

Análise de séries temporais aplicada aos lucros trimestrais da Klabin

Eduardo C. Barbosa^{1†}, Sara Silvério², Paulo César Emiliano¹, Maurício S. Lacerda³

¹Docente do Departamento de Estatística, Universidade Federal de Viçosa (UFV).

²Mestranda do Programa de Estatística Aplicada e Biometria, Universidade Federal de Viçosa (UFV).

³Docente do Departamento de Apoio ao Ensino, Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (DAPE/IFRO).

Resumo: A Klabin S.A. é uma empresa centenária, líder de mercado e também a maior produtora e exportadora de papéis para embalagens do país. A diferença entre as receitas e os custos/despesas se traduz no lucro líquido da companhia, principal fonte de interesse dos investidores, uma vez que a cotação das ações na bolsa de valores é influenciada diretamente pelo lucro da companhia. Nesse sentido, o objetivo deste trabalho é estudar a série histórica de lucros trimestrais da Klabin, analisar seus padrões temporais e ajustar um modelo para prever valores futuros. O modelo ARMA(1,1) com a inclusão de duas variáveis indicadoras e um par de componentes harmônicos foi utilizado no intuito de modelar o efeito de um outlier aditivo, de uma quebra estrutural e de uma periodicidade, respectivamente. Os erros gerados pelo modelo podem ser considerados normais e possuem características de ruído branco. A previsão gerada pelo modelo para os dois primeiros trimestres do ano de 2023 produziu um erro percentual médio absoluto ou MAPE de aproximadamente 7%.

Palavras-chave: Valores discrepantes; quebra estrutural; periodicidade; frequência.

Time series analysis applied to Klabin's quarterly profits

Abstract: Klabin S.A. is a century-old company, a market leader, and also the largest producer and exporter of packaging paper in the country. The difference between revenues and costs/expenses translates into the company's net profit, the main point of interest for investors, as the stock market share prices are directly influenced by the company's profit. In this regard, the goal of this study is to analyze the historical series of Klabin's quarterly profits, examine their temporal patterns, and fit a model to predict future values. The ARMA(1,1) model, including two indicator variables and a pair of harmonic components, was used to model the effect of an additive outlier, structural break, and periodicity, respectively. The errors generated by the model can be considered normal and exhibit characteristics of white noise. The model's forecast for the first two quarters of the year 2023 resulted in an approximately 7% Mean Absolute Percentage Error (MAPE).

Keywords: Outliers; structural break; periodic; frequency.

Introdução

A Klabin é uma das maiores empresas do Brasil no setor de papel e celulose, com uma presença significativa no mercado global. Fundada em 1899, a empresa tem uma longa história na indústria e se tornou uma das principais produtoras e exportadoras de papel e celulose do país. A Klabin atua também em diversos segmentos relacionados à indústria de papel e celulose, como na produção de embalagens de papelão ondulado, cartões e sacos industriais. Ela oferece uma ampla gama de produtos para atender às necessidades de diversos setores, como alimentos, produtos de higiene e limpeza, agricultura, entre outros. Além de operar em todo o Brasil, a Klabin tem uma presença internacional crescente. Ela exporta seus produtos para vários países e mantém escritórios e operações em diferentes partes do mundo (LAFER, 2020).

[†]Autor correspondente: duardo.barbosa@ufv.br.

Adicionalmente, as ações da Klabin são negociadas na Bolsa de Valores de São Paulo (B3) sob os *tickers* “KLBN3 (Ação ordinária ou ON)”, “KLBN4 (Ação Preferencial ou PN)” e “KLBN11 (Ação Unit)” e fazem parte do Índice Ibovespa, o que indica que suas ações possuem alta liquidez e elevado valor de mercado.

Neste sentido, o principal objetivo desse trabalho foi modelar, via análise de séries temporais, os lucros históricos e trimestrais da Klabin. Além de analisar os padrões temporais da série, pretende-se estabelecer previsões futuras para os lucros dos dois últimos trimestres do ano de 2023. A justificativa para o desenvolvimento deste trabalho é que, em geral, as cotações das ações na bolsa de valores são positivamente correlacionadas com os lucros das respectivas companhias. Logo, possuir estimativas dos lucros futuros de uma companhia implica em entender melhor a direção da cotação de suas ações, o que possibilita aos investidores tomar melhores decisões.

Referencial teórico

A análise de séries temporais é uma metodologia estatística frequentemente utilizada para analisar padrões temporais e estabelecer projeções futuras, embasadas em teorias de que o comportamento do mercado, ao longo do tempo, consiste na repetição de tendências, sazonalidades/ciclos e variações aleatórias, de forma semelhante com o ocorrido no passado (MORETTIN; TOLOI, 2006). A metodologia Box e Jenkins, delineada por Box e Jenkins (1970), produz uma classe de modelos estocásticos de previsão, denominada modelos autorregressivos integrados e de médias móveis, ou ARIMA. É possível ainda adicionar ao modelo variáveis explicativas ou regressoras, com o intuito de remover tendências ou sazonalidades determinísticas, bem como efeitos intervencionais de *outliers* e quebras estruturais.

Modelo autorregressivo - AR(p)

De acordo com Morettin e Tolo (2006), o modelo $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$, para $t = 1, 2, \dots, n$, é um processo autorregressivo de ordem p ou AR(p). Neste modelo os valores da série temporal Y_t são representados por uma média geral μ mais uma soma de p observações anteriores Y_{t-1}, \dots, Y_{t-p} ponderadas pelos valores de ϕ_1, \dots, ϕ_p , que são parâmetros tradicionalmente estimados por máxima verossimilhança. Adicionalmente, ε_t representa o erro aleatório e não observável do modelo, o qual supõe-se ser um ruído branco, isto é, $\varepsilon_t \sim RB(0, \sigma^2)$.

Modelos de médias móveis - MA(q)

O modelo $Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$, para $t = 1, 2, \dots, n$, é um processo de médias móveis de ordem q , ou MA(q) (MORETTIN; TOLOI, 2006). Neste modelo os valores da série temporal Y_t são representados por uma média geral μ mais uma soma de q erros ou choques aleatórios anteriores $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$, ponderados pelos valores de $\theta_1, \dots, \theta_q$, que são parâmetros estimados por máxima verossimilhança. Adicionalmente, ε_t representa o erro aleatório e não observável do modelo, o qual supõe-se ser um ruído branco.

Modelo autorregressivo e de médias móveis - ARMA(p, q)

O modelo ARMA de ordem p, q , sendo p a ordem da parte autorregressiva e q a ordem da parte médias móveis do modelo, é representado da seguinte forma $Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$. Assim como nos modelos AR(p) e MA(q), ε_t são erros ruído branco. Os parâmetros do modelo ARMA(p, q) são, também, estimados pelo método da máxima verossimilhança. Vale destacar que quando diferenças matemáticas são necessárias para tornar a série original estacionária, tem-se então um modelo ARIMA(p, d, q). Adicionalmente, um modelo ARIMA($p, 0, q$) equivale a um modelo ARMA(p, q).

Regressão harmônica

Em diversos estudos, a série temporal sob análise pode apresentar periodicidade, isto é, sazonalidade (s inferior a 1 ano) ou ciclicidade (s superior a 1 ano). Segundo Percival e Walden (1993), uma forma de descobrir a periodicidade de uma série é por meio do método do periodograma. Após encontrar a periodicidade s dos dados é possível incluir no modelo ARIMA(p, d, q) variáveis regressoras harmônicas (senos e cossenos) para então estimar o efeito dessa componente. O modelo é definido como

$$Y_t = \beta_0 + \beta_1 \sin(wt) + \beta_2 \cos(wt) + a_t,$$

em que $w = \frac{2\pi}{s}$, $s = \frac{n}{f}$, $f = \max\{I(f)\}$ e $a_t \sim \text{ARIMA}(p, d, q)$.

Análise de intervenção

O termo análise de intervenção foi introduzido por Glass (1972), baseado na teoria estabelecida por Box e Tiao (1965), que já utilizavam estes métodos com uma outra denominação. O modelo de análise de intervenção é definido como:

$$Y_t = \sum_{i=1}^k v_i(B)X_{i,t} + \eta_t,$$

em que Y_t são os valores da série temporal e $v_i(B)$ é a função de transferência que permite inserir no modelo o efeito das intervenções através de até k variáveis indicadoras e, na representação estocástica do modelo $\eta_t \sim \text{ARIMA}(p, d, q)$. A seguir, apresenta-se duas das principais formas de se intervir em uma série utilizando uma função impulso e degrau, respectivamente:

$$X_{1t} = \begin{cases} 1, & \text{se } t = T_1 \\ 0, & \text{se } t \neq T_1 \end{cases} \quad \text{e} \quad X_{2t} = \begin{cases} 1, & \text{se } t \geq T_2 \\ 0, & \text{se } t < T_2 \end{cases}.$$

Metodologia

A série temporal avaliada possui $n = 48$ observações, sendo que cada uma dessas observações corresponde ao lucro líquido (em bilhões de reais) auferido no final de um determinado trimestre pela Klabin. Mais especificamente, a amostra compreende dados que vão do 1º trimestre de 2011 até o 4º trimestre de 2022. Essas informações foram extraídas dos relatórios administrativos disponíveis na página de RI (Relação com Investidores) da companhia (<https://ri.klabin.com.br>).

A análise inicial consistiu em avaliar, via periodograma (PPERIVAL; WALDEN, 1993), se a série original apresentava periodicidade. Investigou-se também a presença de uma quebra estrutural (ou *breakpoint*) na série e de um *outlier* aditivo, isto é, de um evento que cause apenas um choque pontual na série. A análise de quebra estrutural foi realizada através do teste de Bai e Perron (1998) e, para testar o efeito do *outlier* aditivo, foi empregado o teste de Chang et al. (1988).

O efeito das componentes (tendência e periodicidade) e dos eventos mencionados anteriormente (quebra estrutural e *outlier* aditivo) foram removidos da série através do ajuste de um modelo de regressão linear múltipla, utilizando como variáveis explicativas um par de componentes harmônicos e duas variáveis indicadoras do tipo impulso e degrau.

Posteriormente, para modelar a autocorrelação residual remanescente, foram ajustados modelos da classe ARIMA(p, d, q). A ordem do modelo foi selecionada visando minimizar o Critério de Informação Bayesiano ou BIC (SCHWARZ, 1978), pelo fato deste critério ser mais apropriado para séries temporais (EMILIANO et al., 2014), sendo o critério definido por:

$$\text{BIC} = -2 \ln L(\hat{\theta}) + k \ln(n),$$

em que $L(\hat{\theta})$ é o valor da função de verossimilhança ajustada do modelo, k é o número de parâmetros a serem estimados e n é o número de observações da amostra.

Além do BIC, estimou-se a variância residual ($\widehat{\sigma^2}$) e o Erro Percentual Médio Absoluto (*Mean Absolute Percentage Error* - MAPE) do modelo, este último definido como:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|,$$

em que Y_t é a observação real no tempo t e \hat{Y}_t é a previsão estimada para o tempo t .

Por fim, foi realizada a análise dos resíduos do modelo selecionado. O objetivo desta análise é verificar se os resíduos gerados pelo modelo possuem características de ruído branco. Nesse sentido, foram analisados os gráficos da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP) dos resíduos. Adicionalmente, foi aplicado o teste de Ljung e Box (1978) para as séries de resíduos. O intuito da aplicação desse teste é verificar se um conjunto de m autocorrelações pode ser considerado estatisticamente igual a zero, sendo a estatística do teste dada por:

$$Q(K) = n(n+2) \sum_{j=1}^m \frac{\hat{r}_j^2}{(n-j)},$$

sendo \hat{r}_j a autocorrelação estimada da série no *lag* j ; n é a quantidade de observações e; m é o número de *lags* que estão sendo testados.

As análises estatísticas foram realizadas no *software* livre R (R CORE TEAM, 2023) utilizando os pacotes: *forecast* (ajuste de modelos ARIMA), *TSA* (teste para *outlier* aditivo), *strucchange* (teste para quebra estrutural), *lmtest* (significância de coeficientes ARIMA) e *descomponer* (Periodograma).

Resultados e discussão

A Figura 1 ilustra a série histórica de lucros trimestrais da Klabin. Observe que no 1° trimestre do ano de 2020 (observação 37) a série em estudo apresenta um valor atípico ou uma observação que se destoa das demais. Nesse caso, essa observação corresponde a um prejuízo trimestral de cerca de 3 bilhões de reais, possivelmente causado pelos efeitos iniciais da pandemia da Covid 19. O teste de Chang et al. (1988) reforça a hipótese de um *outlier* aditivo nesse trimestre, uma vez que o módulo da estatística do referido teste (bilateral) excedeu ao valor do quantil da distribuição normal padrão, correspondente a um nível de significância de 5%. Ou seja, $|\lambda| = |-5,54| = 5,54 > 1,96$.

A Figura 2 ilustra a decomposição da série nas suas componentes de tendência, periodicidade e parte aleatória.

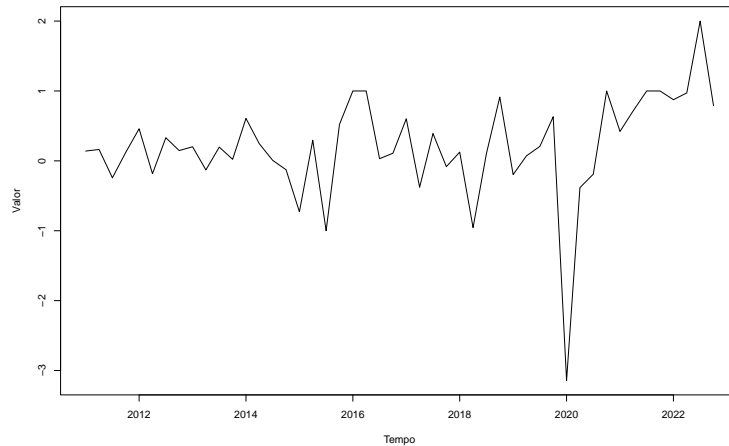
Observe que até a ocorrência da observação atípica, no 1° trimestre de 2020, a série parece oscilar em torno de um nível constante, sem a presença de tendência. Porém, posteriormente, o nível da série aumenta gradualmente, até atingir um patamar superior de lucros.

Nesse sentido, o teste de Bai e Perron (1998) foi utilizado e o resultado de sua aplicação indicou a existência de uma quebra estrutural ou *breakpoint* (menor valor de BIC = 108,15) na série, sendo observada a partir da observação 39 (3° trimestre de 2020). A Figura 3 mostra os resultados do teste de Bai e Perron (1998) e a Figura 4 ilustra a quebra estrutural.

A Figura 2 demonstra, também, que a componente periódica é muito bem definida na série, o que era esperado, uma vez que os lucros da KLABIN são influenciados pelo preço da celulose, uma *commoditie* cíclica. O periodograma da série é apresentado na Figura 5.

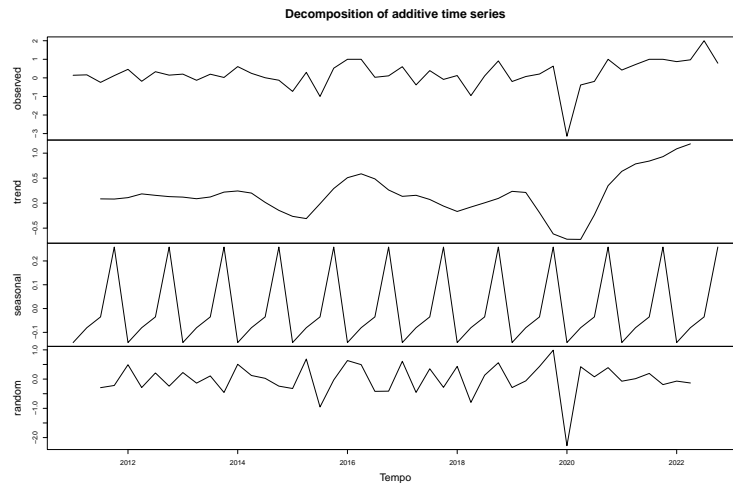
Note que a série apresenta maior intensidade para a frequência $f = 4$ (ou $\frac{f}{n} = \frac{4}{48} \approx 0,083$), o que equivale a um período de $s = \frac{n}{f} = \frac{48}{4} = 12$ trimestres. Logo, a frequência angular é

Figura 1: Série temporal de lucros trimestrais da Klabin.



Fonte: Autores.

Figura 2: Decomposição da série temporal em componentes.



Fonte: Autores.

$w = \frac{2\pi}{s} = \frac{2\pi}{12} = 0,52$. A Figura 6 apresenta o modelo harmônico ajustado para a frequência angular de $w = 0,52$.

Detectados os efeitos presentes na série, a Tabela 1 informa o valor da variância residual estimada ($\widehat{\sigma^2}$) e do Critério de Informação Bayesiano (BIC) para os três modelos ajustados:

Tabela 1: Variância residual e BIC para os modelos candidatos.

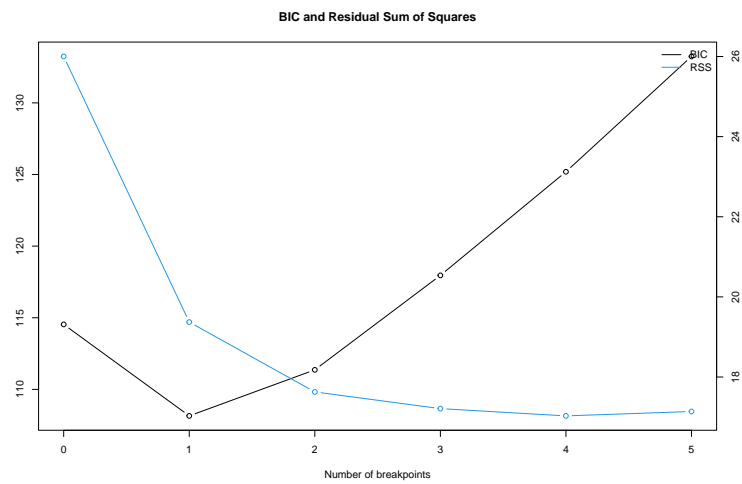
Modelos	$\widehat{\sigma^2}$	BIC
M_1	0,20	71,80
M_2	0,17	70,71
M_3	0,13	65,99

Fonte: Autores.

A especificação dos três modelos é dada a seguir, sendo que o modelo 3, um modelo ARMA(1, 1) com quatro variáveis regressoras, foi selecionado por obter os menores valores de $\widehat{\sigma^2}$ e BIC:

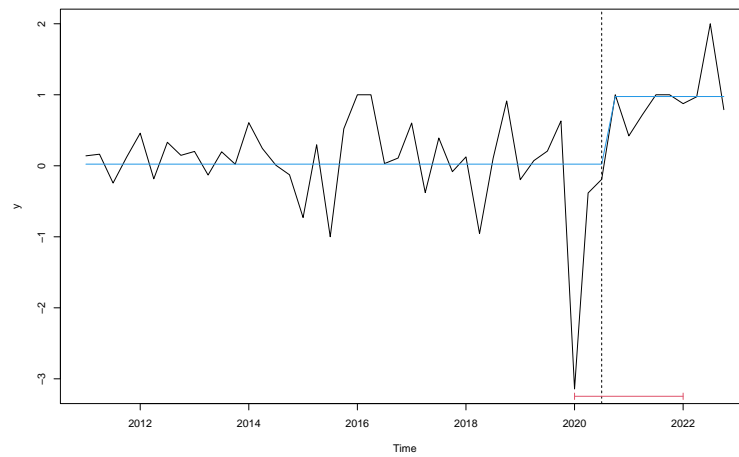
$$M_1 : \widehat{Y}_t = 0,11 - 3,25X_{1t} + 0,87X_{2t} + a_t, \quad a_t \sim RB(0; 0, 20)$$

Figura 3: Resultado do teste de Bai e Perron (1998).



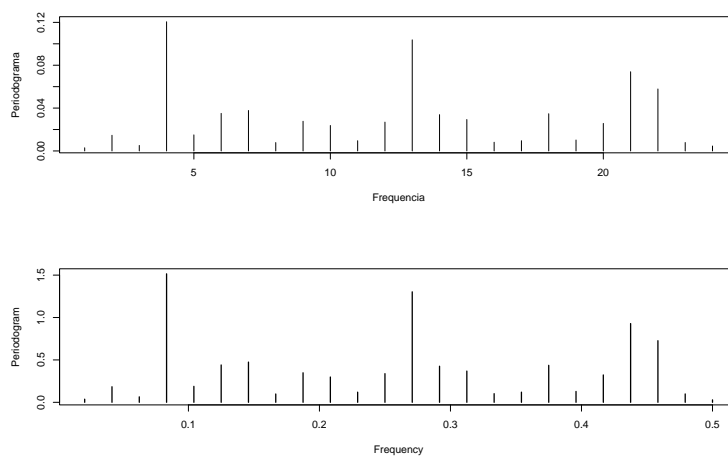
Fonte: Autores.

Figura 4: Quebra estrutural nos dados.



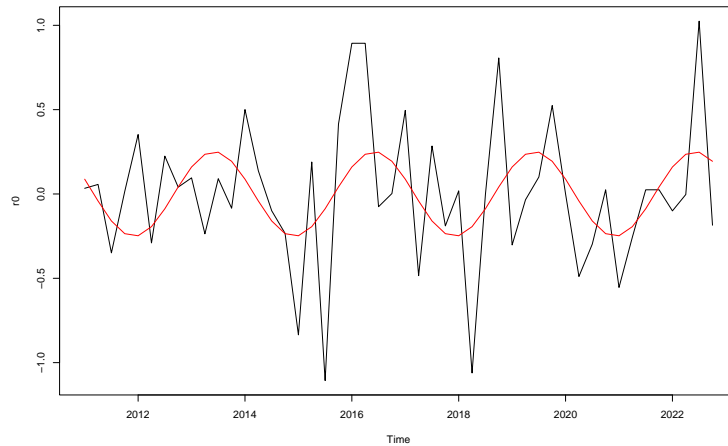
Fonte: Autores.

Figura 5: Periodograma da série.



Fonte: Autores.

Figura 6: Periodicidade de $s = 12$ trimestres ajustada.



Fonte: Autores.

$$M_2 : \hat{Y}_t = 0,11 - 3,34X_{1t} + 0,85X_{2t} - 0,16 \sin(0,52t) + 0,20 \cos(0,52t) + a_t, \quad a_t \sim RB(0; 0, 17)$$

$$M_3 : \hat{Y}_t = 0,12 - 3,75X_{1t} + 0,80X_{2t} - 0,16 \sin(0,52t) + 0,20 \cos(0,52t) + a_t$$

$$a_t = 0,43a_{t-1} - 0,99\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim RB(0; 0, 13),$$

em que X_{1t} e X_{2t} são, respectivamente:

$$X_{1t} = \begin{cases} 1, & \text{se } t = 37 \\ 0, & \text{se } t \neq 37 \end{cases} \text{ e } X_{2t} = \begin{cases} 1, & \text{se } t \geq 39 \\ 0, & \text{se } t < 39 \end{cases}.$$

Já a Tabela 2 ilustra o resultado completo da estimação, por máxima verossimilhança, dos parâmetros do modelo M_3 , inclusive com o teste de significância de Wald.

Tabela 2: Parâmetros estimados para o modelo M_3 .

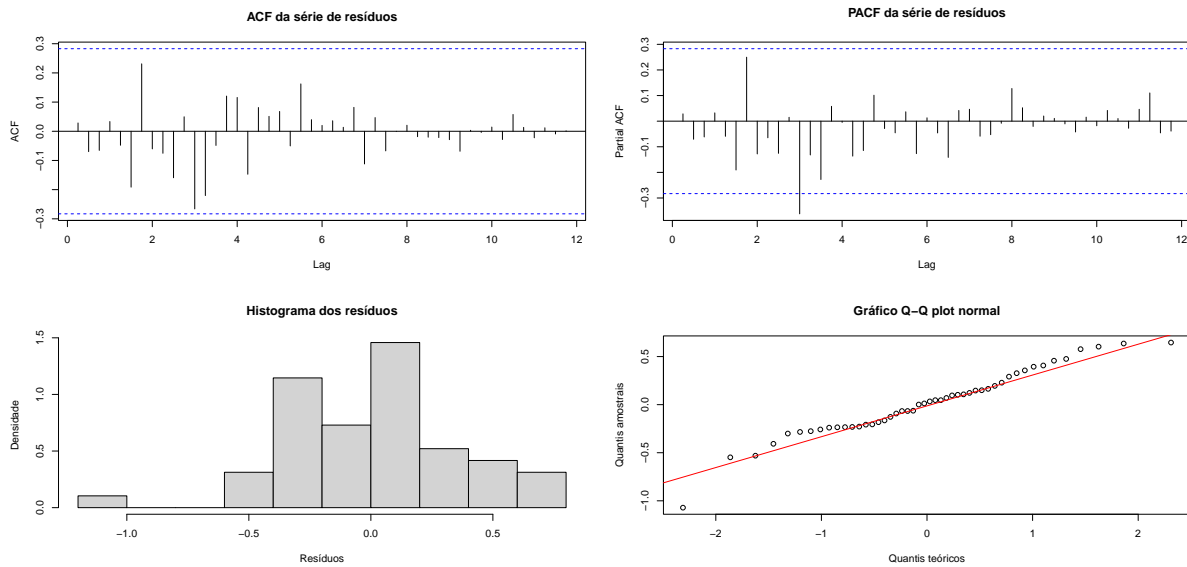
Parâmetros	Estimativas	Erro padrão	z_{cal}	valor p
β_0	0,1240332	0,0087178	14,2276	2.2e-16 *
β_1	-3,7497935	0,3078735	-12,1797	2.2e-16 *
β_2	-0,1567544	0,0543225	-2,8856	0,0039064 *
β_3	0,1993001	0,0570671	3,4924	0,0004787 *
ϕ_1	0,4332009	0,1394452	3,1066	0,0018925 *
θ_1	-0,9999879	0,0562522	-17,7769	2.2e-16 *

Fonte: Autores.

A Figura 7 ilustra o diagnóstico dos resíduos do modelo M_3 , que aparentam ser não correlacionados de acordo com os gráficos das funções de autocorrelação simples e parcial. Além disso, o QQ-Plot e histograma de probabilidades sugerem a normalidade dos resíduos. Essas conclusões são corroboradas pelo teste Ljung-Box (valor $p \approx 58\%$) e Shapiro-Wilk (valor $p \approx 17\%$).

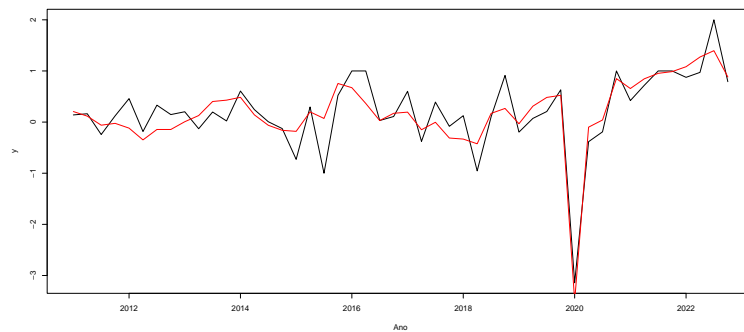
Por fim, a Figura 8 apresenta o ajuste do modelo M_3 aos dados históricos.

As Tabelas 3 e 4 informam, respectivamente, o valor de MAPE produzido pelo modelo ajustado para as duas observações disponíveis no ano de 2023 (1º e 2º trimestres de 2023) e as previsões, com os intervalos de 95% de confiança, para o 3º e 4º trimestres.

Figura 7: Diagnóstico dos resíduos do modelo M_3 .

Fonte: Autores.

Figura 8: Série (linha preta) e valores ajustados (linha vermelha).



Fonte: Autores.

Tabela 3: Valores reais, estimados e MAPE.

Período	Y_t	\hat{Y}_t	$\frac{Y_t - \hat{Y}_t}{Y_t} \times 100$
1º tri./2023	1,0000	0,9646	3,53%
2º tri./2023	0,9731	0,8642	11,19%
$MAPE \approx 7,36\%$			

Tabela 4: Previsão para 3º e 4º trimestres de 2023.

Período	LI	\hat{Y}_t	LS
3º tri./2023	-0,0856	0,7565	1,5987
4º tri./2023	-0,1619	0,6835	1,5289

Conclusões

O modelo ajustado conseguiu captar os efeitos intervencionais na série e também o seu padrão periódico, oriundo do ciclo da *commoditie* celulose. As previsões indicaram uma queda no lucro da Klabin, trimestre a trimestre, para o ano de 2023. Logo, as ações da companhia podem apresentar queda na cotação ao longo deste período, possivelmente por conta de uma virada do ciclo.

Agradecimento

Os autores agradecem Capes, CNPq, Fapemig e a Universidade Federal de Viçosa que deram suporte financeiro para realização deste artigo.

Referências

- BAI, J., PERRON, P. Estimating and testing linear models with multiple structural changes. *Econometrica*, v. 66, n. 1, p. 47-78, 1998.
- BOX, G. E. P.; JENKINS, G. M. *Time series analysis forecasting and control*. San Francisco: Holden-Day, 1970, 537 p.
- BOX, G. E. P.; TIAO, G. C. A change in level of non stationary time-series. *Biometrika*, London, v. 52, n. 1, p. 181-192, 1965.
- CHANG, I. H.; TIAO, G. C.; CHEN, C. Estimation of time series parameters in the presence of outliers. *Technometrics*, v. 30, n. 2, p. 193-204, 1988.
- EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. Information criteria: How do they behave in different models?. *Computational Statistics & Data Analysis*, v. 69, p. 141-153, 2014.
- GLASS, G. V. Estimating the effects of intervention into a non-stationary time-series. *American Educational Research Journal*, v. 9, n. 3, p. 463-477, 1972.
- LAFER, C. A Saga das famílias Lafer e Klabin no Brasil e o pioneiro Maurício Klabin. *Entrevista a Roberta Sundfeld*, 2020.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297-303, 1978.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2. ed. São Paulo: E. Blücher, 2006. 538 p.
- PERCIVAL, D.; WALDEN, A. *Spectral analysis for physical applications*. Cambridge: Cambridge University Press, 1993. 602 p.
- R CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2023.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v. 6, n. 2, p. 461-464, 1978.