Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities

Jules Sadefo-Kamdem, Rose Bandolo Essomba, James Njong Berinyuy

Deep Learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities

Jules Sadefo Kamdem (MRE EA 7491, University of Montpellier)
Rose Bandolo Essomba (AIMS Cameroon)
James Berinyuy Njong (AIMS Cameroon)

July 13, 2020

Abstract

Over the past few years, the application of deep learning models to finance has received much attention from investors and researchers. Our work continues this trend, presenting an application of a Deep learning model, long-term short-term memory (LSTM), for the forecasting of commodity prices. The obtained results predict with great accuracy the prices of commodities including crude oil price (98.2% of accuracy), Brent price(88.2%), Silver price (98.2%) and wheat price (97.8%). We also analyse the impact of the coronavirus on the variability of the commodity prices. This involved checking at the correlation and the causality with the Ganger Causality method. Our results reveal that the coronavirus impacts the recent variability of commodity prices through the number of confirmed cases and the total number of deaths. We then investigate a hybrid ARIMA-Wavelet model to forecast the coronavirus spread. This analyses is interesting as a consequence of the strong causal relationship between the coronavirus(number of confirmed cases) and the commodity prices, the prediction of the evolution of COVID-19 can be useful to anticipate the future direction of the commodity prices.

Keywords: *Forecasting, Deep learning, LSTM, ARIMA-Wavelet, commodity market volatility, Covid-19 Spread, Pandemics Risks.*

1 Introduction

The commodity market is an environment where things change quickly and people like to be prepared for unexpected events. This market is characterized by large price fluctuations and involves many players such as brokers, traders, portfolio managers and investors in general, who anticipate each other's actions, especially in times of great uncertainty, making the market noisy and fluctuating [1]. Many African countries that depend heavily on the export of commodities such as precious metals and mineral oil, began trading in the commodity exchange as far back as the 1990s [2]. Over the last decade, these countries have experienced impressive economic growth due to various commodity price booms. Today, with the drop in commodity prices due to the COVID-19 pandemic, the economies of many countries are being severely affected. With countries that fail to diversify their exportation, it's clear that a fall in commodity prices will drastically affect them. Oil exporting countries such as Nigeria, Gabon, Angola and Cameroon have been critically affected by the recent fall in crude oil price. However, with the globalization of commodity markets it is possible to have access to financial derivatives such as futures, forwards and options to hedge against risks related to price volatility. These securities help in transferring the risk associated to commodity exports to investors looking for speculative opportunities. Since these securities are priced based on predicted commodity prices, it is clear that if your prediction is wrong, you won't be able to take the right investment decision and consequently may suffer huge losses. That is why the financial industry is constantly looking for new tools that can successfully predict the price of financial assets to make efficient market decisions.

In the past, several techniques have been used for financial time series forecasting, among which are the traditional predicting methods. As a matter of fact, the Autoregressive Integrated Moving Average (ARIMA) and its variants are the most used in the literature for forecasting stock price series ([3], [4]). These have indicated that the ARIMA model has a strong predictive potential in the short term and has the ability to compete favourably with existing stock price prediction tools. Unfortunately, these models do not consider the latent dynamics existing in the data and they are based on linearity assumptions. Since financial data are not most of time linear, researchers started to look on others tools which can perform better than traditional statistical tools. Classical machine learning then started to emerge. Numerous studies on machine learning models had been developed for financial time series prediction. Among them, the Artificial Neural Network (ANN) model is very

popular due to its ability to learn models from data and deduce solutions from unknown data. Some related works have used the ANN model in comparison with the ARIMA model to predict stock prices [[5],[6],[7],[8],[9]]. The results of these papers showed that the ANN model is the best predictive model with higher accuracy than the ARIMA model. This allows them to conclude that neural networks are a useful tool for predicting stock prices in emerging markets. However, many papers have reported that the ANN model, trained by the back propagation algorithm, has some limitations in forecasting [10], and it can easily converge to the local minimum because of the tremendous noise and complex dimensionality of the stock market data. In view of these limitations, Deep Learning has been proposed to overcome the local convergence issue for non-linear optimization problems and enhance the accuracy of the ANN model.

Deep Learning (DL) is an exceptional case of Artificial Neural Network (ANN) that consists of multiple processing layers and allows high-level abstraction to model data. The main advantage of DL models is the automatic extraction of good input data by means of a general purpose learning procedure. The literature has pointed different types of deep neural network models for financial time series forecasting, the most popular of which are the Convolutional Neural Network (CNN), the Recurrent Neural Network (RNN) and its variants. Convolutional Neural Networks (CNNs) are appropriate for problems such as image recognition. [11] proposed a deep learning method based on a Convolutional Neural Network to forecast the stock price movement of Chinese stock market. The results of their study show that it is somewhat reasonable to use deep learning methods based on Convolutional Neural Network to predict the stock price fluctuation of China. [12] improved the algorithmic trading framework with the proposed deep convolutional neural networks (CNN) to predict the Taiwan Stock Index .The results show that the deep learning technique is effective for trading simulation and have greater potentials to model the noisy financial data. Despite this ability to handle noise of financial times series, CNN models cannot learn temporal dependencies. The Recurrent Neural Networks especially the Long Short Term Memory (LSTM) have been introduced to improve this limitations of CNN models. LSTM models provide multi-variate input, robustness to noise, multi-variate output, automatic feature extraction, modelling the more complex relationships in the data and also, LSTM read an input sequence data into the model as a separate input vector. [13] conducted a study where they compare ARIMA and LSTM for forecasting financial time series. The result shows that LSTM outperform traditional ARIMA model. More specifically, they find the average reduction in error rates for the LSTM between 84 -87 percent when compared to ARIMA indicating the superiority of LSTM to ARIMA. [14] introduced an RNN-based approach to predict stock returns. The idea was to build portfolios by adjusting the threshold levels of returns by internal layers of the RNN built.

The purpose of this study is therefore to use deep learning techniques, to forecast future prices as well as intra-daily volatility of commodities exported by African countries (Aluminum, oil, Brent, gold, wheat, silver) . In the meantime, this brief will address a measure of intra-daily market risk. We will also find the implications of Covid-19 sanitary crisis on commodity market volatilities in the period of January-March 2020. Its primary objectives are :

1. Perform an empirical study of the deep learning tools for computational finance and contrast them with both traditional machine learning and classical time series analysis techniques.

2. Construct deep learning models for predicting intra-day commodity prices in a highly volatile market environment.

3. Using the models constructed, study the impact of COVID-19 on the intra-day commodity prices and check that the market behaviour is consistent with the efficient market hypothesis.

The rest of the paper is organised thus: in Section 2 we discuss about the methodology we will use to forecast the commodity price, more precisely we present in this section the key idea behind the Long Short Term Memory model(LSTM). The prediction of commodity prices and the impact of COVID 19 on the commodity market is the subject of discussion in Section 3. The Section opens with a presentation in details of the forecasting commodity price with LSTM model. In Section 4 we analyse the impact of coronavirus on the variability of the commodity prices. This involved check the correlation and the causality between the two variables using statistical tools such as Pearson correlation and Granger causality . The conclusion of this part drop us in section 5 for the forecasting of the Covid 19 spread in Cameroon and France using the hybrid integrated autoregressive moving average model and the wavelet-based prediction model. The aim of this prediction is to anticipate the future fluctuation of the commodity price based on the prediction of Covid 19 since we find a correlation or causality between the two variables. We end our study with section 6, reserve for discussion and the conclusion.

2   Methodology

LSTM is a type of Recurrent Neural Network (RNN). Recurrent neural network is designed to avoid long-term dependence problems and is well suited for processing and predicting time series. Developed by Sepp Hochreiter and Jurgen Schmidhuber in 1997 [15], the LSTM model is based on a unique set of memory cells that replace neurons in the hidden layer of RNN, and its key is the state of the memory cells. We will see in this part the architecture of LSTM model, his conception and the mathematic it involved.

2.1   LSTM Network Architecture

The LSTM architecture take it origin in the Neural Network model. The idea behind the Neural network is to construct models based on the human biological neurons functional. The neurons are take as input and these inputs are summed with weight to produce what we call activation inputs. If the activation inputs triggers activation function we have an output. The activation function is a mathematical equation that will determine the output of the neural network. This activation function is in charge of transforming the weighted input added to the node into activation of the output for this input. Figure 2.1 present an Neural Network structure.
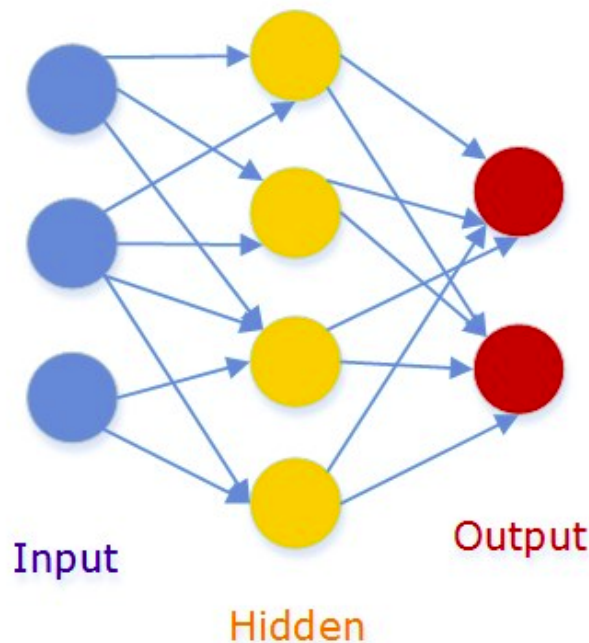

Figure 1: Neural Network structure

Actually, the issue with the Neural networks is the lack of memory. Luckily, the LSTM model don't have this issue. This is one of his big advantage.The LSTM model have in his structure a set of recurrently connected subnets, known as memory cell blocks. Those memory cell blocks facilitate information storage. In every block we have one or more self-connected memory cells and three multiplicative units that are: the input, output and forget gates. Figure 2.1 displays the structure of a LSTM model.

To get a well formed understand what is behind an LSTM model, we will explain step by step the operations of the memory cells.

2.2   The Core Idea Behind LSTM

As we said before, LSTM model is based on connected memory cell blocks. Each memory cell is composed with three sigmoid layers and one tangent hyperbolic(tanh) layer. In figure 2.2 we present the structure of LSTM memory cells.

The first step in the LSTM is to determine which information will be discarded from the model. This decision is ensured by a sigmoid layer called the "forget gate layer" . As we can see from figure 2.2, the forget gate layer examines the output h_{t-1} of the previous moment and the external information x_t of the current moment as inputs and combines them in a long vector $[h_{t-1},\ x_t]$ through σ transformation to become

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f), \tag{1}$$
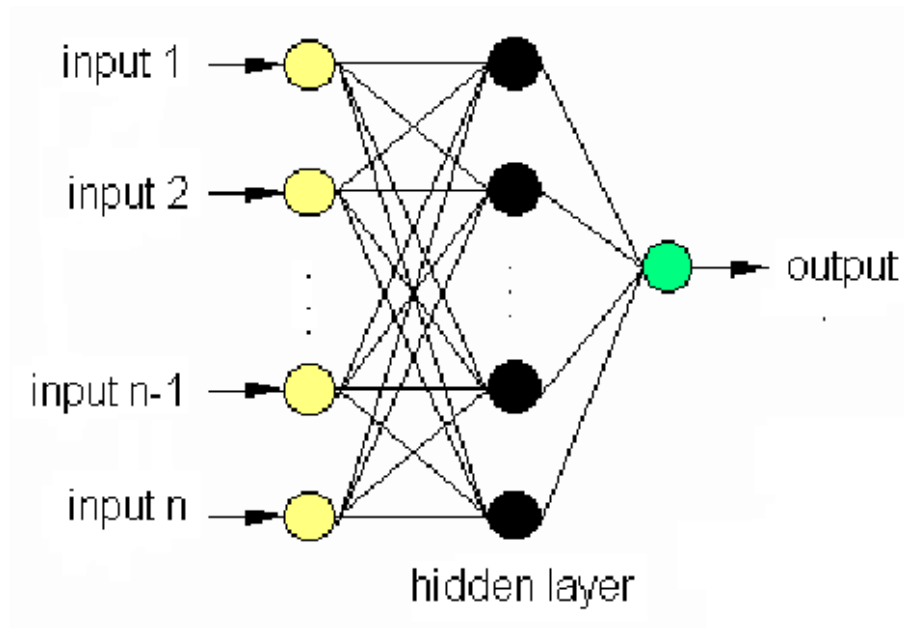
3
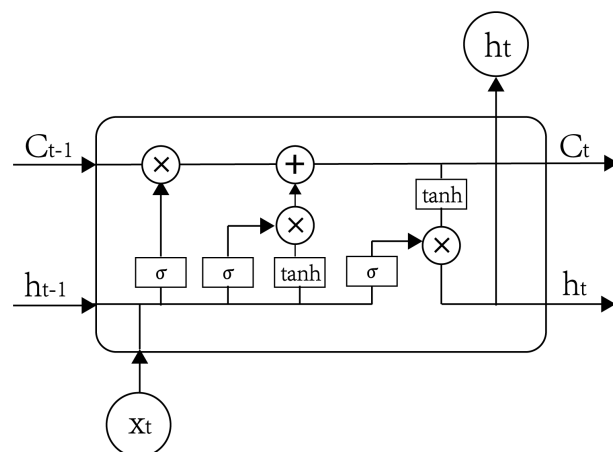
Figure 2: Architecture of long short term(LSTM)
Source: https://www.researchgate.net/figure/LSTM

Figure 3: Structure of long short term memory(LSTM)
Source: https://doi.org/10.1371/journal.pone.0227222.g001

where W_f and b_f are, respectively, the weight matrix and bias of the forgotten gate and $\sigma = \dfrac{1}{1+e^{-z}}$ is the sigmoid function. The forgotten gate's main function is to record how much the cell state C_{t-1} of the previous time is reserved to the cell state $\hat{C}_t$ of the current time. The gate will output a value between 0 and 1 based on h_{t-1} and x_t, where 1 indicates complete reservation and 0 indicates complete discardment.
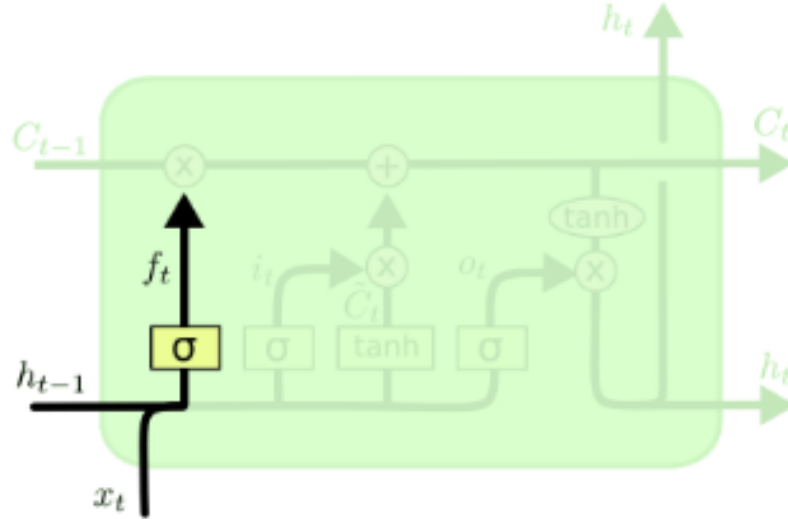
Figure 4: First step visualization in the LSTM
Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

The next step is to decide what new information will be store in the cell state. This is doing in two parts. First, a sigmoid layer called the "input gate layer" find the state of the cell that must be updated as in equation 2 . Next, a hyperbolic tangent (tanh) layer creates a vector of new candidate values, C_t, to control how much new information is added as in equation 3.

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

$$\hat{C}_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{3}$$

The process of this step is shown in figure 2.2

Figure 5: Second step visualization in the LSTM
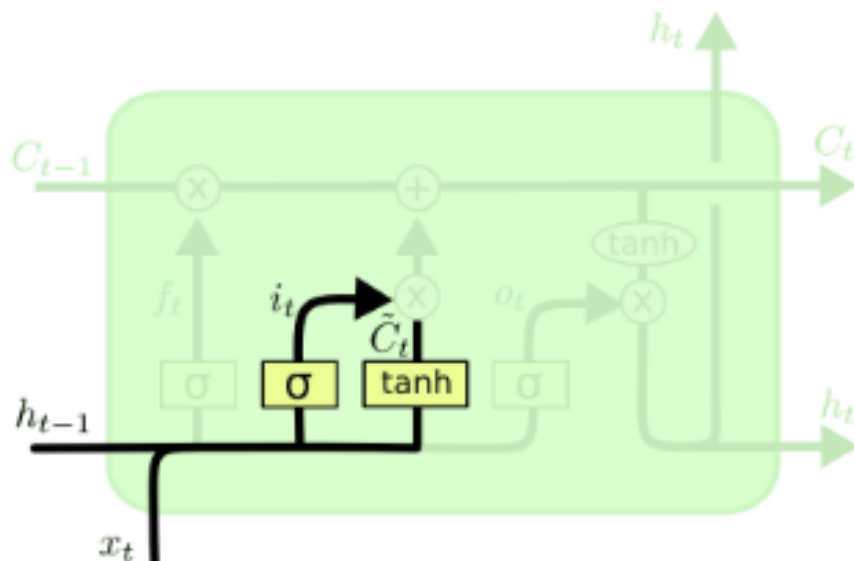**Source:** https://colah.github.io/posts/2015-08-Understanding-LSTMs/

The next step it's then update the cell state of the memory cells. To do that, we multiply the old state C_{t-1} by f_t and forget the things we decided to forget earlier. Then we include $i_t * \hat{C}_t$. These are the new candidate values, scaled in accordance with the value we decided to update for each state. This is represent by equation 4.

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{4}$$
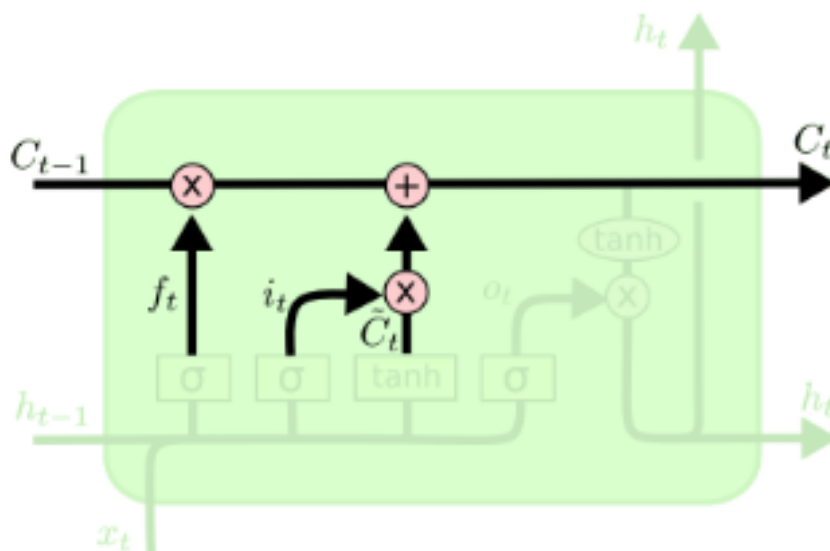
Figure 6: Third step visualization in the LSTM
Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

The last step is to decide what will be consider as output. The output gate regulates how much of the current state of the cell is rejected. The first thing to do is to run a sigmoid layer which will determine the output information.Then secondly, we put the cell in tanh (so that the values are between -1 and 1) and we multiply it by the output of the sigmoid layer to be sure that we will only output the parts we decided to.

$$O_t = \sigma(W_\sigma.[h_{t-1}, x_t] + b_o) \tag{5}$$

The final output value of the cell is formulate as:

$$h_t = O_t * tanh(C_t) \tag{6}$$

We can see in figure 2.2 how it's structured in the model.

Figure 7: Final step visualization in the LSTM
Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

3 Commodity price forecasting and the impact of Covid 19

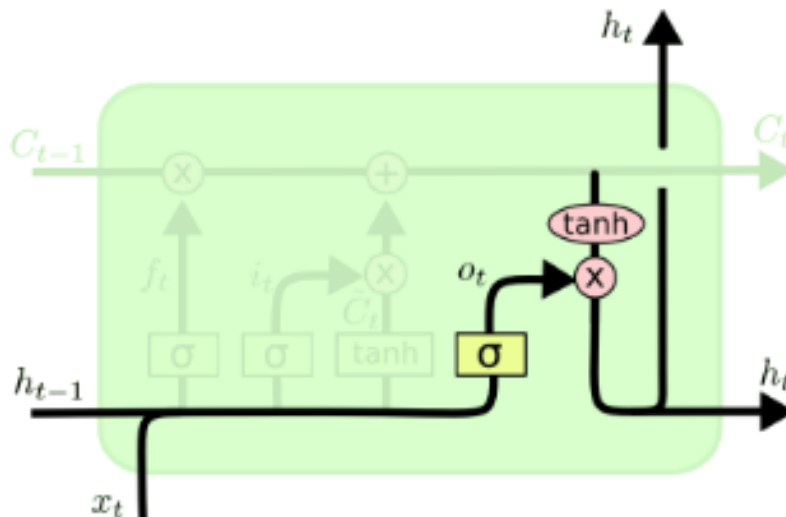Here we turn our attention to exploring the impact of Covid-19 on commodity markets. To this end, we employ tools from the fast growing field of Deep Learning, particularly the LSTM model discussed in section 2, to predict daily prices for several trading commodities such as Brent, Silver, Crude Oil and Wheat. These commodities are of great importance to several African economies and have, in light of the pandemic, left many speculators and exporting countries with many unanswered questions as regards their resistance to the pandemic. Following the price predictions in the early halves of the chapter, we turn our focus to gathering statistical evidence for the impact of the corona virus on the prices of these commodities. As a matter of fact, in section 4, we leverage the strengths of the well known Granger causality test to investigate and show evidence of the claimed impact of the virus on these commodities.

3.1 Price forecasting with the LSTM model.

As a matter of fact, with its ability to remember both long term and short term values, the LSTM models have proved very rewarding for the treatment of financial time series, thereby becoming the preferred Deep Learning tool for time series analysis. Furthermore, their ability to handle very noisy data and the their independence from the linearity assumption makes them go to models for analysis of commodity prices. In what follows, we construct LSTM models for predicting the prices of Brent, Crude Oil, Silver and Wheat. Our data is sourced from the Boursorama database and contains per minute prices for the four commodities of interest through out the month of April 2020. Because this data is recorded per minute, it's volume lends itself perfectly to the demands of the LSTM model which requires large volumes of data for model training. Furthermore, these records taken from the month of April sit at the heart of the pandemic period the world over and therefore paint a perfect picture of the markets and concerned world economies. Moreover, the data set is void of missing values and which lays to rest all questions regarding it's usability in our studies and also any questions arising from our findings. In addition, because this data is composed of opening prices, closing prices, high prices, low prices and the volume of transactions for the day, it present representation of daily trading information for each commodity.

Our analysis begins with summaries and visual explorations of the acquired data. To this end, we visualize the trend series of the closing prices of each commodity and take them as our goal series for testing the model's prediction precision. Indeed, looking at table 3.1, we notice that the average closing price for crude oil is approximately 15.64. This is a far drop from the average price of 60.40 recorded in November of 2019, prior to the outbreak of the virus. This indicates that the prices have dropped to more than half what they were just a few months earlier, a trend not peculiar to Crude oil alone, as all other three commodities show similar behavior when compared to their November prices. Furthermore, the love values of the standard deviation recorded show

	Mean	Std	Min	25%	50%	75%	Max
Crude Oil	15.64	4.11	0.01	12.55	15.05	19.8	22.47
Brent	24.9	3.92	15.99	20.7	27.01	28.13	28.96
Silver	14.96	0.12	14.75	14.86	14.97	15.08	15.3
Wheat	5.41	0.08	5.24	5.34	5.43	5.46	5.61

Table 1: Summary statistics of closing prices for all four commodities.
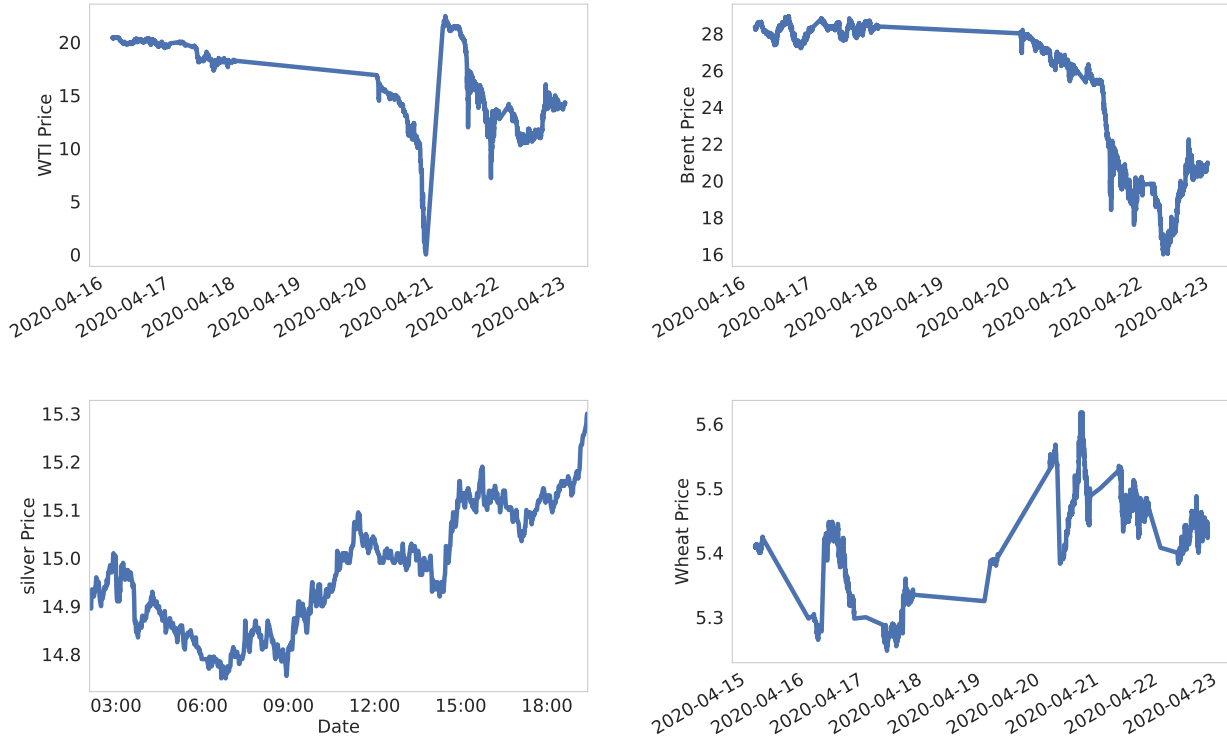

Figure 8: Trends of WTI price, Brent, Silver and Wheat price

that there has been very little variability in the prices of these commodities. Moreover, the relatively small differences between the min and max prices attained through the month support the claim that the prices are significantly low and the market must therefore be suffering from external influences to which these commodities in our study are not resistant.

Now, looking at the trend of the closing crude oil(WTI crude oil) price with the aid of a line graph shown in figure 3.1,we see that from April 16 to April 20, prices fluctuated around 16 and 20 USD. Then, on April 21, there was a sharp drop, followed by a sharp rise the next day. This increase will be followed by another large drop and small variations until the 23th of April. Again, these we claim, pending a causality check, were influenced by news events related to the corona virus coping mechanisms. First with the news of producers running out of storage for crude oil following lock downs around which lead to the fall in prices of futures contracts for May delivery and consequently the sharp drop in the spot prices depicted in figure 3.1. This news however was followed by news of easing of lock downs around most states in the US and other parts of the world which gave the market hope that drivers will soon return to the streets and consequently the sharp rise in spot prices visible in the trend.

Also, turning to Brent and looking at figure 3.1, we observe that the price was just decreasing from the 4th of April till the 22th of April. We also observe a small increase on the 23rd yet the Brent prices still remain very low. We yet again, prior to finding statistical evidence, claim on the basis of the fact that Brent is the most used form of crude oil in Europe that, this observed falling trend in the prices is backed by the fact that most of Europe saw sharp rises in the spread of the virus from the late half of March through early April and up to the late days of April there was very little positive news reaching the markets from these parts. The later days however saw some European cities begin to ease lock downs and this saw a gentle increase in the price of Brent.

Now looking at the closing prices for Silver, we see from figure 3.1 that the prices fluctuate heavily, but interesting enough, from the 22nd of April they tend to increase over the time. This is very interesting resistance demonstrated by this commodity. This will also suggest that, the inability of the other precious commodity to resist the spread of the pandemic influenced investors to move towards this commodity which appears to be

more resistant. It therefore will be very interesting to check if there is any causality between price movements here are the spread of the pandemic, to provide statistical evidence for the claim made here.

Another commodity in our study is wheat, which is one the most important food crops for civilizations around the world. For Africa, wheat accounts for 40 per cent of wheat production and 1.4 per cent at the global level (FAO, 2017). The price of wheat is generally influenced by factors such as climate, yields, imports and the price of oil. Given that we are now witnessing a sharp decline in oil prices, it is clear that wheat prices will not remain indifferent. From figure 3.1 we see that the price of wheat has been relatively low for the early half of the month of April and thereafter experience an admirable increase in the second half of the month. Now, without ignoring the need to provide statistical evidence of causality, we note that wheat is a recreational product, in the sense that it is most consumed by the hospitality industry, which was brought to its knees by the strict lock downs the world over. Knowing that the lock downs eased toward the end of the month and observing the rising (however volatile) trend in the price of wheat, it will therefore be a very interesting probe, to find out if the prices are being driven by the coping mechanisms for the pandemic.

Having used April's records to gain an understanding the dynamics of the commodity market, we now turn to forecasting these future prices in an effort to better prepare decision makers for what is to come. Commodity price prediction has always proved to be an intractable task due to the intrinsic complexity of commodity market mechanisms. In the sequel, we provide details on how we achieve the predicted value and evaluate the performance of our LSTM model. The prediction procedure follows three main steps. The first of these steps combines task such as scaling the data,train the model and updates of the model parameters. The *MinMax* function will allows us to scale the data and obtain values between 0 and 1. For the training, split the data in two halves with an 8:2 ratio and thus us 80% of the data for training the model and 20% for testing. The second step is the validation phase of the process. We use it to adjust the hyperparameters and obtain an optimal model setting. Following the validation, the penultimate step, model testing, uses the half of the data previously uninvolved in the first and second steps. In this step, we use the optimized model to predict the commodity prices. We use the data from the closing price to train the model and predict the price for the next 60 minutes.

The second piece of the prediction procedure is based on well founded mathematical error measurement tools which inform us about the quality of the model. Previous articles have selected several indicators to measure the quality of the model's forecasts of financial market developments [16, 17]. The most classic are the Mean Absolute Percentage Error(MAPE), R ,Theil U and the Root Mean Squared Error(RMSE). MAPE is used to measure the size of the error. It is derived as the relative average of the error. As a precision measure, the MAPE is affected by four problems: equal errors above the true value result in a larger APE; large percentage errors occur when the value of the original series is small; outliers can distort comparisons in empirical studies; and MAPE cannot be directly compared with naive models such as random walk. R is a measure of the linear correlation between two variables. Theil U is a relative measure of the difference between two variables. It squares the deviations to give more weight to large errors and to exaggerate errors. In this article, we follow the approach of [17] and choose the RMSE to measure the predictive accuracy of our model. The Root Mean Squared Error (RMSE) as the evaluation metrics for the absolute value prediction is given by:

$$RMSE = \sqrt{1/N \sum_{i=1}^{N} (y_i^* - y_i)^2} \tag{7}$$

where N is the total number of value price in the closing price data and y_i^* and y_i are respectively the predicted and actual price.First, we predict the WTI oil price. The hyperparameters in our LSTM networks were set as: the number of units in the deep learning layers equaled to 128, batch size equaled to 1, the loss was the mean squared error and random seed equaled 0. We used Adam optimizer with learning rate = 0.01, the bias= 0.1. We also set the number of simulations equaled to 10 with a checkpoint after each epoch (epoch=300) and only saved the best model for prediction. The figure 3.1 below present the result of the forecast for the 4 commodities with the ten simulations. Looking at this graph, it is difficult to see which simulation gave the best accuracy. In fact, the average accuracy of the ten simulations for WTI crude oil, Brent, Silver and Wheat was respectively 76.92, 85.9, 99.6,99.24 . Comparing the different accuracies with each other, we plot the original trend of the price and te best prediction of these deferent commodities in figure 3.1 below. The part in blue after the true trend represent the future value we predicted for the next 60 minutes. The accuracy for WTI crude oil, Brent, Silver and Wheat is respectively 98.2, 88.2, 98.2, 97.8 . These results are quite pretty looking at these deferent accuracies. Unless this impressive results, there is still some remarks we can do looking at each plot. For the first plot, that is the WTI crude oil, we can see that there is not too much deference between the true trend(in black) and the predict trend(in blue) of price.But our model failed to forecast the sharp drop in the price happened on 21 of April. The second plot in figure 3.1 is the Brent oil price forecasting. The obtained accuracy of 88.2 seems pretty good, but looking at the chart, it's only the prices of the last two days of data that were well predicted. For the silver price forecasting the accuracy of 98.2% allows us to conclude that our model performed very well as we can see on the plot. The last plot in figure 3.1 concerned the forecasting for the
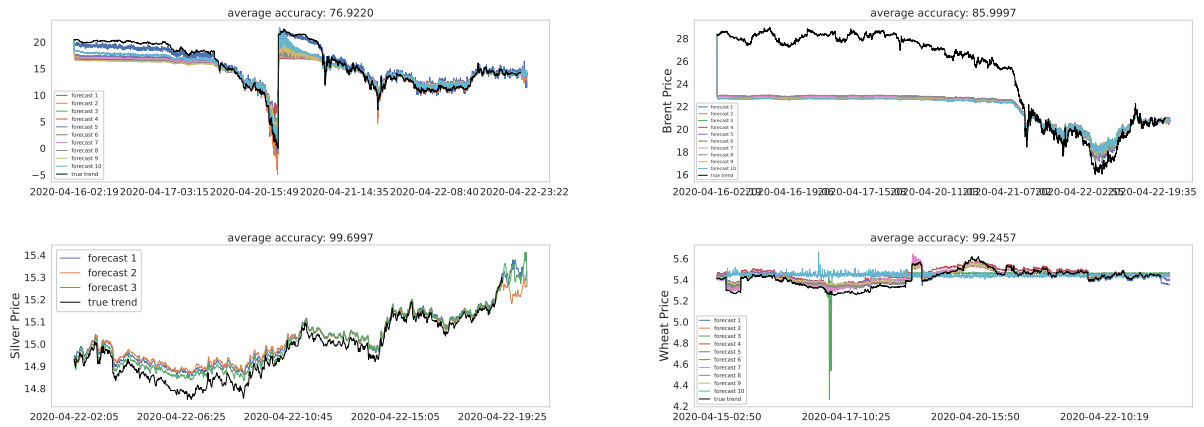
Figure 9: Ten simulations forecast of WTI price, Brent, Silver and Wheat price
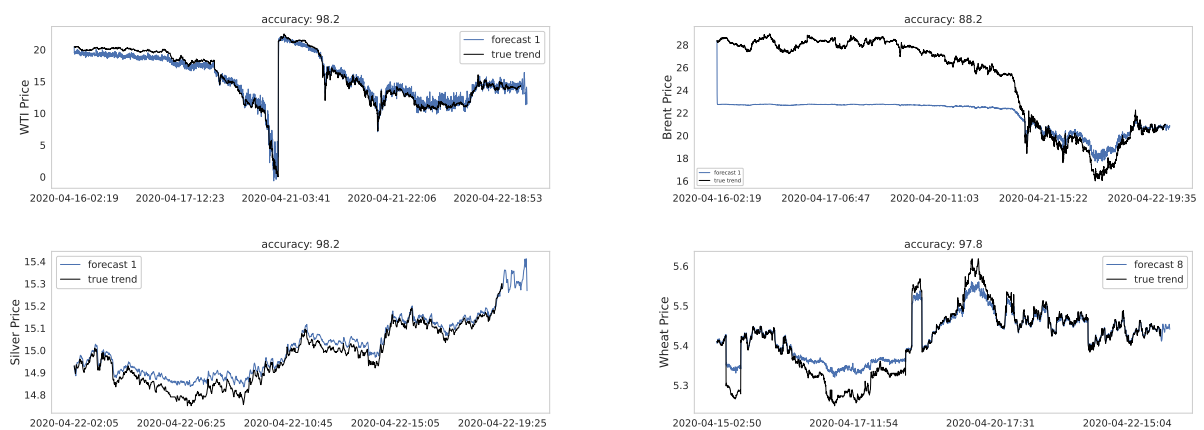

Figure 10: Original price and best prediction prices of WTI oil, Brent, Silver and Wheat

wheat price. The best prediction for the wheat price is given with an accuracy 97.8% . Although this forecast gave a good accuracy , it still fail to give a good predictions for peaks in the wheat prices.

4   Impact of COVID-19 on commodity prices

The commodity market is a meeting ground for most of the worlds economies, bringing together exporting and importing countries with the shared goal of making profits. This market thrives on speculation and trade is heavily dependent on the activities of the individuals in the trading countries. Consequently, it is expected that drastic shifts in the daily routines of these countries will translate into significant movements in the market. However, due to the size of the market, some products may experience direct impacts of these routine shifts while others simple suffer secondary effects. That is, the shifts may cause direct drops or hikes in prices of some commodities and this shift in prices for this impacted set of commodities may push demand in other commodities, thereby increasing the prices of those commodities as well. The primary and secondary phenomena are captured in the statistical concepts of causality and correlation, with the later referring to the secondary effect as opposed to the former which describes or affirms the primary effect.

In this section, we examine the causal relation between the commodities in our study and the corona virus pandemic which has dramatically change routines in countries the world over. The findings in this section shall either refute or support the claims or projections from our analysis earlier. For this purpose we use the COVID-19 dataset [1] which consists of the date, two countries (France and Cameroon), confirmed cases, and total deaths. We combined these data with daily closing price data for 3 commodities [2], Brent, Crude oil WTI and silver. The data starts from 01-01-2020 to 24-04-2020. We will study the correlation and causality of these different Covid 19 information with commodity prices.

4.1   Correlation between Covid Spread France-Cameroon with commodities prices

To measure the degree of association between the movements of two variables, the most common correlation coefficient is the Pearson correlation coefficient. This one allows to measure the linear relationship between two variables. However, in a non-linear relationship, this correlation coefficient is not always an appropriate measure of dependence. In equation 8 we give the correlation coefficient formula.

$$\rho = \frac{cov(X,Y)}{\sigma X \sigma Y} \tag{8}$$

Where ρ is correlation coefficient, cov(X,Y) is the the covariance of X and Y, σX, σY is respectively the standard deviation of X and Y. The range of values of ρ is between -1.0 and 1.0. A correlation of -1.0 means a perfect negative correlation, and a correlation of 1.0 means a perfect positive correlation. A zero value indicates that there is no evidence of a linear relationship between the two variables. Figure 4.1 shows the correlation matrix for these commodities and the corona virus death toll.

Looking at this figure, we notice that Silver price is negative correlated with the number of confirmed cases and the total number of death, but this correlation is not significant. We also see that there is a negative correlation between the prices of WTI crude oil and the number of confirmed case. The correlation coefficient is -0.7 for Cameroon and -0.8 for France. A negative correlation between the Brent prices and the number of confirmed cases is also seen, -0.7 for Cameroon and -0.8 for France. Those two commodities are also negative correlated with the total number of death, -0.6 for Cameroon and -0.7 for France.This strong negative correlation signifies that as the number of cases or the number of death increase, the prices of WTI oil and Brent decrease.

These findings demonstrate that outside Silver, both Crude Oil and Brent have strong linear relationships with the death toll for corona virus. That is an increase in deaths due to corona virus coincides with falling prices for Crude oil and Brent. We can clearly see pictorial representations of this numbers and what they suggest depicted in the line graphs in figure 4.1. It is worthy of note that this is not evidence of causality i.e this is not, if at all, sufficient evidence to conclude that the pandemic is the reason for the down pull in the commodity prices.

In addition, a negative correlation does not automatically mean that the securities must be avoided. The correlation coefficient can help investors diversify their portfolio by considering a combination of investments with negative or low correlation to the commodity market to reduce the risk of volatility in their portfolio.

4.2   Causalities between Covid spread and different commodities: The Granger causality

---

[1] https://github.com/datasets/covid-19, https://ourworldindata.org/covid-deaths?country=CMR+FRA
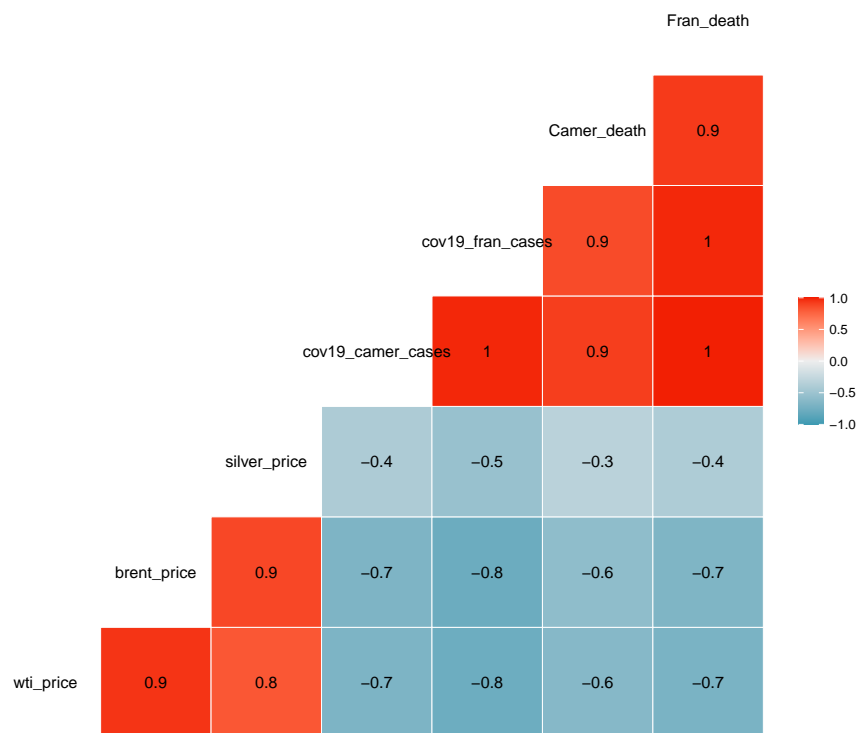[2] https://www.quandl.com/data

Figure 11: Correlation matrix of the Covid 19 dataset and the commodity price dataset
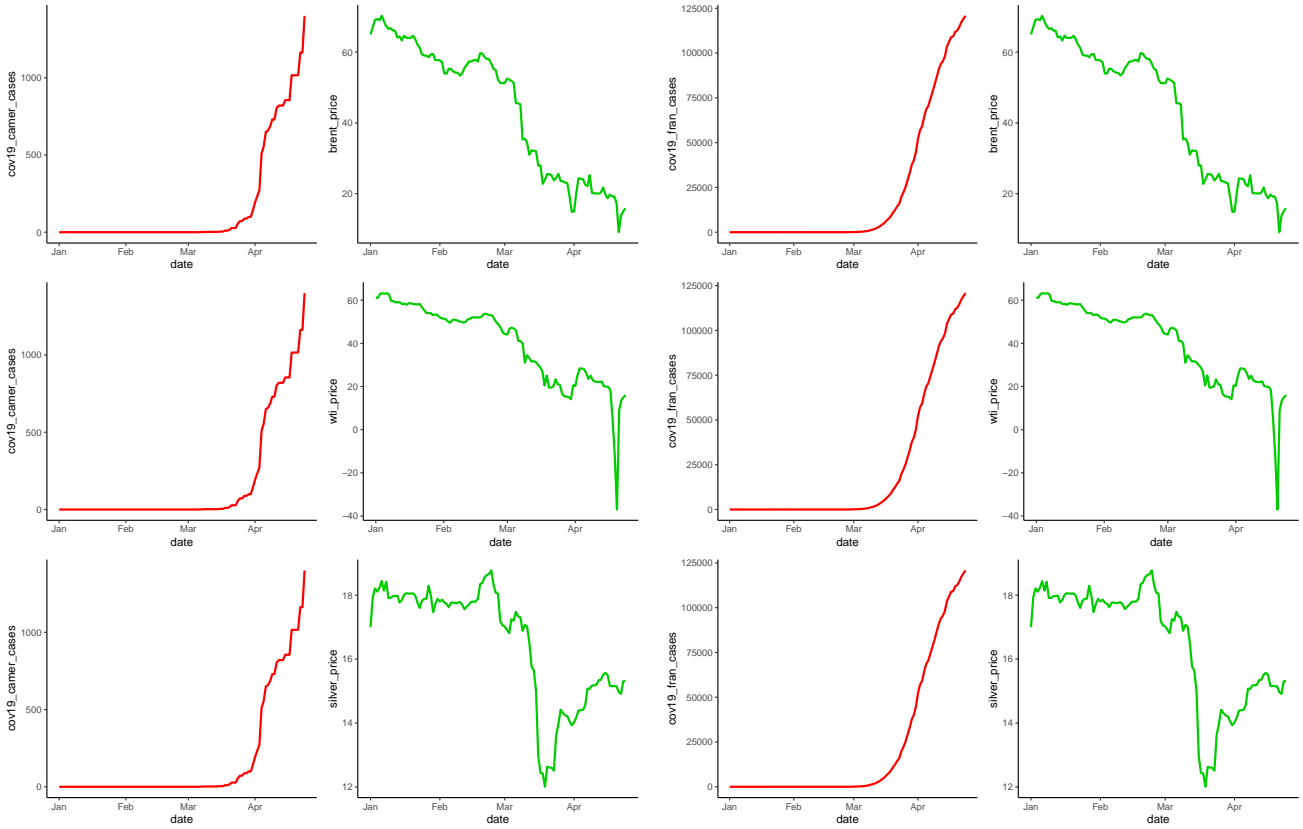
Figure 12: Numbers of cases (Cameroon and France) and commodity prices

Correlation coefficients are not sufficient tools to measure the impact that one variable has on another. It would therefore be both interesting and necessary to see whether there is a causal relationship between the evolution of covid-19 and prices on the commodity market. We therefore turn to the well known Granger causality test to check for the present or lack thereof a causal relationship between the observed price dynamics of these commodities and the death toll of the corona virus.

Granger causality was developed by Granger (1969) [18] to measure the causal influence of one time series on another . Thus, an X time series is said to Granger cause Y if it can be demonstrated that these X values give statistically significant information for future values of Y through a series of t-tests and F-tests on lagged X values. So considering two times series y and x, we are going to test the null hypothesis that x does not Granger-cause y by first finds the proper lagged values of y to include in a univariate autoregression of y:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + ... + a_m y_{t-m} + \epsilon_t \tag{9}$$

next, we set the autoregression of y augmented with the lagged values of x:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + ... + a_m y_{t-m} + b_1 x_{t-1} + ... + b_m x_{t-m} + \epsilon_t \tag{10}$$

The null hypothesis that x does not Granger-cause y is not rejected if and only if no lagged values of x are retained in the regression. The hypothesis can be state as this:

$$H_0 : b_i = 0 \text{ for each i of the element [1,m]}$$
$$H_1 : b_i \neq 0 \text{ for at least 1 i of the element [1,m]}$$

In our analysis, we are trying to determine whether we can say that statistically Covid 19 provides more information about future values of commodity prices than past values of commodity prices alone. This simply means that the two variables are related by some phenomenon. So for this case we only check the partial Granger Causality that is the impact of Covid spread on the commodity price and not the reverse.

In table 4.2 we present the result concerning the relationship between the number of confirmed cases in Cameroon and the commodity prices.

For the first commodity, WTI oil, we have highly significant p-value, from Lag 4 to Lag 6. For Brent, the lag 5 and 6 is significant at 5% and silver is not significant for any of the lagged values of the number of confirmed cases. This table therefore allows us to conclude that the number of cases has an impact only on the prices of WTI and Brent oil. We also analyze the relationship between the number of confirmed cases in France and

Causal direction	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
WTI< −Cases	0.165	0.155	0.032**	0.0003***	0.00***	0.00***
Brent< − Cases	0.745	0.532	0.467	0.109	0.029**	0.024**
Silver< − Cases	0.546	0.556	0.783	0.859	0.892	0.912

Table 2: Relationship between confirmed cases in Cameroon and commodity prices.   Level of significance: *:10%   **:5%   ***:1%

commodity prices. The results are presented in the table 4.2 below. We come to the same conclusion as for Cameroon, namely that the number of cases has an impact only on Brent and WTI oil prices.

Causal direction	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
WTI< −Cases	0.107	0.197	0.269	0.006***	0.015**	0.010**
Brent< −Cases	0.567	0.096*	0.175	0.167	0.039**	0.065*
Silver< − Cases	0.549	0.957	0.986	0.909	0.909	0.978

Table 3: Relationship between confirmed cases in France and commodity prices.   Level of significance: *:10%   **:5%   ***:1%

We also analysed whether the total number of deaths in both countries has an impact on commodity prices. The results are presented in the tables 4.2 and 4.2.,

Causal direction	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
WTI< −Death	0.123	0.00***	0.00***	0.00***	0.00***	0.00***
Brent< − Death	0.787	0.00***	0.00***	0.00***	0.00***	0.00***
Silver< − Death	0.74	0.787	0.831	0.733	0.859	0.917

Table 4: Relationship between confirmed death in Cameroon and commodity prices.   Level of significance: *:10%   **:5%   ***:1%

Table 4.2 presents the results implemented for Cameroon. It can be seen that the total number of deaths is very significant with Brent and WTI prices. On the price of silver, there is still no impact of Covid 19. On table 4.2 below, we show the result for France. The number of death have an impact on the WTI oil and Brent for the Lag 6, and even here the silver price is still not seems to be affected by the number of death.

5   Forecasting of the Corona virus spread

We saw in the previous section that there was a correlation and causality between cumulative confirmed cases of corona virus and commodity price variability. Indeed, the "causality" or "correlation" analysis between the Covid-19 time series (number of deaths per day) and the change in commodity prices was done in order to anticipate the evolution of prices based on a prediction of the evolution of Covid-19. This section therefore focuses on the prediction of the spread of the coronavirus.

There are several techniques to predict the spread of the corona virus. In this document, we use two approaches: the ARIMA-WBF model and the LSTM model. We explained already the LSTM model in section 2.

We follow the methodology used by [19] for the ARIMA-WBF model. It is a hybrid approach based on an integrated autoregressive moving average model and a wavelet-based prediction model. With this model, we want to generate short-term forecasts (ten days in advance) of the number of confirmed daily cases for Cameroon and France. The traditional ARIMA model is well known for its performance in short-term prediction but this forecast is mainly used for linear time series forecasting. With the recent literature demonstrating the excellent performance of the wavelet transformation-based prediction model for modeling non-stationary time series data [20], the combination of the two models could accurately model time series data with complex autocorrelation structures such as those of COVID-19 and reduce the bias and variances of the models' prediction error.

To start the implementation, we first define a ARIMA model as equation 11 shows:

$$y_t = \theta_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q} \tag{11}$$

With y_t the actual number of cases at time t, ϵ_t the random error at time t. The ϕ_i and θj are the ARIMA model coefficients. To build ARIMA model, we follow three main steps:

- Model identification:Consist on the stationarity of the serie.

Causal direction	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
WTI− >Fran_Death	0.183	0.155	0.192	0.335	0.613	0.018**
Brent− > Fran_Death	0.807	0.906	0.882	0.935	968	0.04**
Silver− > Fran_Death	0.691	0.984	0.856	0.917	0.964	0.977

Table 5: Relationship between confirmed Death in France and commodity prices.   Level of significance: *:10%   **:5%   ***:1%

- Parameter estimation: Choose p and q parameters using the autocorrelation function (ACF) and the partial autocorrelation function (PACF)

- Model diagnostics checking: Check the best fitted forecasting model using Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)

After define the ARIMA model the next step is to build the wavelet based forecasting model. The wavelet analysis have the ability to reveal information within the signals in both the time and scale (frequency) domains [19].This model more suited with non stationary data such as epidemic data and climate data. the wavelet model decompose a signal $X = (X_1, ..., X_N)$ to

$$X_t = c_{j,t} + \sum_{j=1}^{J} w_{j,t} \tag{12}$$

where c_j is the smooth version of original signal X and w_j is the detail of X at scale 2^{-j}. For the number of decomposition levels $W_L=$ int[log(n)] is used , where n is the time-series length [3].
The hybrid model is now the combination of those two models using a simply algorithm. The algorithm for implementation of the ARIMA-WBF model is present below.

Proposed Hybrid ARIMA-WBF Model [19]

1) Given a time series of length n, input the in-sample (training) COVID-19 daily cases data.

2) Determine the best ARIMA(p, d, q) model using the in-sample (training) data.

   (i) ARIMA parameters p, d, and q values are selected using the procedures described above.

   (ii) Obtain the predictions using the selected ARIMA(p, d, q) model for the in-sample data and generate required number of out-of-sample forecasts.

   (iii) Obtain the residual series (ϵ_t) by subtracting ARIMA predicted values from the original training series.

3) Train the residual series (ϵ_t) generated by ARIMA by the WBF model.

   (i) Select the number of decomposition level using the formulae $W_L=$ int[log(n)] and boundary is chosen to be "periodic".

   (ii) Obtain in-sample predictions (ϵ_t) using the WBF model and generate required number of out-of-sample forecasts.

4) Final predictions (Y_t^*) are obtained by combining then ARIMA predictions with WBF predictions (ϵ_t^*) for both the training series as well as the out-of-sample forecasts.

Two times series was used: the Covid-19 data for France and the covid-19 data for Cameroon. Data range from January to June 2020 for France and from March to June 2020 for Cameroon. The classical ARIMA was the first thing we implement using "forecast" package in R software. Using AIC and BIC criteria, the best fitted ARIMA models was ARIMA(3,2,1) for Cameroon and ARIMA(3,2,0) for France. We implemented the WBF using "Wavelet Arima" package in R software. We present below in figure 5 the prediction using the ARIMA-WBF and LSTM for the prediction of the number of cases in France and Cameroon.

At it can be se on this figures, the two model have performed well with a little difference. Using Arima-WBF model, we obtained an RMSE of 901.75 for France and 128.99 for Cameroon. Using the LSTM model,we predicted the covid 19 cases with an accuracy of 97.45% for France and 92.13% for Cameroon. The result for the the 10 days forecasting number of cases for Cameroon and France are present below.
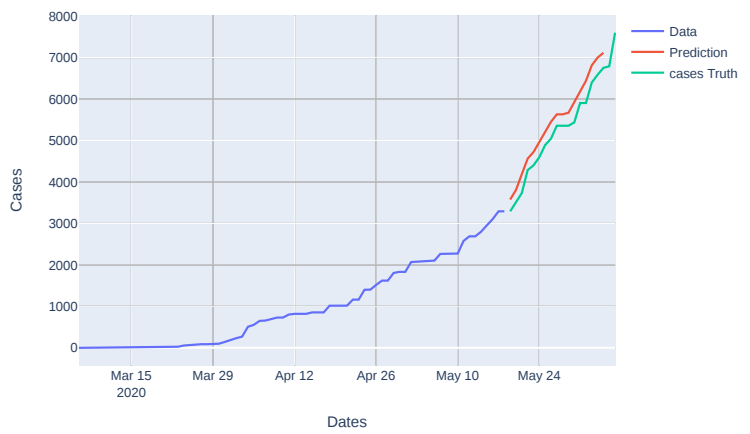
---
[3]More details in Starck et al (1998)
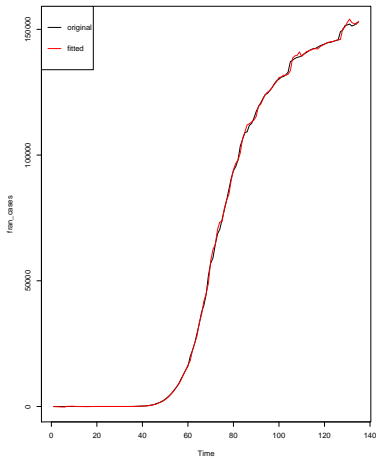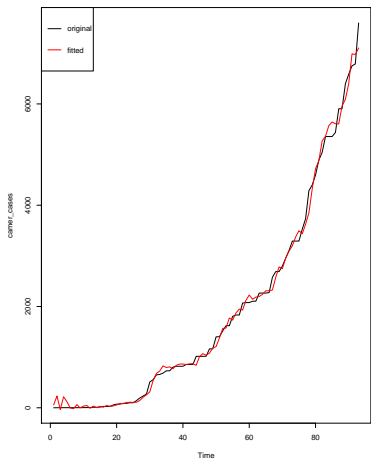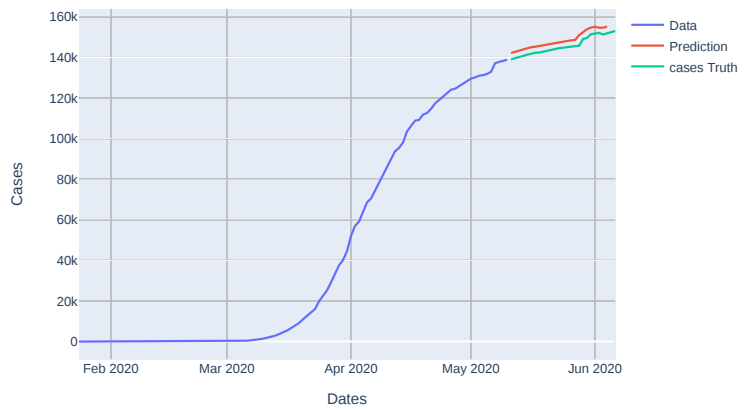
Figure 13: Prediction of Covid-19 cases using ARIMA-WBF(white framework) and LSTM model(Blue framework)

	Dates	Fran_cases	forecast lstm	forecast ARWAV
0	2020-06-06	153055	153055.000000	153055.0
1	2020-06-07	153634	156546.640625	153620.4
2	2020-06-08	153977	158408.437500	154091.0
3	2020-06-09	154188	161289.031250	154468.6
4	2020-06-10	154591	163634.656250	154773.6
5	2020-06-11	155136	166358.359375	155021.9
6	2020-06-12	155561	168925.218750	155225.9
7	2020-06-13	156287	171650.250000	155395.3
8	2020-06-14	156813	174349.593750	155537.8
9	2020-06-15	157220	177132.234375	155659.1
10	2020-06-16	157372	179935.296875	155763.9

Figure 14: 10 days ahead predicted values of Covid-19 cases in France using Arima-WBF and LSTM

	Dates	Camer_cases	forecast lstm	forecast ARWAV
0	2020-06-06	7599	7599.000000	7599.000
1	2020-06-07	7599	7608.708984	7703.082
2	2020-06-08	7908	7959.626465	8044.375
3	2020-06-09	8312	8181.088867	8139.756
4	2020-06-10	8681	8467.666992	8625.458
5	2020-06-11	8929	8739.450195	8840.289
6	2020-06-12	8929	9029.797852	9194.379
7	2020-06-13	9196	9325.341797	9384.143
8	2020-06-14	9572	9632.008789	9664.291
9	2020-06-15	9864	9947.776367	9886.747
10	2020-06-16	9864	10273.917969	10212.188

Figure 15: 10 days ahead predicted values of Covid-19 cases in Cameroon using Arima-WBF and LSTM

6 Discussion and Conclusion

The purpose of this paper was to use the LSTM model for forecasting commodity prices and analyse the impact of coronavirus on the commodity price variability. To achieved this, four main commodities were examined: the West Texas Intermediate (WTI) crude oil, the Brent oil, Silver and Wheat. We chose this commodities because they are among the main commodities exported by African countries and they are also very volatile. Volatility plays an important role in intraday price forecasting. As [21] said, intraday price volatility reflects the underlying elements that determine market prices, such as investor speculation, market news, mergers, product launches, companies that declare profits or losses.

The procedure of construction of our forecasting model passes through three main steps. The first step consisted on scaling the data, train the model and update the model parameters. The MinMax function allowed us to scale the data and obtain values between 0 and 1. For the training, we used the standard, 80% for training data and 20% for testing. The second step was the validation part. We used it to adjust the hyperparameters and obtained an optimal model setting. The last step was the test part, where we used the optimal model to predict the data. We used the data from the closing price to train the model and predict the price 60 steps ahead. As a result, the LSTM model predict the commodity prices with great accuracy.

The second part of our work consisted in analysing the impact of the Corona virus on the commodity price fluctuations and forecasting of coronavirus spread. In this regard, we analysed the correlation between the Covid-19 and the commodity prices. Our results revealed a strong negative relationship between the number of cases and the Brent and WTI crude oil prices. They were also a strong negative correlations between the number of deaths and the Brent and WTI crude oil prices. Silver registered a low correlation coefficient. After the correlation, we study the causality between the Corona virus and the variability of commodity price. The Granger causality has been used to study this causality, and we found that the number of cases and number of deaths have an impact in the variability of commodity price. Based on our results we recommend investors and countries involved in such markets to concentrate their investment on resilient commodities such as silver which was actually resilient to the corona virus.

As future work, this study can be extended with the application of reinforcement learning to take in account information related to The Organization of the Petroleum Exporting Countries (OPEC), that often cause a drop in a price.

References

[1] Ben Moews and Gbenga Ibikunle. Predictive intraday correlations in stable and volatile market environments: Evidence from deep learning. *Physica A: Statistical Mechanics and its Applications*, page 124392, 2020.

[2] Angus Deaton. Commodity prices and growth in africa. *Journal of economic Perspectives*, 13(3):23–40, 1999.

[3] Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 106–112. IEEE, 2014.

[4] Muhammad Aamir and Ani Shabri. Improving crude oil price forecasting accuracy via decomposition and ensemble model by reconstructing the stochastic and deterministic influences. *Advanced Science Letters*, 24(6):4337–4342, 2018.

[5] Halbert White. Economic prediction using neural networks: The case of ibm daily stock returns. 1988.

[6] Nowrouz Kohzadi, Milton S Boyd, Bahman Kermanshahi, and Iebeling Kaastra. A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, 10(2):169–181, 1996.

[7] RK Dase and DD Pawar. Application of artificial neural network for stock market predictions: A review of literature. *International Journal of Machine Intelligence*, 2(2):14–17, 2010.

[8] Qing Cao, Karyl B Leggio, and Marc J Schniederjans. A comparison between fama and french's model and artificial neural networks in predicting the chinese stock market. *Computers & Operations Research*, 32(10):2499–2512, 2005.

[9] Mohamed M Mostafa. Forecasting stock exchange movements using neural networks: Empirical evidence from kuwait. *Expert Systems with Applications*, 37(9):6302–6309, 2010.

[10] Mingyue Qiu and Yu Song. Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5), 2016.

[11] Sheng Chen and Hongxiang He. Stock prediction using convolutional neural network. In *IOP Conference Series: Materials Science and Engineering*, volume 435, page 012026. IOP Publishing, 2018.

[12] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen. Financial time-series data analysis using deep convolutional neural networks. In *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, pages 87–92. IEEE, 2016.

[13] Sima Siami-Namini and Akbar Siami Namin. Forecasting economics and financial time series: Arima vs. lstm. *arXiv preprint arXiv:1803.06386*, 2018.

[14] SI Lee and SJ Yoo. A deep efficient frontier method for optimal investments. *Department of Computer Engineering, Sejong University, Seoul*, 5006, 2017.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Lstm can solve hard long time lag problems. In *Advances in neural information processing systems*, pages 473–479, 1997.

[16] Wei Bao, Jun Yue, and Yulei Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, 12(7), 2017.

[17] Nhi NY Vo, Xuezhong He, Shaowu Liu, and Guandong Xu. Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, 124:113097, 2019.

[18] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[19] Tanujit Chakraborty and Indrajit Ghosh. Real-time forecasts and risk assessment of novel coronavirus (covid-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals*, page 109850, 2020.

[20] Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge university press, 2000.

[21] Aman Thakur. Efficacy of lstms for forecasting market indices: Comparative benchmarking with svm regression.