PERSPECTIVAS DO SETOR EDITORIAL BRASILEIRO ATRAVÉS DE MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS

André Assis de Salles
Universidade Federal do Rio de Janeiro
Centro de Tecnologia – Bloco F – sala F 101 – Ilha do Fundão – Rio – Brasil
e-mail: as@ufrj.br

Carolina Maria Galvão de Souza
Universidade Federal do Rio de Janeiro
e-mail: carolinagalvao@poli.ufrj.br

RESUMO

O presente trabalho tem por objetivo estudar as perspectivas da produção do mercado editorial brasileiro e elaborar um modelo de previsão que permita obter expectativas futuras para o desempenho desse setor. Para este fim, foi considerada a série de faturamento do setor editorial de livros, disponibilizada a partir de 1995 até 2016 e a série histórica da categoria "livros, jornais, revistas e papelaria" gerada pelo IBGE. O referencial teórico deste trabalho está baseado no desenvolvimento de conceitos e modelos da análise de séries temporais. Foram realizados testes estatísticos a fim de examinar os pressupostos de estacionariedade de ambas as séries. Além disso, foram testadas as hipóteses de violação dos pressupostos de normalidade, homocedasticidade e não autocorrelação de resíduos.

PALAVRAS CHAVE. Setor Editorial, Séries Temporais, Metodologia ARIMA.

PO na Área de Estatística, Gestão Financeira, PO na Indústria.

ABSTRACT

This study aims to analyze the perspectives of Brazilian's publishing market and to establish a forecast model to estimate the sector's future expected performance. For that purpose, two time series were employed: publishing revenue time series, available from 1995 to 2016 and the "books, newspaper, magazines and stationary" time series generated by IBGE. This work's theoretical framework is based on time series concepts and analyses and to verify the assumption of stationarity for both time series statistical tests were conducted. Finally, the violation hypothesis of normality, homoscedasticity and non-autocorrelation of residuals were tested.

KEYWORDS. Publishing Market, Time Series, ARIMA Methodology.

OR in Statistics, Financial Management, OR in Industry.

https://proceedings.science/p/85188?lang=pt-br

1. Introdução

A disseminação da leitura é essencial para o desenvolvimento educacional e prosperidade intelectual de uma nação. Embora os resultados da produção editorial tenham sido historicamente negligenciados como objeto de pesquisa e análise, esse comportamento vem se alterando na última década. Halleweell (1985) oferece uma extensa introdução ao panorama histórico da produção e comercialização de livros no Brasil, apontando os principais desafios econômicos, políticos e sociais que o mercado editorial de livros no Brasil enfrentou e enfrenta desde o início de sua história. Como observado por Earp e Kornis (2005), sua cadeia produtiva reúne os setores autoral, editorial, gráfico, produtor de papel e de máquinas gráficas, distribuidor, atacadista, livreiro e bibliotecário, sendo extensa e com interações em vários setores da economia.

A reação do mercado editorial brasileiro frente à abertura da economia nacional da década de 1990 foi objeto de estudo de Korth (2005), que analisa o impacto da globalização sobre as perspectivas tecnológica, política e econômica. A entrada de novas tecnologias no mercado e a maior facilidade de acesso às mesmas acarretou na redução das barreiras à entrada nesse setor. As tecnologias de informação e de informática auxiliaram a gestão desses negócios e rede mundial de computadores se consolidou como importante canal de vendas. A entrada de editoras estrangeiras no mercado brasileiro também se deu a partir dessa abertura econômica, que expôs o setor ao interesse internacional. O modo de entrada das editoras estrangeiras se dá principalmente pela aquisição de editoras nacionais, o que garante uma herança operacional além do catálogo de títulos e autores. A atualização tecnológica do parque gráfico, segundo Korth (2005), foi uma das principais evoluções da indústria editorial brasileira após a abertura econômica no início dos anos noventa. A partir dos anos 90 as barreiras à entrada de novos concorrentes diminuíram: o número de editoras se expandiu ao mesmo tempo em que a indústria de livros se tornou mais competitiva e profissional. O aumento do número de editoras se contrapõe à entrada de grandes grupos editoriais internacionais no mercado brasileiro ao longo dos últimos anos, ocasionando uma maior concentração desse mercado. A Companhia das Letras vendeu 45% de suas ações ao grupo Penguin em 2011. Em 2014 a Objetiva foi adquirida pelo grupo Penguin Random House, resultado da fusão entre a britânica Penguin e da americana Random House em 2012.

Em trabalho mais recente, Bastos e Fleck (2014) analisaram as formas de crescimento da indústria editorial de livros no Brasil entre 1808 e 2012. O período compreendido entre 1990 e 2012 foi considerado a quarta fase de desenvolvimento dessa indústria. Os principais fatores de crescimento dessa etapa tiveram cunho tecnológico, sendo eles: (i) a disseminação de processadores de texto; (ii) a disponibilização de ferramentas digitais de edição e diagramação; (iii) o surgimento do e-commerce como canal de distribuição e (iv) a introdução do livro eletrônico no mercado. Segundo os autores, as políticas públicas também tiveram papel muito relevante no crescimento do setor, com a expansão dos programas de compra de livro pelo governo, o aumento real da renda e a ampliação da educação. Juntos, o Programa Nacional do Livro Didático (PNLD) e o Programa Nacional Biblioteca da Escola (PNBE) corresponderam por mais de 90% do volume de livros vendidos ao governo em termos de faturamento e exemplares. É importante ressaltar que o subsetor de livros didáticos é o que apresentou resultado mais estável ao longo da década passada, tratando-se de um subsetor amplamente amparado pelas vendas realizadas ao governo. Iniciativas federais como o Programa Nacional do Livro Didático (PNLD) e o Fundo Nacional do Desenvolvimento da Educação (FNDE) resultam na compra e distribuição de livros didáticos ao ensino público com periodicidade regular. Esse investimento estatal na compra de livros ainda é visto pelo setor como um fator essencial para o desempenho econômico do mesmo. Uma vez que o governo normalmente realiza encomendas de grande volume e com prazo de pagamento único, Bastos e Fleck (2014) atribuem aos programas atuais de compra de livros grande importância no que tange o aumento do tamanho do mercado real e alívio financeiro às editoras. Mesmo considerando esta supracitada dependência em relação às políticas estatais, trata-se de um setor industrial de margem moderada e para o qual é bastante pertinente a previsão de seu desempenho frente às variações macroeconômicas do país, a fim de traçar diretrizes mercadológicas para o futuro. A influência dos aspectos organizacionais e ambientais no desempenho da indústria editorial brasileira foi objeto de estudo de Fonseca (2013).

Considerando as quatro funções vitais desse setor – a saber: criação, seleção, produção e distribuição, Fonseca (2013) destaca a tecnologia e as políticas públicas como as duas principais forças propulsoras à aceleração do seu crescimento. As políticas públicas se desdobram na compra de livros pelo governo (aumento do mercado real), no aumento real da renda (aumento do mercado nominal), na ampliação da educação, nas isenções de impostos e nas políticas gerais de incentivo à cultura e leitura. No âmbito tecnológico, processador de texto, o e-commerce, a impressão sob demanda e o livro eletrônico impulsionaram o crescimento da indústria no curto e no longo prazo.

Este trabalho tem por objetivo estudar o comportamento do mercado editorial brasileiro e sua perspectiva futuras. Inicialmente foi ponderada a hipótese de criar modelos de regressão que relacionassem o desempenho do setor editorial com os resultados macroeconômicos brasileiros, dada as séries de resultado do PIB e do Índice de Atividade Econômica do Banco Central do Brasil (IBC-Br). No entanto, esses modelos se mostraram fracamente ajustados e pouco promissores, o que parece indicar que o setor editorial se mantém inelástico em relação às séries supramencionadas. Assim foram estudadas as séries de faturamento e de vendas da indústria através de métodos descritos na literatura de séries temporais.

Para tanto, está dividido em cinco seções, incluindo esta introdução. A próxima Seção 2 explicita a metodologia adotada na construção dos modelos de séries temporais utilizados para este trabalho. Enquanto a Seção 3 apresenta as amostras utilizadas e contempla a caracterização dos dados utilizados. Na Seção 4 são reportados e analisados os resultados empíricos obtidos, com os modelos estocásticos construídos. Por fim, a Seção 5 apresenta as considerações finais do trabalho.

2. Abordagem Metodológica

Para consecução dos objetivos deste trabalho foram construídos três modelos de séries temporais que serão descritos no decorrer desta seção, a saber: modelos ARIMA, modelos autoregressivos e de média móvel integrados; modelos SARIMA, modelos ARIMA com componente sazonal; e Modelos autoregressivos com variáveis *dummy*.

2.1. Modelos ARIMA

Os modelos ARIMA representam um conjunto de processos estocásticos que resultam combinação de três componentes, podendo conter os três ou subconjuntos deles, são eles: (i) componente autoregressivo (AR); (ii) componente de médias móveis (MA) e o (iii) componente de integração. De uma forma simplificada, os modelos ARIMA podem ser classificados da seguinte forma:

(i) Modelo autoregressivo (AR) -- um modelo autoregressivo AR(p), Y_t é descrito pela soma ponderada de p fatores passados e pelo choque aleatório no período t (u_t), que pode ser representado pela equação (1), a seguir.

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + u_t \quad (1)$$

(ii) Modelo de médias móveis (MA) -- um modelo de médias móveis MA(q), Y_t resulta da combinação linear dos termos de erro de ruído branco ocorridas no período atual e nos períodos passados, que pode ser representado pela equação (2), a seguir.

$$Y_t = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q} \quad (2)$$

(iii) Modelo autoregressivo de médias móveis (ARMA) -- Quando a série temporal apresenta características tanto do processo AR quando de MA, ou seja, quando Y_t é descrito tanto por seus valores passados quando pelos choques aleatórios passados e atual, construímos um modelo ARMA(p ,q), com p termos autoregressivos e q termos de média móvel. O Modelo ARMA(p, q) pode ser representado pela expressão (3), adiante.

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q} \quad (3)$$

(iv) Modelo autoregressivo integrado de médias móveis (ARIMA) -- Se a série temporal se tornar estacionária após a diferenciarmos d vezes (ou seja, for originalmente uma série integrada de ordem d) e essa série resultante for representada pelo modelo ARMA(p, q), pode ser considerado que a série temporal original é descrita pelo modelo ARIMA(p, d, q).

A metodologia dos modelos ARIMA, ou de Box-Jenkins, é aplicada dado o pressuposto de estacionaridade da série temporal em questão. Quando o modelo estimado por esse método for aplicado para previsões de valores futuros, devemos considerar que suas características são constantes ao longo do tempo (ver Gujarati e Porter (2011)).

Uma vez expostos os processos que compõem esse método, partimos à execução do mesmo. O processo de aplicação da metodologia Box-Jenkins se dá em quatro etapas: identificação, estimação, verificação e previsão. As ferramentas fundamentais para a identificação dos valores mais apropriados de p, d e q ao modelo estimados são os correlogramas obtidos através da função de correlação amostral (ACF) e da função de correlação amostral parcial (PACF). A função de correlação amostral ($\hat{\rho}_k$) é a razão entre a covariância amostral com defasagem k ($\hat{\gamma}_k$) e a variância da amostra ($\hat{\gamma}_0$). Os pares $(k, \hat{\rho}_k)$ denotam a função de autocorrelação entre y_t e y_{t-k}.

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (2)$$

$$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n} \quad (3)$$

$$\hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n} \quad (4)$$

O coeficiente de autocorrelação parcial de ordem k (ϕ_{kk}) mede a correlação entre y_t e y_{t-k} sendo descontada a influência de y_{t-1}, y_{t-2}, y_{t-3},..., y_{t-k-1}. Os pares (k, ϕ_{kk}) denotam a função de autocorrelação parcial. Como observado por Fava (2000), o valor de ϕ_{kk} é dado pelo último coeficiente das autoregressões, apresentadas adiante nas expressões (5), (6) e (7).

$$Y_t = \beta_{11} Y_{t-1} + u_t, \ \phi_{11} = \beta_{11} \quad (5)$$

$$Y_t = \beta_{11} Y_{t-1} + \beta_{22} Y_{t-2} + u_t, \ \phi_{22} = \beta_{22} \quad (6)$$

$$Y_t = \beta_{k1} Y_{t-1} + \beta_{k2} Y_{t-2} + \cdots + \beta_{kk} Y_{t-k} + u_t, \ \phi_{kk} = \beta_{kk} \quad (7)$$

2.2. Modelos SARIMA

Os modelos SARIMA, também designados como modelos ARIMA sazonal, são extensões do modelo ARIMA que considera a presença de sazonalidade na série temporal, representado como SARIMA(p,d,q)(P,D,Q), aonde P, D e Q representam as ordens sazonais de AR, I e MA, respectivamente. De acordo com Becker (2010), $\{X_t\}_{t \in Z}$ pode ser considerado um processo sazonal autoregressivo integrado de média móvel (p,d,q)(P,D,Q)s, onde d e D representam, respectivamente, o grau de diferenciação e o grau de diferenciação sazonal, se este processo estocástico atender à equação a seguir:

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^S)^D(X_t - \mu) = \Theta(B^s)\theta(B)\varepsilon_t \quad (8)$$

onde *B* é o operador de defasagem, de modo que:

$$B^j(X_t) = X_{t-j} \quad (9)$$

$$B^{sj}(X_t) = X_{t-sj} \tag{10}$$

e $(1-B)^d$ e $(1-B^S)^D$ são os operadores de diferença e diferença sazonal, respectivamente. Além disso, μ é a média do processo, $\{\varepsilon_t\}_{t\in Z}$ é um processo de ruído branco e s é a sazonalidade. São $\phi(.)$, $\theta(.)$, $\Phi(.)$ e $\Theta(.)$ polinômios de ordem p, q, P e Q, respectivamente, e são definidos por:

$$\phi(Z) = \sum_{l=0}^{p}(-\phi_l)z^l \tag{11}$$

$$\theta(Z) = \sum_{m=0}^{q}(-\theta_m)z^m \tag{12}$$

$$\Phi(Z) = \sum_{r=0}^{p}(-\Phi_r)z^r \tag{13}$$

$$\Theta(Z) = \sum_{m=0}^{Q}(-\Theta_m)z^m \tag{14}$$

onde ϕ_l, ($1 \le l \le p$), θ_m, ($1 \le m \le q$), Φ_r, ($1 \le r \le P$), Θ_m e ($1 \le l \le Q$) são constantes reais e $\phi_0 = \Phi_0 = \theta_0 = \Theta_0 = -1$.

2.3. Modelos Autoregressivos com Variáveis Dummy

Um modelo autoregressivo Y_t é descrito pela combinação linear de seus p valores passados. No presente trabalho foram elaborados dois modelos, considerando uma e duas defasagens da variável dependente, respectivamente, representados pelas equações (15) e (16), adiante.

$$Y_t = \alpha_1 Y_{t-1} + u_t \tag{15}$$

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + u_t \tag{16}$$

Esses modelos também só poderão ser considerados válidos dado o pressuposto de estacionaridade da série temporal. Nos modelos construídos foram incluídas, além das defasagens supramencionadas, variáveis *dummies*, a fim de captar o efeito qualitativo da sazonalidade na série estudada.

As variáveis binárias ou *dummies* são introduzidas em modelos de regressão nos quais cujos parâmetros podem ser diferentes para algumas observações, ou alguns conjuntos de observações, de uma amostra. Essas variáveis permitem a representação de fatores qualitativos no modelo, ou sazonais. Assume-se a variável igual a 1 para a presença de uma determinada característica de uma variável e igual a 0 para a ausência dessa característica. Desse modo, essas variáveis são "essencialmente um dispositivo para classificar dados em categorias mutuamente excludentes" (ver Gujarati e Porter (2011)). Uma vez introduzida no modelo, a variável binária é tratada como qualquer variável explicativa na estimação de um modelo estocástico. Desse modo, não há distinção em termos das propriedades dos estimadores de mínimos quadrados entre uma variável explicativa e uma variável *dummy*.

A robustez ou qualidade dos estimadores dos mínimos quadrados dependem de pressupostos básicos, para os resíduos ou termos estocásticos, que garantem melhores estimativas lineares não tendenciosas. Dentre esses pressupostos, os principais são: normalidade, homoscedasticidade e não autocorrelação. Assim foram realizados testes estatísticos de hipóteses para cada um desses três pressupostos para todos os modelos estimados neste trabalho. Para normalidade foi realizado o teste de normalidade de Jarque e Bera, para testar a violação da homoscedasticidade o teste de White foi utilizado enquanto para testar a violação da não autocorrelação dos resíduos o teste utilizado foi o de Durbin e Watson. Todos esses três testes estão detalhados em Gujarati e Porter (2011). No que tange os modelos ARIMA e SARIMA, dada a análise dos correlogramas das funções de autocorrelação e de autocorrelação parcial dos

resíduos, se estes resultados não forem significativos estatística e individualmente, pode-se aceitar que os resíduos gerados por tal modelo são aleatórios (ver Gujarati e Porter (2011)).

A fim de selecionar o melhor, dados os vários modelos de previsão elaborados são verificadas algumas medidas de ajustamento. Descritas por Salles (2005), essas medidas figuram como critérios de comparação do ajuste de determinado modelo aos dados de uma série histórica e podem ser de natureza absoluta ou relativa. A seguir são apresentadas as medidas de ajuste que serão aplicadas na comparação dos modelos deste trabalho, onde $\widehat{Y}_i$ representa a informação i da série temporal, ou a i-ésima observação de n observações e $\widehat{Y}_i$ representa cada uma das estimativas das informações da série temporal, ou a i-ésima estimativa da série temporal.

3. Amostra Utilizada – Caracterização dos Dados

3.1. Série de Faturamento

Inicialmente procedeu-se o estudo da série temporal do faturamento anual do setor editorial de livros brasileiro de 1995 a 2016 disponibilizados anualmente pela Pesquisa Produção e Vendas do Setor Editorial Brasileiro, realizada pela Fundação Instituto de Pesquisas Econômicas (FIPE), Sindicato Nacional de Editores de Livros (SNEL) e Câmara Brasileira de Livros (CBL). Foram encontradas dificuldades na obtenção dos registros anteriores ao ano de 2006, tendo estes sido recuperados em fontes diversas. Esses dados são provenientes de uma amostra de empresas cadastradas que fornecem informações referentes a seus principais indicadores de desempenho por meio da resposta de um questionário eletrônico. Desse modo, a Pesquisa Produção e Vendas do Setor Editorial Brasileiro é "uma estimativa, elaborada a partir dos dados de uma amostra de editoras, da performance do setor editorial e de seus subsetores num determinado ano" (FIPE, 2016). Os registros de 2006 e posteriores encontram-se disponíveis no site da SNEL Os dados aqui utilizados foram dessa série histórica com os valores deflacionados pelo índice IPCA, considerando os preços constantes de 2016, e estão mostrados no gráfico apresentado na Figura 1, a seguir.

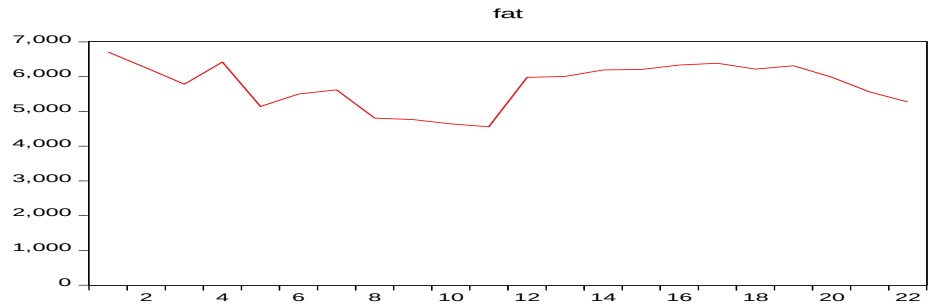


Figura 1: Série do faturamento anual do setor editorial de livros brasileiro.

Além da série temporal original, designada a partir daqui por *fat*, também foram analisadas a série formada por sua primeira diferença e pelo seu retorno logarítmico, respectivamente designadas neste trabalho por *dfat* e *lfat* e obtidas através das fórmulas a seguir.

$$dfat = fat_t - fat_{t-1} \tag{17}$$

$$lfat = ln\left(\frac{fat_t}{fat_{t-1}}\right) \tag{18}$$

A Tabela 1 apresenta os resumos estatísticos das séries supramencionadas. As séries de primeira diferença (*dfat*) e de diferença logarítmica (*lfat*) apresentam desvio padrão relativamente alto dadas suas respectivas médias. Essa relação é mais moderada na série de faturamento original (*fat*). Pode-se observar na Tabela 1 que tanto os coeficientes de assimetria quanto os coeficientes de curtose das três séries de faturamento diferem das de uma distribuição normal, o que é confirmado, em parte, pelo teste de normalidade aqui utilizado. De acordo com os resultados dos testes Jarque Bera realizados, ao nível de significância de 5%, a hipótese nula de

normalidade pode ser rejeitada apenas para a série *lfat*. Na Tabela 1, também, estão apresentados os resultados das inferências dos testes de estacionariedade realizados no software Eviews, utilizando o teste de Dickey Fuller Aumentado, teste ADF. O número de defasagens utilizado foi determinado pelo critério de Akaike, assim como os parâmetros a serem estimados. Ao nível de significância de 5%, não se pode rejeitar a hipótese de estacionariedade para as séries *dfat* e *lfat*. Para a série original de faturamento, a hipótese de estacionariedade pode ser rejeitada.

Tabela 1: Resumo estatístico da série de faturamento, sua primeira diferença e retorno logarítmico.

Estatística	*fat*	*dfat*	*lfat*
Média	5755,23	-68,37	-0,01147
Desvio Padrão	643,77	533,02	0,09651
Assimetria	-0,57047	0,47012	0,65919
Curtose	2,08102	5,02702	5,49719
Jarque-Bera	1,9674	4,3687	6,9773
(p valor)	(0,3739)	(0,1126)	(0,0305)
ADF – valor p	0,0000	0,3507	0,0000
Nº de lags	0	0	0
Observações	22	21	21

3.2. Série de Vendas

Procurando aprofundar análise acerca as tendências do setor além das séries de faturamento, mencionadas anteriormente, também foi analisada a série temporal de "Livros, jornais, revistas e papelaria", gerada pela Pesquisa Mensal do Comércio (PMC) e designada neste trabalho como série de vendas. Essa pesquisa é realizada pelo IBGE e tem como objetivo a produção de indicadores para acompanhamento do comportamento conjuntural do comércio varejista no país. Com periodicidade mensal, esse indicador contempla o período de janeiro de 2003 a maio de 2017, reportando índices de base fixa do setor do comércio varejista supramencionado. Esses índices comparam os níveis nominal e de volume da Receita Bruta de Revenda do mês com a média mensal obtida no ano base de 2016. Embora também abranja as vendas de jornais, revistas e papelarias, essa série representa um forte indicador das tendências de crescimento do mercado editorial, estando representada no gráfico apresentado na Figura 2, a seguir.

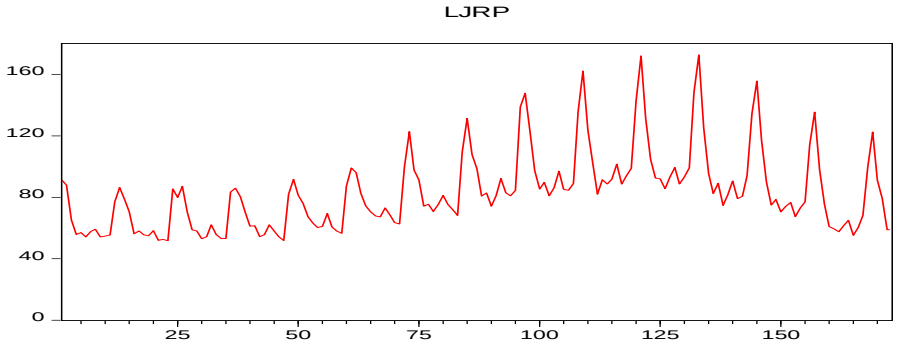


Figura 2: Série da atividade varejista dos setores de "Livros, jornais, revistas e papelaria".

Além da série original de vendas designada por LJRP, também foram analisadas as séries formadas por seus retornos logarítmicos *LJRP_LN1* e *LJRP_LN2*, calculadas pelas expressões (19) e (20), respectivamente, e suas respectivas primeiras diferenças dadas por *LJRP_DIF*, *LJRP_LN1_DF* e *LJRP_LN2_DIF*. Além dessas séries foi utilizada a série da segunda diferença designada por *LJRP_DIF2*. Desse modo, foram realizados resumos estatísticos e testes de estacionariedade para todos os retornos e variações supramencionados e os resultados estão listados na Tabela 2, a seguir.

$$LJRP_{LN1} = ln\left(\frac{LJRP_t}{LJRP_{t-1}}\right) \quad (19)$$

$$LJRP_{LN2} = ln\left(\frac{LJRP_t}{LJRP_{t-1} + LJRP_{t-2}}\right) \quad (20)$$

Tabela 2: Resumo estatístico da série de vendas, sua primeira diferença e retorno logarítmico.

Estatística	*LJRP*	*LJRP_DIF*	*LJRP_DIF2*	*LJRP_LN1*	*LJRP_LN1_DIF*	*LJRP_LN2*	*LJRP_LN2_DIF*
Média	82,8549	-0,1884	0,0187	-0,0026	0,0002	-0,7006	0,0009
Desvio Padrão	25,2549	16,9477	20,8206	0,1748	0,2170	0,2121	0,2165
Assimetria	1,2902	0,5049	-0,8346	0,8506	-0,1358	0,6208	-0,2728
Curtose	4,7476	4,3621	4,8863	3,9541	3,0894	3,0605	3,9068
Jarque-Bera	70,0139	20,6042	45,2026	27,2649	0,5823	11,0099	7,9325
(p-value)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,7474)	(0,0041)	(0,0189)
ADF – valor p	0,4520	0,0893	0,0000	0,1553	0,0000	0,1272	0,0000
Nº de lags	13	12	11	12	12	11	11
Observações	173	172	171	172	171	171	170

Na Tabela 2, acima, pode-se observar que os desvios padrão de todas as séries analisadas são significativamente altos em relação às suas respectivas médias. O comportamento dos coeficientes de assimetria e curtose se mostram bastante variados. Analogamente ao exposto no tópico anterior, as diferenças nos números de observações ocorrem devido aos tratamentos aplicados na série original. Embora tenham sido realizadas transformações logarítmicas e diferenciações, ainda assim não foi possível aceitar a hipótese de normalidade para a maior parte das séries analisadas, dados os resultados dos testes Jarque Bera, exceto para série *LJRP_LN1_DIF* não se rejeita a hipótese nula de normalidade. Na Tabela 2, também, estão apresentados os resultados dos testes de estacionariedade de Dickey Fuller Aumentado e como foi realizado na série de faturamento, o número de defasagens utilizado foi determinado pelo critério de Akaike. Não podemos rejeitar a hipótese de estacionariedade somente para as séries *LJRP_DIF2*, *LJRP_LN1_DIF* e *LJRP_LN2_DIF*.

4. Análise dos Resultados Obtidos

4.1. Série de Faturamento

A série *lfat* não rejeita a hipótese de estacionariedade e foi objeto da metodologia de Box Jenkins. A fim de determinar a classe e a ordem do modelo, bem como definir seus coeficientes apropriados, foram analisadas as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP). A Tabela 3, a seguir, apresenta os três melhores modelos estimados e algumas das principais medidas de ajuste, a saber: o coeficiente de explicação ou determinação (R^2), o critério de Akaike (AIC) e o erro padrão dado (SE). Na Tabela 3 pode-se observar que considerando o R^2 e o R^2 ajustado como critério, nenhum dos modelos aparenta excelente qualidade de ajustamento, embora o melhor resultado tenha sido obtido com o modelo ma(3) ma(4) ma(7), uma vez que este modelo apresenta um mínimo AIC e um mínimo SE. No que se refere aos testes de violação das hipóteses normalidade, não autocorrelação e homocedasticidade observou-se que a hipótese de normalidade dos resíduos não foram rejeitadas ao nível de significância de 5% em nenhum dos modelos analisados. E quanto os resultados dos testes de violação dos pressupostos básicos de homocedasticidade e de não autocorrelação para estimação dos modelo pode-se inferir que a hipótese nula de homocedasticidade, ou ausência de heterocedasticidade, foi rejeitada para todos os modelos de acordo com o teste de White, ao nível de significância de 5%. A hipótese de ausência de autocorrelação serial dos resíduos foi rejeitada para todos os modelos. Assim, os resultados dos testes de violação dos pressupostos básicos, além do teste de estacionariedade não mostram a robustez dos resultados obtidos pelos modelos de previsão do faturamento do setor aqui apresentados.

Tabela 3: Resumo dos modelos estimados para série de faturamento

Modelo	R^2	AIC	SE
ar(3) ma(3) ma(4) ma(7)	0,3343	-1,7049	0,0880
ma(3) ma(4) ma(7)	0,4075	-1,8429	0,0806
ma(3) ma(4) ma(7) ma(9)	0,4116	-1,7813	0,0828

4.2. Série de Vendas

4.2.1. Modelos SARIMA

Conforme mencionado anteriormente, a série de vendas*LJRP_LN1_DIF* não rejeita a hipótese de estacionariedade e de normalidade ao nível de significância de 5%. O gráfico dessa série temporal encontra-se na Figura 3, abaixo.

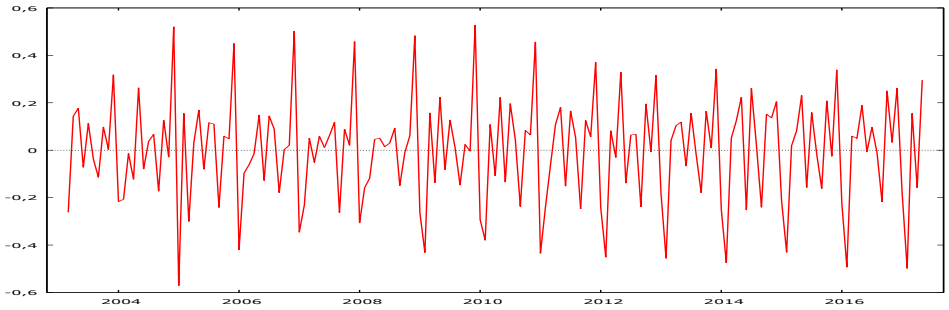


Figura 3: Gráfico da série *LJRP_LN1_DIF*.

Dada a presença de uma componente sazonal foram construídos modelos SARIMA para previsão dessa série temporal. Uma alternativa à desassonalização é a aplicação de modelos que considerem sazonalidade, como o SARIMA. Assim, foram analisadas as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), considerando o comportamento das correlações nas defasagens sazonais. A Tabela 4, a seguir, mostra os resultados dos três melhores modelos estimados através das suas principais medidas de ajuste. Cabe destacar que o modelo SARIMA que apresentou as melhores medidas de ajuste foi ar(1) ma(1) sar(12) sma(12), onde o coeficiente de explicação foi máximo e as medidas AIC (critério de Akaike) e SE (erro padrão) foram mínimas. O gráfico apresentado na Figura 5 mostra os resultados das previsões obtidas para vendas com esse modelo, ou para o modelo MVS2 conforme referenciado na Tabela 4. Pode-se inferir que a hipótese de estacionariedade dos resíduos não é rejeitada em nenhum dos modelos analisados. Apenas no terceiro modelo, o modelo MVS3 a hipótese de normalidade dos resíduos não é aceita ao nível de significância de 5%. Quanto aos testes de violação dos pressupostos básicos de homocedasticidade e de não autocorrelação dos resíduos dos modelos SARIMA estimados. Pode-se inferir que a hipótese nula de homocedasticidade dos resíduos foi rejeitada para todos os modelos, ao nível de significância de 5%. Para a detecção da presença de autocorrelação foram analisadas em conjunto as estatísticas de Durbin Watson. A Tabela 5, a seguir, apresenta os valores observados e estimados do modelo MVS2 para as os meses de junho/17 a setembro/17. Os resultados dos testes de violação dos pressupostos básicos, além do teste de estacionariedade, não indicam a robustez dos resultados obtidos pelos modelos SARIMA de previsão de vendas do setor.

Tabela 4: Resumo dos modelos estimados para série de vendas

	Modelo	R^2	AIC	SE
MVS1	ar(1) ma(1) sar(12)	0,9421	-2,8172	0,0527
MVS2	ar(1) ma(1) sar(12) sma(12)	0,9460	-2,8699	0,0510
MVS3	ar(1) ma(2) sar(12) sma(12)	0,9335	-2,6989	0,0567

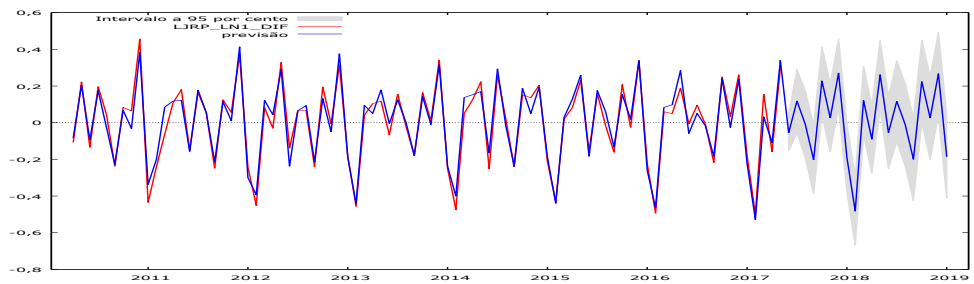


Figura 4: Gráfico do Modelo 2 – ar(1) ma(1) sar(12) sma(12).

Tabela 2: Modelo MVS2 – Valores Observados vs ;Estimados.

Mês	Observado	Estimado	Erro Padrão	LIC	LSC
jun/17	-0,0120	-0,0546	0,0503	-0,1531	0,0440
jul/17	0,0739	0,1186	0,0881	-0,0540	0,2913
ago/17	-0,0505	-0,0091	0,0936	-0,1925	0,1744
set/17	-0,1927	-0,2021	0,0946	-0,3875	-0,0166

4.2.2. Modelos Autoregressivos com Variáveis Dummy

Considerando a série de vendas *LJRP_LN1_DIF* foram elaborados dois modelos autoregressivos: com uma e duas defasagens, respectivamente. Além disso, esses modelos contam com a presença de onze variáveis dummy, representando os meses do ano. Ressalta-se que deve ser incluída sempre uma variável binária a menos (n-1) dado o total de variáveis qualitativas que serão incluídas no modelo (n). No nosso exemplo, o impacto do décimo segundo mês já será captado pelo modelo quando as onze variáveis dummy incluídas assumirem o valor zero, sem que seja necessário incluir outra variável dummy no modelo (o que seria incorreto).

Os resultados para ambos os modelos encontram-se na Tabela 6, a seguir. O modelo com duas defasagens apresentou as melhores medidas de ajuste, tendo o coeficiente de explicação sido máximo e as medidas: AIC critério de Akaike (AIC) e erro padrão (SE) mínimas. O gráfico de ajuste e previsão do modelo com 2 defasagens, referenciado no Tabela 6 como o modelo MARD2, está apresentado adiante na Figuras 6. Na Tabelas 7 pode-se verificar que os valores observados e estimados para as os meses de junho/17 a setembro/17 obtidos com o modelo com duas defasagens ou o modelo MARD2.

Tabela 6: Resumo dos modelos estimados – AR com variáveis dummies

	Modelo	R^2	AIC	SE
MARD1	1 Defasagem	0,8825	-2,2146	0,0771
MARD2	2 Defasagens	0,8942	-2,3031	0,0735

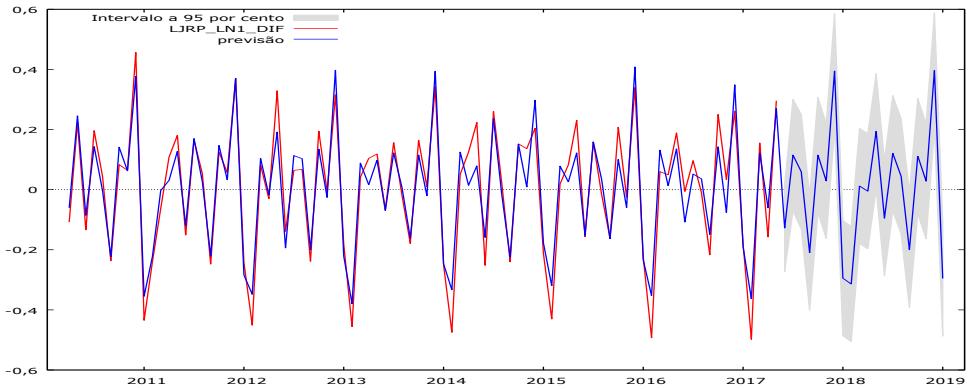


Figura 6: Gráfico do Modelo MARD2 – 02 Defasagem

Tabela 3: Modelo MARD2 modelo com 2 defasagens – Valores observados x estimados.

-	Observado	Estimado	Erro Padrão	LIC	LSC
jun/17	-0,0120	-0,1278	0,0735	-0,2731	0,0174
jul/17	0,0739	0,1151	0,0938	-0,0703	0,3005
ago/17	-0,0505	0,0593	0,0966	-0,1316	0,2502
set/17	-0,1927	-0,2100	0,0966	-0,4009	-0,0191

Em ambos os modelos rejeitaram as hipóteses de normalidade e de homocedasticidade ao nível de significância de 5% e não rejeitaram as hipóteses de estacionariedade para o mesmo nível de significância. Como, também, não foram identificados padrões notáveis nas séries de resíduos, tendo a hipótese de ausência de autocorrelação sido rejeitada para ambos os modelos.

A Tabela 10, a seguir, os resultados dos modelos de previsão das séries de faturamento através das principais medidas de performance. Considerando as medidas absolutas o segundo modelo, ou modelo MF2, é superior ao demais, mas considerando as medidas relativas o modelo MF3, o terceiro modelo se mostra mais adequado. Desse modo, embora não se possa apontar um modelo por consenso, pode-se observar que as medidas absolutas de ambos os modelos estão muito próximas, o que não necessariamente ocorre quando comparamos as medidas relativas, onde o modelo MF2, ou o modelo 2, apresenta vantagem mais significativa. No que se refere aos pressupostos básicos testados, todos atendem aos critérios de normalidade e estacionariedade dos resíduos porém não atendem aos critérios de homocedasticidade e ausência de autocorrelação serial. Na presença de autocorrelação serial, alguns resultados de ajuste dos modelos estimados podem não ser confiáveis e na presença de heterocedasticidade nos resíduos, as inferências acerca os estimadores podem não ser ótimos ou sem robustez.

Tabela 4: Comparativo dos modelos da série de faturamento – Medidas de ajuste.

Modelo de Faturamento		Medidas Absolutas			Medidas Relativas	
		MAD	MSE	SDE	MPE	MAPE
MF1	lfat ar(3) ma(3) ma(4) ma(7)	0,0633	0,0060	0,0774	-0,4788	3,0099
MF2	lfat ma(3) ma(4) ma(7)	0,0563	0,0053	0,0725	0,3429	2,1478
MF3	lfat ma(3) ma(4) ma(7) ma(9)	0,0553	0,0051	0,0717	0,4802	2,4054
Melhor Modelo		3	3	3	2	2

Como pode ser observado na Tabela 11, nenhum modelo para série de vendas tem performance superior em todas as medidas de ajuste. O segundo modelo, ou o modelo referenciado por MV2 possui melhor ajuste considerando as medidas absolutas, mas não há consenso dadas as medidas relativas. No entanto, observamos que os valores do MPE e do MAPE do segundo modelo também estão muito próximos dos melhores ajustes dessas categorias, respectivamente. No que tange ao atendimento aos pressupostos básicos que garantem estimativas ótimas dos cinco modelos os modelos MV1 e MV2 não violam os pressupostos da normalidade enquanto somente o modelo MV3 viola o pressuposto da não autocorrelação dos resíduos, o pressuposto da homoscedasticidade é violada por todos os modelos.

Tabela 5: Comparativo dos modelos da série de vendas – Medidas de ajuste.

Modelo de Vendas		Medidas Absolutas			Medidas Relativas	
		MAD	MSE	SDE	MPE	MAPE
MV1	ar(1) ma(1) sar(12)	0,0398	0,0027	0,0521	0,0324	0,7132
MV2	ar(1) ma(1) sar(12) sma(12)	0,0389	0,0025	0,0503	-0,0346	0,7414
MV3	ar(1) ma(2) sar(12) sma(12)	0,0430	0,0031	0,0558	-0,0300	0,8355
MV4	AR com Dummy - 01 Defasagem	0,0589	0,0055	0,0741	-0,1617	1,2511
MV5	AR com Dummy - 02 Defasagens	0,0556	0,0050	0,0704	-0,2030	1,1294
	Melhor Modelo	2	2	2	3	1

5. Conclusão e Comentários Finais

O presente trabalho teve como principal objetivo determinar e as perspectivas do mercado editorial brasileiro dado modelos de previsão de séries temporais construídos a partir de duas séries distintas: de faturamento e de vendas.

A série de faturamento apresentou um número de observações significativamente menor que a série de vendas, com frequência anual. Por não se tratar de uma série estacionária, os modelos de previsão foram construídos utilizando-se a variável ou série transformada em uma série de retorno logarítmico. Os três modelos ARIMA elaborados apresentaram valores de coeficientes de explicação ajustado pequenos e os gráficos dos valores previstos se mostraram pouco ajustados aos valores reais. Nenhum modelo atendeu aos pressupostos de não autocorrelação e homocedasticidade dos resíduos. Dada a comparação das medidas de ajuste o modelo ARIMA (0,1,7) se mostrou ligeiramente superior aos demais.

No que se refere a série original de vendas, de periodicidade mensal, o critério de estacionariedade também não foi atendido. Desse modo, os modelos foram construídos a partir da série da primeira diferença do seu retorno logarítmico, que atendeu às hipóteses de normalidade e estacionariedade. Devido a presença de sazonalidade, foram construídos modelos SARIMA e autoregressivos com variáveis *dummies*. Nesse caso, o ajuste das previsões aos valores reais foi bastante expressivo e os valores dos coeficientes de explicação ajustados superaram 0,9 para todos os modelos. Dada a comparação das medidas de ajuste e a interpretação desses valores, o modelo SARIMA $(1,1,1)(1,0,1)_{12}$ foi considerado superior aos demais. Embora seus resíduos tenham atendido aos testes das hipóteses de normalidade, estacionariedade e não autocorrelação, violam o pressuposto de homocedasticidade.

Desse modo, uma perspectiva futura para continuidade do estudo desse tema seria a exploração de modelos heterocedásticos de séries temporais, dado que nenhum dos modelos analisados atendeu ao pressuposto de homocedasticidade de resíduos. Como sugestão para futuros trabalhos que possam ser desenvolvidos a partir desta pesquisa tem-se a utilização dos modelos da família ARCH, o que deve resultar em modelos mais robustos e com melhores performances.

Referências

Bastos, L. Fleck, D. (2014) Indústria Editorial de Livros no Brasil: análise panorâmica se seu crescimento. Rio de Janeiro: Coppead.

Becker, M. (2010) Modelos para previsão em séries temporais: uma aplicação para a taxa de desemprego na Região Metropolitana de Porto Alegre. Trabalho de Conclusão de Graduação – Universidade Federal do Rio Grande do Sul.

Earp, F. Kornis, G. (2005) A economia da cadeia produtiva do livro. Rio de Janeiro: BNDES.

Fava, V. L. (2000) Metodologia de Box-Jenkins para Modelos Univariados. In: Vasconcelos, M. A., Alves, D. Manual de Econometria. São Paulo: Atlas.

Fonseca, L. (2013) Crescimento da indústria nacional de livros do Brasil e seus desafios. 232 f. Tese (Mestrado) – Instituto Coppead de Administração, Universidade Federal do Rio de Janeiro.

Fundação Instituto de Pesquisas Econômicas (FIPE) (2017) Desempenho do mercado livreiro: uma análise de 11 anos da pesquisa produção e vendas do setor editorial brasileiro. São Paulo.

Gujarati, D. Porter, D. (2011) Econometria Básica (5ª ed.). Porto Alegre: AMGH Editora.

Hallewell, L. (2012) O Livro no Brasil. Sua História (3ª ed.). São Paulo: Edusp.

Korth, L. (2005) As transformações na indústria editorial de livros no Brasil e os desafios para as empresas brasileiras. 74 f. Dissertação (Metrado) – Escola Brasileira de Administração Pública e de Empresas, Fundação Getúlio Vargas.

Salles, A.A. (2007) Notas de aula – Series Temporais - Previsão - Texto Introdutório. Rio de Janeiro: Escola Politécnica, Universidade Federal do Rio de Janeiro.