

UTILIZANDO A METODOLOGIA BOX & JENKINS PARA PREVER A INFLAÇÃO BRASILEIRA

Janaina Fernanda Battahin¹

RESUMO: Este artigo tem como objetivo apresentar o comportamento futuro da série de Índice Nacional de Preços ao Consumidor Amplo (IPCA), a medida oficial de inflação, para o período que compreende julho de 2019 a junho de 2020. A previsão da inflação é importante para auxiliar as autoridades monetárias a definirem instrumentos de política, já que no caso brasileiro o Banco Central utiliza um sistema de metas de inflação; e também para orientar investimentos no mercado financeiro. Para isso, utilizou-se a metodologia Box-Jenkins com auxílio do software *Gretl*. Vale ressaltar, que o trabalho fez uso do método X-12 ARIMA para atenuar a sazonalidade da série em questão. A obtenção do modelo mais adequado foi baseada na análise de gráficos e em testes estatísticos próprios da metodologia, o qual foi definido o ARMA (1, 12).

Palavras-chave: Índice Geral de Preços; Econometria; Séries Temporais; Metodologia Box & Jenkins; Previsões.

INTRODUÇÃO

As séries temporais podem ser definidas como “um conjunto de valores coletados em períodos regulares ou não de tempo” (MAIA, 2017, p. 291), podendo ser utilizadas para elaborar modelos de causa e efeito e modelos univariados ou multivariados de previsão. O método mais simples de realizar uma previsão de determinada variável é utilizando o último valor da série. Uma previsão só será útil se a variável apresentar o comportamento de um passeio aleatório, ou seja, se seguir o modelo $Y_t = Y_{t-1} + \varepsilon_t$, onde o valor da variável no período t se baseia no valor da variável no período $t - 1$ mais o componente de erro (SARTORIS, 2013).

A previsão de algumas variáveis econômicas como PIB, inflação, taxa de juros, consiste em uma das partes mais importantes da análise econométrica. Um método de previsão que se tornou muito popular é o ARIMA (Processo Autoregressivo Integrado de Média Móvel), proposto por Box e Jenkins (1970) no trabalho intitulado *Times series analysis: forecasting and control*, e por isso, popularmente conhecido como método Box & Jenkins. Essa metodologia enfatiza a

¹ Universidade Estadual Paulista (Unesp), Araraquara – SP – Brasil. Doutoranda do Programa de Pós-Graduação em Economia. janaina.battahin@unesp.br.

análise probabilística ou estocástica e permite que Y_t seja explicado por valores passado ou defasados de Y e dos termos de erro estocástico (GUJARATI; PORTER, 2011).

Barbosa (2019) realizou um estudo prevendo a inflação brasileira através da metodologia de Box & Jenkins e de um modelo baseado na Teoria Econômica, para fins de comparação, utilizando dados entre janeiro de 2002 e outubro de 2018. Santos, Ferreira e Castelar (2016) fizeram uma comparação das previsões da inflação acumulada em 12 meses fornecidas pela pesquisa Focus do Banco Central do Brasil com as previsões obtidas através dos Modelos Autorregressivo Integrado de Média Móvel (ARIMA) e Vetor Autorregressivo (VAR). Coelho Jr. (2004) utilizou a metodologia Box & Jenkins para prever os indicadores econômicos inflação, taxa de juros, taxa de câmbio e Ibovespa, concluindo que apesar de os valores realizados e previstos apresentarem divergências, os modelos estimados previram com moderação os movimentos do mercado.

É muito comum que, principalmente no início de cada ano, as expectativas referentes às tendências econômicas sejam especuladas pela imprensa e grandes consultorias. A previsão do comportamento de variáveis econômicas também tem alto grau de influência no mercado financeiro e serve como norte para os investimentos nesse setor. Mas não são somente esses setores que se importam com o comportamento das variáveis econômicas. Os órgãos públicos oficiais, como por exemplo o Banco do Brasil, realizam suas especulações através do Boletim Focus (VINHAES, 2018).

Nesse sentido, esse artigo tem como objetivo apresentar o comportamento futuro da série de Índice Nacional de Preços ao Consumidor Amplo (IPCA), a medida oficial de inflação, para o período que compreende julho de 2019 a junho de 2020. Calculado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o IPCA é o parâmetro para a política de metas de inflação e abrange as famílias com rendimentos de 1 a 40 salários-mínimos, independente da fonte, residentes nas áreas urbanas das regiões metropolitanas de Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre, além do Distrito Federal e dos municípios de Goiânia e Campo Grande (IBGE, 2022).

Para prever a inflação no período mencionado, utilizou-se a metodologia Box-Jenkins com auxílio do software *Gretl*. A obtenção do modelo mais adequado foi

baseada na análise de gráficos e em testes estatísticos próprios da metodologia. Vale ressaltar, que a série em questão apresentou sazonalidade, tornando necessária a utilização do método X-12-ARIMA para atenuá-la.

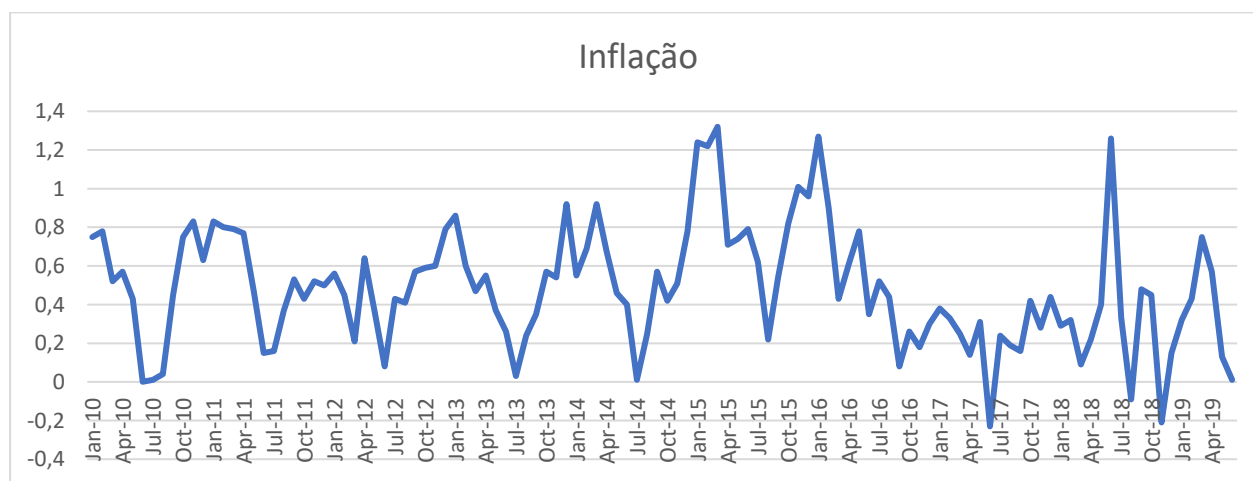
Além desta introdução, o trabalho conta com outras quatro seções. A seção 2 apresenta os dados e a metodologia de previsão utilizada. A seção 3 demonstra os resultados. Por fim, a seção 4 apresenta uma breve discussão e conclusão.

DADOS E METODOLOGIA

DADOS

Para realizar a previsão da inflação, o trabalho utilizou dados obtidos no IPEADATA: obteve-se a série de inflação do Brasil que é medida pelo IPCA entre janeiro 2010 e junho 2019, conforme mostra o gráfico abaixo

Gráfico 1 – IPCA Variação Mensal (%)



Fonte: Elaboração própria com base em dados do IPEADATA (2022).

A partir dos dados do IPCA e utilizando o software econométrico *Gretl* realizou-se uma previsão da inflação para os meses de julho de 2019 a julho de 2020. Para a previsão dessa série temporal utilizou-se a metodologia *Box-Jenkins* que se baseia nos dados passados para estimar previsões. Adiante, serão apresentados alguns conceitos básicos em séries temporais, bem como o método de modelagem utilizado para a realização das previsões.

METODOLOGIA

Estacionariedade

Toda série temporal é gerada por um processo estocástico ou processo aleatório. Em um processo estocástico cada valor de Y de uma série é “extraído aleatoriamente de um conjunto finito ou infinito de valores, com uma determinada distribuição de probabilidade”, ou seja, é “o conjunto de possíveis trajetórias para Y_t ” (MAIA, 2017, p. 292). Os processos estocásticos podem ser definidos como “uma coleção de variáveis aleatórias ordenadas no tempo” (GUJARATI; PORTER, 2011, p. 734). A maioria dos dados econômicos são coletados em “pontos discretos no tempo” e se considerarmos Y como representante da inflação, para os dados teremos: $Y_1, Y_2, Y_3, \dots, Y_n$, em que n denota a última observação, sendo cada um desses Y uma variável aleatória. Assim, utilizamos, em séries temporais, “a realização para extrair inferências sobre o processo estocástico subjacente” (GUJARATI; PORTER, 2011, p. 734).

Para realizar uma análise de séries temporais é importante compreender alguns conceitos e propriedades importantes. Antes de se iniciar a aplicação da metodologia Box-Jenkins deve-se avaliar se a série analisada é estacionária, propriedade desejável para uma série temporal. Dessa forma, é necessário que a série apresente comportamento constante no tempo, já que é muito difícil prever os valores futuros se baseando em valores passados de uma série que mude muito com o tempo. Podemos nos restringir ao conceito de estacionariedade fraca que consiste em um valor de Y que converge para uma média constante, uma variabilidade de Y não muda com o tempo e uma covariância entre valor presente e o valor defasado se mantém a mesma (MAIA, 2017).

Assim, nos modelos de previsão é muito importante que a série seja estacionária, pois:

- Supõe que a relação entre Y_t e Y_{t-1} seja o mesmo em todos os períodos t ;
- Mostra que a série converge para uma média histórica;
- Auxilia na compreensão das relações econômicas;

Uma característica da série estacionária é que a mesma, quando representada por um modelo de regressão de Y_t em função de seus valores defasados, terá a soma de coeficientes referentes as variáveis defasadas menor que 1 (MAIA, 2017). Assim, considerando o modelo de passeio aleatório como

$$Y_t = \rho Y_{t-1} + e_t \quad (1)$$

podemos afirmar que se $\rho < 1$, ou seja, o valor absoluto de ρ for menor do que 1 será possível apontar que a série temporal Y_t é estacionária. No entanto, se $\rho = 1$ a equação se tornará um modelo de passeio aleatório e teremos um problema chamado problema de raiz unitária, ou seja, uma situação de não estacionariedade. Dessa maneira, é importante descobrir se a série temporal trabalhada possui ou não raiz unitária (GUJARATI; PORTER, 2011).

O erro, representado na equação (1) como e_t , é conhecido também como ruído branco e representa uma série com média zero, variância constante e não autocorrelacionada (MAIA, 2017), e pode ser representado pelas equações abaixo

$$E(e_t) = 0 \quad (2)$$

$$Var(e_t) = \gamma_0 = Cov(e_t, e_{t+k}) = \gamma_k = 0 \quad (3)$$

É importante apresentar algumas denominações próprias dos processos estocásticos que servirão para caracterizar “sua natureza estacionária e os componentes que fazem parte de sua especificação” (MAIA, 2017, p. 301). A primeira denominação é o caso mais simples de processo estacionário chamado ruído branco, representado pela equação $y_t = e_t$ e que possui média 0, variância constante e é não correlacionado serialmente. A segunda denominação é a tendência determinística, representada pela equação $y_t = \alpha + \beta t + e_t$ que apresenta média não constante, mas prevista com exatidão quando conhecemos o valor de t . Apresenta-se como uma série com comportamento estacionário pós remoção de tendência, ou seja, apresenta comportamento estacionário em cima de uma tendência no tempo. Outro caso é a tendência determinística com componente autorregressivo estacionário, expressa por $y_t = \alpha + \beta t + \rho Y_{t-1} + e_t$, no qual “choques são absorvidos com o tempo e o processo tende a ser estacionário em torno de uma tendência determinística β_t ” (MAIA, 2017, p. 301-302).

No entanto, o estudo de séries temporais frequentemente apresentará séries não estacionárias, ou seja, séries temporais que seguem um passeio aleatório sem deslocamento, com deslocamento ou com deslocamento e tendência determinística.

O passeio aleatório sem deslocamento pode ser expressado pela equação $Y_t = Y_{t-1} + e_t$ que evidencia quando os choques do presente são absorvidos nos períodos posteriores fazendo com que a série tenha um comportamento imprevisível e com uma variância que “explode” ao longo do tempo. Podemos então transformar a variável Y_t através da primeira diferença de Y ($\Delta Y_t = Y_t - Y_{t-1} = e_t$) e chegar a uma série estacionária. Já o passeio aleatório com deslocamento, $Y_t = \alpha + Y_{t-1} + e_t$, mostra quando os choques do presente são absorvidos nos períodos posteriores e, além disso, apresenta tendência constante de variação em cada período., ou seja, média e variância variam ao longo do tempo. Podemos citar ainda o passeio aleatório com deslocamento e tendência determinística, $Y_t = \alpha + \beta t + Y_{t-1} + e_t$, que é aquele que apresenta comportamento errático, imprevisível e em torno de uma tendência determinística βt , podendo ser transformado em processo estacionário através de transformação algébrica (MAIA, 2017).

Para verificar se uma dada série temporal é estacionário podemos fazer uso de duas ferramentas importantes: a análise gráfica da função de autocorrelação, que fornece uma pista inicial sobre a natureza da série temporal, de modo a ser um ponto de partida intuitivo para que prossigamos com testes mais formais; e o Teste de Raiz Unitária.

A função de autocorrelação “descreve os padrões de relacionamento entre valores presentes e defasados de Y , permitindo identificar a presença de estacionariedade em um processo estocástico” (MAIA, 2017, p. 301). Podemos verificar em que medida valores de um determinado período influenciam ou são influenciados por valores defasados quando observamos as correlações entre valores de um período base (t) e valores defasados ($t - k$) (MAIA, 2017). A função de correlação (FAC) pode ser representada pela seguinte equação

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{DP(Y_t)DP(\rho Y_{t-k})} = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} = \frac{\gamma_k}{\gamma_0} \quad (4)$$

O gráfico da função de autocorrelação em função de k se chama correlograma. Em processos estacionários a correlação entre os valores presentes e defasados tendem a ir desaparecendo conforme a defasagem aumenta (MAIA, 2017). Caso, tenhamos uma série não estacionária os coeficientes de autocorrelação nas diversas defasagens serão altos e vão diminuindo muito lentamente em direção a zero conforme as defasagens vão aumentando (GUJARATI; PORTER, 2011).

Posto isso, cabe apresentar agora o modo mais formal para encontrar evidências da presença ou não de estacionariedade em uma determinada série temporal. O chamado Teste de Raiz Unitária verifica se a “inércia ρ associada ao componente autorregressivo de 1ª ordem Y_{t-1} possui raiz unitária” (MAIA, 2017, p.317). Esse teste pode ser realizado através do teste de Dickey-Fuller e teste de Dickey-Fuller Aumentado (MAIA, 2017).

Segundo Maia (2017), o teste de Dickey-Fuller testa a hipótese nula da existência de que a série não é estacionária. Caso esta hipótese não seja rejeitada, a série possuirá raiz unitária, portanto, não será estacionária. Para evitar o problema da autocorrelação dos resíduos, recomenda-se a utilização do teste ADF (Dickey-Fuller Aumentado), que contempla a equação das defasagens para eliminação do problema de autocorrelação dos resíduos.

Para o Teste de Dickey-Fuller, deve-se conhecer a especificação do processo estocástico gerador da série Y , o que equivale a compreender qual dos modelos abaixo define o processo estocástico que se analisa

- i. Sem constante $Y_t = \rho Y_{t-1} + e_t$
- ii. Com constante $Y_t = \delta + \rho Y_{t-1} + e_t$
- iii. Com constante e tendência determinística $Y_t = \delta + \beta t + \rho Y_{t-1} + e_t$

O teste consiste em transformar as equações, subtraindo o termo Y_{t-1} nos dois lados da equação:

- i. $Y_t = \rho Y_{t-1} + e_t \Rightarrow Y_{t-1} - Y_t = (\rho - 1)Y_{t-1} + e_t \Rightarrow \Delta Y_t = \delta Y_{t-1} + e_t$
- ii. $Y_t = \delta + \rho Y_{t-1} + e_t \Rightarrow \Delta Y_t = \delta + (\rho - 1)Y_{t-1} + e_t \Rightarrow \Delta Y_t = \alpha + \delta Y_{t-1} + e_t$
- iii. $Y_t = \alpha + \beta t + \rho Y_{t-1} + e_t \Rightarrow \Delta Y_t = \alpha + \beta t + (\rho - 1)Y_{t-1} + e_t \Rightarrow \Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + e_t$

A partir disso, testa-se a hipótese nula de que $\rho = 1$, o que equivale a testar a hipótese de que $\delta = 0$, ou seja

$$\begin{aligned} H_0: \delta &= 0 (\text{n\~{a}o estacion\~{a}rio}) \\ H_1: \delta &< 0 (\text{estacion\~{a}rio}) \end{aligned} \quad (5)$$

O problema desse teste é que não considera situações em que os erros sejam autocorrelacionados, o que torna $\hat{\delta}$ viesado. Dessa forma, o Teste de Dickey-Fuller Aumentado apresenta melhor resultado ao considerar as defasagens da variável dependente de ΔY_t entre os regressores para controlar a autocorrelação nos erros (MAIA, 2017). Teremos as seguintes equações transformadas

- i. $\Delta Y_t = \delta Y_{t-1} + \sum_{j=1}^p \Delta Y_{t-j} + e_t$
- ii. $\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{j=1}^p \Delta Y_{t-j} + e_t$
- iii. $\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{j=1}^p \Delta Y_{t-j} + e_t$

As hipóteses serão as mesmas do teste anterior. É importante incluir quantas defasagens (p) forem necessárias para que o erro não apresente correlação serial. Se após esses testes descobrirmos que a série temporal é não estacionária devemos transformá-la em estacionária. Existem dois processos possíveis: processo de diferença estacionária e processo de tendência estacionária. Se uma série temporal tem uma raiz unitária sabemos que suas primeiras diferenças são estacionárias e a solução será tomar suas primeiras diferenças. Já o processo estacionário em tendência é estacionário em torno da linha de tendência e, portanto, para tornar a série estacionária basta regressá-la no tempo (GUJARATI; PORTER, 2011).

A presença de sazonalidade e o Método X-12 ARIMA

A análise da representação gráfica de uma série temporal ao longo do tempo pode revelar padrões de comportamento como tendências de crescimento ou

decrecimento, padrões cíclicos, movimentos estruturais, etc. Segundo Castro (1995, p.14-17) podemos decompor as séries temporais em quatro componentes:

a) Tendência: é o componente que apresenta “variações suaves causadas por fatores que ocorrem sistematicamente através de toda a seção estudada”;

b) Ciclicidade: variações que ocorrem em períodos maiores que um ano, podendo ou não ser periódicos;

c) Sazonalidade: “movimentos oscilatórias, com períodos geralmente menores que um ano”;

d) Erro aleatório (ruído branco): a componente que resulta quando se retira as demais e que é causado por um ruído branco;

Este trabalho se limitará a discutir um componente específico: a sazonalidade; já que a série do IPCA utilizada contou com a presença da mesma. Nesse sentido, pode-se destacar dois modos de atenuação da sazonalidade: o método de diferenças sazonais e o método X-12 ARIMA.

O primeiro método consiste em efetuar diferenças sucessivas da série original onde a ordem da diferença aplicada será igual ao período sazonal investigado (CASTRO, 1995). Já o segundo método realiza a dessazonalização de séries temporais econômicas por meio da aplicação de filtros lineares sucessivos para estimação das componentes não observáveis da série, como tendência e sazonalidade (ARAÚJO; SIQUEIRA; CARDOSO, 2010).

Segundo Yafee e McGee (2000, p.66)

[...] o programa do método X-12 foi desenvolvido pelo U.S. Census e contém algumas inovações em relação ao método X-11 e sua posterior atualização em 1988, o X-11 ARIMA, desenvolvido por E.B. Dagum *et al.* Dagum introduziu o X-11 ARIMA para reduzir a viés de previsão para o final das séries. O novo X-12 contém uma nova sistemática, focada no diagnóstico para avaliar a qualidade de ajustamentos sazonais.

O X-12 Arima realiza a decomposição da série nos componentes ciclo ($C|t$), tendência ($T|t$), sazonalidade ($S|t$) e componente aleatório ($I|t$), que algebricamente é

$$X_t = C_t * T_t * S_t * I_t \quad (6)$$

Assim, ao dividir a série X_t pelo seu componente sazonal S_t o resultado será uma série sazonalmente ajustada. No caso deste trabalho, o software econométrico para a utilização desse método será *Gnu Regression, Econometrics and Time-series Library (Gretl)*.

O método Box & Jenkins

Quando Box e Jenkins publicaram *Times series analysis: forecasting and control* possibilitaram o surgimento de um novo método de previsão que ficou popularmente conhecido como Box-Jenkins. Esse método, intitulado tecnicamente como ARIMA, enfatiza a “análise probabilística, ou estocástica, das propriedades da própria série temporal econômica sob a filosofia *deixe os dados falarem por si mesmos*” (GUJARATI; PORTER, 2011, p. 768). Esse modelo de série temporal possibilita que Y_t seja explicado por valores passados ou defasado de Y e dos termos de erro estocástico (GUJARATI; PORTER, 2011). Para compreender o método ARIMA, faz-se necessário compreender previamente o processo autorregressivo (AR), o processo de média móvel (MA) e o processo autorregressivo de médias móveis (ARMA). Vejamos cada um deles.

Modelo Autorregressivo (AR)

Se supormos que o valor previsto de Y_t é dado pelo seu valor defasado Y_{t-1} , o modelo de previsão seria $Y_t = \alpha + \phi Y_{t-1} + e_t$. Esse modelo é chamado modelo autoregressivo de 1ª ordem ou AR (1), pois leva em conta como variável explicativa apenas a primeira defasagem de Y_t (MAIA, 2017). Nesse modelo “o valor previsto de Y no período t é simplesmente alguma proporção mais um choque aleatório ou perturbação no período t ” e os valores de Y “são expressos em torno dos seus valores médios” (GUJARATI; PORTER; 2011, p. 769). Teremos então um processo autorregressivo de ordem p -ésima, AR (p), que pode ser expressado pela equação seguinte

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (7)$$

Modelo de Médias Móveis (MA)

Esse modelo considera a previsão de uma série a partir da média ponderada das “perturbações aleatórias presentes e passadas”. Um processo de “médias móveis de ordem q , ou $MA(q)$, pressupõe que cada observação corrente 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” (MAIA, 2017, p. 340).

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (8)$$

Modelo autorregressivo e de médias móveis (ARMA)

Se Y possuir características do AR e MA, ou seja, se não puder ser modelado somente como de médias móveis ou somente como autorregressivo será considerado ARMA (MAIA, 2017). Esse processo autorregressivo e de médias móveis de ordem (p, q) pode ser representado pela equação abaixo

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (9)$$

Modelo autorregressivo integrado e médias móveis (ARIMA)

Como já vimos, nem todas as séries são estacionárias. No entanto, os modelos AR, MA e ARMA se aplicam somente a séries estacionárias. Também já vimos que podemos transformar uma série não estacionária em uma série estacionária através da diferenciação ($\Delta Y_t = Y_t - Y_{t-1}$). Podemos diferenciar uma série quantas vezes forem necessárias até torná-la estacionária, e a diferença de ordem d de uma determinada série Y_t ($\Delta^d Y_t$) pode ser dada por:

$$\Delta^d Y_t = \Delta^{d-1} Y_t - \Delta^{d-1} Y_{t-1} \quad (10)$$

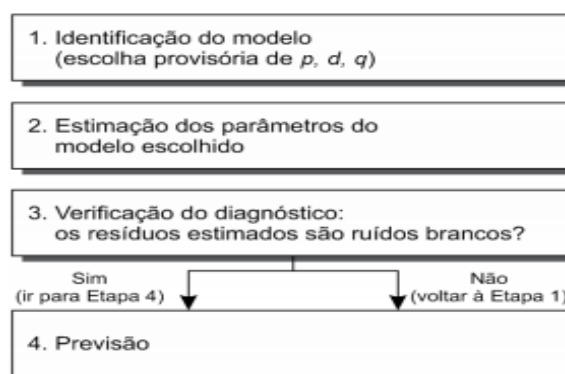
Assim, se for necessário tornar uma série não estacionária em estacionária, podendo diferenciá-la d vezes e aplicar o modelo ARMA (p, q) , chamaremos a série

original de ARIMA (p, d, q), onde p é o número de termos autorregressivos, d é número de vezes que a série será diferenciada para se tornar estacionária e q o número de termos de média móvel. A metodologia Box-Jenkins exige uma série temporal estacionária e que seja estacionária depois de uma ou mais diferenciações (GUJARATI; PORTER, 2011).

O passo a passo da Metodologia Box-Jenkins (BJ)

Para descobrirmos se uma série temporal segue um processo AR puro, MA puro, ARMA ou ARIMA podemos utilizar a metodologia Box-Jenkins que consiste em quatro etapas: identificação, estimação, verificação do diagnóstico e previsão (GUJARATI; PORTER, 2011). Esse processo de análise pode ser explicitado na figura a seguir:

Figura 1 – O método Box-Jenkins



Fonte: Gujarati e Porter (2011, p.772).

- 1) Identificação: nessa etapa serão definidos os valores apropriados de p, d e q através da função de correlação amostral (ACF), da função de correlação parcial (PACF) e dos correlogramas gerados (representam o ACF e o PACF contra a extensão da defasagem)
- 2) Estimação: nesse passo serão estimados os parâmetros dos termos autorregressivos e dos termos de média móvel
- 3) Verificação: esta etapa verificará se o modelo se ajusta aos dados. Na verificação de adequação do modelo dois testes foram realizados: o Teste de autocorrelação residual e o Teste de Box e Pierce. O Teste de

autocorrelação residual consiste em verificar se as autocorrelações dos resíduos estão dentro do intervalo do valor de $\pm \frac{2}{\sqrt{n}}$. É necessário os correlogramas evidenciem que os resíduos estimados são puramente aleatórios, ou seja, que as funções de autocorrelação e autocorrelação parcial mostrem-se bem-comportadas, indicando que o modelo está bem especificado. O teste de Box e Pierce pode confirmar a existência de ruído branco testando duas hipóteses

$$(11) \quad \text{Hipóteses} \quad \left\{ \begin{array}{l} H_0 = \text{os resíduos são ruído branco} \\ H_1 = \text{os resíduos não são ruído branco} \end{array} \right.$$

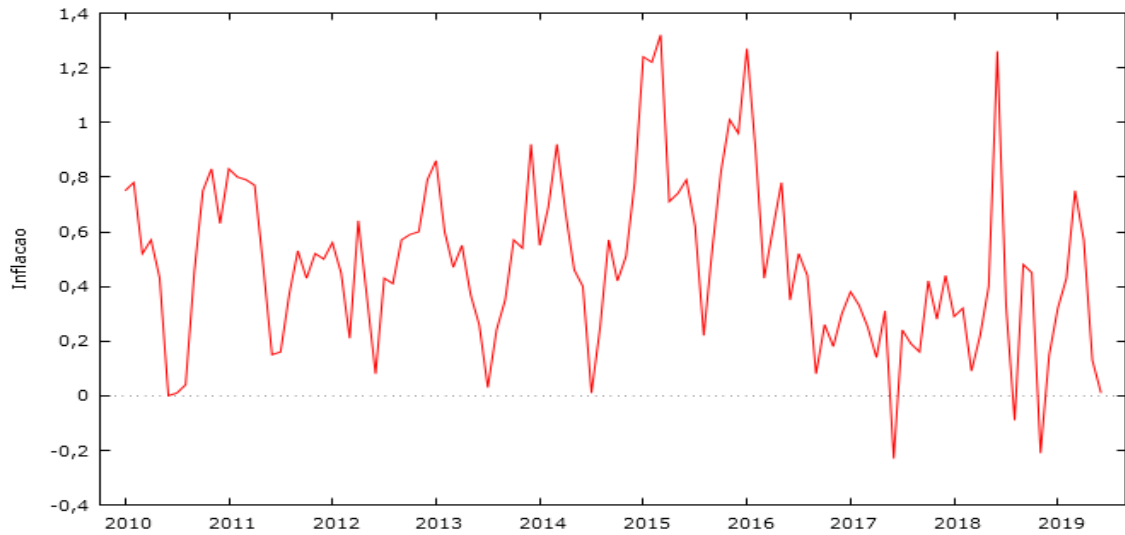
O modelo deve ser ruído branco para então se mostrar adequado. Se $LB < \lambda^2$ não rejeita-se H_0 , ou seja, os resíduos são ruídos branco, sendo o modelo adequado.

- 4) Previsão: após escolher o modelo com melhor ajuste, será realizada uma previsão da série.

RESULTADOS

O primeiro passo da metodologia, aqui utilizada, é a identificação do modelo. Para realizar a análise inicial foi plotado o gráfico da série de inflação conforme apresentado na Figura 1. A análise desse gráfico pode apontou a presença de sazonalidade.

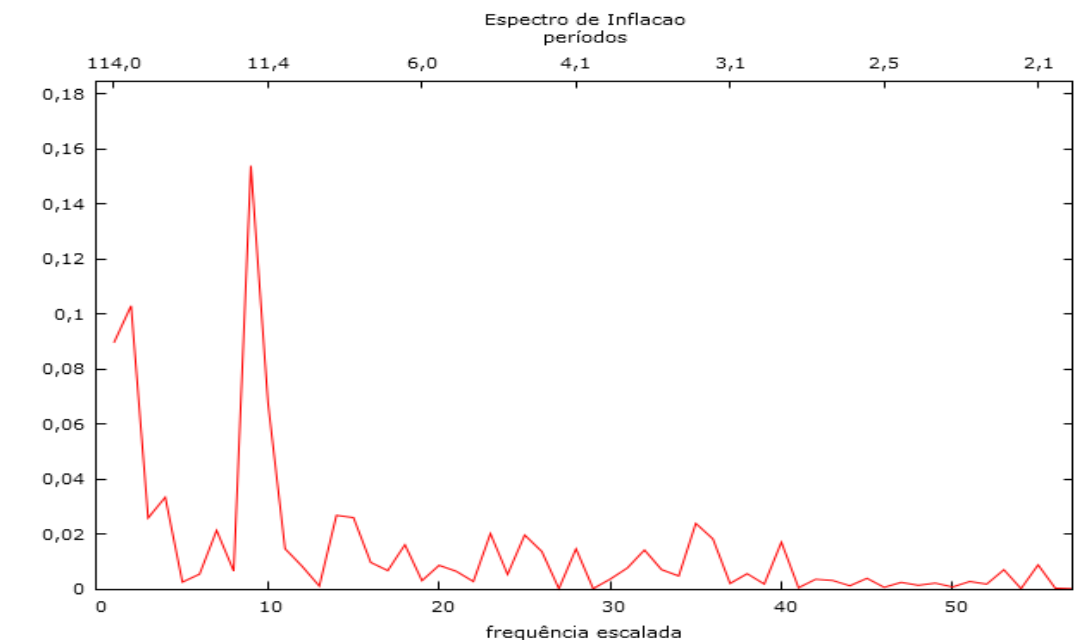
Figura 2 – Série temporal de IPCA de janeiro de 2010 a junho de 2019.



Fonte: Elaboração própria.

O periodograma da série evidenciou essa afirmação ao apresentar uma concentração de frequência (conferindo que existe a sazonalidade) como apontado pela figura 2

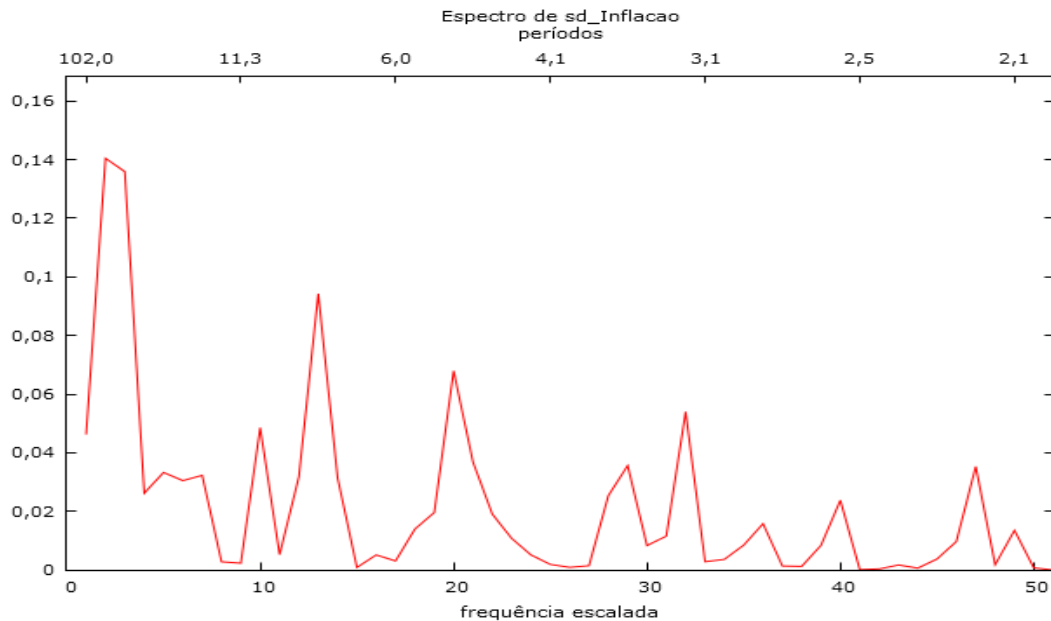
Figura 3 – Periodograma da série temporal de IPCA



Fonte: Elaboração própria.

A sazonalidade foi atenuada acrescentando-se diferenças sazonais na série IPCA. O periodograma, representado pela figura 4, possibilita a verificação desta afirmação.

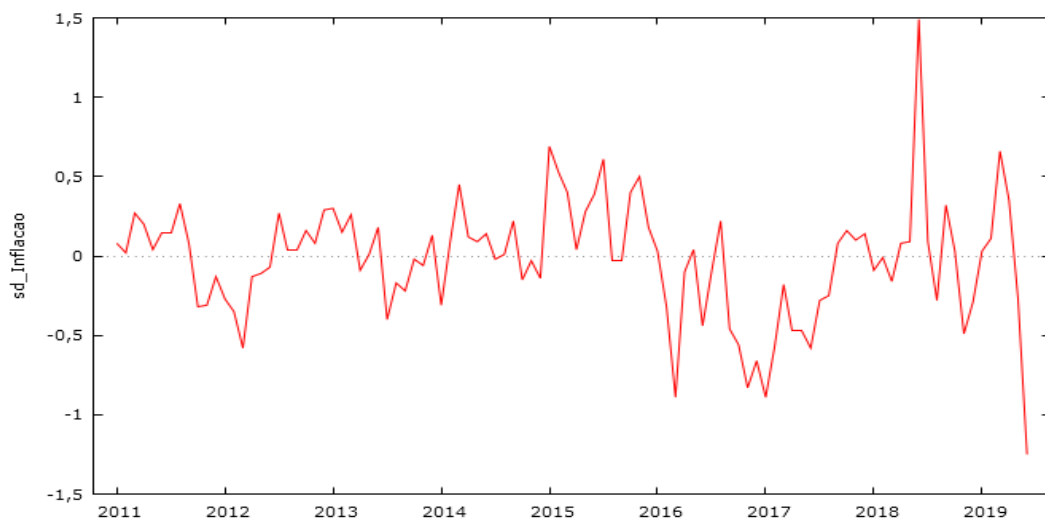
Figura 4 – Periodograma aplicada a primeira diferença sazonal



Fonte: Elaboração própria.

Para verificar se a série, agora com a sazonalidade atenuada, tornou-se estacionária foi realizada uma análise do Gráfico da Série Temporal, da Função Autocorrelação (FAC), da Função Autocorrelação Parcial (FACP) e o Teste de Dickey Fuller Aumentado.

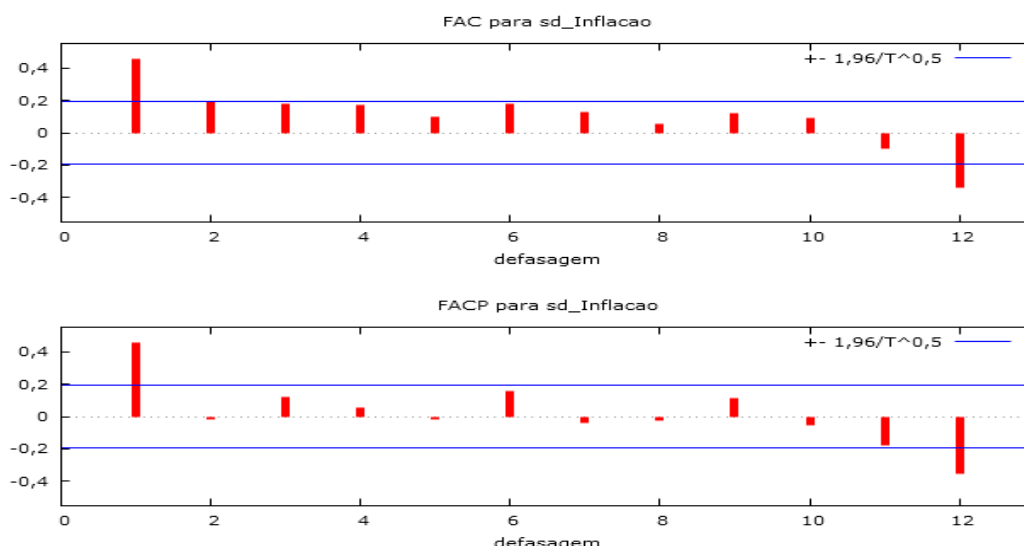
Figura 5 – Gráfico da série temporal de IPCA aplicada a primeira diferença sazonal



Fonte: Elaboração própria.

Analisando a Função Autocorrelação (FAC) e a Função Autocorrelação Parcial (FACP), apresentados na Figura 6, constatou-se que o modelo não apresentou tendência, o que leva a auferir que a série se tornou estacionária. Destaca-se que a série também pode ser considerada estacionária, pois há truncamento abrupto das *lags*.

Figura 6 – Função Autocorrelação (FAC) e Função Autocorrelação Parcial (FACP)



Fonte: Elaboração própria.

O Teste da Raiz Unitária de Dickey-Fuller supondo um grau de significância a 5%, evidencia que a hipótese nula a 5% foi rejeitada, pois o p-valor apresentou valores menores que 5%: o teste com constante apresentou um valor de 0,003416 e o teste com constante e tendência apresentou um valor de 0,01143. Assim a série pode ser considerada estacionária.

Hipóteses $\left\{ \begin{array}{l} H_0 = \text{não é estacionário} \\ H_1 = \text{é estacionário} \end{array} \right.$

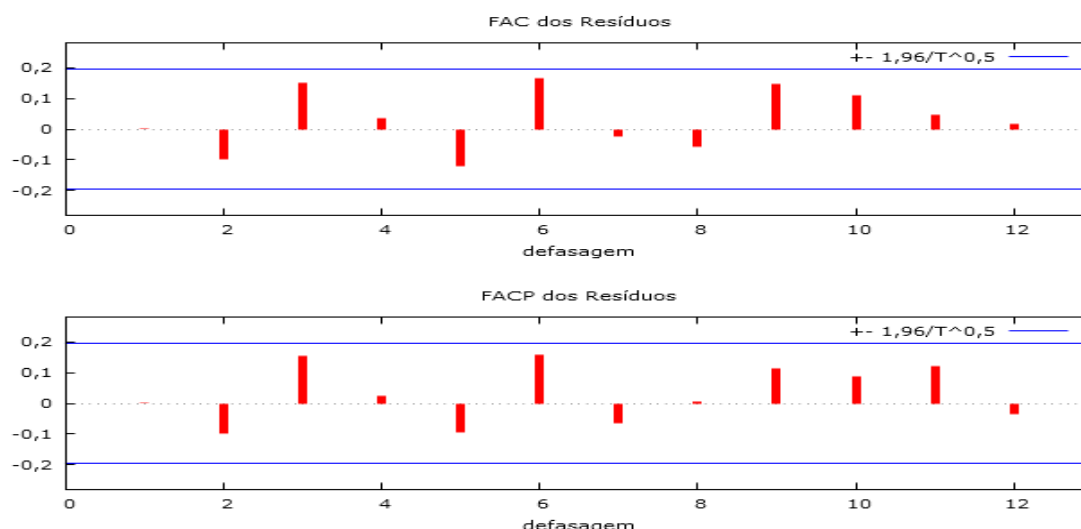
Portanto, com base nas análises do gráfico da série temporal, da FAC, da FACP e do Teste de Dickey-Fuller Aumentado, concluiu-se que a série é estacionária, ou seja, não tem raiz unitária. O gráfico de autocorrelação (FAC) apresentou dependência nos *lags* (padrão trucado) 1 e 12, indicando possíveis quantidades de termos autorregressivos ($p = 1$ e 12). O gráfico de autocorrelação parcial (FACP) também apresentou dependência nos *lags* (padrão trucado) 1 e 12, indicando que quantidade de termos da média móvel deveria ser um ($q = 1$ e 12). Como não foram necessárias diferenciações para tornar a série estacionária o d será igual a zero ($d = 0$). Isso leva à definição inicial do modelo ARMA (1,12). Além disso, foi utilizado outro método relacionado com modelos de séries de tempo denominado de método X-12-ARIMA para a obtenção dos índices sazonais e para uma tentativa de amenizar ainda mais a sazonalidade. Através da estimação do modelo, constatou-se que o valor do coeficiente gerado apresentou um alto nível de significância.

A verificação foi realizada através da análise dos gráficos da FAC e FACP dos resíduos, conhecida como Teste de Autocorrelação Residual. Verificou-se se os resíduos se comportaram como ruídos brancos e apresentaram autocorrelação insignificante. Os gráficos da Figura 7, demonstraram que os resíduos são normalmente distribuídos, o que confirma a inexistência de um padrão de comportamento dos dados; além disso, todos os valores de p foram maiores que 5%, o que mostrou um ruído branco, sendo este modelo o melhor ajustado.

Ao ajustar o modelo ARMA (1,12) verificou-se a significância estatística dos parâmetros. Através do Teste de Autocorrelação Residual pode-se comprovar essa

afirmação de adequação do modelo, já que se constatou que as autocorrelações dos resíduos ficaram dentro do intervalo $[-0,1873; 0,1873]$.

Figura 7 - Função Autocorrelação (FAC) e Função Autocorrelação Parcial (FACP) aplicada as diferenças sazonais e o X-12 ARIMA



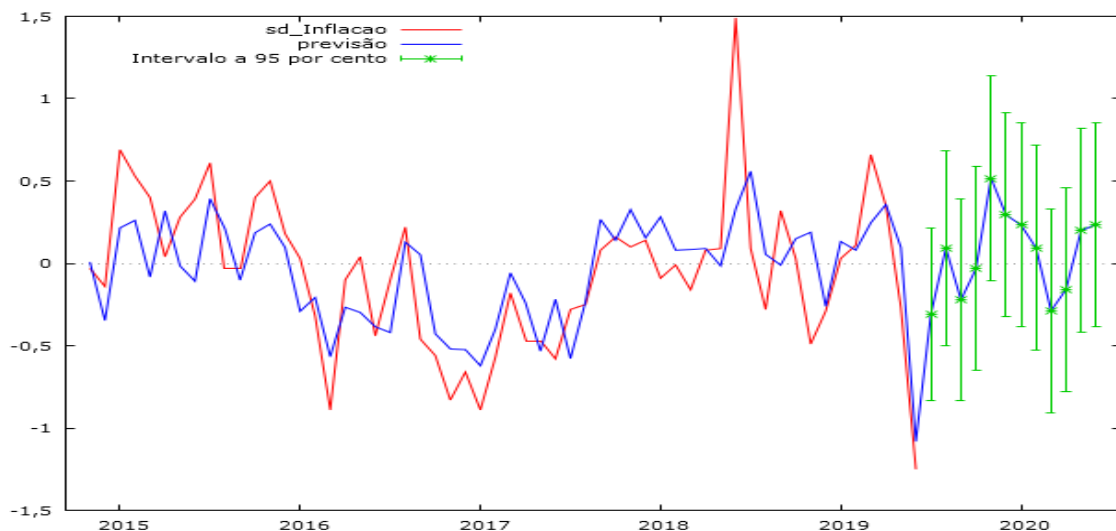
Fonte: Elaboração própria.

Para dar maior consistência a afirmação de adequação dos modelos, realizou-se o Teste de *Box e Pierce*². Como dito anteriormente, o mesmo testa duas hipóteses (11). O modelo deve ser ruído branco para então se mostrar adequado. Como LB foi igual a 12,8707 e λ^2 igual a 0,2310, ou seja, $LB < \lambda^2$, não se rejeitou H_0 , evidenciando que os resíduos são ruídos branco. Portanto, por meio dessas análises pode-se concluir que o modelo é preciso para essa série temporal, o que confirma sua utilização para realizar as previsões.

Após a obtenção do modelo mais adequado, foi realizado a previsão da série em estudo. Através da Figura 8, obteve-se a comparação dos valores reais com os valores previstos, e na Tabela 1, é possível visualizar as previsões de 12 amostras futuras com seus limites de 95% de confiabilidade.

² Conforme explicado na página 9 e 10, o teste de Box e Pierce pode confirmar a existência de ruído branco testando duas hipóteses: as hipóteses H_0 = os resíduos são ruído branco H_1 = os resíduos não são ruído branco. O modelo deve ser ruído branco para então se mostrar adequado. Se $LB < \lambda^2$ não se rejeita H_0 , ou seja, os resíduos são ruídos branco, sendo o modelo adequado.

Figura 8 - Gráfico da Previsão para os meses de julho de 2019 a junho de 2020 da série IPCA.



Fonte: Elaboração própria com base em dados do IPEADATA (2022).

Tabela 1 – Previsão de 12 dados futuros para intervalos de confiança de 95%

Período	Previsão	Erro Padrão
2019.07	-0,31	0,266
2019.08	0,09	0,303
2019.09	-0,22	0,313
2019.10	-0,03	0,316
2019.11	0,52	0,316
2019.12	0,29	0,317
2020.01	0,23	0,317
2020.02	0,09	0,317
2020.03	-0,29	0,317
2020.04	-0,16	0,317
2020.05	0,20	0,317
2020.06	0,23	0,317

Fonte: Elaboração própria.

DISCUSSÃO E CONCLUSÃO

Conseguir prever a inflação é importante principalmente para auxiliar as autoridades monetárias a definirem instrumentos de política e orientar investimentos no mercado financeiro. Através da metodologia Box-Jenkins encontrou-se o modelo ARMA (1,12) que representou o melhor comportamento da série temporal. Vale ressaltar que o grande problema encontrado foi a presença de sazonalidade. Para atenuá-la, a fim de utilizar o método Box e Jenkins que tem como premissa que a série seja estacionária, realizou-se a aplicação de diferença sazonal e o método X-12-ARIMA. A análise dos dados da série de previsão mostrou que o modelo é válido, pois os mesmos encontram-se dentro dos limites aceitáveis, para intervalos de confiança de 95%. Vale ressaltar que os resultados apresentaram meses com deflação.

USING THE BOX & JENKINS METHODOLOGY TO PREVENT BRAZILIAN INFLATION

ABSTRACT: *This article aims to present the future behavior of the National Extended Consumer Price Index (IPCA) series, an official inflation measure, for the period from July 2019 to June 2020. An economy forecast is important to assist as monetary authorities define policy instruments, which have no Brazilian case or the Central Bank uses an inflation targeting system; and also to guide investments in the financial market. For this, use the Box-Jenkins methodology with the aid of the Gretl software. It is worth mentioning that the work made use of the X-12 ARIMA method to mitigate the seasonality of the series in question. A more adequate model was used in the analysis of graphs and in the statistical tests unique to the methodology, or which was defined in ARMA (1, 12).*

Keywords: *General Price Index; Econometrics; Time Series; Box & Jenkins methodology; Forecasts.*

REFERÊNCIAS

ARAÚJO, T. G.; SIQUEIRA, J. C.; CARDOSO, S. C. Ajuste Sazonal do PIB trimestral: X12-ARIMA e Modelo Estrutural: Análise Comparativa dos Resultados. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 19., 2010, São Pedro. **Anais** [...], São Paulo, 2010. Disponível em: <http://www2.ime.unicamp.br/sinape/sites/default/files/Resumo%20estendido%20-%20Thiago%20e%20Julio.pdf>. Acesso em: 4 set. 2022.

BARBOSA, L. M. Prevendo a Inflação Brasileira: Comparação Entre um Modelo Teórico e um Ateórico de Séries Temporais. **Informações Fipe**, São Paulo, 2019. Disponível em: <https://downloads.fipe.org.br/publicacoes/bif/bif462-44-49.pdf>. Acesso em: 9 ago. 2022.

BOX, G. P. E.; JENKINS, G. M. **Times series analysis: forecasting and control**. São Francisco: Holden, 1970.

CASTRO, F.C.C. **Estudo de séries temporais aplicado a perfis de poços de petróleo**. 1995. 117f. Dissertação (Mestrado em Geoengenharia de Reservatórios) – Instituto de Geociências, Universidade Estadual de Campinas, Campinas, 1995.

COELHO JR. **Utilização do método Box-Jenkins para previsão de indicadores econômicos (IPCA, SELIC, Câmbio e Ibovespa)**. 2014. 21f. Monografia (Especialização em Macroeconomia e Finanças) – Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, Curitiba, 2014.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. Porto Alegre: AMGH, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. **Inflação**. Disponível em: <https://www.ibge.gov.br/explica/inflacao.php>. Acesso em: 08 ago. 2022.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA [IPEADATA]. **Sistema Nacional de Índices de Preços ao Consumidor (IBGE/SNIPC)**. 2022. Disponível em: <http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38391>. Acesso em: 30 ago. 2019.

MAIA, A. G. **Econometria: conceitos e aplicações**. São Paulo: Saint Paul, 2017.

SANTOS, C.; FERREIRA, R. T.; CASTELAR, I. Prevendo a inflação no Brasil com grande conjunto de dados: uma aplicação do modelo de fatores comuns. *In: ENCONTRO DE ECONOMIA APLICADA*, 2., 2016, Juiz de Fora. **Anais [...]**, Juiz de Fora: UFJF, 2016.

SARTORIS, A. **Estatística e introdução à econometria**. São Paulo: Saraiva, 2013.

VINHAES, F. A restrita eficácia dos juros no combate à inflação. **Jornal dos Economistas**, Rio de Janeiro, n.346, p.7-8, jun. 2018. Disponível em: <https://www.corecon-rj.org.br/anexos/E3D8B83FDFC97B50530DF8D578B2567F.pdf>. Acesso em: 12 set. 2022.

YAFEE, R.; MCGEE, M. **Introduction to time series analysis and forecasting: with applications of SAS and SPSS**. Cambridge, MA: Academic Press, 2000.