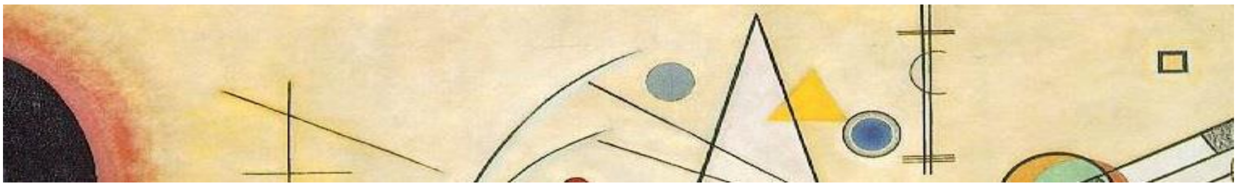


Contemporânea
Contemporary Journal
3(10): 18181-18202, 2023
ISSN: 2447-0961

Artigo

APLICAÇÃO DO MODELO ARIMA À PREVISÃO DO PREÇO DA SOJA COMO FERRAMENTA PARA TOMADA DE DECISÃO

APPLICATION OF THE ARIMA MODEL TO SOYBEAN PRICE PREDICTION AS A TOOL FOR DECISION MAKING

DOI: 10.56083/RCV3N10-083
Recebimento do original: 15/09/2023
Aceitação para publicação: 16/10/2023

Andre Luiz Barros Luchesi
Mestre em Contabilidade
Instituição: Faculdade de Tecnologia Senac Cascavel
Endereço: Rua Recife, 2283, Centro, Cascavel – PR, CEP: 85810-031
E-mail: andreluizluchesi@gmail.com

Maria da Piedade Araújo
Doutora em Economia Aplicada
Instituição: Universidade Estadual do Oeste do Paraná (UNIOESTE)
Endereço: Rua Universitária, 1619, Universitário, Cascavel - PR, CEP: 85819-110
E-mail: madadepi@gmail.com

Silvana Anita Walter
Doutora em Administração
Instituição: Universidade Estadual do Oeste do Paraná (UNIOESTE)
Endereço: Rua Universitária, 1619 - Universitário, Cascavel - PR, CEP: 85819-110
E-mail: silvanaanita.walter@gmail.com

RESUMO: O agronegócio é uma atividade que envolve diversos riscos e incertezas, por essa razão torna-se necessária a utilização de ferramentas para auxiliar a tomada de decisão. Esta pesquisa tem por objetivo analisar o desempenho preditivo do modelo ARIMA para as estimativas de preços a curto e médio prazo, apresentando a trajetória e tendências futuras do preço da soja, com a finalidade de auxiliar os agentes envolvidos na comercialização dessa *commodity*. O estudo utilizou preços mensais da soja comercializada na bolsa de valores de Chicago (CBOT), entre janeiro de 2000 a junho de 2018, e utilizou a metodologia Box-Jenkins para desenvolvimento do modelo. O modelo ARIMA (1,2,1) obteve o melhor ajuste aos dados, apresentando uma taxa de erro de apenas 5,39%, de acordo com o indicador de erro MAPE, um percentual de erro baixo, considerando o mercado agrícola. Os resultados demonstram que esse tipo de modelagem matemática constitui-se em uma ferramenta relevante para tomada de decisão, possibilitando aos gestores elaborarem ações para maximizar seus resultados e minimizar os riscos. No entanto, o intervalo de confiança das previsões geradas ainda é relativamente grande, sendo mais indicada para subsidiar decisões a curto prazo. Embora, suas estimativas a médio prazo devam ser consideradas pelos gestores.

PALAVRAS-CHAVE: Modelagem ARIMA, Previsão, Preço Soja.

ABSTRACT: Agribusiness is an activity that involves several risks and uncertainties, which is why it is necessary to use tools to assist decision-making. This research aims to analyze the predictive performance of the ARIMA model for price estimates in the short and medium term, presenting the trajectory and future trends of soybean price, in order to assist the agents involved in the commercialization of this *commodity*. The study used monthly prices of soybeans traded on the Chicago Stock Exchange (CBOT), between January 2000 and June 2018, and used the Box-Jenkins methodology to develop the model. The ARIMA model (1,2,1) obtained the best fit to the data, presenting an error rate of only 5.39%, according to the MAPA error indicator, a low error percentage, considering the agricultural market. The results show that this type of mathematical modeling is a relevant tool for decision making, enabling managers to develop actions to maximize their results and minimize risks. However, the confidence interval of the generated forecasts is still relatively large, and is more suitable for subsidizing short-term decisions. Although, their medium-term estimates should be considered by managers.

KEYWORDS: ARIMA Modeling, Forecasting, Soybean Price.

1. Introdução

A agricultura vem em crescente processo de modernização, utilizando-se cada vez mais de novas tecnologias. O agronegócio é um dos setores mais dinâmicos da economia brasileira, destacando-se como atividade de alto valor econômico e social, demonstrando contínua evolução nos processos de produção, aumentando sua produtividade e reduzindo custos (CAMPOS, 2007). Entre os vários cultivos praticados no Brasil, a soja se consolidou como umas das principais commodities na agricultura brasileira. Em 2018, a produção de soja foi aproximadamente de 117 milhões de toneladas, atrás apenas dos EUA, maior produtor de soja no mundo, com 119 milhões de toneladas. Estima-se que 54% da produção nacional foi comercializada no mercado externo (CONAB, 2018; USDA, 2018).

No entanto, devido à globalização e outros aspectos inerentes ao agronegócio, a soja se tornou uma atividade de alta complexidade, que apresenta vários riscos e incertezas, principalmente associados a fatores climáticos, mercadológicos e conjunturais. Esses influenciam diretamente na rentabilidade e promovem incertezas aos agricultores e aos demais participantes dessa cadeia de produção (MARTINS e MARTINELLI, 2010). Devido à complexidade e os riscos envolvidos no seu cultivo, torna-se fundamental o desenvolvimento de ferramentas para minimizar os riscos e subsidiar a tomada de decisão dos gestores.

Neste sentido, o preço constitui-se em uma variável importante na determinação da oferta e da demanda; consequentemente, a possibilidade de prever o seu comportamento é de suma importância para os participantes do mercado agrícola, sejam eles compradores, vendedores ou especuladores, pois eles baseiam suas decisões em expectativas de preços (RIBEIRO et al., 2010; SCHWAGER, 1995). Uma das técnicas mais utilizadas para a previsão de preço consiste nos modelos univariados. Esse tipo de modelo econométrico parte do princípio de que o próprio comportamento da variável responderá pela sua dinâmica futura. Ele é conhecido na literatura como o método autorregressivo integrado de médias móveis; ou definido simplesmente por modelo ARIMA.

Diante do exposto, este artigo visa a aplicação de modelos ARIMA (Autorregressivos Integrados de Médias Móveis) nos preços da soja comercializada na bolsa de Chicago (CBOT); buscando responder se o modelo de série temporal ARIMA fornece estimativas de preços futuros confiáveis a curto e médio prazo para auxiliar a tomada de decisão. O objetivo é analisar o desempenho preditivo do modelo ARIMA para estimativas de preços a curto e médio prazo, apresentando a trajetória e tendências futuras do preço da soja, com a finalidade de auxiliar os agentes envolvidos na comercialização dessa commodity.

A presente pesquisa está organizada em cinco seções. A primeira consiste nesta introdução; na seção 2 é apresentado o referencial teórico sobre modelo ARIMA e suas propriedades teóricas; na seção 3 são detalhados os procedimentos metodológicos empregados na pesquisa; na seção 4 são apresentadas as análises e discussões dos resultados; por fim, na seção 5 são destacadas as considerações finais.

2. Referencial Teórico

Para obter sustentação teórica, este referencial apresenta os principais conceitos sobre séries temporais. Em seguida é dado um enfoque ao modelo de séries temporais ARIMA e a metodologia Box-Jenkins. Estudos relacionados também serão apresentados para constatar os resultados já obtidos e a relevância da continuidade da discussão do tema.

2.1 Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo e que apresentam dependência serial; ou seja, as suas observações estão ligadas diretamente umas às outras; portanto, uma série temporal deve considerar a ordem cronológica das observações (MORETIN; TOLOI, 2006). O uso de séries temporais aparece em diversas áreas do conhecimento como, por exemplo: na Economia (preços diários de ações, taxa mensal de desemprego, produção industrial), na Medicina (eletrocardiograma, eletroencefalograma), na Epidemiologia (número mensal de novos casos de meningite), na Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), entre outros (EHLERS, 2005).

Para a análise de uma série temporal, inicialmente deve-se modelar o fenômeno a ser estudado. A partir desse ponto pode ser feita a descrição do comportamento da série, as suas estimativas e finalmente a avaliação dos fatores que influenciam no comportamento da série, tendo em vista a definição da relação causa e efeito. Segundo Morettin e Toloi (2006), os objetivos da análise de séries temporais são: caracterizar os fenômenos que dão origem a série temporal; fazer previsões de valores futuros, sendo que as previsões podem ser a curto e longo prazo; descrever o comportamento da série - neste caso a construção de histogramas e diagramas de dispersão, entre outros, podem ser ferramentas úteis; verificar a existência de tendências, ciclos e variações sazonais e procurar periodicidades relevantes nos dados.

A classe de modelos mais utilizada na análise de séries temporais, com objetivo de prever valores futuros é a dos modelos autorregressivos integrados e de médias móveis, ou simplesmente ARIMA, proposta por Box e Jenkins (1970). Segundo o estudo de Morettin e Toloi (1981), que comparou 10 séries para a economia brasileira, com diversas metodologias de previsão, concluiu que os modelos ARIMA apresentam melhor desempenho que as demais metodologias, exceto para previsões de longo prazo (12 períodos), quando a autorregressão mostrou-se mais eficiente.

Os modelos ARIMA resultam da combinação de três componentes: o componente Auto Regressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA). Uma série temporal pode conter os três componentes ou apenas um subconjunto deles, permitindo modelar a variável resposta por meio de componentes autorregressivos e de médias móveis de forma simultânea, podendo ou não apresentar sazonalidade (FAVA, 2000). Os modelos ARIMA são adequados para descrever séries não estacionárias, ou seja, séries que não possuem média constante no período analisado. Segundo Morettin (2006), o modelo ARIMA pode ser classificado de diferentes maneiras, conforme os parâmetros AR(p) (autorregressivos de ordem p), MA(q) (média móvel de ordem q) e I(d) (integrado de ordem d).

O modelo autorregressivo AR(p) é um processo interativo em que há a identificação da ordem p por meio da função de autocorrelação. A partir daí faz-se a estimativa de um modelo de previsão, bem como a análise dos resíduos para a avaliação da existência de vieses e/ou grandes erros de estimativas. Esse componente "prevê" valores futuros com base em uma combinação linear de valores passados. O modelo de média móvel de ordem q - MA(q) é usado quando há autocorrelação entre os resíduos, ou seja, existe uma relação de dependência entre o conjunto de erro em períodos passados. Esse modelo em série temporal utiliza como previsão para uma determinada observação no futuro, a média das observações passadas. As médias móveis podem ser simples, centradas ou ponderadas. O termo médio móvel é utilizado porque, à medida que a próxima observação está disponível, a média das observações é recalculada, incluindo essa observação no conjunto de observações e desprezando a observação mais antiga.

Por fim, o modelo ARIMA contém o parâmetro de Integração I(d), que representa o componente sazonal ou de tendência da série. Ele é utilizado quando há necessidade de aplicar transformações por meio de diferenças para tornar a série estacionária. O valor do parâmetro (d) representa o número de diferenças necessárias para tornar uma série estacionária. Esse parâmetro também é conhecido como ordem de integração (MARGARIDO, 1998).

Segundo Gujarati (2006) existem alguns pressupostos que precisam ser considerados a fim de obter os melhores resultados na utilização da metodologia ARIMA. O principal refere-se à estacionariedade da série de dados; ou seja, a série deve oscilar em torno de uma média e apresentar variância constante. Caso a série não seja estacionária, será necessário transformar os dados originais. A transformação mais comum consiste em aplicar diferenças sucessivas da série original até se obter uma série estacionária. Para as séries temporais, a estacionariedade é fundamental para previsão de dados futuros, a qual tem como premissa que o futuro se comportará de acordo com o passado (MORETIN; TOLOI, 2006; STOCK; WATSON, 2004).

Nos modelos de séries temporais deve-se verificar a dependência entre os dados. Para isso são utilizadas as funções de autocorrelação (ACF) e autocorrelação parcial (PACF). A função de autocorrelação (ACF) descreve o comportamento de uma variável com base no conhecimento exclusivo dos seus valores passados, medindo a correlação entre pares de valores da série defasados em um e mais períodos. O coeficiente de autocorrelação é uma das principais estatísticas na análise de séries temporais. Esse coeficiente tem como principal característica medir como ela está autocorrelacionada às observações de uma série temporal afastadas k períodos entre si. Além de analisar a correlação total entre pares de observações é importante verificar a correlação simples entre Yt e Yt-k depois de eliminar o efeito que os valores intermediários exercem sobre elas. A função de autocorrelação parcial é uma medida da correlação entre as observações de uma série temporal que são separadas por k unidades de tempo (MONTEGOMERY et al., 1990).

Ao trabalhar com séries temporais esses e outros conceitos devem ser compreendidos pelo pesquisador a fim de que o modelo gerado seja o mais coerente possível. Uma metodologia para o desenvolvimento e ajuste de modelos de series temporais é a metodologia Box-Jenkins, apresentada a seguir.

2.2 Metodologia Box-Jenkins

Segundo Morettin e Toloi (2004) a metodologia de Box e Jenkins constitui-se em ajustar modelos autorregressivos integrados a médias móveis – ARIMA (p, d, q) a um conjunto de informações. Esse conjunto de dados constitui a base para a construção do modelo mais adequado. A construção de um modelo ARIMA baseado na metodologia Box-Jenkins obedece a um ciclo iterativo, composto por quatro etapas: identificação, estimação, verificação e previsão (MORETTIN; TOLOI, 2004).

- **Etapa de identificação:** a identificação busca a ordem dos parâmetros do ARIMA (p,d,q). A identificação da ordem p e q é feita por meio da função de autocorrelação (Fac) e da função de autocorrelação parcial (Facp). Além disso, nesta etapa é verificada a estacionariedade da série, que pode ser realizado por meio de testes para identificação de raiz unitária, como o Dickey-Fuller Aumentado (SAID e DICKEY, 1984) ou Phillips-Perron (PHILLIPS e PERRON, 1988), quando a hipótese alternativa é estacionariedade da série. Também é possível verificar estacionariedade observando a existência de tendência ou sazonalidade no gráfico plotado da série temporal; ou por meio do comportamento vagaroso na queda da Fac ao longo das defasagens;
- **Etapa de estimação:** realiza-se a estimativa dos parâmetros do componente autorregressivo, do componente de médias móveis e da variância;
- **Etapa de verificação:** consiste em analisar, por meio da análise dos resíduos, se o modelo escolhido descreve adequadamente o comportamento da série. Os resíduos devem apresentar ausência de autocorrelação. O teste de Ljung-Box (ENDERS, 1995) busca testar a existência significativa de autocorrelação dos resíduos em algum k defasagem. A inexistência de autocorrelação dos resíduos em nenhuma

defasagem implica a ausência de correlação temporal na estrutura dos resíduos do modelo;

- **Etapa de previsão:** etapa que representa o objetivo principal, realizada apenas se as etapas anteriores forem satisfatórias. Assim, sempre que o modelo não se mostrar adequado, as etapas devem ser repetidas (MORETTIN; TOLOI, 2004).

Quando há um conjunto de diferentes modelos é necessário realizar testes e utilizar critérios para mensurar o ajuste do modelo em relação aos dados, e selecionar os melhores modelos de previsão. Os critérios de seleção para modelos ARIMA mais utilizados são o AIC (Akaike information criterion) e o BIC (Bayesian information criterion).

O critério de informação de Akaike (1974) é uma medida relativa da qualidade de ajuste de um modelo estatístico estimado. Fundamenta-se no conceito de entropia, oferecendo uma medida relativa das informações perdidas, quando um determinado modelo é usado para descrever a realidade. O AIC é um critério que avalia a qualidade do ajuste do modelo paramétrico, estimado pelo método da máxima verossimilhança, no qual o melhor modelo será o que minimize o valor do AIC (BURNHAM; ANDERSON, 2002).

O critério de informação Bayesiano (BIC), também chamado de Critério de Schwarz, foi proposto por Schwarz (1978). Esse método possui um formato semelhante ao AIC, embora tenha um embasamento teórico diferente. Seguindo um ponto de vista bayesiano, é um critério de avaliação de modelos definido em termos da probabilidade. O termo relativo ao número de parâmetros, no caso do BIC, é multiplicado por um fator correspondente ao logaritmo natural do número de observações da amostra (BROCKWELL; DAVIS, 1996; BURNHAM; ANDERSON, 2002).

De forma geral, escolhe-se o modelo que apresente menor valor AIC e BIC. Além de levar em consideração os modelos mais “parcimoniosos”, ou seja, aqueles que apresentam o menor número de parâmetros. Por fim, é recomendável que esses critérios sejam avaliados conjuntamente, pois são complementares e não excludentes (ENDERS, 1995). Além disso, para garantir a correta especificação de um modelo ARIMA é importante analisar os resíduos, verificando se as autocorrelações amostrais dos erros seguem assintoticamente uma distribuição normal, com média zero e variância constante. Dessa forma, se o modelo estiver corretamente especificado, os resíduos não devem apresentar correlação serial, pois todas as correlações dos dados já foram capturadas pelo modelo. Então, os valores das autocorrelações residuais devem estar contidos no intervalo de confiança assintótico de 95%. Em adição ao exame das autocorrelações individuais dos resíduos pode ser utilizado o teste Ljung-Box, que verifica a existência significativa de autocorrelação dos resíduos em algum nível de defasagem. A inexistência de autocorrelação implica na ausência de correlação temporal na estrutura dos resíduos do modelo (ENDERS, 1995; GRANGER; NEWBOLD, 1986).

Por fim, deve-se verificar a acurácia do modelo, mensurando o nível de precisão das previsões geradas. Previsões muito próximas dos dados reais significam erros de precisão baixos; logo, o modelo possui um bom nível de precisão, caso contrário, o modelo deve ser alterado ou ajustado. Para analisar e decidir o melhor modelo de previsão que se ajusta aos dados deve-se levar em conta os indicadores de erro médio (ME), erro quadrático médio (RMSE), erro absoluto médio (MAE), erro percentual médio (MPE), erro percentual absoluto médio (MAPE) e erro médio simétrico absoluto percentual (SMAPE).

3. Procedimentos Metodológicos

Trata-se de pesquisa documental de natureza descritiva com abordagem quantitativa, pois desenvolve um modelo econométrico de série temporal ARIMA para previsão de preço da soja, com base em dados históricos. Os dados utilizados referem-se aos preços mensais da soja comercializada na bolsa de valores de Chicago (CBOT) entre os períodos de janeiro de 2000 a junho de 2018, totalizando 222 observações.

Na presente pesquisa foi utilizada a metodologia Box-Jenkins para o desenvolvimento e ajuste no modelo ARIMA. No entanto, antes de construir um modelo ARIMA, alguns cuidados básicos devem ser tomados: a amostra não deve conter menos que 50 observações, os dados não podem apresentar valores negativos, e a série deve ser estacionária ou passível de estacionariedade (BOX; JENKINS, 1976). Segundo Morettin e Toloi (2006), a análise de séries temporal deve ser precedida da verificação da existência de raiz unitária, a qual permite definir o grau de diferenciação que será necessário para que a série se torne estacionária, caso ela não seja. O teste aplicado para verificar a existência de raiz unitária foi o teste Dickey-Fuller Aumentado, que é amplamente utilizado na literatura.

Depois de verificado esses conceitos básicos, o estudo seguiu a metodologia Box-Jenkins, que compreende quatro etapas: identificação, estimação, verificação de diagnóstico e previsão. Na fase de identificação, verificou-se o correlograma e as funções de autocorrelação e autocorrelação parcial (FAC e FACP). Posteriormente foram definidos os parâmetros de defasagens, médias móveis e diferenças que foram utilizados. Em seguida, vários modelos foram estimados, calculando os valores AIC e BIC, utilizados como critérios de seleção dos modelos. Na fase de diagnóstico os resíduos foram testados para verificar se o modelo ajusta-se aos dados; ou seja, foi verificado se os resíduos se tratam de um ruído branco, quando os resíduos possuem as propriedades estatísticas de média zero, variância constante e não apresentam autocorrelação. Também foi aplicado o teste de Ljung-Box, que verifica a existência significativa de autocorrelação dos resíduos em algum nível de defasagem.

Além disso, buscou-se mensurar o nível de precisão das previsões dos modelos selecionados por meio dos indicadores de erro médio (ME), erro quadrático médio (RMSE), erro absoluto médio (MAE), erro percentual médio (MPE) e erro percentual absoluto médio (MAPE). Por fim, gerou-se a previsão dos preços da soja para seis ciclos à frente.

O presente estudo utilizou o software R Studio para a manipulação dos dados e execução dos procedimentos estatísticos.

4. Resultados

Nesta seção os resultados são apresentados e discutidos conforme a sequência de etapas da metodologia Box-Jenkins. A primeira etapa consiste na identificação do modelo, no qual é verificado o pressuposto da estacionariedade da série e analisado a dependência entre os dados, utilizando as funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

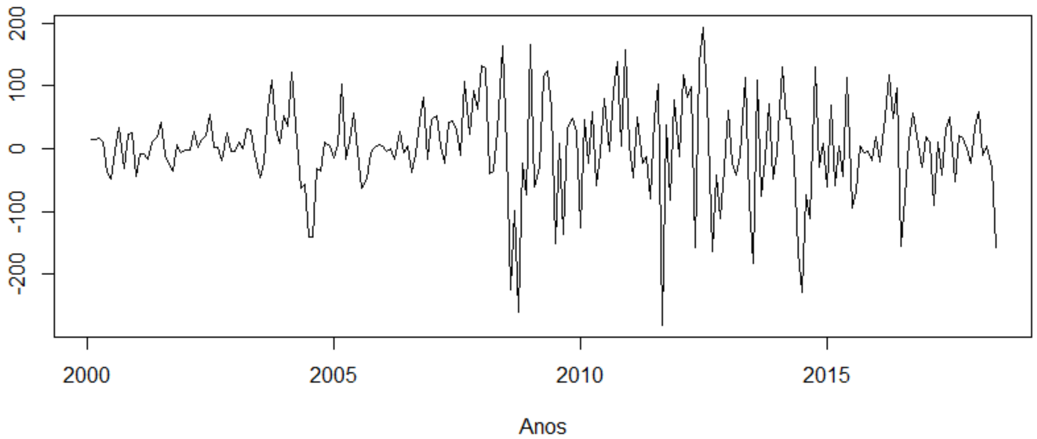
Analisando a série histórica de preço da soja (figura 1), nota-se que a série não oscila em torno da média e não apresenta variância constante, indicando que ela não é estacionária. Para verificar a estacionariedade de forma objetiva foi utilizado o teste de raiz unitária Dickey-Fuller Aumentado, com o qual se constatou que a série em análise é não estacionária ao nível de significância de 5%, pois o p-valor de 0.5638 não rejeita a hipótese nula, de que a série possui raiz unitária.

Figura 1 – Série histórica da cotação da soja na bolsa de Chicago.

Fonte: Elaborado pelos autores (2018).

No entanto, aplicando à diferença em primeira ordem, a série se apresenta estacionária, rejeitando a hipótese nula ao nível de 1% de significância. A figura 2 ilustra o comportamento da série oscilando em torno da média com variância constante.

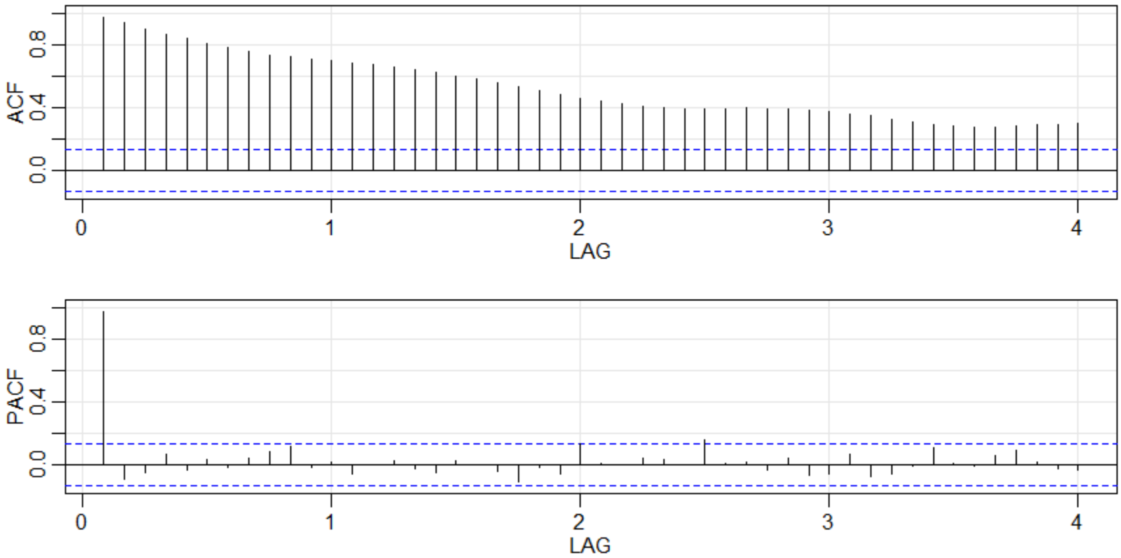
Figura 2 – Série de dados diferenciada em primeira ordem.



Fonte: Elaborado pelos autores (2018).

Analisando o correlograma (figura 3), observa-se que há um decaimento lento na função autocorrelação (ACF), indicando que a série original não é estacionária, conforme verificado anteriormente. E a função autocorrelação parcial (PACF) mostra que existe uma alta correlação com o valor anterior, indicando que o modelo mais apropriado para esses dados seja o ARIMA, com pelo menos uma diferença, e os parâmetros autorregressivos e de média móvel com valores entre zero e um.

Figura 3 – Correlograma de autocorrelação ACF e PACF.



Fonte: Elaborado pelos autores (2018).

A partir da análise do correlograma, procedeu-se à etapa de estimação de modelos, na qual vários modelos ARIMA foram estimados. Nessa etapa todas as combinações dos parâmetros (p,d,q) de 0 a 5 forma estimados. A seleção dos melhores modelos foi baseada no menor valor AIC e BIC. A tabela 1 apresenta os três modelos selecionados e seus respectivos valores de AIC e BIC.

Tabela 1 – Valores AIC e BIC.

Modelo	**AIC**	**BIC**
ARIMA (1,2,1)	2529.442	2539.623
ARIMA (0,2,1)	2530.105	2536.892
ARIMA (0,2,2)	2536.892	2539.925

Fonte: Elaborado pelos autores (2018).

Após selecionar os modelos, foi verificado se os coeficientes de cada um deles são estatisticamente significantes e se os resíduos apresentam distribuição normal; ou seja, se tratam de ruído branco. Para verificar a normalidade dos resíduos foi utilizado o teste Shapiro-Wilk, por meio do qual se constatou que os três modelos rejeitam a hipótese nula, ou seja, os resíduos da amostra seguem a distribuição normal (tabela 2). Ao verificar os

coeficientes dos modelos, apenas os coeficientes Ar(1) e Ma(2) dos modelos ARIMA(1,2,1) e ARIMA(0,2,2), respectivamente, não são estatisticamente significantes, conforme apresenta a tabela 3. No entanto, os valores são muito próximos ao nível de significância de 10%. Por esse motivo, na presente pesquisa optou-se por não remover esses coeficientes.

Tabela 2 – Teste de normalidade Shapiro-Wilk.

Modelo	**Valor**	**p-valor**
ARIMA (1,2,1)	0.95472	1.835 e-6
ARIMA (0,2,1)	0.95299	1.202 e-6
ARIMA (0,2,2)	0.95456	1.764 e-6

Fonte: Elaborado pelos autores (2018).

Quadro 2 – Teste de coeficientes do modelo.

Modelo	**Variável**	**Estimate**	**Std. Error**	**p-valor**
ARIMA (1,2,1)	Ar (1)	0.110886	0.067783	0.1019
	Ma (1)	-1.000000	0.016347	2 e-16 ***
ARIMA (0,2,1)	Ma (1)	0.999998	0.018512	2.2 e-16 ***
ARIMA (0,2,2)	Ma (1)	-0.902451	0.064708	2 e-16 ***
	Ma (2)	-0.097548	0.062535	0.1188

Fonte: Elaborado pelos autores (2018).

Antes de selecionar o melhor modelo para realizar as previsões foi analisado a autocorrelação dos resíduos em vários níveis de defasagem por meio do teste Ljung-Box, e o nível de precisão dos modelos por meio de indicadores de erro, conforme apresenta a tabela 3.

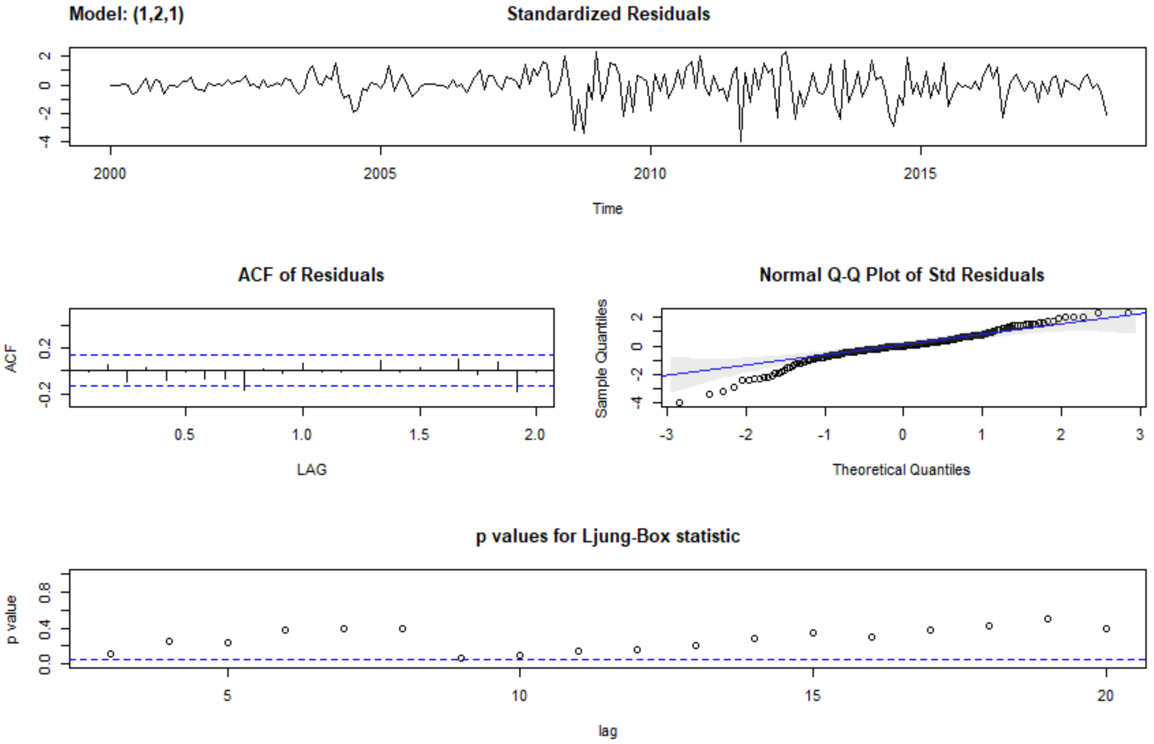
Tabela 3 – Estatísticas de erro.

Modelo	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**
ARIMA (1,2,1)	-1.795	73.690	52.170	-0.325	5.389
ARIMA (0,2,1)	-1.898	74.100	52.094	-0.3594	5.392
ARIMA (0,2,2)	-1.812	73.733	52.172	-0.332	5.391

Fonte: Elaborado pelos autores (2018).

O modelo ARIMA(1,2,1) apresentou o menor valor de erro em todos os indicadores, exceto no indicado de (MAE), no qual o modelo ARIMA(0,2,1) teve o melhor resultado, conforme tabela 4. Portanto, o ARIMA(1,2,1) apresentou-se como o modelo mais preciso entre eles: demonstrando um baixo percentual de erro em suas previsões, como podemos verificar no indicador MAPE, que representa a média percentual da divisão entre o erro de previsão e o valor real; expressando a acurácia do erro em percentagem, que neste caso apresenta 5,389%; ou seja, o modelo possui uma assertividade média de 94,611%, uma taxa de acerto expressiva para cenário analisado. Além disso, os modelos foram submetidos ao teste Ljung-Box, e de acordo com os resultados o modelo ARIMA(1,2,1) apresentou os melhores resultados (figura 4). Desta forma pode-se concluir que o modelo ARIMA (1,2,1) apresenta o melhor ajuste aos dados. A figura 5 ilustra o ajuste do modelo à série de dados original.

Figura 4 – Teste Ljung-Box.



Fonte: Elaborado pelos autores (2018).

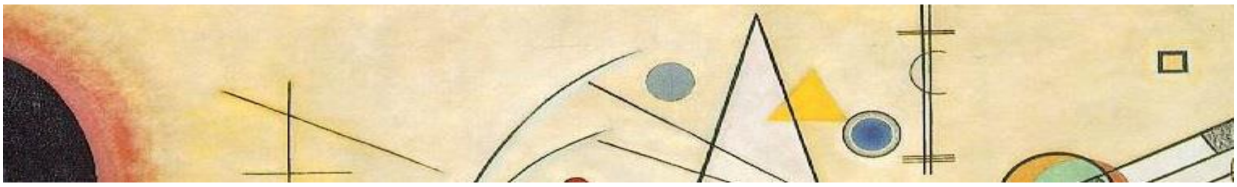
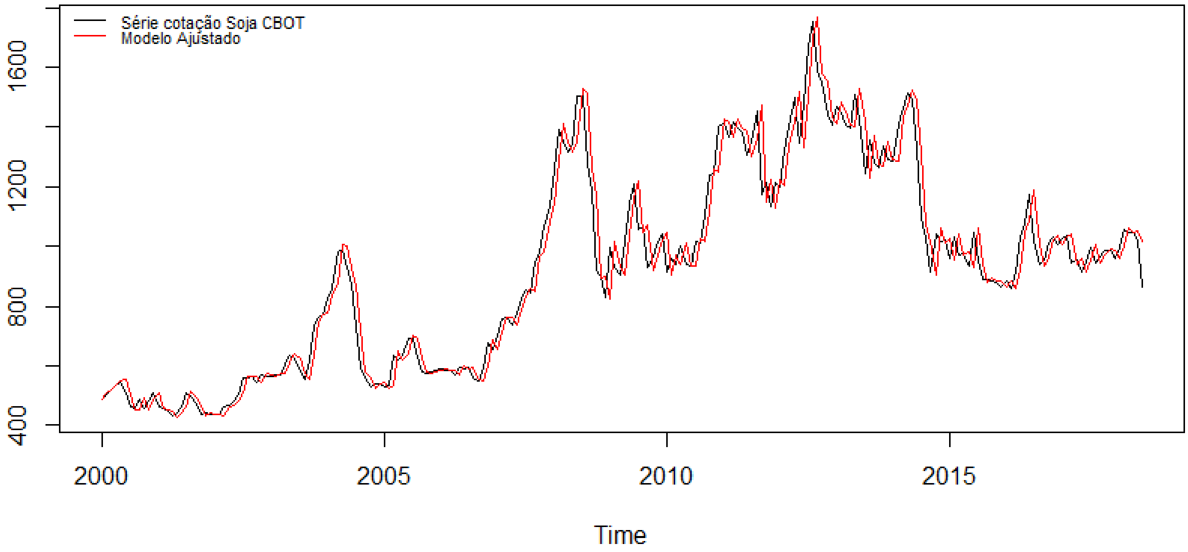


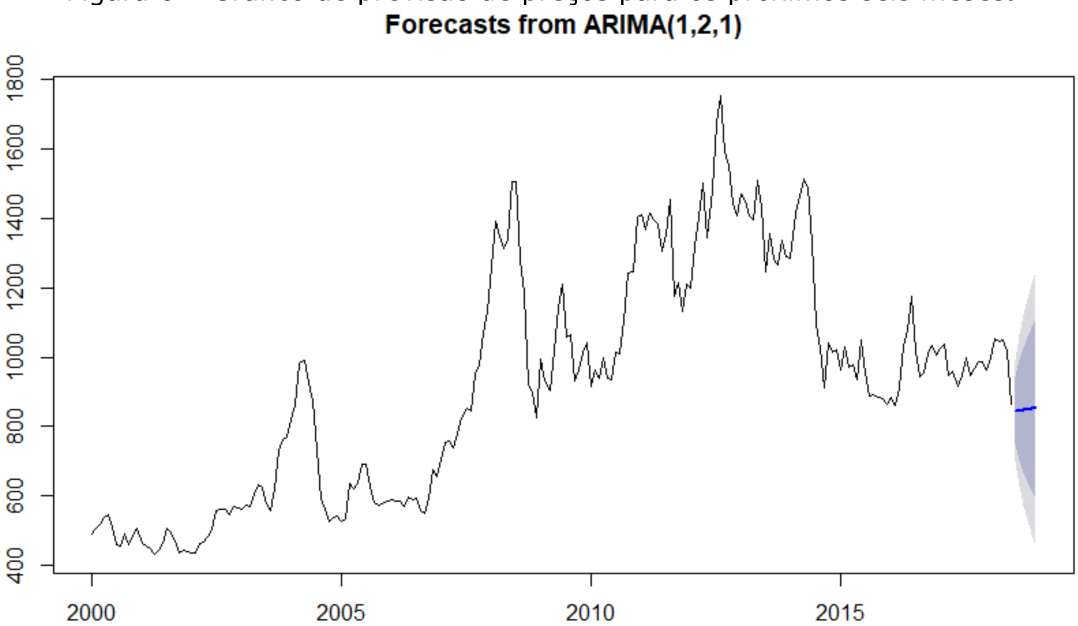
Figura 5 – Ajuste do modelo a série de dados original.



Fonte: Elaborado pelos autores (2018).

A última etapa consiste na previsão, na qual foram estimados os valores futuros da soja para seis ciclos posteriores; ou seja, foi realizada a previsão do preço para os próximos seis meses. A figura 6 ilustra essa previsão e a tabela 4 apresenta os valores previstos e os intervalos ao nível de 80% e 90% de confiança.

Figura 6 – Gráfico de previsão de preços para os próximos seis meses.



Fonte: Elaborado pelos autores (2018).

Tabela 4 – Valores previstos e os intervalos de confiança.

Mês	Estimativa	limite inf. (80)	limite sup. (80)	limite inf. (90)	limite sup. (90)
Jul	846.79	751.71	941.87	701.38	992.20
Ago	846.45	704.02	988.88	628.62	1064.28
Set	847.83	669.27	1026.39	574.75	1120.92
Out	849.41	640.57	1058.25	530.02	1168.80
Nov	851.01	615.54	1086.48	490.89	1211.13
Dez	852.61	593.03	1112.19	455.61	1249.61

Fonte: Elaborado pelos autores (2018).

Analisando os valores previstos, observa-se uma leve tendência de alta do preço para os próximos seis meses. As informações de limite inferior e superior para intervalos de confiança ao nível 80% e 90% são de grande utilidade para análise de riscos, pois garantem que em condições normais de mercado os valores previstos estarão dentro deste intervalo em média 80 e 90 por cento das vezes, de acordo com cada nível de confiança.

De posse dessas informações os participantes do mercado agrícola, (produtores, compradores, vendedores) podem basear suas decisões e definir estratégias para o cenário previsto. Por exemplo, quando a expectativa de preço é negativa, o produtor pode decidir em diminuir a área plantada ou vender a safra antecipada no mercado de futuros. Contudo, quando a expectativa de preço for positiva ele pode aumentar área de cultivo, optar por armazenar a colheita, entre outras ações. Isso vale também para os compradores; empresas que utilizam a soja como matéria prima podem definir estratégias de compras, antecipando ou adiando a compra de acordo com as expectativas futuras.

No entanto, essa modelagem apresenta uma grande amplitude nos intervalos de confiança previstos, como por exemplo, o intervalo para o mês de dezembro ao nível de 90% de confiança vai de 455,61 a 1249,61 reais, uma amplitude de 794 reais, o que dificulta a tomada de decisão a médio prazo (6 meses). Estivas nesse prazo são necessárias para o contexto do agronegócio. A curto prazo a amplitude nos intervalos de confiança é de 190 e 291 reais ao nível de 80% e 90% de confiança, respectivamente, uma amplitude muito menor se comparado as estimativas a médio prazo. Diante disso, pode-se dizer que a modelagem de série temporal ARIMA para a soja é mais indicada para subsidiar decisões a curto prazo; no entanto, as estimativas de médio prazo devem ser consideradas para eventuais análises de riscos e as demais decisões que envolvam riscos e incertezas.

5. Conclusão

O presente trabalho se propôs analisar a eficácia dos modelos univariados em prever valores futuros para a soja a partir de dados históricos, com a finalidade de auxiliar a tomada de decisão. A modelagem ARIMA demonstrou ser uma alternativa eficiente para previsão de valores futuros. O modelo ARIMA (1,2,1) apresentou uma taxa de erro de apenas 5,39%, de acordo com o indicador de erro percentual absoluto médio (MAPE), um percentual de erro baixo considerando o mercado agrícola.

Os resultados demonstram que esse tipo de modelagem matemática se constitui em uma ferramenta relevante para tomada de decisão para os agentes que atuam neste mercado, na medida em que apresentam a tendência dos preços para um horizonte de curto e médio prazo, possibilitando antever o comportamento desta variável. É importante frisar que as previsões não constituem um fim em si, mas apenas representam um meio de fornecer informações relevantes para a tomada de decisão, possibilitando que os gestores elaborem ações para maximizar seus resultados, minimizando os riscos.

No entanto, o intervalo de confiança das previsões geradas ainda é relativamente grande. A modelagem de série temporal ARIMA é mais indicada para subsidiar decisões a curto prazo; embora suas estimativas a médio prazo devam ser consideradas pelos gestores no planejamento a médio e longo prazo. Neste sentido, sugere-se o desenvolvimento de pesquisas visando otimizar a precisão e consequentemente diminuir

amplitude dos intervalos de confiança das estimativas previstas. Uma alternativa é o desenvolvimento de um modelo híbrido, combinando a modelagem ARIMA a outras técnicas de modelagem de séries temporais, entre elas a modelagem por meio de redes neurais.

Referências

BOX, G. E. P.; JENKINS, G. M. **Time series analysis**: forecasting and control. New Jersey: Prentice Hall, 1976.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**, 2nd edition. New York: Springer, 2002.

BURNHAM, K. P.; ANDERSON, D. R. **Model Selection and Multimodel Inference**. 2nd ed. New York: Springer, 2002.

CAMPOS, K. C. **Análise da volatilidade de preços de produtos agropecuários no Brasil**. Revista de Economia e Agronegócio, n. 5, 2007, p. 303-328.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Indicadores da agropecuária**. Brasília: CONAB, 2009. Disponível em: <www.conab.gov.br/info-agro>. Acesso em: 20 jul. 2018.

ENDERS, W. **Applied econometric time series**. New York: John Wiley, 1995.

ENDERS, W. **Rats handbook for econometric time series**. New York: John Wiley, 1996.

EHLERS, R. S. **Análise de Séries Temporais**. Departamento de Estatística, UFPR. Disponível em: <www.est.ufpr.br/~ ehlers/notas> Acesso em: 20 jul. 2018.

FAVA, V. L. **Manual de econometria**. In: VASCONCELOS, M. A. S.; ALVES, D. São Paulo: Editora Atlas, 2000.

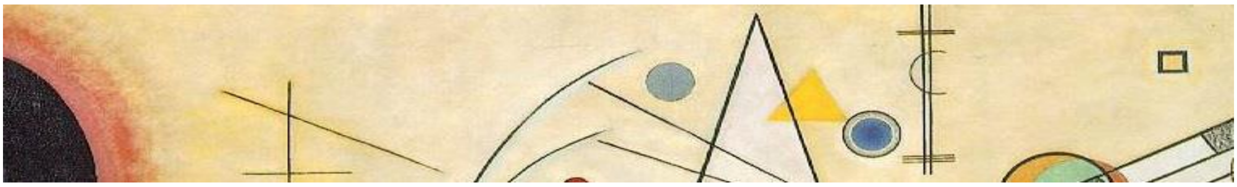
GUJARATI, Damodar N. **Econometria básica**. 4 ed. Rio de Janeiro: Elsevier, 2006.

GRANGER, C.W.J. NEWBOLD, P. **Forecasting Economic Time Series**. Second Edition. Orlando: Academic Press, 1986.

MAKRIDAKIS, S.; HIBON, M. **The M3-Competition:** results, conclusions and implications. International Journal of Forecasting, 2000.

MARGARIDO, M. A. **Transmissão de preços internacionais de suco de laranja para preços ao nível de produtor de laranja no Estado de São Paulo**. São Paulo: IEA. 1998. 127p. (Coleção Estudos Agrícolas, 6/98).

MARTINS, T. M.; MARTINELLI, D. P. **Ciclos e previsão cíclica dos preços de commodities:** um modelo indicador antedecedente para commodity



açúcar. Revista de Administração, Contabilidade e Economia, n. 2, 2010, p. 2-12.

MONTGOMERY, D. C.; JOHNSON, L. A.; GARDINER, J. S. **Forecasting and Time Series Analysis**. 2 ed., New York: McGraw-Hill, 1990.

MORETTIN, P. A.; TOLOI, C. M. **Séries Temporais**. 2. ed. São Paulo: Atual, 2006.

MORETTIN, P. A. **Econometria financeira**: um curso em Séries Temporais financeiras. São Paulo: Departamento de Estatística-Instituto de Matemática e Estatística-USP, 2004.

MORETTIN, P. A., TOLOI, C. M. de C. **Modelos para Previsão de Séries Temporais**. In: 13° Colóquio Brasileiro de Matemática, Rio de Janeiro, 1981.

PHILLIPS, P. C. B.; PERRON P. **Testing for Unit Roots in Time Series Regression**. Biometrika, n.75, 1988.

RIBEIRO, C. O; SOSNOSKI, A. A. K.; OLIVEIRA, S. M. **Um modelo hierárquico para previsão de preços de commodities agrícolas**. Revista Produção On-line, n.10, 2010, p. 719-733.

SAID, S. E.; DICKEY D. **Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order**. Biometrika, n.71, 1984. p. 599-607.

SCHWAGER, J. D. **Fundamental analysis**. New York: John Wiley & Sons,1995.

SIQUEIRA, M. L. **Melhorando a Previsão da Arrecadação Tributária Federal através da Atualização de Modelos de Séries Temporais.** ESAF, 2002 (Monografia – VII Prêmio do Tesouro Nacional).

STOCK, J. H.; WATSON, M. W. **Econometria.** São Paulo: Addison Wesley, 2004.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. **FAS Databases**. Disponível em: < https://www.fas.usda.gov/data >. Acesso em: 01 ago. 2018.