Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Engenharia de Informação

Luciano Henrique Lacerda de Araújo

Modelagem e Previsão de Séries Temporais Financeiras

Trabalho de Graduação III

Santo André – SP

Novembro de 2019

Modelagem e Previsão de Séries Temporais Financeiras

Luciano Henrique Lacerda de Araújo

Relatório submetido como requisito parcial para obtenção do grau de bacharel em Engenharia de Informação

Orientado por Prof. Dr. Claudio José Bordin Júnior

Santo André – SP

Novembro de 2019

Resumo

A extração de informação de uma série temporal - coleção de amostras de valores medidos de um fenômeno no tempo - tem seu estudo baseado no fato de que a natureza de alguns fenômenos permite tomar por hipótese que a amostra observada em um instante é dependente das amostras observadas nos instantes anteriores. Determinar como e quantas medições passadas influenciam, significativamente, a medição atual permite a obtenção de um ou mais modelos matemáticos que forneçam uma descrição razoável do comportamento de uma série. Um modelo adequado permite a estimação de valores futuros da série baseado em informações dos valores passados, possibilitando, portanto, a sua previsão. Este trabalho, com inspiração inicial na teoria de processamento de sinais, busca comparar, a partir de métricas encontradas na literatura, a capacidade de modelos clássicos como ARIMA e GARCH e de modelos de aprendizado de máquina baseados em Redes Neurais Artificiais nas tarefas de ajuste e previsão de séries temporais financeiras utilizando a linguagem MATLAB.

Sumário

1. Introdução

A observação de um fenômeno ou processo, se acompanhada do registro da variação de suas propriedades no decorrer do tempo, permite a construção de uma sequência para cada variável relacionada, constituindo o que, devido à ordenação cronológica (contínua ou discreta) dos dados, denominam-se séries temporais. Devido à natureza dos fenômenos de interesse, os valores da série obtidos até um determinado momento podem fornecer informações sobre seu comportamento futuro. Por exemplo, na observação de fenômenos que estão relacionados a eventos periódicos, podem-se obter informações sobre o valor a ser medido na próxima ocorrência apenas observando-se a medição nos instantes de ocorrências anteriores. O futebol possui eventos periódicos: entre outros fatos, pode-se perceber, pelo gráfico da Figura 1, que o interesse é semelhante no sábado (6 de abril) à noite e no domingo (7 de abril) à tarde, portanto, horários em que ocorrem as partidas de maior audiência; assim, a medição da quantidade de buscas no *Google* no sábado à noite é uma estimativa da medição a ser obtida no domingo à tarde.

Figura 1 - Gráfico (taxa x tempo) da taxa normalizada de buscas no *Google* pelo termo "futebol" feitas no Brasil a cada hora, durante o período compreendido entre os dias 1 e 8 de abril de 2019 [1].

Seja uma série temporal discreta, de duração $n+1$, descrita por um conjunto $\{X(t),\ t \in T\}$, onde o conjunto $T = \{1,...,k,...n\}$ é a coleção dos instantes de medição e $X(k)$ é um valor real medido no instante $k \in T$. A influência de $X(t-1)$, $X(t-2)$, ..., $X(1)$ em $X(t)$ pode ser aproximada por um modelo matemático tanto mais útil quanto melhor este minimize a métrica de erro de ajuste pré-determinada. Uma modelagem com baixa taxa de erros pode possuir, entre outras utilidades, a de previsão, ou seja, ser capaz de estimar razoavelmente os valores futuros da série. O modelo mais adequado depende

das características da série que, por sua vez, depende da natureza do fenômeno ou processo analisado.

Os processos estudados neste presente trabalho são preços de ações advindas da Bolsa de Valores. A escolha desta área foi fundamentada em uma das principais abordagens que teorizam o mercado de ações: a análise técnica assume que o valor do preço de uma ação segue uma tendência cujo comportamento pode ser estimado observando-se o comportamento no passado [2].

2. Fundamentação teórica

2.1 Teoria de processamento de sinais

O sinal de saída (por exemplo, sinal elétrico ou mecânico) de um sistema linear contínuo e invariante no tempo (LCIT), cuja entrada foi um sinal contínuo, pode ser obtido analisando-se o modelo matemático do sistema, no caso LCIT, uma equação diferencial linear de coeficientes constantes que relaciona as funções que modelam os sinais de entrada, *x(t)*, e de saída, *y(t)*. A análise parte do pressuposto que qualquer entrada pode ser representada como combinação linear de impulsos unitários, $\delta(t)$. Logo, devido às propriedades de linearidade e invariância no tempo, a saída correspondente à entrada impulso unitário, denominada resposta ao impulso, *h(t)*, caracteriza o sistema e permite que, conhecido o sinal de entrada, o sinal de saída possa ser calculado pela integral de convolução,

$$y(t) = x(t) * h(t) := \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau. \quad (1)$$

Todavia, a análise pode ser simplificada quando levada para o domínio da frequência [3]. Pode ser demonstrado que sinais do tipo exponencial complexa (e^{st}, onde *s* é um valor complexo e representa a frequência do sinal) são as únicas autofunções de sistemas LCIT, ou seja, a saída é o produto da entrada por uma constante. Uma função qualquer, *f(t)*, pode ser expressa como a soma ponderada de exponenciais complexas de duração infinita. O valor que pondera a exponencial de cada frequência *s* é dado pela função *F(s)*, obtida através da transformada de Laplace de *f(t)*,

$$L\{f(t)\} := \int_{-\infty}^{\infty} f(t)\, e^{-st} dt = F(s), s \in C \quad (2)$$

definida para todo *s* no plano complexo tal que a integral não divirja (região de convergência, RDC). A simplificação mencionada anteriormente advém do teorema que diz que a transformada da função resultante da convolução entre duas funções é igual ao produto das transformadas das respectivas funções [3]:

$$Y(s) = L\{y(t)\} = L\{x(t) * h(t)\} = L\{x(t)\}L\{h(t)\} = X(s)H(s). \quad (3)$$

Assim, o sistema pode ser caracterizado pela sua função de transferência, *H(s)*, uma função que, para cada frequência *s*, calcula o autovalor correspondente à autofunção e^{st},

$$H(s) = \frac{Y(s)}{X(s)}, \quad (4)$$

e o sinal *y(t)* pode ser obtido tomando a transformada de Laplace inversa de *Y(s)*, através, por exemplo, do método das frações parciais, dispositivo prático para qualquer *Y(s)* descrito por uma função racional [3].

Em processamento digital de sinais, são estudados sinais discretos no tempo, cuja característica advém da natureza do processo apreciado ou da amostragem de um sinal contínuo no tempo. Um sistema linear discreto e invariante no tempo (LDIT), de forma análoga aos sistemas LCIT, é descrito por uma equação de diferenças que relaciona o sinal de tempo discreto de entrada, *x[n]*, com o sinal de tempo discreto de saída, *y[n]*, por meio de uma soma ponderada das amostras no instante com as amostras atrasadas até determinado instante passado,

$$y[n] + a_1 y[n-1] + \cdots + a_N y[n-N] - b_0 x[n] - b_1 x[n-1] - \cdots - b_M x[n-M] = 0, \quad (5)$$

sendo dita de ordem max(N,M). Neste caso, o sistema LDIT pode ser caracterizado pela sua resposta ao impulso unitário, *h[n]*, enquanto o sinal de saída, conhecido o sinal de entrada, pode ser determinado pelo somatório de convolução,

$$y[n] = x[n] * h[n] := \sum_{m=-\infty}^{\infty} x[m]h[n-m]. \quad (6)$$

Além disso, a análise no domínio da frequência é possibilitada pela transformada *Z*, para sequências, *f[n]*,

$$Z\{f[n]\} := \sum_{n=-\infty}^{\infty} f[n]z^{-n} = F[z], z \in C, \tag{7}$$

definida para todo *z* no plano complexo tal que a soma não divirja. Tal análise se vale do fato de que, para sistemas LDIT, sinais da forma exponencial infinita (partindo do análogo contínuo),

$$f[n] = e^{sn} = (e^{s})^{n} = z^{n}, ondes, z \in C, \tag{8}$$

são as únicas autofunções, e decompõe o sinal de entrada na soma ponderada de exponenciais de duração infinita. O autovalor para cada componente de frequência *z* é dado por *F[z]*.

Utilizando-se a transformada Z unilateral para sinais causais (*x[n]* = 0, para todo $n < 0$), considerando apenas a resposta de estado nulo (i.e., *y[n]* = 0, para todo $n < 0$), e tomando o teorema de deslocamento no tempo,

$$Z\{x[n-m]\} = z^{-m}Z\{x[n]\} = z^{-m}X[z], \tag{9}$$

bem como as transformadas *Z*

$$Z\{k(\gamma)^{n}u[n]\} = k\sum_{n=0}^{\infty}\left(\frac{z}{\gamma}\right)^{-n} = k\frac{z}{z-\gamma}, |z| > |\gamma|, \tag{10}$$

$$Z\{k\delta[n]\} = k, \tag{11}$$

em que γ e *k* são constantes, é possível transformar a equação de diferenças (Eq. 5) em uma equação algébrica. Aplicando a transformada *Z* em ambos os lados e usando a propriedade da linearidade,

$$Y[z] + \sum_{i=1}^{N} a_i z^{-i}Y[z] - \sum_{l=0}^{M} b_l z^{-l}X[z] = 0, \tag{12}$$

encontraremos a função de transferência para o sistema,

$$H[z] = \frac{Y[z]}{X[z]}. \tag{13}$$

Como a saída, *Y[z]*, no domínio da frequência, pode ser calculada como o produto da entrada em frequência pela função de transferência, e, dependendo da estrutura do sistema, *H[z]* pode apresentar valores muito próximos ao zero para determinados valores de *z*, o sistema (praticamente) elimina essas determinadas componentes da saída; devido a este efeito no domínio da frequência, denominamos o sistema como filtro digital. Tais filtros podem ser divididos entre os que a saída depende apenas da entrada e de um número finito de amostras atrasadas da entrada (filtro FIR – *finite impulse response*),

$$H[z] = \sum_{l=1}^{M} b_l z^{-l} \Leftrightarrow h[t] = \sum_{l=0}^{M} b_l \delta[n-l], \tag{14}$$

e os sistemas cujas saídas dependem das amostras atrasadas da entrada e da saída. Neste último caso, é comum a representação na forma de polos, p_i, e zeros, z_l (Eq. 15). Caso o grau do polinômio denominador seja maior que o grau do numerador (i.e., $N > M$), *H[z]* pode ser escrita, desde que não existam polos repetidos, como uma função polinomial de infinitos termos (Eq. 27), logo, a resposta ao impulso é infinita (filtro IIR – *infinite impulse response*) [4].

$$H[z] = \frac{\sum_{l=0}^{M} b_l z^{-l}}{1 + \sum_{i=1}^{N} a_i z^{-i}} = H_0 \frac{\prod_{l=1}^{M} 1 - z^{-1} z_l}{\prod_{i=1}^{N} 1 - z^{-1} p_i} = \sum_{i=1}^{N} \frac{k_i}{z - p_i} \Leftrightarrow h[n] = \sum_{i=1}^{N} k_i (p_i)^n \, u[n]. \tag{15}$$

Se $M \geq N$, no entanto, esta técnica não pode ser diretamente aplicada: neste caso, através da divisão polinomial, pode-se transformar *H[z]* na soma de uma componente FIR com uma função racional própria (i.e., com $M < N$), e em seguida utilizar a mesma técnica. Para a antitransformada *h[n]* ter valores finitos, os valores dos polos devem estar dentro do círculo de raio unitário; sistemas com tal característica são ditos sistemas LDIT estáveis [4].

Quando uma série temporal é observada no mundo real, o mecanismo de geração da mesma é em geral desconhecido. Assim, tal série pode ser interpretada como uma realização de um processo estocástico. Uma modelagem usual para tais processos é descrevê-los como a saída de um sistema LDIT estável excitado por ruído branco (i.e., um processo i.i.d., de amostras independentes e identicamente distribuídas). A resposta ao impulso de tal sistema LDIT pode então ser estimada analisando-se as propriedades estatísticas das amostras recebidas, com a utilização, por exemplo, da equação de Yule-Walker [5].

2.2 Modelagem e previsão de séries temporais

Um processo estocástico de tempo discreto é uma família $X = \{X(t,\omega),\ t \in T$ e $\omega \in \Omega\}$, onde T é um subconjunto do conjunto dos números naturais e Ω um espaço amostral. Para cada instante, $t \in T$, fixado, $X(t,\omega) \equiv X(\omega)$ é uma variável aleatória definida sobre o espaço amostral Ω. Para cada realização, $\omega \in \Omega$, fixada, $X(t,\omega) \equiv X(t)$ é uma função de t denominada série temporal [5]. Como o processo estocástico em um instante $\{X(t),\ t \in T\}$ é uma variável aleatória (de agora em diante, representado pela notação simplificada X_t), pode-se buscar a distribuição conjunta de $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$; na prática, só se obtém uma realização do processo, de tal modo, devem-se supor algumas restrições: homogeneidade temporal (distribuição conjunta invariante a translações) e limitação na memória do processo (correlação entre amostras diminui com o aumento do intervalo entre os instantes de tempo) [6]. Pode-se supor, também, que o processo é, ao menos, fracamente estacionário: um processo estocástico é dito estacionário se todas as distribuições de dimensões finitas forem alheias a translações no tempo e é dito fracamente estacionário se, e somente se, sua média temporal for constante, sua variância estatística for finita para todo t no domínio e a covariância entre duas amostras for função somente da distância entre estas [6].

Sob tais hipóteses, podem-se utilizar técnicas inspiradas em processamento de sinais para tratar tais séries, atribuindo às mesmas modelos como os de *média móvel*, sob os quais as séries são modeladas com as saídas de filtros FIR excitados por processos i.i.d., ou *autorregressivos* (filtros IIR). Uma vez assumido um mecanismo gerador para a série, ele pode ser utilizado para previsão de valores futuros, de acordo com as etapas da metodologia *Box e Jenkins* para séries estacionárias: identificação do modelo adequado à série, estimação dos parâmetros do modelo, verificação dos resíduos de ajuste e previsão [11][7]. A série diretamente gerada pelo modelo é uma série estimada, $\hat{X}_t$, e a diferença entre os valores das amostras desta e da série real, para cada $t \in T$, é chamada de resíduo do ajuste no instante t. O erro de previsão, ponto de partida para a análise da capacidade de previsão de um modelo, pode ser medido para uma série conhecida quando a série estimada é gerada utilizando valores observados até o instante t-k. Assim, no instante t, o valor previsto para o instante, $\hat{X}_t$, baseado em $X_{t-1}, X_{t-2}, \dots, X_{t-k}$, pode ser comparado ao valor medido no instante, X_t. Desta maneira

são estimados modelos para o processo e aqueles com melhor desempenho de ajuste podem ser selecionados para a previsão de valores futuros.

A métrica de medição de erro mais utilizada é a raiz quadrada do erro médio (RMSE – *Root Mean Squared Error*) (Eq. 17), e uma alternativa independente de escala é o Percentual de Erro Médio Absoluto (MAPE – *Mean Absolute Percent Error*) (Eq. 18). Ambas métricas visam impedir que os valores de erros positivos e negativos (erro por excesso ou por falta, respectivamente) se cancelem [8]. Definindo o erro (Eq. 16) e calculando-o para as uma sequência de *n* amostras em *t*, segue

$$e_t = X_t - \hat{X}_t, \tag{16}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {e_i}^2}, \tag{17}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{e_i}{X_i}\right|\right), \forall x_i \neq 0. \tag{18}$$

O intervalo de confiança da previsão pode ser calculado supondo-se o erro de previsão como um processo Gaussiano i.i.d. com média zero e variância constante. Subtraindo-se da série a sua média amostral e dividindo-se a mesma pelo seu desvio padrão amostral, pode-se transformá-la numa sequência de variáveis aleatórias normais padrão [9].

Para problemas de natureza financeira, as séries podem apresentar características não contidas nas premissas estacionárias, demandando uma metodologia complementar à de *Box e Jenkins*. De modo mais geral, pode-se decompor uma série em sazonalidade (característica periódica), S_t, tendência (ao crescimento ou ao decrescimento), T_t, e irregularidade, I_t, todas, funções do domínio temporal. A sazonalidade tem como modelo mais simples uma senoide com frequência f e amplitude máxima A_0 (Eq. 19); a tendência, uma reta com coeficientes α e β que determinam a taxa de variação e o valor inicial (Eq. 20); e a irregularidade, uma combinação linear de amostras anteriores adicionada de um processo i.i.d. (Eq. 21) [9]. Assim, o valor da série no instante *t* pode ser aproximado genericamente (por exemplo, na Figura 2) de forma mais básica, por

$$S_t = A_0 sin(2\pi f t), \tag{19}$$

$$T_t = \alpha + \beta t, \tag{20}$$

$$I_t = a_1 I_{t-1} + a_2 I_{t-2} + \cdots + a_n I_{t-n} + \varepsilon_t, onde \varepsilon_t = N(\mu, \sigma^2), \tag{21}$$

$$X_t = S_t + T_t + I_t, \tag{22}$$

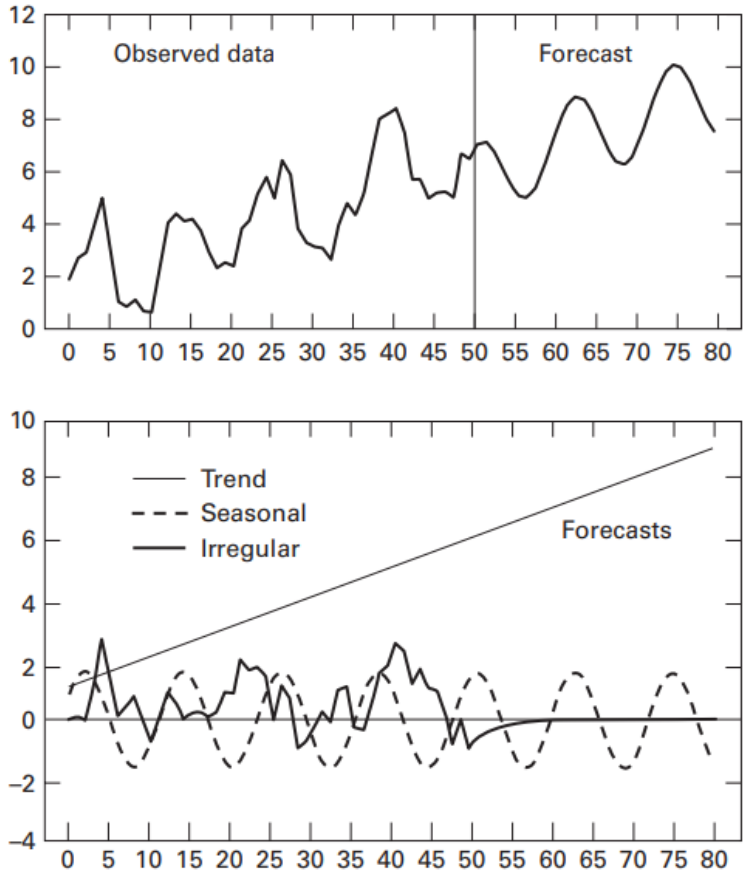

Figura 2 - Série temporal hipotética e sua respectiva previsão, e logo abaixo, decomposição do modelo obtido em parte sazonal (S_t = 1,6sen (tπ/6)), tendência (T_t = 1 + 0.1t) e irregularidade (I_t = $0.7I_{t-1} + \varepsilon_t$)[9].

Para cada uma das fontes de não-estacionariedade descritas anteriormente (tendência e sazonalidade), há modelos que, através de transformações específicas, adéquam, às séries observadas, a metodologia escolhida. Os modelos estudados nas Seções 2.3 e 2.5 buscam modelar a irregularidade, enquanto o modelo da Seção 2.4 contém ajustes para modelar, também, a tendência. Os modelos das Sessões 2.7 e 2.8, *a priori*, contemplam todas as características.

2.3 ARMA

Seja um processo $\{X_t\}$ estritamente estacionário, portanto, tendo média $\mu = E[X_t]$ constante e autocovariância entre duas amostras X_t e $X_{t+\tau}$ dependente apenas da diferença de tempo, τ [10],

$$cov(X_t, X_{t+\tau}) = E[(X_t - \mu)(X_{t+\tau} - \mu)] = \lambda_\tau. \quad (23)$$

Inicialmente, considere que X_t seja linearmente dependente apenas do seu valor anterior, acrescido de uma *inovação*, ε_t, onde ε_t é uma variável aleatória que tem distribuição $N(0, \sigma^2)$, e a é uma constante arbitrária, tal que

$$X_t = a(X_{t-1} - \mu) + \mu + \varepsilon_t. \quad (24)$$

Este processo é denominado autorregressivo de ordem 1, AR(1). Se $|a| < 1$, tem-se um processo (assintoticamente) estacionário. Pode-se definir o operador de atraso B^k, onde

$$B^k X_t = X_{t-k} e B^k \mu = \mu. \quad (25)$$

Assim, pode-se reescrever a Equação 24 como

$$(1 - aB)(X_t - \mu) = \varepsilon_t, \quad (26)$$

e, desde que $|a| < 1$, temos (análogo a um filtro IIR),

$$X_t - \mu = \frac{1}{1 - aB}\varepsilon_t = \sum_{i=0}^{\infty} (aB)^i \varepsilon_t = \sum_{i=0}^{\infty} a^i \left(B^i \varepsilon_t\right) = \sum_{i=0}^{\infty} a^i \varepsilon_{t-i}. \quad (27)$$

A autocorrelação do processo, ρ_τ, pode ser obtida a partir da autocovariância, λ_τ,

$$\rho_\tau := \frac{\lambda_\tau}{\lambda_0} = a^\tau, \quad (28)$$

uma vez que

$$\lambda_\tau = E[(X_t - \mu)(X_{t-\tau} - \mu)] = E[(a(X_{t-1} - \mu) + \mu + \varepsilon_t)(X_{t-\tau} - \mu)] = a\lambda_{\tau-1} = a^\tau \lambda_0. \quad (29)$$

Generalizando, um processo autorregressivo de ordem p, AR(p), de média zero é definido como

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + \varepsilon_t = \left[\sum_{i=1}^{p} a_i B^i X_t\right] + \varepsilon_t. \quad (30)$$

Por sua vez, um processo média móvel de ordem 1, MA(1) é dado por combinação linear dos valores de entrada da variável aleatória e pode ser descrito por

$$X_t = \mu + \varepsilon_t + b\varepsilon_{t-1} = \mu + (1 + bB)\varepsilon_t = \mu + \theta(B)\varepsilon_t. \quad (31)$$

Como $\theta(B) = 1 + bB$ é finito (análogo a um filtro FIR), o processo é sempre estacionário [6][9]. Seja σ^2 a variância de ε_t, tem-se então que

$$\lambda_0 = var(X_t) = (1 + b^2)\sigma^2, \quad (32)$$

$$\lambda_\tau = cov(X_t, X_{t-\tau}) = E[(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t+\tau} + b\varepsilon_{t+\tau-1})]. \quad (33)$$

Devido ao fato de o processo MA(1) relacionar cada amostra apenas à amostra imediatamente anterior, a covariância (consequentemente, a correlação) é nula para τ >1, assim, pode-se calcular

$$\lambda_1 = b\sigma^2, \quad (34)$$

$$\rho_1 = \frac{\lambda_1}{\lambda_0} = \frac{b}{1 + b^2}. \quad (35)$$

Generalizando, um processo média móvel de ordem q, MA(q), e média zero é definido como

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + \cdots + b_p\varepsilon_{t-p}. \quad (36)$$

A união dos processos AR(1) e MA(1) é chamado processo autorregressivo média móvel de ordem (1,1), ou seja, um processo ARMA(1,1) e é dado por

$$X_t - \mu = a(X_{t-1} - \mu) + \varepsilon_t + b\varepsilon_{t-1}, \quad (37)$$

com $|a|<1$, em que

$$\lambda_1 = a\lambda_0 + b\sigma^2, \tag{38}$$

$$\lambda_0 = a\lambda_1 + (1 + ab + b^2)\sigma^2, \tag{39}$$

$$\rho_\tau = \frac{(1+ab)(a+b)}{(1+2ab+b^2)} a^{\tau-1}, \tau \geq 1. \tag{40}$$

Generalizando, um processo autorregressivo de média móvel de ordem (p,q) e média μ é definido

$$X_t - \mu = \sum_{i=1}^{p} a_i (X_{t-i} - \mu) + \sum_{j=0}^{q} b_j \varepsilon_{t-j}. \tag{41}$$

A comparação das autocorrelações da série observada com a autocorrelação teórica dos modelos permite a definição de qual modelo é o mais adequado para a previsão. Esta comparação é possível através de um teste de autocorrelação [10]. Seja a série temporal com n observações $\{X_1, X_2, \ldots, X_n\}$; os coeficientes amostrais de autocorrelação são dados por

$$\rho_s = \frac{\sum_{t=1}^{n-s}(X_t - \mu)(X_{t+s} - \mu)}{\sum_{t=1}^{n}(X_t - \mu)^2}, s \geq 1. \tag{42}$$

Os coeficientes teóricos de autocorrelação compõem a função de autocorrelação (ACF – *Autocorrelation Function*) [9]. A análise comparativa também pode fazer uso da função de autocorrelação parcial (PACF – *Partial Autocorrelation Function*), ϕ_{ss}, [9] que indica a autocorrelação residual àquela proporcionada pelo modelo,

$$\phi_{ss} = \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_{s-j}}, s = 3, 4, 5, \ldots, onde\ \phi_{sj} = \phi_{s-1,j} - \phi_{ss}\phi_{s-1,s-j}. \tag{43}$$

Além disso, $\phi_{11} = \rho_1$ [9]. Tais funções para os modelos considerados podem ser observadas na Tabela 1:

Tabela 1: Modelos e suas respectivas ACF e PACF [9].

Processo	**ACF**	**PACF**
Ruído Branco $\varepsilon_t \sim N(0, \sigma^2)$	$\rho_s = 0, \forall s \neq 0$	$\phi_{ss} = 0, \forall s$
AR(1): $a_1 > 0$	Decaimento geométrico direto: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0, \forall s \geq 2$
AR(1): $a_1 < 0$	Decaimento oscilatório: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0, \forall s \geq 2$
AR(p)	Decai para zero. Coeficientes podem oscilar.	Picos até o atraso p. $\phi_{ss} = 0, \forall s > p$
MA(1): $\beta > 0$	Pico positivo no atraso 1. $\rho_s = 0, \forall s \geq 2$	Decaimento oscilatório: $\phi_{11} > 0$
MA(1): $\beta < 0$	Pico negativo no atraso 1. $\rho_s = 0, \forall s \geq 2$	Decaimento geométrico: $\phi_{11} < 0$
ARMA(1,1): $a_1 > 0$	Decaimento geométrico começando após o atraso 1. Sinal de ρ_1 = sinal de $(a_1 + \beta)$	Decaimento oscilatório após atraso 1. $\phi_{11} = \rho_1$
ARMA(1,1): $a_1 < 0$	Decaimento oscilatório começando após o atraso 1. Sinal de ρ_1 = sinal de $(a_1 + \beta)$	Decaimento geométrico após atraso 1. $\phi_{11} = \rho_1$ e sinal de ϕ_{ss} = sinal de ϕ_{11}
ARMA(p,q)	Decaimento (direto ou oscilatório) começando após o atraso q.	Decaimento (direto ou oscilatório) começando após o atraso p.

2.4 ARIMA

Quando a série temporal obtida empiricamente não é estacionária, pode-se iniciar a análise considerando que tal efeito tem origem na tendência, como visto na Figura 2; nesses casos, os modelos ARMA precisam de alterações.

A presença de tendência pode ser modelada como a presença de uma raiz unitária no polinômio de atraso do processo gerador da série, que pode ser detectada com um teste de raiz unitária, como o Teste de Dickey-Fuller Aumentado (ADF - *Augmented Dickey-Fuller*) [7], caso confirmada a presença, a série pode ser transformada em estacionária tomando dela a diferença de cada amostra, com o uso do modelo ARMA Integrado (ARIMA – *Autoregressive Integrated Moving Average*) [11].

Tomemos o modelo ARMA(p,q) visto na Seção 2.2 (Eq. 27) e coloquemos em termos do operador de atraso, com média zero,

$$\phi(B)X_t = \theta(B)\varepsilon_t, \tag{44}$$

em que

$$\phi(B) = \sum_{i=1}^{p} a_i B^i \, ; \theta(B) = \sum_{j=0}^{q} b_j B^j. \tag{45}$$

Para remover a tendência, podemos substituir cada amostra X_t pela sua respectiva diferença de ordem *d*, indicada pelo operador ∇^d, isto é,

$$W_t = \nabla^d X_t = (1 - B)^d X_t, d \geq 1. \tag{46}$$

Assim, temos o processo ARIMA(*p*,*d*,*q*)

$$\phi(B) W_t = \phi(B) \nabla^d X_t = \theta(B) \varepsilon_t, \tag{47}$$

ou, equivalentemente,

$$\phi(B)(1 - B)^d X_t = \theta(B) \varepsilon_t, \tag{48}$$

em que podemos definir o operador autorregressivo generalizado,

$$\varphi(B) = \phi(B) \nabla^d. \tag{49}$$

Finalmente, dentre diversos modelos concorrentes, há técnicas para a escolha do modelo de maior parcimônia, ou seja, a escolha valoriza não apenas a minimização de erros, mas também os modelos que requeiram menos parâmetros; são os chamados critérios de informação. Dois destes são o *Akaike Information Criterion* (AIC) e o *Schwartz Bayesian Criterion* (SBC) [9], cujas expressões são dadas abaixo, em que *n* é o número de parâmetros usados (*p*+*q* ou *p*+*q*+1, se for usado um termo constante) e *T* é o número de observações usadas.

$$AIC = T ln\left(\sum_i e_i^2\right) + 2n, \tag{50}$$

$$SBC = T ln\left(\sum_i e_i^2\right) + n ln(T). \tag{51}$$

O modelo mais adequado, segundo o AIC ou o SBC, é aquele que minimiza as expressões acima.

2.5 ARCH/GARCH

Os modelos econométricos convencionais assumem que ε_t, portanto, a série, possui variância constante (homocedasticidade). Entretanto, algumas séries apresentam períodos de tranquilidade que se alternam a períodos de oscilação, comportamento que exige o uso de modelo mais complexos [9]. Muitas aplicações não requerem estimativas da variância incondicional, mas sim de variância condicional, portanto, é para esta última que se requer uma nova abordagem.

Para um modelo AR(1), dado pela Equação 24, assuma-se, sem perda de generalidade, que a média seja zero. A variância condicional um passo a frente é dada, considerando que var(ε_t) seja constante igual a σ^2, por

$$var(x_{t+1}|x_t) := E_t[(x_{t+1} - a_0 - a_1 x_t)^2] = E_t[\varepsilon_{t-1}^2] = \sigma^2. \quad (52)$$

A condição acima pode ser generalizada supondo-se, então, que a variância condicional não seja constante. Pode-se modelá-la como [12]

$$\varepsilon_t = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}, v_t = N(0,1), \quad (53)$$

em que v_t e ε_t são processos independentes, ε_t e ε_{t-1}, e v_t e v_{t-1} são descorrelacionados entre si, $\alpha_0 > 0$ e $0 \leq \alpha_1 \leq 1$. Dando origem ao modelo de Heteroscedasticidade Condicional Autorregressiva (ARCH– *Autorregressive Conditional Heteroskedasticity*). Tomando a esperança e variância incondicionais, pode-se verificar que são constantes. Disto segue que não é nas propriedades incondicionais que se estão aplicando alterações, pois

$$E[\varepsilon_t] = E\left[v_t(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{\frac{1}{2}}\right] = E[v_t]E\left[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{\frac{1}{2}}\right] = 0, \quad (54)$$

$$var(\varepsilon_t) = E[\varepsilon_t^2] = E[{v_t}^2(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)] = \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] \underset{t\to\infty}{\to} \frac{\alpha_0}{(1-\alpha_1)}. \quad (55)$$

A esperança e a variância condicionais, por sua vez, são dadas por

$$E[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\dots] = E_{t-1}\left[v_t(\alpha_0+\alpha_1\varepsilon_{t-1}^2)^{\frac{1}{2}}\right] = E_{t-1}[v_t]E_{t-1}\left[(\alpha_0+\alpha_1\varepsilon_{t-1}^2)^{\frac{1}{2}}\right] = 0, \quad (56)$$

$$E[\varepsilon_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\dots] = \alpha_0+\alpha_1\varepsilon_{t-1}^2; \alpha_0>0; 0<\alpha_1<1. \quad (57)$$

Tem-se, portanto, que a variância condicional segue um processo autorregressivo de ordem 1, AR(1), o que constitui um processo denominado ARCH(1). A generalização deste conceito é um processo ARCH(*p*), no qual

$$\varepsilon_t = v_t\sqrt{\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2}, v_t \sim N(0,1). \quad (58)$$

Uma extensão do processo ARCH, fazendo uma analogia com o processo ARMA, ou seja, unindo componentes autorregressivas às componentes média móvel, na variância heteroscedástica é dada por [13]

$$\varepsilon_t = v_t\sqrt{h_t}, \quad (59)$$

$$h_t = \alpha_0+\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2+\sum_{i=1}^{p}\beta_i h_{t-i}. \quad (60)$$

Temos, então, o modelo de processo GARCH (*Generalized autoregressive conditional Heteroskedasticity*) de ordem (p,q) dado pela Equação 59, cuja variância condicional é dada por h_t (Eq. 60). Feita a escolha de um modelo, uma das formas de determinar o adequado, como no estudo de modelos ARMA, consiste em observar que as funções de autocorrelação e autocorrelação parcial devem se comportar como as de um ruído branco; do contrário, os quadrados dos resíduos podem ajudar a identificar a ordem correta do processo GARCH através da construção de um correlograma, obtido calculando-se a variância e as autocorrelações amostrais

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2, \quad (61)$$

$$\rho_i = \frac{\sum_{t=i+1}^{T}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^{T}(\hat{\varepsilon}_t - \hat{\sigma}^2)^2}, \quad (62)$$

como descrito na seção de metodologia.

Os coeficientes do modelo escolhido podem ser estimados utilizando a função log-verossimilhança (*log-likelihood*), assumindo distribuição normal para a série [11]. O princípio de verossimilhança determina que devam ser assumidos, para os parâmetros da distribuição da variável aleatória, valores que maximizem a probabilidade de se obter as amostras particulares observadas. A maximização do logaritmo natural desta função constitui a técnica de estimação por log-verossimilhança [14].

O condicionamento do diagnóstico de adequação de um modelo ao fato de que os resíduos devem ser descorrelacionados [6], ou seja, compor um conjunto de variáveis aleatórias normais independentes e identicamente distribuídas, advém do Teorema do Limite Central, que diz que para um conjunto de amostras aleatórias simples, retiradas de uma população com média e variância finitas, a distribuição da média amostral, consequentemente da variância amostral, se aproximam de uma distribuição normal. O somatório do quadrado de *N* variáveis aleatórias com distribuição normal constitui uma variável aleatória qui-quadrado com n graus de liberdade [6][15]. Desta ideia foram desenvolvidos os testes de hipóteses de *Box-Pierce* e de *Ljung-Box* (Eq. 63) (de melhor desempenho para amostras pequenas) [9].

$$Q = T(T+2)\sum_{i=1}^{n}\frac{\rho_i^2}{(T-i)}\chi^2. \quad (63)$$

O teste de *Ljung-Box* toma por hipótese H_0 que o somatório ponderado do quadrado das n autocorrelações para resíduos do ajuste de um modelo de ordem $p+q$, a estatística *Q*, se aproxima de uma distribuição qui-quadrado com $n-(p+q)$ graus de liberdade; se assim for, os resíduos são descorrelacionados e o modelo escolhido é adequado.

2.6 Redes Neurais Artificiais: introdução

O cérebro humano é notável pela sua capacidade de processar informações novas, em tempo real, utilizando-se de informações processadas anteriormente; além disso, possui uma plasticidade de modo que uma possível consequência do

processamento é alterar sua própria constituição. As unidades estruturais, os neurônios, recebem impulsos nervosos através das suas terminações, denominadas dendritos. Caso o impulso gerado pela combinação das entradas supere um determinado limiar, particular de cada neurônio, sua propagação é conduzida através do axônio, possibilitando a condução do impulso para outros neurônios. Visando obter uma forma de processamento de dados estruturalmente diferente da proporcionada pela computação tradicional, as Redes Neurais Artificiais foram desenvolvidas inspiradas em alguns mecanismos cerebrais, com o intento de apreender capacidades como as supracitadas [16].

Uma Rede Neural Artificial consiste em um conjunto de neurônios artificiais distribuídos em camadas. Estes, por sua vez, são inspirados no neurônio biológico. O modelo básico de um neurônio artificial é composto por um conjunto de *n* sinapses, onde cada sinapse é caracterizada por um peso, w_{ij} (correspondente a *i*-ésima entrada do *j*-ésimo neurônio), atribuído a uma entrada, x_i, um operador somador, uma função de ativação e um parâmetro de corte (*threshold*), Θ_j, além de uma saída, y_j. Este modelo de rede é denominado *perceptron* [16] (Figura 3).

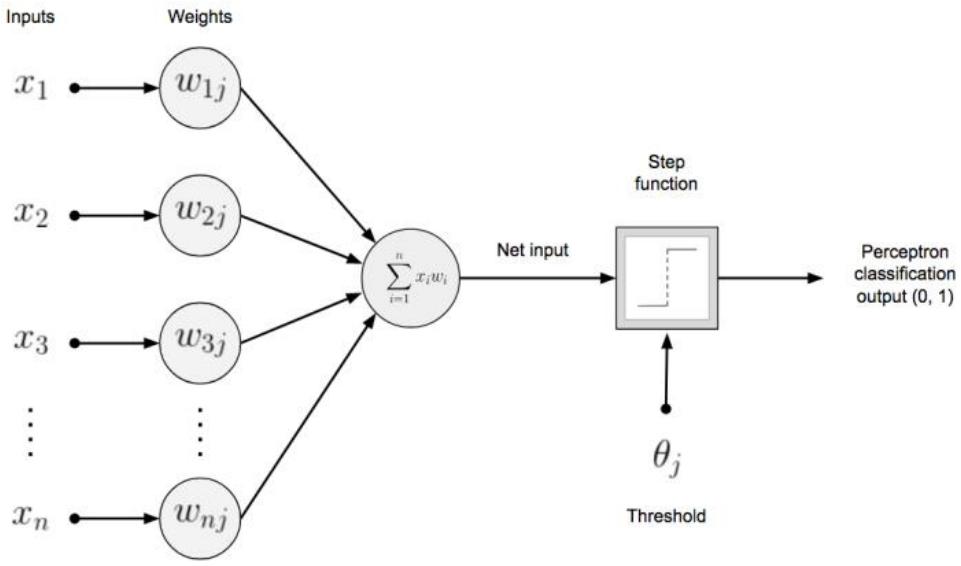

Figura 3 - Modelo de neurônio artificial [16].

Uma das finalidades da função de ativação é evitar o aumento progressivo dos valores de saída ao longo das camadas da rede. Um dos tipos mais simples é a função de limiar, onde a combinação linear das entradas (Eq. 64) é comparada com o parâmetro de corte, se o parâmetro for atingido, o neurônio é ativado e tem saída igual a um, do contrário, zero [16]; assim, para o *j*-ésimo neurônio com *n* entradas,

$$v_j = \sum_{i=1}^{n} x_j w_{ij} - \theta_j, \tag{64}$$

$$y_j = \begin{cases} 1, se v_j \geq 0 \\ 0, se v_j < 0 \end{cases}. \tag{65}$$

Existem duas funções cujos comportamentos se assemelham ao da Equação 65 e, por isto, são muitas vezes utilizadas como função de ativação: a função sigmoide (Eq. 66) e a função tangente hiperbólica (Eq. 67), esta última é mais útil nos casos em que, devido às características do processo analisado, saídas negativas devem fazer parte das camadas de rede [17],

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \tag{66}$$

$$tanh(x) = 2\sigma(2x) - 1 = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{67}$$

Alternativamente, um neurônio pode dispor de uma entrada de valor fixo, denominada *bias*, b_j, que fornece à rede a capacidade de ter uma saída diferente de zero mesmo que todas as entradas sejam iguais a zero, satisfazendo a demanda de certas classes de problemas. Uma entrada x_i pode entrar em mais de um neurônio; um conjunto de neurônios em paralelo compõe uma camada. A saída de cada neurônio pode ser entrada de um ou mais neurônios da camada seguinte: *feed-forward* é o nome dado às redes que possuem uma camada de entrada, uma ou mais camadas intermediárias (também denominadas *hidden layers* – camadas ocultas) e uma camada de saída, com ligações unidirecionais entre os neurônios e as camadas, ou seja, o processo atravessa as camadas na direção da entrada para a saída [18].

A aplicação da rede na tarefa de modelagem e previsão de séries temporais requer, inicialmente, que os pesos w_{ij} sejam adequados às características da série através de um processo de aprendizagem da rede: inserir as entradas e informar à rede as saídas desejadas. Então, a saída obtida é comparada com a saída desejada e atualizam-se os pesos da rede de modo proporcional ao erro medido Em seguida, reinserem-se as entradas, num processo cíclico que se encerrará quando for atingido um parâmetro de erro pré-estabelecido. Para as redes de apenas uma camada, um dos processos de aprendizagem mais simples é a aprendizagem por correção de erro [17]: seja $w_{ij}(n)$, o

peso da sinapse *i* do neurônio *j* no instante *n*, onde sabemos o valor desejado na saída $d_j(n)$, e seja $y_j(n)$ a saída obtida, de modo que é possível calcular o erro (Eq. 68); pode-se obter iterativamente o ajuste, Δw_{ij}, que atualiza o valor do peso e faz com que este tenda na direção de minimizar o erro, onde η é o parâmetro de atualização, como

$$e_j(n) = d_j(n) - y_j(n), \tag{68}$$

$$\Delta w_{ij}(n) = \eta e_j(n) x_j(n), \tag{69}$$

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n). \tag{70}$$

Para uma rede multicamadas (*Multi-layer Perceptron*), utiliza-se o algoritmo de retropropagação (*backpropagation*), resumidamente descrito da seguinte forma: definindo a energia total do erro no instante *n* como metade da soma dos erros quadráticos (Eq. 71), pode-se definir o ajuste $\Delta w_{ij,}$ considerando $v_j(n)$ como a saída do *j*-ésimo neurônio,

$$\xi(n) = \frac{1}{2}\sum_j e_j^2(n), \tag{71}$$

$$v_j(n) = \sum_{i=0}^{n} w_{ij}(n)\, y_i(n), \tag{72}$$

$$\Delta w_{ij} = -\eta \frac{\partial \xi(n)}{\partial v_j(n)} y_i(n). \tag{73}$$

Após o ajuste, a rede pode receber dados completamente novos (ou seja, que não tenham participado do processo de treinamento) e ser utilizada para a previsão das respectivas saídas.

2.7 Redes Neurais Artificiais Autorregressivas

O procedimento de *realimentação* consiste na reinserção de uma saída em uma entrada da mesma camada ou de camadas anteriores através de um laço. Os dois principais tipos de redes neurais artificiais que fazem uso de realimentação são as recorrentes (RNN – *recurrent neural network*) e as autorregressivas (NAR – *Nonlinear*

Autoregressive). A realimentação, em cada tipo, é dada da seguinte maneira: as RNN (Figura 4) fazem uso de conexões recorrentes dentro de sua própria estrutura; por outro lado, as NAR (Figura 5) utilizam a saída atrasada da última camada, bem como as próprias entradas atrasadas, como entradas da camada inicial [19].

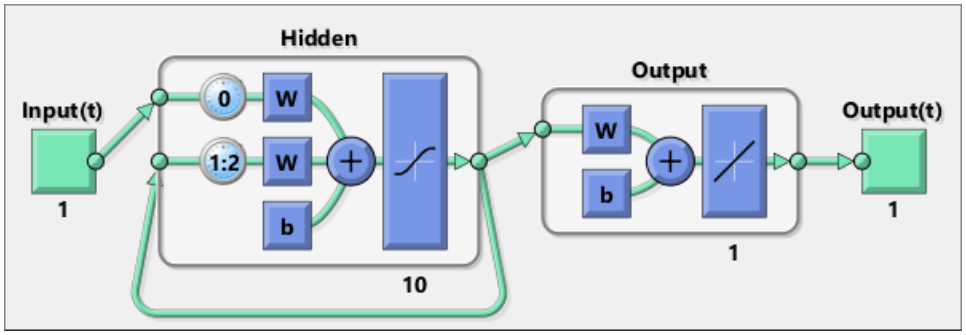

Figura 4 - Modelo de RNN [20].

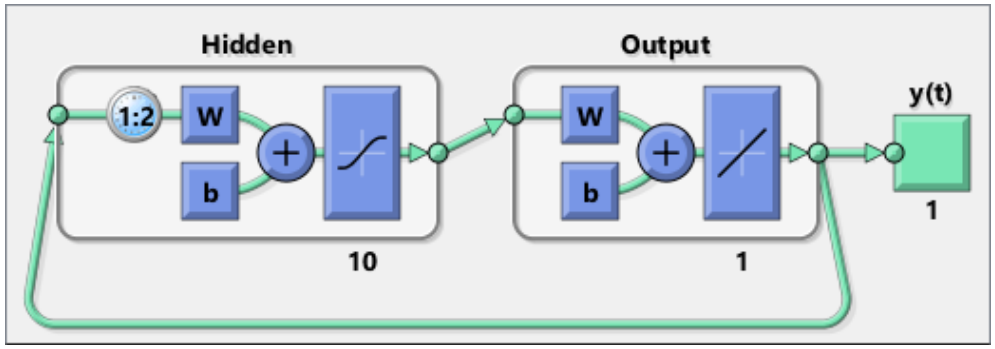

Figura 5 - Modelo de NAR [21].

Neste trabalho, faz-se uso das NAR, nas quais a realimentação contém um operador de atraso por b amostras. Uma rede neural autorregressiva é análoga ao modelo AR(p): o j-ésimo neurônio da camada de entrada recebe as p saídas anteriores como suas entradas, toma sua combinação linear e passa pela função de ativação, fazendo de forma análoga à Equação 63. Assim, tem-se

$$v_j = \sum_{i=1}^{p} \hat{y}(t-i) w_{ij} - \theta_j, \tag{74}$$

em que $\hat{y}(t)$ é a saída da camada de saída no instante t.

2.8 LSTM

Visando resolver alguns problemas inerentes às estruturas recorrentes descritas anteriormente como, por exemplo, o problema de *gradient vanishing* [22],

bem como desenvolver uma estrutura com capacidade de memória maior e mais flexível, foram criadas as redes LSTM (*Long Short-Term Memory*). Trata-se de uma RNN que realimenta seu estado do instante anterior como ilustrado na Figura 6.

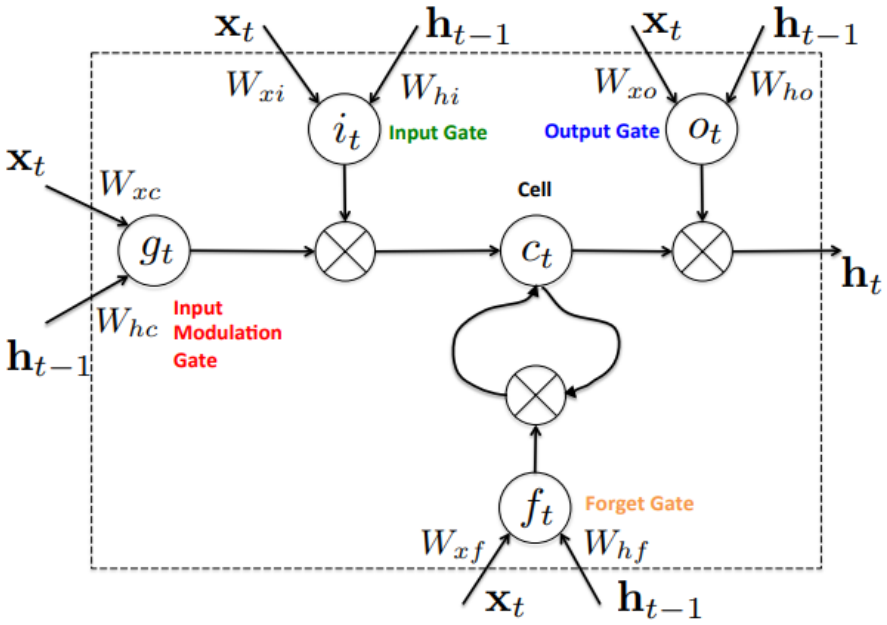

Figura 6 - Modelo de LSTM [22].

Uma LSTM, basicamente, é formada por quatro portas que manipulam a memória, e uma célula, c_t, que é o elemento central, responsável pela retenção da memória. Todas as quatro portas fazem uma combinação linear do vetor de entrada no instante com o vetor de saída da LSTM no instante anterior. Cada porta possui suas matrizes de pesos e vetor de *bias*, os atualiza de forma independente e passa a combinação linear por uma função de ativação.

A saída da função de cada porta tem um objetivo específico na arquitetura da rede: as saídas de duas portas, *Input Modulation Gate*, g_t, e *Input Gate*, $\mathbf{i_t}$, são combinadas e compõem a entrada da célula, adicionando informações úteis a esta. Elas diferem estruturalmente entre si apenas pela função de ativação: a primeira é uma tangente hiperbólica (Eq. 67), pode ser positiva ou negativa, e a segunda uma sigmoide (Eq. 66), assume apenas valores positivos.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{75}$$

$$g_t = tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c); \tag{76}$$

A saída da Forget Gate, f_t, por outro lado, tem a capacidade de alterar o conteúdo da célula, tendo a função de remover as informações que não são mais úteis, ou seja,

limpar a memória quando necessário. Assim, a célula atualiza sua informação condicionando o seu estado anterior ao efeito da *Forget Gate*, e soma à combinação das entradas:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right), \tag{77}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t; \tag{78}$$

em que o operador utilizado na Equação 78 é o produto de *Hadamard* [23], definido para matrizes *A* e *B* de mesmo tamanho *m* x *n* como

$$[A \circ B]_{ij} = [A]_{ij}[B]_{ij,} \forall 1 \leq i \leq m, 1 \leq j \leq n. \tag{79}$$

Por outro lado, a saída da *Output Gate* o_t se combina com a saída da célula (após passagem desta por uma função de ativação sigmoide) constituindo a saída da LSTM, h_t,

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{80}$$

$$h_t = o_t \circ tan\, h(c_t). \tag{81}$$

A atualização das matrizes de pesos da LSTM também pode ser feita de acordo com o algoritmo *backpropagation* [22].

3. Materiais e métodos

3.1 Dados: preços de ações

Os dados foram compostos de históricos de cotações de três ações dentre as disponíveis para *download* no site Yahoo Finanças [24]. Os valores da coluna de valor de abertura de cada arquivo foram numerados e compuseram as séries temporais, objetos de estudo.

As ações estudadas foram os valores históricos da Petroleo Brasileiro S.A. - Petrobras (PETR4.SA) no período 03/01/2000 - 17/04/2019, da Itau Unibanco Holding S.A. (ITUB4.SA) no período 21/12/2000 - 06/11/2019 e da Xiaomi Corporation (1810.HK) no período 08/07/2018 – 06/11/2019. Todas foram separadas em dados de treinamento (para o ajuste do modelo, 85%) e dados de teste (para a avaliação da capacidade de previsão do modelo ajustado, 15%), de acordo com o recomendado pela literatura [12].

Figura 7 – Gráfico preço x tempo das séries estudadas, da esquerda para a direita: PETR4.SA, ITUB4.SA e 1810.HK.

3.2 *Matlab*: *Econometric Toolbox*

O *Matlab* oferece uma caixa de ferramentas de econometria para a modelagem e previsão de séries temporais através de diversos modelos clássicos. O pacote contém funções para análise de escolha entre modelos candidatos, testes de resíduos, criação, estimação, filtragem, previsão, inferência e simulação para modelos ARIMA e GARCH, dentre outros, e inserção de operadores de atraso, dentre outras opções.

3.3 Matlab: *Natural Network Toolbox*

Outra caixa de ferramentas disponível no *Matlab* é a de redes neurais. Esta proporciona uma interface gráfica (Figura 8) que facilita o contato inicial com o pacote, e um conjunto de funções que oferecem mais opções e independência da interface gráfica. O pacote permite a modelagem de redes neurais e o processo de aprendizado (treinamento), possibilitando aplicação em reconhecimento de padrões, clusterização, previsão de séries temporais, por exemplo. A inserção e o processamento de dados podem ser realizados em linhas de código e contam com o apoio visual de diagramas de bloco (Figuras 4 e 5, por exemplo).

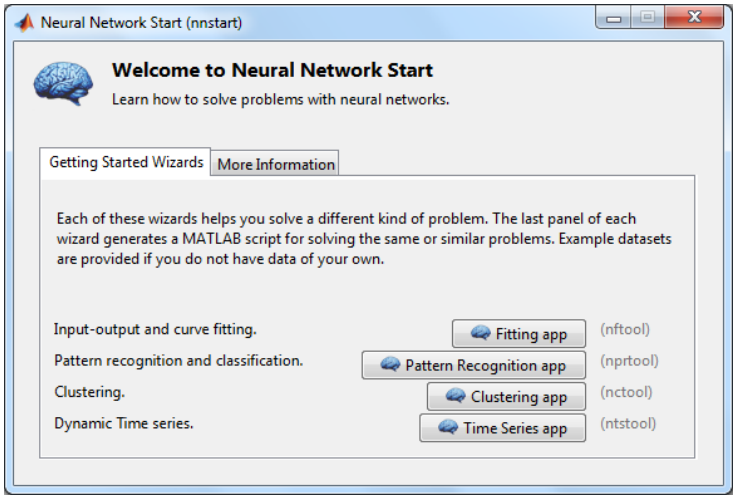

Figura 8 - Interface gráfica disponível no *Neural Network Toolbox*

4. Execução computacional

4.1 ARIMA

A primeira etapa de especificação consistiu em tomar modelos ARIMA. Inicialmente, foi calculada a autocorrelação amostral e a autocorrelação parcial amostral para os dados de treinamento. As funções utilizadas para os cálculos foram "*autocorr*" e "*parcorr*". Foram plotados os correlogramas, de modo a serem estimadas as ordens de autorregressão e média móvel que melhor se ajustem aos dados, seguindo a análise resumida na Tabela 1. Na sequência foi aplicado um teste da raiz unitária ADF (função "*adftest*"). Com os resultados em mãos, foi possível realizar a etapa de identificação, escolhendo as triplas ordenadas (*p*,*d*,*q*) correspondentes às ordens dos modelos.

Com os modelos definidos, iniciou-se a etapa de estimação dos coeficientes de cada modelo utilizando a função "*estimate*" que os calcula pela maximização da função log-verossimilhança [25]. A escolha dos dois modelos mais adequados foi possível com a análise dos valores AIC para cada modelo, calculado com a função "*summarize*".

As séries de resíduos do ajuste dos modelos à série original foram obtidas a partir da função "*infer*", possibilitando a comparação. Inicialmente, os resíduos passaram pelo teste de descorrelação de Ljung-Box (função "*lbqtest*"). Então, foi possível a escolha do melhor modelo para ajuste ao serem comparados os valores de RMSE e MAPE.

Finalmente, os modelos foram usados para a previsão (função "*forecast*"). Os resultados foram comparados com os dados de teste, determinando as séries de erros de previsão, bem como suas medidas de RMSE e MAPE.

4.2 GARCH

O modelo com o melhor desempenho de ajuste na etapa anterior foi selecionado para a tentativa de melhora por meio da alteração da modelagem da variância dos resíduos, de constante, para um modelo GARCH. Para embasar a estimativa da variância dos resíduos, foi calculada a volatilidade histórica (ou volatilidade empírica), definida como a variância amostral de uma janela deslizante de N amostras. Em trabalhos futuros, pode ser estudada a janela de modelo média móvel com ponderação exponencial (EWMA - *Exponentially Weighted* Moving Average) [26]. O tamanho escolhido da janela para percorrer o sinal foi de 21 amostras de largura.

Os correladogramas da volatilidade empírica possibilitaram a escolha de dois modelos para a variância.

Em seguida, empregou-se o mesmo procedimento utilizado para modelos ARIMA: criação do modelo, estimação dos parâmetros por log-verossimilhança, análise dos critérios de informação e escolha das duas mais bem sucedidas dentre as três (o modelo mais bem sucedido da etapa ARIMA seguiu para a comparação com o melhor modelo obtido pela modificação na variância para um modelo GARCH). Em seguida, fez-se a inferência e teste de descorrelação dos resíduos, o cálculo das métricas RMSE e MAPE, da previsão, e o cálculo do erro de previsão e de suas métricas RMSE e MAPE.

4.3 NARNET

Foi utilizada uma NARNET (*Nonlinear Autoregressive Neural Network*), com o uso da função "*narnet*". A rede foi mantida *aberta* para a etapa de ajuste do modelo: isto significa que a realimentação da rede foi efetuada com as amostras disponibilizadas para treinamento, e não com as estimativas geradas pela própria rede.

Para a camada intermediária, foram fixados 10 neurônios. A ordem do operador de atraso foi variada de 1 a 20. A organização dos dados de treinamento, de acordo com o número de operadores de atraso, em saídas desejadas (targets) e entradas, fora feita com o uso da função "preparets", uma vez que a seleção deve percorrer o vetor de dados conforme a rede executa o algoritmo.

A rede foi treinada utilizando a função "*train*", e, após ser atingida a convergência, o vetor de saídas da última sessão da rede foi armazenado. O desempenho da rede (erro médio quadrático) foi obtido com a função "*perform*". Foram separados os resíduos e calculados suas métricas de RMSE e MAPE. Finalmente, a rede foi *fechada* (i.e., realimentada com os valores estimados) e usada para o cálculo da previsão de valores futuros, permitindo assim o cálculo dos erros de previsão e suas métricas RMSE e MAPE.

4.4 LSTM

Com o uso da função "*lstmLayer*" foi construída uma rede LSTM com 100 unidades ocultas (*hidden units*). Os parâmetros como a taxa de aprendizagem inicial e número máximo de épocas foram testados e ajustados para cada série considerada compondo o vetor "*options*", que é parâmetro da função "*trainNetwork*". O tratamento de resíduos de ajuste e erros de previsão, suas métricas, bem como o protocolo que norteou a sequência de tarefas, seguiram a lógica semelhante à vista para as outras etapas desta seção.

5. Resultados

5.1 ARIMA

Os correladogramas calculados fornecendo informações para a escolha, *a priori*, dos modelos mais adequados, estão dispostos na Figura 9.

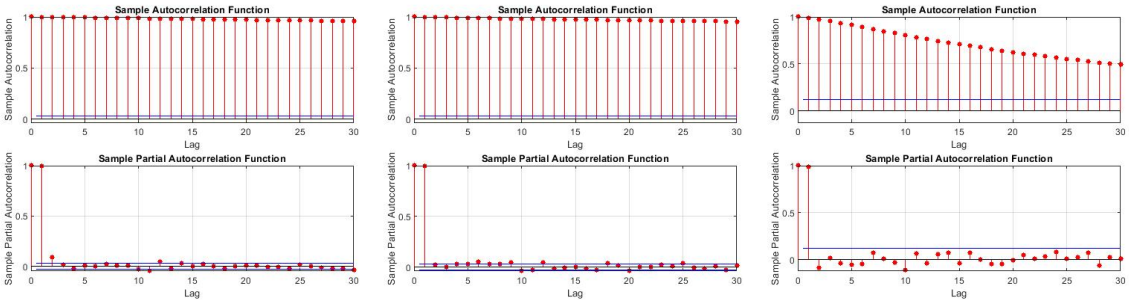

Figura 9– Correladogramas das séries estudadas, da esquerda para a direita: PETR4.SA, ITUB4.SA e 1810.HK.

Estão dispostos na Tabela 2 os dois melhores resultados obtidos com modelos ARIMA para cada série.

Tabela 2: Modelos ARIMA e seus desempenhos.

Série Temporal / Modelo	**Resíduos do ajuste**		**Erros de previsão**	
PETR4.SA	RMSE	MAPE	RMSE	MAPE
ARIMA(2,0,0) – Modelo 1	0.5361	0.0187	7.1418	0.3409
ARIMA(12,0,0) – Modelo 2	0.5343	0.0186	7.0768	0.3373
ITUB4.SA	RMSE	MAPE	RMSE	MAPE
ARIMA(7,0,0) – Modelo 1	0.2909	0.0161	8.5604	0.2368
ARIMA(1,1,0) – Modelo 2	0.2907	0.0161	5.7901	0.1568
1810.HK	RMSE	MAPE	RMSE	MAPE
ARIMA(1,1,0) – Modelo 1	0.3808	0.0215	1.002	0.1005
ARIMA(1,0,1) – Modelo 2	0.3799	0.0213	0.258	0.0237

As previsões 1 e 2, para cada uma das três séries foram realizadas, respectivamente, com os modelos 1 e 2 e os resultados constam nas figuras a seguir.

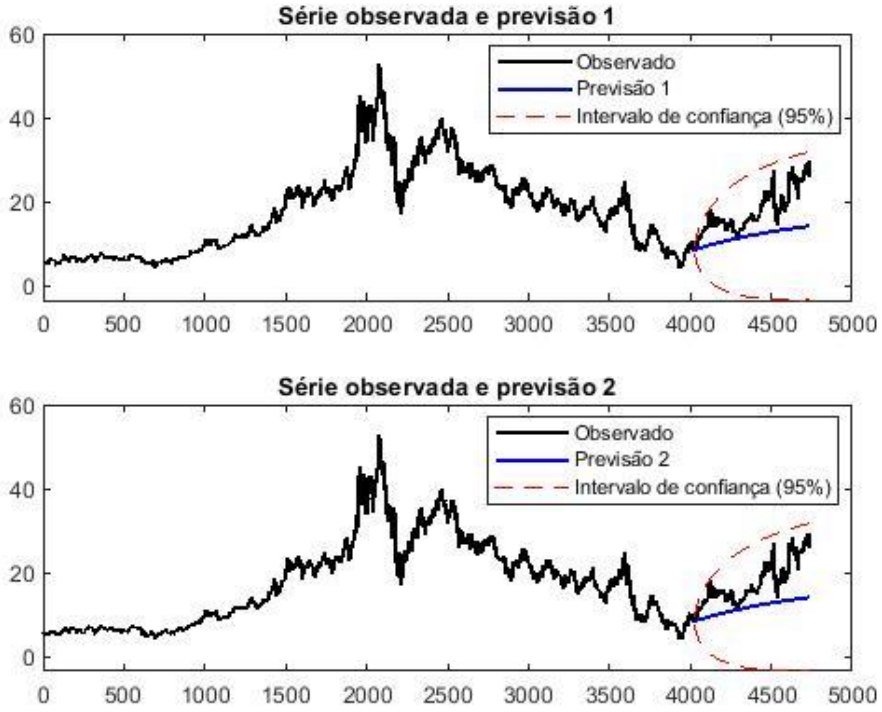

Figura 10 – Previsão (preço x tempo) dos modelos ARIMA em trecho de teste da série PETR4.SA.

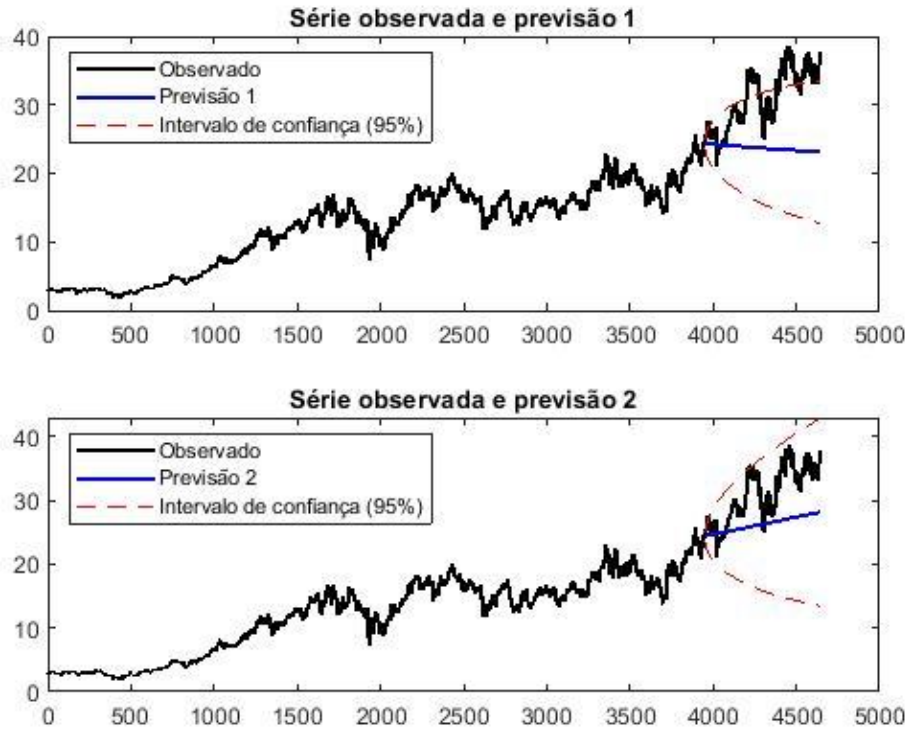

Figura 11 – Previsão (preço x tempo) dos modelos ARIMA em trecho de teste da série ITUB4.SA.

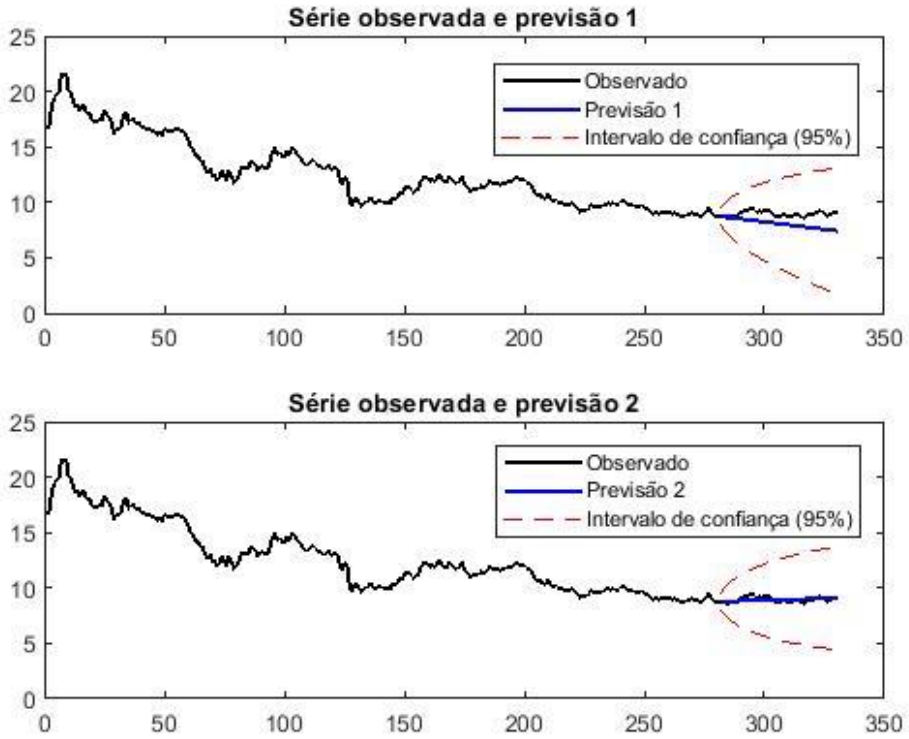

Figura 12 – Previsão (preço x tempo) dos modelos ARIMA em trecho de teste da série 1810.HK.

5.2 ARIMA x GARCH

Foi selecionado o modelo ARIMA de melhor desempenho na etapa anterior para analisar se a mudança da variância no modelo, de constante, para uma versão GARCH melhoraria o desempenho. Na Tabela 3 constam as comparações entre os modelos ARIMA escolhidos e os melhores resultados obtidos com a consideração de heteroscedasticidade.

Tabela 3: Modelos ARIMA-GARCH e seus desempenhos.

Série temporal / Modelo	**Resíduos do ajuste**		**Erros de previsão**	
PETR4.SA	RMSE	MAPE	RMSE	MAPE
ARIMA (12,0,0) – Modelo 1	0.5343	0.0186	7.0768	0.3373
+ GARCH (2,1) – Modelo 2	0.5359	0.0186	8.2012	0.3997
ITUB4.SA	RMSE	MAPE	RMSE	MAPE
ARIMA (1,1,0) – Modelo 1	0.2917	0.0161	5.7901	0.1568
+ GARCH (2,1) – Modelo 2	0.2918	0.0161	6.0112	0.1632
1810.HK	RMSE	MAPE	RMSE	MAPE
ARIMA (1,0,1) – Modelo 1	0.3799	0.0213	0.2580	0.0237
+ GARCH (4,1) – Modelo 2	0.3812	0.0213	0.3610	0.0336

A ausência de melhora no desempenho pode ser constatada nos gráficos dispostos nas figuras a seguir, nas quais, as previsões 1 e 2 foram realizadas com os modelos 1 e 2, respectivamente.

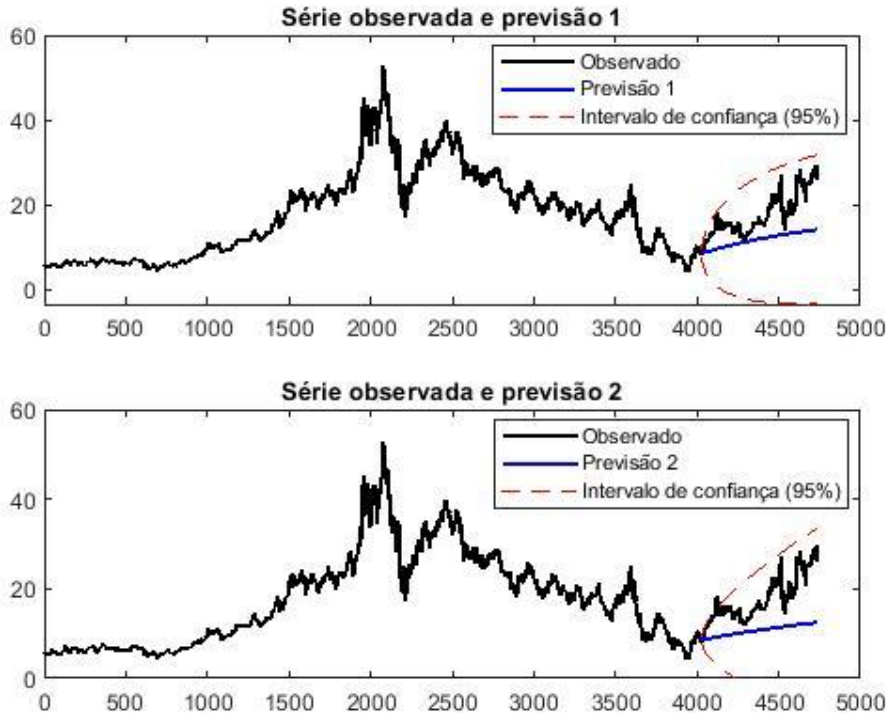

Figura 13 – Previsão (preço x tempo) dos modelos ARIMA com modelo de variância GARCH em trecho de teste da série PETR4.SA.

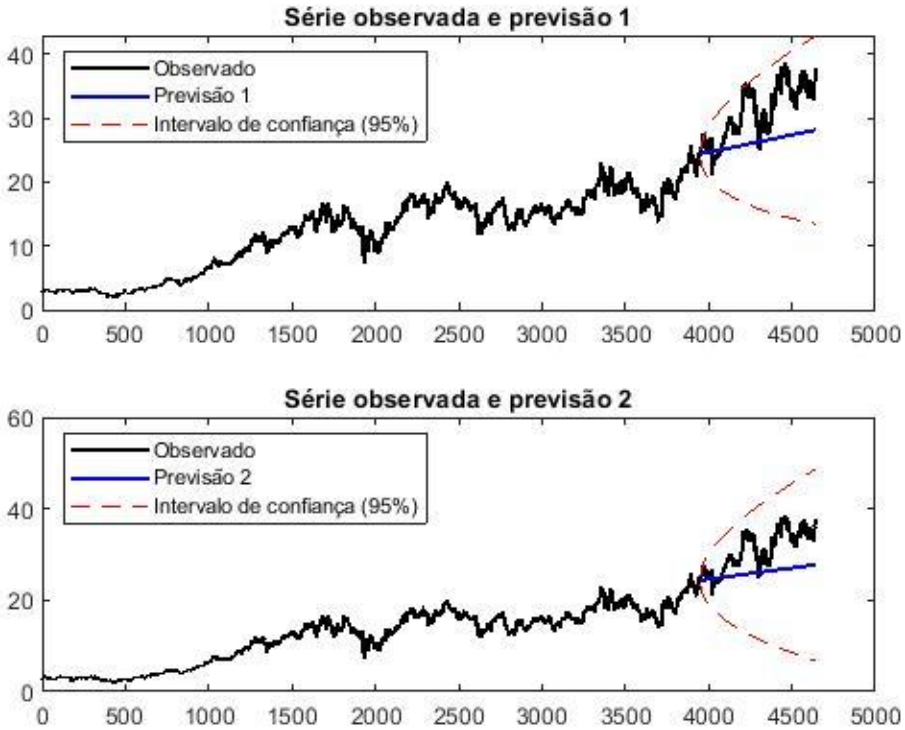

Figura 14 – Previsão (preço x tempo) dos modelos ARIMA com modelo de variância GARCH em trecho de teste da série ITUB4.SA.

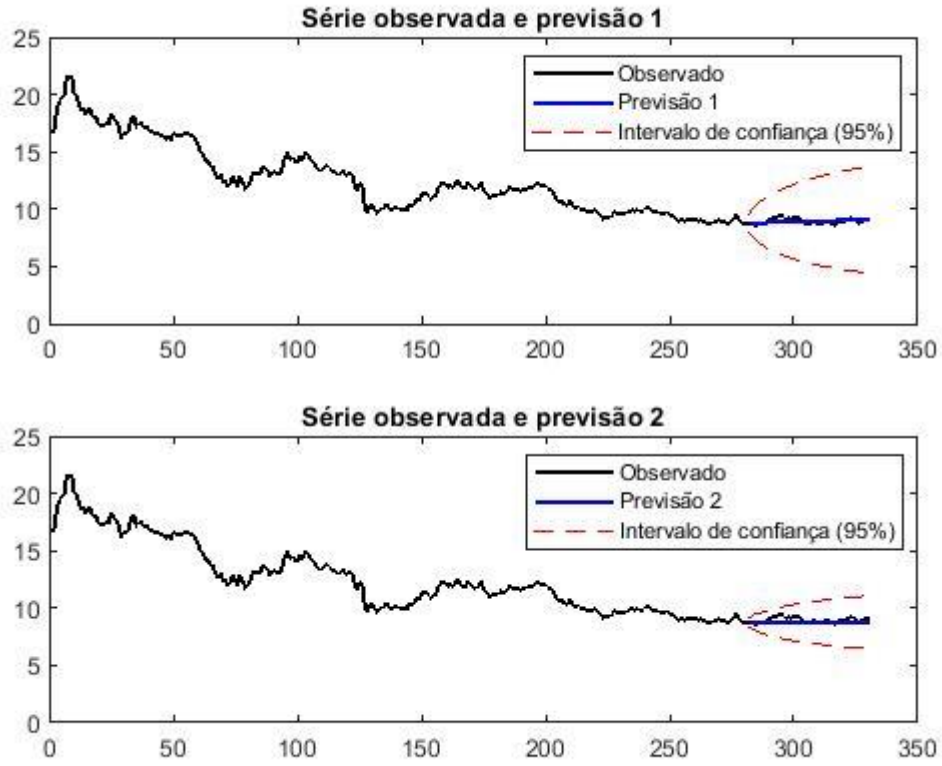

Figura 15 – Previsão (preço x tempo) dos modelos ARIMA com modelo de variância GARCH em trecho de teste da série 1810.HK.

5.3 NARNET

Os resultados para cada ordem do operador de atraso foram coletados e comparados. O desempenho não apresentou melhora proporcional ao aumento da ordem de atraso para nenhuma série, de modo que dentre os 20 níveis executados foram destacados os resultados de melhor ajuste e melhor previsão. O resultado médio (considerando todas as ordens de atraso), para efeito de controle, foi registrado.

Tabela 4: NARNETs e seus desempenhos.

Série temporal / Modelo	**Resíduos do ajuste**		**Erros de previsão**	
PETR4.SA	RMSE	MAPE	RMSE	MAPE
Melhor ajuste: NARNET (atraso de ordem 20)	0.5153	0.0184	4.9000	0.2284
Melhor previsor: NARNET (atraso de ordem 4)	0.5218	0.0185	4.0319	0.1975
Desempenho médio:	0.5295	0.0188	9.3765	0.4651
ITUB4.SA	RMSE	MAPE	RMSE	MAPE
Melhor ajuste: NARNET (atraso de ordem 17)	0.2768	0.0157	12.2925	0.6225
Melhor previsor: NARNET (atraso de ordem 13)	0.2903	0.0162	5.8670	0.1569
Desempenho médio:	0.2920	0.0164	9.2816	0.2895
1810.HK	RMSE	MAPE	RMSE	MAPE
Melhor ajuste: NARNET (atraso de ordem 8)	0.2993	0.0180	0.3303	0.0314
Melhor previsor: NARNET (atraso de ordem 13)	0.3255	0.0193	0.2554	0.0228
Desempenho médio:	0.3572	0.0211	1.0926	0.1134

Para a série PETR4.SA, a rede de melhor ajuste obteve uma capacidade de previsão muito acima da média que, todavia, foi superada na capacidade de previsão por uma outra versão da rede que também apresentou boa capacidade de ajuste. Os resultados de previsão são confirmados pelos gráficos que se encontram na Figura 16.

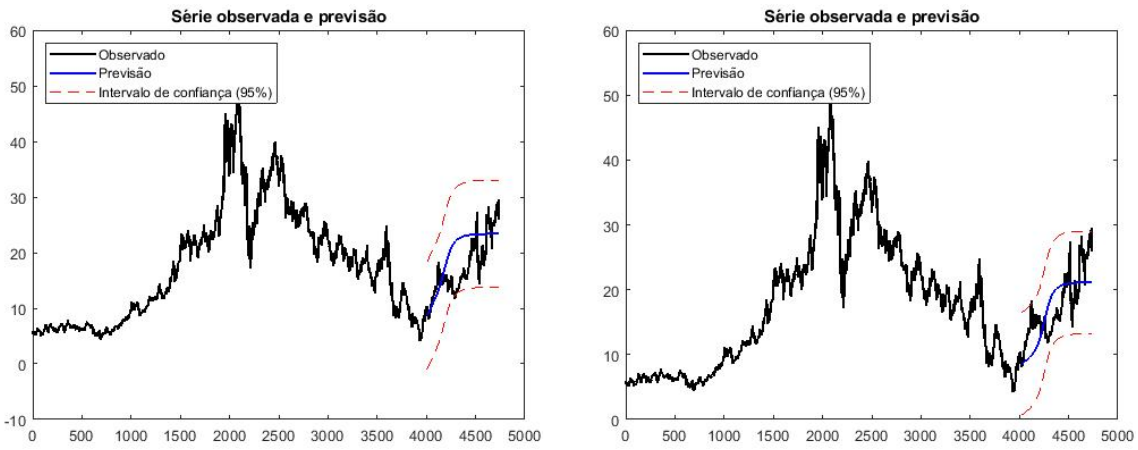

Figura 16 – Previsão (preço x tempo) das NARNET, da esquerda para a direita, de ordem 20 e 4, respectivamente, a de melhor ajuste e a melhor preditora para a série PETR4.SA.

A série ITUB4.SA teve pouco das suas características apreendidas pela rede, de modo que o melhor ajuste não gerou uma previsão satisfatória. Mais ainda: o melhor preditor praticamente não acompanhou a tendência da série (Figura 17).

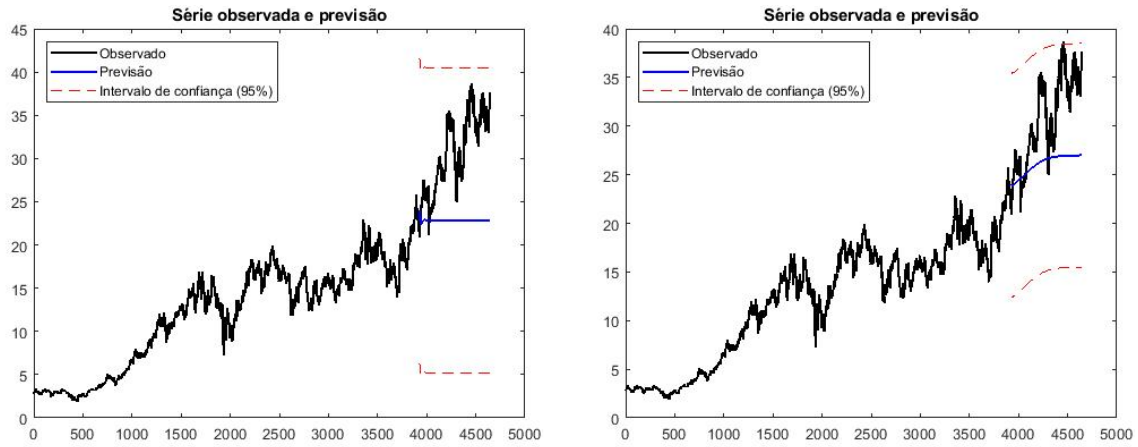

Figura 17 – Previsão (preço x tempo) das NARNET, da esquerda para a direita, de ordem 17 e 13, respectivamente, a de melhor ajuste e a melhor preditora para a série ITUB4.SA.

Finalmente, a série cujas características foram mais bem apreendidas por um modelo NARNET foi a 1810.HK. Ainda que o desempenho médio de ajuste não tenha sido melhor que o da série vista acima, todas as ordens de atraso se ajustaram próximas à média, de modo que a média, e a máxima de previsão foram a melhor dentre as séries estudadas (Figura 18).

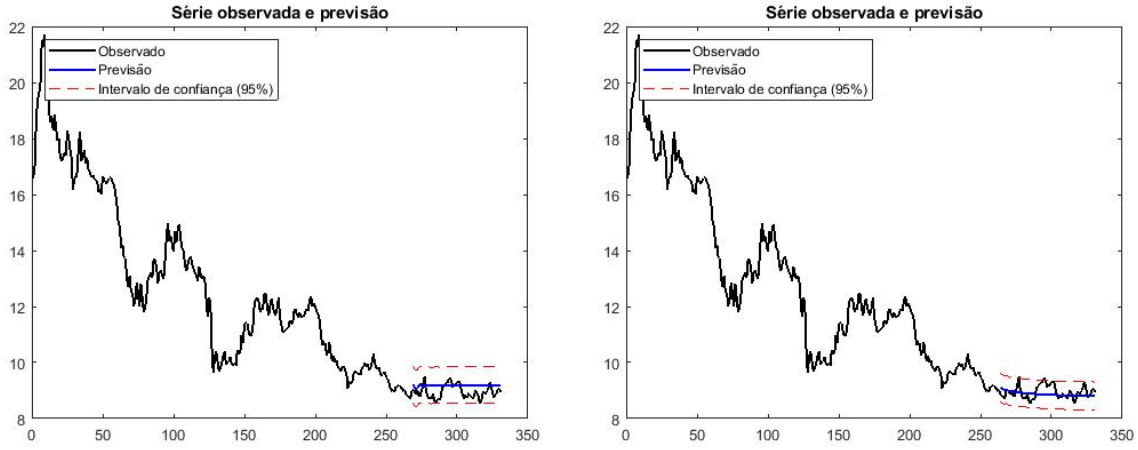

Figura 18 – Previsão (preço x tempo) das NARNET, da esquerda para a direita, de ordem 8 e 13, respectivamente, a de melhor ajuste e a melhor preditora para a série 1810.HK.

5.4 LSTM

Por último, foram utilizadas as LSTM para as tarefas de modelagem e previsão das séries consideradas. Os resultados foram organizados na Tabela 5.

Tabela 5: LSTM e seus desempenhos.

Série temporal	**Resíduos do ajuste**		**Erros de previsão**	
	RMSE	MAPE	RMSE	MAPE
PETR4.SA	0.8337	0.0286	8.6467	0.4608
	RMSE	MAPE	RMSE	MAPE
ITUB4.SA	0.4688	0.0277	8.8567	0.2580
	RMSE	MAPE	RMSE	MAPE
1810.HK	0.4046	0.0231	0.2509	0.0219

A série PETR4.SA (Figura 19) não obteve bom desempenho de previsão, provavelmente, devido ao fato de que o comportamento da série no período de previsão ter tido uma tendência diferente da observada no período de aprendizado. A série ITUB4.SA (Figura 20), mas, principalmente, a série 1810.HK (Figura 21), por apresentarem as mesmas tendências do passado, em maior ou menor grau, tiveram previsões com métrica percentual de erro mais baixas.

Figura 19 – Previsão (preço x tempo) das LSTM em trecho de teste da série PETR4.SA.

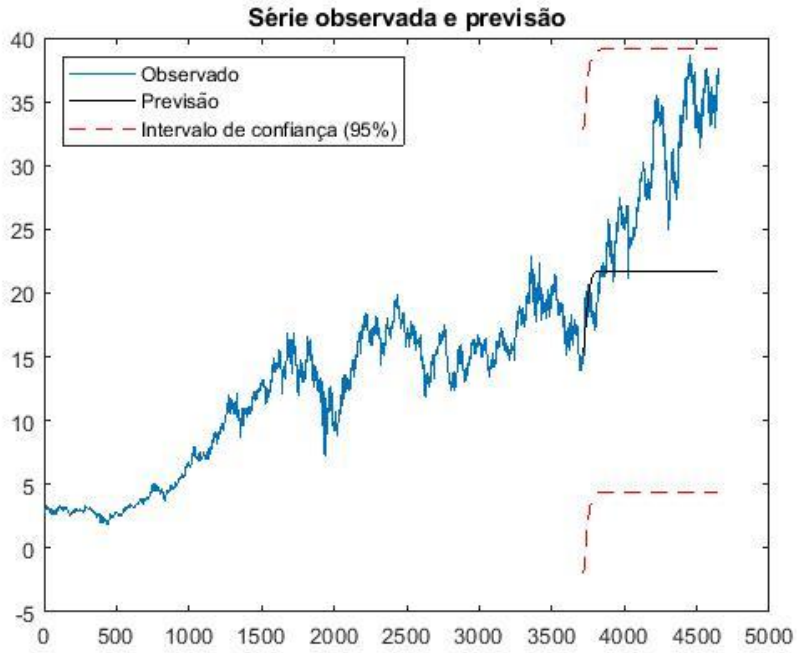

Figura 20 – Previsão (preço x tempo) das LSTM em trecho de teste da série ITUB4.SA.

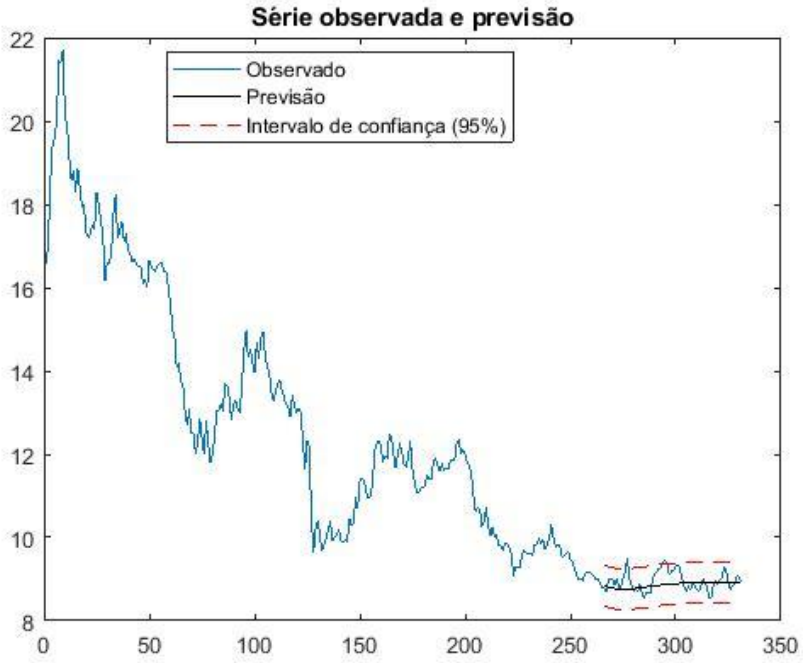

Figura 21 – Previsão (preço x tempo) das LSTM em trecho de teste da série 1810.HK.

5.5 Comparação dos métodos e discussão

Considerando-se o modelo de melhor desempenho para cada técnica, verificou-se capacidade de ajuste das técnicas foi similar, exceto para as LSTM, que não ajustaram satisfatoriamente a série com mudança de tendência.

Para a PETR4.SA, os melhores ajustes apresentaram erro entre 1.84% e 1.86% e raiz quadrada do erro médio em torno de 0,52, ou seja, R$ 0,52 (para uma série que variou entre R$ 5,55 e R$ 52,58). O melhor desempenho foi da NARNET com 20 atrasos na realimentação, atingindo pouco menos de R$ 0,52.

Os melhores modelos para a série ITUB4.SA, com exceção da LSTM, apresentaram erro entre 1,57% e 1,61%, com raiz quadrada do erro médio em torno de R$ 0,29 (a série real tem valores entre R$ 1,89 e R$ 38,67). O melhor desempenho foi da NARNET com 17 atrasos, pouco menos de R$ 0,28.

A série 1810.HK teve, com os melhores modelos, erro entre 1,8% e 2,31%, com raiz quadrada do erro médio variando de pouco menos de R$ 0,30 a pouco mais de R$ 0,40 (a série real tem valores entre R$ 8,53 e R$ 21,70). O melhor desempenho foi da NARNET com 8 atrasos: pouco menos de R$ 0,30.

Tabela 6: Desempenhos dos modelos destacados na tarefa de ajuste.

Série temporal/Modelo	Resíduos do ajuste	
PETR4.SA	RMSE	MAPE
ARIMA (12,0,0)	0.5343	0.0186
ARIMA (12,0,0) + GARCH (2,1)	0.5359	0.0186
NARNET (atraso de ordem 20)	0.5153	0.0184
LSTM	0.8337	0.0286
ITUB4.SA		
ARIMA (1,1,0)	0.2917	0.0161
ARIMA (1,1,0) + GARCH (2,1)	0.2918	0.0161
NARNET (atraso de ordem 17)	0.2768	0.0157
LSTM	0.4688	0.0277
1810.HK		
ARIMA(1,0,1)	0.3799	0.0213
ARIMA (1,0,1) + GARCH (4,1)	0.3812	0.0213
NARNET (atraso de ordem 8)	0.2993	0.0180
LSTM	0.4046	0.0231

A capacidade de previsão apresentou maior variação entre os modelos do que a capacidade de ajuste. As NARNET apresentaram o melhor de desempenho de previsão (Tabela 7).

Os melhores resultados foram, conforme a Tabela 7, aproximadamente: 19,75% e RMSE de R$ 4,03 com NARNET de ordem 4 de atraso para a série PETR4.SA, 15,68% e RMSE de R$ 5,79 com ARIMA(1,1,0) para a série ITUB4.SA e 2,19% e RMSE de R$ 0,25 com LSTM para a série 1810.HK.

Tabela 7: Desempenhos dos modelos destacados na tarefa de previsão.

Série temporal/Modelo	Erros de Previsão	
PETR4.SA	RMSE	MAPE
ARIMA (12,0,0)	7.0768	0.3373
ARIMA (12,0,0) + GARCH (2,1)	8.2012	0.3997
NARNET (atraso de ordem 4)	4.0319	0.1975
LSTM	8.6467	0.4608
ITUB4.SA		
ARIMA (1,1,0)	5.7901	0.1568
ARIMA (1,1,0) + GARCH (2,1)	6.0112	0.1632
NARNET (atraso de ordem 13)	5.8670	0.1569
LSTM	8.8567	0.2580
1810.HK		
ARIMA(1,0,1)	0.2580	0.0237
ARIMA(1,0,1) + GARCH(4,1)	0.3610	0.0336
NARNET (atraso de ordem 13)	0.2554	0.0228
LSTM	0.2509	0.0219

6. Conclusão

Neste trabalho, compararam-se os desempenhos de métodos clássicos e de redes neurais artificiais para ajuste e previsão de valores futuros de séries temporais financeiras. Verificou-se que as redes neurais artificiais têm capacidade de ajuste e previsão possivelmente superior ao de modelos clássicos. As técnicas baseadas em redes neurais, no entanto, em contraste com os métodos clássicos, padecem da falta de métodos quantitativos para a determinação de seus parâmetros, como número de camadas e de neurônios, o que leva a variações imprevisíveis nos resultados obtidos a depender das escolhas heurísticas desses parâmetros.

7. Referências bibliográficas

[1] *Google Trends*. Disponível em: <https://trends.google.com/trends/explore?date=now%207-d&geo=BR&q=futebol>. Acesso em: 08 abr. 2019.

[2] Castro, H. O. P. *Introdução ao Mercado de Capitais*, ed. 10. Instituto Brasileiro de Mercado de Capitais – IBMEC, 1979.

[3] Lathi, B. P. *Sinais e Sistemas Lineares.* ed. 2. Bookman, 2004.

[4] Diniz, P. S. R., Silva, E. A. B. and Netto, S. L. *Processamento Digital de Sinais – Projeto e Análise de Sistemas*. ed. 2. Bookman, 2014.

[5] Hayes, M. H. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[6] Morettin, P. A. and Toloi, C. M. C. *Análise de Séries Temporais*. ed 2. Editora Blucher, 2006.

[7] Bueno, R. L. S. *Econometria de Série Temporais.* ed 2. Cengage Learning, 2008.

[8] *Datascience: 7 Ways Time Series Forecasting Differs from Machine Learning*. Disponível em: <https://www.datascience.com/blog/time-series-forecasting-machine-learning-differences>. Acesso em: 10 mar. 2019.

[9] Enders, W. *Applied Econometric Time Series*. ed.4. Wiley, 2015.

[10] Taylor, S. J. *Modelling Financial Time Series*. ed. 2.World Scientific, 2008.

[11] Box, G. E. P. and Jenkins, G. M. *Time Series Analysis – forecasting and control*. Holden-Day, 1976.

[12] Engle, R. F. *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation* In Econometrica, pgs. 987-1008. vol.50, Jul. 1982.

[13] Bollerslev, T. *Generalized Autoregressive Conditional Heteroskedasticity* In Journal of Econometrics, pgs. 307-327. vol. 31. Fev, 1986.

[14] Bussab, W. O. and Morettin, P. A. *Estatística Básica*. ed. 6. Editora Saraiva, 2010.

[15] Ross, S. *Probabilidade – Um curso moderno com aplicações.* ed. 8. Bookman, 2010.

[16] Haykin, S. *Redes Neurais – Princípios e prática*. ed. 2. Bookman, 2007.

[17] Goodfellow, I., Bengio, Y. and Courville,A. *Deep Learning*. MIT Press, 2016.

[18] Patterson, J. and Gibson, A. *Deep Learning – A Practitioner's approach*. O'Reilly, 2017.

[19] Ruiz, L. G. B. et al. *An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings*. University of Granada, 2016.

[20] *MathWorks Documentation – "layrecnet"* Disponível em: <https://www.mathworks.com/help/deeplearning/ref/layrecnet.html>. Acesso em: 18 set. 2019.

[21] *MathWorks Documentation – "narnet"* Disponível em: <https://www.mathworks.com/help/deeplearning/ref/narnet.html?searchHighlight=narnet&s_tid=doc_srchtitle>. Acesso em: 20 set. 2019.

[22] Hochreiter, S., & Schmidhuber, J. *Long Short-Term Memory*. Neural Computation, 9(8), pgs. 1735–1780. 1997

[23] Million, E.. *The Hadamard Product Elizabeth Million April 12 , 2007 1 Introduction and Basic Results*. 2007.

[24] *Yahoo Finanças: Petróleo Brasileiro S.A. - Petrobras*. Disponível em: <https:// https:// https://br.financas.yahoo.com/quote/PETR4.SA/history?p=PETR4.SA>. Acesso em: 18 abr. 2019.

[25] *MathWorks Documentation – "estimate"* Disponível em: <https://www.mathworks.com/help/econ/arima.estimate.html?searchHighlight=estimate&s_tid=doc_srchtitle>. Acesso em: 18 ago. 2019.

[26] Brooks, C. *Introductory Econometrics for Finance*. Cambridge, UK: Cambridge University Press, 2002.