021.csv")
base_2022 <- read.csv("Bases de Receitas e Despesas/tb_rc_2022.csv")
}

Base_1 <- rbind(base_2001, base_2002, base_2003, base_2004, base_2005, base_2006)

Base_2 <- rbind(base_2007, base_2008, base_2009, base_2010, base_2011, base_2012,
  base_2013, base_2014, base_2015, base_2016, base_2017, base_2018,
  base_2019, base_2020, base_2021, base_2022)

```

```

Base_1 <- Base_1 %>%
  select(ANO, OPERADORA, RECEITA, DESP_AST, DESP_ADM, DESPESAS_C)

Base_2 <- Base_2 %>%
  select(ANO, OPERADORA, RECEITA, DESP_AST, DESP_ADM, DESPESAS_C)

Base <- rbind(Base_1, Base_2) %>%
  filter(OPERADORA == Operadora & ANO>="2012")

Balan_Receita_CE <- matrix(0,length(Base$ANO),12)
Balan_Receita_CA <- matrix(0,length(Base$ANO),12)
Balan_Receita_IND <- matrix(0,length(Base$ANO),12)
Balan_Desp_ast_CE <- matrix(0,length(Base$ANO),12)
Balan_Desp_ast_CA <- matrix(0,length(Base$ANO),12)
Balan_Desp_ast_IND <- matrix(0,length(Base$ANO),12)
Balan_Desp_adm <- matrix(0,length(Base$ANO),12)
Balan_Desp_c <- matrix(0,length(Base$ANO),12)

set.seed(255)

Balanço <- for (i in 1:length(Base$ANO)){
  Balan_Receita_CE[i,1:12] = rnorm(n=12, mean =
(Base$RECEITA[i]/12)*Base_Completa$Coletivo.Empresarial/Base_Completa$Total,
sd=(Base$RECEITA[i]/12)*Base_Completa$Coletivo.Empresarial/Base_Completa$Total*0.20)
  Balan_Receita_CA[i,1:12] = rnorm(n=12, mean =
(Base$RECEITA[i]/12)*Base_Completa$Coletivo.por.Adesão/Base_Completa$Total,
sd=(Base$RECEITA[i]/12)*Base_Completa$Coletivo.por.Adesão/Base_Completa$Total*0.20)
  Balan_Receita_IND[i,1:12] = rnorm(n=12, mean =
(Base$RECEITA[i]/12)*Base_Completa$Individual.ou.Familiar/Base_Completa$Total,
sd=(Base$RECEITA[i]/12)*Base_Completa$Individual.ou.Familiar/Base_Completa$Total*0.20)
  Balan_Desp_ast_CE[i,1:12] = rnorm(n=12, mean =
(Base$DESP_AST[i]/12)*Base_Completa$Coletivo.Empresarial/Base_Completa$Total,
sd=(Base$DESP_AST[i]/12)*Base_Completa$Coletivo.Empresarial/Base_Completa$Total*0.20)
  Balan_Desp_ast_CA[i,1:12] = rnorm(n=12, mean =
(Base$DESP_AST[i]/12)*Base_Completa$Coletivo.por.Adesão/Base_Completa$Total,
sd=(Base$DESP_AST[i]/12)*Base_Completa$Coletivo.por.Adesão/Base_Completa$Total*0.20)
  Balan_Desp_ast_IND[i,1:12] = rnorm(n=12, mean =
(Base$DESP_AST[i]/12)*Base_Completa$Individual.ou.Familiar/Base_Completa$Total,
sd=(Base$DESP_AST[i]/12)*Base_Completa$Individual.ou.Familiar/Base_Completa$Total*0.20)

  Balan_Desp_adm[i,1:12] = rnorm(n=12, mean = Base$DESP_ADM[i]/12,
sd=Base$DESP_ADM[i]/12*0.20)
  Balan_Desp_c[i,1:12] = rnorm(n=12, mean = Base$DESPESAS_C[i]/12,
sd=Base$DESPESAS_C[i]/12*0.20)
}

```

```

}

BR <- as.data.frame(t(Balan_Receita_CE))
Balan_Receita_CE <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20, BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Receita_CA))
Balan_Receita_CA <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20, BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Receita_IND))
Balan_Receita_IND <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20, BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Desp_ast_CE))
Balan_Desp_ast_CE <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20,BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Desp_ast_CA))
Balan_Desp_ast_CA <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20,BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Desp_ast_IND))
Balan_Desp_ast_IND <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20,BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Desp_adm))
Balan_Desp_adm <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20,BR$V21,BR$V22)))

BR <- as.data.frame(t(Balan_Desp_c))
Balan_Desp_c <-
t(rbind(c(BR$V1,BR$V2,BR$V3,BR$V4,BR$V5,BR$V6,BR$V7,BR$V8,BR$V9,
          BR$V10,BR$V11,BR$V12,BR$V13,BR$V14,BR$V15,BR$V16,BR$V17,
          BR$V18,BR$V19,BR$V20,BR$V21,BR$V22)))

```

```

Competência <- as.data.frame(seq(from=as.Date("2012-01-01"), to=as.Date("2022-12-01"),
by="month"))
names(Competência)[1]<-paste("Competência")

Base_Rec_Desp <- as.data.frame(cbind(Balan_Receita_CE, Balan_Receita_CA,
Balan_Receita_IND,
                                Balan_Desp_ast_CE, Balan_Desp_ast_CA, Balan_Desp_ast_IND,
                                Balan_Desp_adm, Balan_Desp_c))

Base_Rec_Desp <- cbind(Competência,Base_Rec_Desp) %>%
  arrange(desc(Competência))

rm(Balanço, BR, Competência, Balan_Desp_adm, Balan_Desp_ast_CE,Balan_Desp_ast_CA,
  Balan_Desp_ast_IND, Balan_Desp_c, Balan_Receita_CE, Balan_Receita_CA,
  Balan_Receita_IND,
  Operadora)
rm(Base, Base_1, Base_2, base_2001, base_2002, base_2003, base_2004,
  base_2005, base_2006, base_2007, base_2008, base_2009, base_2010,
  base_2011, base_2012, base_2013, base_2014, base_2015, base_2016,
  base_2017, base_2018, base_2019, base_2020, base_2021, base_2022)

Balancete <- cbind(Base_Completa,Base_Rec_Desp[2:7])
names(Balancete)[6:11]<-paste(c("Receita Ast CE", "Receita Ast CA", "Receita Ast IND",
  "Despesa Ast CE", "Despesa Ast CA", "Despesa Ast IND"))

Balancete <- Balancete %>%
  mutate("Ticket Médio - CE" = `Receita Ast CE`/Coletivo.Empresarial) %>%
  mutate("Custo Per Capita - CE" = `Despesa Ast CE`/Coletivo.Empresarial) %>%
  mutate("Ticket Médio - CA" = `Receita Ast CA`/Coletivo.por.Adesão) %>%
  mutate("Custo Per Capita - CA" = `Despesa Ast CA`/Coletivo.por.Adesão) %>%
  mutate("Ticket Médio - IND" = `Receita Ast IND`/Individual.ou.Familiar) %>%
  mutate("Custo Per Capita - IND" = `Despesa Ast IND`/Individual.ou.Familiar)

rm(Base_Completa, Base_Rec_Desp, i)

```

8.3. TRABALHANDO OS DADOS DA TAXA SELIC:

```

require(lubridate)
require(dplyr)
library(readr)

SELIC <- read.csv("C:\\Users\\gusta\\Documents\\Iniciacao Cientifica\\codigos/selicDiaria.csv", sep
= ";")

names(SELIC)[2]<-paste("Percentual Diario"); names(SELIC)[3]<-paste("Percentual Anual")
SELIC <- as.data.frame(cbind(SELIC$Date, SELIC$`Percentual Diario`))

```

```

names(SELIC)[1]<-paste("Date"); names(SELIC)[2]<-paste("Percentual Diario")
SELIC <- SELIC[-8750, ]; SELIC$Date <- dmy(SELIC$Date)
SELIC$`Percentual Diario` <- as.numeric(SELIC$`Percentual Diario`)
SELIC$Date <- as.character.Date(SELIC$Date)

SELIC <- SELIC %>%
  filter(Date >= "2011-01-01")

#write.csv(SELIC, "SELIC.txt", row.names = FALSE)

require(yuima)
require(yuimaGUI)

yuimaGUI(theme = "white")

Serie_SELIC <- read_table("Percentual.Diario 5.txt")
Serie_SELIC <- Serie_SELIC[-2]

Serie_SELIC$Date <- format(Serie_SELIC$Date, "%Y-%m")
média <- rowMeans(Serie_SELIC[-1])

Taxa_SELIC <- cbind.data.frame(Serie_SELIC$Date, média)

Taxa_SELIC <- Taxa_SELIC %>%
  group_by(`Serie_SELIC$Date`) %>%
  summarise(sum(as.numeric(média))) %>%
  filter(`Serie_SELIC$Date` <= "2029-12")

#Taxa_SELIC <- Taxa_SELIC$`mean(as.numeric(média))`

Previsao_SELIC = ts(Taxa_SELIC$`sum(as.numeric(média))`,start=2023,frequency=12)

plot(Previsao_SELIC)

SELIC_Observada <- read_table("SELIC_Observada.txt")

SELIC_Observada$Date <- format(SELIC_Observada$Date, "%Y-%m")

Taxa_SELIC <- cbind.data.frame(SELIC_Observada$Date, SELIC_Observada$Percentual.Diario)

Taxa_SELIC <- Taxa_SELIC %>%
  group_by(`SELIC_Observada$Date`) %>%
  summarise(sum(as.numeric(`SELIC_Observada$Percentual.Diario`))) %>%
  filter(`SELIC_Observada$Date` >= "2011-01")

SELIC_Observada = ts(Taxa_SELIC$`sum(as.numeric(`SELIC_Observada$Percentual.Diario`))`,
  start=2011,frequency=12)

```

```

plot(SELIC_Observada)

SELIC <- c(SELIC_Observada, Previsao_SELIC)

SELIC <- ts(SELIC, start=2011,frequency=12)

plot(SELIC)
abline(v=2022)

```

8.4. TRABALHANDO AS PREVISÕES DOS PARÂMETROS:

```

require(forecast)
require(lmtest)
library(readr)
require(lubridate)
require(dplyr)

#Base.Completa <- Balancete %>%
# arrange(desc(desc(Competência)))

#write.csv2(Base.Completa, "Base Completa.csv",row.names = FALSE)

Base.Completa <- read_csv2("Base Completa.csv", col_names = TRUE)

Valor_Futuro = matrix(0,84,1)

#####
#1 Previsão

valores <- Base.Completa$`Ticket Médio - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao

```

```

Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,750))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CE <- c(Valor_Futuro)

#####
#2 Previsão
valores <- Base.Completa$`Custo Per Capita - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))
lines(Previsto, col='blue')
lines(LI, col='red')

```

```

lines(LS, col='red')

Prev_CPC_CE <- c(Valor_Futuro)

#####
#3 Previsão

valores <- Base.Completa$`Ticket Médio - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CA <- c(Valor_Futuro)

#####
#4 Previsão

valores <- Base.Completa$`Custo Per Capita - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

```

```

MRed=arima(valores, order = c(1, 0, 1),seasonal = list(order = c(1, 2, 1)))
coefstest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,500))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_CA <- c(Valor_Futuro)

#####
#5 Previsão

valores <- Base.Completa$`Ticket Médio - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefstest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]

```

```

LI[n+H-H+i] <- Prev$lower[i]
LS[n+H-H+i] <- Prev$upper[i]

Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,800))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_IND <- c(Valor_Futuro)

#####
#6 Previsão

valores <- Base.Completa$`Custo Per Capita - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,600))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_IND <- c(Valor_Futuro)

#####
#7 Previsão

```

```

valores <- Base.Completa$Coletivo.Empresarial
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(2, 1, 0)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores', ylim=c(65000,91000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CE <- c(round(Valor_Futuro))

#####
#8 Previsão

valores <- Base.Completa$Coletivo.por.Adesão
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(1, 0, 0)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))

```

```

# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(10000,25000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CA <- c(round(Valor_Futuro))

#####
#9 Previsão

valores <- Base.Completa$Individual.ou.Familiar
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(18000,26000))

```

```

lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_IND <- c(round(Valor_Futuro))

#####

Serie_SELIC <- read_table("Percentual.Diario 5.txt")
Serie_SELIC <- Serie_SELIC[-2]

Serie_SELIC$Date <- format(Serie_SELIC$Date, "%Y-%m")
média <- rowMeans(Serie_SELIC[-1])

Taxa_SELIC <- cbind.data.frame(Serie_SELIC$Date, média)

Taxa_SELIC <- Taxa_SELIC %>%
  group_by(`Serie_SELIC$Date`) %>%
  summarise(sum(as.numeric(média))) %>%
  filter(`Serie_SELIC$Date` >= "2023-01") %>%
  filter(`Serie_SELIC$Date` <= "2029-12")

Taxa_SELIC <- Taxa_SELIC$`sum(as.numeric(média))`

#####

Base_Previsões <- as.data.frame(cbind(Prev_BENEF_CE,Prev_TM_CE,Prev_CPC_CE,
  Prev_BENEF_CA,Prev_TM_CA,Prev_CPC_CA,
  Prev_BENEF_IND,Prev_TM_IND, Prev_CPC_IND,
  Taxa_SELIC))

rm(MRed, Prev, Valor_Futuro, Dados, H, i, LI, LS, n, Prev_CPC_CA, Taxa_SELIC,
  Prev_CPC_CE, Prev_CPC_IND, Prev_TM_CA, Prev_TM_CE, Prev_TM_IND, média,
  Previsto, valores, VPrev, Prev_BENEF_CA, Prev_BENEF_CE, Prev_BENEF_IND)

```

8.5. TRABALHANDO AS PREVISÕES DOS PARÂMETROS COM AGRAVO NOS CUSTOS PER CAPITA:

```

require(forecast)
require(lmtest)
library(readr)
require(lubridate)
require(dplyr)

#Base.Completa <- Balancete %>%

```

```

# arrange(desc(desc(Competência)))

#write.csv2(Base.Completa, "Base Completa.csv",row.names = FALSE)

Base.Completa <- read_csv2("Base Completa.csv", col_names = TRUE)

Valor_Futuro = matrix(0,84,1)

#####
#1 Previsão

valores <- Base.Completa$`Ticket Médio - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Gráfico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CE <- c(Valor_Futuro)

#####
#2 Previsão
valores <- Base.Completa$`Custo Per Capita - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)

```

```

H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(90))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = LS[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,1000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_CE <- c(Valor_Futuro)

#####
#3 Previsão

valores <- Base.Completa$`Ticket Médio - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)

```

```

LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CA <- c(Valor_Futuro)

#####
#4 Previsão

valores <- Base.Completa$`Custo Per Capita - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 1),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(90))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = LS[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,850))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

```

```

Prev_CPC_CA <- c(Valor_Futuro)

#####
#5 Previsão

valores <- Base.Completa$`Ticket Médio - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefstest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,800))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_IND <- c(Valor_Futuro)

#####
#6 Previsão

valores <- Base.Completa$`Custo Per Capita - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefstest(MRed)

```

MRed

```
# Calculando previsoes
Prev=forecast(MRed, H, level=c(90))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = LS[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,900))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_IND <- c(Valor_Futuro)

#####
#7 Previsão

valores <- Base.Completa$Coletivo.Empresarial
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(2, 1, 0)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]
```

```

Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores', ylim=c(65000,91000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CE <- c(round(Valor_Futuro))

#####
#8 Previsão

valores <- Base.Completa$Coletivo.por.Adesão
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(1, 0, 0)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(10000,25000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CA <- c(round(Valor_Futuro))

#####
#9 Previsão

valores <- Base.Completa$Individual.ou.Familiar

```

```

valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1)))
coefstest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(18000,26000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_IND <- c(round(Valor_Futuro))

#####

Serie_SELIC <- read_table("Percentual.Diario 5.txt")
Serie_SELIC <- Serie_SELIC[-2]

Serie_SELIC$Date <- format(Serie_SELIC$Date, "%Y-%m")
média <- rowMeans(Serie_SELIC[-1])

Taxa_SELIC <- cbind.data.frame(Serie_SELIC$Date, média)

Taxa_SELIC <- Taxa_SELIC %>%
  group_by(`Serie_SELIC$Date`) %>%
  summarise(sum(as.numeric(média))) %>%
  filter(`Serie_SELIC$Date` >= "2023-01") %>%
  filter(`Serie_SELIC$Date` <= "2029-12")

Taxa_SELIC <- Taxa_SELIC$`sum(as.numeric(média))`

```

```
#####
Base_Previsões <- as.data.frame(cbind(Prev_BENEF_CE,Prev_TM_CE,Prev_CPC_CE,
  Prev_BENEF_CA,Prev_TM_CA,Prev_CPC_CA,
  Prev_BENEF_IND,Prev_TM_IND, Prev_CPC_IND,
  Taxa_SELIC))

rm(MRed, Prev, Valor_Futuro, Dados, H, i, LI, LS, n, Prev_CPC_CA, Taxa_SELIC,
  Prev_CPC_CE, Prev_CPC_IND, Prev_TM_CA, Prev_TM_CE, Prev_TM_IND, média,
  Previsto, valores, VPrev, Prev_BENEF_CA, Prev_BENEF_CE, Prev_BENEF_IND)
```

8.6. TRABALHANDO AS PREVISÕES DOS PARÂMETROS COM AGRAVO NA TAXA SELIC:

```
require(forecast)
require(lmtest)
library(readr)
require(lubridate)
require(dplyr)

#Base.Completa <- Balancete %>%
# arrange(desc(desc(Competência)))

#write.csv2(Base.Completa, "Base Completa.csv",row.names = FALSE)

Base.Completa <- read_csv2("Base Completa.csv", col_names = TRUE)

Valor_Futuro = matrix(0,84,1)

#####
#1 Previsão

valores <- Base.Completa$`Ticket Médio - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
```

```

Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,750))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CE <- c(Valor_Futuro)

#####
#2 Previsão
valores <- Base.Completa$`Custo Per Capita - CE`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))

```

```

lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_CE <- c(Valor_Futuro)

#####
#3 Previsão

valores <- Base.Completa$`Ticket Médio - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,650))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_CA <- c(Valor_Futuro)

#####
#4 Previsão

valores <- Base.Completa$`Custo Per Capita - CA`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

```

```

plot(valores)

MRed=arima(valores, order = c(1, 0, 1),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,500))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_CA <- c(Valor_Futuro)

#####
#5 Previsão

valores <- Base.Completa$`Ticket Médio - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)

```

```

for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,800))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_TM_IND <- c(Valor_Futuro)

#####
#6 Previsão

valores <- Base.Completa$`Custo Per Capita - IND`
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 0, 0),seasonal = list(order = c(1, 2, 1)))
coefest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(70,600))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_CPC_IND <- c(Valor_Futuro)

```

```
#####
#7 Previsão

valores <- Base.Completa$Coletivo.Empresarial
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(2, 1, 0)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores', ylim=c(65000,91000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CE <- c(round(Valor_Futuro))

#####
#8 Previsão

valores <- Base.Completa$Coletivo.por.Adeseão
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(1, 1, 0),seasonal = list(order = c(1, 0, 0)))
coeftest(MRed)
MRed
```

```

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(10000,25000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_CA <- c(round(Valor_Futuro))

#####
#9 Previsão

valores <- Base.Completa$Individual.ou.Familiar
valores = ts(valores,start=2012,frequency=12)
n = length(valores)
H=12*7

plot(valores)

MRed=arima(valores, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1)))
coeftest(MRed)
MRed

# Calculando previsoes
Prev=forecast(MRed, H, level=c(95))
# Grafico da serie com previsoes e intervalos de previsao
Dados=c(valores,rep(NA,H))
VPrev=ts(Dados,start=2012,frequency=12)
Previsto <- ts(rep(NA,n+H),start=2012,frequency=12)
LI <- ts(rep(NA,n+H),start=2012,frequency=12)
LS <- ts(rep(NA,n+H),start=2012,frequency=12)
for(i in 1:H){
  Previsto[n+H-H+i] <- Prev$mean[i]
  LI[n+H-H+i] <- Prev$lower[i]
  LS[n+H-H+i] <- Prev$upper[i]

  Valor_Futuro[i,1] = Previsto[n+H-H+i]
}

```

```

}
plot(VPrev,type='l',xlab='Ano',ylab='Valores',ylim=c(18000,26000))
lines(Previsto, col='blue')
lines(LI, col='red')
lines(LS, col='red')

Prev_BENEF_IND <- c(round(Valor_Futuro))

#####

Serie_SELIC <- read_table("Maximo_SELIC.txt")

Serie_SELIC$Date <- format(Serie_SELIC$Date, "%Y-%m")

Serie_SELIC <- Serie_SELIC %>%
  group_by(Date) %>%
  summarise(sum(as.numeric(maximo))) %>%
  filter(Date >= "2023-01") %>%
  filter(Date <= "2029-12")

Taxa_SELIC <- Serie_SELIC`sum(as.numeric(maximo))`

#####

Base_Previsões <- as.data.frame(cbind(Prev_BENEF_CE,Prev_TM_CE,Prev_CPC_CE,
  Prev_BENEF_CA,Prev_TM_CA,Prev_CPC_CA,
  Prev_BENEF_IND,Prev_TM_IND, Prev_CPC_IND,
  Taxa_SELIC))

rm(MRed, Prev, Valor_Futuro, Dados, H, i, LI, LS, n, Prev_CPC_CA, Taxa_SELIC,
  Prev_CPC_CE, Prev_CPC_IND, Prev_TM_CA, Prev_TM_CE, Prev_TM_IND,
  Previsto, valores, VPrev, Prev_BENEF_CA, Prev_BENEF_CE, Prev_BENEF_IND)

```

8.7. TRABALHANDO A BASE COM PREVISÕES PARA A ESTIMAÇÃO DO TAP:

```

#tempo em meses T
#N simulações

set.seed(255)

N<-5000
tmax<-84
#Numero de afiliados por plano S ao longo do tempo ate tmax
#Numero de planos np
np<-3
S<-array(NA, dim=c(tmax,np,N))

```

```

for (n in 1:N) {
  S[,n]<-round(c(runif(tmax*np/3,min =
min(Base_Previsões$Prev_BENEF_CE),max=max(Base_Previsões$Prev_BENEF_CE)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_BENEF_CA),max=max(Base_Previsões$Prev_BENEF_CA)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_BENEF_IND),max=max(Base_Previsões$Prev_BENEF_IND))))
}
head(S,2)
#Tiquets por plano ao longo do tempo ate tmax
T<-array(NA, dim=c(tmax,np,N))
for (n in 1:N) {
  T[,n]<-c(runif(tmax*np/3,min =
min(Base_Previsões$Prev_TM_CE),max=max(Base_Previsões$Prev_TM_CE)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_TM_CA),max=max(Base_Previsões$Prev_TM_CA)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_TM_IND),max=max(Base_Previsões$Prev_TM_IND)))
}
head(T,2)
#Premios por plano ao longo do tempo ate tmax
P<-array(NA, dim=c(tmax,np,N))
for (n in 1:N) {
  P[,n]<-c(runif(tmax*np/3,min =
min(Base_Previsões$Prev_CPC_CE),max=max(Base_Previsões$Prev_CPC_CE)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_CPC_CA),max=max(Base_Previsões$Prev_CPC_CA)),
  runif(tmax*np/3,min =
min(Base_Previsões$Prev_CPC_IND),max=max(Base_Previsões$Prev_CPC_IND)))
}
head(P,2)
#Valor presente correspondente a taxa r ao longo do tempo ate tmax

r<-array(NA, dim=c(tmax,N))
for (n in 1:N) {
  r[,n]<- c(runif(tmax,min =
min(Base_Previsões$Taxa_SELIC),max=max(Base_Previsões$Taxa_SELIC)))
}
head(r,2)

VP<-array(NA, dim=c(tmax,N))
for (n in 1:N) {
  for (i in 1:tmax) {
    VP[i,n]<-1/((1+r[i,n])^(i))
  }
}
head(VP,2)
#Simulação, N simulações
h<-array(NA,dim =c(N,tmax))
for (n in 1:N) {

```

```

for (i in 1:np) {
  for (t in 1:tmax){
    h[n,t] <- (S[t,i,n]*(P[t,i,n]-T[t,i,n]))*VP[t,n]
  }
}
}
V<-c()
for (n in 1:N) {
  ini<-0
  V[n]<-ini
  for (t in 1:tmax) {
    V[n]<-V[n]+h[n,t]
  }
}
plot(h[1,],type="l")
hist(V, (breaks=20))
abline(v=0, col="red", lwd = 2)

rm(Balancete, Base_Previsões, Base.Completa)

#####

require(actuar)

agregado <- aggregateDist(method = "normal",model.sev = V,
                          moments = c(mean(V),var(V)))

VaR <- VaR(agregado,conf.level = c(0.95))
TVaR <- CTE(agregado, conf.level = c(0.95))

VaR
TVaR

hist(V, (breaks=20))
abline(v=c(VaR[1]), col="red", lwd = 2)
abline(v=c(TVaR[1]), col="blue", lwd = 2)

```

8.8. TRABALHANDO OS INTERVALOS DE CONFIANÇA DO VAR E TVAR:

```

MonteCarlo<-function(TAP,alpha,a)
{#Calculo do VaR_alpha , com intervalo de confiança de (1-a)%
#para ter intervalos de confiança Montecarlo vamos dividir a amostra em 50 amostras de tamanho
100
n<-50
TAP_amostra<-matrix(TAP, nrow = n)
##### Estimador Monte Carlo do VaR alpha

```

```

VaR_alpha<-mean(apply(TAP_amostra,c(1),quantile,probs=c(1-alpha)))
VaR_alpha_sd<-sd(apply(TAP_amostra,c(1),quantile,probs=c(1-alpha)))
##Intervalo de confiança Monte Carlo com (1-a)x100% de confiança para VaR
a<-0.05
VaR_alpha_IC_sup<-VaR_alpha+(qt(1-a, n-1)*VaR_alpha_sd/sqrt(n))
VaR_alpha_IC_inf<-VaR_alpha-(qt(1-a, n-1)*VaR_alpha_sd/sqrt(n))
VaR<-t(as.matrix(list(alpha=alpha,VaR=VaR_alpha,VaRInf=VaR_alpha_IC_inf,VaRSup=VaR_alpha
_IC_sup)))
##### Estimador Monte Carlo do VaR alpha
Q<-apply(TAP_amostra,c(1),quantile,probs=c(1-alpha))
TVR<-NULL
for (i in 1:n) {
  q<-which(TAP_amostra[i,]>=Q[i])
  TVR[i]<-mean(TAP_amostra[i,q])
}
TVaR_alpha<-mean(TVR)
TVaR_alpha_sd<-sd(TVR)
##Intervalo de confiança Monte Carlo com (1-a)x100% de confiança para TVaR
a<-0.05
TVaR_alpha_IC_sup<-TVaR_alpha+(qt(1-a, n-1)*TVaR_alpha_sd/sqrt(n))
TVaR_alpha_IC_inf<-TVaR_alpha-(qt(1-a, n-1)*TVaR_alpha_sd/sqrt(n))
TVaR<-t(as.matrix(list(alpha=alpha,TVaR=TVaR_alpha,TVaRInf=TVaR_alpha_IC_inf,TVaRsup=T
VaR_alpha_IC_sup)))
RESULTADO<-rbind(VaR, TVaR)
colnames(RESULTADO)<-c("alfa", "Média", "Inf", "Sup")
rownames(RESULTADO)<-c("Var", "TVaR")
return(RESULTADO)
}
#Exemplo
TAP<-V
hist(TAP)
alpha<-0.05
confianca<-0.01 #corresponde ao nivel de confiança do intervalo de (0,99) ou 99%
Vares<-MonteCarlo(TAP,0.05,0.01)
knitr::kable(Vares, "simple",row.names = TRUE)
hist(TAP, main = 'Histograma de TAP', (breaks=20))
abline(v = Vares[,2], col = c('red', 'blue'), lwd = 2, lty = 'dashed')
text(Vares[1,2],0, "VaR", col = "red", adj = c(1.1, -12.5))
text(Vares[2,2],0, "TVaR", col = "blue", adj = c(-0.1, -12.5))

```