

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Inflação de alimentos no Brasil: análise e previsão a partir de
modelo ARIMA**

Gustavo Jun Yakushiji

Trabalho de conclusão de curso apresentado como
parte dos requisitos para obtenção do título de
Engenheiro Agrônomo

**Piracicaba
2022**

Gustavo Jun Yakushiji

Inflação de alimentos no Brasil: análise e previsão a partir de modelo ARIMA

Orientador:

Prof. Dr. **CRISTIAN MARCELO VILLEGAS LOBOS**

Trabalho de conclusão de curso apresentado como
parte dos requisitos para obtenção do título de
Engenheiro Agrônomo

**Piracicaba
2022**

SUMÁRIO

RESUMO.....	4
ABSTRACT.....	5
LISTA DE FIGURAS.....	6
LISTA DE TABELAS.....	9
LISTA DE ABREVIATURAS E SIGLAS.....	11
1 INTRODUÇÃO.....	13
2 REVISÃO BIBLIOGRÁFICA.....	17
2.1 Índice Nacional de Preços ao Consumidor Amplo (IPCA).....	17
2.2 Modelos de previsão.....	23
3 MATERIAL E MÉTODOS.....	35
3.1 Coleta de dados.....	35
3.2 Processamento e manipulação de dados.....	36
3.3 Modelagem ARIMA.....	37
4 RESULTADOS E DISCUSSÃO.....	39
4.1 Categoria Pecuária.....	39
4.2 Categoria Grãos.....	47
4.3 Categoria Hortícolas.....	54
5 CONCLUSÕES.....	61
REFERÊNCIAS.....	63
APÊNDICES.....	67

RESUMO

Inflação de alimentos no Brasil: análise e previsão a partir de modelo ARIMA

A inflação é definida como o aumento dos preços de produtos e serviços dentro de uma economia. No Brasil, um dos índices que medem a inflação ao consumidor é o Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Entre 2007 e 2021, o Índice de Preços de Alimentação e Bebida (IPAB) - um dos grupos que compõem o IPCA - apresentou crescimento acumulado superior ao IPCA, indicando inflação mais acentuada dos alimentos. Em termos sociais, a inflação de alimentos atinge mais intensamente a população de menor renda, cujo dispêndio com alimentação é relativamente mais alto. Atualmente, cerca de 125,2 milhões de brasileiros apresentam algum grau de insegurança alimentar, representando mais da metade da população. Neste contexto, este trabalho se propõe a desenvolver modelos ARIMA para previsão, à curto prazo, da inflação de alimentos ao consumidor no Brasil. Foram avaliadas seis cadeias agropecuárias - bovinocultura de leite, bovinocultura de corte, arroz, feijão, mandioca e tomate - utilizando dados da variação mensal da inflação do IPCA entre julho de 2006 a setembro de 2022. As análises foram realizadas no ambiente computacional R. Para verificar a estacionariedade das séries de inflação, utilizou-se o teste de raiz unitária Dickey-Fuller Aumentado. O melhor modelo ARIMA(p, d, q), para cada cadeia agropecuária, foi definido a partir da função *auto.arima()*, do pacote *forecast*. Por fim, foram realizadas previsões de inflação das cadeias agropecuárias entre outubro de 2022 e março de 2023. Todas as séries de inflação analisadas apresentaram comportamento estacionário. Os modelos com melhores avaliações foram: AR(5) para bovinocultura de leite, MA(1) para bovinocultura de corte, ARMA(1,1) para arroz, MA(2) para feijão, AR(1) para mandioca e ARMA(3,1) para tomate. Verificou-se que, para as cadeias da bovinocultura de leite, do arroz e da mandioca, o ajuste dos valores do modelo apresentou comportamento próximo aos dados reais. Já para as cadeias da bovinocultura de corte, do feijão e do tomate, os modelos se ajustaram aos dados, porém não acompanharam com precisão a volatilidade da inflação. A partir dos modelos, foi possível prever estabilidade na variação de inflação das cadeias da bovinocultura de corte, do arroz e do feijão; aumento de inflação na cadeia da bovinocultura de leite e queda nas cadeias da mandioca e do tomate.

Palavras-chave: Séries temporais, IPCA, Cadeias agropecuárias, Insegurança alimentar

ABSTRACT

Food inflation in Brazil: analysis and forecast using ARIMA model

Inflation is defined as the increase of the prices of products and services in the economy. In Brazil, one of the indices that measure consumer inflation is the Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculated by the Instituto Brasileiro de Geografia e Estatística (IBGE). Between 2007 and 2021, the Índice de Preços de Alimentação e Bebida (IPAB) - one of the groups that compose the IPCA - presented a higher accumulated growth than the IPCA, indicating more pronounced food and beverage inflation. In social terms, food inflation affects more intensely the lower income population, whose expenditure on food is relatively higher. Currently, about 125.2 million Brazilians suffer some degree of food insecurity, representing more than half of the population. In this context, this paper proposes to develop ARIMA models for short-term forecasts of consumer food inflation in Brazil. Six agricultural chains - dairy cattle, beef cattle, rice, beans, cassava and tomatoes - were evaluated, using IPCA data of the monthly variation of inflation from July 2006 to September 2022. The analyses were done in the R computing environment. To verify the stationarity of the inflation series, the Augmented Dickey-Fuller unit root test was used. The best ARIMA(p, d, q) model, for each agricultural chain, was defined using the function *auto.arima()*, from the *forecast* package. Finally, inflation forecasts were made for the agricultural chains between October 2022 and March 2023. The best models for the agricultural chains were: dairy cattle AR(5), beef cattle MA(1), rice ARMA(1,1), beans MA(2), cassava AR(1) and tomatoes ARMA(3,1). It was verified that for the dairy cattle, rice and cassava chains, the adjustment of the model's values presented a behavior close to the real data. However, for the beef cattle, bean, and tomato chains, the model did not accurately follow the volatility of inflation. Based on the models, it was possible to predict stability in the variation of inflation for beef cattle, rice and beans; increase of inflation for dairy cattle and decrease for cassava and tomato.

Keywords: Time series, IPCA, Agricultural chains, Food insecurity

LISTA DE FIGURAS

Figura 1 - Evolução dos níveis de insegurança alimentar no Brasil, 2004 a 2022.....	14
Figura 2 - 16 regiões metropolitanas utilizadas para calcular o IPCA, destacadas no mapa.....	18
Figura 3 - Estrutura de agregação do IPCA, exemplificado com os subgrupos, itens e subitens do grupo de alimentação e bebidas.....	19
Figura 4 - Variação da inflação anual acumulada do índice geral (IPCA) e do Índice de Preços de Alimentação e Bebida (IPAB), entre 1995 e 2021.....	21
Figura 5 - Correlogramas da função de autocorrelação (FAC ou <i>ACF – Autocorrelation function</i>), com intervalo de confiança de 95%. À esquerda, um típico processo não estacionário; à direita, um processo estacionário.....	26
Figura 6 - Padrões típicos de correlogramas FAC (<i>ACF – autocorrelation function</i>) e FACP (<i>PACF – partial autocorrelation function</i>). De cima para baixo, estão representados os modelos AR(1), MA(1) e ARMA(1,1).....	30
Figura 7 - Séries temporais da categoria Pecuária, cadeias Bovinocultura de corte e Bovinocultura de leite, entre julho de 2006 a setembro de 2022.....	39
Figura 8 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia de Bovinocultura de corte.....	40
Figura 9 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia de Bovinocultura de leite.....	41
Figura 10 - Correlograma da FAC dos resíduos da Cadeia da Bovinocultura de corte.....	43
Figura 11 - Correlograma da FAC dos resíduos da Cadeia da Bovinocultura de leite.....	43
Figura 12 - Valores ajustados pelo modelo MA(1) da cadeia da Bovinocultura de corte, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	44
Figura 13 - Valores ajustados pelo modelo AR(5) da cadeia da Bovinocultura de leite, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	44
Figura 14 - Previsão da variação da inflação da cadeia da Bovinocultura de corte, a partir do modelo MA(1), entre outubro de 2022 e março de 2023.....	45

Figura 15 - Previsão da variação da inflação da cadeia da Bovinocultura de leite, a partir do modelo AR(5), entre outubro de 2022 e março de 2023.....	46
Figura 16 - Séries temporais da categoria Grãos, cadeias Arroz e Feijão, entre julho de 2006 a setembro de 2022.....	47
Figura 17 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Arroz.....	48
Figura 18 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Feijão.....	49
Figura 19 - Correlograma da FAC dos resíduos da Cadeia do Arroz.....	50
Figura 20 - Correlograma da FAC dos resíduos da Cadeia do Feijão.....	50
Figura 21 - Ajuste dos valores do modelo ARMA(1,1), da cadeia do Arroz, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	51
Figura 22 - Ajuste dos valores do modelo MA(2), da cadeia do Feijão, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	51
Figura 23 - Previsão da variação da inflação da cadeia do Arroz, a partir do modelo ARMA(1,1), entre outubro de 2022 e março de 2023.....	52
Figura 24 - Previsão da variação da inflação da cadeia do Feijão, a partir do modelo MA(2), entre outubro de 2022 e março de 2023.....	53
Figura 25 - Séries temporais da categoria Hortícolas, cadeias Mandioca e Tomate, entre julho de 2006 a setembro de 2022.....	54
Figura 26 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia da Mandioca.....	55
Figura 27 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Tomate.....	56
Figura 28 - Correlograma da FAC dos resíduos da Cadeia da Mandioca.....	57
Figura 29 - Correlograma da FAC dos resíduos da Cadeia da Tomate.....	57
Figura 30 - Ajuste dos valores do modelo AR(1), da cadeia da Mandioca, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	58

Figura 31 - Ajuste dos valores do modelo ARMA(3,1), da cadeia do Tomate, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.....	58
Figura 32 - Previsão da variação da inflação da cadeia da Mandioca, a partir do modelo AR(1), entre outubro de 2022 e março de 2023.....	59
Figura 33 - Previsão da variação da inflação da cadeia do Tomate, a partir do modelo ARMA(3,1), entre outubro de 2022 e março de 2023.....	60

LISTA DE TABELAS

Tabela 1 - Relação temporal entre edições da POF e períodos do IPCA, a partir de 1997.....	18
Tabela 2 - Estrutura do grupo alimentação e bebidas em níveis de subgrupo, item e subitem, com base na POF 2017-2018.....	20
Tabela 3 - Participação dos grupos do IPCA na despesa da população nas POFs 2002-2003, 2008-2009 e 2017-2018.....	22
Tabela 4 - Contribuição dos grupos do IPCA para a inflação entre 2007 e 2021.....	23
Tabela 5 - Padrões geométricos típicos da FAC e FACP.....	30
Tabela 6 - Descrições das tabelas utilizadas para coleta dos dados de inflação.....	35
Tabela 7 - Categorização das cadeias agropecuárias, de acordo com a origem agrícola.....	36
Tabela 8 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Bovinocultura de corte e Bovinocultura de leite.....	40
Tabela 9 - Coeficientes e erro padrão do modelo MA(1) da cadeia Bovinocultura de corte.....	41
Tabela 10 - Coeficientes e erro padrão do modelo AR(5) da cadeia Bovinocultura de leite.....	42
Tabela 11 - Estatística-Q de Ljung-Box para a categoria Pecuária, cadeias Bovinocultura de corte e Bovinocultura de leite.....	42
Tabela 12 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95% para a cadeia da Bovinocultura de corte, outubro de 2022 a março de 2023.....	45
Tabela 13 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia da Bovinocultura de leite, outubro de 2022 a março de 2023.....	46
Tabela 14 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Arroz e Feijão.....	47
Tabela 15 - Coeficientes e erro padrão do modelo ARMA(1,1) da cadeia do Arroz.....	48
Tabela 16 - Coeficientes e erro padrão do modelo MA(2) da cadeia do Feijão.....	49
Tabela 17 - Estatística-Q de Ljung-Box para a categoria Grãos, cadeias Arroz e Feijão.....	49

Tabela 18 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Arroz, outubro de 2022 a março de 2023.....	52
Tabela 19 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Feijão, outubro de 2022 a março de 2023.....	53
Tabela 20 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Mandioca e Tomate.....	54
Tabela 21 - Coeficientes e erro padrão do modelo AR(1) da cadeia da Mandioca.....	55
Tabela 22 - Coeficientes e erro padrão do modelo ARMA(3,1) da cadeia do Tomate.....	56
Tabela 23 - Estatística-Q de Ljung-Box para a categoria Hortícolas, cadeias Mandioca e Tomate.....	57
Tabela 24 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia da Mandioca, outubro de 2022 a março de 2023.....	59
Tabela 25 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Tomate, outubro de 2022 a março de 2023.....	60

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i> - Critério de Informação Akaike
ARIMA	<i>Autoregressive Integrated Moving Average</i> - Autorregressivo Integrado de Média Móvel
BIC	<i>Bayesian Information Criterion</i> - Critério de Informação Bayesiano
CEPAL	Comissão Econômica para a América Latina e o Caribe
DFA	Dickey-Fuller Aumentado
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
IA	Insegurança Alimentar
IBGE	Instituto Brasileiro de Geografia e Estatística
IPAB	Índice de Preços de Alimentação e Bebida
IPCA	Índice Nacional de Preços ao Consumidor Amplo
ODS	Objetivos de Desenvolvimento Sustentáveis
PNAD	Pesquisa Nacional por Amostra de Domicílios
POF	Pesquisa de Orçamentos Familiares
Rede Penssan	Rede Brasileira de Pesquisa em Soberania Alimentar e Segurança Alimentar e Nutricional
SIDRA	Sistema IBGE de Recuperação Automática
SQR	Soma de quadrados dos resíduos
VIGISAN	Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da COVID-19 no Brasil

1. INTRODUÇÃO

A inflação é definida como o aumento dos preços de produtos e serviços dentro de uma economia, sendo medida a partir de índices de preços (IBGE, 2022a). Esses índices podem ser calculados com base em diferentes cestas de bens e serviços, variando de acordo com os objetivos do levantamento, como exemplo, serem direcionados a um grupo específico, como consumidores ou produtores (MARTINEZ; CERQUEIRA, 2011).

No Brasil, o Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), é um dos índices que medem a inflação ao consumidor. O governo federal brasileiro utiliza o IPCA como índice oficial de inflação, servindo de referência para definir metas inflacionárias e alterações na taxa de juros (IBGE, 2022a).

Ao analisar o IPCA, observa-se que, entre 2007 e 2019, o Índice de Preços de Alimentação e Bebida (IPAB) - um dos nove grupos que compõem o IPCA - apresentou um crescimento acumulado de 155,7%, superior ao IPCA, que se elevou em 103,5%. Durante a pandemia de COVID-19, entre 2020 e 2021, notou-se a mesma tendência, com um aumento do IPAB em 23,2%, enquanto do IPCA, em 15,0% (BACCARIN; OLIVEIRA, 2021).

Em termos sociais, a inflação de alimentos atinge mais intensamente a população de menor renda – com menor poder aquisitivo, cujo dispêndio com alimentação é relativamente mais alto que o de outras camadas sociais. Em consequência, pode ocorrer diminuição da quantidade de alimentos adquirida ou a substituição daqueles com maior qualidade nutricional, pelos com menor, agravando a situação de insegurança alimentar e nutricional (BACCARIN; OLIVEIRA, 2021).

Segundo a Rede Brasileira de Pesquisa em Soberania Alimentar e Segurança Alimentar e Nutricional (REDE PENSSAN, 2022), insegurança alimentar (IA) ocorre quando uma pessoa não tem acesso regular e permanente a alimentos, sendo classificada em três níveis:

- IA leve: ocorre quando há incerteza em relação ao acesso a alimentos em um período curto e/ou quando a qualidade da alimentação é comprometida;

- IA moderada: quando a quantidade de alimentos é insuficiente;
- IA grave: privação no consumo de alimentos - caracterizando a fome em sua forma mais grave.

De acordo com a Pesquisa de Orçamentos Familiares (POF) 2017-2018, 25,3 milhões de domicílios no Brasil apresentavam algum nível de IA - seja leve, moderada ou grave - o que representa 36,7% dos domicílios no Brasil (IBGE, 2019). Comparada às três pesquisas anteriores - realizadas pela Pesquisa Nacional por Amostra de Domicílios (PNAD), a partir da mesma metodologia, a IA em 2017-2018 apresentou o maior índice, tendo um aumento de 33,3% na IA leve em relação a 2004 e 62,2% em relação a 2013, além de um aumento de 76,1% na IA moderada e 43,7% na IA grave, quando comparadas a 2013 (IBGE, 2019; IBGE, 2020a).

Em dezembro de 2020, a Rede Penssan (2021) realizou o "I Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da COVID-19 no Brasil" - I VIGISAN, apontando que 116,8 milhões de pessoas apresentavam algum grau de IA durante o período, retrocedendo a patamares do ano de 2004. Em 2021-2022, no segundo inquérito (II VIGISAN), a Rede Penssan (2022) constatou que a IA se tornou ainda mais presente, com 125,2 milhões de pessoas apresentando algum grau de IA, representando mais da metade da população brasileira - 58,7%. A Figura 1 ilustra a evolução dos níveis de insegurança alimentar no Brasil, entre 2004 e 2022, de acordo com as pesquisas anteriormente citadas.

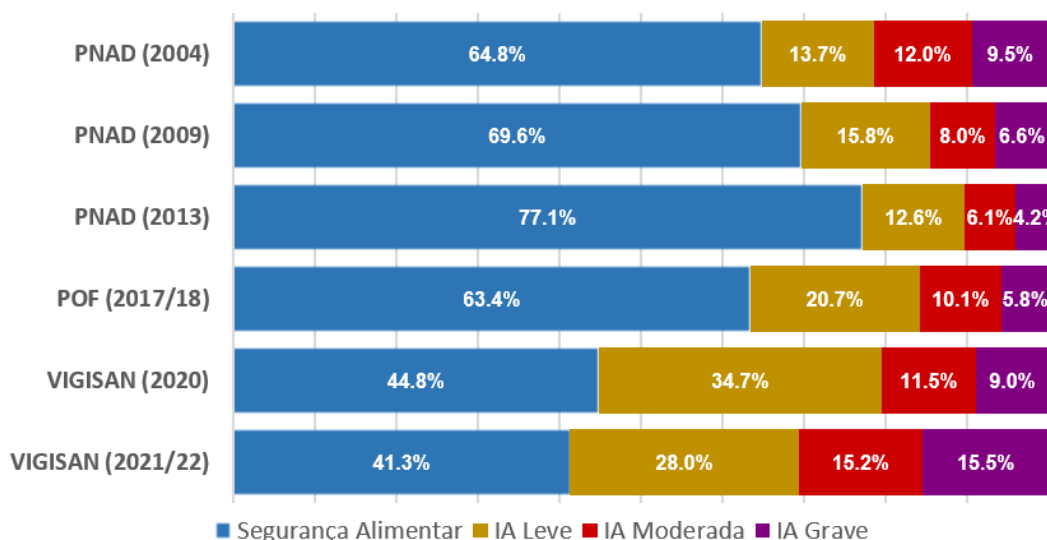


Figura 1 - Evolução dos níveis de insegurança alimentar no Brasil, 2004 a 2022.

Fonte: Elaborado pelo autor, baseado em Rede Penssan (2022).

O crescimento da insegurança alimentar não foi um fenômeno observado somente no cenário brasileiro. No contexto mundial, em 2014, o Índice de Prevalência de Desnutrição havia apresentado significativas melhoras ao longo das últimas duas décadas. Segundo o relatório anual das Nações Unidas de 2015 sobre a fome (FAO; IFAD; WFP, 2015), houve uma redução de 216 milhões de pessoas em situação de desnutrição em relação ao período de 1990-1992. De acordo com o mesmo documento, pôde-se observar um decréscimo acentuado nas regiões em desenvolvimento, mesmo com o significativo crescimento populacional.

Apesar do progresso, a partir de 2015, observou-se uma reversão da tendência, havendo um acréscimo de 60 milhões de pessoas afetadas pela fome no período entre 2015 e 2019, totalizando 690 milhões de pessoas em situação de desnutrição, cerca de 8,9% da população global (FAO; IFAD; UNICEF; WFP; WHO, 2020).

No relatório mais recente das Nações Unidas, em 2021, 56,5 milhões de pessoas passavam fome, enquanto 268 milhões estavam em situação de IA e 823 milhões em situação de desnutrição – mais da metade localizadas no continente asiático; mais de um terço, no africano; e 7,4% na América Latina e Caribe (FAO; IFAD; UNICEF; WFP; WHO, 2022). Segundo o mesmo relatório, desde 2015, a fome quase dobrou na América do Sul, além de ter impulsionado o número de pessoas com IA na América Latina e Caribe em 2020. Ainda, alerta-se para o fato de que o problema de insegurança alimentar e nutricional não está se restringindo apenas à grupos sociais que vivem na pobreza, mas também àqueles que não a vivenciavam antes.

Segundo levantamento da Comissão Econômica para a América Latina e o Caribe (CEPAL, 2020), órgão que atua nos assuntos referentes à América Latina e Caribe, a partir de 2014, observou-se uma queda no PIB *per capita* da região, junto a um aumento brusco no índice de pobreza extrema. A previsão para 2020 indicava uma retração em 5,3% do PIB *per capita*, junto ao crescimento de 2,3 pontos percentuais na extrema pobreza, totalizando 13,5% da população latino-americana e caribenha, sendo este o maior índice registrado nas últimas décadas. Somado a tais indicativos, a CEPAL ainda alertava sobre o aumento dos preços na região, em que se observou uma elevação de 3,3% entre janeiro e maio de 2020, enquanto, globalmente, ocorreu uma queda de 9,1% do mesmo índice.

Tendo em vista este cenário regional e global, a desaceleração e choques econômicos observados em diversos países – além de conflitos e mudanças climáticas - desencadeou em um menor poder de compra da população, principalmente àquelas de menor renda, configurando-se como um dos principais fatores responsáveis pelo aumento dos índices de fome, desnutrição, pobreza extrema e insegurança alimentar (FAO; IFAD; UNICEF; WFP; WHO, 2022; CEPAL, 2020), tornando a meta do “Fome Zero” - um dos 17 Objetivos de Desenvolvimento Sustentáveis (ODS), estipulados na “Agenda 2030 para o Desenvolvimento Sustentável” (ONU, 2015) – um imenso desafio.

Dado o contexto previamente elaborado, este trabalho propõe desenvolver modelos ARIMA para previsão, à curto prazo, da inflação de alimentos ao consumidor no Brasil, utilizando dados de julho de 2006 a setembro de 2022. Por sua vez, como objetivos específicos, almeja-se:

- I. Classificar os alimentos que compõem o IPCA em cadeias agropecuárias;
- II. Desenvolver, selecionar e ajustar modelos ARIMA para a inflação das cadeias agropecuárias;
- III. Realizar previsão da inflação das cadeias agropecuárias, à curto prazo, entre outubro de 2022 e março de 2023.

Com os resultados do trabalho, espera-se proporcionar contribuição acadêmica e de políticas públicas para uma melhor elucidação da temática que afeta diretamente grande parte da população brasileira, principalmente a porção socialmente vulnerável. Conforme Braun (2007), analisar e interpretar tendências recentes e desafios emergentes na situação alimentar – dentre elas, a inflação de alimentos - é essencial para fornecer aos formuladores de políticas as informações necessárias para mobilizar respostas adequadas nos níveis local, nacional, regional e internacional, além de ser fundamental para auxiliar no ajuste adequado das agendas de pesquisa em agricultura, nutrição e saúde.

2. REVISÃO BIBLIOGRÁFICA

2.1 Índice Nacional de Preços ao Consumidor Amplo (IPCA)

2.1.1 Definições

O IPCA - sigla para Índice Nacional de Preços ao Consumidor Amplo - é um dos índices que medem a inflação no Brasil, sendo calculado pelo Instituto Brasileiro de Geografia e Estatística - IBGE. Seu objetivo é medir a variação de preços de uma cesta de produtos e serviços consumida pela população-alvo, sendo ela:

(...) famílias residentes nas áreas urbanas dos estados cobertos pelo índice cujos rendimentos familiares monetários disponíveis mensais, oriundos de qualquer fonte, totalizam montante contido entre 1 e 40 salários-mínimos (IBGE, 2020b; IBGE, 2022a).

O cálculo do IPCA tem como base a Pesquisa de Orçamentos Familiares (POF), também conduzida pelo IBGE, apresentando as edições de 1987/88, 1995/96, 2002/2003, 2008/2009 e 2017/2018 (IBGE, 2020b; IBGE, 2022a). A POF, dentre outras questões, tem como objetivo verificar o que a população consome e quanto do rendimento familiar é gasto em cada produto ou serviço. Com esse levantamento, define-se a cesta padrão da população-alvo de cada região metropolitana pesquisada (IBGE, 2020b; IBGE, 2022a).

Dessa maneira, a partir da cesta padrão, é possível verificar a participação de cada produto na despesa total das famílias, estabelecendo-se uma estrutura de ponderação para cada produto e serviço em cada uma das localidades pesquisadas. Na composição final da estrutura de ponderação, não são computados todos os produtos levantados na pesquisa, mas apenas os bens e serviços que, em termos de despesa, atingem um valor mínimo (IBGE, 2012; IBGE, 2020b).

A Tabela 1 associa as edições da POF com os respectivos períodos utilizados para o cálculo do IPCA. Há sempre uma defasagem entre o levantamento de campo da POF e seu uso na estrutura de ponderação do IPCA, período esse em que são feitas correções e adequações pelos técnicos do IBGE (IBGE, 2020b).

Tabela 1 - Relação temporal entre edições da POF e períodos do IPCA, a partir de 1997.

Edição da POF	Vigência do IPCA
POF 1987-88	Janeiro 1997 - Julho 1999
POF 1996-97	Agosto 1999 - Junho 2006
POF 2002-2003	Julho 2006 - Dezembro 2011
POF 2008-2009	Janeiro 2012 - Dezembro 2019
POF 2017-2018	Janeiro 2020 - Atual

Fonte: Elaborado pelo autor, baseado em IBGE (2020b).

**Figura 2** - 16 regiões metropolitanas utilizadas para calcular o IPCA, destacadas no mapa.

Fonte: Elaborado pelo autor, baseado em IBGE (2020b).

O IPCA é calculado mensalmente a partir de um levantamento de preços em 16 áreas urbanas do país (Figura 2). Todos esses preços são comparados com os preços do mês anterior, cujas médias ponderadas resultam em um único valor que reflete a variação geral de preços ao consumidor no período, valor esse denominado índice oficial de inflação ao consumidor ou índice geral (IBGE, 2020b; IBGE, 2022a).

Destaca-se que as regiões de Vitória (ES) e Campo Grande (MS) foram incluídas no cálculo do IPCA a partir de janeiro de 2014. Mais recentemente, a partir de maio de 2018, outras três regiões foram adicionadas ao cômputo do IPCA, sendo elas Aracaju (SE), Rio Branco (AC) e São Luís (MA) (IBGE, 2020b).

A estrutura do IPCA é organizada em quatro níveis categóricos, sendo o *grupo* o nível mais agregado, em seguida do *subgrupo*, *item* e, por fim, o menor nível de agregação, o *subitem*, que abrange os produtos e serviços efetivamente levantados na POF. A Figura 3 ilustra a estrutura de agregação do IPCA, exemplificado com o grupo alimentação e bebidas.

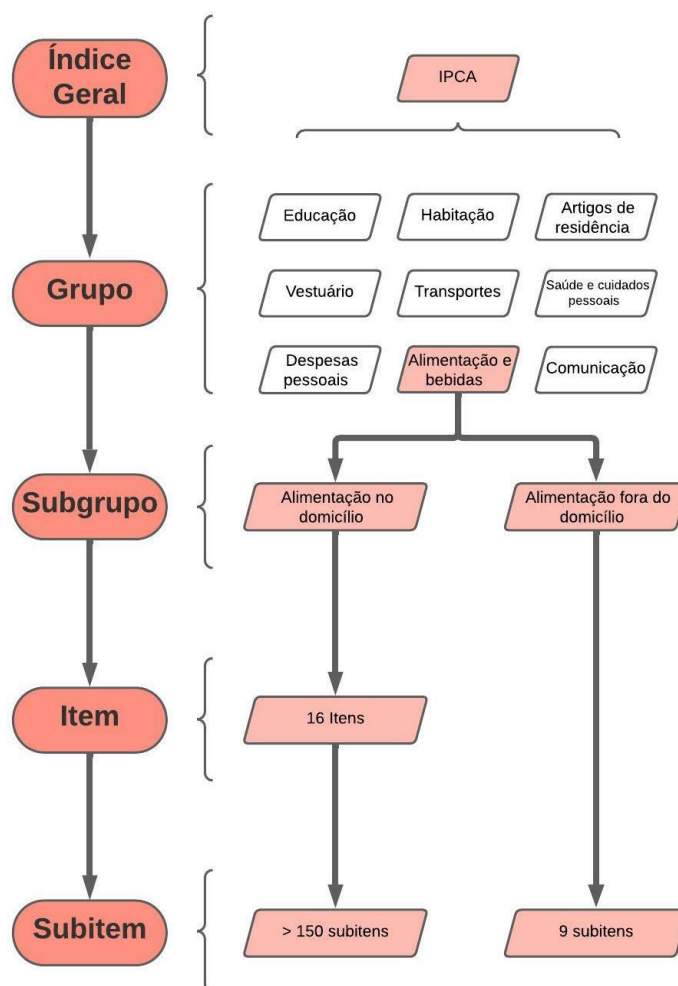


Figura 3 - Estrutura de agregação do IPCA, exemplificado com os subgrupos, itens e subitens do grupo de alimentação e bebidas.

Fonte: Elaborado pelo autor, baseado em IBGE (2020b).

Desde a POF 1996/1997, o IPCA apresenta nove grupos, sendo eles: alimentação e bebidas, educação, habitação, artigos de residência, vestuário,

transportes, saúde e cuidados pessoais, despesas pessoais e comunicação. Cada um dos grupos apresenta suas respectivas subdivisões em subgrupos, itens e subitens (IBGE, 2020b).

O grupo de alimentação e bebidas, referido pela sigla IPAB (Índice de Preços de Alimentos e Bebidas), apresenta dois subgrupos: alimentação no domicílio e alimentação fora do domicílio. O subgrupo alimentação no domicílio contempla 16 itens, os quais abrangem mais de 150 subitens. Já o subgrupo alimentação fora do domicílio não apresenta a categoria item, contemplando 9 subitens a partir da POF 2017-2018 (IBGE, 2020b). A Tabela 2 apresenta as categorias do grupo alimentação e bebidas.

Tabela 2 - Estrutura de agregação do grupo alimentação e bebidas em níveis de subgrupo, item e subitem, com base na POF 2017-2018.

Grupo	Alimentação e Bebidas			
Subgrupo	Alimentação no Domicílio		Alimentação Fora do Domicílio	
Item	Cereais, leguminosas e oleaginosas	Carnes e peixes industrializados	-	
	Farinhas, féculas e massas	Aves e ovos		
	Tubérculos, raízes e legumes	Leites e derivados		
	Açúcares e derivados	Panificados		
	Hortaliças e verduras	Óleos e gorduras		
	Frutas	Bebidas e infusões		
	Carnes	Enlatados e conservas		
	Pescados	Sal e condimentos		
Subitem	>150 subitens		Refeição	Vinho
			Lanche	Doces
			Refrigerante e água mineral	Sorvete

Fonte: Elaborado pelo autor, baseado em IBGE (2020b).

Vale destacar que, entre uma POF e outra, sempre ocorrem alterações no estrato de subitem, sendo alguns produtos incluídos ou excluídos, a depender do consumo das famílias. Além disso, há diferenças na composição de subitens entre as 16 regiões metropolitanas, dada as particularidades de consumo em cada local (IBGE, 2020b).

2.1.2 Panorama da inflação de alimentos no Brasil

No Brasil, pôde-se observar uma alteração no cenário de inflação de alimentos ao longo das últimas décadas. A Figura 4 compara a evolução da inflação anual acumulada do Índice Geral nacional (IPCA) com a do Índice de Preços de Alimentos e Bebidas (IPAB), entre 1995 e 2021.

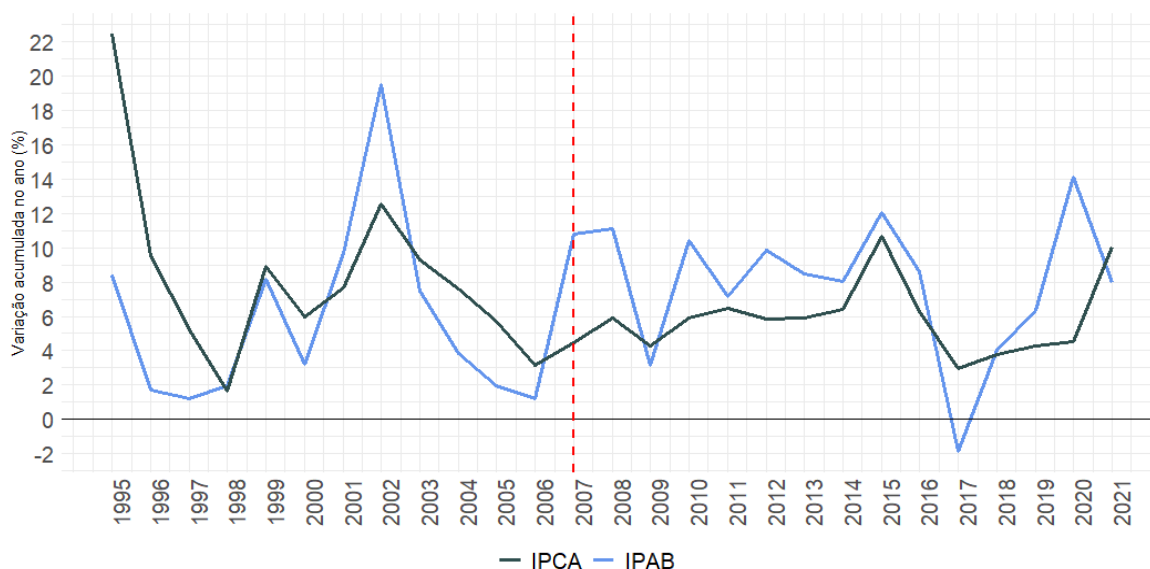


Figura 4 - Variação da inflação anual acumulada do índice geral (IPCA) e do Índice de Preços de Alimentação e Bebida (IPAB), entre 1995 e 2021.

Fonte: Elaborado pelo autor, baseado em IBGE (2022b).

No período de 1995 a 2006, observou-se que, em apenas três anos (1998, 2001 e 2002) o IPAB foi maior que o IPCA. Estudo de Farina e Nunes (2002) aponta que, entre 1994 e 2002, os preços de alimentos apresentaram uma queda real de 20% devido a influência de diferentes mecanismos, como o aumento de produtividade e desenvolvimento da cadeia agroindustrial nacional. De 2001 a 2006, o IPAB continuou a apresentar variação inferior ao IPCA, com um acúmulo de inflação de 50,9%, inferior aos 55,3% do índice geral no período (BACCARIN; OLIVEIRA, 2021).

Contudo, Baccarin e Oliveira (2021) observaram que, a partir de 2007, houve uma inversão no cenário. Entre 2007 e 2019, o IPAB apresentou valores superiores ao IPCA em 11 anos, à exceção dos anos de 2009 e 2017. No período, o IPAB apresentou variação acumulada de 155,68%, enquanto o IPCA variou em 103,48%. Na pandemia de COVID-19, entre 2020 e 2021, constatou-se a mesma tendência observada desde 2007, com o IPAB apresentando variação de 23,15%, superior ao do IPCA, de 15,03%.

Desde a POF 2002-03, o grupo de alimentação e bebidas alternou com o grupo de transportes o posto de agrupamento com maior participação nas despesas da população (Tabela 3). Nas POF 2002-2003, 2008-2009 e 2017-2018, o grupo de alimentação e bebidas apresentou participações de 20,43%, 22,08% e 19,35%, respectivamente (IBGE, 2022b).

Tabela 3 - Participação dos grupos do IPCA na despesa da população nas POFs 2002-2003, 2008-2009 e 2017-2018.

Grupo	POF 2002-2003	POF 2008-2009	POF 2017-2018
Transportes	21,10%	21,95%	20,60%
Alimentação e bebidas	20,43%	22,08%	19,35%
Habitação	13,62%	14,28%	15,59%
Saúde e cuidados pessoais	10,62%	11,08%	13,53%
Despesas pessoais	9,20%	9,19%	10,73%
Educação	7,16%	4,18%	6,15%
Comunicação	6,55%	5,57%	5,71%
Vestuário	6,43%	6,23%	4,58%
Artigos de residência	4,90%	5,43%	3,75%
Índice Geral	100%	100%	100%

Fonte: Elaborado pelo autor, baseado em IBGE (2022b).

Em relação à contribuição dos grupos para a inflação, entre 2007 e 2019, o IPAB representou 31,97% da inflação, mais que o dobro do agrupamento seguinte, o de habitação, com 15,31% (IBGE, 2022b). Durante o período da pandemia, entre 2020 e 2021, o IPAB representou 29,85%, ficando atrás apenas do grupo de transportes, com 30,59%, sendo que ambos os agrupamentos corresponderam a 60,44% da inflação no período (IBGE, 2022b). A Tabela 4 apresenta a contribuição dos grupos para a inflação.

Tabela 4 - Contribuição dos grupos do IPCA para a inflação entre 2007 e 2021.

Grupo	2007 - 2019	2020 - 2021
Transportes	12,68%	30,59%
Alimentação e bebidas	31,97%	29,85%
Habitação	15,31%	19,73%
Saúde e cuidados pessoais	12,84%	4,74%
Artigos de residência	1,71%	4,70%
Despesas pessoais	13,03%	4,15%
Vestuário	4,78%	2,77%
Comunicação	0,77%	1,85%
Educação	6,90%	1,63%
Índice Geral	100%	100%

Fonte: Elaborado pelo autor, baseado em IBGE (2022b).

2.2 Modelos de Previsão

2.2.1 Séries temporais

Séries temporais são conjuntos de dados coletados, sequencialmente, ao longo do tempo. Podem ser utilizados para a formulação de modelos estruturais de causa e efeito, além de modelos univariados ou multivariados de previsão (MAIA, 2013). Nos modelos univariados de previsão, a previsão de um valor futuro é dada em função dos seus valores passados (MAIA, 2013).

Ainda, pode-se definir uma série temporal como uma realização de um processo estocástico – ou processo aleatório – apresentando uma estrutura possível de ser caracterizada e descrita (MAIA, 2013). As inferências sobre a natureza do processo gerador da série temporal são dadas a partir de uma amostra do processo estocástico, sendo essa uma realização particular de todas as possibilidades (GUAJARATI; PORTER, 2011; MAIA, 2013). Assim, embora se observe apenas um valor em um dado período de uma série temporal, deve-se considerar que esta poderia assumir infinitas possibilidades de ocorrência, segundo uma dada distribuição de probabilidade, possibilitando extrair inferências sobre o processo estocástico subjacente (GUAJARATI; PORTER, 2011; MAIA, 2013).

2.2.2 Estacionariedade

As séries temporais requerem algumas propriedades desejáveis para a realização de previsões, baseada em seu comportamento passado. Dentre elas está a estacionariedade, ou seja, comportamento constante ao longo do tempo. Caso contrário, diz-se que a série apresenta comportamento não estacionário (MAIA, 2013). Assumindo que Y seja um processo estocástico e y_t , uma variável aleatória, sendo t o número de n períodos, tem-se:

$$Y = \{y_t, t = 1, 2, \dots, n\}$$

Desta forma, o conceito de estacionariedade pode ser definido de duas maneiras. A primeira é a estacionariedade estrita, que supõe que a distribuição de todos os y_t não mudem ao longo do tempo. Por ser impossível conhecer todas as suas distribuições conjuntas, adota-se, na prática, a segunda definição de estacionariedade: a estacionariedade fraca, que apresenta as seguintes propriedades: (MAIA, 2013)

- I. Média constante: $E(y_t) = \mu$;
- II. Variância constante: $var(y_t) = \sigma^2$;
- III. Covariância entre y_t e seu valor defasado y_{t+k} não depender do tempo, apenas da defasagem: $\gamma_k = E[(y_t - \mu)(y_{t+k} - \mu)]$.

Dessa forma, um processo estocástico fracamente estacionário apresenta média e variância constantes ao longo do tempo, além da covariância entre dois períodos ser independente em relação ao tempo, ou seja, não serem autocorrelacionados (GUAJARATI; PORTER, 2011). Por outro lado, um processo estocástico não estacionário possui variância e/ou média que variam no tempo.

Portanto, a estacionariedade se apresenta como um conceito importante para a formulação de modelos de previsão, uma vez que supõe que o relacionamento entre y_t e seus valores defasados seja o mesmo em todos os períodos t , na qual a série segue uma distribuição de probabilidades previsível, sendo possível a realização de generalizações para períodos futuros (GUAJARATI; PORTER, 2011; MAIA, 2013).

2.2.2.1 Ruído branco

Outro processo estocástico importante para as séries temporais são os processos puramente aleatórios ou ruído branco. O ruído branco apresenta as seguintes propriedades: possui média zero; variância constante; e não apresenta autocorrelação (GUAJARATI; PORTER, 2011; MAIA, 2013). No modelo clássico de regressão linear, o erro (ε_t) é um ruído branco, independente e identicamente distribuído de maneira normal (IIDN), dado por: $\varepsilon_t \sim IIDN(0, \sigma^2)$ e chamado de processo de ruído branco gaussiano (GUAJARATI; PORTER, 2011; MAIA, 2013). Assim, caso o processo estocástico seja um ruído branco, a série será estacionária.

Por outro lado, caso o ruído branco não siga uma distribuição normal, denomina-se como sendo um passeio aleatório, isto é, não estacionária (GUAJARATI; PORTER, 2011).

2.2.2.2 Raiz unitária

Ao representar um modelo de regressão em função de seus valores defasados, se a soma dos coeficientes associados às variáveis defasadas for igual a 1, dizemos que o modelo possui raiz unitária, sendo assim, uma série não estacionária. Caso a soma dos coeficientes for menor que 1, o modelo não apresenta raiz unitária, sendo uma série estacionária (GUAJARATI; PORTER, 2011; MAIA, 2013). Dessa forma, os termos não estacionariedade, passeio aleatório e raiz unitária podem ser tratados como sinônimos (GUAJARATI; PORTER, 2011).

2.2.2.3 Função de autocorrelação (FAC)

A função de autocorrelação (FAC) descreve – a partir do gráfico denominado correlograma - o relacionamento entre valores presentes e valores defasados de uma série temporal, sendo possível verificar a influência dos valores defasados em relação aos valores originais. Dessa forma, pode-se identificar a estacionariedade de um processo estocástico a partir do correlograma da FAC (MAIA, 2013).

A não estacionariedade é verificada quando os valores presentes e defasados apresentam elevada autocorrelação, ou seja, o coeficiente de autocorrelação começa com um valor alto e diminui muito lentamente em direção a zero à medida que a defasagem aumenta, além de apresentar defasagens fora do intervalo de confiança, sendo significativamente diferentes de zero. Já em processos estacionários, a correlação tende a zero, apresentando defasagens dentro do intervalo de confiança, não sendo estatisticamente significantes (GUAJARATI; PORTER, 2011; MAIA, 2013). A Figura 5 ilustra os correlogramas da FAC de processo estacionário (gráfico à esquerda) e não estacionário (gráfico à direita).

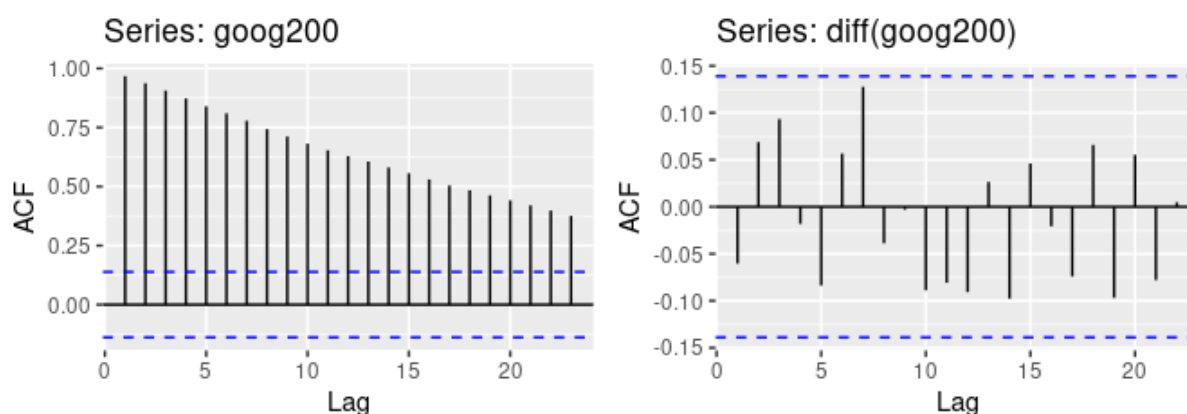


Figura 5 - Correlogramas da função de autocorrelação (FAC ou ACF – *Autocorrelation function*), com intervalo de confiança de 95%. À esquerda, um típico processo não estacionário; à direita, um processo estacionário.

Fonte: Hyndman e Athanasopoulos (2018)

2.2.2.4 Testes de raiz unitária

Em complemento à análise do correlograma da FAC, testes estatísticos podem ser utilizados para verificar a estacionariedade de uma série temporal. A seguir, será apresentado um desses possíveis testes, o de Dickey-Fuller Aumentado (DFA).

O teste de Dickey-Fuller Aumentado (DFA) se baseia no teste precedente dos mesmos autores - o teste Dickey-Fuller (1979) - na qual a hipótese nula é de não estacionariedade e a alternativa, de estacionariedade, tendo como distribuição de probabilidade a estatística tau (τ), apresentando valores críticos definidos de acordo com a metodologia dos autores (GUAJARATI; PORTER, 2011; MAIA, 2013). Contudo,

enquanto o teste de Dickey-Fuller considera que os erros são não correlacionados, o teste DFA considera defasagens da variável dependente entre os regressores como forma de controlar a autocorrelação nos erros (GUAJARATI; PORTER, 2011; MAIA, 2013).

2.2.3 Modelagem ARIMA

O modelo de processo Autorregressivo Integrado de Média Móvel - ARIMA, proposto por Box e Jenkins (1978), é utilizado para prever valores futuros, com base nos valores passados, ou defasados, de uma determinada série temporal (MAIA, 2013), cujo método consiste em ajustar uma curva adequada à determinados dados históricos para realizar previsões (GUAJARATI; PORTER, 2011). Dessa maneira, baseia-se na análise estocástica das propriedades da própria série temporal (GUAJARATI; PORTER, 2011). No caso dos modelos univariados de previsão, a previsão de um valor futuro é dado unicamente em função dos valores passados (MAIA, 2013).

Além disso, os modelos ARIMA pressupõem séries temporais estacionárias ou que possam se tornar estacionárias, mediante a aplicação de uma ou mais diferenças (MAIA, 2013).

Os modelos ARIMA são compostos por três processos: autorregressivo (AR), integrativo (I) e de médias móveis (MA), podendo ser aplicados de forma individual ou a partir de combinações. A seguir, serão definidos os modelos propostos pelos autores, bem como suas combinações.

2.2.3.1 Modelo Autorregressivo (AR)

No modelo autorregressivo (AR), a previsão é realizada com base na combinação linear de valores passados - ou defasados - da série em questão, sendo descrito pela seguinte equação:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t ,$$

em que y_t é a variável estudada, y_{t-p} o seu valor defasado de ordem p , α são os valores médios de y , ϕ_p o parâmetro associado ao p -ésimo valor de y_{t-p} , e ε_t é um erro aleatório com características de ruído branco.

Dessa maneira, temos um modelo autorregressivo y_t de ordem p - AR(p) –, onde a observação corrente de y_t é gerada por uma média ponderada de p observações passadas mais uma perturbação aleatória no período corrente (MAIA, 2013).

2.2.3.2 Modelo de Médias Móveis (MA)

Os modelos de médias móveis (MA) utilizam os erros de previsão anteriores para realizar a previsão de valores futuros. Um processo de médias móveis de ordem q - MA(q) - pressupõe que cada observação de y_t seja gerada por uma média invariável no tempo (μ), mais uma soma ponderada de q observações defasadas das perturbações aleatórias, mais a perturbação do período corrente (ε_t) (MAIA, 2013), isto é:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t ,$$

em que θ_q é o q -ésimo parâmetro associado ao erro de previsão ε_{t-q} .

2.2.3.3 Modelo Autorregressivo e de Médias Móveis (ARMA)

Em um processo autorregressivo e de médias móveis de ordens p e q - ARMA(p, q) - considera-se que cada observação presente de y_t seja gerada pela combinação do processo autorregressivo de ordem p e do processo de médias móveis de ordem q , dada por: (MAIA, 2013)

$$y_t = \delta + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

2.2.3.4 Modelo Autorregressivo Integrado e de Médias Móveis (ARIMA)

O modelo de previsão ARIMA complementa os modelos anteriormente citados ao incluir o processo integrado (I), que possibilita modelar séries não estacionária a partir de diferenciações (MAIA, 2013).

Por definição, a d -ésima diferenciação de uma série y_t ($\Delta^d y_t$) é dada por:

$$\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}$$

Um processo é integrado de ordem $d - I(d)$ – quando, após ser diferenciado d vezes, se torna estacionário (MAIA, 2013). Sendo $\Delta^d y_t$ um processo ARMA (p, q), e após diferenciar d vezes uma série y_t , tem-se que y_t é um processo autorregressivo integrado e de médias móveis de ordens p, d e q - ARIMA (p, d, q), representado pela seguinte expressão: (MAIA, 2013)

$$\Delta^d y_t = \delta + \phi_1 \Delta^d y_{t-1} + \dots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

2.2.3.5 Identificação dos parâmetros do modelo

Os correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) auxiliam a identificar os valores apropriados de p, d e q do modelo ARIMA (p, d, q).

A FAC - discutida anteriormente – é um dos métodos para verificar a estacionariedade - junto aos testes de raiz unitária - logo, permitem identificar o número de diferenciações necessárias para tornar a série estacionária, dessa forma, definir a ordem d do processo integrado – $I(d)$.

A função de autocorrelação parcial (FACP) mede a correlação entre as observações da série temporal, separadas por k períodos, depois de verificar a correlação nas defasagens intermediárias, menores que k . Dessa maneira, mede-se a correlação entre y_t e y_{t-k} após remover o efeito das defasagens intermediárias (GUAJARATI; PORTER, 2011).

Geometricamente, a FAC e a FACP dos processos $AR(p)$, $MA(q)$ e $ARMA(p, q)$ apresentam padrões típicos, o que possibilita identificar as possíveis ordens dos modelos (GUAJARATI; PORTER, 2011). A Tabela 5 descreve os padrões de FAC e FACP, de acordo com o tipo de modelo. Já a Figura 6 ilustra exemplos de correlogramas com padrões típicos de modelos $AR(1)$, $MA(1)$ e $ARMA(1,1)$.

Tabela 5 - Padrões geométricos típicos da FAC e FACP.

Modelo	Padrão típico de FAC	Padrão típico de FACP
$AR(p)$	Declinação exponencial ou padrão de onda senoide amortecida ou ambos	Picos significativos até p defasagens
$MA(q)$	Picos significativos até q defasagens	Declinação exponencial
$ARMA(p, q)$	Queda exponencial	Queda exponencial

Fonte: Elaborada pelo autor, baseado em Gujarati e Porter (2011).

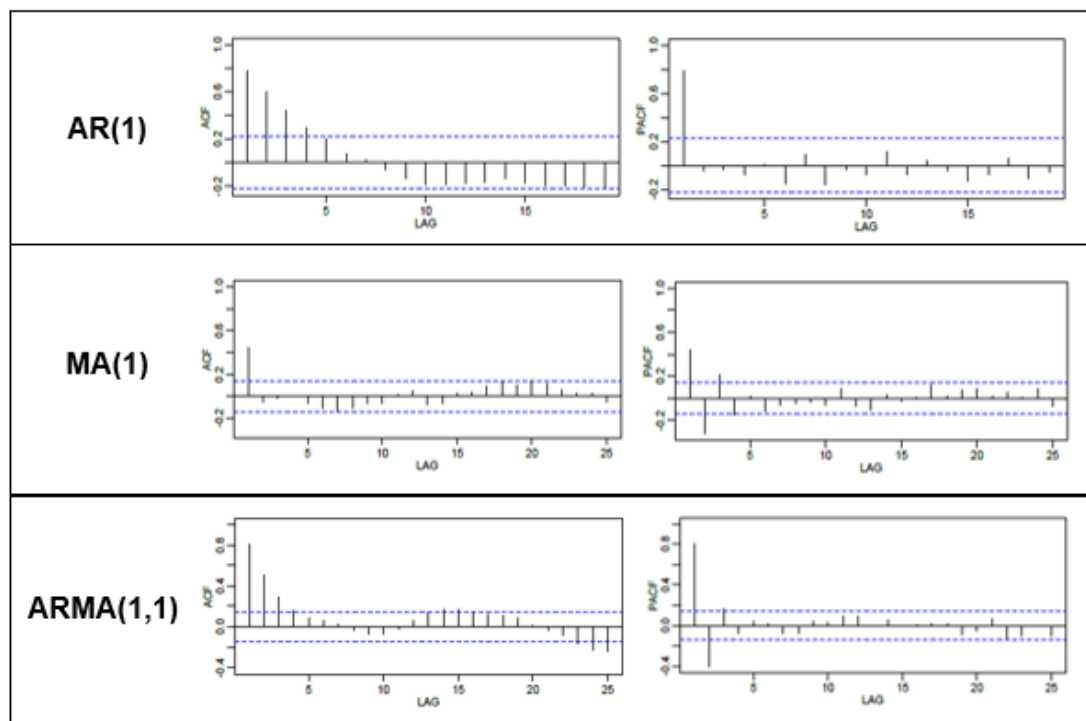


Figura 6 - Padrões típicos de correlogramas FAC (ACF – autocorrelation function) e FACP (PACF – partial autocorrelation function). De cima para baixo, estão representados os modelos $AR(1)$, $MA(1)$ e $ARMA(1,1)$.

Fonte: Elaborado pelo autor, com base em Pennsylvania State University (2022).

2.2.3.6 Função *auto.arima*

Em relação à identificação das ordens do modelo ARIMA a partir dos correlogramas, Gujarati e Porter (2011) e Hyndman e Athanasopoulos (2018) citam o desafio de selecionar os valores apropriados para p , d e q , requerendo habilidade, experiência e prática para o desenvolvimento de modelos de previsão. Para auxiliar na estimação dos valores de p , d e q , diversos softwares possuem funcionalidades que possuem rotinas para estimar o melhor modelo ARIMA.

Como exemplo, o pacote *forecast* (HYNDMAN et al., 2021), disponível no software R (R CORE TEAM, 2022), apresenta a função *auto.arima* (), que utiliza uma variação do algoritmo Hyndman-Khandakar (HYNDMAN; KHANDAKAR, 2008), combinando testes de raiz unitária e minimização de critérios de informação para obter o modelo ARIMA(p , d , q) que apresente as melhores combinações de resultados, a partir de um processo iterativo. Além disso, possui diversos argumentos que possibilitam variações no algoritmo.

De maneira geral, o algoritmo para modelagem automática ARIMA apresenta o seguinte comportamento padrão (HYNDMAN; ATHANASOPOULOS, 2018):

- I. Determinação do número de diferenciações necessárias para tornar a série estacionária, a partir de repetidos testes de raiz unitária KPSS;
 - II. Escolha dos valores de p e q pela minimização do critério de informação AICc.
- Para tanto, as seguintes etapas são realizadas:
- a. Monta-se quatro modelos: ARIMA(0, d ,0), ARIMA(2, d ,2), ARIMA(1, d ,0) e ARIMA(0, d ,1); inclui-se uma constante, exceto se $d = 2$; caso $d \leq 1$, o modelo ARIMA(0, d ,0) - sem uma constante, também será considerado;
 - b. O modelo com menor critério de informação – portanto, o mais bem avaliado - é definido como o “modelo atual”;
 - c. Considera-se a seguinte variação do “modelo atual”: variação de p e/ou q em ± 1 ; após todas as etapas descritas, o melhor modelo escolhido, até então, será o novo “modelo atual”;
 - d. Repetição da etapa II.c até que nenhum critério de informação inferior possa ser encontrado.

Como descrito na etapa II.a, por padrão, o algoritmo utiliza aproximações de modelos para acelerar os testes. Para realizar um maior número de combinações de testes, pode-se utilizar argumentos específicos dentro da própria função. Além disso, é possível utilizar outros testes de raiz unitária, como o de Dickey-Fuller Aumentado e o de Phillips-Perron, além de outros critérios de informação, como o Critério de Informação Akaike e o Critério de Informação Bayesiano de Schwartz.

2.2.3.7 Critérios de seleção de modelos

Diversos critérios de seleção de modelos podem ser empregados, cujo objetivo é selecionar o modelo que melhor descreve os dados observados, a fim de realizar previsão. A seguir serão descritos dois desses critérios, o Critério de Informação Akaike (AIC) e o Critério de Informação Bayesiano de Schwartz (BIC). Ambos os critérios visam minimizar a soma de quadrados dos resíduos (SQR), além de priorizarem a inclusão de um menor número de regressores possíveis (GUAJARATI; PORTER, 2011).

2.2.3.7.1 Critério de Informação Akaike (AIC)

O Critério de Informação Akaike (AIC) impõe uma medida corretiva pelo acréscimo de regressores no modelo, expresso na forma logarítmica por:

$$\ln AIC = \frac{2k}{n} + \ln \frac{SQR}{n}$$

Em que k é o número de regressores e n , o de observações. O modelo que apresentar o menor valor de AIC é o escolhido (GUAJARATI; PORTER, 2011).

2.2.3.7.2 Critério de Informação Bayesiano de Schwartz (BIC)

O Critério de Informação Bayesiano de Schwartz (BIC) se assemelha ao AIC, sendo definido, na forma logarítmica, como:

$$\ln BIC = \frac{k}{n} \ln n + \ln \frac{SQR}{n}$$

Sendo $\frac{k}{n} \ln n$ o fator de correção. O BIC impõe medidas corretivas mais severas do que o AIC e, assim como o outro critério, quanto mais baixo o valor do BIC, mais bem avaliado é o modelo (GUAJARATI; PORTER, 2011).

2.2.3.8 Testes de validação do modelo

Uma vez selecionado o modelo a partir dos critérios de informação, deve-se verificar seu ajuste aos dados da série. Para tanto, pode-se analisar os resíduos do modelo a partir do correlograma da FAC, além de utilizar estatísticas, como exemplo a estatística-Q de Ljung e Box (1978), que verifica se os resíduos do modelo são ruídos brancos. Caso não sejam, deve-se retornar à etapa inicial da modelagem e realizar novas estimativas de modelos (GUAJARATI; PORTER, 2011).

A estatística-Q Ljung-Box (1978) testa se observações ao longo do tempo são aleatórias e independentes, seguindo a distribuição de qui-quadrado com m graus de liberdade. Tem como hipótese nula que as autocorrelações são iguais a zero até k defasagens – ou seja, que os valores são aleatórios e independentes até um certo valor k de defasagem (GUAJARATI; PORTER, 2011). Nesse caso, a estatística é utilizada para a validação do modelo ao verificar se os resíduos são independentes – ruído branco.

3. MATERIAIS E MÉTODOS

As análises foram realizadas no ambiente computacional R (R CORE TEAM, 2022), com auxílio de pacotes contendo funcionalidades específicas para cada etapa. Todos os bancos de dados utilizados, além dos códigos, estão disponíveis em: <https://github.com/gustavojoy/ModeloARIMA>. No Apêndice A, encontra-se o código confeccionado para as análises do trabalho.

3.1 Coleta de dados

Os dados do Índice Nacional de Preços ao Consumidor Amplo (IPCA) foram coletados com o pacote *sidrar* (SIQUEIRA, 2022). Este pacote estabelece conexão direta com os bancos de dados presentes no Sistema IBGE de Recuperação Automática (SIDRA) (IBGE, 2022b), o que possibilita a automatização e flexibilização no acesso aos dados diretamente do R.

Para as análises, utilizou-se a variação mensal da inflação dos subitens presentes no subgrupo alimentação no domicílio – pertencente ao grupo alimentação e bebidas - entre julho de 2006 e setembro de 2022. A Tabela 6 apresenta as descrições completas dos dados obtidos do SIDRA.

Tabela 6 - Descrições das tabelas utilizadas para coleta dos dados de inflação.

Tabela	Descrição	Variáveis utilizadas
2938	IPCA - Variação mensal, acumulada no ano e peso mensal, para o índice geral, grupos, subgrupos, itens e subitens de produtos e serviços (de julho/2006 até dezembro/2011)	Variação mensal dos subitens de alimentação no domicílio à nível nacional, entre julho de 2006 e dezembro de 2011
1419	IPCA - Variação mensal, acumulada no ano, acumulada em 12 meses e peso mensal, para o índice geral, grupos, subgrupos, itens e subitens de produtos e serviços (de janeiro/2012 até dezembro/2019)	Variação mensal dos subitens de alimentação no domicílio à nível nacional, entre janeiro de 2012 e dezembro de 2019
7060	IPCA - Variação mensal, acumulada no ano, acumulada em 12 meses e peso mensal, para o índice geral, grupos, subgrupos, itens e subitens de produtos e serviços (a partir de janeiro/2020)	Variação mensal dos subitens de alimentação no domicílio à nível nacional, entre janeiro de 2020 e setembro de 2022

Fonte: Elaborado pelo autor, baseado em IBGE (2022b).

As estruturas de ponderação foram coletadas do IBGE, a partir das publicações divulgadas para o Índice Nacional de Preços ao Consumidor Amplo (IBGE, 2022c). Para a POF 2002-2003, 2008-2009 e 2017-2018, utilizaram-se os pesos calculados em abril de 2006, janeiro de 2009 e dezembro de 2019, respectivamente.

3.2 Processamento e manipulação de dados

Para o processamento e manipulação de dados, utilizou-se os seguintes pacotes: *lubridate* (SPINU; GROLEMUND; WICKHAM, 2021) e *stringr* (WICKHAM, 2019) – para a manipulação de dados relacionados a datas e do tipo caractere, respectivamente; *dplyr* (WICKHAM et al., 2022) e *tidyr* (WICKHAM, 2021) – para selecionar, filtrar, transformar e agrupar os dados; *ggplot2* (WICKHAM et al., 2016) para a confecção de gráficos; além dos pacotes *knitr* (XIE, 2021) e *kableExtra* (ZHU, 2021) para confecção de tabelas.

Os subitens do subgrupo alimentação no domicílio foram agrupados em cadeias agropecuárias, baseado na classificação proposta por Baccarin, Bueno e da Silva (2015), resultando em 25 cadeias, sendo elas: arroz, avicultura de corte, avicultura de postura, banana, batata inglesa, bovinocultura de corte, bovinocultura de leite, cafeicultura, cana-de-açúcar, cacau, cebola, feijão, indefinido, laranja, maçã, mandioca, milho, outras frutas, pescados, preparado vegetal, soja, suinocultura, tomate, trigo e vegetais in natura.

Dessas 25 cadeias, selecionaram-se seis – arroz, bovinocultura de corte, bovinocultura de leite, feijão, mandioca e tomate - categorizadas de acordo com a origem agrícola: pecuária, grãos e hortícolas. A Tabela 7 ilustra a categorização das cadeias selecionadas.

Tabela 7 - Categorização das cadeias agropecuárias, de acordo com a origem agrícola.

Categorias	Cadeias
Pecuária	Bovinocultura de corte e
	Bovinocultura de leite
Grãos	Arroz e Feijão
Hortícolas	Mandioca e Tomate

Fonte: Elaborado pelo autor.

O cálculo da variação mensal da inflação em cada cadeia foi feito de forma ponderada, baseando nas estruturas de ponderação da POF. Entre 2007 e 2011, a variação da inflação derivou da ponderação presente na POF 2002/03; de 2012 a 2019, na POF 2008/09; e de 2020 em diante, na POF 2017/18.

3.3 Modelagem ARIMA

Para analisar o comportamento das séries, foram construídos gráficos da variação mensal da inflação ao longo do período em análise para cada uma das sete cadeias agropecuárias.

Para verificar a estacionariedade das séries de inflação, foi utilizado o teste de raiz unitária Dickey-Fuller Aumentado (DFA) (DICKEY; FULLER, 1979), utilizando a função *ur.df()* do pacote *urca* (PFAFF, 2008).

Os correlogramas da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) foram confeccionados para explorar as possíveis ordens do modelo ARIMA (p, d, q). Para a confecção dos correlogramas FAC, foi utilizada a função *ggAcf()* e para a FACP, a *ggPacf()*, pertencentes ao pacote *forecast* (HYNDMAN et al., 2021).

A partir da função *auto.arima()*, do pacote *forecast* (HYNDMAN et al., 2021), foram definidos possíveis modelos ARIMA(p, d, q), a fim de testar e definir o melhor modelo para cada cadeia agropecuária. Para tanto, foram utilizados como critério de seleção o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano de Schwarz (BIC), executados pela própria função anteriormente citada.

Uma vez selecionado o melhor modelo para cada cadeia, analisou-se o comportamento dos resíduos dos modelos a partir do correlograma FAC, além da estatística-Q Ljung-Box (LJUNG; BOX, 1978), a partir da função *Box.test()* do pacote *stats* (R CORE TEAM, 2022).

Após a definição dos melhores modelos, foram confeccionados gráficos das séries com os dados observados e os estimados. Posteriormente, foram realizadas as previsões de inflação das cadeias agropecuárias para 6 meses – entre outubro de

2022 e março de 2023 – com a função *forecast()* do pacote *forecast* (HYNDMAN et al., 2021).

4. RESULTADOS E DISCUSSÃO

Nesta seção, serão apresentados e discutidos os resultados das análises. Cada seção está dividida por categoria – Pecuária, Grãos e Hortícolas – as quais abrangem as respectivas cadeias agropecuárias.

4.1 Categoria Pecuária

A Figura 7 ilustra as séries temporais da variação mensal da inflação das cadeias da bovinocultura de corte e bovinocultura de leite, presentes na categoria Pecuária. Ambas as séries, aparentam ser estacionárias, visto que, de maneira geral, apresentam média e variância tendendo à uma constante ao longo do tempo.

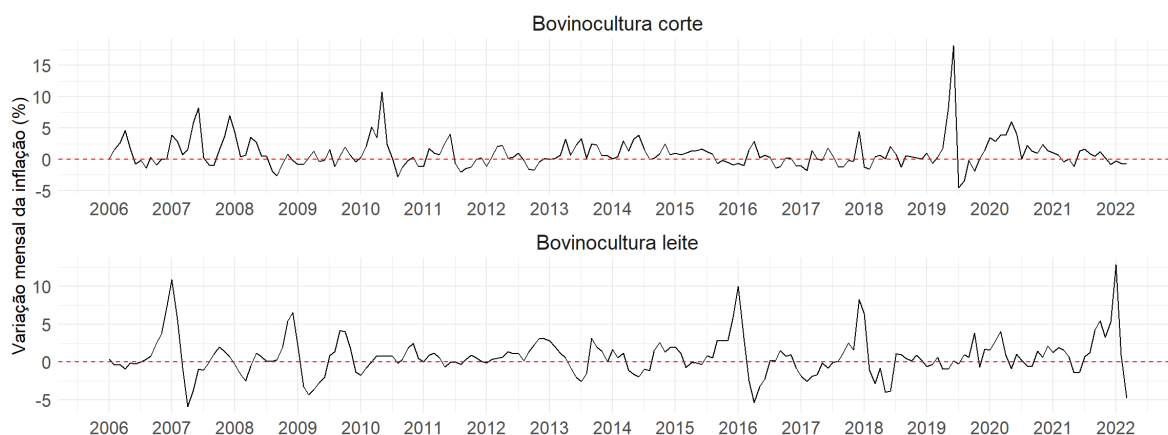


Figura 7 - Séries temporais da categoria Pecuária, cadeias Bovinocultura de corte e Bovinocultura de leite, entre julho de 2006 a setembro de 2022.

Fonte: Elaborado pelo autor, com dados de IBGE (2022b).

O teste Dickey-Fuller Aumentado (DFA) (Tabela 8) forneceu evidências de que nenhuma das cadeias da categoria Pecuária apresentam raiz unitária, rejeitando-se a hipótese nula de presença de raiz unitária, a 1% de nível de significância. Com isso, pode-se confirmar que as séries são estacionárias.

Tabela 8 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Bovinocultura de corte e Bovinocultura de leite.

Cadeia	Estatística τ	Valor crítico (***)	p-valor
Bovinocultura corte	- 8,69 (***)	-3,99	< 0,01 (***)
Bovinocultura leite	- 8,61 (***)	-3,99	< 0,01 (***)

Fonte: Elaborado pelo autor.

Nota: (***) Significativo ao nível de 1,0%. Valores críticos de Dickey e Fuller (1981).

Ao analisar os correlogramas da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP) da cadeia da Bovinocultura de corte (Figura 8), a primeira defasagem da FAC apresentou forte correlação positiva, porém, seguida de uma queda brusca, com valores tendendo a zero, dando indícios de um processo MA(1). Em relação à FACP, novamente, a primeira defasagem apresentou forte correlação positiva, seguida de defasagens tendendo a zero e dentro do intervalo de confiança de 95%, indicando um processo AR(1).

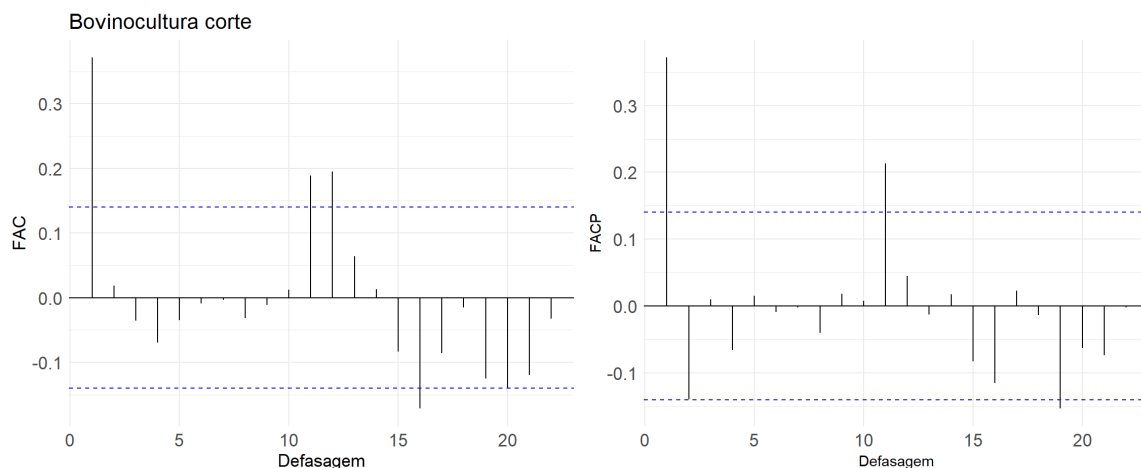


Figura 8 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia de Bovinocultura de corte.

Fonte: Elaborado pelo autor.

Utilizando a função *auto.arima()*, o modelo ARIMA(0, 0, 1) obteve o melhor desempenho para a cadeia da Bovinocultura de corte, apresentando valor de BIC de 872,13 e AIC de 862,31. A Tabela 9 ilustra os coeficientes e erro padrão do modelo.

Tabela 9 - Coeficientes e erro padrão do modelo MA(1) da cadeia Bovinocultura de corte.

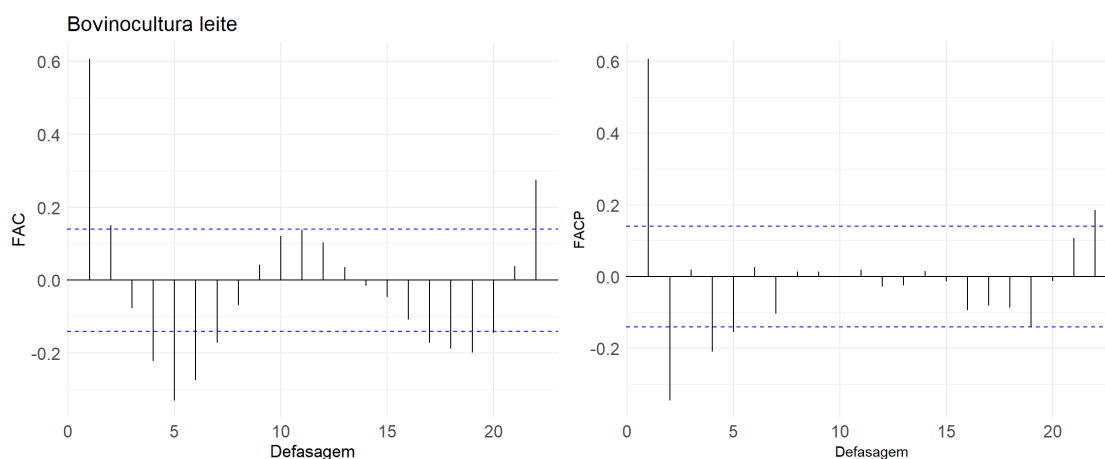
Processos	Coeficientes	Erro padrão
MA(1)	0,4181	0,0645
Média	0,9081	0,2204

Fonte: Elaborada pelo autor.

Com isso, obteve-se um modelo com apenas o processo de médias móveis de primeira ordem – MA(1), representada pela seguinte equação:

$$y_t = 0,9081 + 0,4181 \varepsilon_{t-1} + \varepsilon_t$$

Para a Bovinocultura de leite (Figura 9), a FAC apresentou forte correlação positiva na primeira defasagem, com queda nas duas subsequentes, voltando a apresentar forte correlação – agora negativa, a partir da quarta defasagem, retornando a valores tendendo a zero a partir da oitava defasagem; com esse comportamento, a ordem q do processo MA(q) não é clara. Em relação à FACP, há forte correlação na primeira, segunda, quarta e quinta defasagem, indicando um processo AR(p) com possível segunda ordem.

**Figura 9** - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia de Bovinocultura de leite.

Fonte: Elaborado pelo autor.

A função *auto.arima()* retornou o modelo ARIMA(5,0,0) como o com melhor desempenho, apresentando valor de BIC de 816,70 e AIC de 793,79. A Tabela 10 ilustra os parâmetros e erro padrão obtidos para a Bovinocultura de leite.

Tabela 10 - Coeficientes e erro padrão do modelo AR(5) da cadeia Bovinocultura de leite.

Processos	Coeficientes	Erro padrão
AR(1)	0,7997	0,0699
AR(2)	-0,3760	0,0932
AR(3)	0,0220	0,109
AR(4)	0,0222	0,1071
AR(5)	-0,2330	0,0798
Média	0,6829	0,1679

Fonte: Elaborada pelo autor.

Com isso, a cadeia da Bovinocultura de leite apresentou um modelo autorregressivo de 5ª ordem – AR(5), com a seguinte equação:

$$y_t = 0,6829 + 0,7997 y_{t-1} - 0,3760 y_{t-2} + 0,0220 y_{t-3} + 0,0222 y_{t-4} - 0,2330 y_{t-5} + \varepsilon_t$$

A estatística-Q de Ljung-Box (Tabela 11) nos retorna que os valores dos resíduos não são significantes, não rejeitando a hipótese nula de resíduos independentes. Dessa maneira, tanto os resíduos do modelo da Bovinocultura de corte, quanto os da Bovinocultura de leite, são ruído branco.

Tabela 11 - Estatística-Q de Ljung-Box para a categoria Pecuária, cadeias Bovinocultura de corte e Bovinocultura de leite.

Cadeia	p-valor
Bovinocultura corte	0,87 ^(.)
Bovinocultura leite	0,84 ^(.)

Fonte: Elaborada pelo autor.

Nota: (.) Não significativo a nível de 10% de significância.

O correlograma da FAC dos resíduos (Figura 10 e Figura 11) reitera o que foi constatado pela estatística. No correlograma de ambas as cadeias, os resíduos não apresentam correlação serial, visto que a maioria de suas defasagens estão dentro ou senão, muito próximas, do intervalo de confiança de 95%.

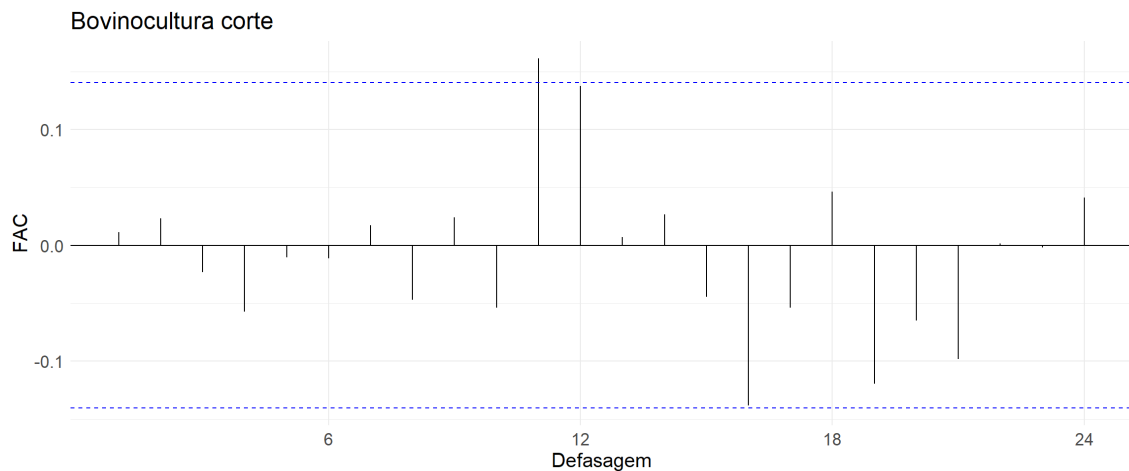


Figura 10 - Correlograma da FAC dos resíduos da Cadeia da Bovinocultura de corte.

Fonte: Elaborada pelo autor.

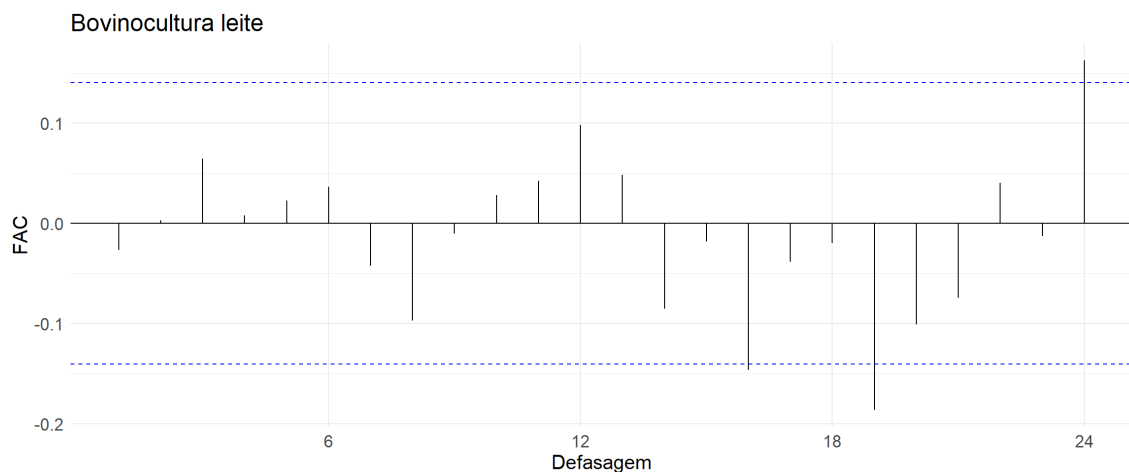


Figura 11 - Correlograma da FAC dos resíduos da Cadeia da Bovinocultura de leite.

Fonte: Elaborada pelo autor.

A Figura 12 ilustra os valores do modelo e os valores observados para a cadeia da Bovinocultura de corte, denotando um comportamento do modelo próximo aos dados reais. Contudo, deve-se considerar que o modelo não previu com grande

exatidão algumas volatilidades dos dados reais, indicando que essa volatilidade pode ser explicada por outros fatores e não pelos valores passados da série.

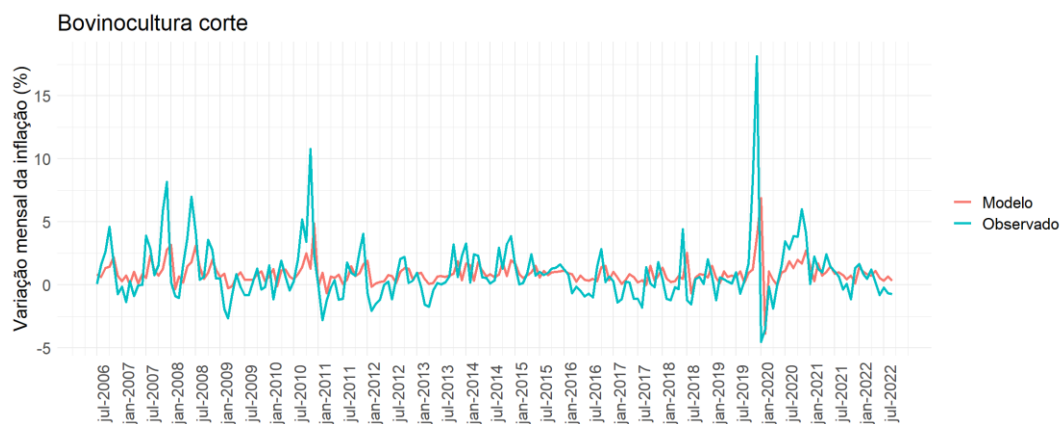


Figura 12 – Valores ajustados pelo modelo MA(1) da cadeia da Bovinocultura de corte, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

Já para a cadeia da Bovinocultura de leite, observa-se que o modelo acompanhou melhor as grandes volatilidades dos dados observados (Figura 13), quando comparado ao modelo da Bovinocultura de corte.



Figura 13 – Valores ajustados pelo modelo AR(5) da cadeia da Bovinocultura de leite, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

A Tabela 12 ilustra as previsões dos valores de inflação realizadas pelo modelo da Bovinocultura de corte, entre outubro de 2022 e março de 2023, de acordo com os intervalos de confiança de 80% e de 95%, além da média dos valores previstos de inflação, utilizados para a confecção do gráfico presente na Figura 14.

Tabela 12 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95% para a cadeia da Bovinocultura de corte, outubro de 2022 a março de 2023.

Período	Valores previstos				
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	Média
Out 2022	-2,34	3,26	-3,82	4,74	0,46
Nov 2022	-2,13	3,94	-3,73	5,55	0,91
Dez 2022	-2,13	3,94	-3,73	5,55	0,91
Jan 2023	-2,13	3,94	-3,73	5,55	0,91
Fev 2023	-2,13	3,94	-3,73	5,55	0,91
Mar 2023	-2,13	3,94	-3,73	5,55	0,91

Fonte: Elaborada pelo autor.

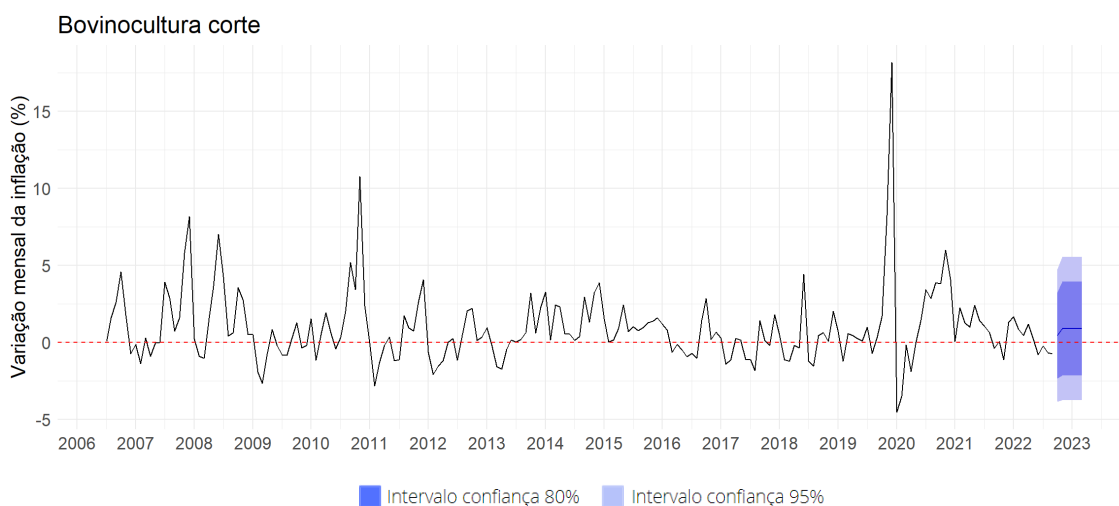


Figura 14 - Previsão da variação da inflação da cadeia da Bovinocultura de corte, a partir do modelo MA(1), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Para a cadeia da Bovinocultura de corte, a tendência média para os próximos seis meses da análise é de pequeno crescimento médio da inflação, com posterior estabilização em 0,91% a partir do mês de novembro, até o mês de março de 2023.

Tabela 13 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia da Bovinocultura de leite, outubro de 2022 a março de 2023.

Período	Valores previstos				
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	Média
Out 2022	-6,34	-1,70	-7,56	-0,47	-4,02
Nov 2022	-4,79	1,14	-6,37	2,72	-1,82
Dez 2022	-5,54	0,52	-7,14	2,13	-2,51
Jan 2023	-4,23	1,84	-5,84	3,45	-1,20
Fev 2023	-1,55	4,55	-3,16	6,16	1,50
Mar 2023	-0,10	6,12	-1,75	7,77	3,01

Fonte: Elaborada pelo autor.

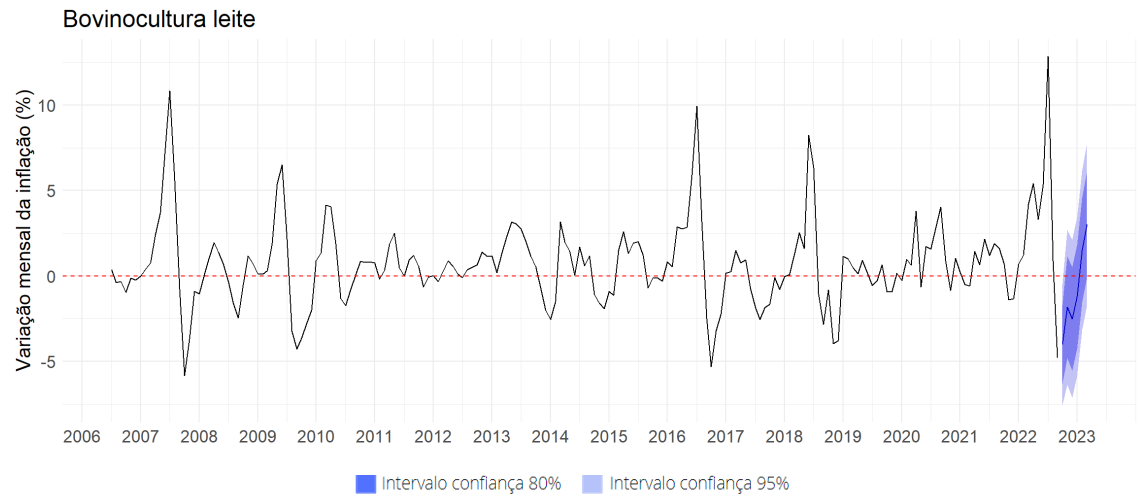


Figura 15 - Previsão da variação da inflação da cadeia da Bovinocultura de leite, a partir do modelo AR(5), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Para a cadeia da Bovinocultura de leite, a Tabela 13 e a Figura 15, trazem os resultados da previsão. A tendência média para os próximos seis meses da análise é de crescimento da variação, partindo do valor de -4,02% no primeiro mês de previsão e chegando a 3,01% em março de 2023.

4.2 Categoria Grãos

A Figura 16 ilustra as séries temporais da variação mensal da inflação das cadeias do Arroz e do Feijão, que pertencem a categoria Grãos. Ambas as séries, aparentam ser estacionárias, visto que, de maneira geral, apresentam média e variância tendendo à uma constante ao longo do tempo. Ademais, na série do Feijão, verifica-se uma grande volatilidade dos preços, enquanto na série do arroz, observa-se variações de inflação mensal atípicas nos anos de 2008, 2012 e 2020, com elevado crescimento do índice.

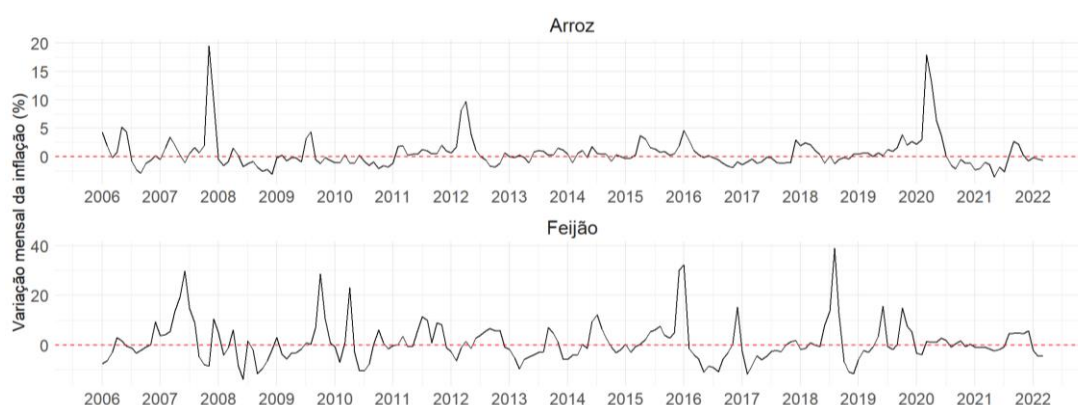


Figura 16 - Séries temporais da categoria Grãos, cadeias Arroz e Feijão, entre julho de 2006 a setembro de 2022.

Fonte: Elaborado pelo autor, com dados de IBGE (2022b).

O teste Dickey-Fuller Aumentado (DFA) (Tabela 14) forneceu evidências de que nenhuma das cadeias da categoria Grãos apresentam raiz unitária, rejeitando-se a hipótese nula de presença de raiz unitária, a 1% de nível de significância. Com isso, pode-se confirmar que as séries são estacionárias.

Tabela 14 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Arroz e Feijão.

Cadeia	Estatística τ	Valor crítico (***)	p-valor
Arroz	- 7,31 (***)	-3,99	< 0,01 (***)
Feijão	- 8,21 (***)	-3,99	< 0,01 (***)

Fonte: Elaborado pelo autor.

Nota: (***) Significativo ao nível de 1,0%. Valores críticos de Dickey e Fuller (1981).

Na cadeia do Arroz, tanto o correlograma FAC, quanto o FACP (Figura 17) apresentam forte correlação nas duas primeiras defasagens, indicando um possível processo de médias móveis MA(2) ou um autorregressivo AR(2).

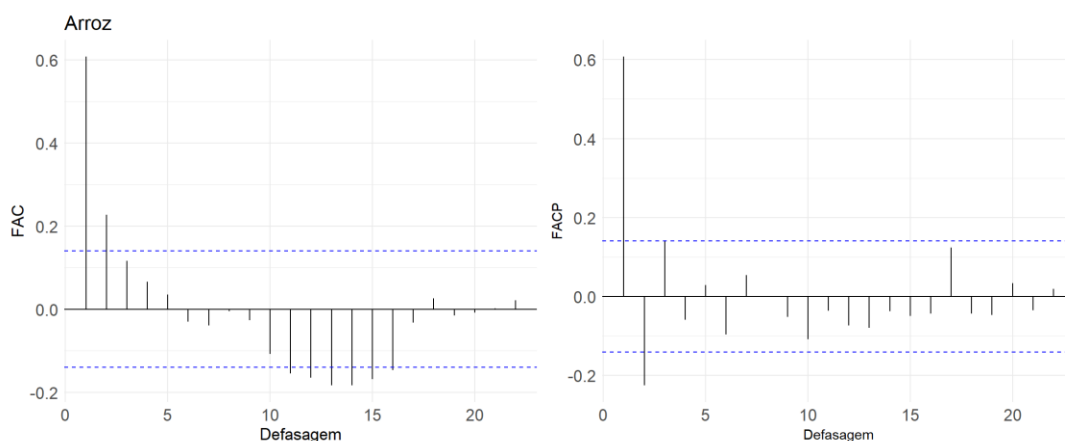


Figura 17 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Arroz.

Fonte: Elaborado pelo autor.

A partir da função *auto.arima()*, obteve-se o modelo ARIMA(1,0,1) com o melhor desempenho para a cadeia do Arroz, apresentando valor de BIC de 876,94 e AIC de 863,84. A Tabela 15 ilustra os coeficientes e erro padrão do modelo.

Tabela 15 - Coeficientes e erro padrão do modelo ARMA(1,1) da cadeia do Arroz.

Processos	Coeficientes	Erro padrão
AR(1)	0,3189	0,1050
MA(1)	0,4890	0,1033
Média	0,6559	0,3380

Fonte: Elaborada pelo autor.

Assim, obteve-se um modelo autorregressivo de médias móveis ARMA(1,1), representado pela seguinte equação:

$$y_t = 0,6559 + 0,3189 y_{t-1} + 0,4890 \varepsilon_{t-1} + \varepsilon_t$$

A FAC da cadeia do Feijão indica um processo MA(1) devido a correlação forte na primeira defasagem ou um processo MA(2), de acordo com o FACP. Ambos os correlogramas estão ilustrados na Figura 18.

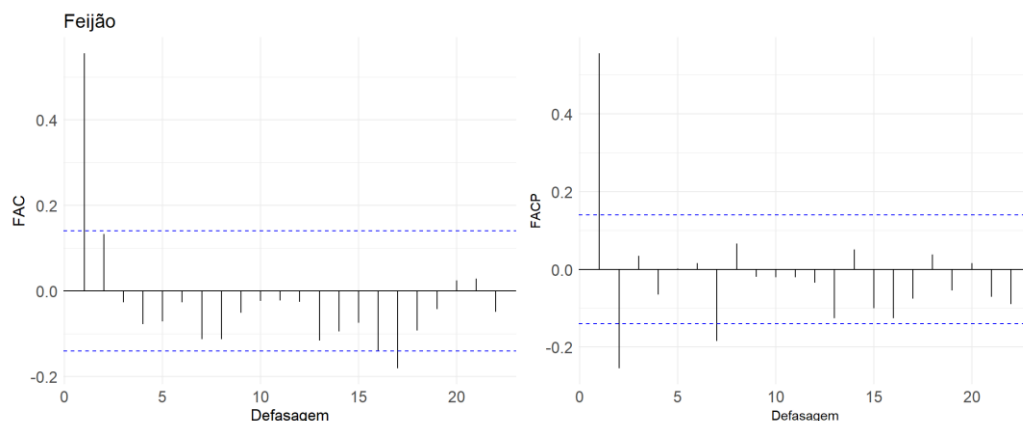


Figura 18 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Feijão.

Fonte: Elaborado pelo autor.

A partir da função *auto.arima()*, obteve-se o modelo ARIMA(0,0,2) com o melhor desempenho, apresentando valor de BIC de 1291,75 e AIC de 1281,93. A Tabela 16 ilustra os coeficientes e erro padrão do modelo.

Tabela 16 - Coeficientes e erro padrão do modelo MA(2) da cadeia do Feijão.

Processos	Coeficientes	Erro padrão
MA(1)	0,7186	0,0711
MA(2)	0,1979	0,0626

Fonte: Elaborada pelo autor.

Com isso, obteve-se um modelo com processo de médias móveis de segunda ordem – MA(2), representada pela seguinte equação:

$$y_t = 0,7186 \varepsilon_{t-1} + 0,1979 \varepsilon_{t-2} + \varepsilon_t$$

A estatística-Q Ljung-Box nos retorna que os valores dos resíduos não são significantes, não rejeitando a hipótese nula de resíduos independentes (Tabela 17). Dessa maneira, tanto os resíduos do modelo do Arroz, quanto os do Feijão são ruído branco.

Tabela 17 - Estatística-Q de Ljung-Box para a categoria Grãos, cadeias Arroz e Feijão.

Cadeia	p-valor
Arroz	0,87 ^(.)
Feijão	0,96 ^(.)

Fonte: Elaborada pelo autor.

Nota: (.) Não significativo a nível de 10% de significância.

O correlograma da FAC dos resíduos (Figura 19 e Figura 20) reitera o que foi constatado pela estatística. No correlograma de ambas as cadeias, os resíduos não apresentam correlação serial, visto que suas defasagens estão dentro do intervalo de confiança de 95%.

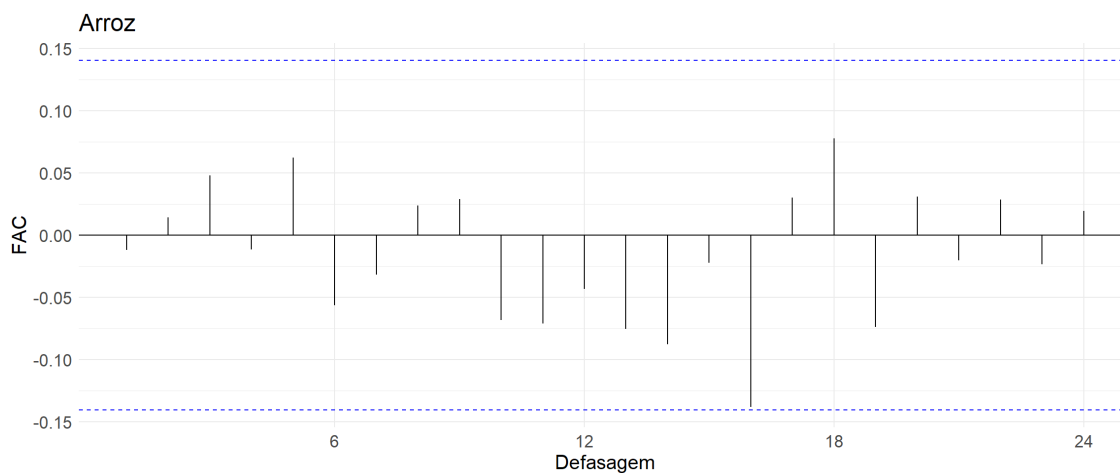


Figura 19 - Correlograma da FAC dos resíduos da Cadeia do Arroz.

Fonte: Elaborada pelo autor.

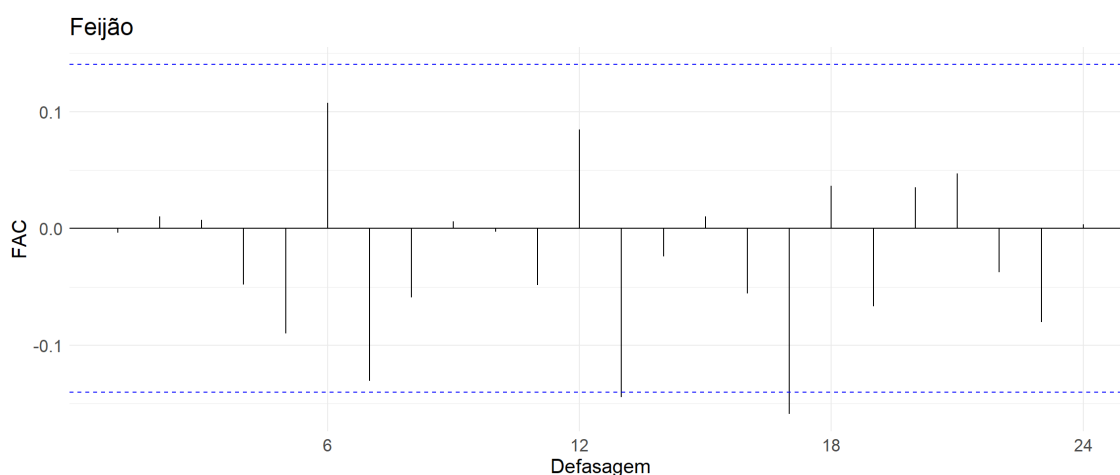


Figura 20 - Correlograma da FAC dos resíduos da Cadeia do Feijão.

Fonte: Elaborada pelo autor.

A Figura 21 ilustra os valores do modelo e os valores observados para a cadeia do Arroz, denotando um comportamento do modelo próximo aos dados reais.

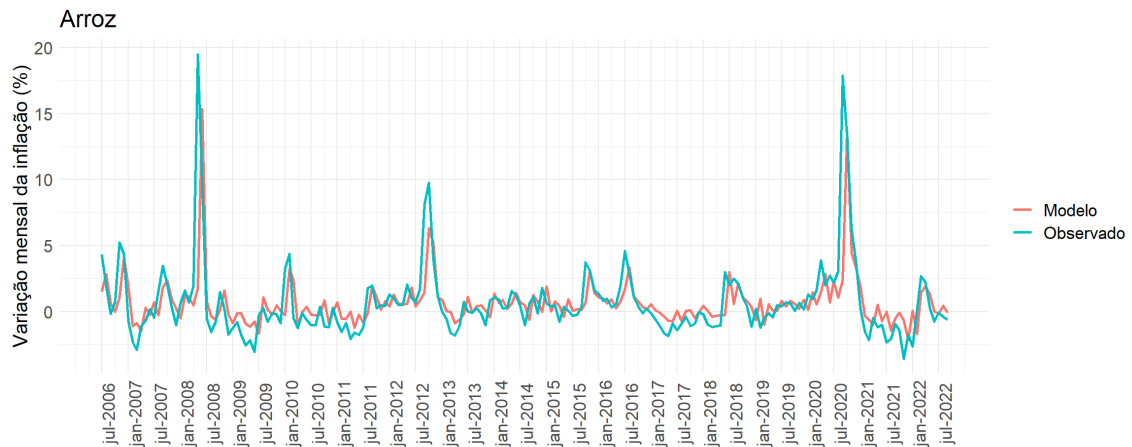


Figura 21 - Ajuste dos valores do modelo ARMA(1,1), da cadeia do Arroz, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

Já para a cadeia do Feijão, observa-se que o modelo apresentou comportamento próximo aos dados reais (Figura 22). Entretanto, previu com menor exatidão algumas volatilidades, quando comparado ao modelo do Arroz.

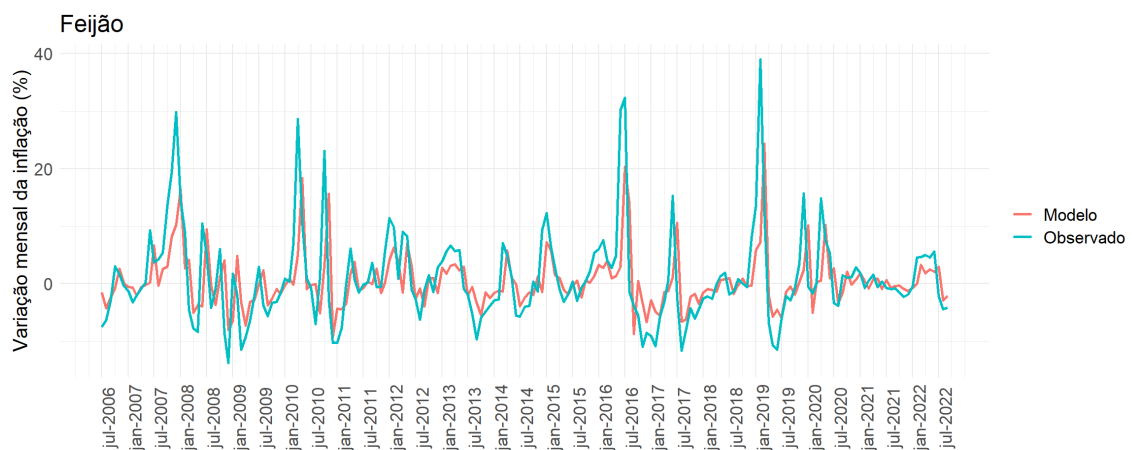


Figura 22 - Ajuste dos valores do modelo MA(2), da cadeia do Feijão, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

A Tabela 18 ilustra as previsões dos valores de inflação realizadas pelo modelo do Arroz, entre outubro de 2022 e março de 2023, de acordo com os intervalos de confiança de 80% e de 95%, além da média dos valores previstos, utilizados para a confecção do gráfico presente na Figura 23.

Tabela 18 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Arroz, outubro de 2022 a março de 2023.

Período	Valores previstos				
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	Média
Out 2022	-2,81	2,79	-4,29	4,27	-0,01
Nov 2022	-3,16	4,04	-5,06	5,95	0,44
Dez 2022	-3,08	4,26	-5,03	6,20	0,59
Jan 2023	-3,04	4,31	-4,99	6,26	0,63
Fev 2023	-3,03	4,33	-4,98	6,28	0,65
Mar 2023	-3,03	4,33	-4,97	6,28	0,65

Fonte: Elaborada pelo autor.

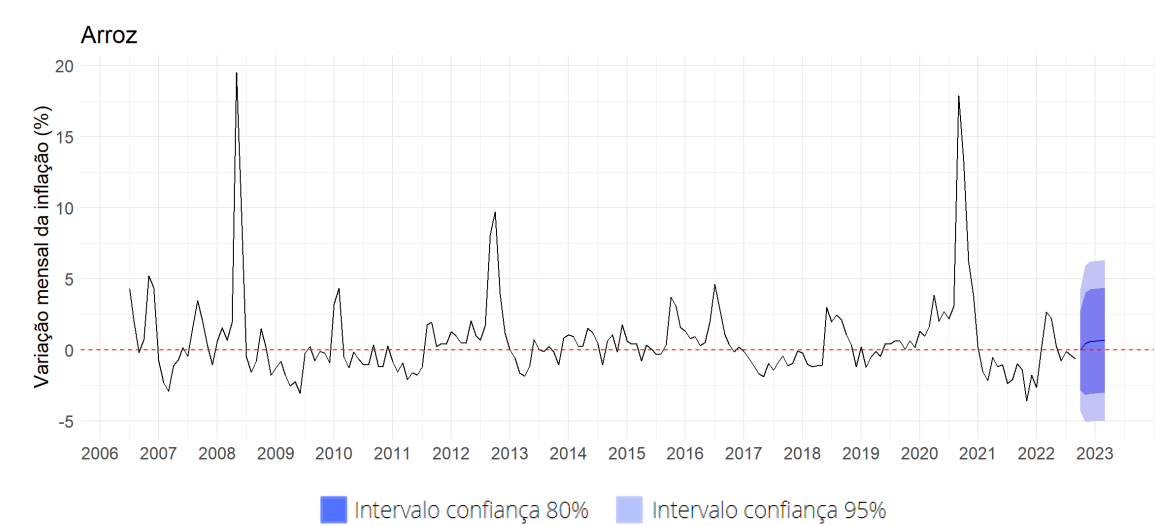


Figura 23 - Previsão da variação da inflação da cadeia do Arroz, a partir do modelo ARMA(1,1), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Para a cadeia do Arroz, a tendência média para os próximos seis meses de análise é de pequeno crescimento médio da inflação, com posterior estabilização em valores próximos de 0,65% a partir do mês de janeiro de 2023.

Tabela 19 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Feijão, outubro de 2022 a março de 2023.

Período	Valores previstos				
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	Média
Out 2022	-10,01	6,39	-14,35	10,74	-1,81
Nov 2022	-10,52	9,68	-15,87	15,03	-0,42
Dez 2022	-10,23	10,23	-15,65	15,65	0,00
Jan 2023	-10,23	10,23	-15,65	15,65	0,00
Fev 2023	-10,23	10,23	-15,65	15,65	0,00
Mar 2023	-10,23	10,23	-15,65	15,65	0,00

Fonte: Elaborada pelo autor.

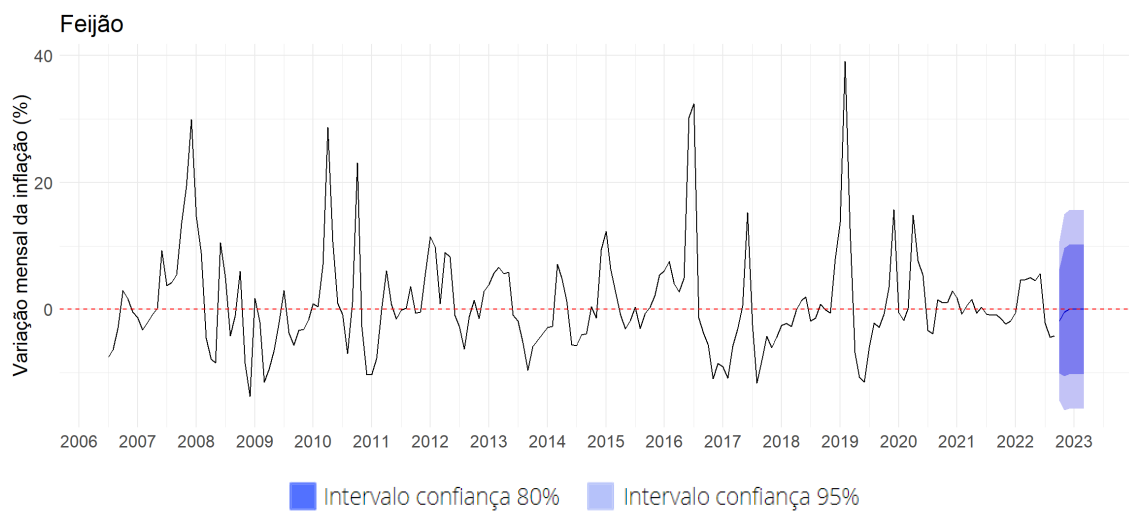


Figura 24 - Previsão da variação da inflação da cadeia do Feijão, a partir do modelo MA(2), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Por fim, a Figura 24 e a Tabela 19 mostram os resultados das previsões para a cadeia do Feijão. A previsão média é de estabilização da inflação em 0% de variação a partir de dezembro de 2022.

4.3 Categoria Hortícolas

A Figura 25 ilustra as séries temporais da variação mensal da inflação das cadeias da Mandioca e do Tomate, que pertencem a categoria Hortícolas. Ambas as séries aparentam ser estacionárias, visto que, de maneira geral, apresentam média e variância tendendo à uma constante ao longo do tempo.

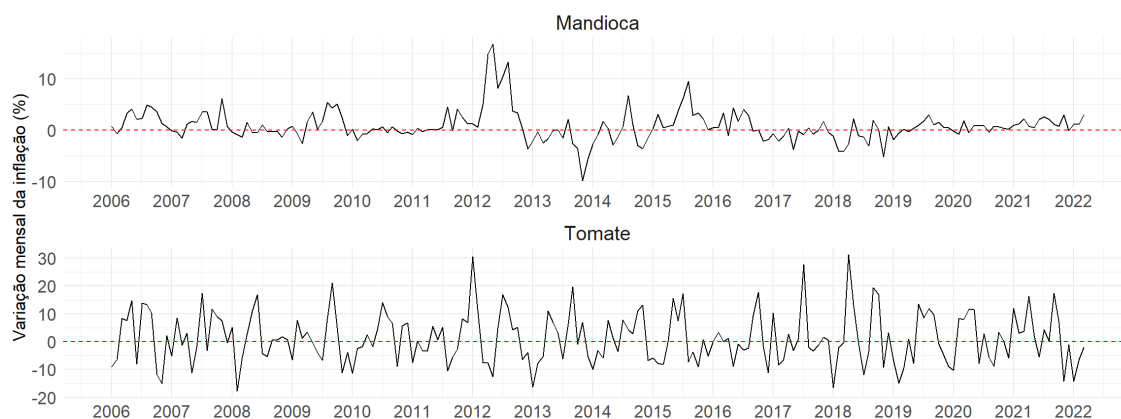


Figura 25 - Séries temporais da categoria Hortícolas, cadeias Mandioca e Tomate, entre julho de 2006 a setembro de 2022.

Fonte: Elaborado pelo autor, com dados de IBGE (2022b).

O teste Dickey-Fuller Aumentado (DFA) (Tabela 20) forneceu evidências de que nenhuma das cadeias da categoria Hortícolas apresenta raiz unitária, rejeitando-se a hipótese nula de presença de raiz unitária, a 1% de nível de significância. Com isso, pode-se confirmar que as séries são estacionárias.

Tabela 20 - Teste de raiz unitária Dickey-Fuller Aumentado para as cadeias Mandioca e Tomate.

Cadeia	Estatística τ	Valor crítico (***)	p-valor
Mandioca	- 5,37 (***)	-3,99	< 0,01 (***)
Tomate	- 10,34 (***)	-3,99	< 0,01 (***)

Fonte: Elaborado pelo autor.

Nota: (***) Significativo ao nível de 1,0%. Valores críticos de Dickey e Fuller (1981).

Na cadeia da Mandioca, as primeiras defasagens da FAC aparentam ter um comportamento de declinação exponencial; já a FACP, tem forte correlação na primeira defasagem. Portanto, aparenta ser um modelo AR(1) (Figura 26).

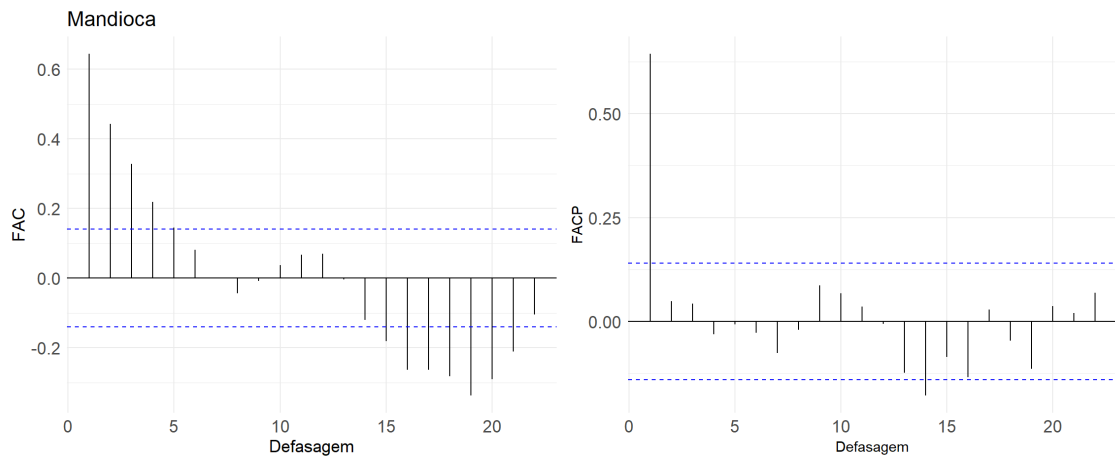


Figura 26 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia da Mandioca.
Fonte: Elaborado pelo autor.

A partir da função *auto.arima()*, obteve-se o modelo ARIMA(1,0,0) com o melhor desempenho para a cadeia de Mandioca, apresentando valor de BIC de 899,87 e AIC de 890,05. A Tabela 21 ilustra os coeficientes e erro padrão do modelo.

Tabela 21 - Coeficientes e erro padrão do modelo AR(1) da cadeia da Mandioca.

Processos	Coeficientes	Erro padrão
AR(1)	0,6430	0,0544
Média	0,8767	0,4635

Fonte: Elaborada pelo autor.

Com isso, a cadeia da Mandioca apresentou um modelo autorregressivo de 1ª ordem – AR(1), com a seguinte equação:

$$y_t = 0,8767 + 0,6430 y_{t-1} + \varepsilon_t$$

A FAC da cadeia do Tomate apresenta forte correlação na primeira, terceira e quarta defasagem. Em relação ao PACF, apresenta forte correlação nas defasagens 1, 3, 4, 7 e 12. A partir dos correlogramas, não fica claro o tipo de modelo, bem como suas ordens (Figura 27).

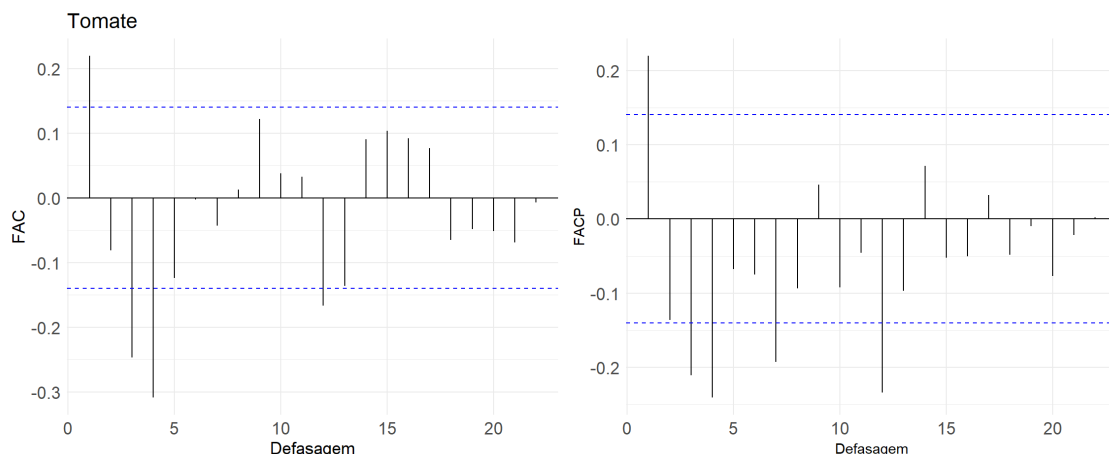


Figura 27 - Correlogramas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da cadeia do Tomate.

Fonte: Elaborado pelo autor.

Para a cadeia do Tomate, a função *auto.arima()* retornou como melhor modelo o ARIMA(3,0,1), apresentando valor de BIC de 1397,04 e AIC de 1377,4. A Tabela 22 ilustra os coeficientes e erro padrão do modelo.

Tabela 22 - Coeficientes e erro padrão do modelo ARMA(3,1) da cadeia do Tomate.

Processos	Coeficientes	Erro padrão
AR(1)	0,9937	0,0857
AR(2)	-0,2725	0,1001
AR(3)	-0,1184	0,0814
MA(1)	-0,9382	0,0552
Média	1,3341	0,1014

Fonte: Elaborada pelo autor.

Assim, a cadeia do Tomate apresentou um modelo autorregressivo de médias móveis ARMA(3,1), representado pela seguinte equação:

$$y_t = 1,3341 + 0,9937 y_{t-1} - 0,2725 y_{t-2} - 0,1184 y_{t-3} - 0,9382 \varepsilon_{t-1} + \varepsilon_t$$

A estatística-Q de Ljung-Box retorna que os valores dos resíduos não são significantes, portanto, não se rejeita a hipótese nula de resíduos independentes (Tabela 23). Dessa maneira, tanto os resíduos do modelo da Mandioca, quanto os do Tomate, são ruído branco.

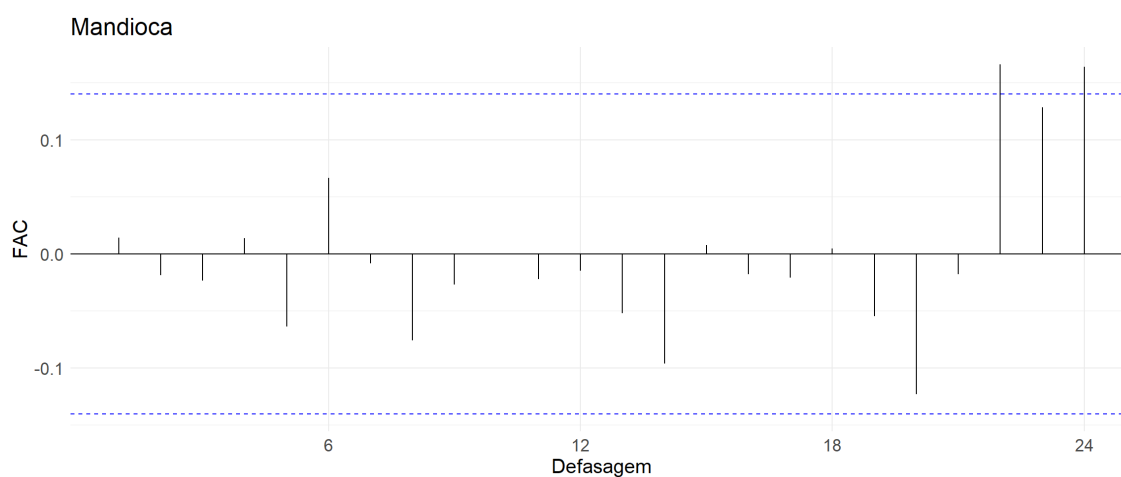
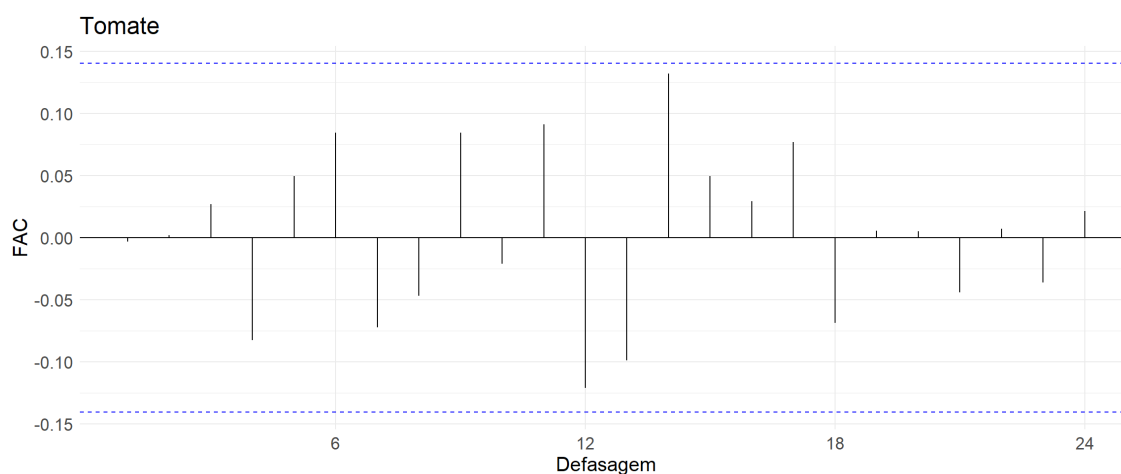
Tabela 23 - Estatística-Q de Ljung-Box para a categoria Hortícolas, cadeias Mandioca e Tomate.

Cadeia	p-valor
Mandioca	0,71 ^(.)
Tomate	0,96 ^(.)

Fonte: Elaborada pelo autor.

Nota: (.) Não significativo a nível de 10% de significância.

O correlograma da FAC dos resíduos (Figura 28 e Figura 29) reitera o que foi constatado pela estatística. No correlograma de ambas as cadeias, os resíduos não apresentam correlação serial, visto que suas defasagens estão dentro ou próximas do intervalo de confiança de 95%.

**Figura 28** - Correlograma da FAC dos resíduos da Cadeia da Mandioca.
Fonte: Elaborada pelo autor.**Figura 29** - Correlograma da FAC dos resíduos da Cadeia da Tomate.
Fonte: Elaborada pelo autor.

A Figura 30 ilustra os valores do modelo e os valores observados para a cadeia da Mandioca, denotando um comportamento próximo aos dados reais.

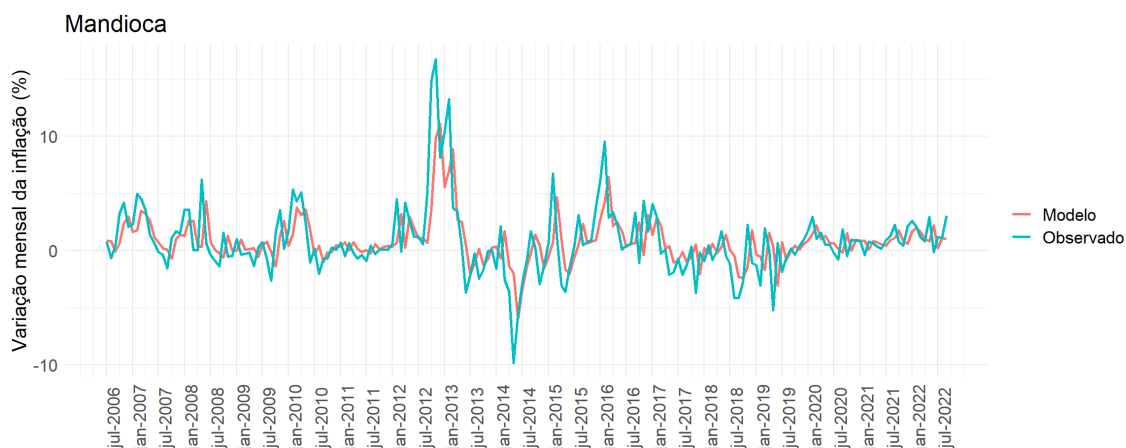


Figura 30 - Ajuste dos valores do modelo AR(1), da cadeia da Mandioca, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

Para a cadeia do Tomate, verifica-se que o modelo não acompanhou as grandes volatilidades da variação mensal da inflação (Figura 31). Isso pode indicar que fatores não ligados aos próprios valores passados da série estejam envolvidos no processo.



Figura 31 - Ajuste dos valores do modelo ARMA(3,1), da cadeia do Tomate, em relação a variação mensal da inflação observada entre julho de 2006 e setembro de 2022.

Fonte: Elaborada pelo autor.

A Tabela 24 ilustra as previsões dos valores de inflação realizadas pelo modelo da Mandioca, entre outubro de 2022 e março de 2023, de acordo com os intervalos de confiança de 80% e de 95%, além da média dos valores previstos, utilizados para a confecção do gráfico presente na Figura 32.

Tabela 24 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos de inflação, de acordo com os intervalos de confiança de 80% e 95%, cadeia da Mandioca, outubro de 2022 a março de 2023.

Período	Valores previstos				
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	Média
Out 2022	-0,72	5,28	-2,31	6,87	2,28
Nov 2022	-1,79	5,35	-3,68	7,24	1,78
Dez 2022	-2,32	5,24	-4,33	7,24	1,46
Jan 2023	-2,61	5,11	-4,66	7,16	1,25
Fev 2023	-2,78	5,01	-4,84	7,08	1,12
Mar 2023	-2,88	4,94	-4,95	7,01	1,03

Fonte: Elaborada pelo autor.

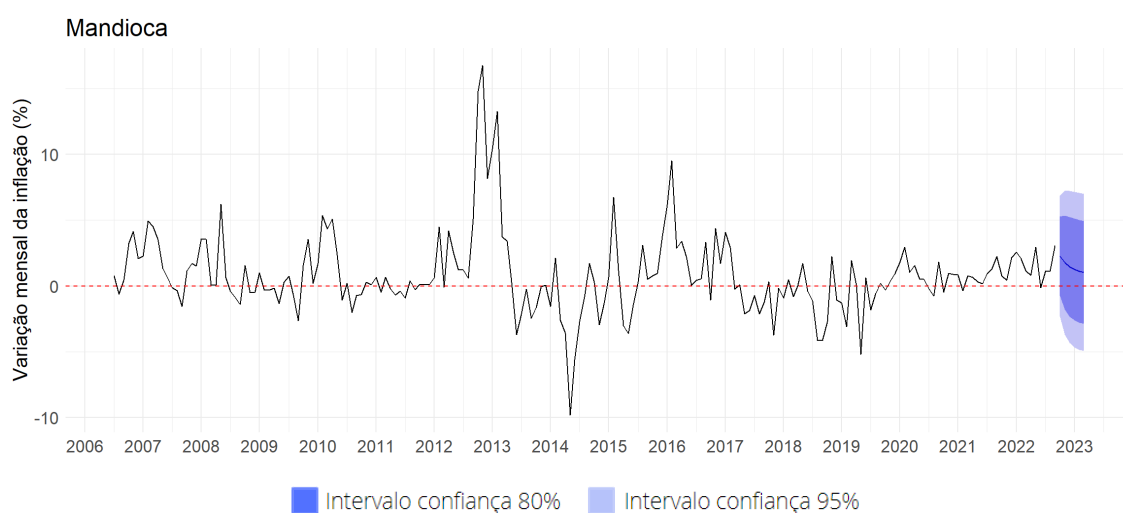


Figura 32 - Previsão da variação da inflação da cadeia da Mandioca, a partir do modelo AR(1), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Para a cadeia da Mandioca, prevê-se uma queda na variação média da inflação, com valor próximo a 1% em março de 2023.

Tabela 25 - Valores médios, inferiores (IC.inf) e superiores (IC.sup) previstos, de acordo com os intervalos de confiança de 80% e 95%, cadeia do Tomate, outubro de 2022 a março de 2023.

Período	Valores previstos				Média
	IC.inf 80%	IC.sup 80%	IC.inf 95%	IC.sup 95%	
Out 2022	-5,12	15,63	-10,61	21,12	5,26
Nov 2022	-3,36	17,42	-8,86	22,92	7,03
Dez 2022	-4,31	16,95	-9,94	22,58	6,32
Jan 2023	-6,96	15,50	-12,91	21,45	4,27
Fev 2023	-9,42	13,86	-15,59	20,02	2,22
Mar 2023	-10,95	12,60	-17,19	18,83	0,82

Fonte: Elaborada pelo autor.

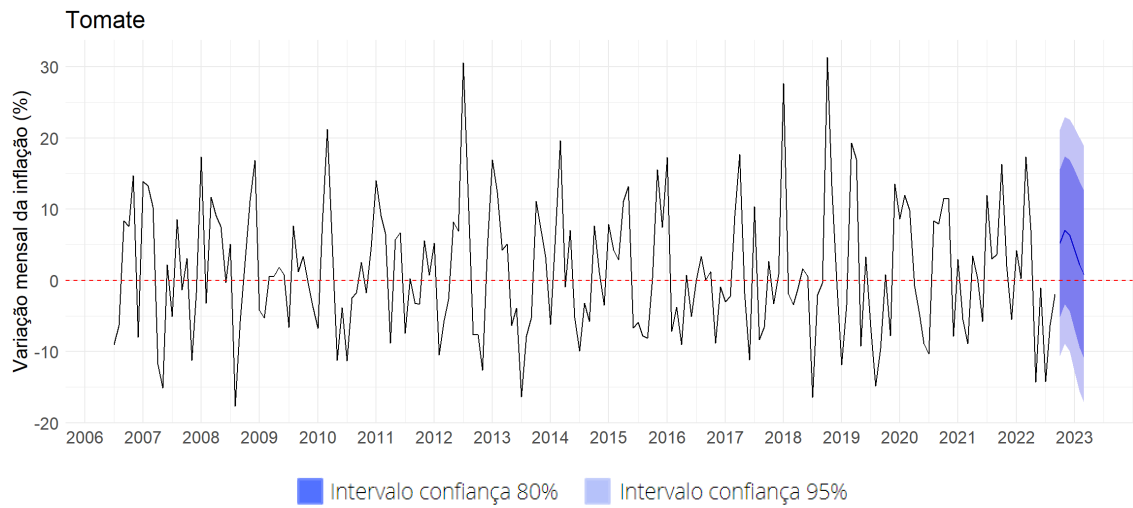


Figura 33 - Previsão da variação da inflação da cadeia do Tomate, a partir do modelo ARMA(3,1), entre outubro de 2022 e março de 2023.

Fonte: Elaborada pelo autor.

Por fim, a Figura 33 e a Tabela 25 mostram os resultados das previsões para a cadeia do Tomate. A previsão média é de queda da variação da inflação, partindo de 5,26% em outubro de 2022, chegando a 0,82% em março de 2023.

5. CONCLUSÕES

As séries de variação mensal da inflação das seis cadeias agropecuárias analisadas, entre julho de 2006 e setembro de 2022, apresentaram comportamento estacionário, com média e variância tendendo à uma constante ao longo do tempo, sendo confirmadas pelo teste de Dickey-Fuller Aumentado. Os melhores modelos ARIMA obtidos para as cadeias agropecuárias foram: Bovinocultura de leite - AR(5); Bovinocultura de corte - MA(1); Arroz - ARMA(1,1); Feijão - MA(2); Mandioca - AR(1); Tomate - ARMA(3,1).

Verificou-se que, para as cadeias da bovinocultura de leite, do arroz e da mandioca, o ajuste dos valores dos modelos apresentaram comportamento próximo aos dados reais. Já para as cadeias da bovinocultura de corte, do feijão e do tomate, o modelo não acompanhou com precisão as acentuadas variações de inflação, indicando que essa volatilidade pode ser explicada por outros fatores e não pelos valores passados da série.

A partir dos modelos, de outubro de 2022 a março de 2023, foi possível prever estabilidade na variação de inflação das cadeias da bovinocultura de corte, do arroz e do feijão; aumento de inflação na cadeia da bovinocultura de leite e queda nas cadeias da mandioca e do tomate.

Dessa maneira, os modelos ARIMA foram bons preditores para grande parte das cadeias agropecuárias estudadas. Contudo, para as cadeias cujo modelo não previu com exatidão as volatilidades dos dados observados, pode-se considerar a aplicação de outros modelos de previsão, como o modelo ARIMA sazonal - ao incluir o fator sazonal aos ajustes e previsão; realizar uma análise multivariada de séries temporais para previsão; ou utilizar outros modelos de previsão, a exemplo dos modelos Vetorial Autorregressivo (VAR) e de Heteroscedasticidade Condicional Autorregressiva (ARCH).

REFERÊNCIAS

- BACCARIN, J.; BUENO, G.; DA SILVA, D.. Cadeias e produtos agropecuários e a inflação brasileira da alimentação no domicílio. **Revista de Política Agrícola**, 2016. Disponível em: <https://seer.sede.embrapa.br/index.php/RPA/article/view/1056/981>. Acesso em: 09 nov. 2022.
- BACCARIN, J.G.; OLIVEIRA, J.A.. Inflação de alimentos no Brasil em período da pandemia da Covid 19, continuidade e mudanças. **Segurança Alimentar e Nutricional**, 2021. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/san/article/view/8661127>. Acesso em: 01 mai. 2022.
- BOX, G.P.E.; JENKINS, G.M.. Time series analysis: forecasting and control. **Rev. Holden Day**, São Francisco: Holden, 1978.
- BRAUN, J.V.. The world food situation: New driving forces and required actions. **International Food Policy Research Institute**, 2007. Disponível em: <http://dx.doi.org/10.2499/0896295303>. Acesso em: 14 set. 2022.
- CEPAL. **Cómo evitar que la crisis del COVID-19 se transforme en una crisis alimentaria**. Acciones urgentes contra el hambre en América Latina y el Caribe. Nações Unidas, 2020. Disponível em: https://www.cepal.org/sites/default/files/publication/files/45702/S2000393_es.pdf. Acesso em: 13 mai. 2022.
- DICKEY, D.A.; FULLER, W.A.. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. **Journal of the American Statistical Association**, 74, p. 427–431, 1979. Disponível em: <http://www.jstor.org/pss/2286348>.
- DICKEY, D.A.; FULLER, W.A.. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. **Econometrica**, 49, 1057–1072, 1981.
- FAO; IFAD; UNICEF; WFP; WHO. **The State of Food Security and Nutrition in the World 2020 (SOFI)**. FAO, IFAD, UNICEF, WFP, WHO, Roma, 2020. Disponível em: <http://www.fao.org/publications/sofi/en/>. Acesso em: 22 jan. 2022.
- FAO; IFAD; UNICEF; WFP; WHO. **The State of Food Security and Nutrition in the World 2022 (SOFI)**. FAO, IFAD, UNICEF, WFP, WHO, Roma, 2022. Disponível em: <https://www.fao.org/publications/sofi/2022/en/>. Acesso em: 25 jan. 2022.
- FAO; IFAD; WFP. **The State of Insecurity in the World**. FAO, Roma, 2015. Disponível em: <http://www.fao.org/3/i4646e/i4646e.pdf>. Acesso em: 02 mai. 2021.

FARINA, E. M. M. Q.; NUNES, R.. A evolução do sistema agroalimentar no Brasil e a redução de preços para o consumidor: os efeitos da atuação dos grandes compradores. **Oficina de la CEPAL en Brasília** (Estudios e Investigaciones), 2002. Disponível em: <https://repositorio.cepal.org/handle/11362/28357>. Acesso em: 13 ago. 2022.

GUAJARATI, D.N.; PORTER, D.C.. **Econometria básica**. 5ª ed. Nova York: AMGH, 2011.

HYNDMAN, R.J.; ATHANASOPOULOS, G.. Forecasting: Principles and practice. **Texts**, 2018. Disponível em: <https://otexts.com/fpp2/arma.html>. Acesso em: 8 out. 2022.

HYNDMAN, R., ATHANASOPOULOS, G., BERGMEIR, C., CACERES, G., CHHAY, L., O'HARAWILD, M., PETROPOULOS, F., RAZBASH, S., WANG, E. E YASMEEEN, F.. **forecast**. Forecasting functions for time series and linear models, 2021. Disponível em: <https://pkg.robjhyndman.com/forecast/>.

HYNDMAN, R. J.; KHANDAKAR, Y.. Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, 27(1), 1–22, 2008. Disponível em: <https://doi.org/10.18637/jss.v027.i03>. Acesso em: 8 nov. 2022.

IBGE. **Para compreender o INPC: um texto simplificado**. Coordenação de Índices de Preços. 6ª ed. Rio de Janeiro, 2012. 56 p. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv88501.pdf>. Acesso em: 30 jan. 2022.

IBGE. **Pesquisa de Orçamento Familiar 2017-2018**. IBGE, Rio de Janeiro, 2019. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101670.pdf>. Acesso em: 15 mai. 2022.

IBGE. **POF 2017-2018**: proporção de domicílios com segurança alimentar fica abaixo do resultado de 2004. Agência IBGE Notícias, Estatísticas Sociais, 17 set. 2020a. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-denoticias/releases/28896-pof-2017-2018-proporcao-de-domicilios-com-seguranca-alimentar-ficaabaixo-do-resultado-de-2004>. Acesso em: 15 mai. 2022.

IBGE. **Sistema Nacional de Índices de Preços ao Consumidor**: Estruturas de ponderação a partir da Pesquisa de Orçamentos Familiares 2017-2018. Rio de Janeiro, 2020b. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101711.pdf>. Acesso em: 30 jan. 2022.

- IBGE. **Pesquisa Nacional por Amostra de Domicílios Contínua** - PNAD Contínua. Séries históricas, 2021. Disponível em: https://www.ibge.gov.br/estatisticas/sociais/trabalho/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?=&t=serieshistoricas&utm_source=landing&utm_medium=explica&utm_campaign=desemprego. Acesso em: 18 mai.2022.
- IBGE. **Inflação**. Instituto Brasileiro de Geografia e Estatística, 2022a. Disponível em: <https://www.ibge.gov.br/explica/inflacao.php>. Acesso em: 18 jul. 2022.
- IBGE. **Sistema IBGE de Recuperação Automática** - SIDRA. Instituto Brasileiro de Geografia e Estatística, 2022b. Disponível em: <https://sidra.ibge.gov.br/pesquisa/snipc/ipca/tabelas>. Acesso em: 27 set. 2022.
- IBGE. **IPCA - Índice Nacional de Preços ao Consumidor Amplo**. IBGE, 2022c. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=downloads>. Acesso em: 13 jan. 2022.
- LJUNG, G. M.; BOX, G. E. P.. On a measure of lack of fit in time series models. **Biometrika**, v. 66, p. 66-72, 1978. Disponível em: <https://doi.org/10.2307/2335207>.
- MAIA, A.G.. **Econometria: conceitos e aplicações**. Campinas: Universidade Estadual de Campinas, 2013.
- MARTINEZ, T. S.; CERQUEIRA, V. dos S. **Estrutura da inflação brasileira: determinantes e desagregação do IPCA**. Repositório do conhecimento IPEA, 2011. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/1422>. Acesso em: 24 jun. 2022.
- ONU. **Transformando Nosso Mundo: A Agenda 2030 para o Desenvolvimento Sustentável**. ONU, Roma, 2015. Disponível em: http://www.itamaraty.gov.br/images/ed_desenvsust/Agenda2030-completo-site.pdf. Acesso em: 13 mai. 2021.
- PENNSYLVANIA STATE UNIVERSITY. **Non-seasonal ARIMA Models**. Stat 510: Applied Time Series Analysis, 2022. Disponível em: <https://online.stat.psu.edu/stat510/lesson/3/3.1>. Acesso em: 13 nov. 2022.
- PFaff B.. **Analysis of Integrated and Cointegrated Time Series with R**. 2ª ed. Nova York: Springer, 2008.

R, CORE TEAM. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. Disponível em: <https://www.R-project.org/>.

REDE PENSSAN. **I VIGISAN**: Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da COVID-19 no Brasil. Brasília, 2021. Disponível em: https://olheparaafome.com.br/VIGISAN_Inseguranca_alimentar.pdf. Acesso em: 15 ago. 2022.

REDE PENSSAN. **II VIGISAN**: Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da COVID-19 no Brasil. Brasília, 2022. Disponível em: <https://olheparaafome.com.br/wp-content/uploads/2022/06/Relatorio-II-VIGISAN-2022.pdf>. Acesso em: 15 ago. 2022.

SIQUEIRA, R. P.. **sidrar**: An Interface to IBGE's SIDRA API. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2022. Disponível em: <https://CRAN.R-project.org/package=sidrar>. Acesso em: 27 jan. 2022.

SPINU, V.; GROLEMUND, G.; WICKHAM, H.. **lubridate**: Make Dealing with Dates a Little Easier, 2021. Disponível em: <https://CRAN.R-project.org/package=lubridate>.

WICKHAM, H. et al.. **ggplot2**: Elegant Graphics for Data Analysis. **Springer-Verlag**, New York, 2016. Disponível em: <https://cran.r-project.org/web/packages/ggplot2/index.html>.

WICKHAM, H.. **stringr**: Simple, Consistent Wrappers for Common String Operations, 2019. Disponível em: <https://CRAN.R-project.org/package=stringr>.

WICKHAM, H.. **tidyr**: Tidy Messy Data, 2021. Disponível em: <https://CRAN.R-project.org/package=tidyr>.

WICKHAM, H. et al.. **dplyr**: A Grammar of Data Manipulation, 2022. Disponível em: <https://CRAN.R-project.org/package=dplyr>.

XIE, Y.. **knitr**: A General-Purpose Package for Dynamic Report Generation in R, 2021. Disponível em: <https://yihui.org/knitr/>.

ZHU, H.. **kableExtra**: Construct Complex Table with kable and Pipe Syntax, 2021. Disponível em: <https://CRAN.R-project.org/package=kableExtra>.

APÊNDICES

APÊNDICE A – Código do software R com processamento, manipulação e análise dos dados

Bibliotecas

```
library(tidyverse); library(lubridate); library(tsibble); library(fabletools); library(feasts); library(tseries);
library(forecast); library(uroot); library(urca)
```

Importação dos dados

```
IPCA_06_22_COMPLETO<- read_rds("IPCA_06_22_FINAL.rds")
```

Processamento e manipulação de dados

Organização dados gerais

```
IPCA_06_22_AL_DOM_BR <- IPCA_06_22_COMPLETO %>%
  filter(NOME_GEO == "Brasil",
         VAR != "Peso mensal",
         COD_SUB_GRUPO %in% c(0, 1, 11, 1101:1116),
         ifelse(test = VAR == "Variação acumulada no ano",
                yes = month(DATE) == month(12),
                no = TRUE)) %>%
  drop_na(POND) %>%
  select(NOME_GEO, ANO, DATE, PERIODO, POF, VAR, CADEIA, NIVEL_PROC, COD_REF,
         COD_SUB_GRUPO, NOME_SUB_GRUPO, ITEM, VALOR, POND)
```

Organização variação mensal IPCA

```
IPCA_06_22_AL_DOM_BR_VARMES <- IPCA_06_22_AL_DOM_BR %>%
  filter(VAR == "Variação mensal",
         CADEIA %in% c("Arroz", "Feijão", "Mandioca", "Tomate", "Bovinocultura corte", "Bovinocultura
  leite")) %>%
  arrange(DATE, CADEIA) %>%
  drop_na(CADEIA) %>%
  mutate(VAL_POND = VALOR * POND) %>%
  group_by(DATE, CADEIA) %>%
  summarise(SOMA_VAL_POND = sum(VAL_POND),
            SOMA_POND = sum(POND),
            VAL_INFLA = round(SOMA_VAL_POND/SOMA_POND, 4)) %>%
  mutate(CAT = case_when(
    CADEIA %in% c("Bovinocultura corte", "Bovinocultura leite") ~ "Pecuária",
    CADEIA %in% c("Arroz", "Feijão") ~ "Grãos",
    CADEIA %in% c("Mandioca", "Tomate") ~ "Hortícolas")) %>%
  ungroup() %>%
  select(DATE, CADEIA, CAT, VALOR)
```

Converter banco de dados para `tsibble`

```
IPCA_TS <- IPCA_06_22_AL_DOM_BR_VARMES %>%
  mutate(DATE = yearmonth(DATE)) %>%
  as_tsibble(index = DATE, key = c(CADEIA, CAT))
```

Análise exploratória

Visualização dos dados

```
cat <- unique(IPCA_TS$CAT)
for(i in 1:length(cat)){
  dados <- IPCA_06_22_AL_DOM_BR_VARMES %>%
    filter(CAT == cat[i])

  graf <- ggplot(data = dados,
```

```

      aes(x = DATE, y = VALOR))+
    geom_line(size = 0.5)+
    geom_hline(yintercept = 0, linetype = 2, color = "red")+
    facet_wrap(~CADEIA, nrow = 2, scales = "free")+
    scale_x_date(breaks = seq(as.Date("2006-07-01"),
                             as.Date("2022-09-01"),
                             by="1 year"),
                date_labels = "%Y")+
    labs(title = "", x="", y = "Variação mensal da inflação (%)")+
    theme_minimal()+
    theme(strip.text.x = element_text(size = 15),
          strip.text.y = element_text(size = 15),
          axis.text.x = element_text(size = 13),
          axis.text.y = element_text(size = 13),
          title = element_text(size = 13),
          plot.title = element_text(hjust = 0.5))

print(graf)
}

# Estacionariedade
## Teste de raiz unitária – Dickey-Fuller Aumentado
cat <- unique(IPCA_TS$CADEIA)
for(i in 1:length(cat)){
  dados <- IPCA_TS %>% filter(CADEIA == cat[i])

  DFA <- urca::ur.df(dados$VALOR, type = "none", lags = 10, selectlags = "BIC") %>% summary()

  print(cat[i])
  print(DFA)
}

# Estimação dos modelos
## Correlograma
cat <- unique(IPCA_TS$CADEIA)
for(i in 1:length(cat)){
  dados <- IPCA_TS %>% filter(CADEIA == cat[i])

  grafACF <- forecast::ggAcf(dados$VALOR, calc.ci = T, level = 95)+
    labs(title = cat[i], x = "Defasagem", y = "FAC")+
    theme_minimal()+
    theme(strip.text.x = element_text(size = 15),
          strip.text.y = element_text(size = 15),
          axis.text.x = element_text(size = 13),
          axis.text.y = element_text(size = 13),
          title = element_text(size = 13))

  grafPACF <- forecast::ggPacf(dados$VALOR)+
    labs(title = "", x = "Defasagem", y = "FACP")+
    theme_minimal()+
    theme(strip.text.x = element_text(size = 15),
          strip.text.y = element_text(size = 15),
          axis.text.x = element_text(size = 13),
          axis.text.y = element_text(size = 13))

  library(patchwork)
  gg <- grafACF + grafPACF
  print(gg)
}

```

Modelagem

Arroz ----

```
FIT_ARROZ <- IPCA_TS %>%
  filter(CADEIA == "Arroz") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_ARROZ
```

Banana ----

```
FIT_BANANA <- IPCA_TS %>%
  filter(CADEIA == "Banana") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_BANANA
```

Bovinocultura corte ----

```
FIT_BOVINOS <- IPCA_TS %>%
  filter(CADEIA == "Bovinocultura corte") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_BOVINOS
```

Bovinocultura leite ----

```
FIT_LEITE <- IPCA_TS %>%
  filter(CADEIA == "Bovinocultura leite") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_LEITE
```

Feijão ----

```
FIT_FEIJAO <- IPCA_TS %>%
  filter(CADEIA == "Feijão") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_FEIJAO
```

Mandioca ----

```
FIT_MANDIOCA <- IPCA_TS %>%
  filter(CADEIA == "Mandioca") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_MANDIOCA
```

Tomate ----

```
FIT_TOMATE <- IPCA_TS %>%
  filter(CADEIA == "Tomate") %>%
  select(VALOR) %>%
  forecast::auto.arima(stationary = TRUE, seasonal = F, trace = TRUE, stepwise=FALSE,
approximation=FALSE)
FIT_TOMATE
```


Diagnóstico do modelo## Visualização dos resíduos do modelo

```

cat <- unique(IPCA_TS$CADEIA)
fit_cad <- list(FIT_ARROZ, FIT_BANANA, FIT_BOVINOS, FIT_MANDIOCA, FIT_FEIJAO, FIT_LEITE,
FIT_TOMATE)
for(i in 1:length(fit_cad)){
  dados <-
  ggAcf(fit_cad[[i]]$residuals)+
  labs(title = cat[i], x = "Defasagem", y = "FAC")+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15))

  print(dados)
}

```

Teste de autocorrelação (Ljung-Box)

```

cat <- unique(IPCA_TS$CADEIA)
fit_cad <- list(FIT_ARROZ, FIT_BANANA, FIT_BOVINOS, FIT_LEITE, FIT_FEIJAO, FIT_MANDIOCA,
FIT_TOMATE)
for(i in 1:length(fit_cad)){
  dados <- stats::Box.test(fit_cad[[i]]$residuals, type = "Ljung-Box")

  print(cat[i])
  print(dados)
}

```

Ajuste dos modelos

```
DATE = unique(IPCA_06_22_AL_DOM_BR_VARMES$DATE)
```

```
# Arroz ----
```

```
MOD_ARROZ <- data.frame(DATE, Observado = FIT_ARROZ[["x"]], Modelo = FIT_ARROZ[["fitted"]])
%>%
  pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

```

```
MOD_ARROZ %>%
```

```

  ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
  geom_line(size = 1.05)+
  scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
  labs(title = "Arroz", color = "", x = "", y = "Variação mensal da inflação (%)")+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13, angle = 90),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15),
        legend.text = element_text(size = 13))

```

```
# Banana ----
```

```

MOD_BANANA <- data.frame(DATE, Observado = FIT_BANANA[["x"]], Modelo =
FIT_BANANA[["fitted"]]) %>%
  pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")
MOD_BANANA %>%
  ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +

```

```

geom_line(size = 1.05)+
scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
labs(title = "Banana", color = "", x = "", y = "Variação mensal da inflação (%))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13, angle = 90),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15),
      legend.text = element_text(size = 13))

# Bovinocultura corte ----
MOD_BOVINOS <- data.frame(DATE, Observado = FIT_BOVINOS[["x"]], Modelo =
FIT_BOVINOS[["fitted"]]) %>%
pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

MOD_BOVINOS %>%
ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
geom_line(size = 1.05)+
scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
labs(title = "Bovinocultura corte", color = "", x = "", y = "Variação mensal da inflação (%))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13, angle = 90),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15),
      legend.text = element_text(size = 13))

# Bovinocultura leite ----
MOD_LEITE <- data.frame(DATE, Observado = FIT_LEITE[["x"]], Modelo = FIT_LEITE[["fitted"]]) %>%
pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

MOD_LEITE %>%
ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
geom_line(size = 1.05)+
scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
labs(title = "Bovinocultura leite", color = "", x = "", y = "Variação mensal da inflação (%))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13, angle = 90),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15),
      legend.text = element_text(size = 13))

# Feijão ----
MOD_FEIJAO <- data.frame(DATE, Observado = FIT_FEIJAO[["x"]], Modelo =
FIT_FEIJAO[["fitted"]]) %>%
pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

```

```

MOD_FEIJAO %>%
  ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
  geom_line(size = 1.05)+
  scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
  labs(title = "Feijão", color = "", x = "", y = "Variação mensal da inflação (%)")+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13, angle = 90),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15),
        legend.text = element_text(size = 13))

# Mandioca ----
MOD_MANDIOCA <- data.frame(DATE, Observado = FIT_MANDIOCA[["x"]], Modelo =
FIT_MANDIOCA[["fitted"]]) %>%
  pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

MOD_MANDIOCA %>%
  ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
  geom_line(size = 1.05)+
  scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
  labs(title = "Mandioca", color = "", x = "", y = "Variação mensal da inflação (%)")+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13, angle = 90),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15),
        legend.text = element_text(size = 13))

# Tomate ----
MOD_TOMATE <- data.frame(DATE, Observado = FIT_TOMATE[["x"]], Modelo =
FIT_TOMATE[["fitted"]]) %>%
  pivot_longer(cols = c(Observado, Modelo), names_to = "VAR", values_to = "VALOR")

MOD_TOMATE %>%
  ggplot(aes(x = DATE, y = VALOR, color = VAR, group = VAR)) +
  geom_line(size = 1.05)+
  scale_x_date(breaks = seq(as.Date("2006-07-01"),
                           as.Date("2022-09-01"),
                           by="6 month"),
              date_labels = "%b-%Y")+
  labs(title = "Tomate", color = "", x = "", y = "Variação mensal da inflação (%)")+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13, angle = 90),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15),
        legend.text = element_text(size = 13))

```

```

# Previsão
# Arroz ----
FIT_ARROZ %>%
  forecast::forecast(h = 6)
  autoplot()+
  geom_hline(yintercept = 0, color = "red", linetype = 2)+
  scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15))+
  labs(title = "Arroz", x = "", y = "Variação mensal da inflação (%)")

# Banana ----
FIT_BANANA %>%
  forecast::forecast(h = 6)
  autoplot()+
  geom_hline(yintercept = 0, color = "red", linetype = 2)+
  scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15))+
  labs(title = "Banana", x = "", y = "Variação mensal da inflação (%)")

# Bovinocultura corte ----
FIT_BOVINOS %>%
  forecast::forecast(h = 6)
  autoplot()+
  geom_hline(yintercept = 0, color = "red", linetype = 2)+
  scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15))+
  labs(title = "Bovinocultura corte", x = "", y = "Variação mensal da inflação (%)")

# Bovinocultura leite ----
FIT_LEITE %>%
  forecast::forecast(h = 6)
  autoplot()+
  geom_hline(yintercept = 0, color = "red", linetype = 2)+
  scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
  theme_minimal()+
  theme(strip.text.x = element_text(size = 15),
        strip.text.y = element_text(size = 15),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        title = element_text(size = 15))+
  labs(title = "Bovinocultura leite", x = "", y = "Variação mensal da inflação (%)")

# Feijão ----
FIT_FEIJAO %>%
  forecast::forecast(h = 6)

```

```

autoplots()+
geom_hline(yintercept = 0, color = "red", linetype = 2)+
scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15))+
labs(title = "Feijão", x = "", y = "Variação mensal da inflação (%)")

# Mandioca ----
FIT_MANDIOCA %>%
forecast::forecast(h = 6)
autoplots()+
geom_hline(yintercept = 0, color = "red", linetype = 2)+
scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15))+
labs(title = "Mandioca", x = "", y = "Variação mensal da inflação (%)")

# Tomate ----
FIT_TOMATE %>%
forecast::forecast(h = 6)
autoplots()+
geom_hline(yintercept = 0, color = "red", linetype = 2)+
scale_x_continuous(breaks = seq(2006, 2023, by = 1))+
theme_minimal()+
theme(strip.text.x = element_text(size = 15),
      strip.text.y = element_text(size = 15),
      axis.text.x = element_text(size = 13),
      axis.text.y = element_text(size = 13),
      title = element_text(size = 15))+
labs(title = "Tomate", x = "", y = "Variação mensal da inflação (%)")

```