Meng Sui, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 10, Issue 8, (Series-II) August 2020, pp. 27-39

RESEARCH ARTICLE OPEN ACCESS

Implementing the Decomposition Model in Gold Price Volatility Analyzing, Forecasting and Trading

Meng Sui [1,2], Erick W. Rengifo [1,3], Eduardo Court [4]
1. Corresponding Author.
2. Department of Economics and the Center for International Policy Studies, Fordham University, 441 E. Fordham Rd., Dealy Hall E513, Bronx, NY 10458.
3. Department of Economics and the Center for International Policy Studies, Fordham University, 441 E. Fordham Rd., Dealy Hall E513, Bronx, NY 10458.
4. Department of Finance, ESAN University, Alonso de Molina 1652, Monterrico, Surco, Lima, Peru.

ABSTRACT
Over the last several decades, investment demand for gold has significantlyincreased, playing a critical role inthe determination of gold's price. This study intends to extend the research on current price oscillationsby applying an econometric techniqueEnsemble Empirical Mode Decomposition (EEMD) which helps sorting out the short- and long-termeffectsof gold price changes. In so doing, highlight the ETFs contribution to the explanation of the recent price volatility.
In addition and based on the statistical criterion of MSE, RMSE, and SMAPEwe show that gold price forecasts obtained with the EEMD are more accurate. We show,using weekly and monthly data, that machine learning models (Multi Layer Perceptrons, Extreme Learning Machines) integrated with the EEMD provides a much better several-steps ahead forecasting results than using only the original data price; We further apply this model into a trading strategy with daily and intraday data and discuss the price forecasting problem of a multi-class (buy, hold or sell decisionscompared with the previous studies which define this issue as a simple binary-class – buy or sell) classification problem. We examine two algorithms, kernel SVM and Random Forest to simulate the EEMD model with 13 technical indicators. We show that both algorithms could generatesubstantially better returns than a simple buy - hold strategy.
***Keywords*:** Price forecasting,Price decomposition, Ensemble Empirical Mode Decomposition (EEMD), Machine Learning models, Trading strategy.
JEL Classification:*C30, C53, G17*

Date of Submission: 28-07-2020 Date of Acceptance: 12-08-2020

I. INTRODUCTION

Gold has been considered as a precious metal during the development of most human societies. As one of the most non-reactive metal, gold is benign in all natural and industrial environments. With relatively constant supply, gold has also been considered as a safe haven asset[1], especially duringeconomic recessionary periods. Nowadays,with continued global geopolitical dilemmas and macroeconomic uncertainty, gold is enjoying a renewed prominence as a financial asset, as a tool to hedge against currency devaluation, protect against inflation or deflation and,as a natural asset for portfolio diversification[2].

This study seeks to explainthe long-term trend in an environment characterized with an increased short-term price volatility by splitting gold's price in its commodity and financial properties. Our results show that the ETFs'activities have a strong and significant impact onthe short-term price of gold but not in the long-term.The second part of this paper showshowto implement the EEMD approach into forecasting and trading. We show that with this approach, theaccuracy of the price forecastingimproves, allowingus to achieve higher returns. Our study could be useful to the

[1]A weak (strong) safe haven asset is defined for the first time in Baur and Lucey (2010) paper. According to them, an asset can be seen as a weak (strong) safe haven asset if it is uncorrelated (positively correlated) with inflation rates in extreme adverse market situations.

[2]World Gold Council: https://www.gold.org/

portfolio manager and national government in terms ofrisk reduction (hedging) and investment decision making.

Unlike fiat currency or other assets, gold has an intrinsic value as a precious metaland can maintain its value throughout time. Gold, as a financial instrument, is used by individuals and firmsas a store of wealth, to preserve and pass on their wealth from one generation to another. As a commodity product, gold also has electrical conductivity properties with a high corrosion resistance rate (Tully and Lucey,2007). Given these properties, gold is widely used in modern technologies that include the fabrication of electronics, telecommunications, dentistry, and chemicals.

Seen as a traditional store of value, gold is traded primarily in dollars. The strength of dollar is found to be a strong short-term determinant of gold's price (Sjaastad, 2008; Tully and Lucey, 2007; Levin and Wright, 2006; Kaufman and Winters, 1989). The correlation between gold and US dollar strength is opposite for most of the time. When the value of the dollar falls (depreciates) against other currencies (from 1998 to 2008), people turntowards gold, which drives up its price. Gold price nearly tripled, reaching the $1,000 per ounce milestone in 2008 and hitting around the $1800 mark in 2012. After bottoming at $1050 per ounce in 2015, the price of gold rallied to $1365 per ounce in July 2016 and has been traded in a range between $1200 and $1500 per ounce thereafter during the last three years.

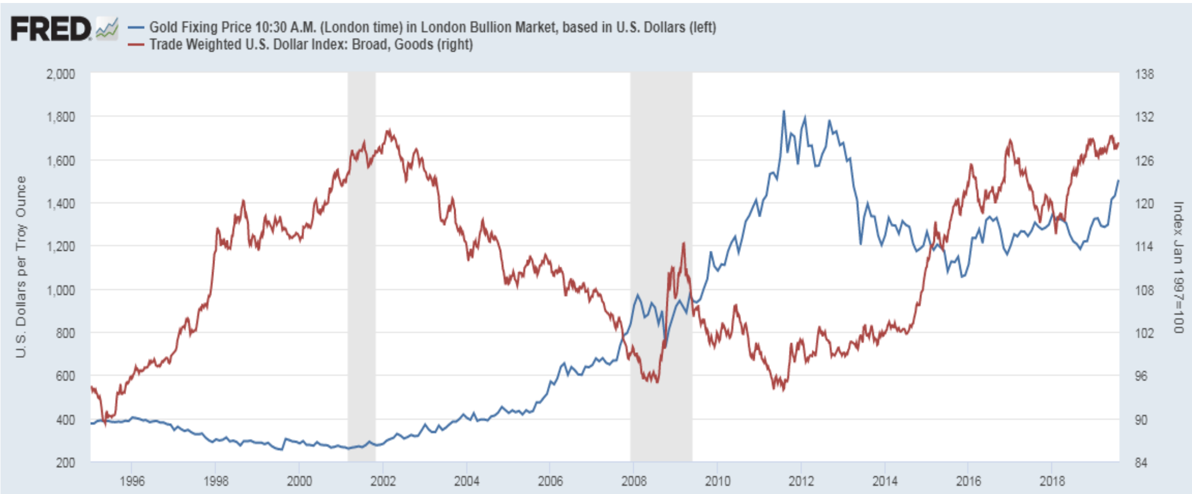

Figure 1: Gold price in dollar terms with daily data from 1/1/1995 to 8/9/2019. The blue line (left axis) represents the gold price in dollars, the red line (right axis) shows the trade weighted dollar index;Source: ICE Benchmark Administration Limited (IBA)

Historically, gold has served as both a legitimate hedge against inflation(Ghosh et al., 2004; Adrangi, Chatrath and Raffiee, 2003; Worthington and Pahlavani, 2007) and as an integral part of a diversified investment portfolio with negative correlation to the most of other equity assets.Baur and Lucey (2010) mentioned that gold, as an alternative investment, should be allocated to the portfoliosto improve the expected portfolio return and mitigateitsrisk.

Jewelry is gold's largest demand component. Based on Bloomberg data, jewelry accounts for more than half of the total gold's demand (commodity property). Investments for physical gold (bar and coin) and exchange-traded funds follow at around 30% (monetary property).Finally, net central-bank buying comes to around 10%.Gold holds its value not only during economic-financial uncertain times, but in times with high geopolitical uncertainty because people flock to gold when world tensions rise. This last property is of special interest in a world with continued global geopolitical and macroeconomic uncertainty, such as the one we are currently living on. In this sense, gold appeal as asafe haven asset and liquidity provider may shine bright and lead to an increasing investment demand.

The rest of the paper is organized as follows. In Section 2, we present the literature on gold pricing and the macroeconomic drivers of the price of gold. In Section 3, we providethe motivation for using the EEMD (Ensemble Empirical Mode Decomposition) model for this research project, describe the data, and use the decompositionmethod to analyze gold's price volatility and the price driverswithin different frequency ranges. The implication for using the EEMD model in price forecasting and daily tradingis presented in section 4and section 5. In the last section we present a summary of our main findings and some concluding remarks pertaining to future work.

II. LITERATURE REVIEW AND THE EEMD MODEL

There are numerous researchersworking on quantitative models to forecast the price of gold (Baur, Beckmann and Czudaj, 2016; Aye et al., 2015); Others look for fundamental and intrinsic value of gold based on different micro or macro foundations(Bertus& Stanhouse, 2001; Bialkowski et al., 2015, Lucey and O'Connor, 2013).Moreover,other authors are testing the dynamic relationship between exchange rate, crude oil, gold commodity, and gold mining stock prices (Shahbaz, Balcilar and Ozdemir 2017; Batten et al., 2017). At this point however, the explanation of changes in gold prices is mixed.Gold price contains not only changes in the intrinsic value of gold from the laws of demand and supply but alsoit captures the concurrent changes in global exchange rates,(exchange rate hedge, Capie et al. 2005) as well it captures the effects from diverse sources of risk (geopolitical or country risks, for example).

In several studies, the gold price is assumed to be in steady state and, based on this, several authors try to estimate the standard price level under corner interior solution[3]. Dowd, K. (1997) derived a representative agent model than the one used by Mcdermott (1987) or Dowd and Sampson (1993). However, these authorsobtainedan inverse Gibson paradox[4] conclusion based on different assumptions. Bordo and Ellson (1985) studied gold price determination using simulationsderived from theoretical results. However, all these studies offer no explanation about neither the way price changes in the transition path nor reveal any evidence about price level behavior outside the steady state.

In another research approach price bubble testing heavily based on the use of the Markov regime-switching models, the ADF test gives mixed resultsin terms ofthe gold price bubbles existence. Bialkowski et al., (2015) use a set of macroeconomic drivers and find no speculative bubbles in the gold price which is consistent with Bertus and Stanhouse (2001) who also derive a gold's supply and demand model explicitly. However, these researchers contradict the findingsof Went et al., (2012), Homm & Breitung (2012) who find evidence supporting the existence of the gold price bubble hypothesis. Lucey and O'Connor (2013), take a deep dive and use the gold lease rate for the first time in the literature as a measure of fundamental value. However, these authors get conflicting results given different month lease rates. One reason for these contradictions is thatall these tests are based on certain assumptions for the fundamental determinant of the value of gold. As expected,the results are sensitive to the (unobserved) fundamental value specification.

In aforesaid literature, several models that are widely developed are usually based on the stationarity assumption of gold's price. However, the price of gold,same as other related financial indices, usually trends non-stationary (i.e. they have been trending over the last decades), meaning that it turns out that defining linear price relationships can generate spurious results. Besides, even though theequations used fits the data, the preselected functional forms are still subjective. Therefore, a challenge for research is to rule out the stationarity assumption constraints and to look for more light on dynamic issues to enable us to explain and forecast gold price with economic variables. Aiming to avoidlinear and stationary assumption as seen in traditional data analysis methods,we use a posteriori-defined statistics model, namely the Ensemble Empirical Mode Decomposition (EEMD), a method of temporal and spatial analysis based on Empirical Mode Decomposition (EMD)(Wu and Huang, 2005).

EMD was first introduced by Huang et al. (1998) and an approach entirely nonparametric (Huang et al. 2003). The method has been exhaustively tested and proved that, the EMD method could generate better results than those from any of the traditional analysis methods (for example Auto-regressive Integrated Moving average (ARIMA) process). Additionally, the EMDapproach canshow the true physical meanings of the examined data (Huang et al., 1996, 1998).

The frequent appearance of mode mixing[5] is the main inevitabledrawbackinthe EMD approach.

[3]Tangency point as the optimal result, given constraints.

[4]Gibson's Paradox states that interest rates are highly correlated to wholesale prices but had little correlation to the rate of inflation.

[5]Mode mixing means that a single intrinsic mode function (IMF) either consists of signals of widely disparate scales or made of signals of a similar scale residing in different IMF components.

It happensdue to signal intermittency, as discussed by Huang et al. (1996, 1998). To mitigate the scale separation problem without bringing in additionalsubjective assumptions, Wu and Huang (2004) extend EMD and proposed EEMD, which redefines the true intrinsic mode function (IMF) components as the mean of an ensemble of shifting procedure, each consisting of the signal plus a finite amplitude white noise, thus pertaining to the ensemble averages as the true intrinsic modes instead. The zero mean of the noise is averaged out over enough trials (Wu and Huang, 2005).

The first part of our research is closely related to Zhang et al. (2008), who used the EEMD approach to decompose and reconstruct crude oil price into three components: short-term fluctuations, supply-demand changes effect, and the trend. In this paper, we use the same model to sort out these various effects in gold price and then apply the decomposition results into forecasting and trading as shown in the second part of this paper. We also examine the short-term causality effect between ETF and gold price fluctuations.

III. METHODOLOGY AND INTERPRETATION

Zhang et al. (2007) argue that the data-driven models including Auto-regressive Integrated Moving average (ARIMA), Autoregressive Conditional Heteroscedasticity (ARCH) or Vector Regressionsusually implement well in short term forecasting scheme, but fail to do so inthe explanation offundamental driving forces that move pricesand usually these models loss their long run forecasting power. Besides, it is subjective even if the dataset fits well with preselected function, as there lack of support that the experimental data follows the priori subjective selected functional forms.

This dilemma can be clarifiedwithEEMD approach. The core of the EEMD model is to decompose the price data into a set of independent intrinsic modes, each represented as an intrinsic mode function (IMF) with two criteria, the same number of extrema and zero-crossings or differencing at most by one. Each IMF at any time point is local zero meanaccording to Huang et al. (2003).

The IMFs contain the variable amplitude and frequency at different times. It is extracted through a sifting process from high frequency (shorter period) to low frequency (longer period). Therefore, this method can help with interpreting the different data properties within different time ranges and fluctuation frequencieswithout any priori subjective assumptions. In the EEMD approach, a different normal distribution random data series is added to the original dataset during each iteration procedure. The new data series is then decomposed toobtain IMFs. Then, the ensemble mean of the corresponding IMFs are the final results.Thus,the scales are naturally distinctwith the ensemble mean and the white noise will cancel itself out via ensemble averaging procedure. Refer to Huang et al. (2003) for complete details[6].

Unlike normal goods, the price of gold has no apparent relationshipwith gold's demand and supply deficit records. The correlation between these two datasets is around -0.3 (from Q1 2010 to Q1 2019). As shown in figure 2, there is no clear path within these two parameters in between.

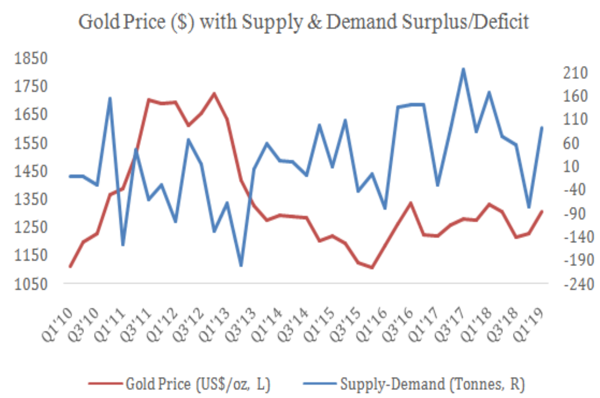

Figure 2: Gold Price with Supply & Demand Surplus/Deficit. Quarterly data from Q12010 to Q1 2019. Data corresponds to gold price in dollars showed on the left side in red; Gold supply minus demand (in Tonnes) showed on the right side in blue. Surplus (positive) means gold supply is above gold demand and vice versa.

3.1 Data and EEMD model decomposition

We use monthly data from December 1969 to May 2019 (594 observations). The data was obtained from the World Gold Council. EEMD produces seven IMFs and one trend series from the original price dataset. A larger index of the IMFs represents lower frequency and larger amplitudes. This reflects the idea that the change in gold price is small and fast in high frequency data (short run) but large and slow in low frequency data (long run). The way to split the IMFs into high or low frequency categories is pointed out by Zhang et al. (2007), who use a boundary line with a t-test to identify each IMF. The split point is identified at which the mean

[6]In this paper, we set the number of ensemble members to 100 and the standard deviation of white noise series is set to 0.1 as did in paper Wu and Huang, 2004, 2005; Zhang et al. 2008.

value significantly departs from zero. Our results, based on this rule, show that the high frequency results have a range within around 1 year, and that low frequency results represent longer periods. In what follows, we combine IMF1 to IMF3 into high frequency series and IMF4 to IMF7 as low frequency series according to the 1-year breakpoint standard.

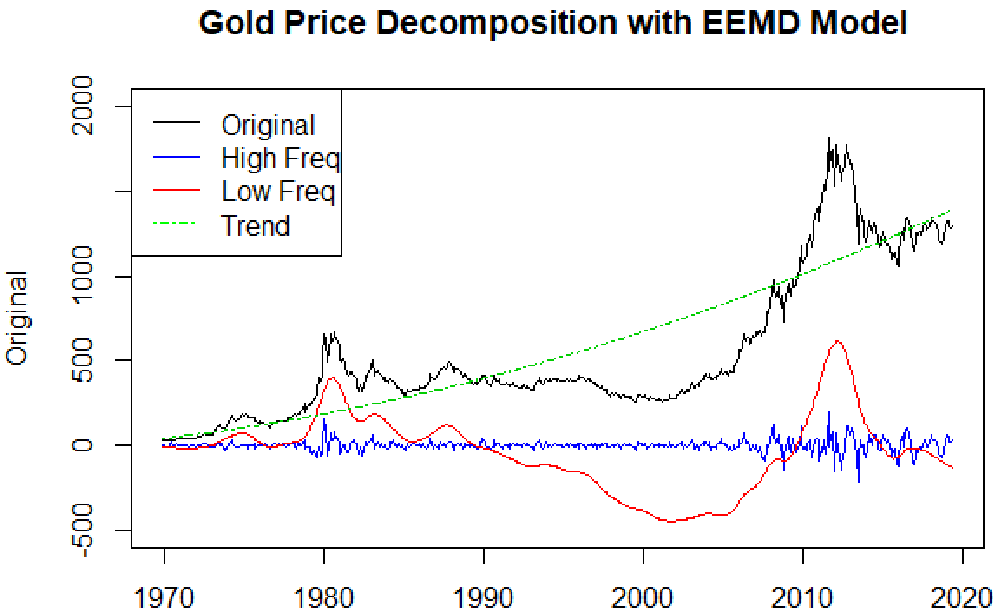

Figure 3: Gold Price with Supply & Demand Surplus/Deficit. Quarterly data from Q12010 to Q1 2019. Data corresponds to gold price in dollars showed on the left side in red; Gold supply minus demand (in Tonnes) showed on the right side in blue. Surplus (positive) means gold supply is above gold demand and vice versa.

Through the EEMD method we can separate the price of gold into a small number of independent and meaningfulIMFs based on different time scales. Thecomponents are identified as high frequency series (short term fluctuations for less than a year in our case) caused by speculation and unforeseen economic events. This is adherent to gold's monetary property as an investment asset used for speculation and hedging. Low frequency series (longer than one year) that reflect the reaction from time to time to a shock of a significant event, supply-demand disequilibrium. This follows gold's monetary property independent from its speculative characteristic.And a long-term trend related to its scarcity linked to its commodity property and the increasing jewelry demand.With limited total supply, the trending price is supported by the expanding gold demand and the influence of inflation and other economic activities. The analysis of these IMFs can assist to analyze the volatility and gold price formation from a new perspective.

Current presence of various innovative financial instruments involved in today's financial market, price behavior, driven by mixed hedge and investment strategies using ETFs, futures contracts, options, and lastly cryptocurrency, becomes more and more volatile. In recent years there is a significant sign of the increase in gold's monetary property as an investment tool (speculation and hedging). We will zoom into the last ten years' weekly data to test this conjecture.

3.2 Data and EEMD model decomposition

To focus on the recently observed price volatility, we use the other two datasets, weekly gold price and weekly ETF holdings (oz) from 10/19/2007 to 10/13/2017. The data wasobtained from the World Gold Council. In this case a total of 522 observations are included. We run the same model for both datasets and, create seven IMFs and one trend series for each of the datasets. The combined EEMD results are shown in Figure 4.

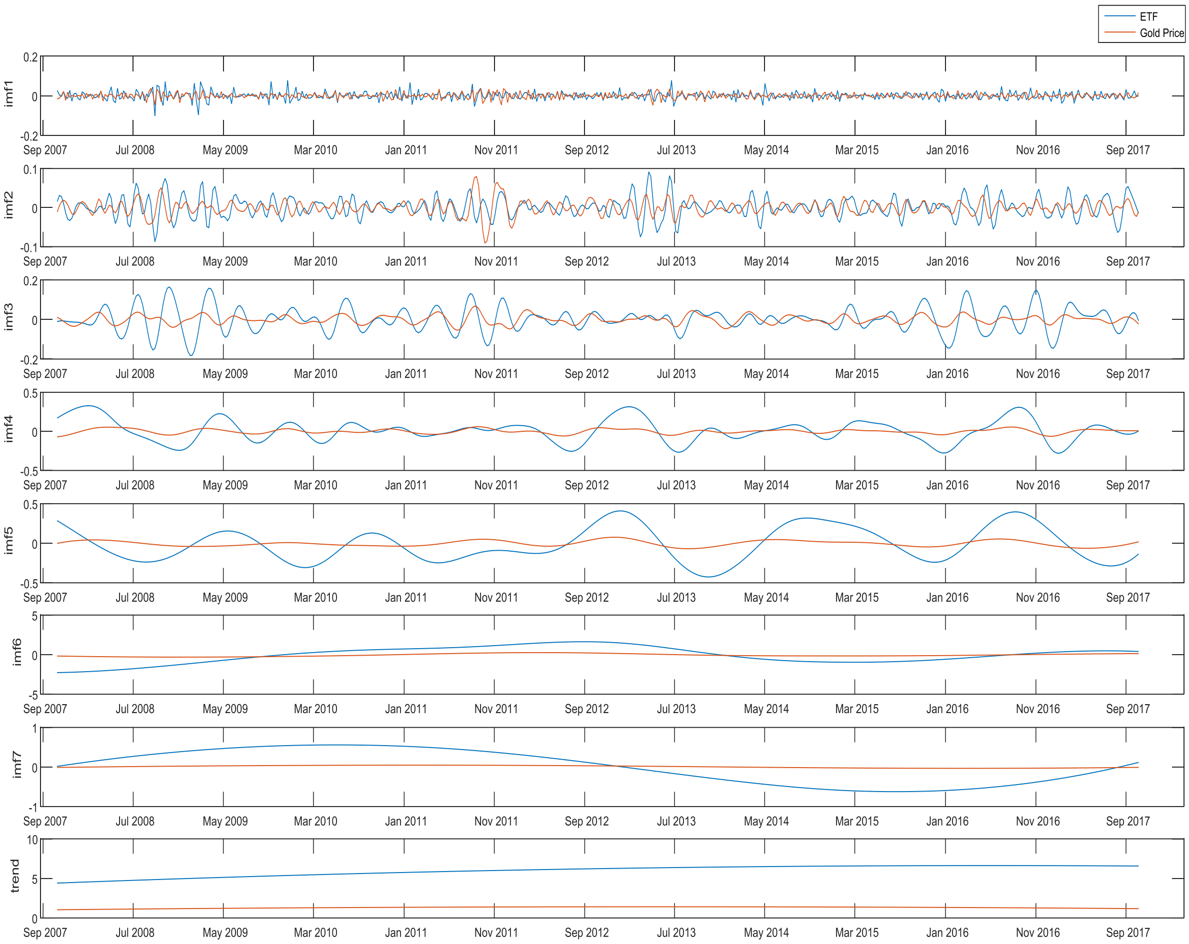

Figure 4: Gold Price and ETF holding decomposition with EEMD Model. The blue lines represent the ETF IMFs and the red lines are IMFs for Gold price. We use weekly data from 10/19/2007 to 10/13/2017.

As explained in the previous sections, purchases and sales of ETFs constitute a significant part of gold's demand in terms of its monetary property, as such appears to be driving gold's short-term price movements. High frequency volatility in recent weeks increased amid higher ETF trading volatility. As it is apparent, in the short term, the volatility of gold price and ETFs overlap to some extent. The EEMD results show that the price of gold and ETFs have very similar variations for IMF1 to IMF2, which mainly represent the high-frequency variations, at a scale of 2.98 to 7.15 weeks.

In this paper, we are also interested in analyzing whether there exists any causality relationship between gold and gold- ETFs holdings. For this, we further performed the granger causality test. The impulse response function (IRF) results are presented in figure 5.

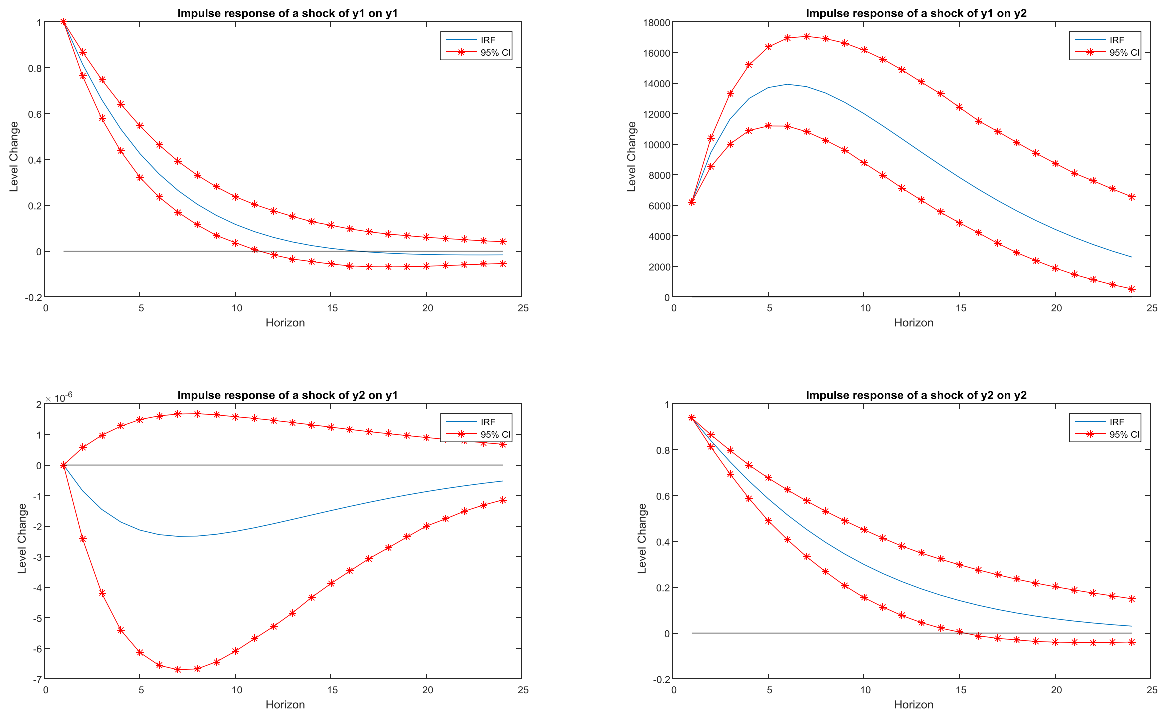

Figure 5: Impulse Response Function results. In this figure, Y1 represents the high frequency component of the gold's price in the red dotted line and, Y2 represents the high frequency component of ETF holdings in the blue line.

Figure 5 demonstrates that while the effect of high frequency ETFs on high frequency gold price (bottom left) is negative but not statistically significant; the rest of the IRF showa significant and positive response[7].In addition, the granger causality test resultimplying a unidirectional causality relationship fromthe price of gold to ETF in high frequency components[8]. While in terms of the low frequency components, gold's price and ETF have a bidirectional granger-causality.This goes in hand with the apparent observed fact that ETFs' traders tend to usemomentum trading strategies to chase short term price movements (high frequency) and,that the volatility change from ETFs volumes flow back to affect gold's price in longer terms (low frequency).

[7]The IRF results for low frequency datasets are available upon request.

[8]For the high frequency datasets, the F-statistic of granger causality test equals to 0.2785 and 0 respectively, meaning that only the high frequency component of gold's price granger cause ETF. While for the low frequency components, the granger causality test equals to 0.0004 and 0.00009 respectively, meaning the low frequency component of gold's price and ETF have a bidirectional granger-causality.

IV. FORECASTING RESULTS WITH MLP AND ELM

The primary goal in this section is to provide a practical implication for including EEMD into price forecasting procedure, aiming to show ifthe EEMD integrated models can providea better(several-steps ahead)forecasting resultswith weekly and monthly data.

We apply two machine learning models, Multi Layer Perceptrons (MLP) and Extreme Learning Machines (ELM) to exam the performance of the prediction model with EEMD approach included. We choose only two approaches because the main point in this paper is to see if EEMD can improve the forecasting accuracy but not to find the best forecasting model (this is left for future research). These two approaches are picked based on the result of Ahmed (2010) paper.

Ahmed (2010) examines the forecasts ofthe 1045 series from the M3 competition. According to his results, MLP generates the best performance regarding to the accuracy indicator among eight other Machine learning models[9]. The

[9]The paper compares the result from eight different machine learning models: Multi-Layer Perceptron (MLP); Bayesian Neural Network (BNN), Radial

sMAPE (%)value for MLP is approximately two digits lower than the others. For this reason, we include the MLP in this paper. Moreover, we add an additional model,namely the Extreme Learning Machines (ELM),which is a rising star in the forecasting areaand can provide a superior solutions than Support Vector Machines (SVM) or Least-square Support Vector Machine (LS-SVM) in regression problems (Huang et al. 2012).ELM is a nonparametric machine learning model with no strong assumptions about the form of the mapping function. Unlike MLP, which assumes a logit mapping function in the process, ELM is free to learn any functional form from the training data.

We apply monthly and weekly data from January 2000 to December 2018, the rolling window for the forecasting period is 3, 6, and 13weeks for weekly data and,3, 6, and 12 months for monthly data. We use data from 2000 to 2003 to train the model, thenthe testing date starts from January 2004 for both monthly and weekly subset data groups. We present the benchmark result as the output from the MLP and ELM model with the original gold price. We then use the same price data to firstly run the EEMD models beforeapplyingthe MLP or ELM model with the same forecasting periodsfor each IMF. We then sum all forecasting results from each IMF to see the differencein forecasting resultswhenthe EEMD modelis included. The results are shown in Table 1.

Table 1: Gold Price Forecasting Result

Method		Weekly			Monthly		
	Forecast Period	3	6	13	3	6	12
	Date Span	1/04 - 12/18	1/04 - 12/18	1/04 - 12/18	1/05 - 12/18	1/05 - 12/18	1/05 - 12/18
	# of Obs.	780	780	780	168	168	168
MLP - EEMD	MSE	1,223	2,662	6,381	4,339	7,089	16,711
	RMSE	34.98	51.60	79.88	65.87	84.19	129.27
	SMAPE	2.39	3.31	5.66	4.47	5.48	9.32
	Corr w RealPrice	0.99	0.98	0.97	0.95	0.92	0.89
MLP - Original	MSE	1,674	3,109	5,702	8,176	11,794	22,286
	RMSE	40.92	55.76	75.51	90.42	108.60	149.29
	SMAPE	2.78	3.65	5.22	5.87	7.35	9.33
	Corr w RealPrice	0.99	0.99	0.98	0.97	0.95	0.92
ELM - EEMD	MSE	1,824	4,301	10,746	3,487	7,677	16,406
	RMSE	42.71	65.58	103.66	59.05	87.62	128.08
	SMAPE	2.99	4.48	7.74	4.20	6.00	9.37
	Corr w RealPrice	0.99	0.98	0.98	0.95	0.92	0.87
ELM - Original	MSE	2,368	4,924	10,320	7,346	11,146	23,509
	RMSE	48.66	70.17	101.59	85.71	105.58	153.33
	SMAPE	3.49	4.98	7.66	5.72	7.49	10.38
	Corr w RealPrice	0.99	0.99	0.97	0.97	0.96	0.92

Basis Functions (RBF); Kernel Regression; K-Nearest Neighbor Regression (KNN); CART Regression Trees, Support Vector Regression (SVR) and Gaussian Processes (GP).

In this table, we use different rolling periods (3, 6, and 13 weeks or months) to forecast based on MLP and ELM models and, compare the result with the output when EEMD is included.

Overall and in terms of MSE, RMSE, sMAPE,both MLP and ELM model generate much better forecasting results with EEMD included. In our case, MLP outperforms theELM in most of the cases. Up to here, we have reason to believe that it is possible that EEMD could help to smooth out the high frequency noise and to better identify theshort-term price path. We also plot out the ELM weekly results with 13 weeks rolling time window in Figure 5. The red line (ELM with EEMD model included) follows the black line (real price) more closely than the blue line (ELM model only). We further zoom into the year 2009-2015 (Figure 6, right panel) to see if the updated model can capture the short-term trends in most of the cases. We choose this specific period because gold pricesignificantly increases from 2009 to 2011 (uptrend case), stay relatively constant for around 3 years then dramatically decreases (downtrend case) thereafter. It is even more obviousin this graph that the blue line (ELM model only) diverges away from the black line (real price) while the red line (EEMD model plus the ELM model) traces the price trend closely from year to year. MLP displays the same behavior for this period. These discoveries facilitated our assumption that it is amendable and preciseto include the EEMD when it comes to the short-term price trend prediction. More well designed and comprehensive experimental workis guaranteed in the future study.

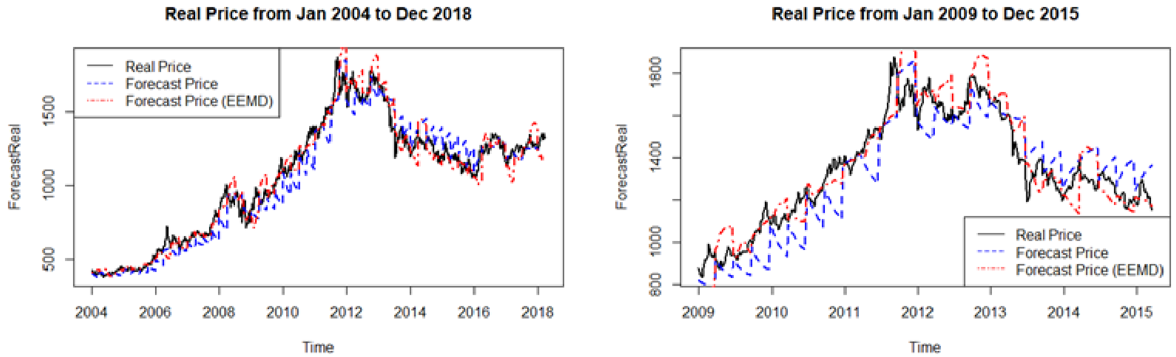

Figure 6: Forecasting results. We use monthly data from 2000 to 2018 to test the difference between ELM model and the improvement EEMD model can make in terms of time series forecasting results. The output with the EEMD model included is shown in the red line, the real gold price shows in the black line and the forecasting model (ELM) with no EEMD model result is shown in the blue line.

V. TRADING IMPLEMENTATION

Motivated by the EEMD model, we propose a two-step time series model to forecasting the next period's (daily and intraday) price direction. We make buy, hold or sell decision based on the local minimum (buy) and local maximum (sell) point used in the first step of the EEMD approach. The second step is to use the kernel SVM and Random Forest approach to replicate the model that is close to the EEMD output with 13 technical indicators. Moreover, current studies have employedthe price direction predicting problem as a binary one (buy or sell) based on the change of prices. Our main contribution in this framework is that we include a third potential decision (hold) and amend the price forecasting problem into a multi-classification issue from the simple binary classification.

To train the model, researchers assign buy if the next day's price is higher than the current level and sell vice versa. The problem with this approach is that this can generatemassive meaningless trades which would induce huge transaction costs and potentially damage capital gains. To avoid this situation, in our paper, the trade action is based only on local minimum and maximum theory from EEMD mode. The buy signal is defined as a local minimum point and a sell signal will be generated if it is a maximum point instead. In other cases, we define the point as a hold, meaning we do not do anything at a specific time since the signal is not clear. In this way, we can eliminate the amount of meaningless trade and reduce the transaction cost albeit miss a few trade opportunities in between.

There is one more benefit we should touch on this specific model we are building; this process sets a limit to the accumulated shares on hold. To be specific, the buy and sell pointsare activated in turns which means that we will not accumulate too many shares at each instant (this is consistent with the two IMF criteria required in the EEMD model). This is very important in high frequency trading, since a sharp price change can rule out all profits within seconds given a large position on hold. In a volatile market like the ones observed in current days, it is more important to adhere to more prudent trading strategies (such as the one we are proposing).Ideally, if our testing result is close to the training model, we can eliminate the volume risk to some extent as well.

To have a better picture of what we are trying to do, we plot a sample with our decision rule. As shown in Figure 7, we will buy when the color of the dot is green (local minimum point) and sell when it is red (local maximum point). If it is blue means we keep our current position and do not trade at that time. The red and green dots appear in sequence means we are eliminating the accumulated shares as mentioned above. Besides, with the hold condition included, we can knock out the unnecessary trade and be more effective, which can simultaneously decrease the transaction fees dramatically.

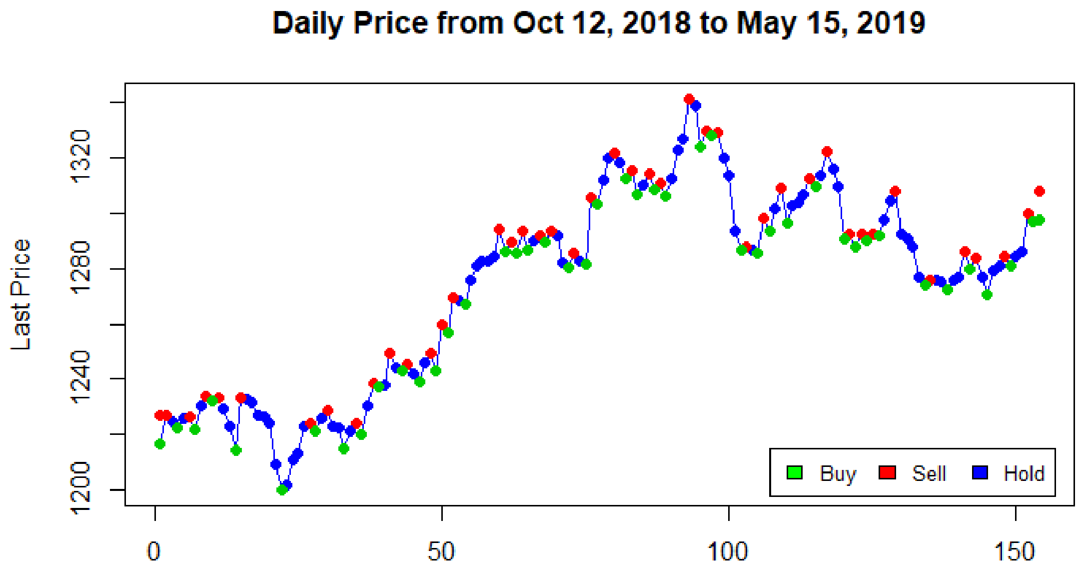

Figure 7: Decision Rule. We use daily data from Oct 12, 2018 to May 15, 2019 to show the trading action with the EEMD model's set up. The green dots represent the buy point, the red dots are for sell, the blue dots mean we should hold our current position and do nothing. The green, red and blue dots are defined from the EEMD model.

We apply two algorithms to build the model,kernel SVM (Support Vector Machine) and Random Forest, then do forecast respectively. We split the dataset into two parts, apply the local minimum and local maximum theory to define the buy, sell and hold points in our sample period, and use 13 technical indicators with kernel SVM and Random Forest approaches to replicate the model with the training dataset. We then use the testing dataset to examine the model output.

In this section, we use daily and intraday(5minutes and 1minute) data to exam the model effects. Our study covers the period of May 12, 2000 to May 15, 2019 (total 4954 observations of daily data); April 16, 2019 to May 14, 2019 (total 5547 observations of 5-minute data; and 26,344 observations of 1-minute data). We train our model with data from May 2000 to July 2015 (daily), April 16 to May 8, 2019 (5 minute and 1 minute) then test the model with data thereafter to see the forecasting results. As mentioned earlier, this training period contains the uptrend, downtrend, and relative constant period which is important to build the model that covers all these scenarios. We use buy and hold strategy as a benchmark, assuming we hold 100 shares in the whole testing period.

To replicate the EEMD model, we apply the widely used short term trend technical indicators since we are only interested in the nextperiod's price direction and not in the long-term one. The technical indicators picked in this paper are: Williams %R, Stochastic momentum, moving average convergence (MACD); Correlation, Change of Correlation. On balance volume (OBV); Relative Strength Index (RSI), Rate of Change (ROC), etc.The model output is shown in Table 2.

Table 2: Gold Price Forecasting Result

Strategy	EndProfit	MaxProfit	MaxLoss	MaxVolume	MinVolume	SharpeRatio
Daily Data (Beginning price is $1094.24, ending price is $1297.29)						
BuyHold	20,305	27,209	(4,314)	100	-	0.51
KernelSVM	976,706	1,534,608	(401,814)	13,300	100	0.51
RandomForest	1,350,116	1,752,959	(142,426)	10,300	100	0.51
EEMD	351,045	351,045	-	200	-	11.64
5min Data (Beginning price is $1282.6, ending price is $1297.15)						
BuyHold	1,455	2,058	(284)	100	-	9.80
KernelSVM	79,866	118,199	(8,924)	8,000	-	9.94

RandomForest	104,265	140,995	(9,751)	8,500	(100)	10.77
EEMD	15,334	15,334	-	-	(200)	181.17
1min Data (Beginning price is $1280.96, ending price is $1297.02)						
BuyHold	1,606	2,232	(38)	100	-	11.48
KernelSVM	61,483	67,175	(6,045)	4,100	(400)	14.77
RandomForest	278,560	354,725	(43,959)	20,800	(100)	11.47
EEMD	33,587	33,587	-	200	-	363.48

We present the strategies'P&L performance during the sample period. The 2nd column shows the ending profit under each strategy. The 3rdand 4thcolumns show the maximum and minimum profits we reach within the sample period. We do not hold restrictions for the strategy, if the signal is buy, we then include 100 long shares, decrease 100 shares if the signal is sell. The last two columns show the maximum and minimum volume we reach in this process.

We see from table 2 that all our approaches could generate a better profit than a simple buy and hold strategy even in a bullish period. In addition, EEMD model is the most prudent strategy as it is accumulating the profit gradually and liquidate the shares on hold at a fast speed (the maximum share on hold is no more than 200 shares at any given point of time). KernelSVM generates closer results to the EEMD model for 1min data, and Random Forest performs well in a daily and 5min time interval. Figure 8 shows the P&L fluctuation in a subsample of the period used for daily data.It shows that the EEMD model generates consistent positive returns in this sample period, it can be a compliment model to the buy-hold strategy with lower risk. Random Forest and kernel SVM model follow the price trend periodicallybut generate much higher profit at the end of the period. One possible reason could be that we do not have restrictions on volume on hold.

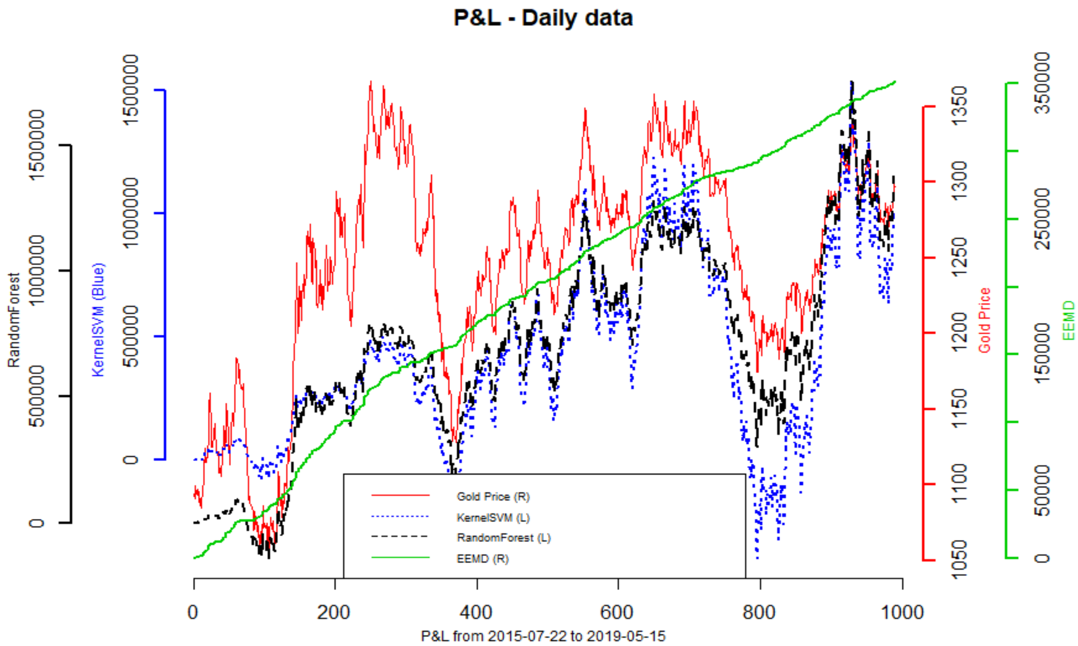

Figure 8: Accumulated P&L with daily data. The accumulated profits from Random Forest are presented in the left axis in black and the Kernel SVM is also in the left axis in blue; the EEMD strategy is presented in the right axis in green and the gold price is in the right axis in red. Wecut out a section of the testing period (July 22, 2015 to May 15, 2019) to show the P&L fluctuations in this picture. Accumulated P&L with intraday data is available upon request.

The experimental results lead to the conclusion that our proposed model can be used in real time tradingas it is likely to yield much better profit. Our next work is about the improvement of the selection of technical indicators used inbuilding the model to better mimic the EEMD model output.

Besides, to verify the model, a larger set of data is necessary to analyze and further improve the accuracy of the model output and the expected investment returns.

VI. DISCUSSION AND CONCLUSION

As an alternative currency to fiat currencies, gold is well positioned to benefit from the trend of increased global quantitative easing and rising levels of sovereign debt.Due to gold's intrinsic complexity properties, it is almost impossible to produce a consistently accurateprice forecasting result. In this paper we examine the double nature of gold as a commodity and financial investment with the EEMD model and focus on the model application for data within different time frames.

Based on the EEMD results, we see that the high frequency dataset has no serious impact on the gold price trend,but these eventsaccumulate and gradually become the fundamental impulsion for driving the gold price up in the short run. Given this conclusion, we follow two machine learning models, Extreme Learning Machine (ELM) and Multi Layer Perceptrons (MLP) to obtain several-steps ahead forecasts with weekly and monthly data, not just the one-step ones as did in most of the papers. We test and prove that the result is more precise with the EEMD model included since all error measurement indicators show with a lower value in this case (smaller value of MSE, RMSE, and SMAPE). Toward the higher frequency data (daily and intraday), we pose the forecasting problem as a direction predicting exercise and introduce the hold condition based on the EEMD model local minimum and local maximum setup. We replicate the EEMD model with 13 short-term trend indicators through Kernel SVM and Random Forest to solve this multi-class classification problem. The experimental results lead to the conclusion that our model can be successfully used as a real-time trading model and generate much higher profit with more efficient trading transactions than a simple buy and hold strategy even in a bullish market.

REFERENCE

[1]. Adrangi, B., Chatrath, A., & Raffiee, K. (2003). Economic activity, inflation, and hedging: The case of gold and silver investments. Journal of Wealth Management, 6(2), 60–77.
[2]. Ahmed, N., Atiya, A., Gayar, N.,& El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Econometric Reviews, 29(5–6):594–621.
[3]. Aye, G., Gupta, R., Hammoudeh S., & Kim, W.J. (2015). Forecasting the price of gold using dynamic model averaging. International Review of Financial Analysis, 41, 257–266.
[4]. Baur, D.G., Beckmann, J., Czudaj, R. (2016). A melting pot – Gold price forecasts under model and parameter uncertainty. International Review of Financial Analysis, 48, 282–291.
[5]. Baur, D., & Lucey, B. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. Financial Review, 45(2), 217-229.
[6]. Batten, J.A., Ciner, C., Kosedag, A., Lucey, B.M. (2017). Is the price of gold to gold mining stocks asymmetric. Economic Modelling, 60, 402-407.
[7]. Bertus, M., & Stanhouse, B. (2001). Rational speculative bubbles in the gold futures market: an application of dynamic factor analysis. The Journal of Futures Markets, 21(1), 79-108.
[8]. Bialkowski, J., Bohl, M.T., Stephan, P.M., & Wisniewski, T.P. (2015). The gold price in times of crisis. International Review of Financial Analysis, 41, 329-339.
[9]. Bordo, Michael D., and Richard W. Ellson. (1985). A Model of the Classical Gold Standard with Depletion. Journal of Monetary Economics, 16, 109-20.
[10]. Capie, F., Mills, T. C., Wood, G. (2005). Gold as a hedge against the dollar. J. Int. Financ. Mark. Inst. Money 15(4), 343–352.
[11]. Dowd, Kevin, and Anthony A. Sampson. (1993). A New Model of the Gold Standard. Canadian Journal of Economics 26, 380-91.
[12]. Ghosh, D., Levin, E., Macmillan, P., & Wright, R. (2004). Gold as an inflation hedge. Studies in Economics and Finance, 22, 1-25.
[13]. Homm, U., & Breitung, J. (2012). Testing for speculative bubbles in stockmarkets: A comparison of alternative methods. Journal of Financial Econometrics, 10(1), 198–231.
[14]. Huang, N. E., Long, S.R., and Shen, Z. (1996). The mechanism for frequency downshift in nonlinear wave evolution. Adv. Appl. Mech., 32, 59–111.
[15]. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C., Liu, H.H., (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series

analysis. Proc. R. Soc. London, Ser. A, 454, 903–995.

[16]. Huang, N.E., Wu, M.L., Qu, W.D., Long, S.R., Shen, S.S.P. (2003). Applications of Hilbert–Huang transform to non- stationary financial time series analysis. Applied Stochastic Models in Business and Industry 19, 245–268.

[17]. Huang, G.B., Zhou, H.M., Ding, X., and Zhang, R. (2012). Extreme Learning Machine for Regression and Multiclass Classification. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 42 (2), 513–529.

[18]. Kaufman, T., & Winters, R. (1989). The price of gold: a simple model. Resource Policy, 15(4), 309-313.

[19]. Levin, E., & Wright, D. (2006). Short-run and long-run determinants of the price of gold. World Gold Council. Research Study No. 32.

[20]. Lucey, B.M., O'Connor, F.A. (2013). Do bubbles occur in the gold price? An investigation of gold lease rates and Markov Switching models. Borsa Istanbul Review, 13, 53-63.

[21]. McDermott, John. (1987). Adding Exhaustibility to the Traditional Theory of the Gold Standard. Journal of Macroeconomics, 9, 545-66.

[22]. Ming, L., Yang, S., Cheng, C. (2016). The double nature of the price of gold - A quantitative analysis based on Ensemble Empirical Mode Decomposition. Resources Policy 47, 125-131.

[23]. Shahbaz, M., Balcilar, M., Ozdemir, Z.A. (2017). Does oil predict gold? A nonparametric causality-in-quantiles approach. Resources Policy, 52, 257-265.

[24]. Sjaastad, L. A. (2008). The price of gold and the exchange rates: Once again. Resource Policy, 33(2), 118–124.

[25]. Tully, E., Lucey, B.M. (2007). A power GARCH examination of the gold market. Research in International Business and Finance, 21, 316-325.

[26]. Went, P., Jirasakuldech, B., & Emekter, R. (2012). Rational speculative bubbles and commodities markets: Application of duration dependence test. Applied Financial Economics, 22(7), 581–596.

[27]. Worthington, A.C., & Pahlavani, M. (2007). Gold investment as an inflationary hedge: Cointegration evidence with allowance for endogenous structural breaks. Applied Financial Economics Letters, 3(4), 259–262.

[28]. Wu, Z., Huang, N.E. (2005). Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method (Center for Ocean-Land-Atmosphere Studies, Calverton, MD), Technical Report 193, p.51. http://www.iges.org/pubs/tech.html.

[29]. Wu, Z., Huang, N.E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. Proceedings of the Royal Society of London. A 460, 1597–1611.

[30]. Zhang, X., Lai, K.K., & Wang, S.Y. (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Economics, 30(3), 905–918.

Meng Sui, et. al. "Implementing the Decomposition Model in Gold Price Volatility Analyzing, Forecasting and Trading." *International Journal of Engineering Research and Applications (IJERA)*, vol.10 (08), 2020, pp. 27-39