

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas do Departamento de Estatística
Programa de Especialização em Estatística

Leandro Bernardo Nascimento

**Modelagem de séries temporais para projeção do volume de concessão de
empréstimo consignado para aposentados e pensionistas do INSS**

Belo Horizonte

2022

Leandro Bernardo Nascimento

**Modelagem de séries temporais para projeção de concessão de empréstimo
consignado para aposentados e pensionistas do INSS**

Monografia apresentada ao Programa de Especialização em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como requisito para obtenção do título de Especialista em Estatística.

Orientadora: Prof^a. Dra. Ela Mercedes Medrano de Toscano

Coorientador: Prof. Dr. Luís Alberto Toscano Medrano

Belo Horizonte

2022

2022, Leandro Bernardo Nascimento.
Todos os direitos reservados

Nascimento, Leandro Bernardo.

N244m

Modelagem de séries temporais para projeção do volume de concessão de empréstimo consignado para aposentados e pensionistas do INSS [recurso eletrônico] / Leandro Bernardo Nascimento – 2022.

1 recurso online (81 f. il., color.) : pdf.

Orientador: Ela Mercedes Medrano de Toscano

Coorientador: Luís Alberto Toscano Medrano.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 79-81

1. Estatística. 2. Análise de séries temporais. 3. Consignação em pagamento – Brasil. 4. R (Linguagem de programação de computador). I. Toscano, Ela Mercedes Medrano de. II. Toscano Medrano, Luís Alberto. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. IV. Título.

CDU 519.2(043)



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 - Belo Horizonte - MG

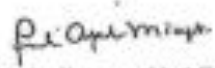
E-mail: pgest@ufmg.br
Tel: 3409-5923 - FAX: 3409-5924

ATA DO 248º TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE LEANDRO BERNARDO NASCIMENTO.

Aos sete dias do mês de novembro de 2022, às 08:30 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno Leandro Bernardo Nascimento, intitulado: "Modelagem de séries temporais para projeção da concessão de empréstimo consignado para aposentados e pensionistas do INSS", como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Professora Ela Mercedes Medrano - Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 07 de novembro de 2022.


Prof.ª Ela Mercedes Medrano (Orientadora)
Departamento de Estatística / ICEX / UFMG


Prof. Luis Alberto Toscano M. (Co-Orientador)
UPRRJ


Prof.ª Sueli Aparecida Mingoti
Departamento de Estatística / ICEX / UFMG


Prof. Mario Ernesto Piscoya Diaz
IME / UFG

Dedico à minha família, às minhas amadas filhas Bruna e Gabriela (razões da minha vida) e à minha esposa Vanessa, cujo apoio tornou possível essa conquista.

Dedico também aos meus pais, Clever e Suzana, meus maiores exemplos e grandes fontes de inspiração. Também dedico essa conquista a República Santa Casa (onde os fracos não têm vez).

AGRADECIMENTOS

Em primeiro lugar a Deus.

Agradeço enormemente aos meus orientadores, professora Ela Mercedes e professor Luís Alberto, pelo apoio, direcionamento, compreensão pelas minhas falhas e acima de tudo pela confiança de que eu conseguiria.

Também agradeço a todos os colegas de classe nesta jornada e a todos os professores do curso de especialização em estatística da UFMG, que apesar das adversidades lutam para multiplicar o conhecimento.

Agradeço aos meus pais por sempre incentivarem os estudos e a busca pelo conhecimento.

Em especial a minha amada esposa Vanessa, que sempre me incentiva na busca pelo conhecimento, no crescimento pessoal e profissional e que sacrificou precioso tempo para que eu pudesse me dedicar aos estudos.

RESUMO

O presente trabalho visa utilizar modelos estatísticos de séries temporais para projetar, com satisfatório grau de confiança, utilizando o software R, o volume de concessão de crédito consignado para aposentados e pensionistas do INSS, tendo em vista a grande relevância desta modalidade e deste nicho de tomadores de crédito dentre as várias modalidades de concessão de crédito. A série analisada compreendeu o período de maio de 2011 a dezembro de 2021, e após análise estatística da série e através do método de Box *and* Jenkins, foram ajustados diferentes modelos visando a comparação das estatísticas dos ajustes e das estatísticas dos erros de previsão para encontrar o melhor modelo que acompanhe a série. A grande variabilidade no período final da série foi um dificultador na busca pelo melhor modelo. Os resultados obtidos indicam que a utilização de variáveis exógenas pode contribuir para uma melhor previsão da série, sendo o modelo SARIMAX com sazonalidade no tempo $t=12$, o que apresentou a melhor projeção.

Palavras-chaves: séries temporais; previsão; crédito consignado; SARIMA; SARIMAX.

ABSTRACT

The present work aims to use statistical time series models to project, with a certified degree of confidence, using the R software, the volume of payroll loans granted to INSS retirees and pensioners, taking into account the degree of fidelity of this modality and this niche of credit takers within various credit granting modalities. The series followed comprised the period from May 2011 to December 2021, and after statistical analysis of the series and through the Box and Jenkins method, different models were adjusted, observing the comparison of the adjustment statistics and the forecast error statistics to find the best model that accompanies the series. The great variability in the final period of the series hindered the search for the best model. The results obtained indicate that the use of exogenous variables can contribute to a better forecast of the series, being the SARIMAX model with seasonality in time $t=12$, which presented the best projection.

Key-words: time series; forecasting; consigned credit; SARIMA; SARIMAX.

LISTA DE FIGURAS

Figura 1 - Fluxograma Box-Jenkins para Modelagem ARIMA	29
Figura 2 - Matriz Correlação.....	58

LISTA DE GRÁFICOS

Gráfico 1 - Crescimento percentual da Concessão de Crédito (ano base 2016).....	15
Gráficos 2 - Evolução Mensal das Concessões de Crédito Consignado do INSS	37
Gráficos 3 - Evolução da Taxa Média Mensal de Juros das Operações de Crédito Consignado INSS.....	40
Gráfico 4 - Evolução Taxa Selic	41
Gráfico 5 - Evolução IPCA	41
Gráficos 6 - Prazo Médio Mensal das Concessões de Crédito Consignado INSS	42
Gráficos 7 - Percentual de Inadimplência da Carteira de Crédito Consignado INSS	44
Gráficos 8 - Percentual da Carteira de Crédito Consignado INSS com Atraso entre 15 e 90 dias.....	46
Gráficos 9 - Evolução do PIB Mensal.....	47
Gráficos 10 - Selic Acumulada no Mês	49
Gráficos 11 - Evolução do salário-mínimo.....	51
Gráficos 12 - IPCA Variação Mensal.....	53
Gráfico 13 - Diagrama Dispersão Concessão Crédito INSS x Taxa Juros.....	55
Gráfico 14 - Diagrama Dispersão Concessão Crédito INSS x Prazo Concessão	55
Gráfico 15 - Diagrama Dispersão Concessão Crédito INSS x Percentual Inadimplência Carteira INSS	56
Gráfico 16 - Diagrama Dispersão Concessão Crédito INSS x Percentual Carteira em Atraso INSS.....	56
Gráfico 17 - Diagrama Dispersão Concessão Crédito INSS x PIB Mensal	56
Gráfico 18 - Diagrama Dispersão Concessão Crédito INSS x Selic Acumulada Mês	56
Gráfico 19 - Diagrama Dispersão Concessão Crédito INSS x Salário Mínimo.....	57
Gráfico 20 - Diagrama Dispersão Concessão Crédito INSS x IPCA Variação Mensal	57
Gráfico 21 - Correlograma FAC Série Temporal Concessão Crédito INSS	61
Gráfico 22 - Correlograma FACP Série Temporal Concessão Crédito INSS	61
Gráficos 23 - Séries, respectivamente, somente 1 diferença não sazonal, com 1 diferença não sazonal e 1 diferença sazonal e somente 1 diferença sazonal.....	61
Gráficos 24 - FAC e FACP dos Resíduos do Modelo 1 (95% de confiança).....	64
Gráficos 25 - FAC e FACP dos Resíduos do Modelo 1 99% de confiança	64
Gráficos 26 - Resíduos Modelo 1	65
Gráficos 27 - Funções FAC e FACP para os resíduos do Modelo 2	67
Gráficos 28 - Resíduos do Modelo 2	68
Gráficos 29 - FAC e FACP dos Resíduos do Modelo 3 (95% confiança).....	70
Gráficos 30 - FAC e FACP dos Resíduos do Modelo 3 (99% confiança).....	71
Gráficos 31 - Resíduos do Modelo 3	71
Gráficos 32 - Estimativas Período Amostral dos Modelos Ajustados	74
Gráfico 33 - Previsões dos Modelos Ajustados	75

LISTA DE QUADROS

Quadro 1 - Saída Computacional do Modelo	63
Quadro 2 - Saída Computacional do Modelo 2	66
Quadro 3 - Saída Computacional do Modelo 3	69

LISTA DE TABELAS

Tabela 1 - Estatísticas Descritivas Consignado INSS	38
Tabela 2 - Estatísticas Descritivas Taxa Juros Concessão INSS.....	41
Tabela 3 - Estatísticas Descritivas Prazo Concessão INSS em meses	43
Tabela 4 - Estatísticas Descritivas Percentual Inadimplência Carteira INSS	45
Tabela 5 - Estatísticas Descritivas Percentual da Carteira em Atraso INSS	46
Tabela 6 - Estatísticas Descritivas PIB Mensal em milhões de reais	48
Tabela 7 - Estatísticas Descritivas Selic Acumulada no Mês	50
Tabela 8 - Estatísticas Descritivas salário-mínimo	52
Tabela 9 - Estatísticas Descritivas IPCA Variação Mensal.....	54
Tabela 10 - Análise Inflação da Variância	60
Tabela 11 - Defasagens teste de Portmanteau (Q).....	72
Tabela 12 - Resumo das Estatísticas do Ajuste dos Modelos, março/2011 a junho/2021.....	73
Tabela 13 - Resumo das Estatísticas dos Erros.....	76

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Justificativa	14
1.2	Objetivo Geral	15
1.3	Objetivos Específicos	16
2	METODOLOGIA	16
2.1	Conceito de Séries Temporais	16
2.1.1	Definição de Séries Temporais	17
2.1.2	Processo Estocástico	19
2.1.3	Processo Estocástico Estacionário	19
2.1.4	Função de Autocorrelação e Função de Autocorrelação Parcial	20
2.1.5	Processo Ruído Branco Gaussiano	22
2.2	Modelos de Séries Temporais	22
2.2.1	Modelos ARMA	22
2.2.2	Diagnósticos de Modelos	26
2.2.3	Modelos ARIMA	26
2.2.4	Modelos SARIMA	27
2.2.5	Metodologia Box & Jenkins	28
2.3	Modelos Multivariados de Séries Temporais	29
2.3.1	Regressão Dinâmica	29
2.3.2	Modelos ARIMAX	32
2.4	Medidas de Acurácia	34
2.4.1	Critério de Informação de Akaike (AIC)	34
2.4.2	Critério de Informação de Bayesiano (BIC)	35
2.4.3	Estatísticas de Erros	35
3	ANÁLISE DE DADOS	36
3.1	Análise Exploratória das Séries de Dados	36
3.1.1	Evolução Mensal das Concessões de Crédito Consignado do INSS (EMCCC)	36
3.1.2	Taxa Média Mensal de Juros das Operações de Crédito Consignado INSS (TMMJ OCC)	39
3.1.3	Prazo Médio Mensal das Concessões de Crédito Consignado INSS (PMMCCC)	41
3.1.4	Inadimplência da Carteira de Crédito Consignado INSS (INAD CCC)	43
3.1.5	Percentual da Carteira de Crédito Consignado INSS com Atraso entre 15 e 90 dias (PCCC ATRASO)	45
3.1.6	PIB Mensal	47
3.1.7	Taxa Básica de Juros - Selic Acumulada no Mês (SELIC)	48
3.1.8	Salário-Mínimo	50
3.1.9	Índice Nacional de Preços ao Consumidor-amplio (IPCA)	52
3.2	Correlação entre séries	54
3.3	Estimação dos parâmetros dos Modelos	61
3.4	Valores da Estatística de Ajuste dos Modelos	72

3.5	Previsões para o período de validação	74
3.6	Estatísticas dos Erros de Previsão	76
4	CONSIDERAÇÕES FINAIS.....	77
	REFERÊNCIAS.....	79

1 INTRODUÇÃO

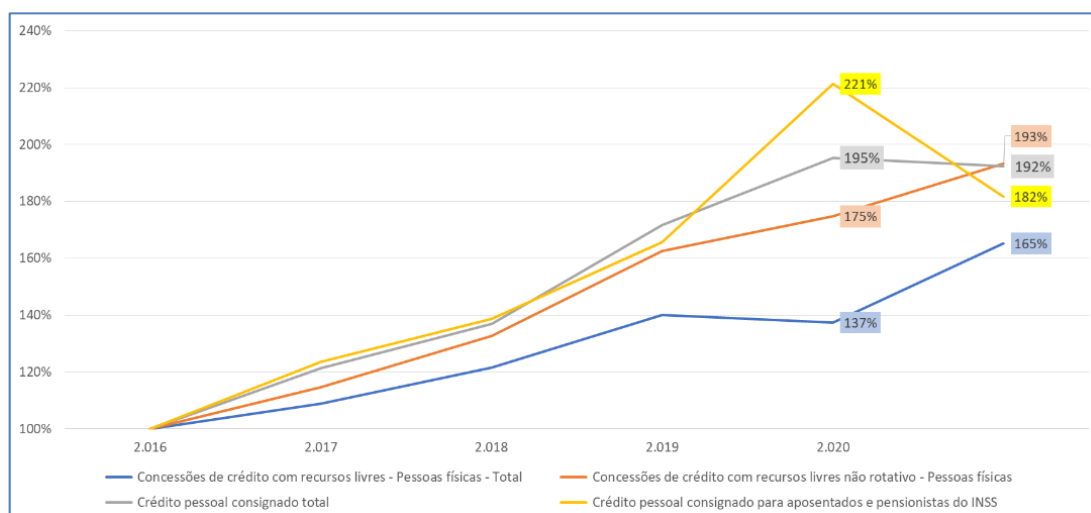
O crédito consignado trata-se de relevante modalidade de crédito para pessoa física com recursos livres dentre as possibilidades existentes no mercado financeiro nacional. É uma modalidade de crédito que possui baixo risco de inadimplência, devido ao desconto da parcela direto em folha de pagamento, e baixa burocracia, com facilidade de análise e aprovação. Desta forma essa modalidade de crédito se mostra vantajosa tanto para os tomadores quanto para as instituições financeiras.

O crédito consignado pode ser solicitado por funcionários de empresas privadas, servidores públicos e aposentados e pensionistas do INSS, desde que haja a formalização de um convênio entre a instituição financeira e estes órgãos. As operações de crédito consignado então podem ser agrupadas pelos setores de atuação dos convênios, podendo privados, públicos e o INSS.

Entender a dinâmica e conseguir projetar a evolução do volume de concessão pode trazer ganhos estratégicos para as instituições financeiras, o que por sua vez pode acarretar condições ainda melhores para o público tomador desta modalidade de crédito.

1.1 Justificativa

Conforme mostra o gráfico 1, de janeiro de 2016 até dezembro de 2020, a concessão de crédito não rotativo pessoal com recursos livres para pessoa física (linha laranja) apresentou crescimento maior que o crescimento do total da concessão de crédito para pessoa física (linha azul). Neste mesmo período o crédito consignado total (linha cinza) apresentou crescimento maior que o crescimento do crédito não rotativo e por sua vez o crescimento da concessão de crédito consignado do INSS (linha amarela) foi maior que o crescimento do crédito consignado total.

Gráfico 1 - Crescimento percentual da Concessão de Crédito (ano base 2016)

Fonte: BACEN (2016; 2020).

Em 2021, sob forte efeito da pandemia do Covid19, vemos uma retração da concessão de crédito consignado do INSS enquanto as concessões de crédito com recursos livres total e com recursos livres não rotativos apresentaram crescimento.

Desta forma, é possível inferir o grau de relevância que a modalidade de crédito consignado para os aposentados e pensionistas do INSS tem tomado no cenário de crédito nacional. Trata-se de importante mecanismo de acesso ao crédito no mercado financeiro.

Projetar o comportamento desta modalidade de crédito pode colaborar na definição de metas e na elaboração de estratégias de médio e longo prazos, utilizando as variáveis disponibilizadas no site do Banco Central, identificando a correlação delas com a variável resposta, se necessário, retirar as que não forem significativas, para assim realizar o ajuste do modelo com as variáveis preditoras restantes.

1.2 Objetivo Geral

Utilizar de conceitos de modelagem de séries de temporais para projetar a concessão de crédito consignado para aposentados e pensionistas do INSS ao longo do tempo.

1.3 Objetivos Específicos

- Identificar o comportamento do volume concedido de crédito consignado para aposentados e pensionistas do INSS expressa em R\$ milhões, compreendendo o período de março de 2011 a dezembro de 2021;
- Dentre as variáveis disponibilizadas no site do Banco Central, identificar àquelas que têm maior impacto na concessão do crédito consignado para aposentados e pensionistas do INSS;
- Ajustar modelos estatísticos de fácil aplicação dentro de instituições que auxiliem no pensamento estratégico frente ao mercado.

2 METODOLOGIA

2.1 Conceito de Séries Temporais

Todo o arcabouço teórico utilizado neste trabalho tem como embasamento as análises feitas ao longo do estudo, realizadas de forma criteriosa.

Os dados coletados para a realização do presente trabalho, devido disponibilidade, contemplam o período de março de 2011 a dezembro de 2021, extraídos do Sistema Gerenciados de Séries Temporais (SGS) do site do Banco Central do Brasil (BACEN), tendo como variável resposta a evolução mensal das concessões de crédito consignado para aposentados e pensionistas do INSS (EMCCC), expressa em R\$ milhões, a nível de Brasil.

Para acompanhar a variabilidade desta variável, e tentar fazer previsões em função do passado da série foram selecionados modelos estocásticos chamados de modelos autorregressivos e de médias móveis (ARIMA). Para tanto, a série foi separada em dois períodos, o período amostral, que compreende os meses de março de 2011 até junho de 2021, que servirá para ajustar os modelos estatísticos, e o período que abrange os seis meses finais de série, de julho de 2021 a dezembro do mesmo ano, será utilizado na avaliação da capacidade de predição futura do modelo.

Para avaliar a associação de variáveis econômicas na variabilidade da série EMCCC, foram selecionadas 8 variáveis como possíveis explicativas, pois possuem correlação em teoria, que posteriormente será avaliada essa correlação com medidas estatísticas adequadas, se, que também tiveram como fonte de dados o SGS do BACEN. As variáveis analisadas como possíveis explicativas foram:

- Taxa média mensal de juros das operações de crédito consignado INSS (TMMJ OCC) - (% a.m.);
- Prazo médio mensal das concessões de crédito consignado INSS (PMMCCC) - (Meses);
- Inadimplência da carteira de crédito consignado INSS (INAD CCC) - (%);
- Percentual da carteira de crédito consignado INSS com atraso entre 15 e 90 dias (PCCC ATRASO) - (%);
- PIB mensal - Valores correntes (R\$ milhões);
- Taxa básica de juros - Selic acumulada no mês (SELIC) - (% a.m.);
- Salário-mínimo - (R\$);
- Índice nacional de preços ao consumidor-amplo (IPCA) - (Var. % mensal).

O objetivo na escolha destas variáveis foi o de levantar a relação de indicadores econômicos abrangentes com a variável resposta. Através do uso de técnicas de análise exploratória iremos analisar primeiramente cada variável de forma individual, visando entender seu comportamento. Posteriormente, será analisada também a relação da variável dependente com cada uma das variáveis candidatas a variável explicativa e através de análises de correlações e do fator de inflação da variância (VIF), onde iremos apontar qual ou quais variáveis reúne(m) condições de melhor explicar o comportamento da variável resposta EMCCC, sendo as demais variáveis descartadas.

2.1.1 Definição de Séries Temporais

Fisher (1982), nos diz que uma série temporal é um conjunto de observações de uma variável dispostas sequencialmente no tempo. Podemos dizer então que se trata de um conjunto de dados observados em intervalos de tempo equidistantes e que apresentam uma dependência serial entre eles.

Uma série é representada como segue:

$$\{Y_t; t = 1, 2, 3 \dots\}, \{Y_t\} \text{ ou } Y_t \quad (1)$$

Conforme o conjunto obtido, pode-se classificar a série em contínua ou discreta. Uma série é discreta quando o conjunto de observações no tempo for finito ou infinito

enumerável, isto é, possui características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores e, assim, somente fazem sentido valores inteiros. Geralmente são o resultado de contagens, como por exemplo número de filhos ou número de ações vendidas de uma empresa. Porém a série será contínua, quando o conjunto obtido for infinito não enumerável, a série é contínua, isto é, possui características mensuráveis que assumem valores em uma escala contínua (na reta real), para as quais valores fracionais fazem sentido. Usualmente devem ser medidas através de algum instrumento. Exemplos: peso (balança), altura (régua) e tempo (relógio). A maioria dos estudos que se valem do tratamento de séries temporais, utilizam séries discretas, onde as observações são geradas em um intervalo de tempo com amplitude constante.

Morettin e Tolo (2006) enumeram alguns exemplos de séries temporais:

1. Valores diários de poluição na cidade de São Paulo;
2. Valores mensais de temperatura na cidade de Cananéia-SP;
3. Índices diários da Bolsa de Valores de São Paulo;
4. Precipitação atmosférica anual na cidade de Fortaleza;
5. Número médio anual de manchas solares;
6. Registro mensal das marés no porto de Santos com o uso de um marégrafo.

Com relação a classificação das séries acima listadas, pode-se dizer que as séries 1, 2, 3, 4 e 5 são temporais discretas, enquanto a série 6 é contínua.

Pode-se dizer que o que chamamos de série temporal é uma trajetória, dentre muitas que poderiam ter sido observadas. Essas trajetórias podem ser expressas em curvas, que geralmente, não são idênticas. Estas curvas são chamadas de trajetórias do processo físico que está sendo observado e este (o processo estocástico) nada mais é do que o conjunto de todas as possíveis trajetórias que podem ser observadas. Cada trajetória é também chamada uma série temporal.

2.1.2 Processo Estocástico

Um processo estocástico é uma **família de variáveis aleatórias** indexadas no tempo, que denotamos como segue:

$$\{ Y_t, t = 1, 2, \dots \}$$

Assumindo valores no espaço de índices T, pode ser discreto ou contínuo.

2.1.3 Processo Estocástico Estacionário

Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que a série seja estacionária. Isso quer dizer que ela se desenvolve aleatoriamente no tempo, em torno de uma média constante, refletindo alguma forma de equilíbrio (Morettin; Toloi, 2006). Isso significa dizer que o comportamento da série não se altera com o passar do tempo, e sua média não muda ao longo do tempo.

Para que as inferências estatísticas sejam feitas a cerca de um processo estocástico, se faz necessário que o processo mantenha suas propriedades estatísticas constantes ao longo do tempo, ou seja, o processo precisa estar em um estado de equilíbrio estatístico, (Box; Jenkins; Reinsel, 2008).

Segundo Ehlers (2007), o processo estacionário é uma importante classe dos processos estocásticos podendo ser estritamente estacionário ou fracamente estacionário.

Diz-se que um processo estocástico é estacionário quando o processo apresenta o equilíbrio estatístico, onde a distribuição de probabilidades de Y_t não é afetada pelo tempo. Em suma, para que um processo estocástico apresente estacionariedade é necessário que a média do processo estocástico seja constante, que a variância também seja constante e a covariância entre Y_t e Y_{t-k} dependente da variável k. Isso pode ser expresso na expressão abaixo:

$$\begin{aligned} E(Y_t) &= \mu & (2) \\ \text{Var}(Y_t) &= E[(Y_t - \mu)^2] = \sigma^2 \\ \text{Cov}(Y_t, Y_{t-k}) &= E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma_k \end{aligned}$$

Em geral, é mais comum a existência de séries apresentando crescimento e/ou decréscimo conforme o tempo passa, especialmente na economia, onde é pouco provável encontrar uma série originada no tempo sob forma linear. “Um simples exame de algumas das variáveis mais comuns na economia, como por exemplo, preço, produto nacional, renda, vendas, consumo, investimento e tantas outras, mostra que suas variações apresentam um comportamento não estacionário evolutivo no decorrer do tempo” (Fisher, 1982).

Em Morettin e Tolo (2006), vemos que a maioria dos procedimentos de análise estatística de séries temporais supõem que estas sejam estacionárias. A necessidade de transformar os dados originais se dá quando a série é não estacionária, e basicamente há duas razões para se transformar os dados originais: para estabilizar a variância e para tornar aditivo o efeito sazonal, que pode ser definido como efeitos que afetam a série de forma cíclica ao longo de um determinado período.

Nos casos, em que a série não é estacionária em nível, pode-se aplicar d diferenças sucessivas até obter uma série estacionária. Desta forma uma diferenciação de ordem d pode ser expressa assim:

$$Z_t = (1 - B)^d Y_t \quad (3)$$

O operador de defasagem B é o termo usado para designar o número de períodos defasados a uma observação precedente. O operador de defasagem B por k pode ser definido por:

$$B^k Y_t = Y_{t-k}, \quad (4)$$

Onde B^k representa a defasagem de Y_t por k períodos.

2.1.4 Função de Autocorrelação e Função de Autocorrelação Parcial

A função de autocorrelação (FAC) mede o grau de correlação de uma variável, em um dado instante, consigo mesma, em um instante de tempo posterior, ela permite que se analise o grau de irregularidade de um sinal. A função de autocorrelação parcial (FACP) é a correlação entre as observações em uma série temporal que não é

contabilizada por todos os intervalos mais curtos entre as observações. Onde a FAC e FACP auxiliam no ajuste do modelo e comportamento de séries temporais. A correlação entre duas variáveis aleatórias Y_t e Y_{t-k} distanciadas por k períodos é chamada de autocorrelação de ordem k , assim temos:

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)Var(Y_{t-k})} = \frac{\gamma_k}{\gamma_0} \quad (5)$$

Onde ρ_k , representa a função de autocorrelação, $Cov(Y_t, Y_{t-k})$ a autocovariância entre Y_t e Y_{t-k} , por fim o termo $Var(Y_t)$ se refere a variância da variável Y_t .

Frente a suposição de estacionariedade temos as propriedades da função de autocorrelação, sendo $\rho_0 = 1$. Para todo k , a autocorrelação compreende $|\rho_k| \leq 1$ e $\rho_{-k} = \rho_k$. Bartleet (1946) mostra que, para processos estacionários, a FAC estimada segue uma distribuição $\hat{\rho}_k \sim Normal(0, S_{\hat{\rho}_k})$, sendo $\hat{\rho}_k$ uma estimativa da autocorrelação de ordem k , com isso o desvio padrão $S_{\hat{\rho}_k}$ é definido por:

$$S_{\hat{\rho}_k} \cong \sqrt{\frac{1}{n} (1 + 2\hat{\rho}_1^2 + \dots + 2\hat{\rho}_{k-1}^2)} \quad (6)$$

Para verificarmos a estrutura de dependência entre Y_t e Y_{t-k} , as informações da função de autocorrelação devem ser associadas ao conceito de autocorrelação parcial. Assim, depois de removidas as dependências lineares das variáveis intermediárias $Y_{t+1}, Y_{t+2}, Y_{t+3}, \dots, Y_{t+(k-1)}$, se obtém a correlação condicional, que é a autocorrelação parcial. Ela funciona de forma a mensurar a correlação linear entre Y_t e Y_{t-k} após a remoção dos efeitos das outras defasagens 1, 2, ..., $k-1$. A autocorrelação parcial no *lag* k pode ser estimada a partir do coeficiente de regressão:

$$\begin{aligned} Y_t &= \phi_{k1}Y_{t-1} + \dots + \phi_{kk}Y_{t-k} + a_t \\ \phi_{kk} &= Corr(Y_t, Y_{t+k} | Y_{t+1}, \dots, Y_{t+k-1}) \end{aligned} \quad (7)$$

A estimação da função de autocorrelação parcial com distribuição normal sendo $\hat{\phi}_{kk} \sim Normal(0, S_{\hat{\phi}_{11}})$, com o desvio padrão S definido por:

$$S_{\phi_{kk}} \cong \sqrt{\frac{1}{n}} \quad (8)$$

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{22} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & \hat{\rho}_2 \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{vmatrix}}$$

2.1.5 Processo Ruído Branco Gaussiano

Um dos processos estocásticos mais utilizados no estudo de séries temporais é justamente o ruído branco gaussiano. Ele indica uma sequência de variáveis aleatórias independentes e igualmente distribuídas com média igual zero e variância constante e não correlacionadas. Quando as variáveis aleatórias seguem uma distribuição normal, é dito que trata-se de um ruído branco gaussiano, desta forma, satisfazendo: $a_t = N(\mu, \sigma_a^2), \forall t = 1, 2, \dots$. Assim temos $E(a_t) = \mu = 0$, $Var(a_t) = \sigma_a^2$ e $Cov(a_t, a_{t-k}) = 0$.

Ao se trabalhar com séries temporais, um dos principais objetivos é ajustar modelos onde resíduos tenham correlações iguais a zero, para que se tenha o que é denominado de processo ruído branco.

2.2 Modelos de Séries Temporais

2.2.1 Modelos ARMA

A inclusão de termos autorregressivos e de médias móveis surgiu como uma alternativa para simplificar um modelo com grande número de parâmetros (Morettin; Tolo, 2006). Considerando uma série temporal Y_t estacionária que apresenta uma estrutura de dependência serial entre observações, o objetivo será encontrar o melhor modelo que descreva esta dependência. Um modelo de regressão em séries temporais pode ser descrito conforme abaixo:

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \dots + b_p X_{pt} + a_t \text{ para } t = 1, 2, \dots, n \quad (9)$$

Onde Y_t pode ser representado por uma função das variáveis X_{1t}, X_{2t}, X_{pt} que são variáveis explicativas medidas no tempo t , $b_0, b_1 \dots b_p$ são os coeficientes da regressão linear e a_t é o termo do erro aleatório.

Em alguns casos a introdução de variáveis explicativas podem auxiliar na construção de melhores modelos e se as variáveis explicativas são representadas pela a mesma variável Y_t defasada no tempo, assim temos que: $X_{1t} = Y_{t-1}$, $X_{2t} = Y_{t-2}$, ..., $X_{pt} = Y_{t-p}$. Assim o modelo pode ser representado da seguinte forma:

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + a_t \text{ para } t = 1, 2, \dots, n \quad (10)$$

Este modelo é chamado de modelo auto-regressivo. Nele a suposição básica da independência dos erros (resíduos) pode ser facilmente violada, desde que as variáveis explicativas ou regressoras usualmente tem uma relação de dependência. Se as variáveis explicativas são representadas pelos erros defasados no tempo, temos o modelo com seguinte equação:

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (11)$$

Aqui se evidencia a relação de dependência entre os sucessivos erros defasados no tempo. Este modelo é chamado modelo de médias móveis.

Se nos modelos auto-regressivos é possível acoplar os modelos de médias móveis, se chega a uma forma geral desta classe de modelos, chamados de modelos auto-regressivos e de médias móveis.

Um modelo que representa a série temporal Y_t dependendo da série defasada ao passo p atrás Y_{t-p} é chamado de modelo auto-regressivo de ordem p , ou AR(P), dado pela seguinte equação:

$$Y_t = \theta_0 + \phi Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad (12)$$

Onde o termo constante $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ μ é a média do processo; ϕ_j é o j -ésimo parâmetro autoregressivo, $\{a_t\}$ é um processo Ruído Branco com média zero $E(a_t) = 0$ e variância constante $Var(a_t) = \sigma^2$.

A seguir vemos as restrições dos parâmetros auto-regressivos que definem a condição de estacionariedade dos modelos:

- Para $p = 1$, $-1 < \phi < 1$
- Para $p = 2$, as seguintes três condições definem a região de estacionariedade:

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad e - 1 < \phi_2 < 1$$

Quando $p > 2$, as condições para os coeficientes da FAC são mais complexas. Considerando os sinais dos coeficientes, as correlações podem apresentar decaimento alternado. A função FACP apresenta as p primeiras correlações significativamente diferentes de zero.

De maneira geral, espera-se que a função de autocorrelação FAC de uma série temporal estacionária que seja modelada por um processo auto regressivo AR(p) amortecida para zero, isto é, espera-se que a série decaia exponencialmente. Já a função de autocorrelação parcial FACP, espera-se que seja aproximadamente zero para todas as ordens superiores à ordem p do processo, isto é, espera-se que a série seja truncada (cortada) na defasagem p .

Um modelo de médias móveis de ordem q pode ser descrito pela relação a seguir:

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (13)$$

Onde: μ é a média do processo; θ_j é o j -ésimo parâmetro da componente médias móveis; $\{a_t\}$ é um processo ruído branco com média zero $E(a_t) = 0$ e variância constante $Var(a_t) = \sigma^2$.

Abaixo vemos as restrições dos parâmetros θ_i que definem a condição de invertibilidade dos modelos:

- Para $q = 1$, $-1 < \theta_1 < 1$
- Para $q = 2$, as seguintes três condições definem a região de estacionariedade:

$$\theta_1 + \theta_2 < 1, \quad \theta_2 - \theta_1 < 1, \quad e - 1 < \theta_2 < 1$$

Para $q > 2$ as condições para os coeficientes são mais complicadas. Para um processo estacionário de médias móveis MA(q), em geral, espera-se que os q

primeiros coeficientes da função de autocorrelação simples (FAC) sejam diferentes de zero, e os restantes iguais a zero. Já os coeficientes da função de autocorrelação parcial (FACP), devem apresentar um decaimento amortecido para zero.

O comportamento das funções FAC e FACP de um processo de média móvel de ordem 1 (MA(1) puro) são exatamente ao contrário das de um processo AR(1) puro. Ou seja, a FAC de um processo MA(1) é nula após o *lag* 1 e a FACP decai exponencialmente para zero (as autocorrelações parciais não caem bruscamente, mas, decrescem, amortecendo gradualmente para zero, mas as autocorrelações devem ser iguais a zero a partir da segunda, inclusive).

Os modelos autorregressivos são muito populares em algumas áreas como a Economia, onde é natural pensar o valor de alguma variável no instante t como função de valores defasados da mesma variável.

Segundo Morettin e Tolo (2006), para muitas séries encontradas na prática, se quisermos um modelo com um número não muito grande de parâmetros, a inclusão de termos autorregressivos e de médias móveis é a solução mais adequada.

Aos modelos autorregressivos pode-se efetivamente ser acoplados os modelos de médias móveis:

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (14)$$

Fazendo uso do operador de retardo B temos outra forma de representar os modelos ARMA(p,q):

$$\Phi(B)Y_t = \Theta_0 + \Theta(B)a_t \quad (15)$$

Onde:

- Θ_0 é a constante do processo;
- $\Phi(B) = (1 - \theta_1 B - \dots - \theta_p B^p) = 0$ é o polinômio auto-regressivo de ordem “p”, condição de estacionariedade do processo;
- $\Theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0$ é o polinômio de médias móveis de ordem “q”, condição de invertibilidade do processo.

- $\{a_t\}$ é um processo Ruído branco com média zero $E(a_t) = 0$ e variância constante $Var(a_t) = \sigma^2$. Este modelo é uma forma geral da classe de modelos chamados, modelos auto-regressivos e de médias móveis.

2.2.2 Diagnósticos de Modelos

O teste de Portmanteau é comumente usado na análise de séries temporais para avaliar a qualidade do ajuste (Wei,1994). As estatísticas de teste de Portmanteau são baseadas em autocorrelações residuais e se o modelo ajustado é considerado uma boa representação dos dados quando tais autocorrelações são conjuntamente desprezíveis. As hipóteses nula e alternativa do teste de Portmanteau são:

$$\begin{cases} H_0: \rho_1 = \rho_2 = \dots = \rho_i = 0 \\ H_1: \text{pelo menos um } \rho_i \neq 0 \end{cases}$$

A rejeição da hipótese nula indica que há evidências de má especificação do modelo.

2.2.3 Modelos ARIMA

O modelo ARIMA (Auto Regressivo Integrado de Média Móvel), conhecido também como metodologia Box-Jenkins, é próprio para previsões de séries temporais. Este modelo é popular em consequência de sua generalidade, ele pode lidar com séries estacionárias ou não.

Comumente, não será necessário calcular mais de uma diferença para que a série se torne estacionária. Assim, podemos dizer que uma diferenciação de ordem “d” aplicada à série não estacionária, $Z_t = (1 - B)^d Y_t$ se torna estacionária e produz a série estacionária Z_t que pode ser estudada por modelos ARMA(p,q):

$$\Phi_p(B)Z_t = \Theta_0 + \Theta_q(B)a_t \quad (16)$$

Então o modelo para a série Y_t é da forma, onde:

$$\Phi_p(B)(1 - B)^d Y_t = \Theta_0 + \Theta_q(B)a_t \quad (17)$$

Onde:

- $\theta_0 = \mu(\phi_1 - \phi_2 - \dots - \phi_p)$ é a constante do processo.
- $\Phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p) = 0$ é o polinômio auto-regressivo de ordem “p” com raízes fora do círculo unitário, condição de estacionariedade do processo;
- $\Theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0$ é o polinômio de médias móveis de ordem “q” com raízes fora do círculo unitário, condição de invertibilidade do processo.
- $\{a_t\}$ é um processo Ruído branco com média zero $E(a_t) = 0$ e variância constante $Var(a_t) = \sigma^2$. Este modelo é uma forma geral da classe de modelos chamados, modelos auto-regressivos integrados de médias móveis ARIMA(p,d,q).

2.2.4 Modelos SARIMA

A complexidade final que pode ser introduzida aos modelos ARIMA visa tratar o problema sazonalidade. Box and Jenkins (1970) generalizaram o modelo ARIMA sazonal, denominado SARIMA, dado por:

$$\phi(B)\Phi(B^s)Z_t = \theta(B)\Theta(B^s)a_t \quad (18)$$

onde:

$$\begin{aligned} \phi(B) &= (1 - \phi_1 B - \dots - \phi_p B^p) \\ \Phi(B^s) &= (1 - \Phi_s B^s - \dots - \phi \Phi_p B^{Ps}) \\ Z_t &= (1 - B)^d Y_t \\ \theta(B) &= (1 + \theta_1 B + \dots + \theta_q B^q) \\ \Theta(B^s) &= (1 + \Theta_s B^s + \dots + \Theta_Q B^{Qs}) \end{aligned}$$

Segundo Ehlers (2007), este modelo é chamado SARIMA multiplicativo de ordem $(p, d, q) \times (P, D, Q)_s$. Apesar de à primeira vista parecer complicado na prática, os valores de d e D em geral não serão maiores do que 1 e um número pequeno de coeficientes será suficiente. Por exemplo, com P=1 tem-se que:

$$\Phi(B^s) = (1 - \Phi_s B^s) \quad (19)$$

Isso significa que Z_t depende de Z_{t-s} . A série Z_t é formada a partir da série original tomando-se diferenças simples para remover a tendência e diferenças sazonais para remover a sazonalidade. O modelo SARIMA (1,0,0)(0,1,1)[12] para dados mensais, apresenta um termo autorregressivo e um termo média móvel sazonal modelando a primeira diferença sazonal. Segundo Fisher (1982), SARIMA é a simbologia para um modelo ARIMA sazonal multiplicativo de ordem $(p, d, q) \times (P, D, Q)_s$. O modelo pode ser escrito como:

$$(1 - \phi B)(1 - B^s)Z_t = (1 - \theta B^s)a_t \quad (20)$$

E assim desenvolvendo os produtos obtém-se:

$$Z_t = Z_{t-s} + \phi Z_{t-1} - \phi Z_{t-(s+1)} + a_t + \phi a_{t-s} \quad (21)$$

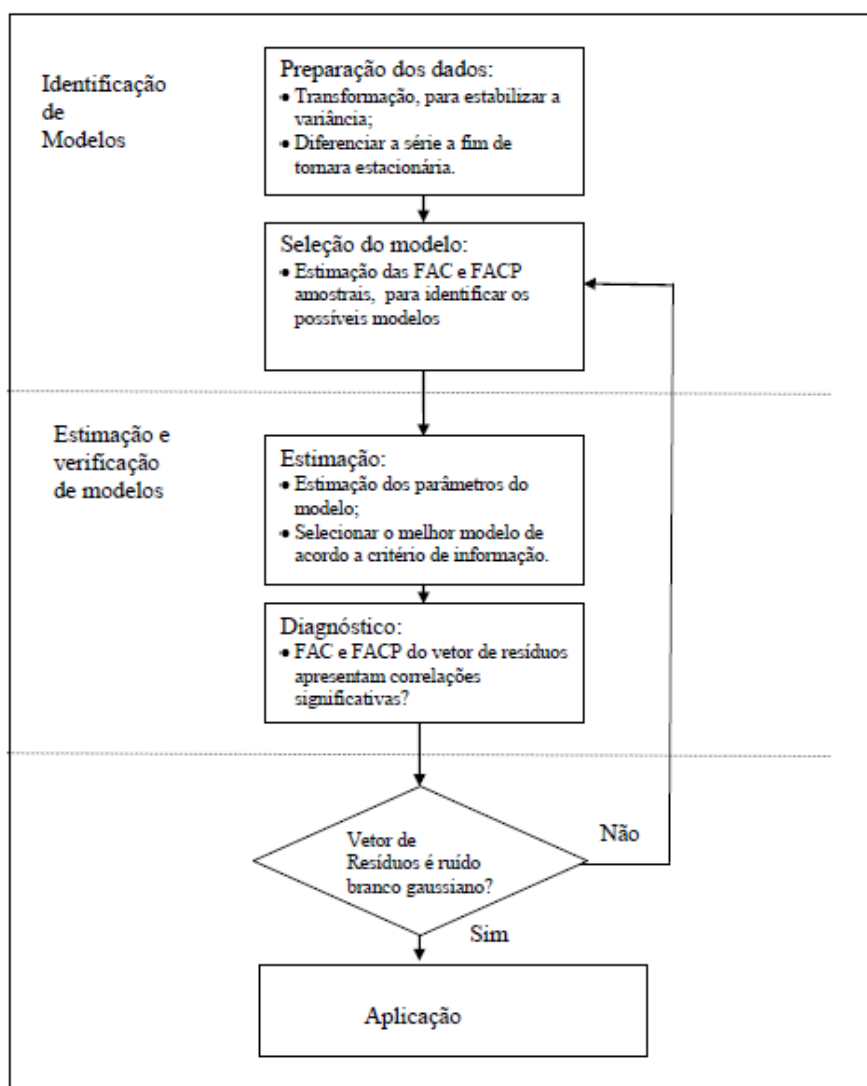
Assim, Z_t depende de Z_{t-1} , Z_{t-s} e $Z_{t-(s+1)}$ além do erro no tempo $t-s$. Para finalizar, ao ajustar um modelo sazonal aos dados, a primeira tarefa é especificar os valores de d e D que tornam a série (aproximadamente) estacionária e remove a maior parte da sazonalidade.

2.2.5 Metodologia Box & Jenkins

Segundo Fisher (1982), o tratamento desenvolvido por Box e Jenkins (1970) para a análise das séries de temporais está baseado no fato de que, embora seus respectivos valores no tempo x , apresentem correlação serial, cada um deles pode ser considerado como gerado por uma sequência de choques " a_t ", $t \in T$, aleatórios e independentes entre si, cada um possuindo uma determinada distribuição, com $\mu_{(t)} = 0$ e variância constante σ_a^2 .

Box e Jenkins (1970) sugere o procedimento descrito na Figura 1 visando facilitar a modelagem de séries temporais usando os modelos ARIMA.

Figura 1 - Fluxograma Box-Jenkins para Modelagem ARIMA



Fonte: Box e Jenkins (1970).

2.3 Modelos Multivariados de Séries Temporais

2.3.1 Regressão Dinâmica

Para Stellwagen e Goodrich (2011), a regressão dinâmica é um modelo de regressão que combina a dinâmica de séries temporais e os efeitos de variáveis explicativas. O modelo incorpora defasagens da variável dependente e também defasagens das variáveis exógenas e do termo autoregressivo de erro (Hyndman e Athanasopoulos, 2016).

Segundo Barros e Souza (1995), ao se tentar modelar uma série temporal através de um modelo de regressão, a hipótese de independência dos ruídos não é realista,

e os resultados e testes estatísticos usados nos modelos de regressão não são válidos. Por exemplo, segundo os autores, algumas das consequências da autocorrelação dos resíduos são:

1. Os estimadores usuais por mínimos quadrados são ainda não tendenciosos, mas não têm variância mínima.
2. Os estimadores da variância e dos erros padrões dos coeficientes da regressão são subestimados, o que levaria à conclusão de que os estimadores são mais precisos do que na realidade.
3. Os intervalos de confiança para os parâmetros da regressão e os testes de hipóteses relacionados a estes intervalos perdem a validade, como uma consequência direta de 2.

Desta forma, estes três motivos implicam na necessidade de procurar procedimentos para tratar o problema de autocorrelação dos erros, pois ignorá-los leva, em geral, a inúmeras conclusões falsas.

Em particular, dado que a suposição de independência dos erros nem sempre é satisfeita no contexto de séries temporais, os modelos de regressão dinâmica estendem os modelos usuais de regressão ao levantarem esta restrição.

Na escolha do modelo apropriado, deve-se levar em consideração não apenas a significância dos parâmetros, mas também sua consistência com a teoria econômica. Por exemplo, a concessão de crédito é afetada pela taxa. Um aumento na taxa tende a reduzir a busca por crédito. Logo, se o modelo de regressão encontrado para explicar a concessão de crédito pela variação da taxa apresenta coeficiente positivo para a variável taxa, deve-se desconfiar por mais bem ajustado que esteja o modelo, porque, em geral, a relação sugerida pelo modelo não é coerente. Ou seja, na escolha de um modelo de regressão dinâmica, não é necessário apenas encontrar um ajuste de parâmetros adequado, mas fundamentalmente faz-se imprescindível verificar a coerência dos coeficientes estimados.

Segundo Goodrich (1989), modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente indicar que não se pode supor a independência dos erros.

Os modelos de regressão dinâmica podem ser descritos pela equação:

$$\phi(B)Y_t = \beta X_t + a_t \quad (22)$$

Onde:

Y_t = variável dependente no tempo t

β = vetor de coeficientes das variáveis causais, que será estimado por máxima verossimilhança.

X_t = vetor de variáveis causais (exógenas) no instante t

a_t = ruído aleatório associado ao modelo, onde supomos que os a_t são independentes e identicamente distribuídos com distribuição $N(0, \sigma^2)$

$\phi(\beta)$ = polinômio autoregressivo de ordem p , isto é:

$\phi(\beta) = 1 - \phi_1 B_1 - \phi_2 B_2 - \dots - \phi_p B_p$ sendo B o operador de atraso.

A estrutura do modelo de regressão dinâmica permite considerar como elementos X_t tanto variáveis causais como também suas defasagens.

A presença do polinômio $\phi(B)$ no modelo traz uma grande flexibilidade desta classe de modelos, mas, ao mesmo tempo, dificulta a procura por um modelo adequado. Observa-se que, se $\phi(B) = 1$, não existem defasagens da variável dependente e a interpretação do modelo é muito simples, pois as variáveis causais influenciam diretamente a variável que será explicada pelo modelo. Ao contrário, quando $\phi(B) \neq 1$, o modelo pode ser usado para representar relações mais complicadas.

Para a construção de um modelo de regressão dinâmica parte-se de um modelo simples, com poucas variáveis, refina-o, utilizando métodos estatísticos para verificar a adequação das incógnitas e inclui novas variáveis até encontrar um modelo apropriado. Às vezes, no entanto, o impacto de uma variável incluída em um modelo de regressão dinâmica não será simples e imediato (Hyndman e Athanasopoulos, 2016). Por isso, a construção de um modelo de regressão dinâmica é muitas vezes um procedimento difícil, porque além de escolher as variáveis a serem incluídas no modelo, também é necessário definir os lags (defasagens) destas variáveis.

As previsões geradas por um modelo de regressão dinâmica dependem não só de valores passados da própria série, mas também dos valores previstos para as variáveis causais. Logo, para se obter as previsões da série Y_t para $Y_{t+1}, Y_{t+2}, Y_{t+3}$ etc,

é necessário fornecer ao modelo os valores futuros do vetor de variáveis causais X_t . Logo, as previsões destas variáveis causais devem ser adequadas para que o modelo de regressão dinâmica não gere valores futuros para Y_t inadequados. Por outro lado, esta condição possibilita a criação de cenários futuros para as variáveis causais, possibilitando analisar os possíveis valores futuros da variável de interesse em diversas condições.

Diferentemente dos modelos univariados, onde a sazonalidade e os ciclos são fenômenos limitados à variável dependente, no modelo de regressão dinâmica, a sazonalidade e os ciclos podem ser tratados de forma muito flexível e distinta, vinculada a uma ou mais variáveis, endógenas ou exógenas, através das defasagens. Não obstante, a sazonalidade e os ciclos também podem ser introduzidos através de variáveis *dummy*. Esta flexibilidade se traduz em grande potencial de adaptação às condições e/ou expectativas para a construção e análise do modelo funcional.

2.3.2 Modelos ARIMAX

Como o objetivo do presente estudo é estimar com boa precisão a evolução do volume de crédito consignado no INSS levando em consideração variáveis econômicas de domínio público iremos buscar um modelo que leva em consideração a influência de variáveis preditoras.

A metodologia de Box-Tiao, a qual é comumente chamada de ARIMAX (Box e Tiao, 1975), representa uma expansão dos modelos Auto-Regressivos Integrados à Média Móvel (ARIMA) da metodologia Box-Jenkins (Box *et al.*, 2015). O modelo ARIMAX é considerado multivariado que acrescenta um componente linear em função das observações das covariáveis (também chamadas de variáveis exógenas). A principal diferença entre os dois modelos é que o ARIMAX possui, além dos parâmetros auto-regressivos e de médias móveis, a entrada exógena (Bennet, 2014). O modelo ARIMAX pode ser entendido como a combinação dos modelos Auto-Regressivo AR(p), Integrado (d), Média-Móvel MA(q) e Exógeno X(r), e que pode então ser simbolizado por ARIMAX(p,d,q,r).

- AR representa uma regressão da variável contra valores passados
- MA representa a média móvel ponderada dos erros de previsão anteriores

- o X_t representa uma variável ou um conjunto de variáveis explicativas, com isso, o termo r corresponde à ordem da equação que representa a entrada exógena.

De forma simplificada, pode-se representar o modelo por:

$$Y_t = b + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^r \beta_j X_{t-j} + \sum_{j=1}^q \theta_j a_{t-j} \quad (23)$$

Em que Y_t é variável dependente no tempo t e que se pretende projetar; b é uma constante; Y_{t-i} é a variável dependente defasada por i passos de tempo; ϕ_i é o coeficiente de Y_{t-i} , p é o número máximo de intervalos de tempo; x_j representa as variáveis exógenas; β_j representa os coeficientes das variáveis exógenas; r é o número máximo de variáveis exógenas; θ_j é o coeficiente do termo de a_{t-j} que, por sua vez, representa o erro no tempo t defasado de j , e a_t é a componente de erro do modelo, com $a_t \sim N(0, \sigma^2)$. Os coeficientes dos modelos são estimados por regressão linear. Material para aprofundamento pode ser encontrados em (Yogarajah *et al.*, 2013).

Nessa monografia as ordens dos modelos ARIMAX, ou seja, os números que darão origem ao funcionamento das funções foram encontrados através da função, `auto.arima` e a escolha do melhor modelo ARIMAX para fornecer os melhores ajustes aos dados observados levou em consideração alguns critérios que serão melhor detalhados na seção 4.4. Na prática, através do uso destes critérios buscou-se encontrar o modelo mais parcimonioso, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta.

Para avaliar se os modelos da metodologia de Box-Tiao, ajustados possuem viabilidade para a previsão do volume de crédito consignado no INSS foi realizado uma análise de resíduos. Conforme Shumway e Stoffer (2006), para que os modelos ARIMAX sejam viáveis no tocante aos ajustes dos dados observados, o termo de erro ε_t que representa uma média da diferença entre a função estimada e os valores reais, desse modelo deve se comportar como um ruído branco, ou seja, possuir média zero, variância constante (são homocedásticos) e os erros são independentes. E ainda o termo ε_t deve seguir uma distribuição normal. Para identificação desses pressupostos

existem alguns testes estatísticos possíveis de serem aplicados sobre os resíduos estimados.

2.4 Medidas de Acurácia

Em muitos casos, no estudo de séries temporais, é possível encontrar vários modelos adequados no que se refere ao comportamento dos resíduos estimados. Uma forma de comparação entre estes modelos é utilizar os chamados critérios de informação que levam em conta não apenas a qualidade do ajuste, mas também penalizam a inclusão de parâmetros extras. Assim, um modelo com mais parâmetros pode ter um melhor ajuste, mas não necessariamente será preferível em termos de critério de informação (Ehlers, 2007). Pretende-se selecionar o modelo cujo critério de informação calculado seja mínimo.

2.4.1 Critério de Informação de Akaike (AIC)

A regra mais utilizada em séries temporais é o chamado critério de informação de Akaike (1973). Esse critério é baseado na identificação do menor valor para o coeficiente AIC (*Akaike Information Criterion* do inglês cuja a tradução em português é Critério de Informação Akaike), o qual teoricamente resulta no melhor modelo do ponto de vista estatístico (Brockwell e Davis, 1991). Geralmente é dado por:

$$AIC = -2\log L(\hat{\theta}) + 2(m) \quad (24)$$

Onde m é o número de parâmetros a serem estimados pelo modelo e $L(\hat{\theta})$ a função de máxima verossimilhança. Para dados distribuídos normalmente e utilizando-se estimativas de máxima verossimilhança dos parâmetros, tem-se que:

$$AIC = -2\log(\hat{\sigma}_a^2) + 2m \quad (25)$$

Onde $\hat{\sigma}_a^2 = \left(\frac{1}{n}\right) \Sigma \hat{a}_t^2$.

A presença de m na fórmula da estatística AIC, tem por objetivo penalizar os modelos com muitos parâmetros, tendo em vista que modelos mais parcimoniosos devem ser privilegiados, por apresentarem menor número de parâmetros a serem estimados.

2.4.2 Critério de Informação de Bayesiano (BIC)

A comparação de modelos também poderá ser realizada com o Critério de Informação Bayesiano (BIC) como aproximações do fator de Bayes. O BIC é basicamente uma modificação do AIC, na forma de penalizar a inclusão de parâmetros extras e é descrito por:

$$BIC = -2 \log f(x_n | \theta) + m[\log n] \quad (26)$$

Onde $f(x_n | \theta)$ é o modelo escolhido, m é o número de parâmetros a serem estimados e n é o número de observações da amostra. Igualmente ao critério AIC, o modelo mais adequado será aquele que tiver o menor valor da estatística BIC. Observa-se que este critério de informação penaliza bem mais a inclusão de parâmetros do que o AIC e, portanto, tende a selecionar modelos mais parcimoniosos. (Ehlers, 2007)

2.4.3 Estatísticas de Erros

Seja a série temporal Y_t $t \in T$, modelada por uma abordagem que resultou uma estimativa dos valores ajustados \hat{Y}_t por esta abordagem. Existem algumas medidas de precisão geralmente usadas, cuja escala depende da escala dos dados. Estes são úteis na comparação de métodos diferentes com o mesmo conjunto de dados, mas não deve ser usado, por exemplo, entre os conjuntos de dados que tem escalas diferentes. As escalas de medidas dependentes mais utilizadas baseiam-se no erro absoluto ou erros quadrados: A acuidade do método de previsão utilizado é comumente mensurada pelas medidas listadas a seguir:

- Erro Quadrado Médio (MSE) = $\frac{1}{n} \sum_{j=T+1}^n e_j^2$
- Raiz do Erro Quadrático Médio (RMSE) = \sqrt{MSE}
- Erro Absoluto Médio (MAE) ou Desvio Absoluto Médio (MAD) = $\frac{1}{n} \sum_{j=T+1}^n |e_j|$
- Erro Percentual Absoluto Médio (MAPE) = $\left[\frac{1}{n} \sum_{j=1}^n \left| \frac{e_j}{y_i} \right| \right] 100\%$

Estas medidas são automaticamente calculadas nos principais softwares dedicados a análises de séries temporais, inclusive no R, que foi o software utilizado na execução deste trabalho.

3 ANÁLISE DE DADOS

3.1 Análise Exploratória das Séries de Dados

Como mencionado na seção 2.1, foram extraídas do banco de séries temporais do Bacen 8 possíveis variáveis explicativas visando, através de critérios estatísticos, levantar qual ou quais poderiam ajudar a explicar a variabilidade da série EMCCC, no período de março de 2011 a dezembro de 2021, onde os dados são distribuídos mensalmente. Trata-se então de dados de domínio público.

3.1.1 Evolução Mensal das Concessões de Crédito Consignado do INSS (EMCCC)

Iniciaremos a análise exploratória dos dados pela variável resposta. Está expressa em R\$ milhões e compreende o período de março de 2011 a dezembro de 2021. Trata-se do volume de desembolsos referente as operações mensais de concessão de empréstimos para aposentados e pensionistas do Instituto Nacional de Seguro Social (INSS), na modalidade de crédito consignado, isto é, com retenção de parte do benefício do contratante pelo INSS para o pagamento das parcelas do empréstimo (desconto em folha de pagamento) ocorridas no mês de referência, expresso em milhões de reais (R\$ MM).

Primeiro analisaremos as estatísticas descritivas da série em questão para então passar a análise temporal.

Os dados da amostra, as Concessões de Crédito Consignado INSS, apresentam um valor médio de R\$ 4.824 milhões, com um desvio padrão de R\$ 2.031. Temos 50% dos valores acima de R\$ 4.295 milhões, 25% do volume concedido está entre 1.948 e 3.452, e 25% estão acima de R\$ 5.862. O histograma mostra visualmente que há uma distribuição assimétrica positiva (a direita) onde a mediana é menor que média. Observa-se no Gráfico 2.5 que os meses de janeiro, setembro e dezembro estão um pouco abaixo da média no período compreendido pela série.

Gráficos 2 - Evolução Mensal das Concessões de Crédito Consignado do INSS

Gráfico 2.1: Série Temporal Concessão Crédito

INSS

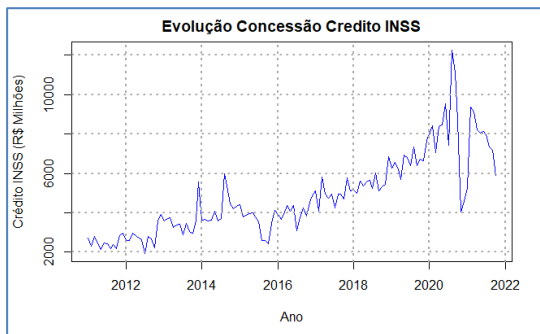


Gráfico 2.2: Histograma Concessão

Consignado INSS

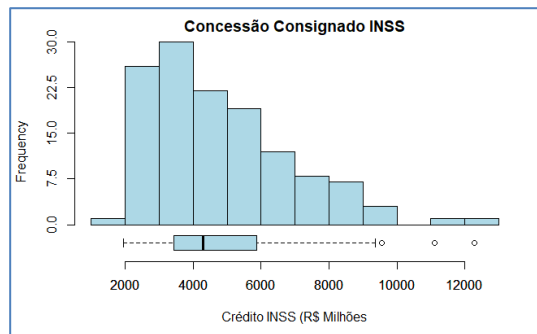


Gráfico 2.3: Box Plot Ano Concessão Crédito

INSS

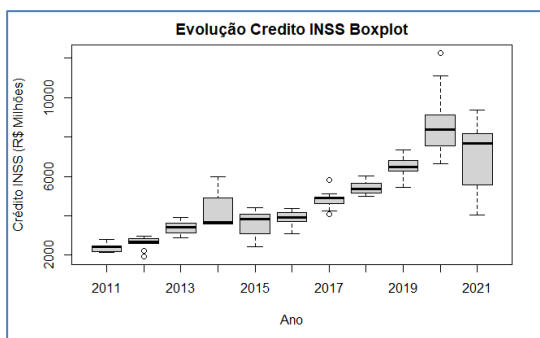


Gráfico 2.4: Box Plot Mês Concessão Crédito

INSS

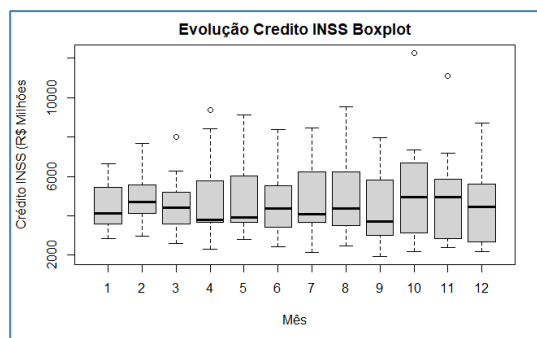


Gráfico 2.5: Subsérie Sazonal Consignado

INSS

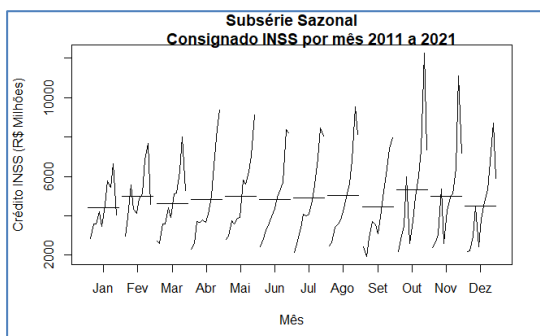


Gráfico 2.6: Distribuição Consignado INSS

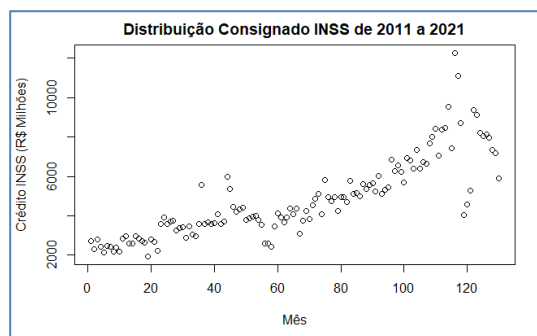


Tabela 1 - Estatísticas Descritivas Consignado INSS

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
1.948	3.451,75	4.295	4.824,5	5.861,75	12.273	4.125.056	2031,02

Fonte: Elaborado pelo autor.

A evolução da concessão de crédito INSS por ano, se observa uma menor variabilidade em 2011 e 2012, e maior variabilidade em 2014, 2020 (que apresenta uma observação atípica) e principalmente em 2021, onde já podemos observar efeitos da pandemia da COVID-19. Na evolução de crédito por mês, se observa menor variabilidade nos três primeiros meses do ano e maior variabilidade nos últimos três meses, com valores discrepantes nos meses de março, abril, outubro e novembro. No final do ano de 2014, mais especificamente em outubro de 2014, o governo Federal aumentou o prazo máximo para o pagamento das operações de empréstimo (e também de cartão de crédito consignado) relativas ao crédito consignado dos aposentados e pensionistas do INSS, ampliando de 60 para 72 meses. Esse aumento do prazo faz com que mais valor possa ser liberado na operação de empréstimo mediante as mesmas condições de taxa, parcela e margem consignável. Já em 2020, devido a pandemia da COVID-19, o governo Federal aumentou a margem consignável de 30% para 35%, visando aquecer a economia através do incentivo ao crédito. Essa alteração entrou em vigor em outubro e inicialmente teria validade até 31 de dezembro do mesmo ano. Porém, em março de 2021 o governo novamente elevou a margem para 35%, com validade até dezembro de 2021.

A análise exploratória da série nos mostra dois períodos com tendência de crescimento, o primeiro começa em meados de 2012 e vai até o final de 2014 e o segundo período com tendência de crescimento vai de 2016 até 2020. Em 2015 tivemos queda do volume de crédito, foi um ano marcado por forte crise política o que fez agravar ainda mais a crise econômica, uma das mais intensas dos últimos anos. O desequilíbrio das contas públicas levou o governo a adotar cortes nos gastos públicos, e este ajuste fiscal teve impacto nos benefícios da previdência. Vale lembrar que ao final de 2015 foi aberto o processo de impeachment da então presidente Dilma Rousseff, culminando no seu afastamento em agosto de 2016, período que marca a retomada da concessão de crédito consignado na série, que durou até o final de 2020.

É possível notar claramente o impulso na concessão do crédito consignado no final do ano de 2020 devido ao aumento da margem consignável. Desta forma o começo do ano de 2021 é marcado por uma retração nas operações de concessão do crédito. Ainda sentindo a necessidade de estimular a economia, o governo novamente eleva a margem consignável no INSS para 35%, em março de 2021, com validade até dezembro do mesmo ano, assim as operações de concessão retomam o viés de alta até o final do ano.

Considerando o período inteiro analisado, vemos que a série não varia aleatoriamente em torno da média e não varia aleatoriamente de forma constante, então não é uma série estacionária.

3.1.2 Taxa Média Mensal de Juros das Operações de Crédito Consignado INSS (TMMJ OCC)

Trata-se do valor médio mensal dos juros das operações de concessão de crédito consignado ao público do INSS ponderada pelos respectivos valores desembolsados, expresso em percentual e está expressa em percentual mensal (%a.m.).

A taxa média mensal de juros das operações de consignado do INSS da amostra apresenta um valor médio de 1,99% a.m., sendo que 50% das observações estão acima ou abaixo de 2,03% a.m. Vale ressaltar que não se trata da taxa média ponderada no período em questão, mas sim da média simples dos registros, estes sim ponderados pelo volume de operação. O desvio padrão da amostra ficou em 0,23%. A análise gráfica do histograma mostra visualmente que há uma distribuição aparentemente assimétrica negativa (a esquerda) onde a mediana é maior que a média.

A menor taxa média registrada na amostra foi de 1,55% a.m. e a maior foi de 2,35% a.m. Temos que 25% das taxas registradas possuem valor entre 1,55% a.m. e 1,86% a.m. e outros 25% das taxas observadas estão entre 2,12% a.m. e 2,35% a.m.

A evolução anualizada da taxa média mostra que ano de 2011 apresentou o maior valor da série histórica e que o ano de 2020 apresentou o menor valor da série. Pode-se ver que em determinados anos tivemos pouca ou quase nenhuma dispersão da série dentro dos anos. Isso se dá pelo caráter regulatório da taxa para as operações com o INSS, onde o teto da taxa é definido pelo Governo Federal. Ao final do mês de março de 2012, mais precisamente no dia 23 de março, entrou em vigor a resolução que aprovava a redução do teto dos juros para as operações de empréstimo consignado para aposentado e pensionista do INSS. O limite máximo de juros nas

operações consignadas no INSS caiu de 2,34% a.m. para 2,14% a.m. Em 2017, em 29 de setembro, tivemos nova redução do teto, desta vez de 2,14% para 2,08%. Em março de 2020, em razão do agravamento da crise econômica causada pela pandemia, o teto da taxa caiu de 2,08% para 1,80%.

Gráficos 3 - Evolução da Taxa Média Mensal de Juros das Operações de Crédito Consignado INSS

Gráfico 3.1: Série Temporal Taxa Juros Concessão INSS

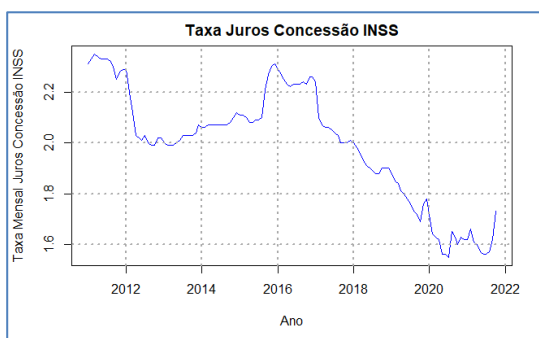


Gráfico 3.2: Histograma Taxa Juros Concessão INSS

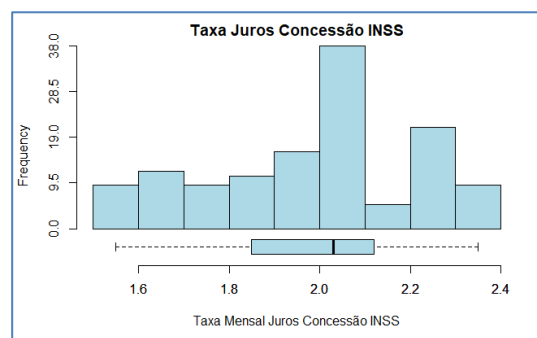


Gráfico 3.3: Box Plot Ano Taxa Juros Concessão INSS

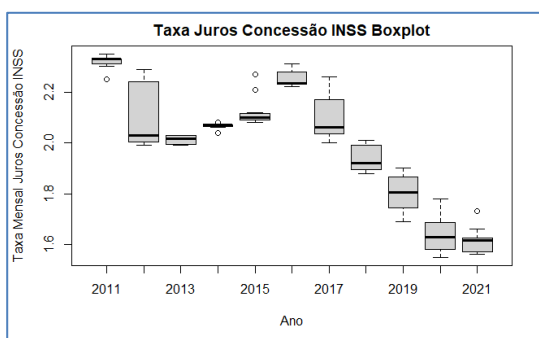


Gráfico 3.4: Box Plot Mês Taxa Juros Concessão INSS

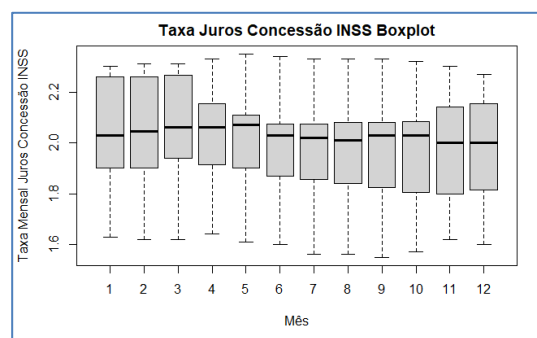


Gráfico 3.5: Subsérie Sazonal Taxa Juros Concessão INSS

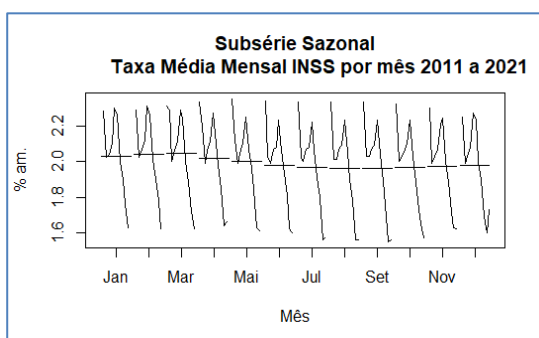
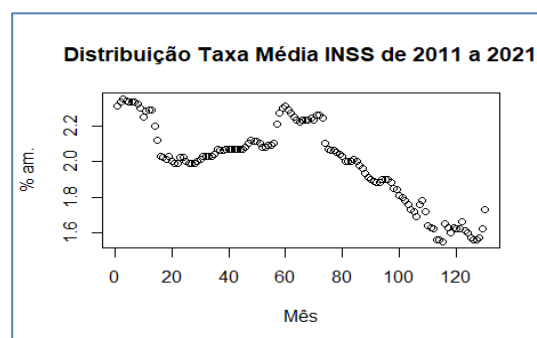


Gráfico 3.6: Distribuição Taxa Juros Concessão INSS



E em dezembro de 2021 foi anunciado que a partir de janeiro de 2022 o teto da taxa passaria para 2,14%. Essa subida da taxa teto pode ser explicada pelo aumento da taxa básica de juros, a Selic, que por sua vez apresentou gradual elevação ao longo de 2021 com o objetivo de conter os índices inflacionários. Essas alterações são facilmente percebidas na evolução da série.

Tabela 2 - Estatísticas Descritivas Taxa Juros Concessão INSS

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
1,55	1,8575	2,03	1,9922	2,1175	2,35	0,0512	0,2263

Fonte: Elaborado pelo autor.

Gráfico 4 - Evolução Taxa Selic



Fonte: Banco Central (2017; 2022)

Gráfico 5 - Evolução IPCA



Fonte: IBGE (2020, 2021).

Os anos de 2014, 2015 e 2021 apresentaram *outliers*. É possível identificar ainda uma tendência de crescimento, mesmo que tímida, da taxa média entre os anos de 2013 e 2015, talvez pelo impacto negativo no cenário econômico causa pelo conturbado momento político do período, e entre 2016 e 2020 temos uma clara tendência de queda, ocasionado principalmente pela gradual queda dos juros Selic no período.

3.1.3 Prazo Médio Mensal das Concessões de Crédito Consignado INSS (PMMCCC)

Trata-se da média dos prazos a decorres (número de dias corridos existentes entre a data da concessão do crédito e a data da última parcela, dividido por 30) das

operações de concessão de crédito consignado ao público do INSS, ponderada pelo valor das concessões, expresso em meses.

Gráficos 6 - Prazo Médio Mensal das Concessões de Crédito Consignado INSS

Gráfico 6.1: Série Temporal Prazo Concessão

INSS

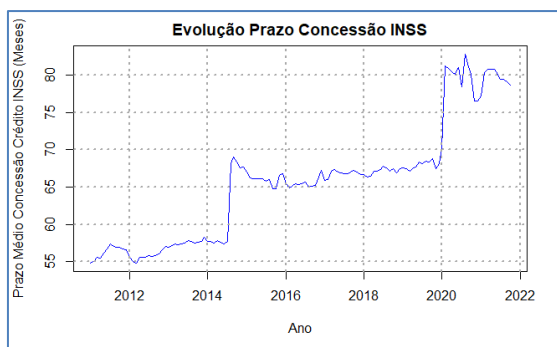


Gráfico 6.2: Histograma Prazo Concessão

INSS

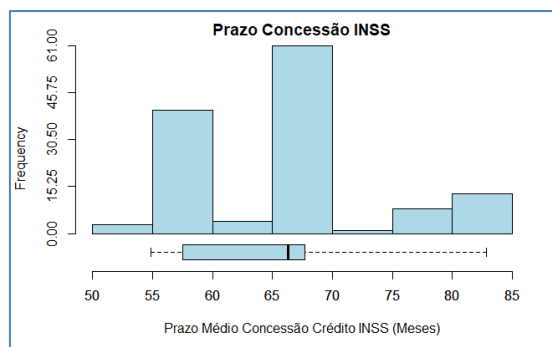


Gráfico 6.3: Box Plot Ano Prazo Concessão

INSS

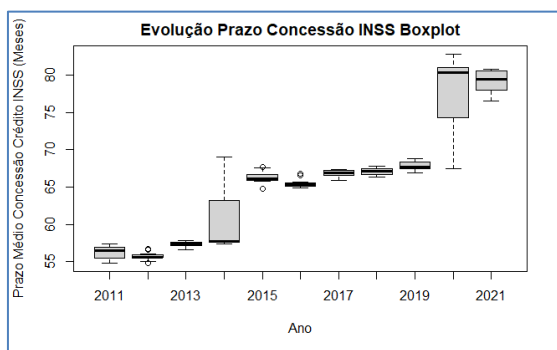


Gráfico 6.4: Box Plot Mês Prazo Concessão

INSS

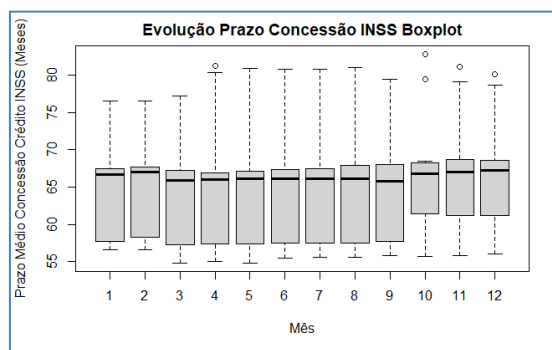


Gráfico 6.5: Subsérie Sazonal Prazo Concessão

INSS

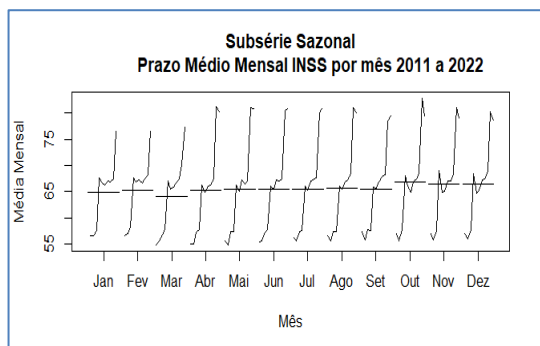


Gráfico 6.6: Distribuição Prazo Concessão INSS

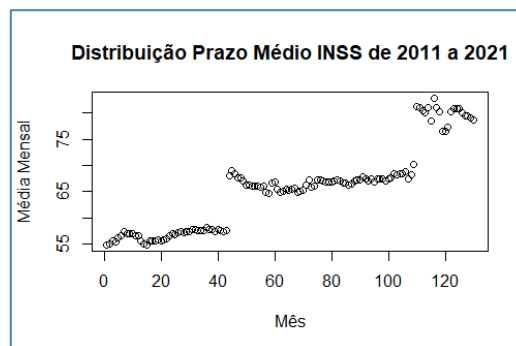


Tabela 3 - Estatísticas Descritivas Prazo Concessão INSS em meses

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
54,87	57,54	66,33	65,5865	67,73	82,8	61,2613	7,8270

Fonte: Elaborado pelo autor.

O prazo médio mensal das operações de consignado do INSS da amostra apresenta uma média de aproximadamente 65,6 meses, já o valor central da amostra foi de 66,3 meses. Assim como na análise da taxa média, vale aqui ressaltar que não se trata do prazo médio mensal ponderada no período, mas da média simples dos registros, estes sim ponderados pelo volume de operação. O desvio padrão foi de 7,8 meses. Os 25% menores prazos médios registrados estão entre 54,87 e 57,54 meses. Já os registros 25% maiores estão entre 67,73 e 82,80 meses. A análise gráfica do histograma mostra visualmente que há uma distribuição aparentemente assimétrica negativa (a esquerda) onde a mediana é maior que média

O teto do prazo médio também é regulado por normativos governamentais, da mesma forma que o teto da taxa, desta forma, fora dos períodos da alteração da regra, não temos variações consideráveis ao longo do período. O que fica evidente nos anos de 2014 e 2020, quando tivemos alteração do teto do prazo. Em outubro de 2014 o teto do prazo aumentou de 60 para 72 meses e em março de 2020 o teto foi aumentado de 72 para 84 meses. Ambas as mudanças se deram com o intuito de liberar mais crédito.

3.1.4 Inadimplência da Carteira de Crédito Consignado INSS (INAD CCC)

Trata-se do somatório do saldo das operações de crédito consignado ao público do INSS com atraso acima de 90 (noventa) dias e não baixados para prejuízo, dividido pelo saldo total de crédito da carteira, expresso em percentual, expresso em percentual (%).

A inadimplência da Carteira da amostra apresenta um valor médio de 2,03%, com um desvio padrão de 0,35 pp. A mediana da amostra foi de 1,92%, sendo que 25% dos registros inferiores estão entre 1,56% e 1,75%, e os 25% maiores estão entre 2,30% e 2,84%. O histograma mostra visualmente que há uma distribuição assimétrica positiva (a direita) onde a mediana é menor que média.

A evolução da inadimplência da carteira do INSS por ano mostra pouco variabilidade dentro dos anos, 2013, 2019 e principalmente 2021 foram anos que apresentam uma maior variabilidade. Na evolução por mês, se observa variabilidade similar nos 12 meses, sem valores discrepantes.

Gráficos 7 - Percentual de Inadimplência da Carteira de Crédito Consignado INSS

Gráfico 7.1: Série Temporal Percentual Inadimplência Carteira INSS

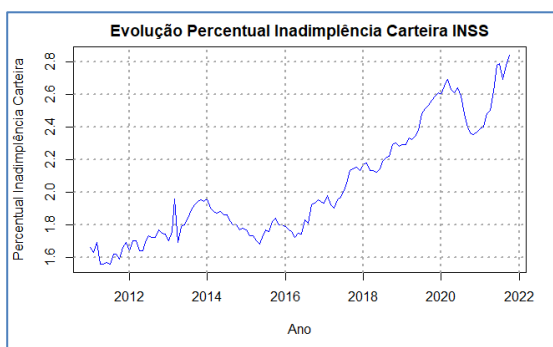


Gráfico 7.2: Histograma Percentual Inadimplência Carteira INSS

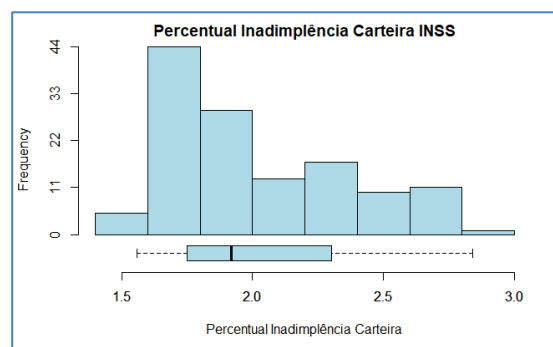


Gráfico 7.3: Box Plot Ano Percentual Inadimplência Carteira INSS

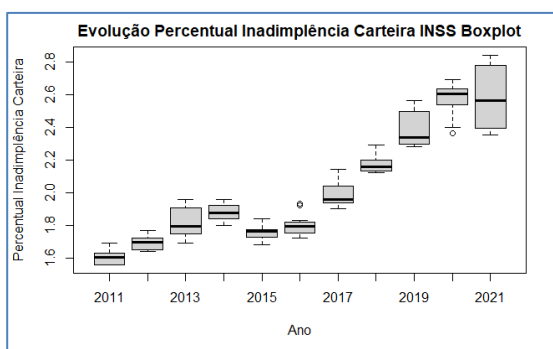


Gráfico 7.4: Box Plot Mês Percentual Inadimplência Carteira INSS

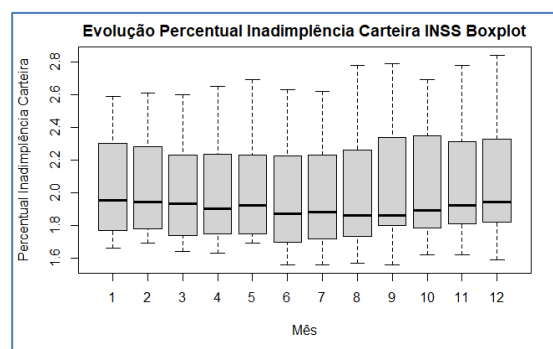


Gráfico 7.5: Subsérie Sazonal Percentual Inadimplência Carteira INSS

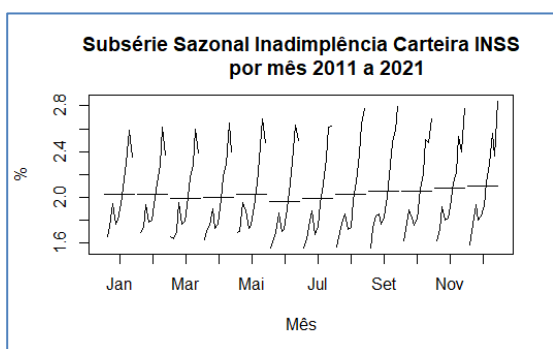


Gráfico 7.6: Distribuição Percentual Inadimplência Carteira INSS

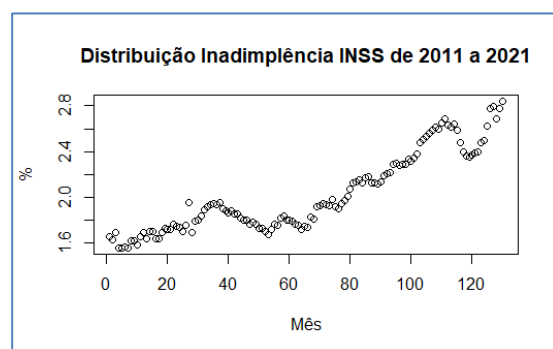


Tabela 4 - Estatísticas Descritivas Percentual Inadimplência Carteira INSS

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
1,56	1,7525	1,92	2,0288	2,2975	2,84	0,1196	0,3459

Fonte: Elaborado pelo autor.

A análise exploratória da série nos mostra que após uma queda da inadimplência entre 2014 e 2015 o período entre 2016 e o começo de 2020 apresentou tendência de crescimento, interrompida um recuo da inadimplência ao longo do ano de 2020. A série volta a apresentar crescimento em 2021.

3.1.5 Percentual da Carteira de Crédito Consignado INSS com Atraso entre 15 e 90 dias (PCCC ATRASO)

Trata-se do somatório do saldo das operações de crédito consignado ao público do INSS com atraso de 15 a 90 dias, inclusive, dividido pelo saldo total de crédito da carteira, expresso em percentual, expresso em percentual (%).

A evolução do percentual da carteira em atraso entre 15 e 90 dias por ano mostra uma variabilidade alta nos anos de 2012, 2014 e 2018, sendo que em 2012 e 2014 tivemos a ocorrência de valores superiores discrepantes. Os demais anos apresentaram pouca variabilidade. É possível notar a queda da mediana anual no período de 2011 a 2017 e um crescimento entre 2017 e 2019. Apresentando certa estabilidade nos anos de 2020 e 2021. Na evolução por mês, se observa variabilidade similar nos 12 meses, sem valores discrepantes, mas nos meses de fevereiro e abril tivemos a ocorrência de valores altos, devido um *outlier* em 2012 e um alto valor registrado em 2018.

A análise exploratória da série nos mostra que apesar dos picos registrados em 2012 e 2014, tivemos uma tendência de queda entre os anos de 2011 e 2017 e entre 2017 e 2019 tivemos tendência de aumento.

O percentual da carteira em atraso entre 15 e 90 dias da amostra apresenta um valor médio de 1,49%, com um desvio padrão de 0,33 pp. A mediana da amostra ficou muito próxima da média, 1,52%, sendo que 25% dos registros estão entre 0,93% e 1,19%, e 25% estão entre 1,74% e 2,45%. O histograma mostra visualmente que há uma distribuição simétrica onde a mediana é igual a média.

Gráficos 8 - Percentual da Carteira de Crédito Consignado INSS com Atraso entre 15 e 90 dias

Gráfico 8.1: Série Temporal Percentual da Carteira em Atraso INSS

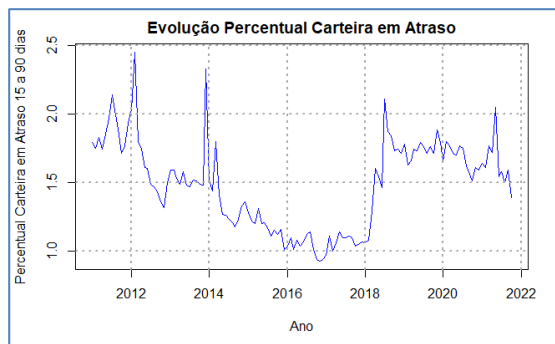


Gráfico 8.2: Histograma Percentual da Carteira em Atraso INSS

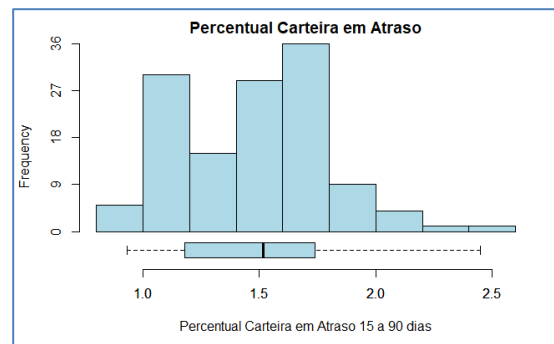


Gráfico 8.3: Box Plot Ano Percentual da Carteira em Atraso INSS

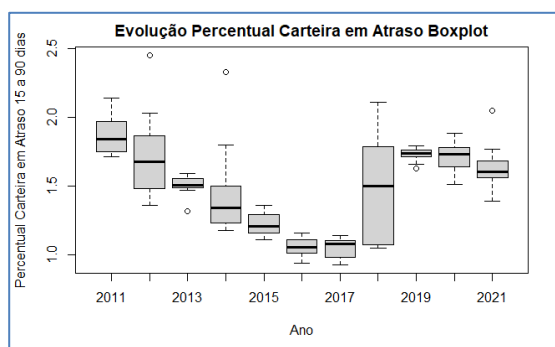


Gráfico 8.4: Box Plot Mês Percentual da Carteira em Atraso INSS

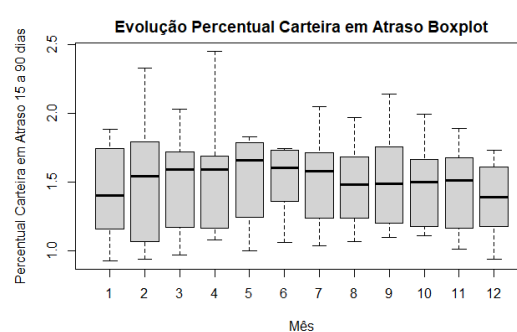


Gráfico 8.5: Subsérie Sazonal Percentual da Carteira em Atraso INSS

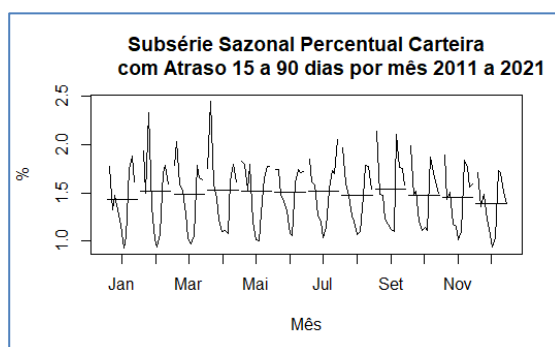


Gráfico 8.6: Distribuição Percentual da Carteira em Atraso INSS

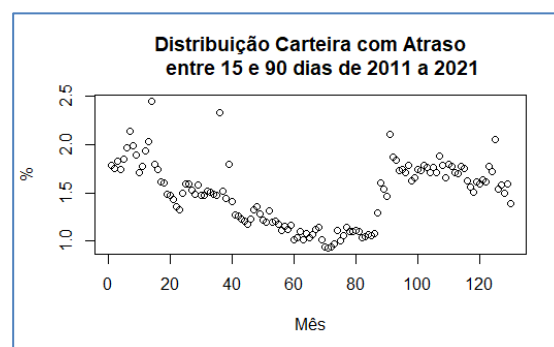


Tabela 5 - Estatísticas Descritivas Percentual da Carteira em Atraso INSS

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
0,93	1,185	1,515	1,4868	1,74	2,45	0,1051	0,3242

Fonte: Elaborado pelo autor.

3.1.6 PIB Mensal

Trata-se indicador mensal estimado do Produto Interno Bruto via interpolação de valores trimestrais já divulgados pelo IBGE, expresso em milhões de reais (R\$ MM). O menor e o maior valor registrados foram R\$ 354.617 milhões e R\$ 762.978 milhões, em 2011 e 2021 respectivamente. Em determinados anos tivemos pouca ou quase nenhuma dispersão dentro do ano, a exceção é em 2020, que apresentou uma maior dispersão em relação aos demais anos. Novamente vale ressaltar que em 2020, a pandemia da COVID-19 impactou a economia do país.

Gráficos 9 - Evolução do PIB Mensal

Gráfico 9.1: Série Temporal PIB Mensal

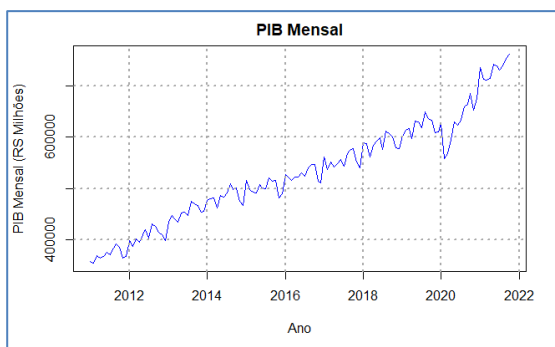


Gráfico 9.2: Histograma PIB Mensal

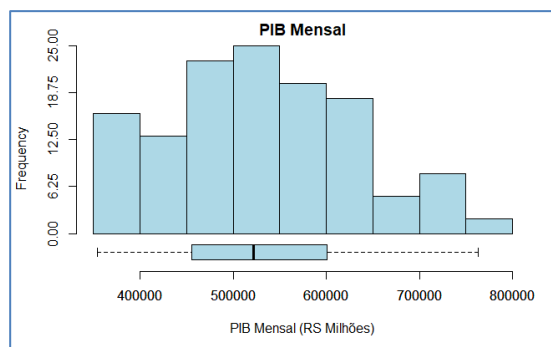


Gráfico 9.3: Box Plot Ano PIB Mensal

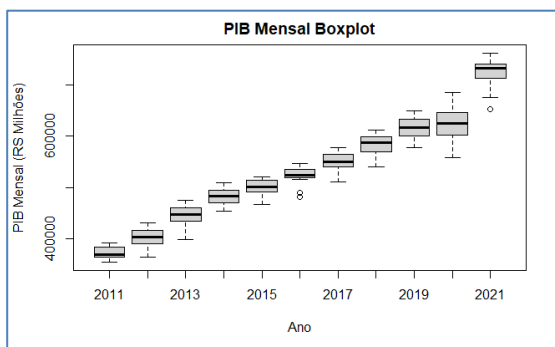


Gráfico 9.4: Box Plot Mês PIB Mensal

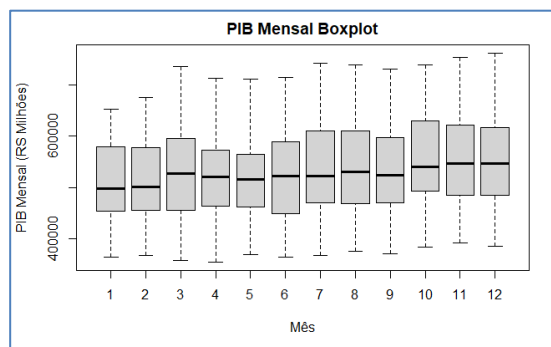


Gráfico 9.5: Subsérie Sazonal PIB Mensal

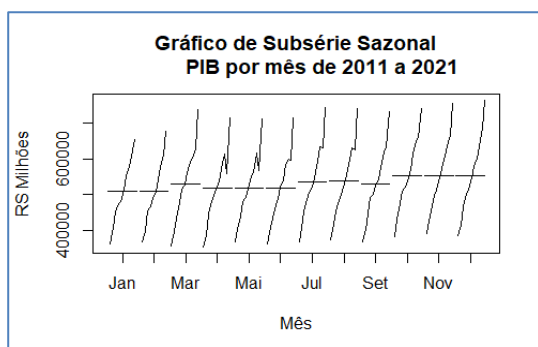
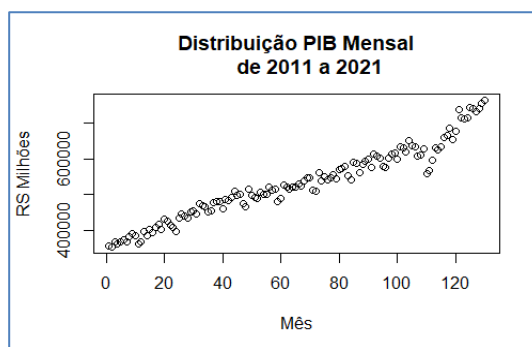


Gráfico 9.6: Distribuição PIB Mensal



É possível identificar uma clara tendência de crescimento, o que também pode ser confirmado pelo gráfico de linha. Vemos essa tendência de crescimento praticamente em toda a série, até 2019. Após o período atípico no começo de 2020, a série volta a apresentar crescimento.

Tabela 6 - Estatísticas Descritivas PIB Mensal em milhões de reais

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
354.617,8	456.957,4	521.882,2	531.021,3	600.197,7	762.978,4	10240786842	101.196,8

Fonte: Elaborado pelo autor.

O PIB médio da amostra apresenta um valor R\$ 531.021 milhões e a mediana de R\$ 521.882 milhões. Valores muito próximos, o que indica que não há a existência de valores extremos na amostra. O desvio padrão da amostra ficou foi de R\$ 101.196 milhões. E a análise gráfica do histograma mostra visualmente que há uma distribuição aparentemente simétrica.

3.1.7 Taxa Básica de Juros - Selic Acumulada no Mês (SELIC)

Trata-se do valor averiguado nas operações de empréstimos de um dia entre os bancos que usam títulos públicos federais como garantia, sendo considerada a taxa básica de juros da economia no Brasil. A sigla SELIC é a abreviação de Sistema Especial de Liquidação e Custódia, expressa em percentual ao mês (% a.m.).

A taxa Selic da amostra apresenta um valor médio de 0,69% a.m., com um desvio padrão de 0,29 pp. A mediana da amostra apresentou resultado de 0,71% a.m., sendo que 25% dos registros inferiores estão entre 0,13% e 0,49% a.m., e os 25% maiores registros estão entre 0,94% e 1,22%. O histograma mostra visualmente que há uma distribuição assimétrica negativa (a esquerda).

Gráficos 10 - Selic Acumulada no Mês

Gráfico 10.1: Série Temporal Selic Acumulada Mês

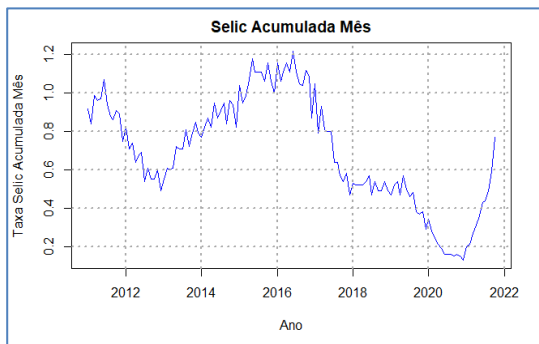


Gráfico 10.2: Histograma Selic Acumulada Mês

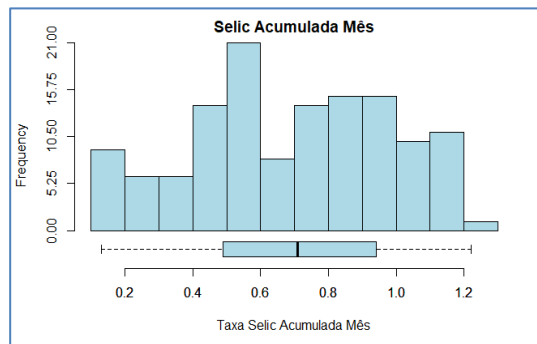


Gráfico 10.3: Box Plot Ano Selic Acumulada Mês

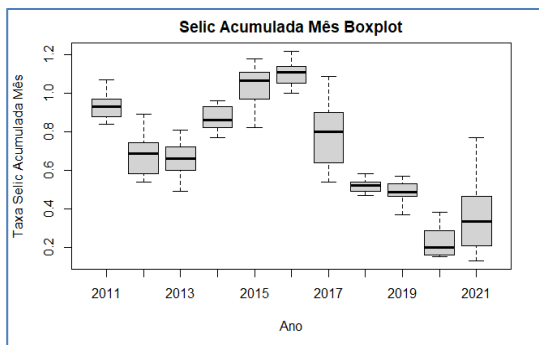


Gráfico 10.4: Box Plot Mês Selic Acumulada Mês

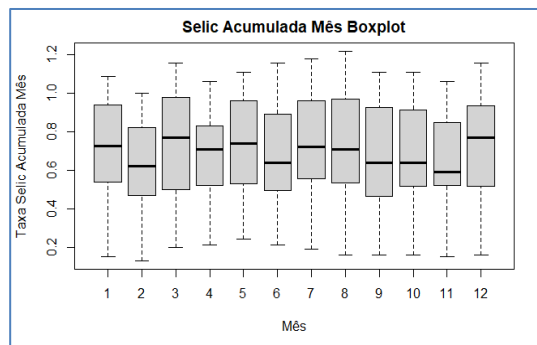


Gráfico 10.5: Subsérie Sazonal Selic Acumulada Mês

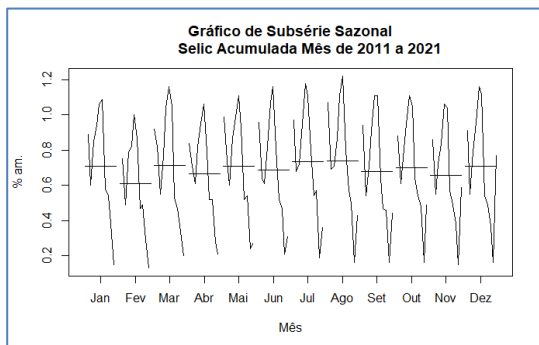


Gráfico 10.6: Distribuição Selic Acumulada Mês

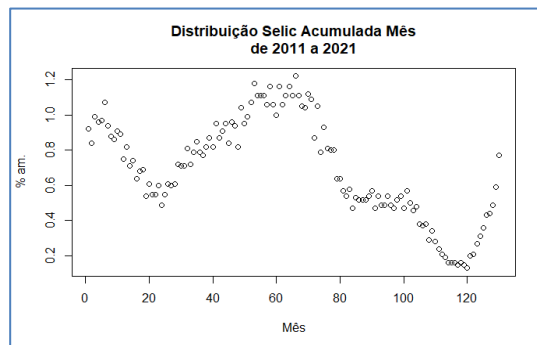


Tabela 7 - Estatísticas Descritivas Selic Acumulada no Mês

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
0,13	0,49	0,71	0,6938	0,9375	1,22	0,2882	0,2882

Fonte: Elaborado pelo autor.

A evolução da Selic em geral mostra pouca variabilidade dentro dos anos. Os anos de 2017 e 2021 apresentaram maior variação. Ao longo de 2017 foram oito cortes consecutivos na taxa básica de juros que reduziram a Selic de 13,75% ao fim de 2016 para 7,00% em dezembro de 2017, o menor patamar da história até então. Já em 2021, novamente lembrando o impacto da pandemia na economia, foram sete aumentos consecutivos. Após manter a taxa básica de juros em 2,00% na sua primeira reunião do ano, em janeiro, o Copom (comitê de política monetária) aumentou a Selic nas outras 7 reuniões que ocorrem ao longo do ano, fechando 2021 em 8,75%. Na evolução por mês é observado uma variabilidade similar nos 12 meses, com certa discrepância para os meses de janeiro, fevereiro e abril. Não foi observado valores discrepantes em nenhum mês.

A análise exploratória da série nos apresenta um período de queda da Selic de 2011 até o final de 2012. De 2013 até o começo de 2017 tivemos um período de aumento da taxa, o que pode ser creditado à crise político-econômica que o país enfrentou na ocasião, culminando com o impeachment da, então, presidente Dilma Rousseff, em agosto de 2016. Em 2017 tivemos sucessivas queda da Selic. Já em 2018 e 2019 tivemos praticamente uma estabilização da taxa básica e do final de 2019 até o final da série em 2020 tivemos novamente um período de queda, onde a taxa chegou ao seu menor nível histórico até o momento. Já em 2021 tivemos um novo período de aumento.

3.1.8 Salário-Mínimo

Trata-se do valor mais baixo de salário que os empregadores podem legalmente pagar aos seus funcionários pelo tempo e esforço gastos na produção de bens e serviços no âmbito nacional, expresso em unidade monetária contábil (R\$). O salário-mínimo sofre aumentos em fevereiro de cada ano e se mantém estável até o início do ano seguinte.

O salário-mínimo médio da amostra apresenta um resultado de R\$ 847 e a mediana de R\$ 880 com um desvio padrão da amostra ficou foi de R\$ 173. A análise gráfica do histograma mostra visualmente que há uma distribuição aparentemente assimétrica negativa.

Gráficos 11 - Evolução do salário-mínimo

Gráfico 11.1: Série Temporal Salário Mínimo

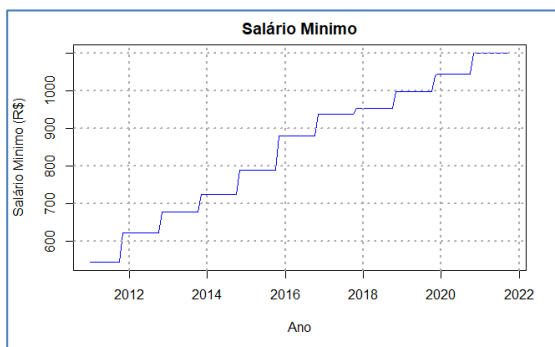


Gráfico 11.2: Histograma Salário Mínimo

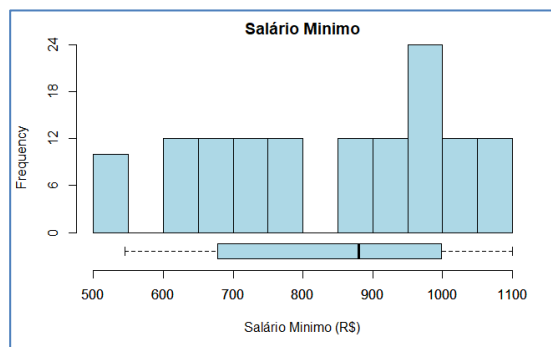


Gráfico 11.3: Box Plot Ano Salário Mínimo

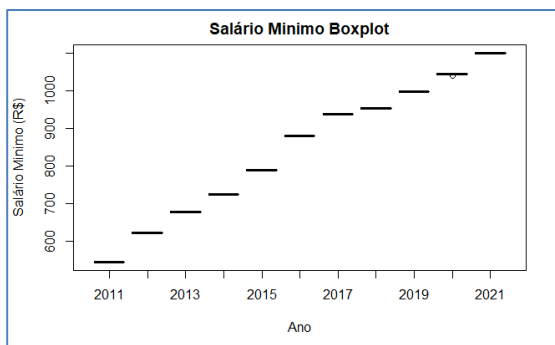


Gráfico 11.4: Box Plot Mês Salário Mínimo

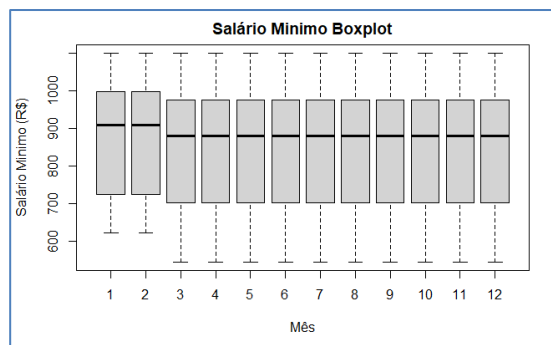


Gráfico 11.5: Subsérie Sazonal Salário Mínimo

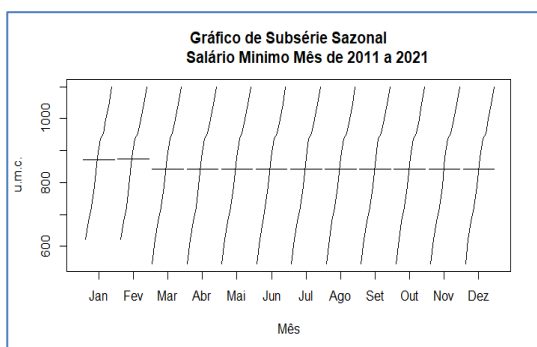


Gráfico 11.6: Distribuição Salário Mínimo

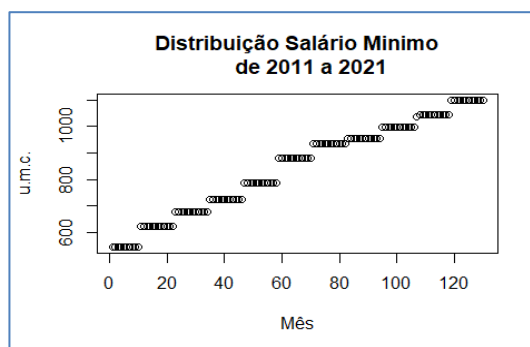


Tabela 8 - Estatísticas Descritivas salário-mínimo

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
545	678	880	847,3538	998	1100	29791,16	172,6012

Fonte: Elaborado pelo autor.

O menor e o maior valor registrados na série foram R\$ 545 e R\$ 1.100, em 2011 e 2021 respectivamente. Com exceção de 2020, onde tivemos dois aumentos do salário-mínimo, um em janeiro e outro em fevereiro (pois a inflação superou a projeção do governo), nos demais anos da série tivemos um aumento por ano.

É possível identificar uma clara tendência de crescimento, o que também pode ser confirmado pelo gráfico de linha. Vemos essa tendência de crescimento praticamente em toda a série, pois anualmente o salário-mínimo é corrigido de acordo com a inflação no período.

3.1.9 Índice Nacional de Preços ao Consumidor-amplio (IPCA)

Trata-se da variação percentual mensal do índice brasileiro oficial para identificar a variação de preços no comércio sob a ótica do consumidor, sendo utilizado para averiguar a inflação ou deflação, isto é, o poder de compra da população, expresso em percentual (%).

A variação do IPCA por ano mostra uma variabilidade alta nos anos de 2015 e 2016, período do auge da crise política, e em 2019 e 2020. Em 2019 a alta variação do IPCA se deu principalmente pelo aumento do preço da carne, *commoditie* muito valorizada devido ao aumento da exportação para a China e à desvalorização do real frente ao dólar. Em 2020 pode-se creditar essa alta variabilidade a crise gerada pela pandemia. Os demais anos não apresentaram tanta variabilidade. É possível notar valores discrepantes nos anos de 2017 e 2018. Na evolução por mês é possível observar uma certa sazonalidade, com meses com maior variação do índice e outros meses com pouca variação.

A análise exploratória nos mostra uma variação do IPCA menor no começo com um aumento no decorrer da série, até 2015 e 2016 com forte variabilidade. Após um período de maior estabilidade do índice de inflação entre o final 2016 e o final de 2017,

voltamos a ter um período com maior variabilidade do índice entre os anos de 2018 e 2020. E voltando a ter uma menor variabilidade do índice em 2021.

Gráficos 12 - IPCA Variação Mensal

Gráfico 12.1: Série Temporal IPCA Variação Mensal

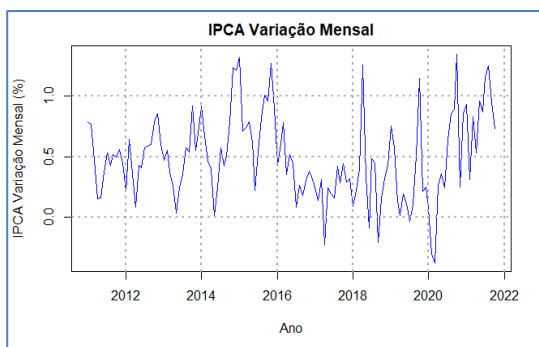


Gráfico 12.2: Histograma Concessão IPCA Variação Mensal

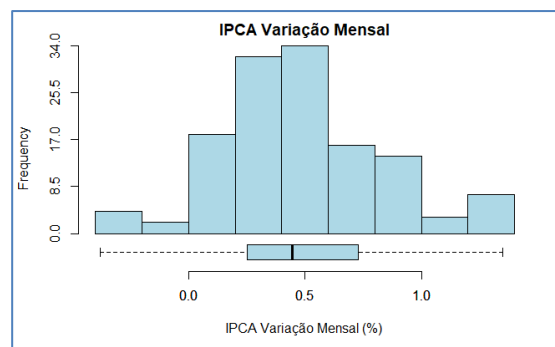


Gráfico 12.3: Box Plot Ano IPCA Variação Mensal

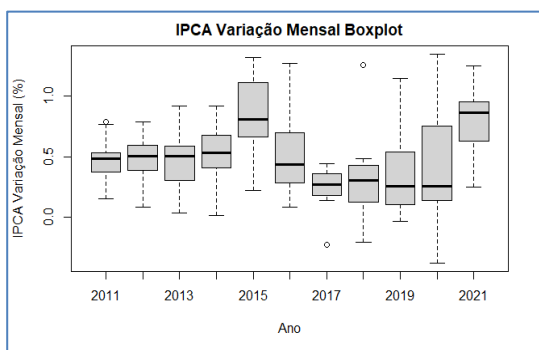


Gráfico 12.4: Box Plot Mês IPCA Variação Mensal

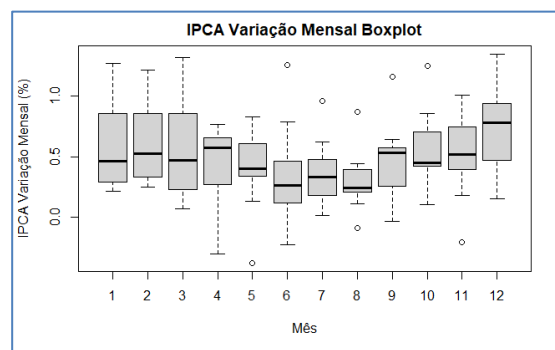


Gráfico 12.5: Subsérie Sazonal IPCA Variação Mensal

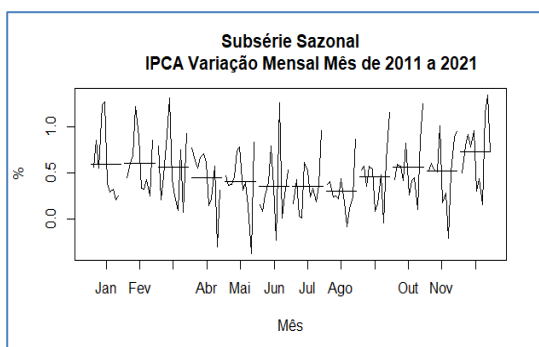


Gráfico 12.6: Distribuição IPCA Variação Mensal

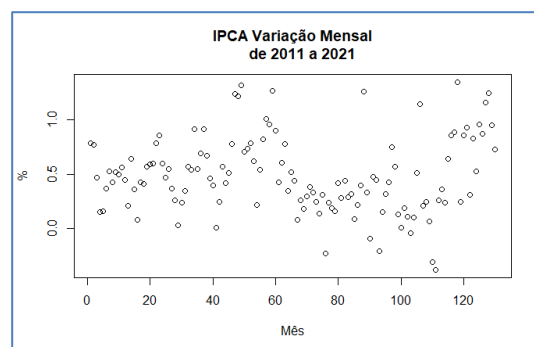


Tabela 9 - Estatísticas Descritivas IPCA Variação Mensal

Mínimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo	Variância	Desvio Padrão
0,38	0,25	0,445	0,4891	0,725	1,35	0,1221	0,3494

Fonte: Elaborado pelo autor.

A variação do índice nacional de preço ao consumidor amplo, o IPCA, apresentou um valor médio na amostra de 0,49%, com um desvio padrão de 0,35%. A mediana da amostra ficou muito próxima da média, 0,45%, indicando a não existência de valores discrepantes. O histograma mostra visualmente que há uma distribuição simétrica.

3.2 Correlação entre séries

Um método estatístico amplamente utilizado para estudar o grau de relacionamento entre essas variáveis é a análise de correlação. De acordo com LIRA (2004), ela mede a intensidade e a direção da relação linear ou não-linear entre duas ou mais variáveis e permite saber se as alterações sofridas por uma delas são acompanhadas por alterações nas demais.

Um dos coeficientes de correlação mais simples é o coeficiente de Pearson, que mede a correlação linear entre duas variáveis e é definido por SANTOS (2015) como:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y - \bar{y})^2}} \quad (27)$$

Onde:

- x_i e y_i = Valores medidos de ambas as variáveis
- \bar{x} e \bar{y} = Médias amostrais de ambas as variáveis

O coeficiente de correlação quantifica a força de associação linear entre duas variáveis, portanto, descreve quão bem uma linha reta se ajustaria através de uma nuvem de pontos. O coeficiente de correlação pode variar entre -1 (correlação negativa perfeita) e +1 (correlação positiva perfeita). Valores negativos do coeficiente de correlação indicam uma correlação do tipo inversa, isto é, quando x aumenta y

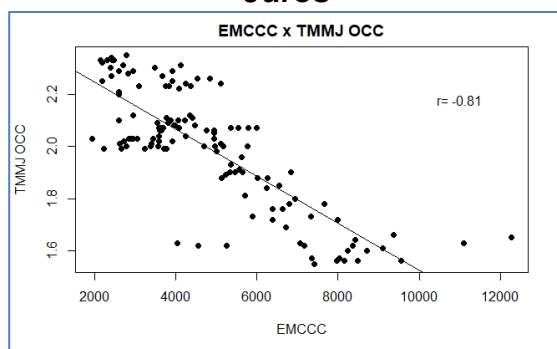
diminui. Valores positivos do coeficiente de correlação ocorrem quando x e y variam no mesmo sentido, isto é, quando x aumenta y aumenta ou quando x diminui y também diminui (SANTOS, 2015).

Para Dancey e Reidy (2006) valores de coeficiente de correlação até 0,30 devem ser considerados fracos, entre 0,40 e 0,60 moderados e acima de 0,70 fortes.

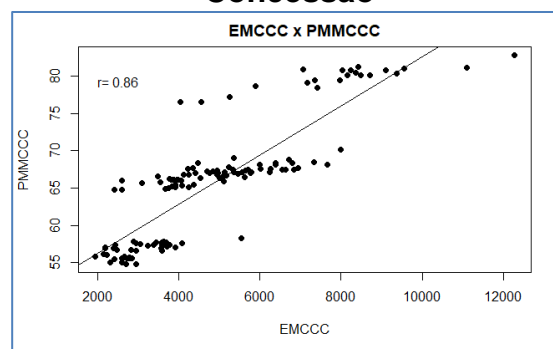
Desta forma, foi calculado o coeficiente de correlação entre as variáveis a fim de descartar para a elaboração do modelo aquelas que não se apresentam significativas. Tendo como finalidade definir as variáveis que serão utilizadas no modelo ARIMAX, com sazonalidade.

A seguir, com o auxílio de diagramas de dispersão será descrito o tipo de relação existente entre a variável dependente Y_t no instante t com as variáveis explicativas.

**Gráfico 13 - Diagrama Dispersão
Concessão Crédito INSS x Taxa
Juros**

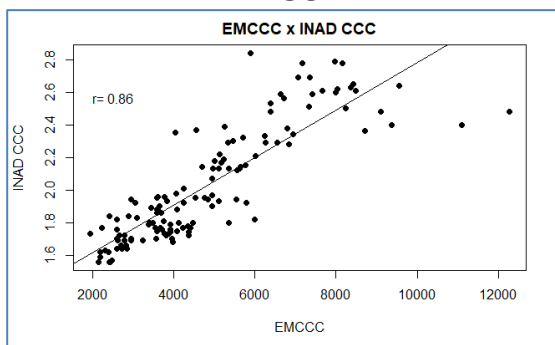


**Gráfico 14 - Diagrama Dispersão
Concessão Crédito INSS x Prazo
Concessão**

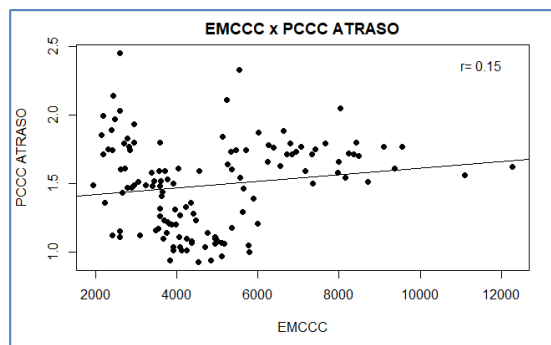


Pela análise do gráfico de dispersão e pelo resultado do coeficiente de variação linear pode-se concluir que a correlação entre a variável Taxa média mensal de juros das operações de crédito consignado e a variável resposta Evolução mensal das concessões de crédito consignado do INSS é negativa forte, ou seja, a variável dependente descesse quase que na mesma proporção que a variável independente.

**Gráfico 15 - Diagrama Dispersão
Concessão Crédito INSS x
Percentual Inadimplência Carteira
INSS**



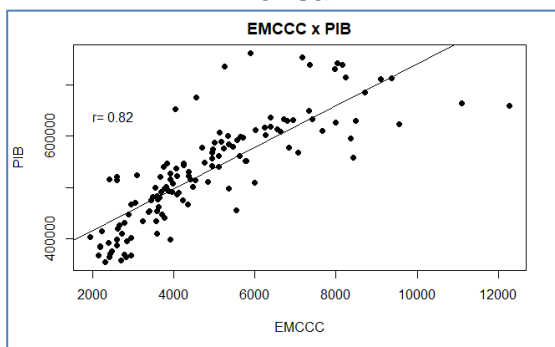
**Gráfico 16 - Diagrama Dispersão
Concessão Crédito INSS x
Percentual Carteira em Atraso INSS**



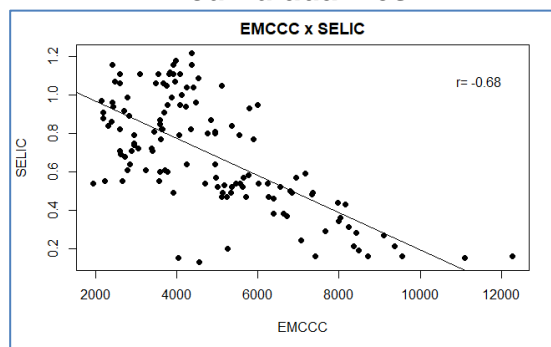
A variável Inadimplência da carteira de crédito consignado do INSS apresenta correlação positiva forte. A variável percentual da carteira de crédito consignado em atraso entre 15 e 90 dias apresenta uma correlação positiva desprezível.

Já a variável Prazo médio mensal das concessões de crédito consignado do INSS apresenta correlação também forte, porém positiva com a variável resposta.

**Gráfico 17 - Diagrama Dispersão
Concessão Crédito INSS x PIB
Mensal**

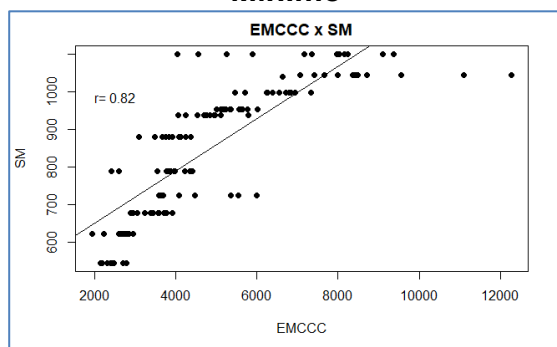


**Gráfico 18 - Diagrama Dispersão
Concessão Crédito INSS x Selic
Acumulada Mês**

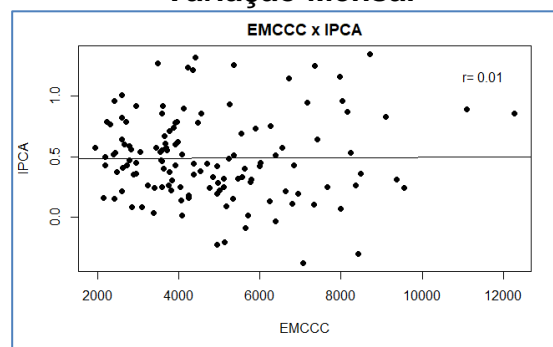


A correlação da variável independente PIB com a variável dependente é positiva forte. Já a correlação da variável Selic é negativa moderada, quase forte.

**Gráfico 19 - Diagrama Dispersão
Concessão Crédito INSS x Salário
Mínimo**



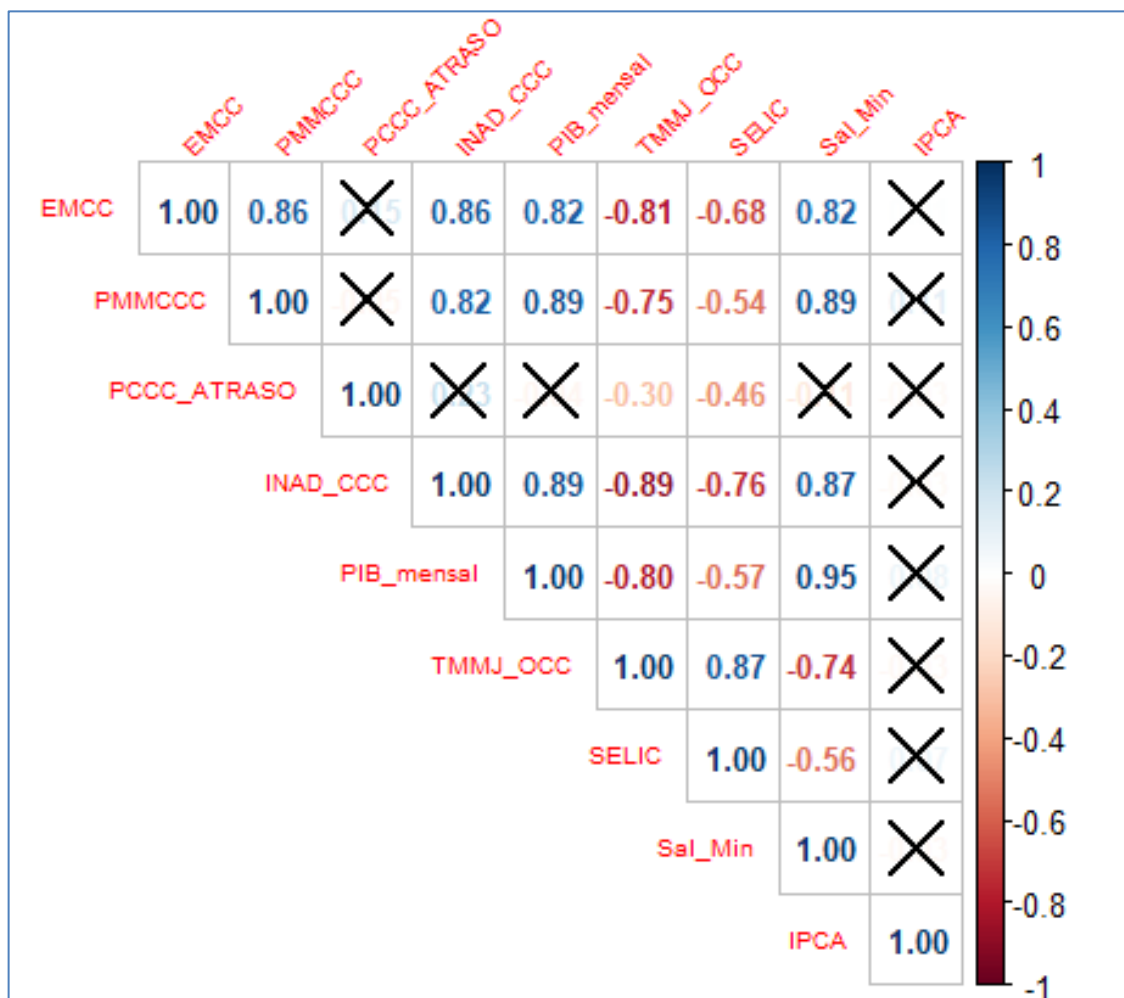
**Gráfico 20 - Diagrama Dispersão
Concessão Crédito INSS x IPCA
Variação Mensal**



A variável explicativa “salário-mínimo” apresenta correlação positiva forte com a variável resposta, enquanto a variável IPCA apresenta correlação negativa desprezível.

Analisando a Figura 2, pode-se concluir que as variáveis TMMJ OCC ($r = -0,81$, $p\text{-valor} = 0,0000$), PMMCCC ($r = 0,86$, $p\text{-valor} = 0,0000$), INAD CCC ($r = 0,86$, $p\text{-valor} = 0,0000$), PIB ($r = 0,82$, $p\text{-valor} = 0,0000$) e SM ($r = 0,82$, $p\text{-valor} = 0,0000$) apresentaram correlação significativa com a variável resposta. A variável Selic apresentou uma correlação moderada para forte ($r = 0,68$, $p\text{-valor} = 0,0000$), em contrapartida as variáveis PCCC_Atraso ($r = 0,15$, $p\text{-valor} = 0,0822$) e IPCA ($r = 0,01$, $p\text{-valor} = 0,9516$) apresentaram correlação não significativa, considerando o nível de significância de 5%. Os campos marcados com X apresentam $p\text{-valor}$ maior que 5%, indicando que a relação entre as variáveis não é estatisticamente significativa para o nível de significância adotado.

Figura 2 - Matriz Correlação



Assim, as correlações das variáveis PCCC_ATRASO (Percentual da carteira de crédito consignado INSS com atraso entre 15 e 90 dias) e IPCA se apresentaram não significativas em relação a variável resposta, como apresentado no parágrafo anterior.

3.2.1 Cálculo do VIF – Variance Inflation Factor

Um dos pressupostos em um modelo de regressão é a independência entre as variáveis. Hair (2005) descreve que o ideal seria obter diversas variáveis totalmente independentes, para que não ocorra multicolinearidade, que ocorre quando há correlação significativa entre duas ou mais variáveis preditoras, mas que estas sejam altamente correlacionadas com a variável resposta. Os efeitos da multicolinearidade são referentes à explicação e estimação que, segundo Corrar (2009), tende a distorcer os coeficientes angulares estimados para as variáveis que a apresentam,

prejudicando a capacidade preditiva do modelo e a compreensão do real efeito das variáveis independentes sobre o comportamento da variável resposta.

Em outras palavras pode-se dizer que multicolinearidade em regressão é uma condição que ocorre quando algumas variáveis preditoras no modelo estão correlacionadas entre si, sendo que uma forte multicolinearidade pode aumentar a variância dos coeficientes de regressão.

Uma forma de avaliar se existe multicolinearidade entre as variáveis preditoras é aplicar o fator de inflação da variância (VIF) (o inverso do VIF é chamado de coeficiente de tolerância), cuja fórmula é:

$$VIF_i = \frac{1}{1-R_i^2}, \quad i = 1, 2, \dots, p \quad (28)$$

Onde:

- p = número das variáveis preditoras
- R_i^2 = coeficiente de correlação múltipla, resultante da regressão de X_i nos outros $p - 1$ regressores

Assim, para medir a multicolinearidade das variáveis explicativas foi calculado o fator de inflação da variância (VIF), que mede o quanto a variância de um coeficiente de regressão estimado aumenta se seus preditores estão correlacionados.

A multicolinearidade forte ($VIF > 5$) é problemática porque pode aumentar a variância dos coeficientes de regressão, tornando-os instáveis. Coeficientes instáveis podem parecer insignificantes mesmo quando existe uma relação significativa entre o preditor e a resposta. Outro ponto negativo é que coeficientes preditores altamente correlacionados variam fortemente de amostra para amostra. Além disso, a remoção de quaisquer termos altamente correlacionados do modelo afetará significativamente os coeficientes estimados dos outros termos altamente correlacionados, afetando os termos das variáveis não correlacionadas, sendo que estes podem até mesmo ter o sinal errado, do ponto de vista da relação esperada com a variável resposta.

Se todos os VIFs forem 1 não há multicolinearidade, mas se algum resultado for maior do que 1, os preditores estão correlacionados. Segundo Alain Zuur et. al. (2009) quando o resultado do VIF for maior que 5 o coeficiente de regressão para esse termo

não é estimado de maneira apropriada. Desta forma, iremos retirar esta variável da base e calcular novamente a VIF.

Como na análise de correlação as variáveis PCCC_Atraso (p-valor = 0,0822) e IPCA (p-valor = 0,9516) se apresentaram como sendo não estatisticamente significativas para um nível $\alpha=0,05$ em relação a variável resposta, iremos desconsiderá-las na análise da VIF.

Tabela 10 - Análise Inflação da Variância

Variável	1ª Round	2ª Round	3ª Round	4ª Round
PMMCCC	6	6	5	5
INAD_CCC	10	9,97	8	
PIB_mensal	18			
TMMJ_OCC	14	10,18		
SELIC	6	5	3	1
Sal_Min	15	8	7	5

Fonte: Elaborado pelo autor.

Na primeira rodada de análise da VIF, todas as 6 variáveis apresentaram resultado acima de 5. Desta forma iremos retirar a variável com maior resultado de VIF e rodar o modelo novamente.

Uma segunda rodada, com a retirada da variável PIB_mensal, o resultado da inflação das variáveis sofreu alteração, todas apresentaram queda. Porém quatro variáveis apontadas na primeira rodada persistem com resultado acima de 5.

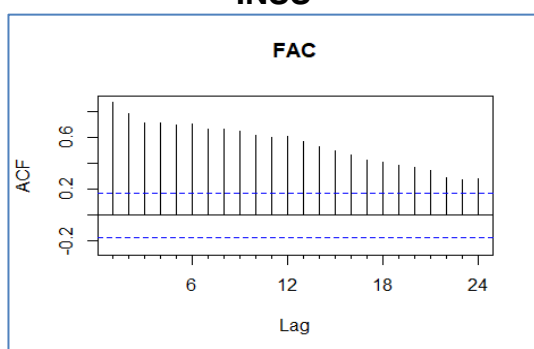
Uma terceira rodada sem a variável TMMJ_OCC resultou na queda da inflação das variáveis que permaneceram. Na quarta rodada de cálculo da VIF retirando a variável INAD_CCC resultou uma queda do VIF.

Após então quatro rodadas de análise da VIF a variável Salário-Mínimo Mensal, passou a apresentar resultado igual a 5. Desta forma, iremos considerar as 3 variáveis restantes que irão compor o modelo de regressão. Ambas apresentam boa correlação com a variável resposta e p-valor menor que 0,0001 como abordado na seção 5.2, indicando que a relação é estatisticamente significativa para um nível $\alpha=0,05$.

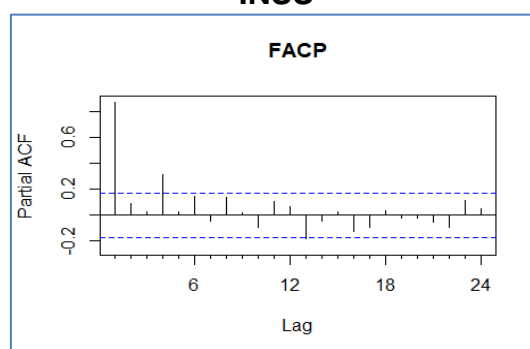
3.3 Estimação dos parâmetros dos Modelos

Os Gráficos 21 e 22, que apresenta o correlograma da autocorrelação (FAC) e autocorrelação parcial (FACP) referente a série Concessões de crédito consignado do INSS, analisando-os verifica-se que a série não é estacionária em relação ao nível pois tem uma tendência crescente. Logo, para uso dos modelos ARIMA, se faz necessário aplicar uma diferenciação na série.

**Gráfico 21 - Correlograma FAC
Série Temporal Concessão Crédito
INSS**



**Gráfico 22 - Correlograma FACP
Série Temporal Concessão Crédito
INSS**



Gráficos 23 - Séries, respectivamente, somente 1 diferença não sazonal, com 1 diferença não sazonal e 1 diferença sazonal e somente 1 diferença sazonal

Gráfico 23.1: Gráfico Diferença Série Concessão INSS com 1 diferença não Sazonal

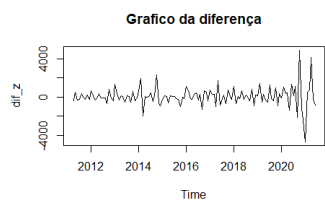


Gráfico 23.2: Gráfico Diferença Série Concessão INSS 1 diferença não sazonal e 1 diferença sazonal

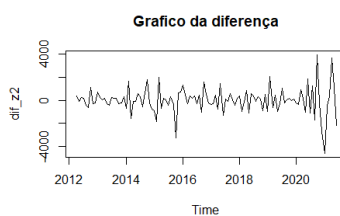


Gráfico 23.3: Gráfico Diferença Série Concessão INSS com 1 diferença sazonal

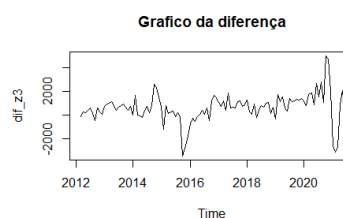


Gráfico 23.4: Correlograma
FAC Série Temporal
Concessão Crédito INSS com
1 diferença não Sazonal

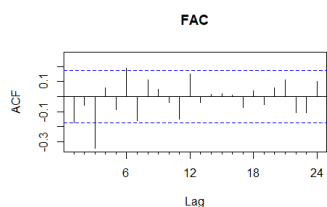


Gráfico 23.5: Correlograma
FAC Série Temporal
Concessão Crédito INSS com
1 diferença não sazonal e 1
diferença sazonal

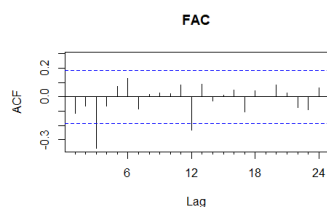


Gráfico 23.6: Correlograma
FAC Série Temporal
Concessão Crédito INSS com 1
diferença sazonal

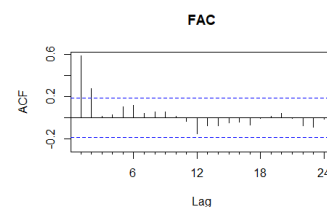


Gráfico 23.7: Correlograma
FACP Série Temporal
Concessão Crédito INSS com
1 diferença não Sazonal

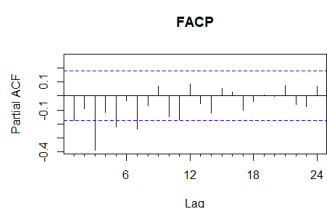


Gráfico 23.8: Correlograma
FACP Série Temporal
Concessão Crédito INSS com
1 diferença não sazonal e 1
diferença sazonal

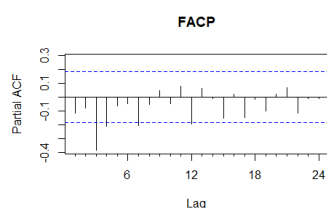
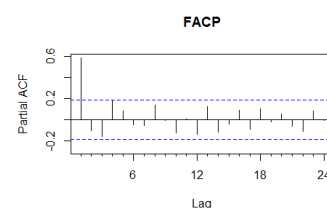


Gráfico 23.9: Correlograma
FACP Série Temporal
Concessão Crédito INSS com 1
diferença sazonal



Testando as diferenças com e sem sazonalidade no Gráfico 23, observa-se que a série se torna estacionária com 1 diferença não sazonal. A partir da análise de FAC e FACP da série diferenciada (dif. simples) observa-se que tem média móvel (MA(3)) e possivelmente a série tem sazonalidade de 12 e fator autoregressivo na parte sazonal de 1.

Pela análise dos correlogramas verificou-se os possíveis valores no fator MA(3), com uma diferença e fator AR(0) na parte sazonal verifica-se a necessidade do fator autoregressivo com valor 1 e com sazonalidade de 12, portanto, será utilizado também o método de mínimos quadrados para estimação foram ajustados dos parâmetros do modelo visando a comparação das estatísticas dos ajustes dos modelos e das estatísticas dos erros de previsão.

Na separação da série em período amostral e de validação para construção dos modelos, iremos considerar de março de 2011 a junho de 2021 como o período amostral, e assim, de julho a dezembro de 2021 teremos o período de validação.

3.3.1 Modelo 1

O primeiro modelo a ser apresentado é o modelo ARIMA (1,1,1). É um modelo com 1 diferenciação, 1 coeficiente auto-regressivo e 1 coeficiente de média móvel não sazonais.

Quadro 1 - Saída Computacional do Modelo

```
>modelo1<- Arima(Credito,order=c(1,1,1),seasonal=list(order=c(0,0,0),period=12),include.constant=F)
Series: Credito
ARIMA(1,1,1)

Coefficients:
      ar1      ma1
      0.4681 -0.8534
s.e.  0.1023  0.0500

sigma^2 = 866001:  log likelihood = -1014.59
AIC=2035.18  AICc=2035.39  BIC=2043.62

>coeftest(modelo1)
z test of coefficients:

      Estimate Std. Error  z value      Pr(>|z|)
ar1  0.468093   0.102258   4.5776      0.000004704 ***
ma1 -0.853363   0.050032  -17.0564 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>shapiro.test(padronizados1)
Shapiro-Wilk normality test

data:  padronizados1
W = 0.82252, p-value = 0.00000000006326

>sumario_Box.test
      [,1]      [,2]      [,3]      [,4]
Defasagem    12     24     36     48
Chi-cuadrada 24.1058 31.14697 33.91588 39.50337
Valor P      0.01967738 0.14959 0.5680726 0.8037855
```

A estimativa dos coeficientes da série usando o modelo ARIMA descrito nesta seção, observando-se que tanto o coeficiente auto-regressivo como o de média móvel ajustado pelo modelo são significativos, a um nível de significância de 5%.

Nos gráficos 24, é possível verificar que há correlação significativa nos resíduos, ou seja, possui uma correlação diferente de zero, com um nível de confiança de 95%, sendo este um indicativo que os resíduos do modelo não seguem o processo ruído branco.

Gráficos 24 - FAC e FACP dos Resíduos do Modelo 1 (95% de confiança)

Gráfico 24.1: Correlograma FAC Resíduos
Modelo1(95% confiança)

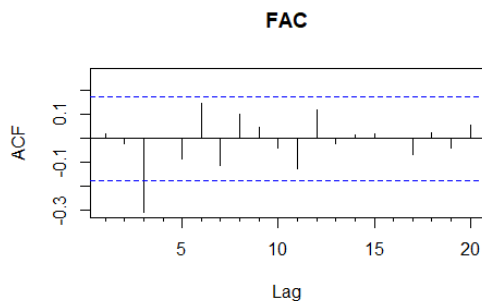
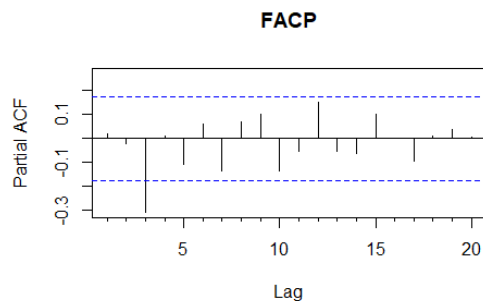


Gráfico 24.2: Correlograma FACP Resíduos
Modelo1(95% confiança)



Gráficos 25 - FAC e FACP dos Resíduos do Modelo 1 99% de confiança

Gráfico 25.1: Correlograma FAC Resíduos
Modelo1(99% confiança)

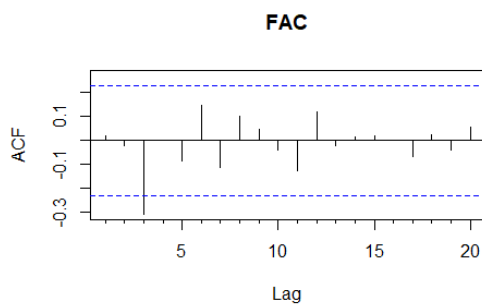
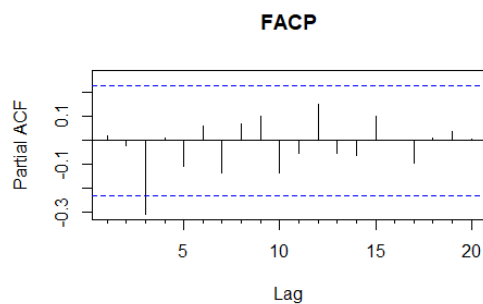


Gráfico 25.2: Correlograma FACP Resíduos
Modelo1(99% confiança)



Verificando as FAC e FACP dos resíduos do Modelo 1, nos gráficos 25, para o intervalo de confiança de 99%, os correlogramas dos resíduos indicam que não se trata de um ruído branco.

Analisando como são distribuídos e quanto a normalidade dos resíduos do Modelo 1, nos gráficos 26, observa-se que o histograma tende a distribuição normal e o q-q plot também apresenta uma tendência a normalidade, eliminando os valores discrepantes, porém, o teste de shapiro-wilk ($p\text{-valor} = 0,0000$) rejeitou a hipótese de que os resíduos do modelo 2, tenham distribuição normal.

Gráficos 26 - Resíduos Modelo 1

Gráfico 26.1: Distribuição Resíduos x Ordem dos dados Modelo1

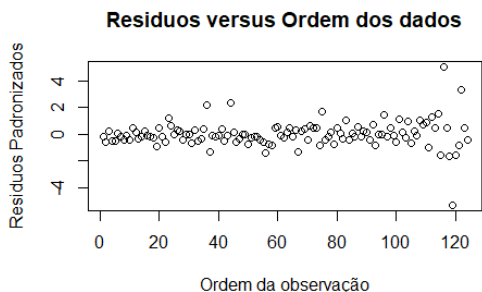
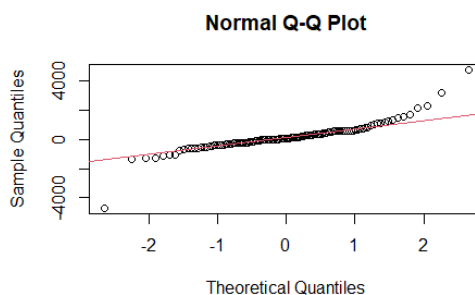


Gráfico 26.2: Histograma Resíduos Modelo1



Gráfico 26.3: Normalidade Modelo1



Pode-se verificar também pelo teste de Portmanteau, se o vetor de resíduos do modelo ajustado é um ruído branco, donde se o p -valor $> 0,05$ não rejeitamos H_0 , indicando que possui um processo ruído branco. Realizado o teste de Portmanteau para os resíduos do modelo 1, observou-se que o p -valor foi superior a 0,05 em todas as defasagens, exceto nas primeiras 12 defasagens onde este apresentou um p -valor baixo (p -valor $< 0,05$), mostrando que não se tem evidências para rejeitar a hipótese de que o erro segue um processo ruído branco, mesmo com as 12 primeiras defasagens sendo significativas (p -valor $< 0,05$), como pode ser comprovado no Quadro 1.

Mesmo não chegando a um ruído branco o modelo 1 foi selecionado por acompanhar bem a série de interesse.

3.3.2 Modelo 2

Em algumas situações pode ser muito difícil inferir um modelo aderente a partir da FAC e da FACP. Desta forma, a função auto.arima, que faz parte do

pacote *forecast* do software R. Onde a função retorna o modelo escolhido (conforme os valores de p , d e q), os termos autorregressivos e de média móvel, de “drift” (termo constante em modelo que tem $d > 0$ devido à sazonalidade), intercepto (média do termo $(1-L)^d X_t$, próximo à média da série histórica) e dos regressores (se for o caso), assim como o erro/desvio padrão (s.e. – erro padrão) de cada valor apresentado. Como o desvio padrão mostra o quanto de variação ou dispersão existente em relação à média, se todos os pontos estiverem exatamente sobre a linha de regressão, então a regressão explicaria toda a variação, ou seja, s.e. seria igual a zero. A função também imprime os critérios de informação AIC e BIC, e a variância (σ^2), que é definida como o quadrado do desvio padrão. Esta função do R usa uma variação do algoritmo desenvolvido por Hyndman e Khandakar (2008) que combina testes de raiz unitária e minimização do AIC para obter um modelo ARIMA.

O comando `auto.arima` indicou o modelo SARIMA (0,1,3)(1,0,0)[12] com intercepto. Trata-se de modelo com 1 diferenciação simples e 3 coeficientes de médias móveis na parte não sazonal e 1 coeficiente auto-regressivo na parte sazonal, com período de sazonalidade igual a 12, indicando que a série tem uma dependência no tempo t com o valor de $t = 12$.

A estimativa da série usando o modelo nos mostra que todos os coeficientes e o intercepto são significativos a um nível de significância de 5%.

Quadro 2 - Saída Computacional do Modelo 2

```
> modelo2<- auto.arima(z)
> modelo2
Series: z
ARIMA(0,1,3) (1,0,0) [12] with drift
Coefficients:
      ma1      ma2      ma3      sar1      drift
-0.3391 -0.1993 -0.3799  0.3486  47.2701
s.e.    0.0860  0.0985  0.0909  0.1229  11.1111

sigma^2 = 735036:  log likelihood = -1004.11
AIC=2020.22  AICc=2020.94  BIC=2037.09
>coeftest(modelo2)
z test of coefficients:
      Estimate Std. Error z value  Pr(>|z|)
ma1  -0.339101  0.086032 -3.9416  0.00008096 ***
```

```

ma2  -0.199261  0.098497 -2.0230  0.043071 *
ma3  -0.379866  0.090922 -4.1780  0.00002941 ***
sar1  0.348634  0.122950  2.8356  0.004574 **
drift 47.270100 11.111069  4.2543  0.00002097 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>shapiro.test(padronizados2)
Shapiro-Wilk normality test

data: residuos1
W = 0.87058, p-value = 0.000000005248
>sumario_Box.test

```

	[,1]	[,2]	[,3]	[,4]
Defasagem	12	24	36	48
Chi-cuadrada	8.28927	15.78988	17.87708	24.2561
Valor P	0.7621361	0.8955105	0.9950265	0.9983133

Os gráficos 27, que se refere as funções FAC e FACP dos resíduos indicam que pode não se tratar de um ruído branco, pois em uma das defasagens foi ultrapassado os respectivos limites da estimação intervalar, indicando a necessidade de inclusão de um parâmetro para eliminar essa correlação apresentada nos resíduos, com um nível de confiança de 95%.

Gráficos 27 - Funções FAC e FACP para os resíduos do Modelo 2

Gráfico 27.1: Correlograma FAC Resíduos
Modelo2 (95% confiança)

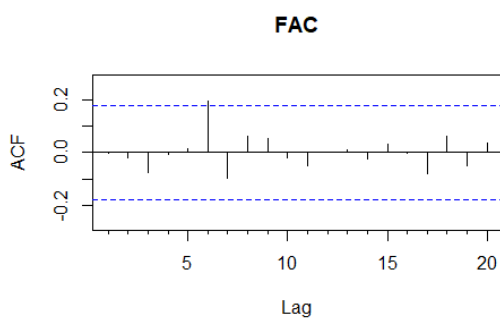
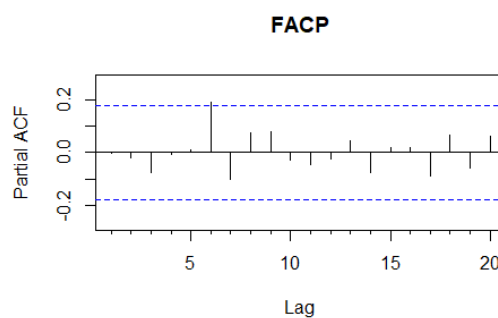


Gráfico 27.2: Correlograma FACP Resíduos
Modelo2 (95% confiança)



Para a análise dos resíduos está-se utilizando resíduos padronizados. No histograma é possível observar uma distribuição que tende a normal, sendo um

indicativo de normalidade, porém o Gráfico 28.3, mostra alguns pontos muito afastados da reta, o que indica que os resíduos podem não possuir distribuição normal, porém se desconsiderados os valores discrepantes, avalia-se que os resíduos tendem a distribuição normal.

Gráficos 28 - Resíduos do Modelo 2

Gráfico 28.1: Distribuição Resíduos x Ordem dos dados Modelo2

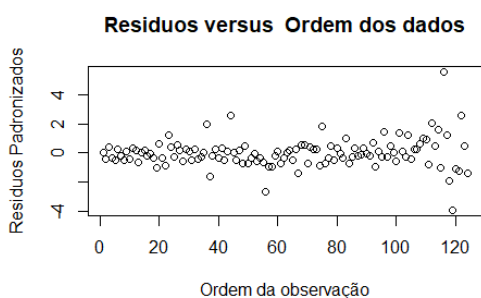


Gráfico 28.2: Histograma Resíduos Modelo2

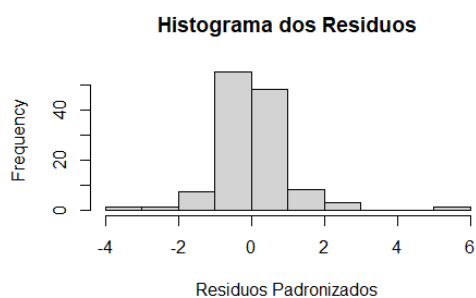
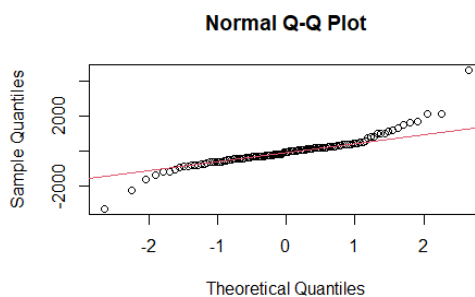


Gráfico 28.3: Normalidade Modelo1



O teste de normalidade de Shapiro-Wilk indica que podemos rejeitar a hipótese de que os resíduos são normalmente distribuídos (p -valor = 0,0000), mostrando que os resíduos não possuem distribuição normal com um nível de significância de 5%.

Pelo teste de Portmanteau pode-se verificar se o vetor de resíduos do modelo é um ruído branco. Sendo o p -valor acima de 0,05 então, não rejeitamos H_0 (ruído branco). O modelo não rejeita a hipótese de apresentar ruído branco, isto é, não rejeita a hipótese de correlação nula entre os resíduos.

3.3.3 Modelo 3

Como um dos objetivos do presente estudo é o de se utilizar variáveis de domínio público como preditoras que contribuam para melhor explicar o comportamento da série de concessões de empréstimo consignado do INSS, com a inclusão das variáveis preditoras que não foram retiradas no VIF (Tabela 10), sendo elas a PMMCCC, SELIC e o salário-mínimo, para a construção do modelo então foi ajustado modelos com estas variáveis preditoras.

Seguindo então a metodologia de regressão dinâmica, onde se indica incluir e retirar variáveis até o ajuste de um modelo considerado bom, dentro das premissas anteriormente citadas.

Após as análise feitas, o melhor modelo ajustado se mostrou o modelo SARIMA (0,0,2)(0,0,1)[12], com duas variáveis preditoras, a variável SELIC e Salário Mínimo. Outros modelos, apesar de até apresentarem bons resultados estatísticos, não apresentaram previsões próximas aos valores reais que estão nos dados para validação e por isso foram desconsiderados.

Trata-se de um modelo com 2 variáveis preditoras, 2 coeficientes de médias móveis não sazonais e 1 coeficiente de média móvel sazonal, com período de sazonalidade igual a 12, indicando que a série tem uma dependência no tempo $t=12$, como já verificado anteriormente. No Quadro 3, temos os resultados do modelo.

Quadro 3 - Saída Computacional do Modelo 3

```
>modelo3
Series: zx[, 1] # Variável EMCCC
Regression with ARIMA(0,0,2)(0,0,1)[12] 69rros

Coefficients:
      ma1      ma2      sma1      SELIC      Sal_Min
      0.5661  0.3981  0.245  -1682.495    7.0989
s.e.    0.0837  0.0881  0.105    455.729    0.4067

sigma^2 = 768506:  log likelihood = -1014.27
AIC=2040.53  AICc=2041.25  BIC=2057.45

>coeftest(modelo3)
z test of coefficients:
      Estimate  Std. Error z value      Pr(>|z|)
```

```

ma1      0.566150    0.083741  6.7607    0.000000000001373 ***
ma2      0.398063    0.088143  4.5161    0.00000629951356 ***
sma1     0.245044    0.105030  2.3331    0.0196432 *
SELIC   -1682.494536   455.729024 -3.6919    0.0002226 ***
Sal_Min   7.098936    0.406695 17.4552 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>shapiro.test(padronizados3)
Shapiro-Wilk normality test
data:  padronizados3
W = 0.8673, p-value = 0.000000003762

>sumario_Box.test

```

	[,1]	[,2]	[,3]	[,4]
Defasagem	12	24	36	48
Chi-cuadrada	8.812506	14.40641	19.48709	23.54984
Valor P	0.718856	0.9369326	0.9887134	0.9988469

As estimativas dos valores da série usando o modelo nos mostra todos os coeficientes são significativos a um nível de significância de 5% ($p\text{-valor} < 0,05$).

Os correlogramas dos resíduos do Modelo 3 com 95% de confiança, apresentados nos gráficos 29, mostram que não se trata de um ruído branco, pois tivemos correlação que ultrapassou os respectivos limites da estimação intervalar.

Gráficos 29 - FAC e FACP dos Resíduos do Modelo 3 (95% confiança)

Gráfico 29.1: Correlograma FAC Resíduos
Modelo3 (95% confiança)

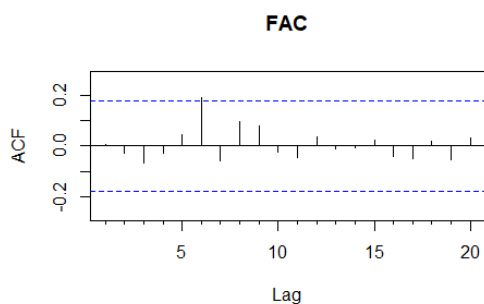
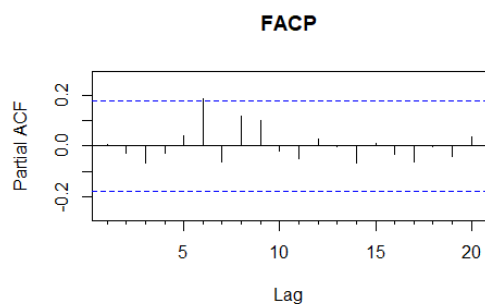


Gráfico 29.2: Correlograma FACP Resíduos
Modelo3 (95% confiança)



Porém para um intervalo de confiança de 99% os correlogramas dos resíduos indicam um ruído branco, como pode ser observado nos gráficos 30.

Gráficos 30 - FAC e FACP dos Resíduos do Modelo 3 (99% confiança)

Gráfico 30.1: Correlograma FAC Resíduos
Modelo3 (99% confiança)

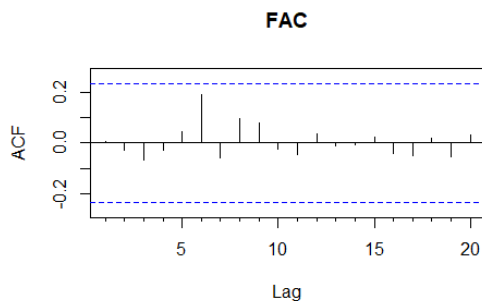
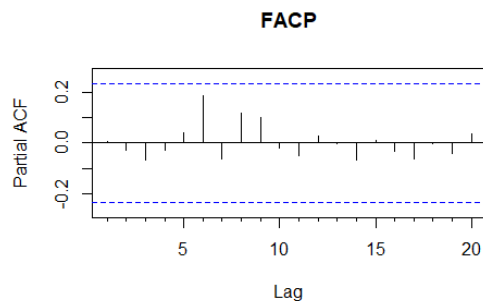


Gráfico 30.2: Correlograma FACP Resíduos
Modelo3 (99% confiança)



Os gráficos 31, referente aos resíduos do Modelo 3, mostra a partir do Gráfico 31.1 que os resíduos indicam possuir uma distribuição com média zero e com variância constante. No Gráfico 31.2 a distribuição aparenta ser simétrica, sendo um indicativo de normalidade, através do Gráfico 31.3 normalidade, tem-se mais um indício de normalidade dos resíduos, pois a suposição de normalidade não foi violada.

Gráficos 31 - Resíduos do Modelo 3

Gráfico 31.1: Distribuição Resíduos x Ordem
dos dados Modelo3

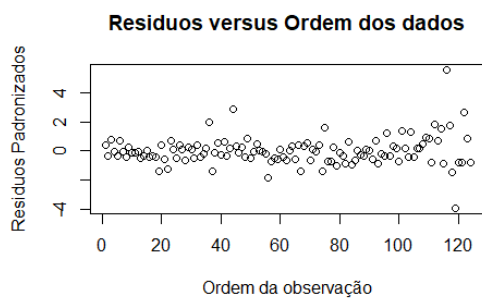


Gráfico 31.2: Histograma Resíduos Modelo3

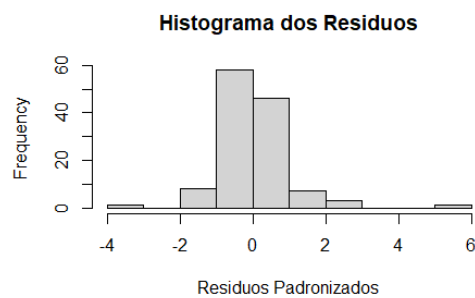
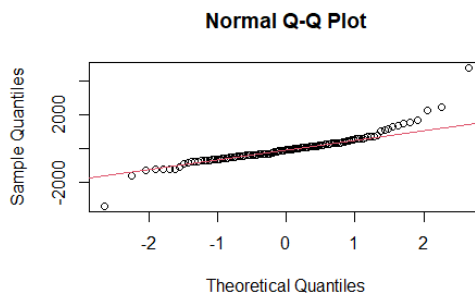


Gráfico 31.3: Normalidade Modelo3



O teste de normalidade de Shapiro-Wilk (onde a hipótese nula é de normalidade) indica que podemos rejeitar a hipótese de que os resíduos são normalmente distribuídos (p -valor = 0,0000), provavelmente o teste rejeitou a hipótese pelos valores discrepantes vistos no Gráfico 31.3, porém, verifica-se que sem esses valores, os resíduos estão distribuídos ao longo da reta de normalidade, o que indica a normalidade.

Na análise dos resíduos, pelo teste de Portmanteau, assim como no modelo anterior, o teste não rejeita a hipótese de apresentar ruído branco nos resíduos do modelo 3, isto é, não rejeita a hipótese de correlação nula entre os resíduos.

3.4 Valores da Estatística de Ajuste dos Modelos

Como descrito na seção 2, através dos estimadores usuais de mínimos quadrados é possível identificar o melhor modelo, para tanto serão utilizados os critérios AIC e BIC, cujo o objetivo é encontrar as ordens k e l que minimizam os critérios para a determinação das ordens p e q .

A seguir será apresentada tabela comparativa com os valores das estatísticas de Ajuste dos Modelos. Na análise dos resíduos, apresentados na Tabela 11, pelo teste de Portmanteau, os modelos 2 e 3 não rejeitaram a hipótese de apresentarem ruído branco, isto é, não rejeita a hipótese de correlação nula entre os resíduos. Já o modelo 1 rejeitou a hipótese de apresentar ruído branco nas primeiras 12 defasagens (P-Valor baixo).

Tabela 11 - Defasagens teste de Portmanteau (Q)

Defasagem	12	24	36	48
M1 ARIMA (1,1,1)	0,0197	0,1496	0,5681	0,8038
M2 SARIMA (0,1,3)(1,0,0)[12]	0,7621	0,8955	0,9950	0,9983
M3 SARIMAX (0,0,1)(0,0,2)[12]	0,7187	0,9369	0,9887	0,9988

Fonte: Elaborado pelo autor.

Então pelas estatísticas do modelo e pela análise dos resíduos, descritos na tabela 12, vê-se que o modelo mais aderente tende a ser o Modelo 2, o modelo SARIMA (0,1,3)(1,0,0)[12], pois apresentou os menores resultados de BIC e AIC.

Nos gráficos 32, que apresenta os gráficos dos valores estimados contra os dados reais podemos observar que os valores preditos acompanham muito bem a série real em todos os modelos, ficando por vezes os dados, praticamente sobrepostos, porém os modelos 2 e 3 entendem melhor a tendência da série quando comparado ao modelo 1.

Tabela 12 - Resumo das Estatísticas do Ajuste dos Modelos, março/2011 a junho/2021

Estatísticas	M1 ARIMA (1,1,1)	M2 SARIMA (0,1,3)(1,0,0)[12]	M3 SARIMAX (0,0,2)(0,0,1)[12]
Estatísticas do Modelo			
Constante(<i>b</i>)		47,27	
ϕ (ar1)	0,4681		
θ (ma1)	-0,8534	-0,3391	0,5661
θ (ma2)		-0,1993	0,3981
θ (ma3)		-0,3799	
Φ			
ϕ (sar1)		0,3486	
θ (sma1)			0,2450
β_1 (SELIC)			-1.682,50
β_2 (Salário Mínimo)			7,10
AIC	2.035,18	2.020,22	2.040,53
Schwarz (BIC)	2.043,62	2.037,09	2.057,45
Estatísticas dos Resíduos			
Média	154,51	-14,9400	-14,7300
Desvio padrão	909,86	839,6020	862,1500
Assimetria	0,2700	1,1096	1,3200

Curtose	13,7900	11,2634	11,6410
Teste Shapiro-wilk (p-valor)	0,0000	0,0000	0,0000

Fonte: Elaborado pelo autor.

Gráficos 32 - Estimativas Período Amostral dos Modelos Ajustados

Gráfico 32.1: Estimativas Período Amostral

Modelo1

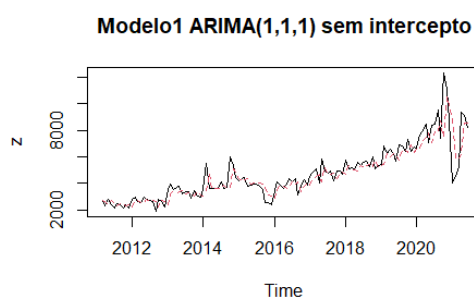


Gráfico 32.2: Estimativas Período Amostral

Modelo2

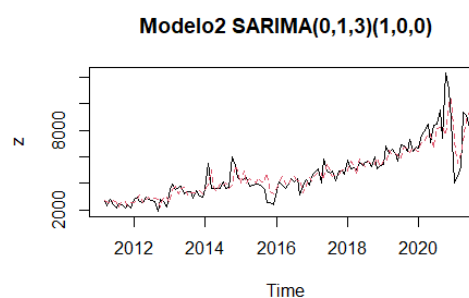
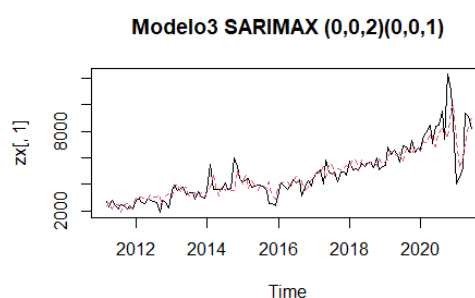


Gráfico 32.3: Estimativas Período Amostral Modelo3



3.5 Previsões para o período de validação

Fazendo uso dos modelos ajustados, foram calculadas as previsões dos últimos 6 meses da série original, compreendendo o período de julho/2021 a dezembro/2021, onde os dados desses meses não foram utilizados para ajustar os modelos e serão utilizados para avaliar a capacidade de predição dos mesmos, com seus limites superior e inferior usando 95% de confiança.

Nos gráficos 33, é mostrado os valores reais da EMCCC utilizados para validação (linha preta), os valores preditos pelos Modelos (linha vermelha) e o intervalo de confiança de 95% (linha verde), incluindo os limites inferior e superior, em cada respectivo Gráfico, vemos que a projeção não foi aderente para o Modelo 2 (Gráfico

33.2), pois é nítido que os dados reais a partir do mês de outubro não ficam dentro dos limites estabelecidos pelo intervalo de confiança de 95% para os modelos em questão, porém se for considerado somente os três primeiros meses de validação o Modelo 2 é que mais se aproxima dos dados reais.

Gráficos 33 - Previsões dos Modelos Ajustados

Gráfico 33.1: Previsão Período Validação do Modelo1 ARIMA (1,1,1)

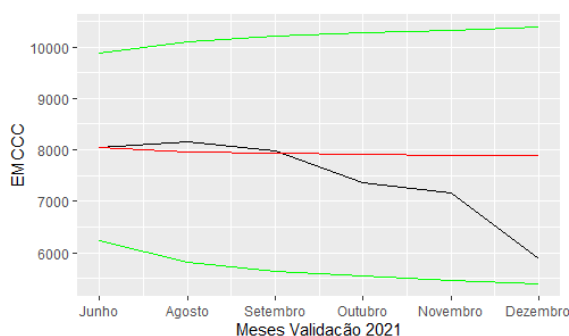


Gráfico 33.2: Previsão Período Validação do Modelo2 SARIMA (0,1,3)(1,0,0)[12]

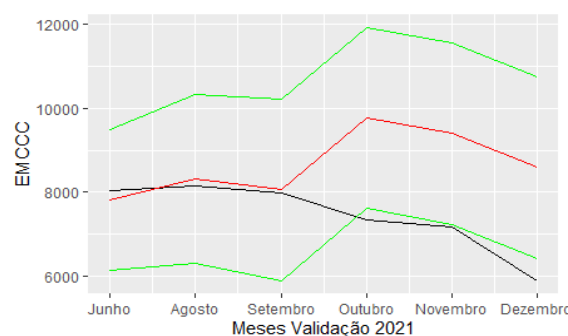
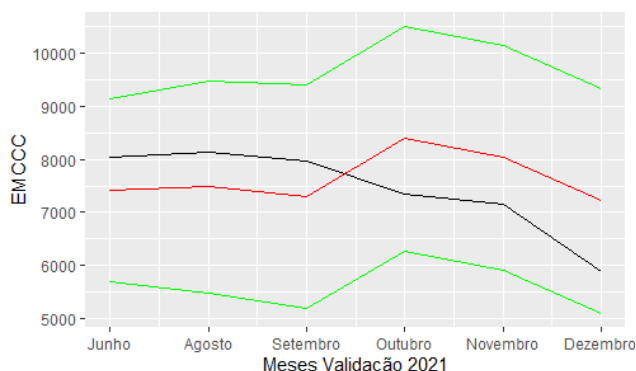


Gráfico 33.3: Previsão Período Validação do Modelo3 SARIMAX (0,0,2)(0,0,1)[12]



Para os modelos 1 e 3, os dados reais ficaram dentro dos limites de 95% de confiança, estabelecidos pelos próprios modelos, considerando estes dois modelos os dois fazem previsões bem próximas dos dados reais, porém o Modelo 3, é o que entende melhor a tendência da série no período de validação, ficando próximo aos dados reais em todo o período compreendido, somente errando a tendência nos meses de setembro à outubro.

3.6 Estatísticas dos Erros de Previsão

Por fim, iremos analisar as estatísticas dos erros do modelo. Através da tabela 13 é possível observar que pelo MPE as previsões encontradas para o Modelo1 estão acima dos valores reais, cerca de 6,50%, para o Modelo2 cerca de 13,36% e para o Modelo 3 aproximadamente 0,08%.

Já pelo MAPE, podemos verificar que no Modelo 1 aproximadamente 7,48% dos dados contêm erros de ajuste, no Modelo 2 o 14,35% dos dados contêm erros de ajustes e no Modelo 3 10,72%.

Pelo MAE, é possível concluir que o desvio padrão médio do Modelo 1 é de 557 unidades, no Modelo 2 é de 545 unidades e no Modelo 3 é de 570 unidades.

Tabela 13 - Resumo das Estatísticas dos Erros

	M1 ARIMA (1,1,1)	M2 SARIMA (0,1,3)(1,0,0)[12]	M3 SARIMAX (0,0,2)(0,0,1)[12]
Período Amostral, março/2011-junho-2021			
ME	-154,505	14,940	14,730
RMSE	919,266	836,343	858,789
MAE	557,105	545,387	570,291
MPE	-3,668	0,753	0,583
MAPE	12,051	11,401	12,073
Período Validação, julho/2021-dezembro/2021			
ME	513,093	1.231,398	18,344
RMSE	906,494	1.747,625	801,801
MAE	590,819	1.308,650	792,173
MPE	6,501	13,364	0,081
MAPE	7,478	14,353	10,725

Fonte: Elaborado pelo autor.

Desta forma, pode-se dizer que os resultados do Modelo 3 vão ao encontro da análise gráfica, podendo ser o Modelo 3 considerado aderente.

4 CONSIDERAÇÕES FINAIS

A análise comparativa das estatísticas de ajuste dos modelos deu indícios de que o Modelo 2 SARIMA (0,1,3)(1,0,0)[12] seria o mais adequado, pois apresenta menor valor BIC e AIC, de acordo com os dados analisados na série EMCCC.

Porém, no comparativo das estatísticas dos erros de previsão o Modelo 2 não figura como o mais indicado, pois não apresenta os menores resultados, o que até corrobora com a análise gráfica do Gráfico 33 referente aos dados reais do período de validação jul/2021-dez/2021 em comparação com os dados previstos pelos modelos ajustados.

Na avaliação das previsões e estatísticas dos erros envolvidos, é possível concluir que o Modelo 3 SARIMAX (0,0,1)(0,0,2)[12] apresenta melhor resultado de previsão considerando os 6 meses do período de validação, embora o Modelo 2 SARIMA (0,1,3)(1,0,0)[12], apresenta as melhores previsões para os três primeiros meses do período de validação, com isso, o Modelo 2 é melhor para prever valores próximos e o Modelo 3 é mais eficiente para prever valores em maiores períodos de tempo.

Levando-se em consideração o principal objetivo do presente estudo, que é o de projetar, com satisfatório grau de confiança, o volume de concessão de crédito consignado para aposentados e pensionistas do INSS através de técnicas de modelagem de séries temporais utilizando variáveis preditoras de domínio público, podemos concluir que este trabalho teve êxito, de forma que ao ajustarmos o modelo SARIMAX a série de interesse foi possível prever de forma satisfatória o comportamento da concessão de crédito consignado do INSS. Vale destacar que captar o impacto da pandemia de Covid19 na concessão de crédito foi um desafio a mais deste trabalho. Ressaltando que a série de interesse continha um total de 124 observações no período de março/2011 a jun/2021, além das 06 observações do ano de 2021 (jul/2021-dez/2021), utilizadas como validadoras para a previsão dos modelos ajustados, onde no Modelo 3 a previsão se aproximou do real e se manteve dentro do intervalo de confiança com 95% de confiança.

Em seguida, ao se ajustar os modelos de regressão linear múltipla pode-se também considerar que foi possível atender de forma satisfatória os objetivos secundários, uma vez que foi possível identificar as variáveis que têm maior impacto na concessão do crédito consignado no INSS, através do VIF. Mesmo os resíduos do modelo ajustado não apresentando normalidade segundo o teste de Shapiro-wilk, mas

na análise do q-q plot, verificou-se que a distribuição dos resíduos tende a normal, não considerando alguns valores discrepantes. É possível inferir que o Modelo 3 serve como bom parâmetro para aplicação estratégica no dia a dia que tenha interesse na variável EMCCC.

Vale ressaltar que para se chegar ao modelo final, mesmo que este não apresente normalidade nos resíduos, foram testados outros inúmeros modelos, sejam eles de ajuste de série temporal bem como de regressão linear múltipla com as variáveis SELIC e salário-mínimo, resultadas das análises VIF e sentido para o modelo. De fato, o modelo escolhido (Modelo 3) apresentados neste trabalho foram os que apresentaram previsões melhores, onde eles têm suas características para previsões em um curto período de tempo e longos períodos, onde foi privilegiada a facilidade de se aplicar este modelo em um ambiente corporativo em detrimento de modelos que seguissem todos os pressupostos estatísticos, mas que não possuíssem aplicabilidade real.

Evidentemente ainda é possível enriquecer este trabalho incluindo no modelo variáveis que sejam relevantes, correlacionadas com a variável EMCCC que é a variável resposta deste estudo. Variáveis estas que são informações que instituições que atuam na área de crédito devem ter em suas bases de dados. Pode-se também aplicar técnicas estatísticas diferenciadas das utilizadas neste trabalho, sejam técnicas mais avançadas aplicadas ao estudo de séries temporais, de regressão multivariada, ou ainda, técnicas diferentes das citadas como de *machine learning* ou redes neurais artificiais por exemplo.

REFERÊNCIAS

ABHAHAM, Bovas; LEDOLTER, Johannes. **Statistical Methods for Forecasting**. Nova Jersey: John Wiley & Sons, 1983.

BARTLEE, M. S. On the theoretical specification of sampling properties of autocorrelation time series. *In: Symposium on Autocorrelation in Time Series*, v.8, 1946. Oxford. **Anais [...]**. Oxford: Oxford University Press, 1946.

BANCO CENTRAL DO BRASIL. **SGS – Sistema Gerenciador de Séries Temporais**. Disponível em: <https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>. Acesso em: 05 jul. 2022.

BARROS, M., SOUZA, R. C. **Regressão Dinâmica. Núcleo de Estatística Computacional**. PUC-Rio, 1995.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control**. 5 ed. Nova York: HoldenDay ,1976.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control, San Francisco**. 5 ed. Nova York: Holden-Day, 1970.

BOX G. E. P.; JENKINS, G. M.; REINSEL G. C. **Time Series Analysis, Forecasting and Control**. 4 ed. Nova York: Wiley-Interscience. 2008.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. New York: John Wiley & Sons, 2015.

BOX, G. E.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. **Journal of the American Statistical Association**, Massachusetts v. 65, n.332, p. 1509-1526, 1970.

BOX, G. E. P.; TIAO, G. Intervention Analysis with Applications to Economic and Environmental Problems. **Journal of the American Statistical Association**, Massachusetts, v. 70, p.70–79, 1975.

BROCKWELL, P. J.; DAVIS, R. A. **Time Series: Theory and Methods**. New York, Springer, 1991.

CORRAR, L.; PAULO, E.; FILHO, J. **Análise Multivariada para os Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 1995.

DANCEY, C.; REIDY, J. **Estatística Sem Matemática para Psicologia: Usando SPSS para Windows**. Porto Alegre: Artmed, 2006.

EHLERS, R. S. **Análise de Séries Temporais**. Departamento de Estatística, UFPR, 2017. Disponível em: <http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>. Acesso em: 05 jul. 2021.

FERREIRA, F. B. F. **Estudo da demanda pelo crédito imobiliário habitacional destinado à aquisição de imóveis residenciais novos no período de 2006 a 2016**. Instituto de Ciências Exatas, UFMG, 2018.

FISHER, S. **Séries Univariantes de Tempo – Metodologia de Box & Jenkins**. Porto Alegre: Fundação de Economia Estatística, 1982.

GOODRICH, R. L., **Applied Statistical Forecasting**. Belmont: Business Forecast Systems, 1989.

HAIR, J.; ANDERSON, R.; TATHAM, R.; BLACK, W. **Análise Multivariada de Dados**. Porto Alegre: Bookman, 2005.

HYNDMAN, R. J., KOEHLER, A. B. Another look at measures of forecast accuracy. **Internacional Journal of Forecasting**, v. 22, p. 18, 2005.

HYNDMAN, R. J., ATHANASOPOULOS, G. **Forecasting: principles and practice**. [s.l.] O Texts, 2014.

HYNDMAN, R. J. et al. **Frecast**: Forecasting functions for time series and linear models. [s.l.], 2014. R package version 5.1. Disponível em: CRAN.R-project.org/package=forecast. Acesso em: 10 set. 2024.

HYNDMAN, R. J., ATHANASOPOULOS, G. **Forecasting: principles and practice**. [s.l.] O Texts, 2016.

JOHNSTON, J. e DINARDO, J. **Econometric Methods**. New York: John Wiley, 1998.

LIRA, S. A. **Análise de correlação**: abordagem teórica e de construção dos coeficientes com aplicações. Dissertação (Mestrado em Ciências) – Faculdade em Métodos Numéricos em Engenharia dos Setores, Ciências Exatas e Tecnologia Universidade Federal do Paraná. Curitiba, p. 209, 2004.

MORETTIN, P. A.; TOLOI, C. M. C. **Modelos para Previsão de Séries Temporais**. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1981.

MORETTIN, P. A.; TOLOI, C. M. C. **Previsão de Séries Temporais**. São Paulo: Atual Editora, 1987.

MORETTIN, P. A.; TOLOI, C. M. C. **Análises de Séries Temporais**. São Paulo: Blucher, 2006.

R CORE TEAM. R: **A Language and Environment for Statistical Computing**. Vienna, 2013. Disponível em: www.R-project.org/. Acesso em: 10 set. 2024.

SANTOS, A. M. **Análise de Regressão**. Disponível em: <http://www.pgsc.ufma.br/arquivos/aula1.pdf>. Acesso em: 20 set. 2022.

SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications**. Nova York: Springer, 2006.

STELLWAGEN, E.; GOODRICH, R. **Forecast Pro Statistical Reference Manual**. Belmont: Business Forecast Systems, 2011.

WEI, W. W. S. **Time Series Analysis**. Nova York: Addison-Wesley Publishing Company, 1994.

YOGARAJAH, B.; ELANKUMARAN, C.; VIGNESWARAN, R. Application of ARIMAX model for forecasting paddy production in Trincomalee district in Sri Lanka. Proceedings of the Third International Symposium, 2013. **Anais** [...]. Oluvil: SEUSL, 2013.

ZUUR, A. F. et. al. **Mixed effects models and extensions in ecology with R**. Nova York: Springer, 2009.