

Artigo original

DOI: <https://doi.org/10.5281/zenodo.13625559>**PREVISÃO DO COMPORTAMENTO DOS PREÇOS DO LEITE USANDO
MODELOS ARIMA E SARIMA***PREDICTING THE BEHAVIOR OF MILK PRICES USING ARIMA AND SARIMA MODELS*Eunice Henriques Pereira Vilela¹ Antonio Sérgio Torres Penedo² **RESUMO**

Considerando a relevância econômica e social da cadeia produtiva do leite para o Brasil e, especialmente para a região do Triângulo Mineiro e Alto Paranaíba (TMAP) – principal bacia produtora de leite do país, bem como o fato do leite ser uma das *commodities* agrícolas com a maior volatilidade de preços, este trabalho teve por objetivo desenvolver um modelo de previsão do comportamento dos preços do leite na região do TMAP. Para tanto, foi utilizado um Modelo Autorregressivo Integrado e de Médias Móveis na sua forma simples (ARIMA) e com componente sazonal (SARIMA). Os resultados indicaram que o modelo que apresentou o melhor desempenho foi o modelo SARIMA (0,1,1)(2,1,0)₁₂, com um MSE de 0,285 e um R² de 0,557. Também foi possível observar que, dentre os modelos ARIMA e SARIMA testados, os modelos com componente sazonal apresentaram desempenho superior aos modelos sem a sua inclusão, corroborando a existência de comportamento sazonal nos preços do leite.

PALAVRAS-CHAVE: Leite. Previsão de Preços. ARIMA.**ABSTRACT**

Considering the economic and social relevance of the milk production chain for Brazil, and especially for the Triângulo Mineiro and Alto Paranaíba (TMAP) region – the main milk producing basin in the country, as well as the fact that milk is one of the agricultural commodities with the highest price volatility, this work aimed to develop a model to predict the behavior of milk prices in the TMAP region. For this purpose, an Integrated Autoregressive and Moving Average Model was used in its simple form (ARIMA) and with a seasonal component (SARIMA). The results indicated that the model that presented the best performance was the SARIMA(0,1,1)(2,1,0)₁₂ model,

Autor corresponde: Eunice Henriques Pereira Vilela, eunice.hvilela@gmail.com

1,2 Universidade Federal de Uberlândia

with an MSE of 0.285 and an R^2 of 0.557. It was also possible to observe that, among the ARIMA and SARIMA models tested, the models with a seasonal component performed better than the models without its inclusion, corroborating the existence of seasonal behavior in milk prices.

KEYWORDS: Milk. Price Forecast. ARIMA.

INTRODUÇÃO

O agronegócio é um dos setores de maior relevância no cenário brasileiro apresentando desdobramentos significativos no âmbito do comércio internacional e reconhecimento como um vetor crucial do crescimento econômico. A produção de leite se caracteriza como uma das principais atividades agropecuárias praticadas no Brasil, com relevância tanto do ponto de vista econômico quanto social. Sua cadeia produtiva é capaz de gerar impactos em diversos setores da economia, uma vez que apresenta forte correlação e dependência com outras cadeias produtivas.

No Brasil, a produção de leite está presente em praticamente todos os municípios e em quase 90% das propriedades rurais, e envolve mais de um milhão de produtores no campo, além de gerar outros milhões de empregos nos demais segmentos da cadeia (ROCHA, CARVALHO; RESENDE, 2020). Essa produção concentra-se especialmente em algumas regiões, como é o caso da mesorregião do Triângulo Mineiro e Alto Paranaíba (TMAP) – principal bacia produtora de leite do país e responsável por aproximadamente 7% da produção nacional.

Um dos grandes desafios dessa atividade é a volatilidade dos preços. As oscilações dos preços desestabilizam o mercado econômico e afetam as decisões de gestão de produtores, intermediários e consumidores. Nesse cenário, o leite é considerado como uma das *commodities* agrícolas mais voláteis no mercado internacional (ACOSTA; IHLE; ROBLES, 2014; BORAWSKI et al., 2020). Essa volatilidade é, em grande parte, causada pelas próprias características do setor como a extrema perecibilidade do produto e de seus derivados, a elasticidade da demanda e o comportamento sazonal da produção. Porém, fatores como as mudanças nas políticas ou questões regulatórias e o aumento da exposição aos mercados internacionais também têm contribuído para o aumento da volatilidade em diversos mercados (DONG; DU; GOLD, 2011).

As mudanças climáticas também se tornaram, ao longo das últimas décadas, fatores que exercem influência sobre os preços do leite. Isto se dá pela variação de curto prazo no volume produzido decorrente de oscilações na temperatura e na precipitação que podem restringir a disponibilidade de pastagens para a alimentação do rebanho, bem como, provocar choques aleatórios nos custos de insumos de produção – como energia elétrica e ração, por exemplo (NICHOLSON; FIDDAMAN, 2003).

Assim, o aumento na volatilidade dos preços do leite não é mais percebido como um fenômeno temporário, mas como a nova tendência dos mercados globais. Para contornar as dificuldades impostas, é necessário que, cada vez mais, produtores e formuladores de políticas públicas busquem aprimorar seu conhecimento a respeito do funcionamento dos mercados, bem como do comportamento dos preços. O estudo e conhecimento a respeito do comportamento dos preços, bem como dos fatores que o influenciam tornam-se fundamentais tanto para o desenho de políticas públicas apropriadas para reduzir o nível de volatilidade dos preços do leite, quanto para subsidiar a tomada de decisão em condições incertas (ACOSTA; IHLE; ROBLES, 2014; HANSEN; LI, 2017).

Nessa perspectiva, a negociação em mercados futuros vem sendo defendida como uma possível e interessante ferramenta de gerenciamento de risco. No entanto, para que os produtores possam negociar neste tipo de mercado, é necessário que consigam fazer estimativas de preços à longo prazo de modo a subsidiar a tomada de decisão. Hansen e Li (2017), afirmam que compreensão do comportamento dos preços é, nesses casos, um elemento crítico que pode influenciar significativamente os retornos dos participantes do mercado de laticínios. Com este intuito, o interesse por modelos de previsão de preços tem se tornado crescente.

Considerando tal cenário este trabalho se propõe a desenvolver um modelo de previsão do comportamento dos preços do leite na região do Triângulo Mineiro e Alto Paranaíba. A região foi escolhida como objeto de estudo, uma vez que se trata da principal bacia produtora de leite do Brasil. O modelo utilizado adota é uma ARIMA (Autorregressivo Integrado e de Médias Móveis) na sua forma padrão e com a inclusão do componente sazonal (SARIMA).

Além desta introdução, o trabalho está dividido em 5 sessões, que compreendem o referencial teórico utilizado para a construção do trabalho, os aspectos metodológicos pertinentes a elaboração da pesquisa, a apresentação dos resultados obtidos e as considerações finais sobre os achados da pesquisa.

REFERENCIAL TEÓRICO

A cadeia produtiva do agronegócio apresenta uma complexidade inerente devido aos diversos fatores envolvidos em sua produção e comercialização, o que acarreta inúmeros riscos e incertezas no setor (RODRIGUES et al. 2018). A eliminação desses riscos, contudo, não viável e nem desejável, uma vez que poderia também limitar a possibilidade de maiores rendimentos. Dessa forma faz-se fundamental que os produtores utilizem ferramentas para a gestão de riscos (MOREIRA; PROTIL; SILVA, 2014).

Uma das tarefas cruciais para o gerenciamento de riscos é a previsão de preços dos produtos agrícolas, fator importante para todo o processo de tomada de decisão (RODRIGUES et al. 2018). Todavia, como explicam Pinheiro e Sena (2017), o fato de as atividades agropecuárias serem caracterizadas por movimentos cíclicos, elevada volatilidade e sofrerem influência de vários fatores do mercado constitui-se num obstáculo para sua previsibilidade.

Nesse sentido, Shyian et al. (2021) explicam que a previsão é um elemento importante do processo de tomada de decisão econômica em qualquer nível hierárquico de gestão e em qualquer área da atividade humana, inclusive econômica. Sendo especialmente importante em áreas da economia com produção sazonal e um período significativo de rotação de capital. Tais áreas incluem a agricultura e a pecuária.

Esta última possui características próprias no desenvolvimento e operação, uma das quais é um longo período de giro do capital investido. Portanto, a previsão de indicadores financeiros e econômicos de suas atividades é relevante para que as empresas planejem suas receitas e despesas futuras das atividades operacionais e desenvolvam a estratégia correta. De modo que a previsão dos preços dos produtos pela administração da empresa cria condições favoráveis para a tomada de decisões fundamentadas (SHYIAN et al., 2021).

A previsão dos preços passa, portanto, a ser uma das principais ferramentas para execução do planejamento e avaliação das atividades agrícolas uma vez que é fator de decisão dos investimentos. No caso do leite, a previsão dos preços futuros é importante para a gestão de risco, por possibilitar o uso dos mercados futuros e de opções para gerenciar a variação no risco do preço do leite, a gestão oportuna de possíveis déficits de fluxo de caixa, e as decisões de expansão ou descarte do rebanho – tendo em vista que as decisões para aumentar a produção têm um lapso de tempo; de modo a tornar fazendas mais resilientes aos ciclos de preços (HANSEN; LI, 2017).

As previsões desempenham também um papel fundamental na definição de políticas econômicas adequadas. A estimação do preço futuro do leite pode subsidiar as autoridades na formulação de políticas para conter a instabilidade de preços, reduzir os riscos do mercado e aumentar o bem-estar dos produtores e consumidores (FAHIMY ET AL., 2011; SHAHRIARY; MIR, 2016). Vários métodos são usados e explorados para prever variáveis econômicas, sendo os mais tradicionais a regressão ou técnica polinomial, a regressão espontânea, a média animada, os modelos de Box-Jenkins, os modelos estruturais, os modelos de séries temporais (SHAHRIARY; MIR, 2016).

Dentre essas técnicas, a mais popular é a metodologia Box-Jenkins, utilizada para a construção de modelos paramétricos para séries temporais univariadas. Tais modelos são popularmente conhecidos como modelos ARIMA (Autorregressivos Integrados e de Médias Móveis). A metodologia de Box Jenkins pressupõe que as informações estão contidas na própria série temporal. Assim, os modelos Autorregressivos Integrados e de Médias Móveis (ARIMA), são modelos matemáticos que possuem por objetivo compreender o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal e, com o embasamento nesse comportamento, realizar previsões futuras. Ou seja, é o comportamento da variável em estudo que responderá por sua dinâmica futura.

A relação temporal considerada pela metodologia de Box e Jenkins (1976) é representada por um conjunto de características estocásticas, que incorporam três componentes, também denominados “filtros”. Assim o modelo é denotado como ARIMA (p, d, q), onde p representa o componente Autorregressivo (AR), d é o filtro

de Integração (I) e q o componente de Médias Móveis (MA). Outra possibilidade é a incorporação de componentes sazonais no modelo, adotando um Autorregressivo Integrado e de Médias Móveis Sazonais (SARIMA), definido como SARIMA(p,d,q)(P,D,Q), onde P é o Componente Autorregressivo Sazonal, D é o termo de diferenciação sazonal do modelo, e Q é o Componente de Médias Móveis Sazonais (GUJARATI; PORTER, 2008).

Através de revisão da literatura foram encontrados dos trabalhos de Fachinello e Bacchi (2006); Fachinello e Ponchio (2006); Alves, Sousa e Ervilha (2014); Shahriary e Mir (2016); e Bergmann, O'connor e Thümmel (2019) que buscaram realizar a previsão de preços no mercado do leite adotando os métodos ARIMA ou SARIMA. De modo geral as conclusões desses trabalhos apontam para uma elevada capacidade preditiva dos modelos, configurando assim boas alternativas para a previsão do comportamento dos preços no mercado do leite.

MÉTODO

Nesta seção são descritos os procedimentos metodológicos utilizados para a construção dos modelos de previsão do comportamento de preços do leite pago ao produtor na mesorregião do Triângulo Mineiro e Alto Paranaíba. A operacionalização do Modelo ARIMA é composta por quatro etapas: identificação, estimação, diagnóstico e previsão.

Na etapa de identificação, busca-se verificar se a série é estacionária, para tanto são utilizados os Testes de Dickey Fuller Aumentado (ADF) e o Teste de Phillips-Perron (PP). Caso seja identificada a não estacionariedade da série é necessário transformar os dados em séries temporais estacionárias, obtendo a primeira diferença da série temporal, antes de construir o modelo preditivo. O nível de integração (ou diferenciação) da série será denotado pelo componente I . De modo que, caso seja identificado que a série é estacionária em nível, o I (d) do modelo ARIMA assumirá a ordem 0; se a série não for estacionária em nível, mas tornar-se quando diferenciada em primeira ordem, o I assumirá o valor 1, e assim sucessivamente.

Nessa etapa, também é necessário verificar se a série apresenta padrão sazonal, o que leva à escolha da adoção de um modelo ARIMA, ou ARIMA sazonal (SARIMA). Para verificar se a série apresenta comportamento sazonal, é utilizada a

análise dos gráficos decompostos por meio do método de Decomposição de Tendência Sazonal (STL), e a análise das funções de autocorrelação (FAC) e função de autocorrelação parcial (FACP). Nos correlogramas da FAC e da FACP é necessário observar as características do comportamento da série. Caso a série apresente um comportamento senoidal, isto indica a existência de um comportamento sazonal na série de preços.

As assinaturas AR e MA também devem ser determinadas usando gráficos de função de autocorrelação (FAC) e função de autocorrelação parcial (FACP). A FAC define a defasagem do MA e a FACP define a defasagem do AR. No primeiro caso, sabe-se que a função de autocorrelação decai com o aumento de defasagens, e a função de autocorrelação parcial é truncada a partir da defasagem p . No segundo caso, ocorre o inverso: a função de autocorrelação é truncada na defasagem q , e a função de autocorrelação parcial decai. No caso de uma ARMA p, q , ambas as funções decaem a partir da defasagem de truncagem.

A partir da análise do comportamento dessas funções devem ser definidos os parâmetros de ajuste dos modelos potenciais e selecionar o melhor modelo usando critérios adequados. O modelo selecionado deve ser aquele que apresenta os menores valores dos critérios de informação (PEPPLE; HARRISON, 2017). Esse processo pode ser realizado por meio de sucessivas estimações e da análise de suas respectivas estatísticas t e dos critérios de informação, buscando identificar o modelo que apresente os melhores resultados.

Uma vez selecionada a estrutura de modelo ARIMA que apresenta os melhores resultados, procede-se a sua estimação. Após a estimação do modelo é necessário proceder à etapa de diagnóstico dos resultados. Nessa etapa devem ser analisadas as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) dos resíduos da estimação, buscando identificar se estes apresentam características de ruído branco. De acordo com Bueno (2012), uma sequência é um ruído branco se cada valor nela tiver média zero, variância constante e não for correlacionado a qualquer realização da própria série (autocorrelação igual a zero).

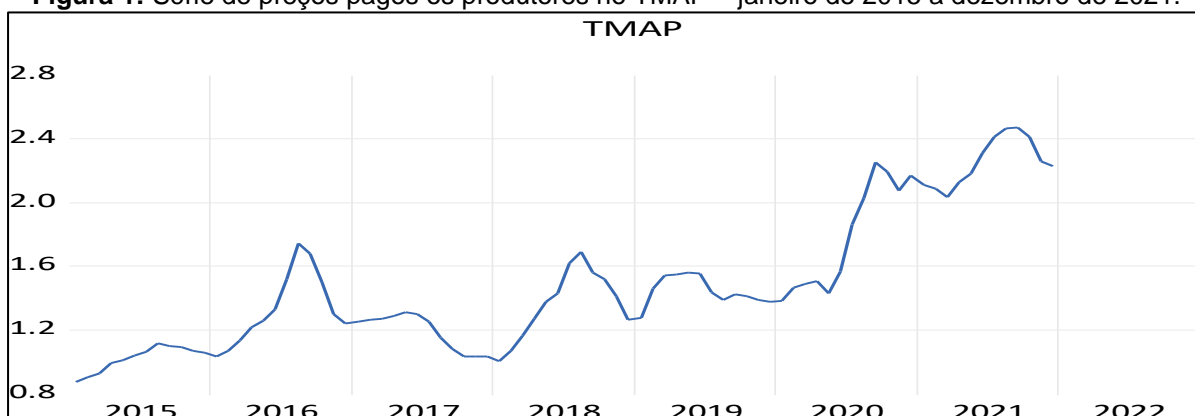
Na etapa de diagnóstico também devem ser verificados eventuais problemas de autocorrelação e heterocedasticidade do modelo, por meio da análise da estatística de Durbin-Watson e da aplicação do Teste de ARCH. Caso os resíduos

apresentem características de ruído branco e não sejam identificados problemas de autocorrelação e heterocedasticidade no modelo, o passo seguinte consiste na utilização do modelo para realizar as previsões. Por fim, é necessário analisar a confiabilidade dos valores previstos, para tanto são utilizadas como métricas o coeficiente de correlação (R^2), e o Erro Médio Quadrado (MSE).

RESULTADOS E DISCUSSÃO

Como explicado no tópico anterior, o primeiro passo para a operacionalização de um modelo ARIMA consiste na identificação das características dos dados, que é feita em duas etapas. A primeira consiste em verificar se a série é ou não estacionária, para tanto foi inicialmente utilizada a análise gráfica, que pode ser observada na Figura 1:

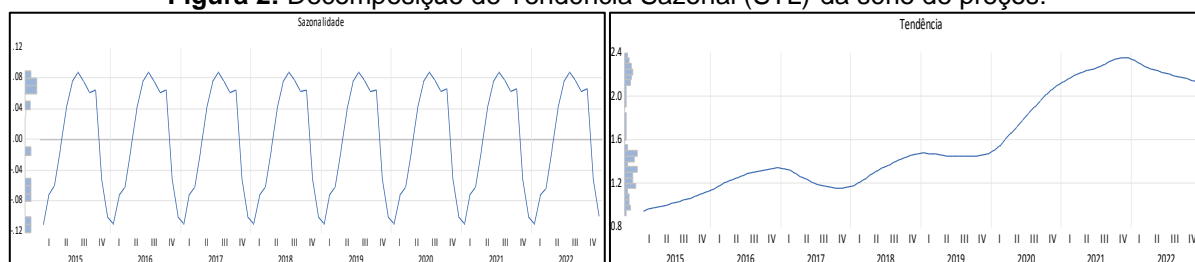
Figura 1: Série de preços pagos os produtores no TMAP – janeiro de 2015 a dezembro de 2021.

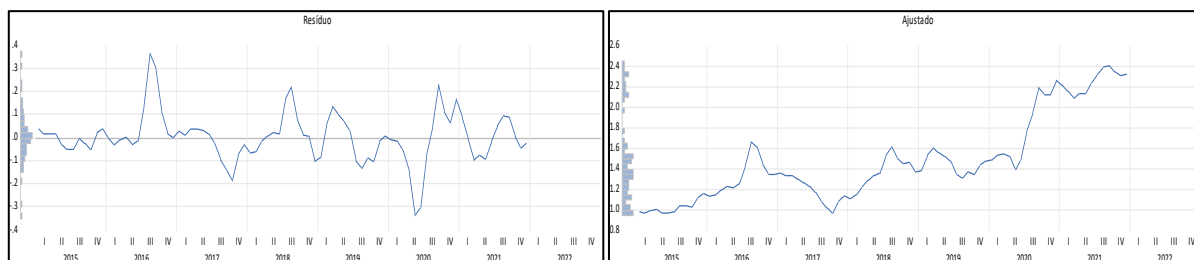


Fonte: Elaboração própria com base nos resultados.

Por meio da análise gráfica é possível observar que preços do leite pago aos produtores no TMAP apresentam um comportamento irregular com uma tendência de alta ao longo do período analisado. Para melhor analisar esse comportamento a série foi decomposta por meio do método de Decomposição de Tendência Sazonal (STL), como pode ser observado na Figura 2:

Figura 2: Decomposição de Tendência Sazonal (STL) da série de preços:





Fonte: Elaboração própria com base nos resultados.

A análise gráfica das séries obtidas por meio da Decomposição de Tendência Sazonal (STL), permite afirmar que os preços do leite pagos aos produtores no TMAP são influenciados por fatores sazonais, apresentando um comportamento de alta dos preços no início segundo trimestre do ano que perdura até o fim do terceiro trimestre – que coincide com o período de estiagem na região, e uma redução expressiva dos preços no início do quarto trimestre, com os preços permanecendo reduzidos até meados do primeiro trimestre do ano subsequente. Também é possível observar que os preços do leite pago aos produtores no TMAP apresentam uma tendência de alta irregular ao longo do período analisado, sendo esse um comportamento típico de uma série não estacionária.

A segunda parte da identificação dos dados consistiu em verificar se a série de preços é estacionária. Para a verificar a presença de raiz unitária na série de preços foram aplicados os Testes de Dickey Fuller Aumentado (ADF) e Philips Peron (PP), em nível e na primeira diferença. Seus resultados são apresentados na Tabela 1:

Tabela 1: Resultados do Teste de Raiz Unitária - Modelo ARIMA

Critério	ADF				Philips Peron	
	Nível		1ª dif		Nível	1ª dif
	SC	AIC	SC	AIC		
TMAP	-3,195*	-1,987	-5,125*	-5,125*	-2,763	-4,979*

Legenda: (Nível) equação do teste em nível, (1ªDif) equação do teste na primeira diferença, (AIC) critério de informação de Akaike Modificado, (SC) critério de informação de Schwarz.

Fonte: Elaboração Própria com base nos resultados.

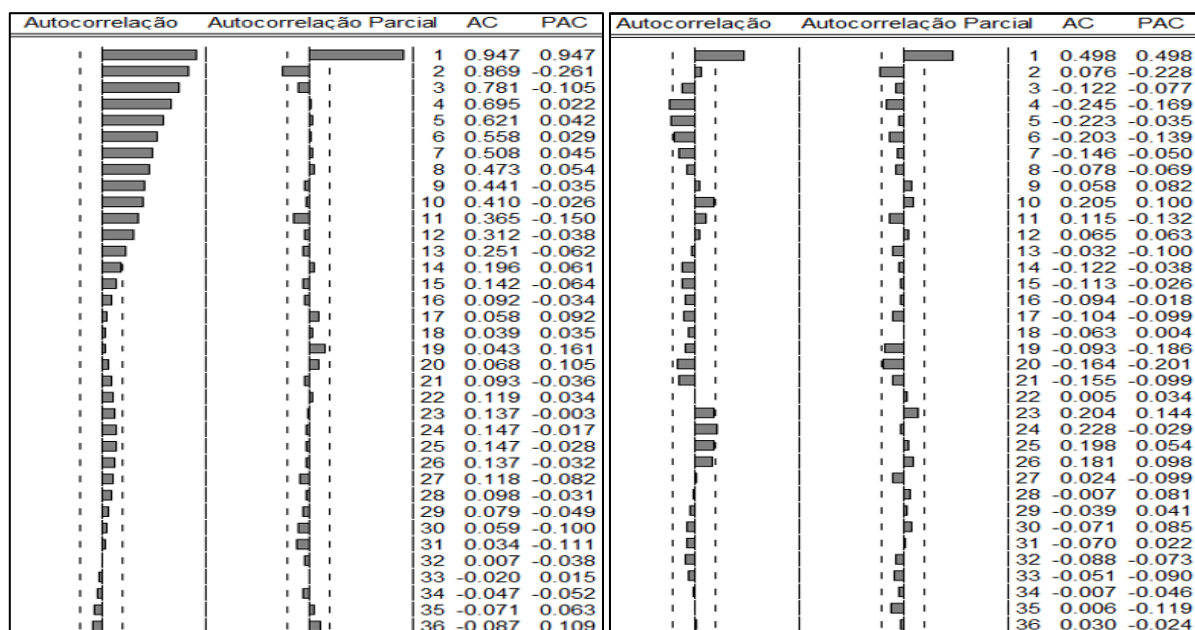
Tanto os resultados do Teste ADF quanto do Philips Peron, indicam que a série de preços é não estacionária, porém torna-se quando diferenciada em primeira ordem. Esse resultado indica que, para a estimação do modelo ARIMA, a série deve ser utilizada na sua forma diferenciada. Portanto, o parâmetro “I” do modelo deve ter valor 1, indicando o número de vezes que a série deve ser diferenciada.

Identificada a ordem de integração da série (d), o próximo passo foi identificar o número de termos autorregressivos (p) e o número de termos de média móvel (q) no modelo. Para isso, foi usada a Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP). Como pode ser observado na Figura 3, a FAC inicia com um valor alto e diminui gradualmente, tendo 13 defasagens fora do intervalo de confiança de 95%, o que deixa os coeficientes estatisticamente diferentes de zero. O que é um indicativo de não estacionariedade. Observa-se também pela Figura 3 que a FACP possui uma defasagem fora do intervalo de confiança. Diante desse resultado, obteve-se também a FAC e FACP da série de preços em primeira diferença (Figura 4).

Os resultados mostraram que a FAC da primeira diferença da série não apresenta características de uma série não estacionária, corroborando os resultados dos testes de raiz unitária. Outra característica observada é a presença de comportamento senoidal, o que sugere a existência de um comportamento sazonal na série de preços, corroborando os resultados da Decomposição de Tendência Sazonal (STL).

Figura 3: FAC e FACP da série em nível:

Figura 4: FAC e FACP da série em diferença:



Fonte: Resultados da pesquisa.

Uma vez identificada a necessidade de se trabalhar com a série de preços diferenciada em primeira ordem e a presença de comportamento sazonal, procede-se a segunda etapa da operacionalização do modelo, que consiste na estimação dos parâmetros do modelo a ser adotado. Nesse processo, foram estimados 8 modelos, cujos resultados são apresentados na Tabela 2:

Tabela 2: Resultados da Estimação dos Modelos ARIMA

Variável	ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,1,1)	ARIMA (2,1,1)	ARIMA (2,1,2)	ARIMA (0,1,1)
SIGMA	0,006*	0,006*	0,006*	0,006*	0,006*	0,006*	0,006*
AR(1)	0,509*	0,616*	0,603*	0,257	1,189	1,167	
AR(2)		-0,201*	-0,171		-0,491*	0,445*	
AR(3)			-0,061				
MA(1)				0,357	-0,617*	-0,071	0,557*
MA(2)						0,006*	
R ²	0,241	0,275	0,277	0,270	0,288	0,289	0,257
R ² ajustado	0,232	0,256	0,249	0,252	0,261	0,252	0,248
Critério de Inf. de Akaike	-2,188	-2,208	-2,188	-2,202	-2,203	-2,179	-2,201
Critério de Inf. de Schwarz	-2,129	-2,121	-2,071	-2,115	-2,086	-2,034	-2,150
Critério Inf. Hannan-Quinn	-2,165	-2,173	-2,141	-2,167	-2,155	-2,121	-2,185
Estatística de Durbin-Watson	1,773	2,006	1,998	1,977	1,961	1,984	1,853

Fonte: Elaboração própria com base nos resultados.

O primeiro modelo testado foi um modelo ARIMA (1,1,0), e apresentou um coeficiente estatisticamente significativo ao nível de 1% do componente autorregressivo (AR), de 0,509, e um R² de 0,241 e um R² ajustado. O segundo

modelo testado foi um ARIMA (2,1,0), que também apresentou um coeficiente estatisticamente significativo ao nível de 1% do componente AR (1) de 0,616, e um coeficiente estatisticamente significativo ao nível de 5% do componente AR (2) de -0,210. O R^2 do modelo foi de 0,275 e o R^2 ajustado foi de 0,256 apresentado uma melhora de desempenho em relação ao modelo ARIMA (1,1,0). Com relação aos critérios de informação, o valor do critério de Schwarz diminuiu (de -2,129 para -2,121), já os dos critérios de Akaike e Hannan-Quinn aumentaram (de -2,188 para -2,208, e de -2,165 para 2,173, respectivamente) o que indica uma melhora no ajuste do modelo.

Uma vez que houve melhora do ajuste com a inclusão do componente AR (2), testou-se a inclusão do componente AR (3), de modo que o terceiro modelo testado foi um ARIMA (3,1,0). Nesse modelo, os coeficientes dos componentes AR (1) e AR (2) continuaram sendo estatisticamente significativos ao nível de 5%, já o componente AR(3) não apresentou significância estatística ao nível de 5%. O R^2 aumentou para 0,277, porém o R^2 ajustado diminuiu para 0,249, e todos os critérios de informação apresentaram redução em suas estatísticas, o que indica que a inclusão do componente AR (3) não melhora o desempenho do modelo.

O quarto e o quinto modelos testados tiveram a inclusão do componente de médias móveis (MA), sendo o quarto modelo um ARIMA (1,1,1) e o quinto um ARIMA (2,1,1). No modelo ARIMA (1,1,1) o componente MA (1) não apresentou um coeficiente estatisticamente significativo, e tanto os valores de R^2 e R^2 ajustado quando os dos critérios de informação apresentação uma elevação em relação ao primeiro modelo testado, o que indica que a inclusão do componente de médias móveis (MA) melhorou o desempenho do modelo. O mesmo ocorre quando se compara as estatísticas do quinto modelo em relação ao terceiro. Já quando se compara o desempenho dos modelos ARIMA (1,1,1) e ARIMA (2,1,1), observa-se que todos os valores do R^2 , do R^2 ajustado e do critério de informação de Akaike aumentam, já os valores dos critérios de informação de Schwarz e Hannan-Quinn se reduzem.

No sexto modelo testado houve a inclusão do componente MA(2), sendo este, portanto, um modelo ARIMA (2,1,2). Nesse caso, houve uma elevação do R^2 e uma redução do R^2 ajustado e dos três critérios de informação, o que indica que a inclusão

do componente MA(2) não melhora o desempenho do modelo. O sétimo modelo testado foi um ARIMA (0,0,1), este apresentou o maior valor dos critérios de Schwarz e Hanann-Quinn com relação a todos os demais modelos já testados, do entanto, tanto o R^2 quanto o R^2 ajustado e o valor do critério de Akaike diminuíram, apresentando um desempenho inferior aos modelos ARIMA (2,1,0), ARIMA (1,1,1) e ARIMA (2,1,1).

Para validar os resultados identificados anteriormente, foi utilizada uma função do *Eviews 12* denominada *automatic ARIMA forecasting* (previsão ARIMA automática), que realiza o teste com diversos estrutura de modelo ARIMA e ranqueia os resultados de acordo com o critério de informação selecionado. Foi testada também a possibilidade de inclusão do componente Autorregressivo Sazonal (SAR) e do componente de Médias Móveis Sazonais (SMA), configurando a possibilidade da adoção de um modelo ARIMA com componente sazonal (SARIMA).

Conforme os resultados apresentados na Tabela 3, foram testados 81 modelos ARIMA e SARIMA, com composições de até 2 componentes Autorregressivos (AR), 2 componentes de Médias Móveis (MA), 2 componentes Autorregressivos Sazonais (SAR) e 2 componentes de Médias Móveis Sazonais (SMA).

Tabela 3: Resultados da Aplicação da Previsão ARIMA Automática.

Variável Dependente Selecionada			D(TMAP)
Amostra:	2015M01 a 2021M12	Obs:	83
Modelo Máximo:	(2,2) 1 (2,2)	Modelos ARMA Estimados:	81
Modelo Selecionado:	(0,1) (0,0)	Valor SIC:	-2,150
Modelo	AIC	SIC**	HQ
(0,1) (0,0)	-2,208	-2,150*	-2,185*
(1,0) (0,0)	-2,188	-2,129	-2,165
(2,0) (0,0)	-2,208	-2,121	-2,173
(1,1) (0,0)	-2,202	-2,115	-2,167
(0,2) (0,0)	-2,199	-2,112	-2,164
(0,1) (1,1)	-2,225*	-2,109	-2,179
(0,1) (1,0)	-2,195	-2,108	-2,160
(0,1)(0,1)	-2,192	-2,105	-2,157
(0,1) (2,0)	-2,212	-2,096	-2,165
(2,1) (0,0)	-2,203	-2,087	-2,157

Legenda: (AIC) critério de informação de Akaike Modificado, (SIC) critério de informação de Schwarz, (HQ) critério de informação de Hannan-Quinn.

Nota:* Modelo com melhor desempenho de acordo com o critério; ** critério de ranqueamento adotado.

Fonte: Elaboração própria com base nos resultados.

Os resultados indicaram que, conforme os critérios de informação de Schwarz e Hannan-Quinn o modelo que apresentou os melhores resultados foi o ARIMA (0,1,1), e de acordo com o critério de informação de Akaike o melhor resultado foi obtido pelo modelo SARIMA (0,1,1) (1,1,1), esse modelo também foi apontado pelo critério de Hannan-Quinn como o segundo melhor desempenho. Já o segundo modelo que apresentou o melhor desempenho de acordo com o critério de Schwarz foi o ARIMA (1,1,0), e de acordo com o critério de Akaike, o segundo modelo que apresentou o melhor desempenho foi o modelo SARIMA (0,1,1)(2,1,0)₁₂.

Assim, optou-se por estimar os modelos SARIMA (0,1,1) (1,1,1)₁₂ e SARIMA (0,1,1) (2,1,0)₁₂, fazer a o diagnóstico e a previsão utilizando esses quatro modelos e comparar o desempenho destes na etapa de previsão. Os resultados da estimação dos modelos SARIMA (0,1,1) (1,1,1)₁₂ e SARIMA (0,1,1) (2,1,0)₁₂, bem como dos modelos ARIMA (1,1,0) e ARIMA(0,1,1) são apresentados na Tabela 4.

Comparando os resultados dos modelos testados, todos apresentaram coeficientes estatisticamente significantes ao nível de 5% para os componentes Autorregressivos (AR), de Médias Móveis (MA), Autorregressivos Sazonais (SAR) e de Médias Móveis Sazonais (SMA) testados. Com relação ao desempenho dos modelos, o Modelo ARIMA (1,1,0) apresentou um R² de 0,241 e um R² ajustado de 0,232, o Modelo ARIMA(0,1,1,) apresentou um R² de 0,257 e um R² ajustado de 0,248, o Modelo SARIMA (0,1,1) (1,1,1)₁₂ apresentou um R² de 0,391 e um R² ajustado de 0,368, sendo portanto aquele que apresentou o melhor desempenho entre os testados, e o Modelo SARIMA (0,1,1)(2,1,0)₁₂, apresentou um R² de 0,311 e um R² ajustado de 0,285.

Tabela 4: Resultados da Estimação dos Modelos ARIMA e SARIMA

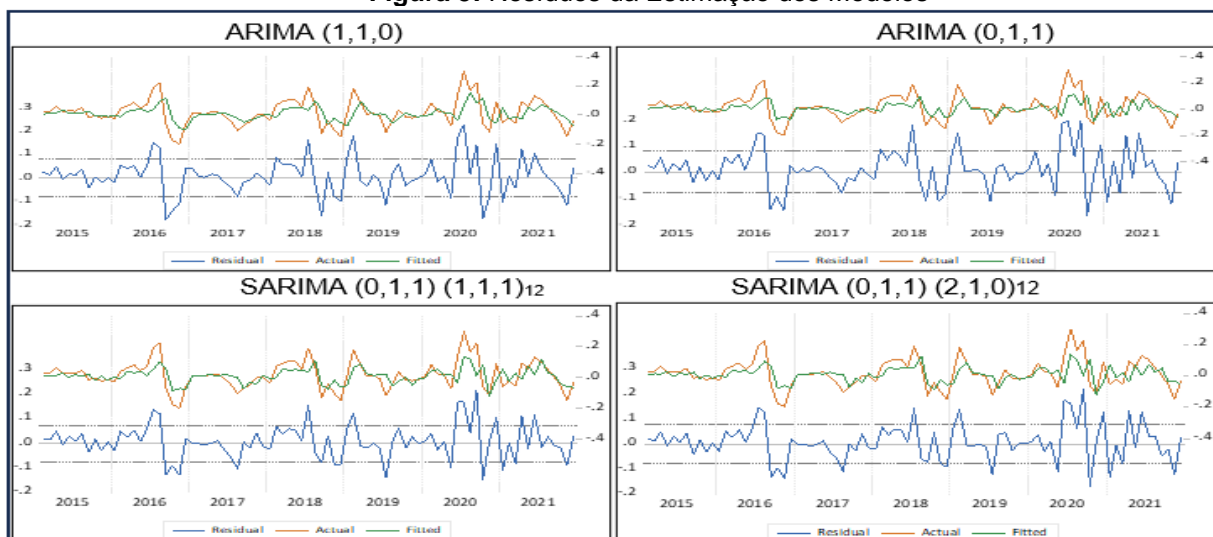
Variável	ARIMA (1,1,0)	ARIMA (0,1,1)	SARIMA (0,1,1)(1,1,1) ₁₂	SARIMA (0,1,1)(2,1,0) ₁₂
SIGMA	0,006*	0,006*	0,005*	0,006*
AR(1)	0,509*			
SAR(12)			1,000*	0,107
SAR(24)				0,264*
MA(1)		0,557*	0,556*	0,549*
SMA(12)			-0,999*	

R ²	0,241	0,257	0,391	0,311
R ² ajustado	0,232	0,248	0,368	0,285
Crítério de Inf. de Akaike	-2,188	-2,201	-2,225	-2,212
Crítério de Inf. de Schwarz	-2,129	-2,150	-2,109	-2,096
Crítério de Inf. de Hannan-Quinn	-2,165	-2,185	-2,179	-2,165
Estatística de Durbin-Watson	1,773	1,853	1,945	1,950

Fonte: Elaboração própria com base nos resultados.

Outro indicativo importante que pode ser observado é que todos os modelos apresentaram resultados da estatística de Durbin-Watson dentro do intervalo de confiança do teste, de modo que é possível afirmar que os modelos não apresentam indícios de autocorrelação. O passo seguinte no diagnóstico, consistiu em analisar os resíduos da regressão. A Figura 5 apresenta o comportamento dos resíduos da estimação:

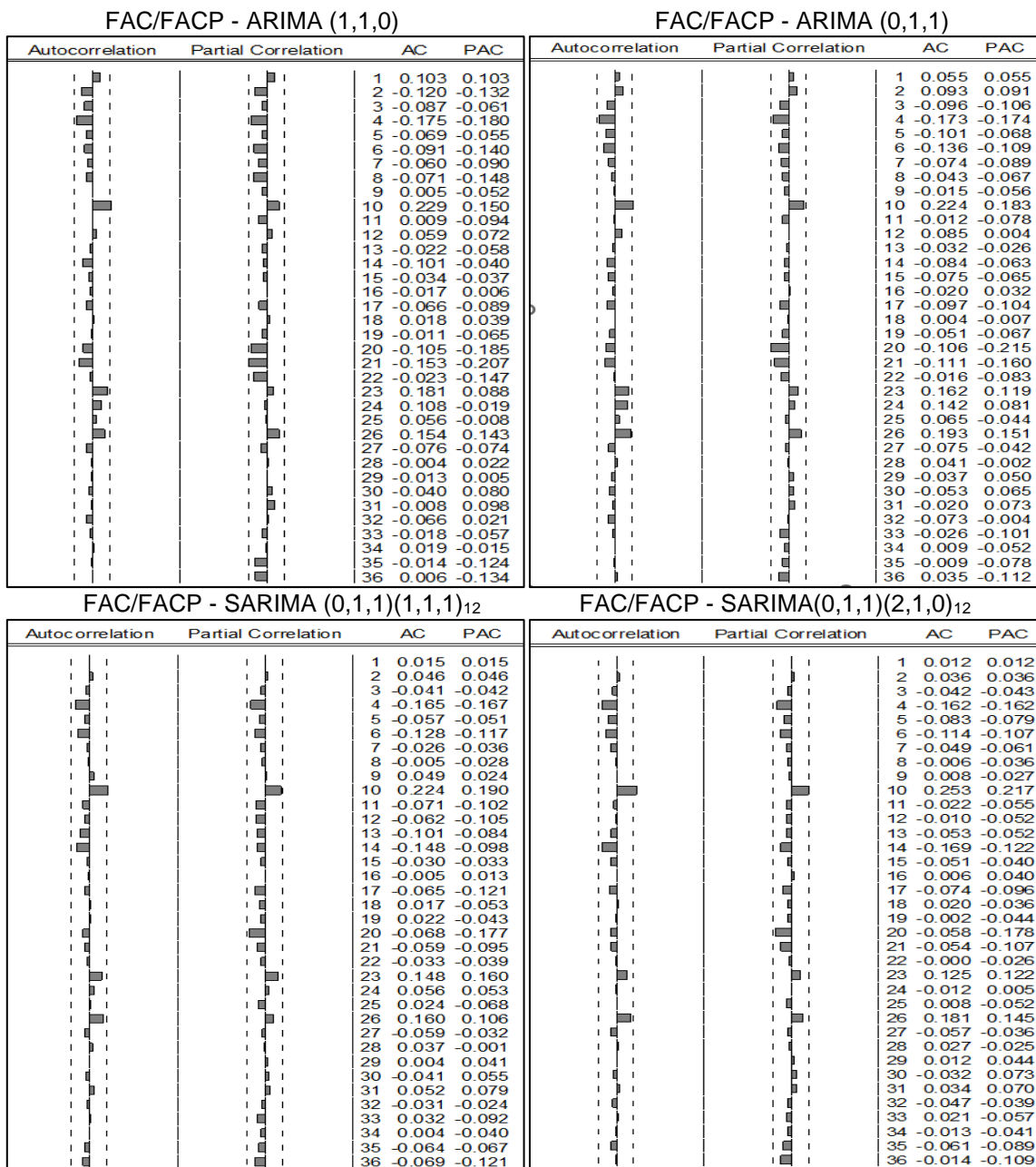
Figura 5: Resíduos da Estimação dos Modelos



Fonte: Resultados da pesquisa.

A análise gráfica permite observar que, em todos os modelos testados, os resíduos das estimações são estáveis, ou seja, flutuam em torno de zero. No entanto, observa-se que estes apresentam certa volatilidade. Para aprofundar a análise dos resíduos da estimação, foram analisados os correlogramas das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP), que podem ser observadas no Figura 6.

Figura 6: FAC e FACP dos resíduos da estimação dos Modelos:



Fonte: Resultados da pesquisa.

Os correlogramas das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP), mostram que, para todos os modelos testados, os coeficientes dos resíduos estão dentro do intervalo, o que indica que estes não são estatisticamente significativos, ou seja, o comportamento da FAC e da FACP dos resíduos das estimações se comportam como um ruído branco, o que se trata de uma propriedade desejável.

Por fim, para verificar a existência de heterocedasticidade nos resíduos da estimação dos modelos, foi aplicado o teste estatístico ARCH, cujos resultados estão dispostos na Tabela 5:

Tabela 5: Resultados do Teste ARCH

	ARIMA (1,1,0)	ARIMA (0,1,1)	SARIMA (0,1,1)(1,1,1) ₁₂	SARIMA (0,1,1)(2,1,0) ₁₂
Estatística F	6,429	5,613	6,029	8,723
Obs. R ²	6,099	5,376	5,746	8,062
Prob.F(1,80)	0,0132	0,020	0,016	0,041
Prob. Chi. ²	0,0135	0,0204	0,016	0,045
C	0,004*	0,004*	0,004*	0,004*
RESID ² (-1)	0,272*	0,256*	0,264*	0,313*
R ²	0,074	0,065	0,070	0,098
R ² Ajustado	0,062	0,054	0,058	0,087
Obs	82	82	82	82

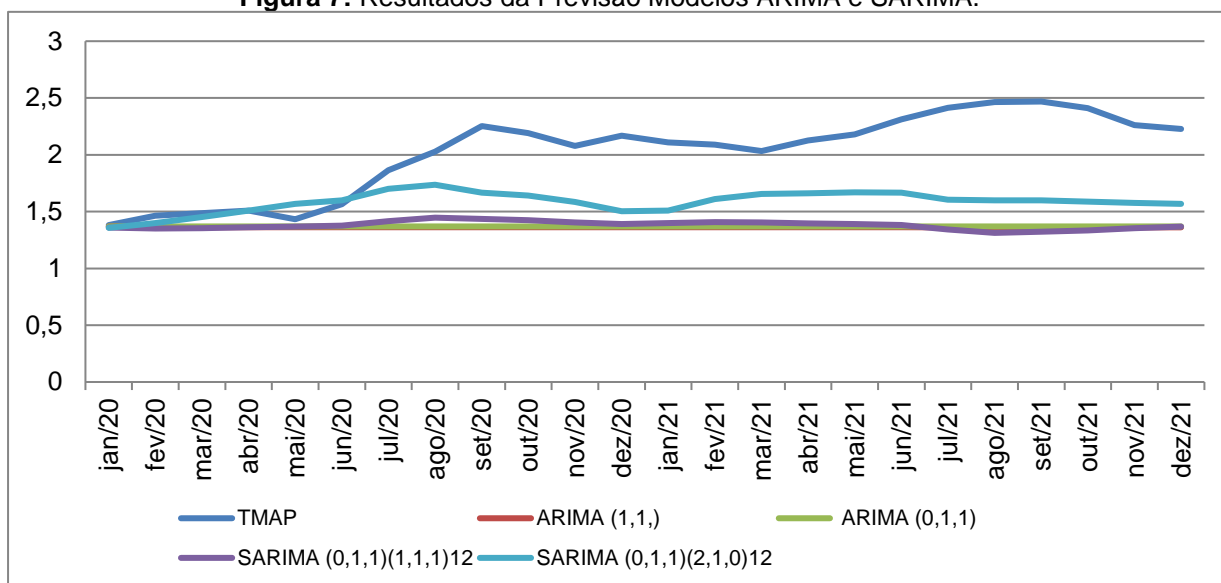
Fonte: Elaboração própria com base nos resultados

O teste de heterocedasticidade ARCH, apresentou uma estatística *t* de 0,0123 para o Modelo ARIMA (1,1,0), de 0,020 para o Modelo ARIMA(0,1,1), de 0,016 para o Modelo SARIMA (0,1,1)(1,1,1)₁₂, e de 0,041 para o Modelo SARIMA (0,1,1)(2,1,0)₁₂, o que permite aceitar, ao nível de significância de 1%, a hipótese nula do teste - presença de homoscedasticidade.

Assim, uma vez realizado a diagnóstico dos modelos e descartada a presença de autocorrelação e heterocedasticidade em seus resíduos, bem como identificado o comportamento de ruído branco destes, o passo seguinte consiste utilização dos modelos para realizar previsões e na análise dos seus resultados.

As previsões foram realizadas para o período de janeiro de 2020 a dezembro de 2021, utilizando o método de previsão dinâmico. Os valores previstos pelos modelos, juntamente com os preços praticados no TMAP no período de janeiro de 2020 a dezembro de 2021 são apresentados na Figura 7:

Figura 7: Resultados da Previsão Modelos ARIMA e SARIMA.



Fonte: Resultados da Pesquisa.

Analisando os valores pelos modelos em comparação com os preços pagos aos produtores do Triângulo Mineiro e Alto Paranaíba, observa-se que o modelo que apresentou o melhor desempenho preditivo foi o modelo SARIMA(0,1,1)(2,1,0)₁₂. Este modelo conseguiu prever os movimentos de alta e baixa, apresentando, no entanto, variações de menor magnitude do que aquelas efetivamente ocorridas no período. Esse desempenho foi corroborado pelas estatísticas das métricas de erro adotadas (MSE e R²) cujos resultados podem ser observados na Tabela 6:

Tabela 6: Estatísticas de Desempenhos dos Modelos ARIMA e SARIMA

Modelo	MSE	R2
ARIMA (1,1,)	0,5565	-0,6733
ARIMA (0,1,1)	0,5458	0,0000
SARIMA (0,1,1)(1,1,1) ₁₂	0,5349	-0,0401
SARIMA (0,1,1)(2,1,0) ₁₂	0,2850	0,5570

Fonte: Elaboração própria com base nos resultados.

De acordo com as métricas de desempenho consideradas, o modelo que apresentou o melhor desempenho preditivo foi o modelo SARIMA (0,1,1) (2,1,0)₁₂, com um MSE de 0,285 e um coeficiente de correlação de 0,557. Já o segundo melhor desempenho, medido pelo MSE, foi o do modelo SARIMA (0,1,1) (1,1,1)₁₂, com um erro médio quadrado (MSE) de 0,5349, este modelo, já quando medido pelo coeficiente de correlação (R²) houve uma grande redução de desempenho preditivo,

apresentando um R^2 de -0,0401, valor que, por ser muito próximo de zero, indica uma baixa capacidade do modelo de prever o preço do leite pago aos produtores no TMAP.

CONSIDERAÇÕES FINAIS

A partir dos resultados obtidos, considera-se que o objetivo proposto para este artigo foi atingido, de modo que foi possível elaborar um modelo que apresenta desempenho satisfatório para realizar previsões para o comportamento dos preços do leite no Triângulo Mineiro e Alto Paranaíba. Dessa forma, é possível afirmar que utilização de modelos de autorregressivos para previsão de preços é uma alternativa viável para auxiliar os produtores no planejamento de suas atividades de modo a mitigar os riscos provocados pela flutuação dos preços do leite. Destaca-se também os modelos com componente sazonal apresentaram desempenho superior aos modelos sem a sua inclusão, corroborando a existência de comportamento sazonal nos preços do leite apontada por Marin, Cavalheiro e Anschau (2011), Nicholson e Stephenson (2015) e Paura e Arhipova (2016).

REFERÊNCIAS

ACOSTA, A.; IHLE, R.; ROBLES, M. Spatial price transmission of soaring milk prices from global to domestic markets. **Agribusiness**, v. 30, n. 1, p. 64-73, 2014

ALVES, F.F.; SOUSA, L.V. C.; ERVILHA, G.T. Planejamento e previsão do preço do leite em Minas Gerais: análise empírica com base no modelo X12-ARIMA. **Revista de Economia e Agronegócio – REA**, v. 12, n 1, 2, 3, p.115–134, 2015.

BERGMANN, D.; O’CONNOR, D.; THÜMMEL, A. An evaluation of point and density forecasts for selected EU farm gate milk prices. **International Journal of Food and Agricultural Economics (IJFAEC)**, v. 6, n. 1128-2019-535, p. 23-53, 2019.

BÓRAWSKI, P.; PAWLEWICZ, A.; PARZONKO, A.; HARPER, J.; HOLDEN, L. Factors shaping cow’s milk production in the EU. **Sustainability**, v. 12, n. 1, p. 420, 2020.

BUENO, R. L. S. **Econometria de séries temporais**. São Paulo: Cengage Learning, 2012.

DONG, F.; DU, X.; GOULD, B. W. Milk Price Volatility and its Determinants. **Agricultural and Applied Economics Association**, 2011.

FACHINELLO, A. L.; BACCHI, M. R. P. Determinação de um modelo de previsão univariado para preços de leite pagos aos produtores em Santa Catarina. **Análise Econômica**, v. 24, n. 46, 2006.

FAHINELLO, A. L.; PONCHIO, L. A. Modelo de previsão univariado para preços de leite pagos aos produtos nas principais regiões brasileiras. In: XLIV **Congresso Brasileiro de Economia e Sociologia Rural**, 2006, Fortaleza. Anais... Fortaleza: SOBER, 2006. CD-ROM.

FAHIMY, D. S.; SALARPOUR, M.; SOUBOOHI, M. Comparisons Of The Ability Of Predicting Anfis Fuzzy-Neural Model With Ann Neural Network And Arima Auto Regression (Case Study Of Weekly Price Of Egg). **Journal of Agricultural Economics and Development**, 74, 188-204, 2011.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5ª ed. Porto Alegre: AMGH, 2011.

HANSEN, B. G.; LI, Y. An analysis of past world market prices of feed and milk and predictions for the future. **Agribusiness**, v. 33, n. 2, p. 175-193, 2017.

MARIN, S. R.; CAVALHEIRO, A. G.; ANSCHAU, D. Sazonalidade do preço do leite no Rio Grande do Sul (1986-2009). **Ciência Rural**, v. 41, p. 361-364, 2011.

MOREIRA, V.; PROTIL, R.; SILVA, C. L. Gestão dos riscos de mercado do agronegócio no contexto das cooperativas agroindustriais. Anais: **Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural**, Campo Grande, 2014.

NICHOLSON, C. F.; FIDDAMAN, T. **Dairy policy and price volatility**. 2003.

PAURA, L.; ARHIPOVA, I. Analysis of the milk production and milk price in Latvia. **Procedia Economics and Finance**, v. 39, p. 39-43, 2016.

PEPPLE, S. U.; HARRISON, E. E. Comparative performance of Garch and Sarima techniques in the modeling of Nigerian board money. **CARD Int J Soc Sci Confl Manag**, v. 2, n. 4, p. 258-270, 2017.

PINHEIRO, C. A. O.; SENNA, V. de. Previsão de preços através de redes neurais e análise espectral: evidências para o mercado futuro das commodities açúcar e soja. **Custos e @gronegócio on-line**, v. 13, n. 4, p. 103-128, 2017.

ROCHA, D. T.; CARVALHO, G. R.; DE RESENDE, J. C. **Cadeia produtiva do leite no Brasil: produção primária**. Embrapa Gado de Leite - Circular Técnica (INFOTECA-E), 2020.

RODRIGUES, L. S.; REZENDE, S. O.; MOURA, M. F.; MARCACINI, R. M. Agribusiness Time Series Forecasting using Perceptually Important Events. In: 2018 **XLIV Latin American Computer Conference (CLEI)**. IEEE, 2018. p. 268-277.

SHAHRIARY, G.; MIR, Y. Application of artificial neural network model in predicting price of milk in Iran. **Modern Applied Science**, v. 10, n. 4, p. 173-178, 2016.

SHYIAN, N.; MOSKALENKO, V.; SHABINSKYI, O.; PECHKO, V. Milk price modeling and forecasting. **Agricultural and Resource Economics: International Scientific E-Journal**, v. 7, n. 1868-2021-1004, p. 81-95, 2021.

- **Financiamento:**

Este trabalho foi financiado pelo seguinte órgão de fomento: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Extraído de Tese de Doutorado “Comportamento e previsão dos preços do leite no Triângulo Mineiro e Alto Paranaíba” apresentada ao Programa de Pós-graduação em Administração (PPGAdm) da Universidade Federal de Uberlândia, na cidade de Uberlândia (BR), no ano de 2023. Disponível em: <https://repositorio.ufu.br/handle/123456789/38508>