



Forecasting commodity prices: empirical evidence using deep learning tools

Hachmi Ben Ameer¹ · Sahbi Boubaker² · Zied Ftiti³ · Wael Louhichi⁴ · Kais Tissaoui^{5,6}

Accepted: 8 November 2022 / Published online: 20 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Since the last two decades, financial markets have exhibited several transformations owing to recurring crises episodes that has led to the development of alternative assets. Particularly, the commodity market has attracted attention from investors and hedgers. However, the operational research stream has also developed substantially based on the growth of the artificial intelligence field, which includes machine learning and deep learning. The choice of algorithms in both machine learning and deep learning is case-sensitive. Hence, AI practitioners should first attempt solutions related to machine learning algorithms, and if such solutions are unsatisfactory, they must apply deep learning algorithms. Using this perspective, this study aims to investigate the potential of various deep learning basic algorithms for forecasting selected commodity prices. Formally, we use the Bloomberg Commodity Index (noted by the Global Aggregate Index) and its five component indices: Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex. Based on daily data from January 2002 (the beginning wave of commodity markets' financialization) to December 2020, results show the effectiveness of the Long Short-Term Memory method as a forecasting tool and the superiority of the Bloomberg Livestock Subindex and Bloomberg Industrial Metals Subindex for assessing other commodities' indices. These findings is important in term for investors in term of risk management as well as policymakers in adjusting public policy, especially during Russian-Ukrainian war.

✉ Hachmi Ben Ameer
hechba@yahoo.fr

¹ INSEEC Grande Ecole, Omnes Education, Paris, France

² Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 21959, Saudi Arabia

³ OCRE-Lab, EDC Paris Business School, Paris, France

⁴ ESSCA School of Management, Paris, France

⁵ University of Ha'il, Applied College, Hail City PO Box 2440, Saudi Arabia

⁶ Faculty of Economic Sciences and Management of Tunis, University of Tunis El Manar, the International Finance Group, Tunis, Tunisia

Keywords Commodity markets · Forecasting · Deep learning · Bloomberg Commodity Index · Performance metrics

1 Introduction

Since commodity markets were financialized in 2002, commodities' market capitalization has been growing substantially. There are several reasons why these commodity assets have attracted policymakers as well as investors and hedgers. Bernanke¹ (2008) claimed that policymakers should underscore “*the importance for policy of both forecasting commodity price changes and understanding the factors that drive those changes*” and highlighted the role of commodity prices in influencing inflation and macroeconomic environments. Similarly, the Editorial of the September 2008 issue of the Monthly Bulletin of the European Central Bank also stated, “[...] *Rapidly rising prices for globally traded commodities have been the major source of the relatively high rates of inflation we have experienced in recent years, underscoring the importance for policy of both forecasting commodity price changes and understanding the factors that drive those changes.*” Moreover, commodity asset management requires investors to play an important role regarding portfolio management (Bodie & Rosansky, 1980; Gospodinov & Ng, 2013; Lintner, 1983; Marquis & Cunningham, 1990). Recent empirical studies have highlighted the important role of commodity assets in portfolio diversification and asset allocation (Ftiti et al., 2016; Kablan et al., 2017; Klein, 2017). Studies also support the ability of commodity assets to serve as safe-havens and hedges (Madani & Ftiti, 2022).

Therefore, forecasting the dynamic of commodity markets is crucial for investors, hedgers, and policymakers. During the last two decades, commodity markets have experienced various instances of high volatility where its fundamental driving force is related to commodities' demand and supply. Regarding commodity assets, the spot prices traduce the current and demand conditions as well as their future expectations, as these assets are storable. From a macroeconomic point of view, the dynamic of commodity markets is important for developing economies, as they are often strongly dependent on commodity exportation (Ftiti et al., 2016; Kablan et al., 2017). Additionally, it affects various investment channels of developed economies (Nguyen & Walther, 2020).

Literature regarding commodity price forecasting and predictability have evolved considerably in recent years. The first generation of literature considered commodity prices to be predictable based on macroeconomic factors including inflation and industrial production (Karali & Power, 2013), macroeconomic news (Smales, 2017), economic uncertainty (Fang et al., 2019), and financial drivers such as the VIX measure, the default return spread, Treasury—Eurodollar spread, and bond markets (Asgharian et al., 2015; Prokopczuk et al., 2017). A recent study (Gargano & Timmermann, 2014) used a large set of commodity prices and assessed the out-of-sample predictability of commodity prices based on macroeconomic and financial variables. They found that commodity currencies have a predictive power at short forecast horizons. However, economic growth and the investment–capital ratio wield predictive power at relatively longer horizons.

The second generation of literature is based on the use of traditional and conventional time series models to forecast the commodity prices. Dooley and Lenihan (2005) employed a time

¹ “Outstanding Issues in the Analysis of Inflation,” Speech by Ben S. Bernanke for the Federal Reserve Bank of Boston’s 53rd Annual Economic Conference, Chatham, Massachusetts, June 9, 2008. <https://www.federalreserve.gov/newsevents/speech/bernanke20080609a.htm>.

series ARIMA and lagged-forward price modelling to forecast future lead and zinc prices and showed the effectiveness of the ARIMA model. Some recent studies have challenged the accuracy of predictions based on past events and have supported the use of stochastic models, which are characterized by pre-established and better-defined boundaries of forecasting prices (Ahrens & Sharma, 1997; Berck & Roberts, 1996; Lee et al., 2006; Slade, 1988). More recently, Szarek et al. (2020) proposed a stochastic model with time varying specificity and non-Gaussian distribution. They found that this new class of models can consider commodity markets' time dependency and that its asymmetric distribution can perform better than traditional time series models. Some other studies have also employed a stochastic model based on the Bayesian approach. Kostrzewski and Kostrzewska (2019) make comparisons between the Bayesian approach, considering time-varying parameters, latent volatility, and jump, and the non-Bayesian individual autoregressive models, and three averaging schemes, for spot prices forecast and showed the superiority of the Bayesian stochastic volatility model. The last generation of literature is based on operational research tools including artificial intelligence (AI), machine learning (ML), and deep learning (DL) systems. Panella et al. (2012) investigated the forecasting of prices of crude oil, coal, natural gas, and electricity prices based on a ML approach, that is, neural networks. Using various algorithms based on a mixture of the Gaussian neural network, authors show that the optimal model can be identified using a hierarchical constructive procedure. Narayan et al. (2013) demonstrated the parsimoniousness and self-calibration of their proposed model to forecast future oil future prices based on a regime-switching framework using hidden Markov filtering algorithms. However, studies that have used DL models to forecast commodity prices remain scarce. Kamdem et al. (2020) applied DL models to forecast the commodities' prices during the COVID-19 pandemic crisis and showed that the Long Short-Term Memory (LSTM) model predicts commodity prices accurately. Our study is based on the recent trend of operational research literature and aims to investigate the best tool for commodity prices forecasting.

Formally, AI, which includes computerized tools that mimic human intelligence, is a relatively new concept for solving complex problems. Recently, more complex ML algorithms are being categorized into the DL field. The choice of algorithms from either DL or ML is case-sensitive. However, when faced with a learning problem, AI practitioners should test ML algorithms, and if the solutions are unsatisfactory, they must use DL algorithms. In this backdrop, many authors that have studied commodity markets have indicated that the interaction between the commodities and political, financial, and macroeconomic factors (e.g., political events, supply and demand, financial market, and exchange rates) is the main source of the existence of many complex characteristics in commodities series data such as asymmetry, non-linear dynamics, chaotic pattern, long memory, non-stationary, and heterogeneity (Dong et al., 2018; Karasu & Altan, 2022). Hence, achieving an accurate forecast requires overcoming these different complex features by using the adequate forecasting tools (Karasu & Altan, 2022).

Additionally, Zhang & Ci (2020), Kamdem et al. (2020), Chen et al. (2021), and Karasu & Altan (2022) illustrated that the conventional econometric models and classical ML models may not forecast commodity prices with good accuracy. They showed that classical ML tools like Artificial Neural Networks (ANNs) are facing difficulties dealing with complex dimensionality along with their slow convergence rate. Hence, DL is currently being preferred to cope with the local convergence problem for non-linear optimization and to enhance the performance of basic ANN models. Thus, it can be concluded that complex-structure algorithms can provide better results. Therefore, this study aims to investigate the potential of various DL basic algorithms for forecasting selected commodities' prices. The issue of

commodity price forecasting is no longer as important as it is today based on the Russian-Ukrainian crisis. The forecasting results can be utilized to properly address the optimal portfolio management problem for investors and public policy adjustment in term food safety from policymakers perspective.

We aim to forecast commodity prices based on DL tools. These DL approaches are mainly based on creating more sophisticated hierarchical architectures. The application of DL tools demonstrates superiority in performance over classical ML algorithms. Unlike classical feed-forward artificial neural networks (FFANN) where the information is transferred from the input to the output in one direction (forward), DL has the ability to process the past information and then processes the data in two directions (forward and backward).

This study utilized a set of commodity assets, that is, the Bloomberg Commodity Index (noted by the Global Aggregated Index) and its five component indices: Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex. We show the LSTM method's effectiveness as a forecasting tool and the superiority of the Bloomberg Energy Subindex and Bloomberg Industrial Metals Subindex with regard to other commodity indices.

The remainder of the paper is organized as follows: Sect. 2 presents the methodology; Sect. 3 discusses the results; and Sect. 4 concludes.

2 Methodology

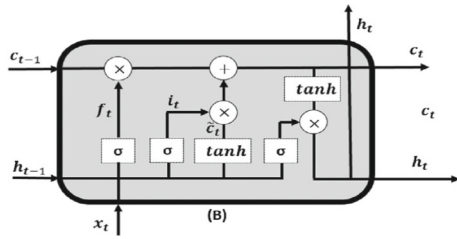
AI is the branch of computer science including computerized tools to mimic and simulate the human intelligence through a machine with the aim of solving complex problems. ML, as a special sub-class of AI where computers can learn from a given set of data and then can generalize to new data (that are not used during the learning phase). Recently, more complex ML algorithms are categorized under the field of DL.

2.1 Overview of deep learning (DL) algorithms

DL is a special case of artificial neural networks (ANNs) structurally including several layers and having the ability to extract the best features from previous data. It has successful applications in various fields of research including image processing and time-series forecasting. Although classical forecasting tools (ARFIMA and FFANN) have been largely utilized in forecasting with significant levels of accuracy, they are drawbacks in using them. For example: data should involve linear behavior in case of ARFIMA, and FFANN might be trapped into local minima solutions when trained by descent algorithms (Kamdem et al., 2020). To overcome these issues, DL has emerged as an efficient tool in forecasting time-series in fluctuating environments since it is able to consider hidden and latent dynamics of the data.

During the last few years, DL has been used extensively in various domains. For example: ArunKumar et al. (2021) forecasted COVID-19 cumulative cases, recoveries, and deaths over different countries; Kamdem et al. (2020) analyzed the effect of COVID-19 on the prices' volatilities of commodities; Lago et al. (2018) and Memarzadeh and Keynia (2021) studied electricity price forecasting; Liu et al. (2021) focused on bitcoin price prediction; Vidal and Krisjanpoller (2020) examined gold volatility; and Alameer et al. (2020) studied coal price forecasting.

Fig. 1 Basic structure of an LSTM. Source: ArunKumar et al. (2021)



Due to the high ability of DL to forecast time-series, we investigate the performance of various DL algorithms in predicting the prices of six commodities, namely, the Bloomberg Commodity Index, Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex.

2.2 Deep learning techniques

In this section, we briefly review the architecture, parameter settings, and training algorithms of the four DL techniques to investigate the simple Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), LSTM, and Convolutional Neural Network (CNN).

2.2.1 Long short-term memory (LSTM)

LSTM is a sub-class of DL used to model complex relationships, and learn from experience using long-range series (Alameer et al. 2021). It is assumed to solve various problems of classical ANN such as vanishing/exploding gradients, and the concept of forget gates is then used to accomplish this task. The basic structure of an LSTM cell is provided in Fig. 1. Thus, an LSTM cell includes input, output, and forget gates. The input gate is responsible of controlling which data is to be accepted and then transfer it to the cell. The amount of information neglected (and then prevented from passing to the cell) is controlled by the forget gate. This principle is among one of the key features of the LSTM algorithm. The data is then transferred to the output gate, which is responsible for generating the cell output and state. Mathematically, the procedure for LSTM is expressed as follows (Boubaker et al., 2021):

$$f_t = \alpha(w_f.[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \alpha(w_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \alpha(w_o.[h_{t-1}, x_t] + b_o) \tag{3}$$

$$C_t = f_t C_{t-1} + i_t.th(w_c.[h_{t-1}, x_t] + b_c) \tag{4}$$

$$h_t = O_t.\beta(C_t) \tag{5}$$

2.2.2 Gated recurrent unit (GRU)

As shown in Fig. 2, the GRU is similar to the LSTM except that it does not include a forget

Fig. 2 Basic structure of GRU cell. *Source:* ArunKumar et al. (2021)

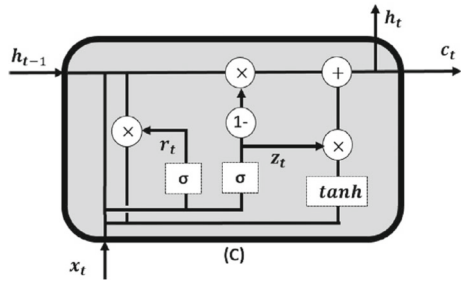
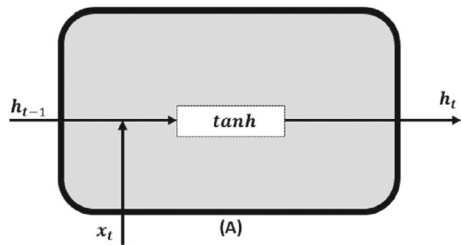


Fig. 3 Basic structure of an RNN cell. *Source:* ArunKumar et al. (2021)



gate. With this relatively simple structure, a GRU model training is computationally more tractable than an LSTM model. GRU is infact based on the idea of updating the cell state.

2.2.3 Recurrent neural network (RNN)

The simple RNNs are improved versions of the classical ANNs (AruKumar et al., 2021). Unlike ANNs, RNNs have the ability of remembering the input data features. They use the concept of cell state which keeps the previous data features and uses them to calculate the new cell state. RNN structure is obviously simpler than both GRU and LSTM. The RNN structure is shown in Fig. 3 below.

2.2.4 Convolution neural network (CNN)

CNNs are a type of deep neural networks that were initially designed for computer vision applications such as data classification and object recognition. CNNs have the ability to extract relevant features from high dimensional data (Koprinska, Wu & Wang, 2018). Although they were designed for two-dimensional data problems, they were later adapted successfully to one-dimensional problems such as in Boubaker et al. (2021) where they succeeded in predicting daily global horizontal irradiation (GHI). Other relevant applications are evapotranspiration time-series (Lucas et al., 2020) and energy time-series (Koprinska, Wu & Wang, 2018).

The principle of working of CNNs is based on three cascaded layers: a pooling layer sandwiched between a convolutional layer and a fully connected layer.²

² More details about CNNs can be found in (Lucas et al., 2020) and the references therein.

Table 1 Optimal hyperparameters of RNN, GRU, and LSTM deep neural networks used in this study

Parameter	Value	Meaning
Units	100	Number of cells in a DL structure
Activation	Relu	Returns a value if the input is greater than 0 and returns 0.00 otherwise
Batch-size	512	Used to distinguish the common features of the input data
Epochs	100	Number of “iterations” for the training phase
Scaler	Min–Max	Allows to scale the data within the interval [-1,1]
Optimizer	Adam	The adaptive moment (Adam) estimation is an improved descent algorithm for training deep neural networks

2.2.5 Deep learning algorithms hyperparameters’ setting

Despite their ability to forecast time-series, DL algorithms (like any other neural network) have the drawback of difficult hyperparameters’ setting and the dependence on many factors (Lago et al., 2018). DL parameter tuning is always based on a trial-and-error procedure. This is mainly due to the fact that DL algorithms are not generically extendable to other case studies. A parameter set may work well for a dataset and may not work for another set even in the same applied research field. In this study, we have conducted several trial-and-error runs for using the three DL algorithms. The performance metrics (will be discussed later in this paper) for the commodity prices under investigation were recorded. A careful study of those performance indicators helped us in tuning the DL parameters. Since our study is mainly related to the comparison of three DL algorithms’ ability in forecasting commodity prices, the model parameters are kept the same in order to make sense of the comparison. The adopted optimal parameters are shown in Table 1.

2.3 Robustness check

We compare our proposed DL model with the traditional ML and time series models such as the ANN (Fig. 4) and autoregressive fractionally integrated moving average models (ARFIMA (p,d,q)) to show the superiority of the former as compared to the later approaches. First, the ARFIMA (p,d,q) models are time-series models that generalize ARIMA (p,d,q) by allowing to take non-integer values of the differencing parameter d and can be described as follows:

$$\Delta^d y_t = \delta + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{i=0}^q \sigma_i \varepsilon_{t-i} + d + e \quad (6)$$

where $y(t)$ is the output (the commodities indices); $y(t - i)$ are the previous values of the output (the commodities indices); δ is a constant and d is the order of integration; φ_i , and σ_i denote the coefficient parameters; k , p , and q are the maximum time lag related to output sequence and residuals, respectively and $e(t)$ is an error term. To find the best fit of ARFIMA (p,d,q) for each series of commodities indices, we employ a mixed approach composed of two methods: First, the box and Jenkins method to identify the autoregressive (AR) part and the moving average (MA) part based on the partial autoregressive functions (PACF) and the autoregressive functions (ACF). The Akaike Information Criteria (AIC) is used to choose the best models from the various possible ARFIMA (p,d,q) models. Second, we use the ANN

model where the commodity variable $y(t)$ is defined as a function of its historical values as follows (Herrera et al., 2019):

$$y(t) = g(y(t - 1), y(t - 2), \dots, y(t - p)) + e(t) \tag{7}$$

where $e(t)$ is an error value following usually a normal distribution. In the ANN paradigm, the commodity model inputs are its p previous values (called patterns) and the value at the current day is considered as the target. The main objective of the ANN model is to map sets of patterns to their corresponding targets regardless of the nature of the hidden relationships.

We first adopt the multi-layer feed-forward neural network to model the relationship between the prices of some selected commodities and their history. A basic FF-ANN operates according to the following equations:

$$\begin{aligned} a^1 &= f^1(W^{1.1}X + b^1) \\ a^2 &= f^2(W^{2.1}a^1 + b^2) \\ &\vdots \\ a^j &= f^j(W^{j,j-1}a^{j-1} + b^j) \\ &\vdots \\ a^N &= f^N(W^{N,N-1}a^{N-1} + b^N) \end{aligned} \tag{8}$$

where:

$a^j, j = 1: N$ are the outputs of the respective layers; a^1 is the output of the input layer; and a^N is the output of the output layer. $W^{j,j-1}$ are the weights of the j th layer. b^j are the biases of the j th layer. f^j is the transfer function of the j th layer.

The number of hidden layers as well as their numbers of neurons are usually determined using trial-and-error procedure (Jnr et al., 2021). This is considered as the main drawback of ANNs. The objective of the training process is to minimize the error between the targets and the ANN outputs as described below:

$$\left(\varepsilon^2(W_i, bi) = \sum_{t=1}^N (y(t) - g(y(t - 1), y(t - 2), \dots, y(t - p)) + e(t))^2 \right) \tag{9}$$

During the training phase, the training algorithm updates iteratively the ANN weights and biases as in the following equation:

$$W(k) = W(k - 1) + \alpha \Delta W(k - 1)X^T \tag{10}$$

where k is the iteration index, α the learning rate, and $\Delta W(k - 1)$ is the error function. For the implementation of ANN, various activation functions and training algorithms exist. However, the choice of the suitable function and training algorithm are case-sensitive. Hence, after trying several combinations of the ANN structure and the training algorithm, the parameters adopted are summarized in Table 2.

2.4 Performance metrics

The dataset has been divided into 80% for training and 20% for validation (Kamdem et al., 2020). Three performance metrics have been used to measure the quality of the developed forecasters (Liu et al., 2021) and are given as follows:

Table 2 Parameters of the used ANN algorithm

Parameter	Our choice
Number of input layers	1
Number of hidden layers	2
Number of neurons in hidden layers	5
Activation functions	Linear function for all layers
Training algorithm	Resilient-propagation

- Table 4 displays the mean absolute percentage error

$$MAPE = \frac{100}{N} \sum_{t=1}^N \frac{|y(t) - \hat{y}(t)|}{\bar{y}} \quad (11)$$

- The coefficient of determination (R^2)

$$R^2 = 1 - \frac{\frac{1}{N_2} \sum_{t=1}^N (y(t) - \hat{y}(t))^2}{\frac{1}{N} \sum_{t=1}^N (y(t) - \bar{y})^2} \quad (12)$$

- The root mean square error

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t))^2} \quad (13)$$

- Mean absolute deviation

$$MAD = \frac{\sum |y(t) - \hat{y}(t)|}{N} \quad (14)$$

where $\hat{y}(t)$ and $y(t)$ denote the forecasted and the observed (real) commodity price at day t , respectively; \bar{y} is the average value of the same price; and N is the size of validation sample.

3 Empirical results

3.1 Data

Our daily data from January, 2002 (the beginning wave of commodity markets' financialization) to December, 2020 covers a set of commodity assets, that is, the Bloomberg Commodity Index (noted by the Global Aggregated Index) and its five component indices: Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Live-stock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex. Therefore, we obtained 5473 observations for the first four series and 5270 observations for the last two indices (the Bloomberg Industrial Metals Subindex and Bloomberg Energy Subindex). For the first set of variables, the sample was subdivided into two subsamples. The first subsample (5000 observations) was used for the training algorithms whereas the second subsample (472 observations) was used for the validation algorithms. For the last two

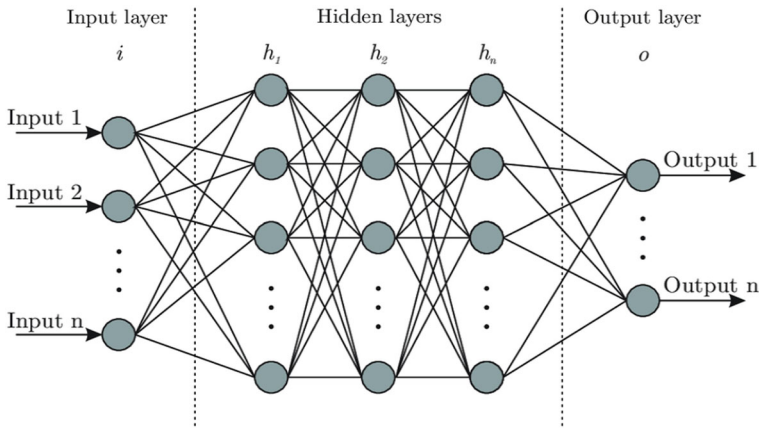


Fig. 4 Structure of an artificial neural network (ANN)

Table 3 Descriptive statistics

	Mean	Max	Min	S.D	Skew	Kurt
<i>Bloomberg Commodity Index</i>	122.580	237.953	59.479	35.024	0.428	2.522
<i>Bloomberg Agriculture Subindex</i>	62.907	101.822	34.152	13.843	0.248	2.744
<i>Bloomberg Precious Metals Subindex</i>	143.631	306.846	49.663	63.466	0.148	2.106
<i>Bloomberg Livestock Subindex</i>	47.737	84.390	16.209	19.303	0.393	1.589
<i>Bloomberg Industrial Metals Subindex</i>	126.730	266.763	53.902	46.677	0.620	2.693
<i>Bloomberg Energy Subindex</i>	148.177	516.656	15.469	109.983	0.887	2.926

Table 3 indicates the mean, standard deviation (SD), skewness, kurtosis, minimum (Min.), maximum (Max.) Bloomberg Commodity Index and its five components indices: Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex for the sample period from January, 2002 to December, 2020.

variables of the commodities, the training phase was based on the 5000 observations and the validation phase was based on 269 observations.

Table 3 shows that the skewness values for all indices differ from 0, that is, the distributions occur in an asymmetrical pattern. Furthermore, the values of kurtosis test was much lower than 3. This shows that, for all the variables, the distribution of the commodity assets had lighter tails than those of normal distributions. This result is synonymous of the acceptance of the existence of non-linearity in the data.

Figure 5 plots the autocorrelation pattern of all the series which shows that all the variables of the commodity variables had a unit-root problem. The series data varied further away from zero. Thus, all the series had long-memory behavior. Ftiti et al. (2020) illustrate that the existence of non-linear structures and long-memory patterns in dynamic systems can be attributed to the inside variables of the system itself. Hence, a non-linear dynamic approach should be used as part of forecasting to deal with it, since this type of model does not necessitate a transformation of the original data. However, the authors also indicate that implementing the classical approaches as linear tools could overcome the problem of unit root before the commencement of the forecasting process.

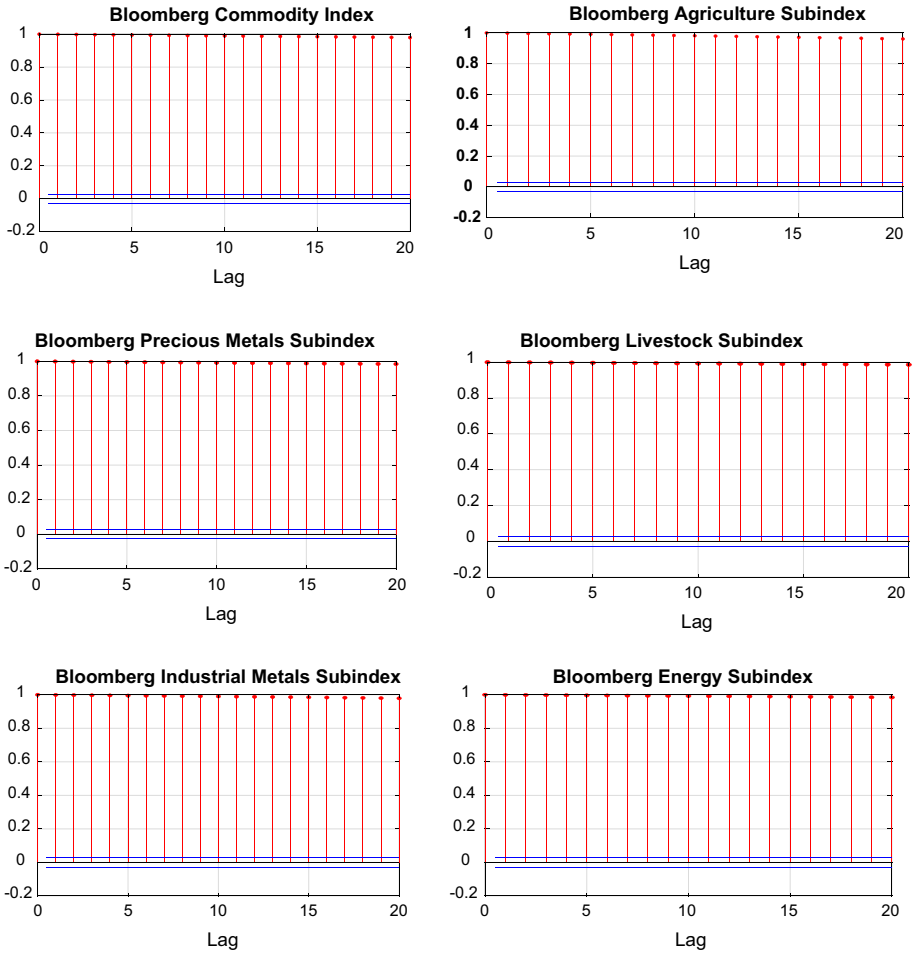


Fig. 5 Autocorrelation test

3.2 Forecasting results

Figure 6 plots all six forecasts for the commodity indices using the validation sets. The different complex DL models—LSTM, RNN, GRU, and CNN—highlighted that all curves of the DL models (forecasted values) were located near to the blue curve (actual values) for all the commodities indices. Clearly, the suggested DL tools exhibited a strong follow-up behavior during the validation phase. More specifically, for the LSTM, RNN, and GRU, the lines representing the actual price and the forecasted price completely correspond with each other for all commodities assets, even during the period where the price of commodities varies largely. Although the CNN method provided a good accuracy for Bloomberg Commodity Index, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, and Bloomberg Industrial Metals Subindex, it fails to accurately predict Bloomberg Agriculture Subindex and Bloomberg Energy Subindex. According to Fig. 6, we can observe that there is no significant gap between the true curve (in blue) and the predicted curve (in yellow) of all

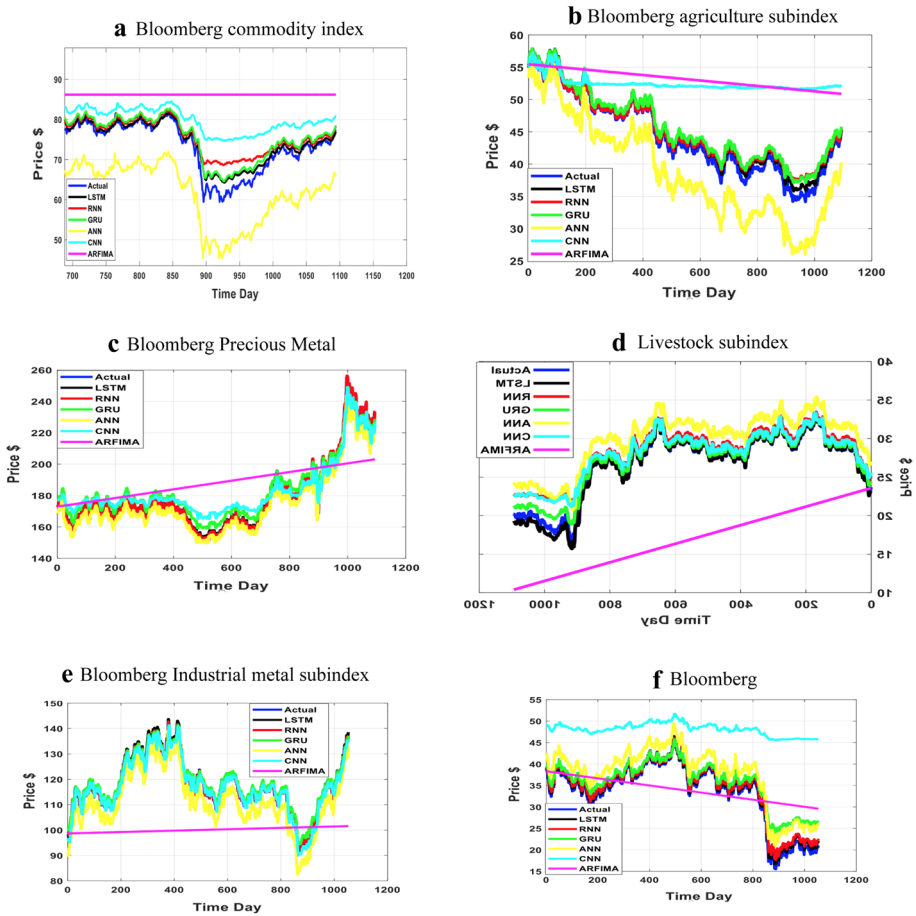


Fig. 6 Validation phase: Plot of actual and forecasted commodities indices

indices, except in the Bloomberg Commodity Index and Bloomberg Agriculture Subindex where the traditional ANN method is applied for forecasting. Additionally, the ARFIMA tool fails to provide accurate predictions. Hence, all the DL methods (except for the CNN model) have shown a similar tendency and forecasts with an overall accuracy on the whole validation period. Additionally, findings show that DL methods dominate traditional ML and classical time series tools in predicting commodities indices with good accuracy. Hence, the proposed LSTM, DRU, and RNN models are the best models for prediction.

However, for the Bloomberg Commodity Index, Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex and Bloomberg Industrial Metals Subindex, all DL models succeed to generate curves of current price closer to forecast curves during the tranquil period.³ Additionally, this good accuracy is also evident during the pandemic period where the prices fluctuate significantly. We indicate that the superiority of LSTM to forecast the commodities indices compared to RNN, GRU and CNN models during COVID-19 outbreak

³ Period before Wednesday March 11th, 2020 that is the date of announcement that the new Corona virus, which is spreading throughout the world, is a “global epidemic.

period. The potential reason behind this is that the LSTM is able to solve problems related to long-term and short-term dependency memory (Memarzadeh & Keynia, 2021).

Table 4 reports the outputs of evaluation metrics (Mean absolute percentage error (MAPE), the root mean squared error (RMSE), and the determination coefficient (R^2)) of the proposed forecast tools across the different commodities. It shows that the DL models offer a better forecast when compared to ANN and ARFIMA models. For the Bloomberg Commodity Index, Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, and Bloomberg Industrial Metals Subindex, the LSTM model provides lesser RMSE values when compared to the competitor's models. However, the RMSE of RNN model is lesser than that of the LSTM model for Bloomberg Livestock Subindex. For the rest of commodities, the GRU model has a lesser RMSE value than that of LSTM and RNN models. In terms of MAPE metric, the LSTM performed well with lesser MAPE than that of the rest of models in case of Bloomberg Commodity Index, Bloomberg Agriculture Subindex, Bloomberg Livestock Subindex, and Bloomberg Energy Subindex. However, the RNN model for Bloomberg Precious Metals Subindex and Bloomberg Industrial Metals Subindex has lesser MAPE values when compared to that of the LSTM model. Table 4 provides evidence that the LSTM model outperformed competitor models with higher values of R^2 for all commodities indices, except for the Bloomberg Energy Subindex where the ANN model indicates a higher R^2 . Hence, these results indicate that the LSTM model shows its dominance according to the performance metrics and is the best forecaster tool compared to the other DL, ML, I and time-series models. Thus, we also compare the commodities indices based on the LSTM tool. We found that the according to the RMSE and R^2 metric, the Bloomberg Precious Metals Subindex is the index that is well forecasted by its lagged values and it dominates all other indices. Furthermore, the LSTM findings indicate that the Bloomberg Livestock Subindex has the smaller value of MAPE which highlights that this asset is dominant compared to the other five indices. Finally, the lower MAD metric generated by LSTM also indicates its outperformance as forecasting tool. Additionally, MAD results are in line with the MAPE results for the Bloomberg Livestock Subindex which indicates that it is dominant as compared to the remaining five indices (Tables 4).

Figure 7 depicts the distributed plots of the forecasted values of all the commodity indices. The bisector curves indicate a perfect forecast having great accuracy. For all the LSTM, RNN and GRU forecasting tools, we provide evidence that the points on or close to the bisector denoted the best correspondence between the forecasted values and the actual values of different commodities indices. However, the CNN and ARFIMA models performed poorly. We found that the points are far on the bisector, which indicates a significant gap between the predicted and actual values of the various commodity indices.

The validity of the results is emphasized when we forecast for the Bloomberg Livestock Subindex and Bloomberg Industrial Metals Subindex by using the LSTM model. We confirm the effectiveness of the LSTM method as a forecasting tool and the superiority of Bloomberg Livestock Subindex and Bloomberg Industrial Metals Subindex compared to the other commodity indices. These findings are consistent with Kamdem et al. (2020) that proved the superiority of LSTM model in forecasting the commodities prices. More interestingly, our study complete the abundant literature related to commodity prices forecasting through showing the superiority of Deep Learning approaches compared to machine learning as well as conventional approaches (Fig. 6).

Table 4 Comparison of the value of RMSE, R2, and MAPE between the proposed models

Models	Bloomberg Commodity Index			Bloomberg Agriculture Subindex			Bloomberg Precious Metals Subindex					
	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)
LSTM	1.27	0.97	1.06	0.85	0.98	0.98	1.90	0.85	0.72	0.99	2.03	1.30
RNN	2.23	0.90	1.80	1.45	1.25	0.96	2.20	0.98	2.11	0.99	0.75	1.35
GRU	1.85	0.93	1.90	1.53	1.48	0.94	2.92	1.30	5.11	0.95	2.58	4.62
CNN	4.92	0.54	4.61	3.71	9.59	- 1.41	18.11	8.07	7.27	0.89	3.41	6.12
ANN	7.22	0.96	4.95	10.00	3.75	0.93	4.78	5.48	6.32	0.99	3.73	4.57
ARFIMA	9.23	- 0.63	8.19	6.59	10.02	- 1.63	19.87	6.59	19.54	0.22	8.89	6.59
Models	Bloomberg Livestock Subindex			Bloomberg Industrial Metals Subindex			Bloomberg Energy Subindex					
	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)	RMSE (\$)	R2 (%)	MAPE (%)	MAD (\$)
LSTM	1.42	0.98	0.51	0.38	0.82	0.99	1.25	0.96	2.31	0.98	0.95	0.77
RNN	1.73	0.83	5.04	1.35	1.26	0.99	0.83	0.97	1.24	0.97	3.11	1.04
GRU	0.79	0.96	2.21	0.59	1.73	0.97	1.17	1.38	3.24	0.80	7.22	2.41
CNN	1.63	0.84	4.51	1.21	2.14	0.96	1.43	1.68	15.94	- 3.72	44.30	14.77
ANN	2.75	0.98	5.30	3.01	6.40	0.98	4.65	5.97	9.22	0.99	4.27	4.21
ARFIMA	10.33	- 5.22	36.78	9.89	20.66	- 2.78	15.50	18.26	6.09	0.31	14.56	4.86

Table 4 displays the mean absolute percentage error(MAPE(%)), the coefficient of determination (R2(%)), the root mean square error(RMSE(\$)), the mean absolute deviation(MAD(\$)). Deep learning techniques are: Long short-term memory (LSTM), Gated recurrent unit (GRU), recurrent neural network (RNN), and Convolution neural network (CNN). The traditional tools are the artificial neural network (ANN) and the autoregressive fractionally integrated moving average models (ARFIMA). Bloomberg Commodity Index and its five components indices: Bloomberg Agriculture Subindex, Bloomberg Precious Metals Subindex, Bloomberg Livestock Subindex, Bloomberg Industrial Metals Subindex, and Bloomberg Energy Subindex for the sample period from January, 2002 to December, 2020

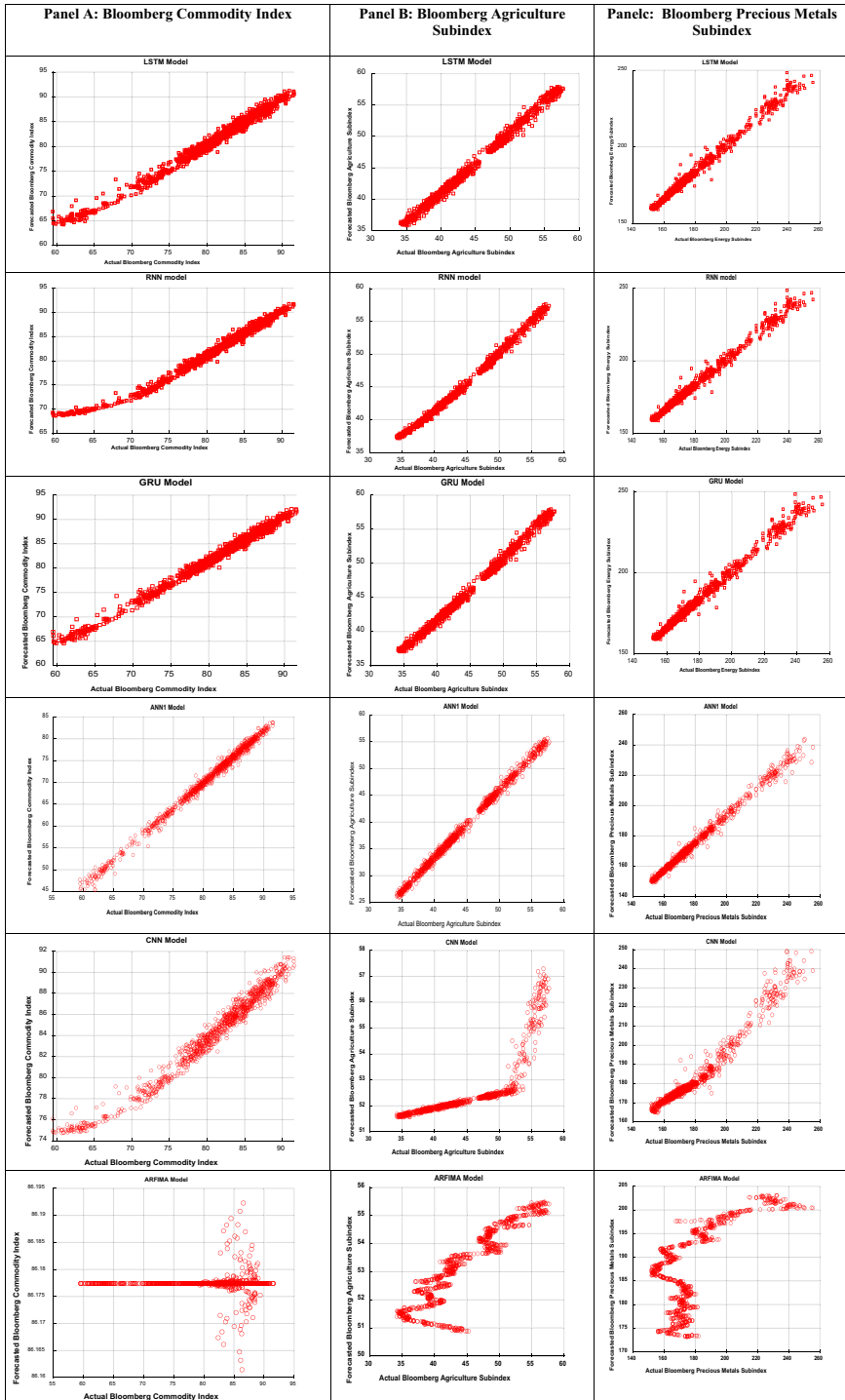


Fig. 7 Dispersed plot of forecasted values

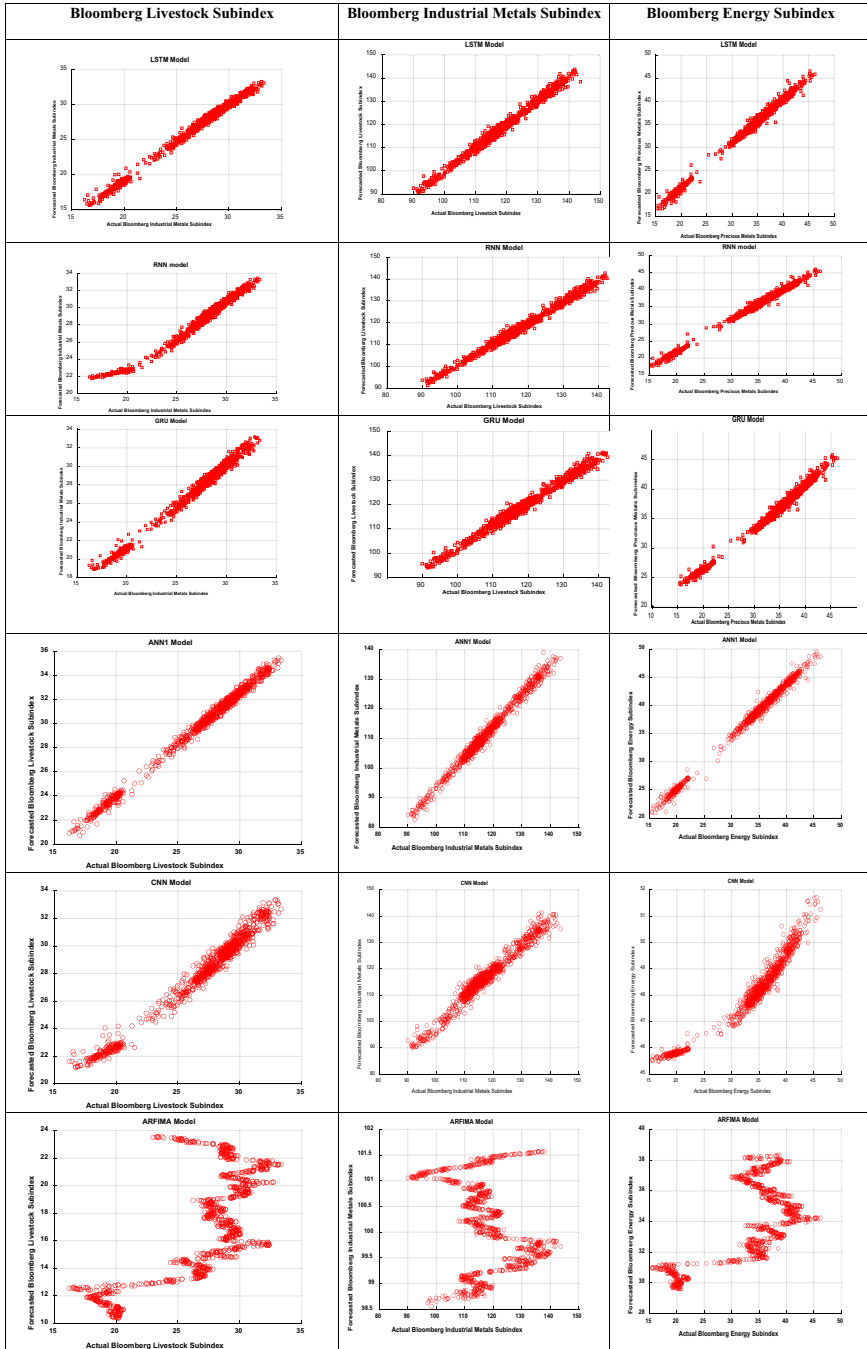


Fig. 7 continued

4 Conclusion and implications

In this study, we focused on the aspect of forecasting commodity prices, as they play an important role in providing implications for economic situations. We show the main mechanisms that can cause commodity prices to affect the economy and financial systems. Particularly, we discussed the theoretical motivation regarding why both developing and developed countries show a great interest in following the behavior and pattern of the main commodity markets. Therefore, the forecasting of commodity prices is a vital activity for investors, hedgers, and policymakers.

Our study proposed the use of various DL tools and showed that the LSTM method is the best-fit model. We highlight the superiority of Bloomberg Livestock Subindex and Bloomberg Industrial Metals Subindex compared to other commodity indices.

This study will provide value-added theoretical and practical contribution to the literature. In terms of theoretical implications, our results contribute to the larger area of financialization of commodity markets–portfolio diversification relationship. Yet, an accurate forecast of commodities assets using DL models permit to monitor commodity markets that are exposed to structural adjustments in return distribution characteristics as they relate to other types of traded assets. As a result, this will allow realizing an optimal allocation of financial capital and hedging demand in commodity markets, and consequently achieve diversification benefits. Additionally, the good fit of forecasted values of commodities indices indicate an attractive risk–return trade off in commodity markets. This is important as it guarantees the continuity of the commodity financialization process.

Concerning managerial implications, the outperformance of DL models in forecasting commodity prices can help policy makers in commodity markets to learn from commodity price dynamics, even in fluctuating environments. This allows them to effectively monitor the macroeconomic situation and inflation. Good forecasting accuracy of commodity indices also motivates investors to perform efficient portfolio management based on commodity futures management to achieve portfolio diversification and optimal asset allocation. Furthermore, DL tools show a stronger ability to accurately capture the complex structure in commodities indices. This means that they are able to predict commodities' price, despite the fact that the sample used does have crisis periods. The good appropriateness of these methods will prompt investors to profit from these complex tools for the forecasting of commodities prices and other risky assets with good accuracy.

References

- Ahrens, W. A., & Sharma, V. R. (1997). Trends in natural resource commodity prices: Deterministic or stochastic? *Journal of Environmental Economics and Management*, 33(1), 59–74. <https://doi.org/10.1006/jeeem.1996.0980>
- Alameer, Z., Fathalla, A., Li, K., Ye, H., & Jianhua, Z. (2020). Multistep-ahead Forecasting of Coal Prices Using a Hybrid Deep Learning Model. *Resources Policy*, 65.
- ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2021). Forecasting of COVID-19 using deep layer Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) cells. *Chaos, Solitons & Fractals*, 146, 110861.
- Asgharian, H., Christiansen, C., & Hou, A. J. (2015). Effects of macroeconomic uncertainty on the stock and bond markets. *Finance Research Letters*, 13, 10–16. <https://doi.org/10.1016/j.frl.2015.03.008>
- Berck, P., & Roberts, M. (1996). Natural resource prices: Will they ever turn up? *Journal of Environmental Economics and Management*, 31(1), 65–78. <https://doi.org/10.1006/jeeem.1996.0032>
- Bodie, Z., & Rosansky, V. I. (1980). Risk and return in commodity futures. *Financial Analysts Journal*, 36(3), 27–39.

- Boubaker, S., Benghanem, M., Mellit, A., Lefza, A., Kahouli, O., & Kolsi, L. (2021). Deep neural networks for predicting solar radiation at hail region, Saudi Arabia. *IEEE Access*, 9, 36719–36729.
- Chen, S.S., Choubey, B., Singh, V. (2021). A neural network-based price sensitive recommender model to predict customer choices based on price effect. *Journal of Retailing and Consumer Services*, 61.
- Dong, K., Sun, R., & Dong, X. (2018). CO2 emissions, natural gas and renewables, economic growth: assessing the evidence from China. *Science of the Total Environment*, 640, 293–302
- Dooley, G., & Lenihan, H. (2005). An assessment of time series methods in metal price forecasting. *Resources Policy*, 30(3), 208–217. <https://doi.org/10.1016/j.resourpol.2005.08.007>
- Fang, L., Bouri, E., Gupta, R., & Roubaud, D. (2019). Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin? *International Review of Financial Analysis*, 61, 29–36. <https://doi.org/10.1016/j.irfa.2018.12.010>
- Fiti, Z., Kablan, S., & Guesmi, K. (2016). What can we learn about commodity and credit cycles? Evidence from African commodity-exporting countries. *Energy Economics*, 60, 313–324. <https://doi.org/10.1016/j.eneco.2016.10.011>
- Fiti, Z., Tissaoui, K., & Boubaker, S. (2020). On the relationship between oil and gas markets: A new forecasting framework based on a machine learning approach. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-020-03652-2>
- Gargano, A., & Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, 30(3), 825–843. <https://doi.org/10.1016/j.ijforecast.2013.09.003>
- Gospodinov, N., & Ng, S. (2013). Commodity prices, convenience yields, and inflation. *Review of Economics and Statistics*, 95(1), 206–219.
- Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J., & Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy*, 179, 214–221.
- Jnr, E. O. N., Ziggah, Y. Y., & Relvas, S. (2021). Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting. *Sustainable Cities and Society*, 66, 102679.
- Kablan, S., Fiti, Z., & Guesmi, K. (2017). Commodity price cycles and financial pressures in African commodities exporters. *Emerging Markets Review*, 30, 215–231. <https://doi.org/10.1016/j.ememar.2016.05.005>
- Kamdem, J. S., Essomba, R. B., & Berinyuy, J. N. (2020). Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities. *Chaos, Solitons & Fractals*, 140, 110215.
- Karali, B., & Power, G. J. (2013). Short-and long-run determinants of commodity price volatility. *American Journal of Agricultural Economics*, 95(3), 724–738. <https://doi.org/10.1093/ajae/aas122>
- Karasu, S. & Altan, A. (2022). Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization. *Energy*, 242, 122964.
- Klein, T. (2017). Dynamic correlation of precious metals and flight-to-quality in developed markets. *Finance Research Letters*, 23, 283–290. <https://doi.org/10.1016/j.frl.2017.05.002>
- Koprinska, I., Wu, D., & Wang, Z. (2018). Convolutional Neural Networks for Energy Time Series Forecasting. In: *2018 International Joint Conference on Neural Networks (IJCNN)* (pp, 1-8).
- Kostrzewski, M., & Kostrzewska, J. (2019). Probabilistic electricity price forecasting with Bayesian stochastic volatility models. *Energy Economics*, 80, 610–620. <https://doi.org/10.1016/j.eneco.2019.02.004>
- Lago, J., Ridder, F.D., Schutter, B.D. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221, 386–405
- Lee, J., List, J. A., & Strazicich, M. C. (2006). Non-renewable resource prices: Deterministic or stochastic trends? *Journal of Environmental Economics and Management*, 51(3), 354–370. <https://doi.org/10.1016/j.jeem.2005.09.005>
- Lintner, J. (1983): “The Potential Role of Managed Commodity-Financial Futures Accounts (and/or Funds) in Portfolios of Stocks and Bonds. In: Paper presented at the annual conference of the Financial Analysts Federation, Toronto, Canada.
- Liu, X., Lin, Z., & Feng, Z. (2021). Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. *Energy*, 227, 120492.
- Lucas, P., Alves, M., Silva, P., & Guimarães, F. (2020). Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. *Computers and Electronics in Agriculture*, 177.
- Madani, M.A., Fiti, Z. (2022). Is gold a hedge or safe haven against oil and currency market movements? A revisit using multifractal approach. *Annals of Operations Research*, 313, 367–400. <https://doi.org/10.1007/s10479-021-04288-6>
- Marquis, M. H., & Cunningham, S. R. (1990). Is there a role for commodity prices in the design of monetary policy? Some empirical evidence. *Southern Economic Journal*, 57(2), 394–412.

- Memarzadeh, G., & Keynia, F. (2021). Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electric Power Systems Research*, 192, 106995.
- Narayan, P. K., Narayan, S., & Sharma, S. S. (2013). An analysis of commodity markets: What gain for investors? *Journal of Banking & Finance*, 37(10), 3878–3889. <https://doi.org/10.1016/j.jbankfin.2013.07.009>
- Nguyen, D. K., Walther, T. (2020). Modeling and forecasting commodity market volatility with long-term economic and financial variables. *Journal of Forecasting*, 39, 126–142.
- Panella, M., Barcellona, F., & D'eclesia, R. L. (2012). Forecasting energy commodity prices using neural networks. *Advances in Decision Sciences*, 2012
- Prokopczuk, M., Symeonidis, L., & Simen, C. W. (2017). Variance risk in commodity markets. *Journal of Banking & Finance*, 81, 136–149. <https://doi.org/10.1016/j.jbankfin.2017.05.003>
- Slade, M. E. (1988). Grade selection under uncertainty: Least cost last and other anomalies. *Journal of Environmental Economics and Management*, 15(2), 189–205. [https://doi.org/10.1016/0095-0696\(88\)90018-6](https://doi.org/10.1016/0095-0696(88)90018-6)
- Smales, L. A. (2017). Commodity market volatility in the presence of US and Chinese macroeconomic news. *Journal of Commodity Markets*, 7, 15–27. <https://doi.org/10.1016/j.jcomm.2017.06.002>
- Szarek, D., Bielak, Ł., & Wylomańska, A. (2020). Long-term prediction of the metals' prices using non-Gaussian time-inhomogeneous stochastic process. *Physica a: Statistical Mechanics and Its Applications*, 555, 124659. <https://doi.org/10.1016/j.physa.2020.124659>
- Vidal, A., & Kristjanpoller, W., (2020). Gold volatility prediction using a CNN-LSTM approach, *Expert Systems with Applications*, 157, 113481
- Zhang, P., & Ci, B. (2020). Deep belief network for gold price forecasting. *Resources Policy*, 69, 101806

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.