XXIII SemeAd
Seminários em Administração

novembro de 2020
ISSN 2177-3866

Sorte ou Técnica? Um estudo sobre o desempenho do mercado de ações à luz de diferentes técnicas de previsão

AHMED SAMEER EL KHATIB
CENTRO UNIVERSITÁRIO ÁLVARES PENTEADO (FECAP)

Sorte ou Técnica? Um estudo sobre o desempenho do mercado de ações à luz de diferentes técnicas de previsão

1. INTRODUÇÃO

Estudos precedentes, teóricos e empíricos, evidenciaram que existe uma relação positiva entre os mercados financeiros e o crescimento econômico (p.ex., Levine, 1997; Rajan & Zingales, 1998; Rousseau & Watchel, 2000; Beck & Levine, 2003; Guptha & Rao, 2018). Dada à importância dos mercados financeiros, a previsão de retornos das ações ocupa uma posição de destaque na tomada de decisão sobre investimentos. No entanto, os mercados de ações são caracterizados por alta volatilidade (causada por crises financeiras internacionais, locais ou Pandemias, como no recente caso da COVID-19), dinamismo e complexidade (Johnson, Jefferies & Hui, 2003; Cristelli, 2014; Wieland, 2015). Os movimentos nas bolsas de valores são influenciados por vários fatores, como fatores macroeconômicos, eventos internacionais e comportamento humano. Portanto, prever o retorno do preço das ações pode se tornar uma tarefa desafiadora. A lucratividade dos investimentos nas bolsas de valores depende muito da previsibilidade dos movimentos das ações. Caso um modelo ou técnica de previsão possa prever com precisão a direção do mercado, o risco e a incerteza do investimento poderão ser minimizados, por conta da ferramenta e não por um golpe de sorte. Melhorar os fluxos de investimento no mercado de ações também pode ser útil para formuladores de políticas e reguladores na tomada de decisões apropriadas, sejam elas preventivas ou corretivas.

Existem duas escolas de pensamento distintas para avaliação do mercado de ações: a análise fundamentalista e a análise técnica. Os fundamentalistas preveem os preços das ações com base em análises financeiras de empresas ou indústrias. Do outro lado da esteira, os analistas técnicos usam dados históricos de títulos e preveem preços futuros com base no pressuposto de que os preços das ações são determinados pelas forças do mercado e que a história tende a se repetir (Levy, 1967). Essas teorias coexistiram por várias décadas como estratégias para a tomada de decisões de investimento. Essas abordagens foram desafiadas na década de 1960 pela Teoria da Caminhada Aleatória, popularmente conhecida como Hipótese do Mercado Eficiente (Fama, 1970), que propõe que mudanças futuras nos preços das ações não podem ser previstas a partir de mudanças passadas nos preços. Alguns estudos empíricos mostraram a presença do chamado "”passeio aleatório” nos preços das ações (p.ex., Tong, Li & Benkato, 2014; Konak & Seker, 2014; Erdem & Ulucak, 2016). No entanto, a maioria dos estudos empíricos descobriu que os preços das ações são previsíveis (Darrat & Zhong, 2000; Lo & MacKinlay, 2002; Harrison & Moore, 2012; Owido, Onyuma & Owuor, 2013; Radikoko, 2014; Said, 2015; Almudhaf, 2018).

Várias técnicas de previsibilidade estão disponíveis para previsão de séries temporais. Os Modelos de Média Móvel Integrada Autoregresiva (ARIMA) foram propostos por Box e Jenkins (1970) para análise e previsão de séries temporais. Alguns estudos foram realizados empregando modelos ARIMA para prever retornos do mercado de ações (Al-Shaib, 2006; Ojo & Olatayo, 2009; Adebiyi & Oluinka, 2014; Mondal, Shit & Goswami, 2014). Poucos estudos descobriram que os modelos ARIMA produziam previsões inferiores para dados de séries temporais financeiras (Zhang, 2003; Adebiyi & Oluinka, 2014; Khandelwal & Adhikari, 2015). Para explicar as não linearidades resultantes de mudanças de regime nas economias, alguns pesquisadores usaram Modelos de Série Temporal com Mudança de Regime de Markov (*Markov Switching Models)* e Modelo Auto-Regressivo com Limiar (TAR), assumindo processos estacionários não lineares para prever os preços das ações (Hamilton, 1989; Tong, 1990).

Tasy (1989) propôs um procedimento de construção de modelo simples, mas amplamente aplicável, para Modelos Autoregressivos com Limiar, bem como um teste para não-linearidade de limiar. Gooijer (1998) considerou a troca de regime em um modelo de Média Móvel (MA) e usou critérios de validação para a seleção do Modelo SETAR (*Self-Exciting Threshold AutoRegressive*). Alguns estudos empíricos comparando métodos diferentes com o SETAR descobriram que esse método produzia resultados superiores aos modelos lineares (p,e., Clements & Smith, 1999; Boero & Marrocu, 2002; Boero, 2003; Firat, 2017).

No final dos anos 80, uma classe de modelos de Inteligência Artificial (IA), como Rede *Feedforward*, *Backpropagation* e modelos de redes neurais recorrentes, foram introduzidos para fins de previsão no mercado de ações. As características distintivas das Redes Neurais Artificiais (RNA) são que elas são orientadas por dados não-lineares e auto-adaptáveis, além de possuírem poucas premissas a priori. Isso torna as RNA valiosas e atraentes para a previsão de séries temporais financeiras. Entre os modelos de RNA, a rede neural *Feedforward* com uma única camada oculta se tornou a mais popular para prever retornos do mercado de ações (Zhang, 2003). Muitos estudos mostraram que esses modelos produzem previsões mais precisas em comparação com modelos ingênuos e lineares (p.e., Ghiassi et al., 2005; Mostafa, 2010; Qosta et al., 2016; Aras & Kocakoc, 2016).

Além disso, existem diversos modelos de redes neurais para prever retornos do mercado de ações. Lu e Wu (2011) usaram o modelo de rede neural do Controlador de Articulação do Modelo Cerebelar (CAMC NN) para prever os índices de mercado de ações do Nikkei 225 e da Bolsa de Valores de Taiwan. Os resultados mostraram que o CAMC NN produziu previsões mais precisas do que os modelos de regressão vetorial e Rede Neural de Propagação Retro Propação (BPNN). Guresen, Kayakutlu e Daim (2011) observaram que os modelos clássicos de RNA e a percepção de multicamadas (MLP) superaram os modelos da classe GARCH para o índice NASDAQ. Lahmiri (2016) empregou Redes Neurais de Regressão Geral (GRNN) baseadas em Decomposição de Modo Variacional (VMD) para quatro conjuntos de dados econômicos e financeiros e descobriu que os modelos VMD-GRNN superaram o modelo ARIMA e outros modelos de redes neurais. A Rede Neural Polinomial Condensada baseada em algoritmo genético de Nayak e Misra (2018) (GA-CPNN) melhorou a precisão dos índices de previsão do preço de ações quando comparados aos modelos de Rede Neural de Função de Base Radial (RBFNN), *Perceptron* Multicamada e Algoritmo Genético (MLP-GA). Zhong e Enke (2019) observaram que técnicas como redes neurais profundas usando Análise de Componentes Principais ou *Principal Component Analysis* (PCA) e redes neurais artificiais tiveram um desempenho melhor do que os modelos tradicionais. No entanto, a maioria dos estudos descobriu que os modelos tradicionais de RNA, bem como os modelos de RNA combinados com modelos lineares, produzem previsões mais precisas do que outros modelos (p.e., Asadi, Tavakoli & Hejazi, 2010; Wang, Wang, Zhang & Guo, 2011; Khandelwal & Adhikari, 2015; Mallikarjuna, Arti & Guptha, 2017).

Recentemente, Modelos no Domínio da Frequência, como Análise Espectral, *Wavelets* e Transformadas de *Fourier*, foram propostos para melhorar a precisão da previsão de séries temporais financeiras. Uma técnica amplamente utilizada é a Análise de Espectro Singular ou *Singular Spectrum Analysis* (SSA), que é um método não paramétrico robusto, sem suposições prévias sobre os dados (Golyandina, Nekrutkin & Zhigljavsky, 2001; Hassani, Soofi & Zhiglavsky, 2013a). O SSA decompõe os dados da série temporal em seus componentes e depois reconstrói a série deixando o componente de ruído aleatório antes de usar a série reconstruída para prever os pontos futuros da série (Hassani, 2007; Ghodsi & Omer, 2014). Como a maioria dos conjuntos de dados de séries temporais financeiras não exibe padrões puramente lineares nem puramente não-lineares, a combinação de técnicas lineares e não-lineares, isto é, técnicas híbridas para modelar estruturas de dados complexas para maior precisão, foi proposta (Asadi et al., 2010; Khashei & Bijari, 2010; Khashei & Bijari, 2012; Khandelwal & Adhikari, 2015; Ince & Trafalis, 2017). Khashei e Hajirahimi (2017) compararam modelos lineares e não lineares com Modelos Híbridos (*Hybrid Model* - HM) e concluíram que os modelos híbridos apresentam desempenho melhor que os modelos individuais.

Apenas alguns estudos têm procurado encontrar um método adequado para prever o retorno das ações de um grupo de mercados. Guidolin, Hyde, McMillan e Ono (2009) avaliaram o desempenho de modelos lineares e não lineares para prever o retorno de ativos financeiros dos países do G7. Eles descobriram que os modelos não lineares, como os TAR e STAR, tiveram um desempenho melhor que os modelos lineares no caso de retorno de ativos nos EUA e no Reino Unido. Enquanto isso, modelos lineares simples eram melhores para retornos de ativos nos mercados francês, alemão e italiano. Isso sugere que nenhum modelo é adequado para prever os retornos de todos os mercados de ações. Awajan et al. (2018) compararam o desempenho de vários métodos de previsão aplicando-os a seis bolsas de valores e descobriram que o método empírico de Decomposição do Método de *Holt-Winters* (EMD-HW) forneceu previsões mais precisas do que outros modelos.

Embora existam várias técnicas para prever retornos do mercado de ações, nenhum método único pode ser empregado uniformemente para os retornos de todos os mercados e pode-se afirmar que não existe previsibilidade com base do fator sorte. A literatura indica que não há consenso entre os pesquisadores sobre as técnicas de previsão de retorno do mercado de ações. O presente estudo, portanto, teve como objetivo avaliar diferentes técnicas de previsão, a saber: modelos ARIMA, SETAR, ANN, SSA e HM, representando inteligência linear, não linear, inteligência artificial (IA), domínio da frequência e métodos híbridos, respectivamente, aplicando-os a mercados de ações individuais. Este estudo também examinou a adequação de diferentes métodos de previsão para cada categoria dos mercados de ações mundiais - a saber, desenvolvidos, emergentes e de fronteira. Encontrar um método único que possa produzir previsões ideais para todos os mercados pode ajudar os investidores a economizar tempo e recursos e tomar melhores decisões. Este estudo é útil principalmente para investidores internacionais e investidores institucionais estrangeiros que desejam minimizar riscos e diversificar suas carteiras, com o objetivo de maximizar lucros. Dessa forma, os objetivos do presente estudo, que serão alcançados ao longo das próximas seções, são:

Objetivo 1: Prever retornos do mercado de ações usando inteligência linear, não linear, artificial, domínio da frequência e métodos híbridos;

Objetivo 2: Encontrar as técnicas de previsão mais apropriadas entre as cinco técnicas exploradas para mercados desenvolvidos, emergentes e de fronteira; e

Objetivo 3: Verificar se uma única técnica pode ser aplicada a todos os mercados (desenvolvidos, emergentes e de fronteira) para obter previsões ideais.

2. DADOS E METODOLOGIA

De acordo com os objetivos deste estudo, consideramos três tipos de mercados, baseados na classificação do *Morgan Stanley Capital International* (MSCI, 2019): desenvolvidos, emergentes e de fronteira. Os índices de mercado adotados para a categoria desenvolvida são Austrália (ASX 200), Canadá (TSX *Composite*), França (CAC 40), Alemanha (DAX), Japão (NIKKEI 225), Coréia do Sul (KOSPI), Suíça (SMI), Estados Unidos Reino Unido (FTSE 100) e Estados Unidos (S&P 500). Os mercados emergentes são Brasil (BOVE-SPA), China (SSEC), Egito (EGX 30), Índia (SENSEX), Indonésia (IDX), México (BMV IPC), Rússia (MOEX), Rússia (MOEX), África do Sul (JSE 40), Tailândia (SET) e Turquia (BIST 100). Por fim, os da categoria fronteira são Argentina (S&P MERVAL), Estônia (TSEG), Quênia (NSE 20), Sri Lanka (CSE AS) e Tunísia (TUNINDEX). Os preços diários de fechamento desses índices para o período de 1º de janeiro de 2000 a 31 de dezembro de 2019 foram obtidos no site www.investing.com.

Os retornos dos ativos (R_t) foram calculados a partir dos preços de fechamento de todos os índices, usando a fórmula apresentada na Equação 1 a seguir:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} * 100 \tag{1}$$

Onde, P_t é o preço do ativo no período atual e P_{t-1} é o preço de um ativo no período anterior.

2.1 Média móvel integrada autorregressiva (ARIMA)

Propostos por George Box e Gwilym Jenkins em 1970, os modelos ARIMA estão entre os modelos lineares mais populares. Nos modelos ARIMA, o valor futuro de uma variável é obtido através de uma função linear de algumas observações anteriores da variável e de alguns erros aleatórios. O processo que gera a série temporal é apresentado na Equação 2 a seguir:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots.., \emptyset_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \ldots.., \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2}$$

Onde y_t é a variável que será explicada no tempo t; c é a constante ou o intercepto; ϕi (i = 1, 2,… p) e θj (j = 1, 2,… q) são os parâmetros do modelo; p e q são números inteiros e são frequentemente referidos como ordens AR e MA do modelo, respectivamente; e e_t é o termo do erro. O pressuposto relativo aos erros aleatórios ε_t é que eles são distribuídos de forma independente e idêntica com um zero médio e uma variação constante de σ^2. Esse modelo envolve um processo iterativo de três etapas de identificação, estimativa e verificação de diagnóstico.

A etapa de identificação envolve a especificação de um modelo experimental, decidindo a ordem dos termos de AR (p) e MA (q). Depois que um modelo experimental é especificado, os parâmetros do modelo devem ser estimados, de modo que a medida geral de erros seja minimizada, o que geralmente é feito com um procedimento de otimização não linear.

Após a estimativa dos parâmetros, deve-se verificar o diagnóstico da adequação do modelo, o que envolve testar se as suposições do modelo sobre os erros ε_t são atendidas. Se o modelo for adequado, pode-se proceder à previsão; caso contrário, um novo modelo experimental deve ser identificado após a estimativa do parâmetro e a verificação do modelo. Esse processo com três etapas deve ser repetido até que um modelo satisfatório seja selecionado para prever os dados.

2.2 Self-Exciting Threshold Autoregressive (SETAR)

O modelo SETAR, desenvolvido por Tong (1983), é um tipo de modelo autoregressivo que pode ser aplicado aos dados de séries temporais. Este modelo tem mais flexibilidade nos parâmetros que apresentam comportamento de mudança de regime (Watier & Richardson, 1995). A alternância de regime neste modelo é baseada na autodinâmica da variável dependente. Em outras palavras, o valor limite no modelo SETAR está relacionado à variável endógena, enquanto no Modelo TAR está relacionado a uma variável exógena. Este modelo assume um processo autoregressivo diferente, de acordo com valores limite específicos. Os modelos SETAR têm a vantagem de capturar um fenômeno não linear comumente observado que não pode ser capturado por modelos lineares como suavização exponencial e modelos ARIMA. Um modelo de autorregressão de limite pode ser transformado em um modelo SETAR se a variável de limite for tomada como um valor atrasado da própria série temporal. O modelo SETAR com dois regimes é na Equação 3 a seguir:

$$y_t = \begin{cases} \alpha_0 + \sum_{i=1}^{p} \alpha_i\, y_{t-i} + \varepsilon_t i_f y_{t-d} \leq \tau \\ \beta_0 + \sum_{i=1}^{p} \beta_i\, y_{t-i} + \varepsilon_t i_f y_{t-d} > \tau \end{cases} \tag{3}$$

Onde α_i e β_i são coeficientes autoregressivos, p é a ordem do modelo SETAR, d é o parâmetro de atraso e y_{t-d} é a variável de limite, ε_t é uma série de variáveis aleatórias independentes e identicamente distribuídos com média 0 e variação $\sigma^2\varepsilon$. τ é o valor do limiar e, se o valor de τ for conhecido, as observações podem ser separadas com base em seu valor em comparação com o limiar, ou seja, se y_{t-d} está abaixo ou acima do limiar. Dessa forma, usando o método dos mínimos quadrados ordinários, o modelo de RA é estimado (Ismail & Isa, 2006). O valor limite deve ser determinado juntamente com outros parâmetros do modelo SETAR, pois o valor limite é desconhecido em geral.

2.3 Redes Neurais Artificiais (RNA)

As redes neurais artificiais são uma das estruturas de computação flexíveis que podem ser usadas para modelar uma ampla variedade de dados não lineares. As principais vantagens dos modelos de RNA são que eles são aproximadores orientados a dados e universais, que podem aproximar uma grande classe de funções com grande precisão. Esse processo de criação de modelo não exige nenhuma suposição anterior sobre o formato do modelo, pois as características dos dados determinam o modelo de rede. Uma rede neural *feedforward* com uma única camada oculta é um dos métodos mais amplamente utilizados para prever os dados de séries temporais (Zhang, 2003). A estrutura do modelo é definida por uma rede de três camadas de unidades de processamento simples conectadas por links acíclicos. A representação matemática da relação entre a saída yt e as entradas (y_{t-1}, y_{t-2}, ... y_{t-p}) pode ser definida conforme Equação 4 a seguir:

$$y_t = \mathrm{w}_0 + \sum_{j=1}^{q} w_{j.g}\left(w_{o,j} + \sum_{i=1}^{p} w_{ij} \cdot y_{t-1}\right) + \varepsilon_t \tag{4}$$

Onde w_j (j = 0, 1, 2,…, q) e w_{ij} (i = 0, 1, 2,… .., p; j = 0, 1, 2,…, q) são os pesos de conexão ou o parâmetros de modelo, p é o número de nós de entrada e q é o número de nós ocultos. A função de transferência da camada oculta é dada pela função logística apresentada na Equação 5 a seguir:

$$Sig\,(x) = \frac{1}{1+\exp(-x)} \tag{5}$$

Portanto, o modelo de RNA na equação 4 executa o mapeamento funcional não linear de observações passadas (y_{t-1}, y_{t-2}, ... y_{t-p}) para o valor futuro y_t, isto é,

$$y_t = f\left(y_{t-1}, \ldots \ldots y_{t-p,} w\right) + \varepsilon_t \tag{6}$$

Onde f é uma função determinada pela estrutura da rede e pelos pesos de conexão, e w é um vetor de todos os parâmetros. Assim, esse modelo de rede neural é semelhante a um modelo autorregressivo com funcionalidade não linear. A escolha do valor de q depende dos dados, pois não existe um *procudere* padrão para determinar esse parâmetro específico. Outra tarefa vital da modelagem de RNA é a escolha da dimensão do vetor de entrada e o número de observações atrasadas, p. Este talvez seja o parâmetro mais crucial a ser estimado em um modelo de rede neural artificial, pois a determinação da estrutura de autocorrelação não linear da série temporal depende desse parâmetro. No entanto, não existe uma regra geral que possa ser seguida para selecionar o valor de p. Portanto, frequentemente são realizados ensaios para selecionar um valor ideal de p e q. Após especificar a estrutura da rede com os parâmetros p, e q, está pronto para treinamento. Isso é feito com algoritmos de otimização não linear eficientes, como algoritmos de descida de gradiente e algoritmos de gradiente conjugado, além do algoritmo básico de treinamento de retropropagação (Hung, 1993). Nas RNAs, as funções de ativação mais usadas são as funções sigmóides. Recentemente, no aprendizado profundo, várias outras funções foram sugeridas como alternativas à função sigmóide, como a função tangente hiperbólica (*tanh*), unidades lineares retificadas (*ReLU*), softmax e gaussiana. Essas funções são fornecidas abaixo. A função tangente hiperbólica (*tanh*) é uma das alternativas à função sigmóide. Pode ser definida como:

$$tanh\,(x) = \frac{1-e^{-2x}}{1+ e^{-2x}} \tag{7}$$

Essa função é semelhante à função sigmóide, no entanto, compacta o número de valor real para um intervalo entre - 1 e 1; ou seja, tanh (x) ∈ (−1, 1). Unidades lineares retificadas (ReLU) são definidas como:

$$f\,(x) = \max(0, x) \tag{8}$$

Onde x é a entrada para um neurônio. Em outras palavras, a ativação é simplesmente configurada no limite de zero. O intervalo da ReLU está entre 0 e ∞. A função *softmax*, também chamada de função exponencial normalizada, é uma generalização da função logística que 'comprime' um vetor K-dimensional *Z* de valores reais aleatórios para um vetor K-dimensional σ (z) de valores reais no intervalo [0, 1], que somam 1. A função é definida como:

$$\sigma\,(\mathrm{z})_j = \frac{e^{z\,j}}{\sum_{k-1}^{k} e^{zk}}, j = 1,2, \ldots \ldots, k \tag{9}$$

As funções de ativação gaussiana são curvas em forma de sino que são contínuas. A saída do nó é interpretada dependendo da proximidade da entrada líquida de um valor médio escolhido, ou seja, é interpretada em termos de associação à classe (1 ou 0). A função é definida como:

$$f\,(x) = \frac{1}{\sqrt{2\pi\,\sigma}}\, e^{\frac{-(x-\mu)^2}{2\sigma 2}} \tag{10}$$

2.4 Análise de Espectro Singular (SSA)

Alguns estudos empregaram o método SSA para prever séries temporais financeiras (Hassani, Soofi & Zhiglavsky, 2013b; Ghodsi & Omer, 2014). O método SSA compreende dois estágios, um é decomposição e o outro é reconstrução. No primeiro estágio, a série temporal é decomposta para separar o sinal e o ruído; depois, no segundo estágio, a série com menos ruído é reconstruída e aplicada à previsão usando as seguintes algumas etapas (Hassani, 2007).

A primeira das etapas é a Incorporação. A incorporação pode ser considerada um mapeamento que transfere uma série temporal unidimensional $Y_N = (y_1,\ldots, y_N)$ para uma série multidimensional $X_1,\ldots, X_K$ com vetores $X_i = (y_i,\ldots, y_i + L - 1)^T \in R^L$, onde L ($2 \leq L \leq N - 1$) é o comprimento da janela e $K = N - L + 1$. O resultado dessa etapa é a matriz de trajetória.

$$X = [X_1, \ldots \ldots, X_K] = (X_{ij})_{i,j=1}^{L,K}$$

Na etapa 2, temos a Decomposição do Valor Singular (SVD). Nesta etapa, o SVD de X é implementado. Denotado por $\lambda_1 \ldots .$, λ_L os autovalores de XX^T dispostos em ordem decrescente ($\lambda_1,, \geq \ldots \geq \lambda_L \geq 0$) e por $U_1 \ldots .$,, U_L os vetores próprios correspondentes. O SVD de X pode ser escrito como $X = X_1 + \ldots + X_L$, onde, $X_i = \sqrt{\lambda_i} U_i V_i^T$.

$\tilde{y}_i$ para $i = 1, \ldots .., N$

$$y_t = \left\{ \sum_{j=1}^{L-1} \alpha_j \, y_{i-j}, para \; i = N+1, \ldots . N + h' \right\} \tag{11}$$

Onde $\tilde{y}_i$ ($i = 1; \ldots; N$) é a série reconstruída, e o vetor $A = (\alpha_1,\ldots, \alpha_{L-1})$ pode ser calculado por:

$$A = \frac{1}{1-v^2} \sum_{i=1}^{r} \pi_i u_i^{\nabla} \tag{12}$$

2.5 Modelo Híbrido (HM)

Modelos puramente lineares ou puramente não lineares podem não ser adequados para prever retornos de ações, uma vez que os retornos são de natureza complexa. Até RNAs baseadas em dados produziram resultados mistos na previsão dos dados de séries temporais. Por exemplo, Denton (1995) usou dados simulados e descobriu que, quando há multicolinearidade ou outliers nos dados, as redes neurais podem prever os dados melhor do que os modelos de regressão linear. O tamanho da amostra e o nível de ruído desempenham um papel crucial na determinação do desempenho das RNAs para problemas de regressão linear (Markham & Rakes, 1998). Portanto, pode não ser útil aplicar RNAs para todos os tipos de dados.

Dadas as complexidades nos dados do mercado de ações, um método que pode lidar com os dados lineares e não lineares, ou seja, o modelo híbrido pode ser uma alternativa para a previsão. Aspectos lineares e não lineares dos padrões subjacentes nos dados podem ser capturados através da combinação de diferentes modelos. Pode ser útil considerar dados de séries temporais que consistem em estrutura de autocorrelação linear e um componente não linear. Isso é,

$$y_t = L_t + N_t \tag{13}$$

Onde L_t representa o componente linear e N_t denota o componente não linear. Inicialmente, devemos aplicar um modelo linear para os dados e, em seguida, os resíduos do modelo linear conteria apenas a relação não linear. Esses resíduos podem ser definidos como:

$$e_t = y_t - y_t + \bar{Ḷ}_t \tag{14}$$

Onde $\bar{Ḷ}_t$ é o valor previsto no tempo *t* da relação estimada da equação13. Os resíduos são muito cruciais no diagnóstico da adequação dos modelos lineares, porque a presença de correlação linear nos resíduos indica a inadequação do modelo linear. Além disso, qualquer padrão não linear significativo nos resíduos também indica a limitação no modelo linear. Relações não lineares podem ser descobertas modelando resíduos usando RNAs. O modelo de RNA para resíduos com *n* nós de entrada será:

$$e_t = f(e_{t-1,}. e_{t-2,}. \ldots.. e_{t-n}) + \varepsilon_{\mathrm{t}} \tag{15}$$

Onde f é uma função não linear determinada pela rede neural e ε_t é o erro aleatório. Denotando a previsão da equação 13como $\dot{N}_t$, a previsão combinada será:

$$e_t = f(e_{t-1,}. e_{t-2,}. \ldots.. e_{t-n} \tag{16}$$

Onde $\hat{Y}_t$ é o valor estimado do modelo híbrido, que é uma combinação de modelos lineares e não lineares. Utilizamos a razão erro inverso da média quadrada da previsão (MSFE) para determinar os pesos ótimos para os modelos híbridos, pois é um método amplamente utilizado com uma base teórica robusta (Bates & Granger, 1969). Para os modelos *M*, a previsão combinada do passo *h* à frente é:

$$\hat{Y}_{t+h} = \sum_{m=1}^{M} w_{m,h,t} \hat{Y}_{t+h,m'} \tag{17}$$

$$w_{m,h,t} = \frac{\left(\frac{1}{msfe_{m,h,y}}\right)^k}{\sum_{j-1}^{m}\left(\frac{1}{msfe_{m,h,t}}\right)^{k'}} \tag{18}$$

Onde $\hat{Y}_{t+h,m}$ é o ponto previsto para h passos à frente no tempo t do modelo m. Em resumo, esse método híbrido contém duas etapas. O primeiro passo é empregar o ARIMA para modelar a parte linear dos dados. O segundo passo é aplicar a RNA para modelar os resíduos obtidos do ARIMA, esses resíduos têm informações sobre a não linearidade nos dados. Os resultados do modelo ANN podem ser usados como previsões para os termos de erro do modelo ARIMA. Dessa forma, o modelo híbrido abrange as características dos modelos ARIMA e ANN na modelagem de dados de séries temporais. Assim, pode ser benéfico empregar modelos híbridos para melhorar a precisão das previsões.

2.6 Medidas de desempenho de previsão

A precisão das previsões indica quão bem um modelo de previsão prevê a variável escolhida. Diferentes medidas de precisão são usadas para validar a adequação de um modelo para um determinado conjunto de dados. Existem várias medidas de precisão na literatura, como erro médio (EM), erro absoluto médio (MAE), erro percentual absoluto médio (MAPE), erro quadrático médio (MSE) e erro quadrático médio da raiz (RMSE). Neste estudo, usamos o RMSE por ser um dos métodos mais adequados para medir a precisão da previsão de dados na mesma escala, e esse critério tem sido empregado em vários estudos anteriores (Lu & Wu, 2011; Wang et al., 2011;

Hyndman & Athanasopoulos, 2015; Makridakis et al., 2015). Chai e Draxler (2014) também sugeriram que o RMSE é uma medida adequada para modelos com erros distribuídos normalmente. O presente estudo constatou que os erros na maioria dos modelos seguem a distribuição normal. Se Y_t for a observação real para o período t e F_t for a previsão para o mesmo período, o erro será definido como:

$$e_t = y_t - F_t, \tag{19}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|, \tag{20}$$

$$MPE = \frac{1}{n}\sum_{t=1}^{n} PE_t, \tag{21}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|PE_t|, \tag{22}$$

Onde,

$$PE_t = \left(\frac{y_t - F_t}{y_t}\right) * 100 \tag{23}$$

O erro quadrado médio (MSE) será dado por:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t{}^2, \tag{24}$$

O erro quadrático médio da raiz (RMSE) será dado por:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t{}^2} \tag{25}$$

3. RESULTADOS EMPÍRICOS

Os resultados empíricos, incluindo estatísticas descritivas e medidas de desempenho de vários métodos de previsão para retorno de ações em mercados desenvolvidos, emergentes e de fronteira serão apresentados nesta seção.

3.1 Estatísticas descritivas dos retornos das ações

As Tabelas 1, 2 e 3 apresentam as estatísticas resumidas (por exemplo, média, desvio padrão, assimetria, curtose, estatística JB) para retornos dos mercados de ações desenvolvidos, emergentes e de fronteira, respectivamente. A partir dessas tabelas, podemos ver que os retornos médios em todos os mercados são positivos, indicando retornos positivos gerais sobre os investimentos durante o período considerado para este estudo.

Os valores de curtose das séries de retorno de todos os mercados são superiores a 3, indicando que todas as séries são leptocúrticas - ou seja, possuem caudas grossas, o que é um fenômeno comum nos retornos das ações (Bouchauda & Potters, 2001; Humala, 2013; Mallikarjuna, Guptha & Rao, 2018). O teste de Jarque-Bera mostrou que as séries não são normalmente distribuídas. Outra característica importante, do teste de Tsay (1989), é que os retornos de todos os mercados são não-lineares.

Tabela 1. Estatística Descritiva para Mercados Desenvolvidos

País	Média	Desvio Padrão	Skewness	Kurtosis	Teste Jarque–Bera	Teste Tsay
Austrália	0,016900	0,981318	-0,366178	8,451780	5.732,678 (0,00000)	Não-linear
Canada	0,016719	1,043102	-0,466783	13.58932	21.277,95 (0,00000)	Não-linear
França	0,005790	1,429676	0,141474	8,712746	6.275,928 (0,00000)	Não-linear
Alemanha	0,021703	1,475373	0,097060	8,111260	4.987,257 (0,00000)	Não-linear
Japão	0,019562	1,501151	-0,211389	9,413287	7.642,163 (0,00000)	Não-linear
Coreia do Sul	0,040316	1,384506	-0,350392	9,724636	8.473,831 (0,00000)	Não-linear
Suíça	0,007731	1,174780	-0,012284	10,18851	9.766,624 (0,00000)	Não-linear
Reino Unido	0,009243	1,171826	-0,004334	9,918071	9.061,453 (0,00000)	Não-linear
Estados Unidos	0,020694	1,147998	-0,089045	12,10219	15.630,08 (0,00000)	Não-linear

Fonte: Estatísticas da Pesquisa (2020).

3.2 Resultados dos métodos de previsão

Antes de aplicar os métodos de previsão, dividimos os dados no conjunto de treinamento e no conjunto de testes; usamos 80% dos dados para treinar os modelos e os 20% restantes para testar os modelos. Para prever os retornos usando o modelo ARIMA (p, d, q), foi necessário verificar a estacionariedade para obter inferências válidas.

Para testar a estacionariedade das séries de retornos, empregamos os testes aumentados de Dickey-Fuller (1979) e Phillips-Perron (1988); os resultados mostraram que os retornos de todos os mercados eram estacionários.

Tabela 2. Estatísticas Descritivas para Mercados Emergentes

País	Média	Desvio Padrão	Skewness	Kurtosis	Teste Jarque–Bera	Teste Tsay
África do Sul	0,047775	1,298747	0,036830	6,299922	2.044,612 (0,00000)	Não-linear
Brasil	0,055547	1,769091	0,075096	7,347967	3.439,261 (0,00000)	Não-linear
China	0,016499	1,585054	-0,218775	7,684998	4.025,909 (0,00000)	Não-linear
Egito	0,079649	1,658833	-0,121416	13,15053	18.891,65 (0,00000)	Não-linear
Índia	0,059777	1,414324	0,120611	12,83913	17.795,20 (0,00000)	Não-linear
Indonésia	0,070676	1,326156	-0,501572	9,570707	8.083,189 (0,00000)	Não-linear
México	0,051820	1,213204	0,153714	9,356521	7.644,352 (0,00000)	Não-linear
Rússia	0,082735	1,965327	0,370012	24,33804	85.492,66 (0,00000)	Não-linear
Tailândia	0,047775	1,263998	-0,520070	13,23751	2.044,612 (0,00000)	Não-linear
Turquia	0,069655	1,973356	-0,037463	9,689365	8.426,657 (0,00000)	Não-linear

Fonte: Estatísticas da Pesquisa (2020).

Determinamos o comprimento ótimo do atraso para a média autoregressiva (*p*) e a média móvel (*q*) componentes que utilizam o Critério de Informação de *Akaike* (AIC). Observamos diferentes ordens de AR e MA para séries diferentes e apresentá-las juntamente com os valores RMSE nas Tabelas 4, 5 e 6. No modelo SETAR, a série exibiu tendências não lineares, e identificamos dois regimes pelos valores mínimos de AIC. Em seguida, o modelo foi usado para prever os retornos dos mercados. Para prever retornos de ações usando o modelo ANN, empregamos redes neurais de *feedforward*, pois muitos estudos demonstraram que eles se encaixam bem com dados de retorno de ativos (Zhang, 2003; Qiu, Song & Akagi, 2016).

Empregamos um modelo recorrente de análise de espectro único (RSSA) para prever os retornos após decompor e reconstruir a série de retornos originais, seguindo as quatro etapas envolvidas na previsão com SSA: incorporação, reconstrução, agrupamento e média diagonal. Para o modelo híbrido, que é uma combinação de ARIMA e ANN, ajustamos o modelo empregando a razão de erro médio de previsão quadrada inversa (MSFE) amplamente utilizada (Bates & Granger, 1969; Winkler & Makridakis, 1983) para atribuir melhor as ponderações para os modelos na previsão.

Tabela 3. Estatísticas Descritivas para Mercados de Fronteira

País	Média	Desvio Padrão	Skewness	Kurtosis	Teste Jarque–Bera	Teste Tsay
Argentina	0,120985	2,162861	0,018419	7,202958	3.230,70 (0,00000)	Não-linear
Estônia	0,052646	1,036820	0,316893	14,29612	24.406,12 (0,00000)	Não-linear
Quênia	0,012726	0,831383	0,555709	15,16038	28.057,55 (0,00000)	Não-linear
Sri Lanka	0,066494	1,128874	0,993308	43,75095	300.319,20 (0,00000)	Não-linear
Tunísia	0,035152	1,292467	0,603647	20,46153	59.408,84 (0,00000)	Não-linear

Fonte: Estatísticas da Pesquisa (2020).

Nas Tabelas 4, 5 e 6, podemos observar que nenhum método isolado teve desempenho uniforme em todos os mercados. No entanto, o modelo não linear (ou seja, SETAR) teve um desempenho melhor que os outros modelos, produzindo previsões ideais para 10 mercados (ou seja, quatro mercados desenvolvidos, quatro emergentes e dois de fronteira). Esse resultado contrasta com Guidolin et al. (2009).

No caso de mercados desenvolvidos, o modelo SETAR produziu previsões ótimas para quatro dos nove mercados (Austrália, França, Japão e Suíça). As Tabelas 4, 5 e 6 apresentam os valores RMSE dos conjuntos de testes da série de previsões para todas as técnicas (ou seja, ARIMA, SETAR, SSA, ANN e HM) para mercados desenvolvidos, emergentes e de fronteira, respectivamente.

O modelo com o menor RMSE foi escolhido como o modelo mais apropriado. Além disso, testamos a significância do RMSE usando o teste de Diebold-Marino (1995) e constatou que o RMSE de todos os modelos era significativo, exceto para o Japão, África do Sul e Sri Lanka.

Tabela 4. Valores RMSE dos modelos de previsão para mercados desenvolvidos

País	ARIMA	SETAR	ANN	SSA	HM
Alemanha	ARIMA (3,0,3) 1,137963	1,140388	1,156347	1,139859	1,138008
Austrália	ARIMA (1,0,0) 0,839708	0,8371075	0,839961	0,853739	0,8398178
Canadá	ARIMA (4,0,4) 0,7160948	0,7235669	0,7279448	0,721191	0,7178712
Coreia do Sul	ARIMA (1,0,2) 1,312612	0,7845301	0,7816591	0,7975102	0,7815836
Estados Unidos	ARIMA (2,0,0) 0,8197795	0,8125207	0,8171281	0,8187795	0,8119642
França	ARIMA (2,0,3) 1,104531	1,1044509	1,121098	1,133334	1,107893
Japão	ARIMA (1,0,1) 1,312612	1,307090	1,312699	1,319925	1,312697
Reino Unido	ARIMA (3,0,2) 0,9002485	0,9003106	0,9093658	0,9194308	0,9057976
Suíça	ARIMA (3,0,3) 0,9262604	0,9252304	0,9262753	0,935211	0,9262085

Fonte: Autor (2020).

O modelo ARIMA foi ideal para o Canadá, Alemanha e Reino Unido, e o modelo HM foi ideal para a Coréia do Sul e os EUA. Assim, podemos dizer que modelos não lineares são mais adequados para mercados desenvolvidos. Enquanto isso, os modelos ANN e SSA não são de todo úteis para mercados desenvolvidos, uma vez que não forneceram previsões ideais.

Tabela 5. Valores RMSE dos modelos de previsão para mercados emergentes

País	ARIMA	SETAR	ANN	SSA	HM
África do Sul	ARIMA (3,0,1) 1,062336	1,064449	1,063028	1,066819	1,061791
Brasil	ARIMA (2,0,1) 1,441972	1,442167	1,442900	1,470025	1,442262
China	ARIMA (3,0,3) 1,555623	1,570942	1,56378	1,555431	1,554521
Egito	ARIMA (0,0,1) 1,357198	1,324729	1,368285	1,385264	1,357206
Índia	ARIMA (3,0,1) 0,838730	0,844534	0,839839	0,851503	0,838335
Indonésia	ARIMA (1,0,0) 0,919355	0,926838	0,918685	0,921310	0,918988
México	ARIMA (2,0,1) 0,850496	0,848814	0,867506	0,851756	0,848878
Rússia	ARIMA (3,0,4) 0,996021	0,993842	1,008659	1,006565	1,000470
Tailândia	ARIMA (2,0,2) 0,756856	0,754199	0,770766	0,775997	0,762275
Turquia	ARIMA (2,0,2) 1,280623	1,286060	1,292737	1,291581	1,283447

Fonte: Autor (2020).

Para mercados emergentes, o modelo SETAR foi considerado adequado para quatro mercados (Egito, México, Rússia, Tailândia). Os modelos HM foram apropriados para três mercados (China, Índia e África do Sul) e os modelos ARIMA para dois (Brasil e Turquia). O modelo de RNA era apropriado para apenas um mercado (Indonésia), enquanto o modelo de SSA não era adequado para nenhum mercado emergente. Embora nenhum modelo único fosse adequado para todos os mercados emergentes, os modelos SETAR e HM eram relativamente mais úteis. Em relação aos mercados fronteiriços, o SETAR foi adequado para Argentina e Quênia, ARIMA para Estônia e Sri Lanka e SSA para Tunísia. Os modelos ANN e HM não eram apropriados para nenhum mercado.

Tabela 6. Valores RMSE dos modelos de previsão para mercados de fronteira

País	ARIMA	SETAR	ANN	SSA	HM
Argentina	ARIMA (1,0,0) 2,058513	2,046473	2,059416	2,067785	2,061399
Estônia	ARIMA (1,0,2) 0,554262	0,576389	0,593064	0,567850	0,574015
Quênia	ARIMA (2,0,2) 0,679084	0,656696	0,656707	0,680819	0,684070
Sri Lanka	ARIMA (0,0,2) 0,443346	0,449314	0,444426	0,473222	0,445437
Tunísia	ARIMA (1,0,0) 0,460538	0,44934	0,618865	0,443756	0,464829

Fonte: Autor (2020).

Dos vinte e quatro índices do mercado de ações, o modelo SETAR produziu previsões ideais para dez, ARIMA para sete, modelos HM para cinco e modelos ANN e SSA para um mercado cada. A partir desses resultados, podemos observar que modelos não lineares são mais úteis para mercados desenvolvidos, emergentes e de fronteira. Outra observação interessante é que os modelos de IA e domínio de frequência foram considerados adequados apenas para um mercado cada. Assim, podemos dizer que, mesmo com os avanços nos modelos de IA e no domínio da frequência, os modelos estatísticos tradicionais não se tornaram obsoletos; eles ainda são úteis e, de fato, são melhores que os modelos de IA e de domínio de frequência para prever dados financeiros de séries temporais.

4. CONSIDERAÇÕES FINAIS

Ao longo dos anos, as bolsas de valores tornaram-se caminhos alternativos para excedentes de fundos entre investidores individuais e institucionais, principalmente após a globalização e a integração dos mercados financeiros mundiais. Dado o risco inerente, a incerteza e a natureza dinâmica dos mercados de ações, prever com precisão o retorno das ações pode ajudar a minimizar os riscos dos investidores. Assim, as técnicas de previsão podem ajudar na melhor tomada de decisão de investimento, uma vez que prever retorno de ações não é um golpe de sorte, mas sim a aplicação de combinadas técnicas de previsão, que se aperfeiçoam com grande dinamismo.

Este estudo considerou dados diários dos retornos do mercado de ações durante o período de 1º de janeiro de 2000 a 31 de dezembro de 2019 para comparar técnicas de previsão (ou seja, modelos ARIMA, SETAR, ANN, SSA e HM) representando lineares, não lineares, IA, domínio de frequência e híbridos métodos. Foram coletados os índices de ações de 24 bolsas de valores em três categorias de mercado (nove desenvolvidas, dez emergentes e cinco de fronteira) para encontrar técnicas de previsão adequadas para cada categoria.

Os resultados mostraram que nenhuma técnica de previsão única forneceu uma previsão ideal para todos os mercados. No entanto, o SETAR teve melhor desempenho em dez mercados, ARIMA em sete, HM em cinco e ANN e SSA em um mercado cada. As técnicas SETAR e ARIMA podem, portanto, ser consideradas as vencedoras na previsão de retornos do mercado de ações para mercados desenvolvidos, emergentes e de fronteira, pois esses dois métodos forneceram previsões ideais para dezessete dos vinte e quatro mercados.

Para pesquisas futuras, recomenda-se o uso das técnicas de previsão estudadas neste artigo para captar sua precisão apenas em momentos de crise financeira internacional (p.ex., crise da dívida dos países da América Latina em 1920; bolha imobiliária e de ações no Japão em 1985; crise dos mercados emergentes em 1994 ou crise mundial do suprime em 2008) ou ainda destacando apenas períodos de volatilidade causados por Endemias ou Pandemias, como no caso recente da COVID-19.

REFERÊNCIAS

Adebiyi AA, Oluinka A (2014) Comparision of ARIMA and artificial neural network models for stock market prediction. Journal of Applied Mathematics.

Almudhaf F (2018) Predictability, Price bubbles, and efficiency in the Indonesian stock-market. Bull Indones Econ Stud 54(1): 113–124

Al-Shaib M (2006) The predictability of the Amman stock exchange using Univariate autoregressive integrated moving average (ARIMA) model. Journal of Economic and Administrative Sciences 22(2):17–35

Aras S, Kocakoc ID (2016) A new model selection strategy in time series forecasting with artificial neural networks. IHTS Neurocomputing 174:974–987

Asadi S, Tavakoli A, Hejazi SR (2010) A new hybrid for improvement of auto-regressive integrated moving average models applying particle swarm optimization. Expert Syst Appl 39:5332–5337

Awajan AM, Ismail MT, Wadi SA (2018) Improving forecasting accuracy for stock market data using EMD-HW bagging. PLoS One 13(7):1–20

Bates JM, Granger CWJ (1969) The combination of forecasts. Operational Research Society 20(4):451–468

Beck T, Levine R (2003) Stock markets, banks and growth: panel evidence. J Bank Financ 28:423–442

Boero G (2003) The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts. Int J Forecast 20:305–320

Boero G, Marrocu E (2002) The performance of non-linear exchange rate models: a forecasting comparison. J Forecast 21(7): 513–542

Bouchauda JP, Potters M (2001) More stylized facts of financial markets: leverage effect and downside correlations. Physica A 299:60–70

Clements MP, Smith J (1999) A Monte Carlo study of the forecasting performance of empirical SETAR models. J Appl Econ 14:124–141

Cristelli M (2014) Complexity in financial markets. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319- 00723-6

Darrat AF, Zhong M (2000) On testing the random walk hypothesis a model Comparision approach. The Financial Review 35: 105–124

Denton JW (1995) How good are neural networks for causal forecasting? The Journal of Business Forecasting Methods and Systems 14(2):17–23

Dickey D, Fuller W (1979) Distribution of the estimators for autoregressive time series with a unit root. Journal of American Statistical Association 74(366):427–431

Diebold FX, Marino RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13(3):134–144

Erdem E, Ulucak R (2016) Efficiency of stock exchange markets in G7 countries: bootstrap causality approach. Economics World 4(1):17–24

Fama EF (1970) Efficient capital markets:a review of theory and empirical work. J Financ 25(2):383–417

Firat EH (2017) SETAR (self-exciting threshold autoregressive) non-linear currency Modelling in EUR/USD, EUR/TRY and USD/ TRY parities. Mathematics and Statistics 5(1):33–55

Ghiassi M, Saidane H, Zimbra DK (2005) A dynamic artificial neural network model for forecasting series events. Int J Forecast 21:341–362

Ghodsi Z, Omer HN (2014) Forecasting energy data using singular Spectrum analysis in the presence of outlier(s). International Journal of Energy and Statistics 2(2):125–136

Golyandina N, Nekrutkin V, Zhigljavsky A (2001) Analysis of time series structure SSA and related techniques. Chapman and Hall/CRC, Newyork

Gooijer DJ (1998) On threshold moving-average models. J Time Ser Anal 19(1):1–18

Guptha SK, Rao RP (2018) The causal relationship between financial development and economic growth experience with BRICS economies. Journal of Social and Economic Development 20(2):308–326

Guresen E, Kayakutlu G, Daim TU (2011) Using artificial neural network models in stock market index prediction. Expert Syst Appl 38:10389–10397

Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384

Harrison B, Moore M (2012) Stock market efficiency, non-linearity, thin trading and asymmetric information in MENA stock markets. Economic Issues 17(1):77–93

Hassani H (2007) Singular spectrum analysis: methodology and comparison. Journal of Data Science 5(2):239–257

Hassani H, Soofi A, Zhiglavsky A (2013a) Forecasting UK industrial production with multivariate singular Spectrum analysis. J Forecast 32(5):395–408

Hassani H, Soofi A, Zhiglavsky A (2013b) Predicting inflation dynamics with singular Spectrum analysis. J R Stat Soc 176(3): 743–760

Humala A (2013) Some stylized facts of return in the foreign exchange and stock markets in Peru. Stud Econ Financ 30(2):139–158

Hung SL, Adeli H (1993) Parallel backpropagation algorithms on CRAY Y-MP8/864 supercomputer. Neurocomputing 5(6):287–302

Hyndman R, Athanasopoulos G (2015) Forecasting principles and practice. Otexts, Melbourne. Available at: https://otexts.com/ fpp3/. Accessed 20 Mar 2019.

Ince H, Trafalis TB (2017) A hybrid forecasting model for stock market prediction. Economic Computation and Economic Cybernetics Studies and Research 21:263–280

Ismail MT, Isa Z (2006) Modelling exchange rate using regime switching models. Sains Malaysiana 35(2):55–62

Johnson NF, Jefferies P, Hui PM (2003) Financial market complexity. Oxford University Press, Oxford

Khandelwal I, Adhikari R (2015) Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. Procedia Computer Science 48:173–179

Khashei M, Bijari M (2010) An artificial neural network model for time series forecasting. Expert Syst Appl 37:479–489

Khashei M, Bijari M (2012) A new class of hybrid models for time series forecasting. Expert Syst Appl 39:4344–4357

Khashei M, Hajirahimi Z (2017) Performance evaluation of series and parallel strategies for financial time series forecasting. Financial Innovation 3(24):1–24

Konak F, Seker Y (2014) The efficiency of developed markets: empirical evidence from FTSE 100. J Adv Manag Sci 2(1):29–32

Lahmiri S (2016) A variational mode decomposition approach for analysis and forecasting of economic and financial time series. Expert Syst Appl 55:268–273

Levine R (1997) Financial development and economic growth: views and agenda. J Econ Lit 35:688–726 Levy RA (1967) The theory of random walks: a study of findings. Am Econ 11(2):34–48

Lo AW, Mackinlay AC (2002) An non-random walk down Wall street. Princeton University Press, Princeton

Lu CJ, Wu JY (2011) An efficient CMAC neural network for stock index forecasting. Expert Syst Appl 38:15194–15201

Makridakis S, Wheelwright SC, Hyndman RJ (2015) Forecasting: methods and applications. Wiley India, New Delhi

Mallikarjuna M, Arti G, Rao RP (2018) Forecasting stock returns of selected sectors of Indian capital market. SS International Journal of Economics and Management 8(6):111–126

Markham IS, Rakes TR (1998) The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. Comput Oper Res 25:251–263

Mondal P, Shit L, Goswami S (2014) Study of effectiveness of time series Modelling (ARIMA) in forecasting stock prices. International Journal of Computer Science, Engineering and Applications 4(2):13–29

Mostafa MM (2010) Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait. Expert Syst Appl 37:6302–6309

MSCI (2018) MSCI Announces the Results of Its Annual Market Classification Review. Available at: https://www.msci.com/ market-classification. Accessed 25 Mar 2019

Nayak SC, Misra BB (2018) Estimating stock closing indices using a GA-weighted condensed polynomial neural network. Financial Innovation 4(21):1–22

Ojo JF, Olatayo TO (2009) ON the estimation and performance of subset of autoregressive integrated moving average models. Eur J Sci Res 28:287–293

Phillips PCB, Perron P (1988) Testing for unit roots in time series regression. Biometrika 75:335–346

Qiu M, Song Y, Akagi F (2016) Application of artificial neural network for the prediction of stock market returns the case of the Japanese stock market. Chaos, Solitons and Fractals 85:1–7

Radikoko I (2014) Testing weak-form market efficiency on the TSX. J Appl Bus Res 30(3):647–658 Rajan R, Zingales L (1998) Financial dependence and growth. Am Econ Rev 88:559–586

Rousseau PL, Watchel P (2000) Equity markets and growth: cross-country evidence on timing and outcomes, 1980-1995. J Bank Financ 24(12):1933–1957

Said A (2015) The efficiency of the Russian stock market: a revisit of the random walk hypothesis. Academy of Accounting and Financial Studies Journal 19(1):42–48

Tong T, Li B, Benkato O (2014) Revisiting the weak form efficiency of the Australian stock market. Corp Ownersh Control 11(2):21–28

Tsay R (1989) Testing and modeling threshold autoregressive processes. Journal of American Statistical Association 84:231–240

Wang JZ, Wang JJ, Zhang ZG, Guo SP (2011) Forecasting stock indices with backpropagation neural network. Expert Syst Appl 38:14346–14355

Watier L, Richardson S (1995) Modelling of an epidemiological time series by a threshold autoregressive model. Journal of Royal Statistical Society 44(3):353–364

Wieland OL (2015) Modern financial markets and the complexity of financial innovation. Universal Journal of Accounting and Finance 3(3):117–125

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175

Zhong X, Enke D (2019) Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial Innovation 5(4):1–20