METODOLOGIA BOX-JENKINS: UM EXEMPLO DE APLICAÇÃO NA PREVISÃO DO PIB NACIONAL

PAULO SIGA THOMAZ[1]; VIVIANE LEITE DIAS DE MATTOS[2]

[1]Universidade Federal do Rio Grande (FURG) – paulosigathomaz@gmail.com
[2]Universidade Federal do Rio Grande (FURG) – viviane.leite.mattos@gmail.com

1. INTRODUÇÃO

A previsão de séries temporais é importante, pois auxilia economistas, engenheiros, administradores e profissionais do mercado financeiro em tomadas de decisões. Nesse contexto, a metodologia Box-Jenkins se destaca por produzir modelos simples e que, em geral, apresentam bons resultados. Esses modelos, denominados Auto-Regressivos Integrados de Médias Móveis (ARIMA), podem ser empregados para ajuste e previsão de séries temporais nas mais diversas áreas. Cunha & Margarido (1999), por exemplo, os utilizaram com o objetivo de avaliar o impacto dos planos econômicos na inflação, entre os anos de 1971 e 1998. Os autores identificaram que o plano real foi o único a ter sucesso em estabilizar a inflação em longo prazo. Em outro estudo, Yip et al (2013) os aplicaram para prever o custo de manutenção de equipamentos da construção civil, obtendo, segundo os autores, resultados satisfatórios. Na área ambiental, Taneja et. al (2016) utilizaram os modelos ARIMA com sazonalidade com a intenção de prever a profundidade óptica de aerossol na Índia. Nesse estudo, os modelos foram considerados satisfatórios, embora tenham sido incapazes de simular valores extremos.

O objetivo deste trabalho é apresentar uma síntese da operacionalização das diversas etapas da metodologia Box-Jenkins, envolvendo os processos de identificação dos parâmetros, avaliação da qualidade do ajuste, diagnóstico de resíduos e previsão. Para tal, é realizada uma aplicação utilizando a série temporal do Produto Interno Bruno (PIB) nacional.

2. METODOLOGIA

A metodologia Box-Jenkins consiste em explicar uma variável através de seus valores passados e perturbações aleatórios passadas (SARTORIS, 2013, p. 279). Segundo Bueno (2008, p. 30), um modelo ARMA (p, q) pode ser escrito como:

$$y_t = c + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j},$$

onde, y_t é o valor observado no *t-ésimo* tempo; c é uma constante; i é a defasagem do processo autoregressivo $(i = 1, 2, \ldots, p)$; j é a defasagem do processo de médias móveis $(j = 1, 2, \ldots, q)$; $\emptyset_i$ é o coeficiente da parcela auto-regressiva na *i-ésima* defasagem; θ_j é o coeficiente de médias móveis na *j-ésima* defasagem; p é a ordem do processo auto regressivo; q é a ordem do processo de médias móveis; ε_t é o erro no *i-ésimo* tempo.

Para que o modelo ARMA possa ser aplicado, a série deve ser necessariamente estacionária, ou seja, deve possuir média constante, variância constante e covariância independente do tempo (BUENO, 2008, p.15). Se a série não suprir esse requisito, é possível torná-la estacionária pelo processo de diferenciação, o qual consiste em computar a diferença entre observações consecutivas até que a série atinja a estacionariedade. A combinação do processo auto-regressivo de médias móveis e de diferenciação resulta no modelo ARIMA (p,d,q), onde d indica a ordem de diferenciação.

Para verificar se uma série é estacionária, pode ser utilizado o teste Aumentado de Dickey-Fuller (ADF), que avalia a hipótese nula de que a série não é estacionária (HYNDMAN & ATHANASOPOULOS, 2013). Com a série estacionária e a ordem de diferenciação (d) definida, parte-se para a determinação das ordens do processo auto regressivo (p) e de médias móveis (q). Essas ordens podem ser obtidas iterativamente, analisando-se a significância dos coeficientes do modelo, os gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP). A FAC leva em consideração a relação implícita entre as observações, enquanto a FACP considera cada observação individualmente (BUENO, 2008, p.40-42). Os modelos selecionados nesta etapa são comparados por meio dos Critérios de Informação de Akaike e Hannan-Quinn. Esses critérios são utilizados para identificar o modelo com melhor ajuste e menor complexidade, o qual, dentre os modelos escolhidos, será aquele com os menores valores para o critério de informação (YIP ET AL., 2014, p.32).

A etapa final da escolha do modelo consiste no diagnóstico dos resíduos, onde são avaliadas a normalidade e a presença de autocorrelação entre os resíduos estimados, além da heterocedasticidade. O teste de Jarque-Bera é utilizado para avaliar a normalidade, testando a hipótese nula de que os momentos da série estimada são iguais aos da distribuição normal (BUENO, 2008, p.71). O teste de Ljung-Box avalia a autocorrelação, nesse caso a ausência de autocorrelação entre os resíduos é considerada como hipótese nula (HYNDMAN & ATHANASOPOULOS, 2013). Por fim, é aplicado o teste de ARCH-LM para verificar a presença de heteroscedasticidade, testando a hipótese nula de que os resíduos são homoscedásticos (BUENO, 2008, p.73). Se os resíduos do modelo estimado apresentam distribuição normal, ausência de autocorrelação e homoscedasticidade, ele pode ser usado para previsão.

Na avaliação da qualidade de previsão do modelo foram utilizadas as medidas de erro: raiz do erro quadrático médio, erro absoluto médio e erro percentual absoluto médio.

3. APLICAÇÃO

Como exemplo de aplicação da metodologia Box-Jenkins, foi realizada a modelagem da série histórica deflacionada e sazonalmente ajustada do Produto Interno Bruto (PIB) nacional a preços de mercado, ou seja, com exclusão dos impostos sobre os produtos. Os dados foram obtidos a partir do banco de dados do Instituto Brasileiro de Geografia e Estatística (IBGE). A série é composta de informações trimestrais, partindo do primeiro trimestre de 1996 até o último trimestre de 2016. Os dois últimos anos (8 observações) são utilizados para validação do modelo de previsão estimado. A base da série foi fixada no primeiro trimestre de 1996.

O teste Aumentado de Dickey-Fuller resultou em valor p igual a 0.5830, dessa forma a hipótese nula de não estacionariedade não é rejeitada. Realizou-se, então, a diferenciação da série e o teste foi mais uma vez aplicado, resultando em valor próximo de zero, permitindo que a hipótese nula seja rejeitada na primeira diferença. Portando a ordem de diferenciação da série é d=1.

A Tabela 1 apresenta os resultados dos Critérios de Informação de Akaike (CIA) e Hannan-Quinn (CHQ), além dos testes de Jarque Bera (JB), Ljung-Box (LB) e ARCH-LM (ALM) para a análise dos resíduos estimados. Como os critérios de informação dos modelos chegaram a valores próximos entre si e os diagnósticos de resíduos foram satisfatórios, todos os modelos podem ser usados para previsão. Portanto, para a escolha do melhor modelo foram utilizadas as medidas de erro de previsão: a raíz do erro quadrático médio (REQM), erro absoluto médio (EAM) e erro percentual absoluto médio (EPAM), cujos resultados estão apresentados na Tabela 2.

Tabela 1 – Resultados dos critérios de informação e diagnóstico dos resíduos.

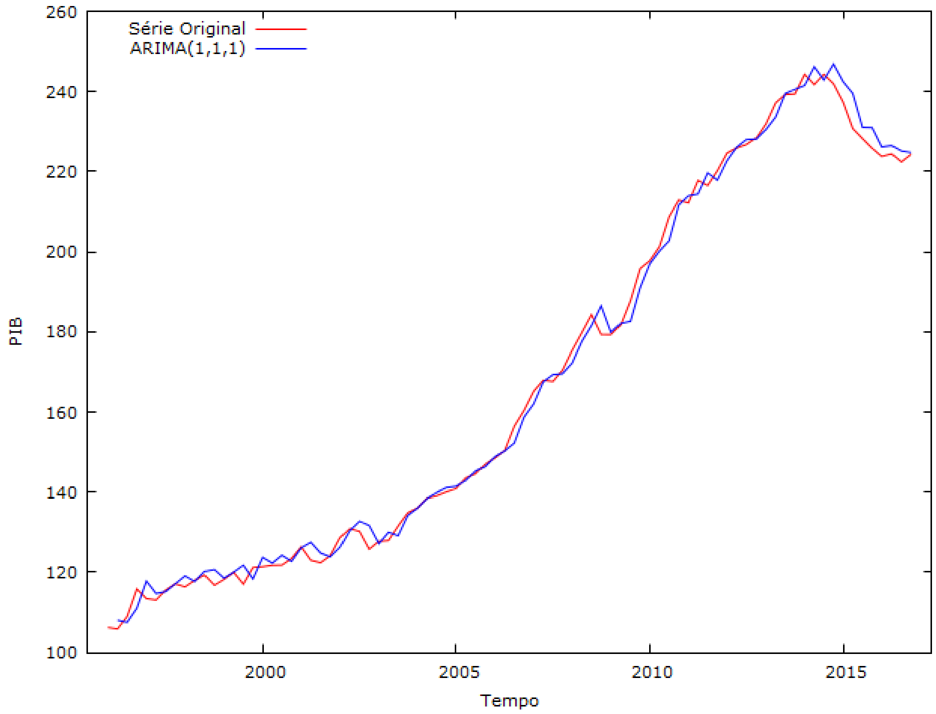
Modelo	**CIA**	**CHQ**	**JB (valor p)**	**LB (valor p)**	**ALM (valor p)**
Arima(1,1,1)	364,52	368,22	0,6781	0,9360	0,5533
Arima(1,1,0)	365,29	368,06	0,6288	0,6556	0,6973
Arima(0,1,1)	365,27	368,04	0,6238	0,6520	0,7049

Tabela 2 – Resultados dos erros de previsão.

Modelo	**REQM**	**EAM**	**EPAM**
Arima(1,1,1)	4,4158	3,6929	1,6140
Arima(1,1,0)	4,4391	3,7907	1,6561
Arima(0,1,1)	4,4380	3,7936	1,6575

É possível observar que o modelo ARIMA (1,1,1) apresenta os menores erros, portanto este é melhor modelo ARIMA para a previsão da série do PIB nacional. No entanto, todos os modelos apresentaram bons resultados e podem ser aplicados. A Figura 1 apresenta o modelo ARIMA (1,1,1) ajustado e a série temporal original.

Figura 1 – Modelo de previsão e série original.



Estes achados concordam com as considerações de Yip et al. (2013) e Taneja et al (2016), os quais afirmam que a abordagem Box-Jenkins consiste em uma modelagem simples, mas que apresenta resultados satisfatórios em diversas áreas de estudo.

4. CONCLUSÕES

Este trabalho apresentou uma síntese acompanhada de uma aplicação da metodologia Box-Jenkins na análise de séries temporais. Os modelos obtidos foram satisfatórios, concordando com trabalhos de outros autores que utilizaram esta metodologia na modelagem de outras séries temporais. O presente estudo faz parte de um estudo maior que objetiva avaliar o desempenho de modelos Auto-Regressivo de Médias Móveis (ARMA) e de modelos Auto-Regressivos de Médias Móveis Generalizados (GARMA), quando aplicados a séries econômicas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BUENO, R. de. L. da S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2008.

SARTORIS, A. **Estatística e Introdução a Econometria**. São Paulo: Editora Saraiva, 2013.

HYNDMAN, R.J.; ATHANASOPOULOS, G. ***Forecasting: principles and practive.*** Melbourne: OTexts, 2013. Acessado em 27 set. 2017. Online. Disponível em: < https://www.otexts.org/fpp>.

YIP, H.; FAN, H.; CHAING, Y. *Predicting the maintenance cost of construction equipment: comparison between general regression neural network and Box-Jenkins time series models.* ***Automation in Construction***, n.38, p. 30-38, 2013.

TANEJA, K.; AHMAD, S; AHMAD, K; ATTRI; S. D. *Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modelling approach.* ***Atmospheric Pollution Research.*** n. 7, p. 585-596, 2016.

CUNHA, MS.; MARGARIDO, MA. Avaliação dos impactos dos planos de estabilização pós-1986 sobre o índice geral de preços (IGP): uma aplicação da metodologia Box-Jenkins. **Revista Científica do Instituto de Economia Agrícola – IEA.** n. 46, p. 1-18, 1999.