Forecasting gold returns using principal component analysis from a large number of predictors

Fredrik Allgén
Supervisor: Joakim Westerlund

Master Thesis I, 2022

Lund University School of Economics and Management

Abstract

Gold is known in the financial world to be an important asset in unstable periods, especially as a hedge against inflation. If the gold price can be forecasted, it will be possible to strategically invest in gold rather than acquire it as a last-minute hedge against economic downturns. Although there are many studies on forecasting, few focus directly on gold returns.

This study investigates the role of economic variables in predicting gold returns, followed by a twelve-month forecast to determine its future returns. By principal component analysis, a large number of predictors are extracted down to seven factors, these are; the business cycle, nominal, interest rate, commodity, exchange rate, stock price, and government bond yield factor. The ARMA model is used to predict gold returns with these factors and two additional variables, the Kansas City Fed's financial stress index, and the U.S. economic policy uncertainty index. The available dataset contains data from January 2000 through December 2019. With an in-sample period from January 2000 to December 2009 and an out-of-sample from January 2010 to December 2019.

Three alternative predictive models are compared to evaluate the forecast performance. The original model included all variables, the AR (1) benchmark model, and the model excluding the government bond yield factor. The results from the mean squared forecasted errors showed that the forecasting models containing predictors extracted from the principal component analysis outperformed the benchmark model in forecasting. The forecast for the twelve months preceding the data period showed solely positive returns. Regarding predictive power, the interest rate factor contributed to increasing gold returns. On the contrary, the business cycle factor, nominal factor, and stock market factor all tend to have a negative effect on the return of gold. However, the other variables showed insignificant results; therefore, the evidence was not strong enough to draw additional conclusions.

Keywords: Forecasting, Gold, Principal Component Analysis, ARMA

Table of contents

Introduction

1.1 Background

Gold has been used throughout history for many purposes, but mainly as a tool for indirect exchange and an element to store wealth. Banerjee (2013) argues that gold may have been the start of civilized society since the commodity market is one of the most fundamental markets in the world and was heavily dependent on gold. Nowadays, gold serves several functions in the world economy and is heavily linked to macro and financial variables (Pierdzioch, Risse, & Rohloff, 2014a). It is a financial asset that can easily be converted into cash. Due to its monetary value, central banks use gold as a part of their federal reserve (Gupta, Hammoudeh, Kim, & Simo-Kengne, 2014). It can also be transformed into jewellery and therefore serves an industrial purpose, creating jobs and stimulating the economy. During uncertain times and crises, gold tends to be a popular investment tool to hedge against major currencies and inflation. Therefore, gold is widely accepted as a safe haven for many households (Ghazali, Lean, & Bahari, 2013).

In recent years, the world economy has been showing instability. The COVID-19 pandemic, with global lockdowns and movement restrictions around the world, has negatively affected the economy. More recently, the global economy has suffered from the effects of faster inflation and slower growth as a consequence of the Russian invasion of Ukraine in March 2022. As well as the federal reserve in 2022 raised the interest rate to the most significant hike since 1994 (CNBC 2022). It results in many investors seeking a safe haven such as gold, an investment to retain or increase its value during market turbulence.

Kiraz and Yildiz Üstün (2020) showed that the COVID-19 pandemic affected the world financially similarly to the 2001 oil crisis and 2008 global financial crisis. The graph below shows that gold prices increased significantly during the financial crisis in 2007. Demonstrating the increased demand for gold during uncertain economic times. As well as the upward trend for gold prices under emerging financial crises, shocks and uncertainties. However, since the price of gold depends on future supply and demand and is forward-looking, thus predicting gold price fluctuations is very difficult, according to Dichtl (2019).

Figure 1. Gold price (US dollar)

1.2 Previous Research

There are many published studies on forecasting. However, fewer studies have been published on forecasting gold returns. This has changed in the last few years, with O'Connor et al. (2015) showing that the number of publications on gold returns and predictability has increased substantially. Ismail et al. (2009) attempted to forecast gold prices using multiple linear regression. They showed that by using a set of economic factors such as exchange rates, commodity index, inflation rate, money stock, etc., they accomplished higher predictive accuracy compared to previous studies. Another study by Sjaastad (2008) showed a significant relationship between the gold price and major exchange rates. The main finding was the significant impact on the world market if the US dollar either depreciated or appreciated, resulting in substantial effects on the gold price. Moreover, Shafiee and Topal (2009) tried to forecast the next ten years by investigating the relationship between gold price and key influencing factors. They found empirical evidence that the gold price would be exceptionally high until 2014 and revert to normal by 2018. Pierdzioch et al. (2014b) investigated if the business cycle, represented by the output from the G7 countries, had a predictive power on the gold price change. The results were significant, and the authors could conclude that the business cycle is a good predictor of gold prices. Another important finding by Apergis (2014) was that gold price significantly correlates with nominal and real exchange rates. By an out-of-sample forecast, using daily and quarterly data, Aspergis found empirical evidence that the Australian dollar/U.S. dollar exchange rates affected the gold price. In addition, the author also provides predictive power that both markets were driven by each other.

My literature search only identified a few studies that focused directly on the topic of forecasting gold returns; some are Aye et. (2015), Dichtl (2019), and Pierdzioch et al. (2014b). Pierdzioch et al. (2014b), used quite a few numbers of predictors and only focused on the G7 countries. Aye et al. (2015) used similar predictors as Pierdzioch et al. (2014b) but instead of allowing for model uncertainty, as Pierdzioch et al. (2014a) did, Aye et al. (2015) used a similar nonlinear model for both in-and out-of-sample to try to estimate the relationship between gold returns and the predicting variables. Aye et al. (2015) also allowed for time-varying coefficients. The results found in this study were that the forecast of dynamic model averaging and dynamic model selection outperformed the constant parameter models, as well as the major predictor power for gold returns was financial variables compared to real economic variables. Dichtl (2019) continued the work of Aye et al. (2015) by investigating the forecasting power of gold returns using technical indicators, diffusion indices, and economically motivated restrictions in predictive regressions. Even though the result from the study is that neither of these concepts leads to improvements in the gold returns predictions, the authors still found strong empirical evidence that some variables are suited for forecasting gold returns in various states of the economy. Some variables showed more substantial predictive power in expansive business cycles compared to others that performed better in recessive business cycles. Dichtl (2019) suggests that future research regarding forecasting gold returns should apply a regime-dependent forecast method where the relationships between the variables should depend on some prevailing background regime, such that can be changed by, for example, assessing different countries.

1.3 Purpose

Forecasting is widely used worldwide in various sectors. It is both interesting from a market perspective to measure the efficiency of different financial variables, as well as highly relevant for assets management companies and investors, both private and institutional. Monetary policymakers also use it for informational advantage. Hedge fund managers also try to estimate future inflation, exchange rate, and demand for goods. Therefore, being able to forecast future values is an essential factor for many stakeholders in the financial world.

Countries are getting increasingly dependent on each other. Over the past decades, increased international trade and reduced barriers have allowed the global economy to flourish when countries focus on their special expertise. Yet, this leaves countries more vulnerable to global events than ever before. The COVID-19 pandemic and the Russian invasion of Ukraine are examples of how large events significantly impact our global economy. In the U.S, the effects are seen in the record-high inflation rate and the highest increase in the interest rate in almost 30 years.

When the future is uncertain, investors leave the market searching for safer options. Gold is one such alternative. If its price can be forecasted, we minimize the risk of investing in gold at the wrong time. Instead of being a last-minute exit from unstable markets, gold investments can be a hedge against a downturn while still being a high-yielding asset.

Nowadays, economic data is readily accessible. When fitting a model, generally, more data and different types of variables are better. Adding additional explanatory variables can improve the fit as the R^2 increases. Hence, applying data from worldwide makes the prediction more accurate for the model. Nonetheless, estimating a model using too many predictors may lead to overspecification bias. However, Stock and Watson (2002) found that forecasting time series using a large number of predictors based on principal components is asymptotically efficient. This is an essential solution for economic forecasting with big datasets. The empirical results from the study also showed that the methods with a large number of variables substantially improved the forecasting models compared to the models using fewer variables.

This study seeks to find the models that most accurately forecast gold returns and the variables that are good predictors, in order to determine future gold returns using a twelve-month forecast. This is done through a multi-step ahead forecasting method with a large number of predictors, to achieve better predictability. By using principal component analysis, the data becomes simplified without losing important traits and helps with the overspecification problem.

Methodology

The method used for implementation was the statistical software package STATA 17.

2.1 Log returns

For the dependent variable gold, log returns are used. The reason is that log returns are time additive, so it is simple to obtain the n-period return just by adding all the single period returns until n is reached, as well as that the variable becomes normalized. Therefore, the returns are in a comparable metric. The log returns are calculated by

$$R_t = ln \frac{P_t}{P_{t-1}} \tag{1}$$

Where, P_t is the price of gold and t represents the given time.

2.2 Principal component analysis (PCA)

PCA is commonly used for making predictive models, where many variables are correlated, and the goal is to simplify the data without losing essential traits. In other words, PCA is commonly used to reduce the dimensionality in a data set with a broad number of interrelated variables while keeping as much of the variation in the data set. This is done by computing principal components, a new set of transformed variables that do not correlate. They are ordered from top to bottom, with the most variation present from the original variables (Jolliffe, 2002).

The main objective of PCA is to find a linear combination of the variables with the highest variance. The first step in this is to find a linear function $\alpha_1'x$ of the elements of x with the maximum variance. Where x is a vector of random variables p and $\alpha_1$is a vector of p constants $\alpha_{11,}\alpha_{12}, \dots, \alpha_{1p}$. So

$$\alpha_1'x = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j \tag{2}$$

The second step is to find a linear combination of $\alpha_2'x$ that is not correlated with $\alpha_1'x$ and has the second highest variance. Keep on doing so until the k-th stage of a linear function of $\alpha_k'x$ is found that is not correlated with $\alpha_1'x, \alpha_2'x$,…, $\alpha_{k-1}'x$. This variable, $\alpha_k'x$, is the k-th PC.

A way to determine how suited the data is for a factor in the principal components is by testing the sampling adequacy for the variables by Kaiser-Meyer-Olkin.

$$KMO = \frac{\sum_S r_{ij}^2}{\sum_s (a_{ij}^2 + r_{ij}^2)} \tag{3}$$

Where S is the correlation of variables i and j and a_{ij} is the anti-image correlation (Kaisar 1974).

2.3. Stationarity test

According to Brooks (2014), a requirement in time series analysis is that the variables must be stationary. If not, modeling a regression might result in spurious bias where the regression can show significant results that are incorrect. Dickey and Fuller (1979) developed a procedure testing whether a variable either follows a unit-root process or a random walk. This Augmented Dickey-Fuller (ADF) test is used to test for stationarity. The null hypothesis states that the variable consists of a unit root. On the other, the alternative hypothesis is that the variable was derived by a process of stationarity. Alternatively, for the regression, you can include a trend term and lagged values of the difference of the variable if you exclude the constant (Enders, 2015).

2.4 Autoregressive-moving-average model

The ARMA model is used to predict multi-step ahead future values based on its past values. Where a dependent variable is fit on an independent variable to follow a linear autoregressive moving-average specification, this works through lagged moving averages which smooth the data of time series. An ARMA model is a model that contains both an Autoregressive process (AR) and a Moving Average process (MA). The AR(p) model for y_t can be combined with an MA(q) model for ε_t, leading to the following ARMA(p, q):

$$y_t \ = \ a_1 y_{t-1} + \ldots + a_p y_{t-p} + \ \varepsilon_t - \ \beta_1 \varepsilon_{t-1} - \cdots - \beta_t \varepsilon_{t-q}. \tag{4}$$

(Box, Jenkins, and Reinsel, 2008).

2.5 Forecasting Evaluation

The mean squared forecasting error (MSFE) is used to evaluate the forecasting performance from the ARMA model.

$$MSFE = \frac{\sum_{\tau=\tau_0}^{\tau} [Y_\tau - E(Y_\tau | Data_{\tau-h})]^2}{T - \tau_0 + 1} \tag{5}$$

Where $Data_{\tau-h}$ denotes the information available through a period $\tau - h$, where h is the forecast horizon, and $E(Y_\tau | Data_{\tau-h})$ is the point forecast of Y_τ (Koop & Korobilis 2009).

2.6 White noise

An important concept in forecasting is white noise. If a time series suffers from white noise, it cannot be predicted since the sequence is purely based on random numbers. On the other hand, the errors from the forecasts has to be white noise, otherwise the predictive model needs improvement.

White noise is explained as if different values are collected at various moments and times, and these values are not correlated with each other. In other words, white noise is a stationary random process that has zero autocorrelation. Box and Pierce (1970) developed a test for white noise called the portmanteau test, later refined by Ljung and Box (1978). The test is conducted by testing for autocorrelation in the residuals of a model.

$$Q = n(n+2) \sum_{j=1}^{m} \frac{1}{n-j} \hat{p}(j) \rightarrow \chi_m^2 \tag{6}$$

This relies on the fact that if $x(1), \dots, x(n)$ is a realization from a white noise process. As well as m indicated the number of autocorrelations that are calculated and shows the convergence in distribution to a chi-squared distribution.

3. Data

3.1 The data set

The range of the data is from January 2000 to December 2019. Which in total consists of two hundred forty monthly distinct time periods. Monthly data is selected according to Aye et al. (2015). With monthly data, returns are at least approximately normally distributed compared to more high-frequency data. The range of the data set is also based on Aye et al. (2015) but also includes all recent available data. This period covers the global financial crisis of 2008, the oil crisis of 2014, and the beginning of the COVID-19 pandemic. However, the data does not cover the after works of the pandemic, the start of the Ukraine war 2022, and the increasing interest rates from the Fed. The start and end points of the sample are purely chosen by the availability of the data and primarily the availability of data from China. The entire data set comes from secondary sources, shown in table 1. The data was collected from each source's own database. Nonetheless, all the data and information collected are from highly credible sources, such as FRED, OECD, Bank of England, etc. and peer-reviewed scientific articles, indicating that the articles maintain academic standards.

3.2 Variables

Thirty-two different variables are used in this study, presented in table 1. This broad number of variables and potential predictors can cause significant statistical problems in the model selection. Therefore, principal component analysis is used a data pre-processing technique to reduce the number of parameters to be estimated. Twenty-nine of the variables are merged into seven categories. The business cycle, nominal, interest rate, commodity price, exchange rate, stock market, and government bond yield factor. For each category, one principal component is extracted from several different variables. For example, in the business cycle factor, one principal component is extracted from the industrial production index from the United States, China, Europe and Japan. In other words, these four different variables are merged into one variable, the business cycle factor, by principal component analysis. This process does not capture the entire variation since all factors are latent. Nevertheless, the blocks of the variables are still more suitable to capture the common component of the variables compared to their parts alone.

Pierdzioch et al. (2014a) have shown that six categories of different global variables significantly impact the price of gold. All the different countries, indexes and financial variables are selected according to Aye et al. (2015). They based their selection of countries, indexes, and financial variables, according to economic relevance, representation of major world economies, and variables that strongly sway commodity markets. The first category of variables is the business cycle factor. This category is represented by the industrial production index from the two largest economies, the United States and China, and an index for Europe and Japan, respectively. Pierdzioch et al. (2014b) have found that the primary economic activity, in the form of an international business cycle, has a significant predictive power of gold price changes. The nominal factor is the next category of variables, which refers to inflation. The CPI represents inflation for the United States, EU, China, and Japan. As well as the money stock for Europe and the United States. As mentioned earlier, the price of gold increases as inflation rises. The third variable is the interest rate factor which includes the policy rate in Japan, the United Kingdom, and the United States. This is because Pindyck & Rotemberg (1990) found that gold prices are sensitive to changes in interest rates. Also, when other commodity prices tend to fluctuate, then gold prices are influenced by this movement. In a commodity boom, if the prices rise, then gold co-moves. As a result, a commodity price factor is represented by a brent oil index, two crude oil indexes, silver spot price, and an all-commodity price index. Moreover, Sjaastad (2008) shows that gold prices are heavily affected by changes in foreign exchange markets. This is explained by gold being considered a currency as it is often a part of foreign reserves. If the exchange rate falls, traders seek safer investments such as gold because of the law of one price. Aspergis (2014) also found that gold price has a positive relationship with the nominal and real exchange rates when investigating the Australian dollar/US dollar. Thus, an exchange rate factor is used that includes the US effective exchange rate, the Yen/Dollar exchange rate, and the Dollar/Euro exchange rate. Another critical factor category that has an impact on gold returns are stocks. Stocks and gold frequently co-move together, but this may depend on the business cycle. To ensure this is taken into account, four stock indexes are used. The 600 biggest companies from Asia, the EU and North America are included, as well as a stock index from MSCI emerging markets. Finally, Aye et al. (2015) demonstrated that financial variables significantly influence gold returns over real economic variables. Therefore, an additional financial variable is included, represented by a government bond yield category from the United States, Japan, and EU. These countries are selected to match the economic relevance and representation of major world economies according to previous variables.

In addition, Baur & Lucey (2010) have shown that the Kansas City Fed's financial stress index and the U.S. economic policy uncertainty index have impacted the gold price returns. This is mainly because these variables deal with fear and anxiety (Ciner et al.,.2013). Since gold is a safe haven, these individual variables show the importance of predictive power for gold price. These two variables are not part of the PCA factors.

3.3 Transformations

To achieve stationary for the different variables, some variables are transformed. The transformations are based respectively on Aye et al. (2015) and Stock & Watson (2005). No transformation is used for the Economic Policy Uncertainty Index for the United States since the individual series is already considered stationary. However, the first difference is required for the government bond yield and the Kansas City Fed's Financial Stress index. For the rest of all the variables, first, the logarithmic variable is generated, and afterward the first difference of the logarithmic variables is created. In the table below, the transformation code 1 is for no transformation, 2 is the first difference and 3 is the first difference of the logarithmic variable.

Table 1. Variables used in the study

Category	***Tcode***	***Description***	***Source***
Business cycle factor	3	EU industrial production	OECD
	3	China industrial production	FRED
	3	US industrial production	FRED
	3	Japan industrial production	FRED
Nominal factor	3	EU CPI	FRED
	3	China CPI	FRED
	3	US CPI	FRED
	3	Japan CPI	FRED
	3	Money stock US	FRED
	3	Money stock EU	FRED
Interest rate factor	3	UK policy rate	Bank of England
	3	EU policy rate	ECB
	3	US policy rate	FRED
	3	Japan policy rate	Ministry of Economy, Trade and Industry
Commodity price factor	3	All commodity price index	FRED
	3	Midland Texas crude oil	FRED
	3	Global price Dubai brent oil	FRED
	3	Global price Dubai crude oil	FRED
	3	Silver fixing price (US)	London Bullion Market Association
Exchange rate factor	3	US exchange rate	FRED
	3	Yen/Dollar	FRED
	3	Dollar/Euro	FRED
Stock market factor	3	STOXX Europé 600	STOXX
	3	STOXX North America 600	STOXX
	3	STOXX Asia 600	STOXX
	3	MSCI emerging	MSCI
Government bond yield factor	2	10 years government bond yield US	FRED
	2	10 years government bond yield EU	FRED
	2	10 years government bond yield Japan	FRED
KCFSI	2	The Kansas City Fed's Financial Stress Index	The Kansas City Fed
EPU US	1	Economic Policy Uncertainty index for United States	Scott Baker, Nicholas Bloom and Steven J. Davis
Gold	3	Gold fixing price (US)	London Bullion Market Association

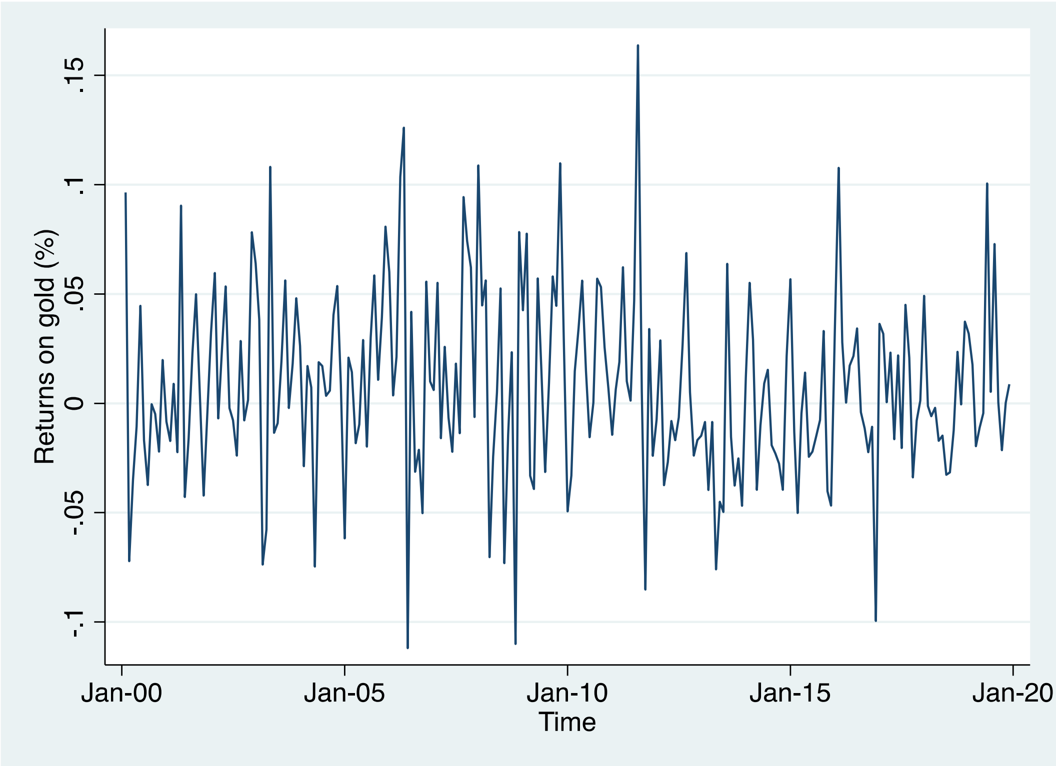

Figure 2. Gold returns (in percentage)

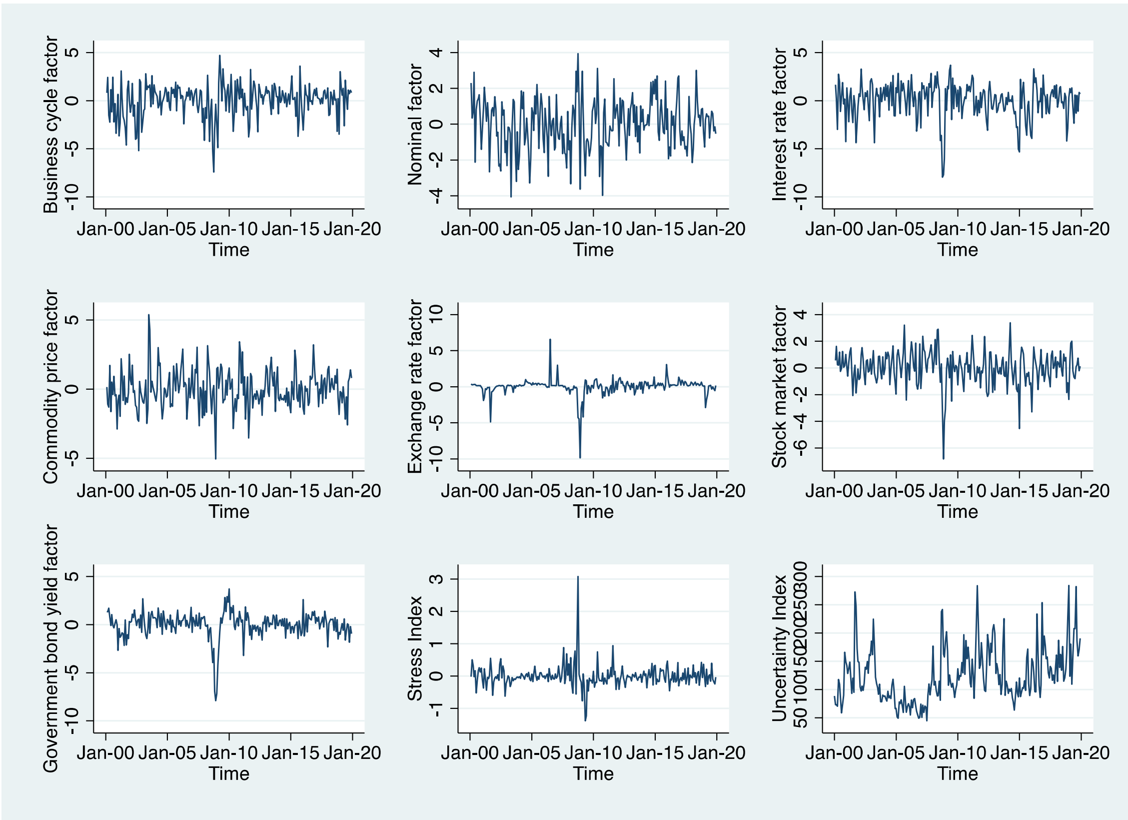

Figure 3. Independent variables

4. Implementation and Results

4.1 Course of action

After the factors are extracted from the principal component analysis, the model selection begins. Now three different ARMA models are computed to forecast the gold returns. Model (A), that includes all the predictors. Furthermore, these predictions are compared to a benchmark model, represented by an AR (1) process of returns on gold, called model (B). This is done to ensure that the forecast outperforms a linear model such as the benchmark. Lastly, a forecast model called (C), excluding the government yield factor component, is compared to the models to test whether financial variables have more substantial predictive power, as Aye et al. (2015) claim.

When fitting the model to forecast, it is essential to specify the optimal lag length. The choice of lag length is often discussed in published literature. Swanson (1998) emphasized the importance of lag length choice based on information criteria. However, this advice has often been ignored in more recent studies according to Shi et al. (2020). Nowadays, many studies choose their lag length arbitrarily, fixing the number of lags at 6 or 12 for simplicity. A modern approach is to use 6, 12 or 24 lags for monthly data, given sufficient data points. In this study, the lag order does not vary across factors and variables. The selected lag order is 6 for both autoregressive and moving averages and is applied to all models, variables and in all subsamples, except for the benchmark model.

The approach for forecasting the returns on gold is based on monthly data on the predictors and applying the sample period from January 2000 to December 2019. Firstly, the in-sample period with the first 120 observations is predicted, where the predicted value refers to the value of Y_t predicted for an observation in the entire sample. Then secondly, the out-of-sample is forecasted, where the forecasted values refer to the value Y_t forecasted for an observation not in the sample but based on previous observations. This is done to evaluate which model has the best forecast performance. By comparing the forecasted values, based on the first 120 estimations, with the real values of the last 120 observations. The forecasting is based on a multi-step ahead method, where the first 120 observations are used to fit the models and then making predictions of the last 120 observations. Lastly, all available data from January 2000 to December 2019 is used to forecast one, six, and twelve months preceding the data period. The results from the forecast are analysed to determine whether gold yields positive returns.

4.2 Results

Firstly, diagnostic testing for the principal components is done. The Kaiser-Meyer-Olkin measure of sampling adequacy takes a value between 0 and 1, which indicated how well the data is suited for the factors in the principal components. A rule of thumb is that if the value is above 0.5, the principal components are justified to be used (Kaiser 1974). As shown in table 2, all values are above 0.5, so it is adequate to use. However, one can observe that the data in the commodity price factor is the most suitable.

Table 2. Kaiser-Meyer-Olkin measure of sampling adequacy

KMO Overall	***Factor***
0.692	Business cycle factor
0.541	Nominal factor
0.573	Interest rate factor
0.814	Commodity price factor
0.572	Exchange rate factor
0.735	Stock market factor
0.588	Government bond yield factor

Table 3 shows how much of the total variation is explained by the first principal component. The stock market factor has the highest total variation explained by 78.8%, and the nominal factor has the lowest with 27.1%. If additional principal components were to be included, the explanatory power would increase. However, according to Aye et al. (2015), the literature standard is to create the factor based on the first principal component of each block of variables. Hence, it was concluded that only the first component would be used for all the factors.

Table 3. Variation explained by PCA

Variation explained	***Factor***
0.460	Business cycle factor
0.271	Nominal factor
0.456	Interest rate factor
0.693	Commodity price factor
0.672	Exchange rate factor
0.788	Stock market factor
0.637	Government bond yield factor

Afterwards, an augmented dickey fuller test is used to test for unit root for the variables. In table 4, it can be observed that none of the variables has a unit root and therefore conclude that they are all stationary.

Table 4. Augmented dickey fuller for the variables

Variable	***Test statistic***	***P-value***
Gold returns	-5.214	0.000
Business cycle factor	-5.082	0.001
Nominal factor	-6.026	0.002
Interest rate factor	-5.984	0.003
Commodity price factor	-7.064	0.004
Exchange rate factor	-3.949	0.005
Stock market factor	-6.043	0.006
Government bond yield factor	-4.762	0.001
Stress index	-3.385	0.009
Uncertainty index	6.190	0.000

Some descriptive statistics are presented in table 5. The values are incredibly high for the uncertainty index compared to the other variables. This is because they are the only variables that are not transformed.

Table 5. Descriptive statistics for the variables

Variable	***Mean***	***Std.dev***
Gold returns	0.008	0.043
Business cycle factor	-5.09E-10	1.776
Nominal factor	2.07E-09	1.419
Interest rate factor	-3.24E-09	1.862
Commodity price factor	-6.12E-10	1.382
Exchange rate factor	-1.87E-10	1.169
Stock market factor	-1.75E-09	1.274
Government bond yield factor	2.61E-10	1.356
Stress index	-3.85E-03	0.343
Uncertainty index	125.729	48.491

In addition, the Portmanteaus test for white noise is performed for all the predicted residuals in all three models. The residuals are the difference between the data and the fitted model. Therefore, the residuals should be white noise to display a good fit for the model. The conclusion from the tests, as shown in tables 6, is that the residuals follow a white noise process in all cases. Since the residuals are random and no general pattern exists. This suggests that if various residuals are gathered at different moments and times, there will not be correlated with each other, which implies that there exists no autocorrelation between the residuals. This shows that the predictive model has good forecasting abilities.

Table 6. Portmanteaus test for the residuals

	Portmanteau (Q) statistic	***Prob > chi2***
Model A	23.448	0.983
Model B	40.086	0.446
Model C	28.970	0.902

To analyze the forecast performance of each model, the mean squared forecast error (MSFE) in percentage is used. The MSFE is compared between all the three models for the last 120 observations. The model with the lowest MSFE has the best forecasting ability since the fitted values are closest to the real data. By observing tables 7, one can see that model A and model C has the same MSFE and that benchmark model has the highest MSFE. However, before rounding up, model A had a slightly lower value. This shows that the forecasting models containing predictors extracted from the principal components analysis outperformed the benchmark model.

Table 7. Mean squared forecasting error

MSFE	***Mean***	***Std.dev***
Model (A)	0.169	0.284
Model (B)	0.178	0.301
Model (C)	0.169	0.284

By observing figure 5, the benchmark model does not capture the returns of gold very well. The benchmark is purely set near zero for all the entire interval. This could be due to the benchmark model being an AR (1), which is purely based on its past value and therefore becoming near zero. For model A, shown in figure 4, the predicted values follow the real observation from the data set very well. When the real observation increases, the prediction tends to follow this and vice versa. However, the prediction tends to underfit the outliers. It is difficult to capture spikes when forecasting; therefore, the predicted values are centered around the mean and display a lower variation. When computing the twelve months ahead forecasts preceding the data period, there are no real observations to compare the forecasted ones with. Therefore, it is hard to know how accurate the estimates are. However, it is a good sign that the prediction is volatile and not just in a straight line compared to the benchmark model. By inspecting the model without the government bond yield factor in figure 6, the graphs look very similar to model A in figure 4.

Table 8 shows that the values for models A and C are almost identical. Mars 2020 is the month where the gold return is forecasted to be the highest, 0.014 for model A and 0.013 for model C. The lowest value for gold returns in the sample is in May 2020, which is 0.004 for model A and 0.003 for model C. On the other hand, for the benchmark model when rounding up, the values are set close to 0.008 for all the precited months. In other words, there are no negative returns in the forecasted sample period twelve months ahead.

Table 8. Twelve months ahead forecast

Date	***Model A***	***Model B***	***Model C***
20-Jan	0.011	0.008	0.012
20-Feb	0.006	0.008	0.006
20-Mar	0.014	0.008	0.013
20-Apr	0.013	0.008	0.013
20-May	0.004	0.008	0.003
20-Jun	0.004	0.008	0.004
20-Jul	0.013	0.008	0.012
20-Aug	0.007	0.008	0.007
20-Sep	0.005	0.008	0.005
20-Oct	0.012	0.008	0.012
20-Nov	0.010	0.008	0.011
20-Dec	0.004	0.008	0.004

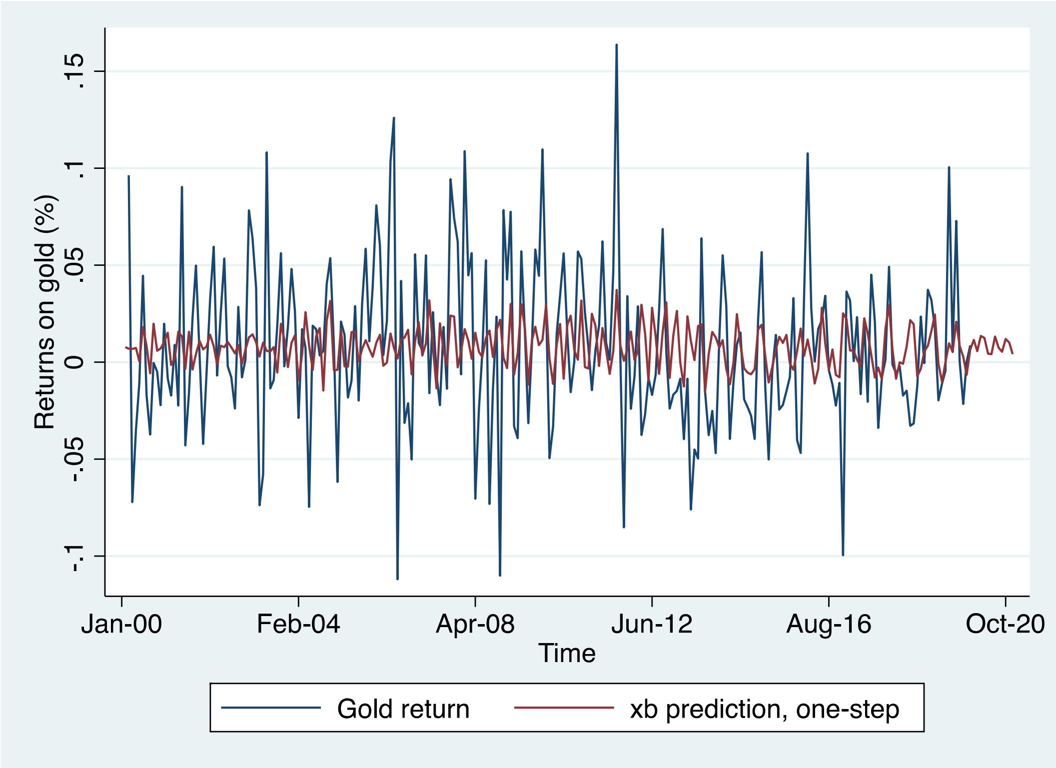

Figure 4. Gold return and forecasted gold returns for model A

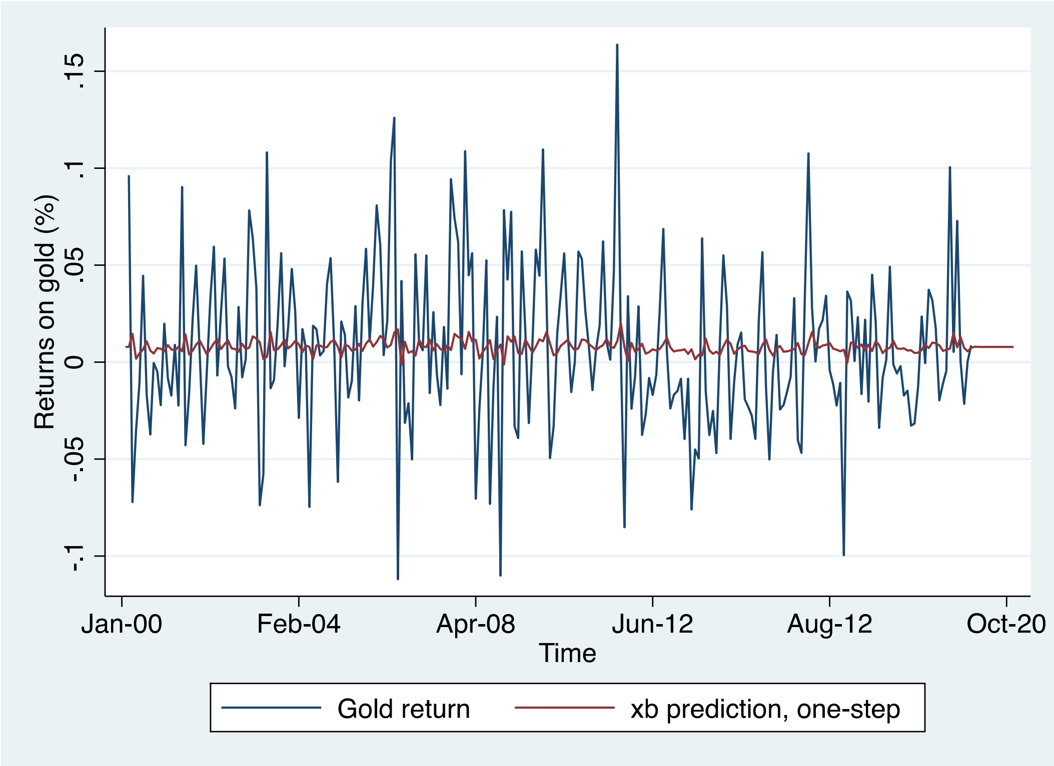

Figure 5, Gold returns and forecasted gold returns for model B

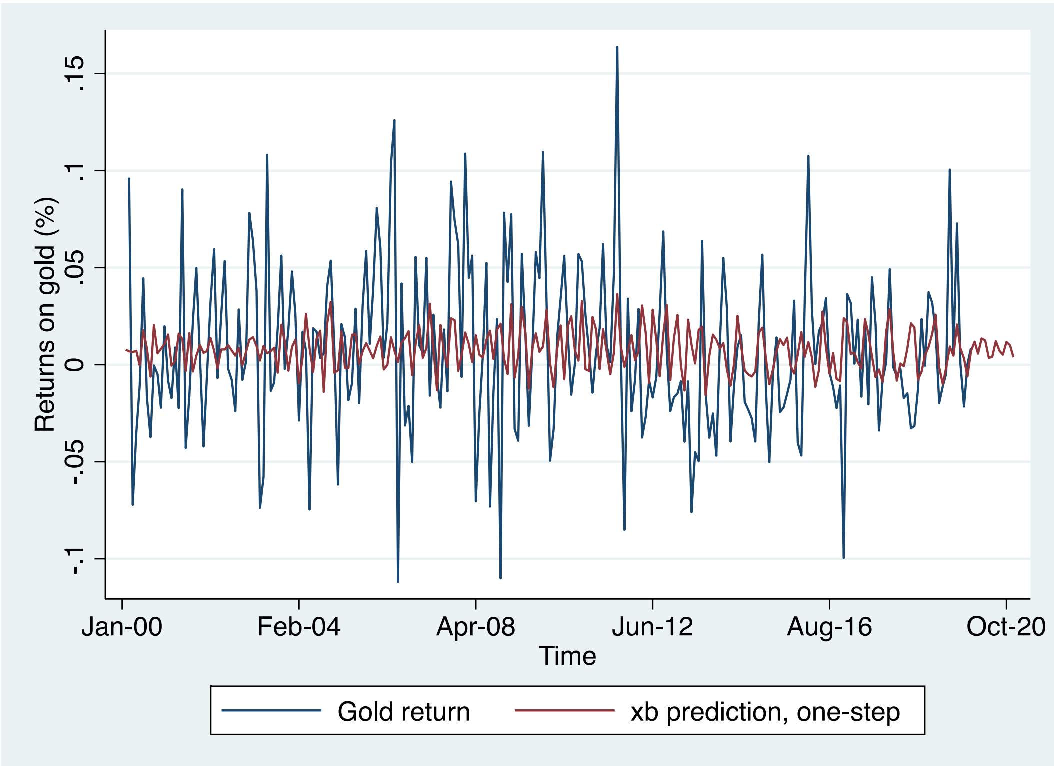

Figure 6 Gold return and forecasted gold returns for model C

Tables 9-11 shows the coefficients and lags from the three different models. This is done to compare the strength of the effect of each explanatory variable for gold returns. Table 9 shows how the first three and last lag of the AR part are positively correlated with gold returns, while the fourth and fifth lags are negatively correlated. For the MA part, the opposite lags are negative and vice versa. However, only the second lag of the AR part and none of the MA part is significant at the 5% level. For the coefficients, only the business cycle, nominal, interest rate and stock market factor are statistically significant. This implies that an increase in the interest rate factor by one percentage unit increases the gold returns by 0.005 percentage units. When the nominal, stock market and business cycle factor increase by one percentage unit, then gold returns decrease by -0.011, -0.004 and -0.005 percentage units, separately.

Table 9. Coefficients from an ARMA (6,6) for model A

Variable	***Coefficients***	***Std.err.***	***P-value***
Business cycle factor	-0.005	0.001	0.000
Nominal factor	-0.011	0.002	0.000
Interest rate factor	0.005	0.002	0.004
Commodity price factor	-0.004	0.002	0.079
Exchange rate factor	0.004	0.002	0.866
Stock market factor	-0.004	0.002	0.034
Government bond yield factor	4.09E-03	0.002	0.862
Stress index	7.77E-04	5.53E-05	0.160
Uncertainty index	0.010	0.009	0.298
Constant	-0.002	0.007	0.800
AR			
L1	0.266	0.713	0.709
L2	1.257	0.278	0.000
L3	0.282	0.747	0.706
L4	-0.978	0.682	0.152
L5	-0.539	0.312	0.084
L6	0.479	0.596	0.422
MA			
L1	-0.195	252.447	0.999
L2	-1.403	317.981	0.996
L3	-0.367	N/A	N/A
L4	1.194	509,631	0.998
L5	0.688	331.887	0.998
L6	-0.664	327.891	0.998

By observing table 10, it shows that the coefficient for the benchmark model is significant. The AR part is positively correlated with the gold return; however, it is not significant at the 5% level. While comparing models A and C, the outcomes are very similar. Expect that the stock market factor is insignificant and that the values differ slightly. However, when it comes to the AR lags, all expect the second lag is positive, but it is only the first lag that is significant. For the MA part, none of the lags are significant.

Table 10. Coefficients from an AR (1) for model B

Variable	***Coefficients***	***Std.err.***	***P-value***
Constant	0.008	0.004	0.010
AR			
L1	0.077	0.057	0.177

Table 11. Coefficients from an ARMA (6,6) for model C

Variable	***Coefficients***	***Std.err.***	***P-value***
Business cycle factor	-0.004	0.001	0.006
Nominal factor	-0.011	0.002	0.000
Interest rate factor	0.006	0.001	0.000
Commodity price factor	-0.004	0.002	0.066
Exchange rate factor	4.39E-04	0.019	0.819
Stock market factor	-0.002	0.002	0.238
Stress index	0.001	5.31E-04	0.114
Uncertainty index	0.011	0.091	0.239
Constant	-0.003	0.007	0.724
AR			
L1	1.986	0.696	0.004
L2	-1.091	1.679	0.516
L3	0.605	1.896	0.750
L4	1.118	1.895	0.555
L5	0.721	1.429	0.614
L6	0.287	0.475	0.546
MA			
L1	2.125	244.094	0.993
L2	1.255	N/A	N/A
L3	-0.712	57.044	0.990
L4	-1.259	125.285	0.992
L5	-0.609	37.310	0.987
L6	-0.193	26.710	0.994

5. Discussion

5.1 Analysis

When investigating the relationship between gold and key influencing factors, the interest rate, exchange rate, government bond yield, stress index, and uncertainty index factor all seem to contribute to increasing gold returns. On the contrary, the business cycle factor, nominal factor, commodity price factor, and stock market factor tend to have a negative effect on the return of gold. However, only the business cycle, nominal, interest rate and stock market factor demonstrated significant p-values in model A. This does not support what Aye et al. (2015) showed, that financial variables have more substantial predictive power for gold returns than real economic variables. In this study, more real economic variables showed predictive power compared to financial variables.

With the financial variables, one can assume that the interest rate factor would not have a particularly strong predictive power during the great recession in 2008 since the U.S economy entered a liquidity trap, which was explained by a zero-lower bound, where the short-term interest rate did not respond to monetary policy changes. However, in this study, the interest rate factor positively impacted the gold returns. This suggests that gold is sensitive to changes in interest rates, which Pindyck & Rotemberg (1990) claims. When it comes to the stock market factors, Aye et al. (2015) argues that stocks and gold frequently co-move together. In this case, the stock market factor had a negative effect on the return of gold, contradicting this.

On the other hand, when it comes to the real economic variables, they showed negative impact on gold returns. The business cycle factor was negative and contradicted what Aye et al. (2015) found. They argue that the business cycle power was very strong through the subperiod between 1999 and 2005, mainly because China's increase in its industrial production led to an increasing influence on the prices of commodities, such as gold. However, a commodity boom appeared due to the rising emerging markets in the early 21st century. The commodity price boom may have triumphed the business cycle. This could explain how the business cycle factor did not show strong predicting power. The nominal factor also had a negative effect on the return of gold. This contradicts what was mentioned earlier, that if inflation rises, the price of gold increases. Showing some evidence that gold might not be a good hedge against inflation.

One of the main results from the various diagnostic tests in the empirical part is that model A, containing all predictors extracted from the principal component analysis outperformed the other two models when evaluating the MSFE. This implies that the government bond yield factor did improve the model's performance, which confirms what Aye et al. (2015) claimed, that financial variables have the most potent predictive power for gold returns. Therefore, by increasing the number of financial predictors in the model, one would expect a more significant outcome. However, the result from this study might have been very different if another financial variable had been chosen instead of the government bond yield factor. On the other hand, Dichtl (2019) argues that even though the mean squared forecast error is an excellent statistical measurement for forecast accuracy, it does not imply that the forecast is necessarily correct.

Another conclusion from the twelve months ahead forecast was that none of the values were negative among this subsample within all three models. Therefore, one could argue that gold would be a good investment since it only yields positive returns. The forecasted returns are also within a small range between 0.013 and 0.003. This shows that the gold returns are predicted not to be very volatile since the interval is small. Therefore, one could support Ghazali, Lean, & Bahari (2013) 's claim that investing in gold is a hedge against unstable economic periods.

However, the former paragraph should be looked upon with some caution. The forecast is only made twelve months ahead, a short period. Therefore, making assumptions about the future is very speculative due to the limited size of the sample. Moreover, returns are typically believed to be very unpredictable. In figure 2, the graph displays how the return of gold is mean reverting, suggesting that in the long run, the returns will sooner or later revert to the mean, or in other words, the average level of the dataset. Nevertheless, the results of this study are somewhat based on insignificant variables. The majority of the coefficients and lags from the ARMA models are not significant at 5%. Thus, making assumptions based on the model is not reliable.

With the insignificant result, it is not possible to draw any conclusion about whether data can show if there is a difference or not. The exact reason for these results is difficult to know. It could be the chosen lag length in the model selection. Instead of setting a fixed number of lags as six or twelve for simplicity, according to Shi et al. (2020), it may have been better to determine the lag length based on an information criterion. Swanson (1998) emphasized this importance and strongly advocated this methodology for models used to describe a high-dimensional reality.

Furthermore, the forecasting technique used may not have been the most optimal one. ARMA models are widely considered a very effective tool for predicting the future, although it is ideal for a univariate model, according to Enders (2015). Considering that a multivariate model with multiple predictors is being used to forecast, the outcome would have been potentially improved if another model had been used. Enders (2015) explains that a vector-autoregressive model is a forecasting model that excels when two or more time series influence each other.

When using an ARMA model to estimate multi-step ahead forecasting, the model allows for time-varying parameters. Aye et al. (2015) found that using time-varying parameters leads to poor out-of-sample forecasting when a large set of predictors are used since time-varying models usually over-fit the in-sample estimation. Although many extensions to these models are developed, such as VAR models, they still suffer from the same limitation. Therefore, according to Aye et al. (2015), a better alternative to forecasting a variable such as the returns of gold is by dynamic model averaging. Since gold is such a volatile variable, a dynamic model that combines the forecasts of a large number of dynamic linear models seems to be ideally suited when predicting the future value. Thereby, the model is allowed to change over time, and the coefficient in each model will evolve over time. Another problem when allowing the coefficients to be estimated unrestricted, is that the conditional variance may be negative. Hence a better estimation of the coefficients could be to use the maximum likelihood estimation instead.

Nonetheless, the nonsignificant result from the study could have been purely based on something different, despite the model selection and forecasting technique. One reason that Aye et al. (2015) received significant results with almost identical variables as predictors compared to this study could be that different sample periods were used. Therefore, this might cause the results of this study to turn out completely different than expected. Moreover, there could have been other outside factors that were impossible to control, which would explain the study's findings. It could also have been due to different measurement errors and biases. For instance, an error may have occurred when collecting the data for the study. Or some kind of selection bias, where the sample being used does not rightly represent a wider population. Hence, analyzing the data from various countries, and not only the countries with the most significant economies.

6. Conclusion

Selecting the optimal forecasting model and predicting future value has always seemed challenging. Gold returns are expected to be unpredictable since the commodity is sensitive to supply and demand shocks. The purpose of this study was to provide evidence on the models that most accurately predict gold returns and which variables can be used as good predictors, in order to determine future gold returns using a twelve-month forecast.

Through principal component analysis, seven economic factors were extracted. The forecasting models depended on these factors and two additional individual variables represented by the Kansas City Fed’s financial stress index and the U.S. economic policy uncertainty index. The ARMA model was used to predict future values with a multi-step ahead method, as it is widely considered one of the best forecasting tools. The sample period was between 2000 and 2019, which included both increases and decreases in the gold price, commodities booms, the great recession, and many changes in the business cycle.

The forecasting models containing predictors extracted from the principal components analysis outperformed the benchmark model in forecasting. However, it was difficult to find the variables that are good predictors. Only four of the predictors from the original model were significant, and the interest rate factor was the only one of these with a positive beta coefficient. When it comes to the forecasted twelve months ahead outside the dataset, they only yielded positive returns.

The results from this study contradict some previous research, which was expected since several predictors and lags were insignificant, and no firm conclusion could be drawn. However, some new insight about forecasting gold returns is provided that could be applied in future research. For further analysis, using a vector-autoregressive model may be more optimal when applying a multivariate model. On the other hand, instead of investigating the returns of gold, one could explore the variance and volatility around gold using GARCH models. Many economic time series often exhibit volatility clustering. Hence, forecasting volatility as a risk measurement instead of the returns would be just as important. Moreover, different sample periods and other variables could be applied to achieve better significance. Hopefully, this will be investigated further in the future as the problem of predicting gold returns is not yet fully solved.

7. References

7.1 Reference list

Apergis, N. (2014). Can gold prices forecast the Australian dollar movements? International Review of Economics & Finance, 29, 75–82.

Aye, Goodness & Gupta, Rangan & Hammoudeh, Shawkat & Kim, Won Joong. (2015). Forecasting the Price of Gold Using Dynamic Model Averaging. International Review of Financial Analysis. 41.

Banerjee, J. Origins of Growing Money (January 2013) (Available at forbesindia.com/printcontent/34515)

Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, Bonds and gold. Financial Review, 45(2), 217–229.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. Time Series Analysis: Forecasting and Control. 4th ed. Hoboken, NJ: Wiley

Box, G. E. P., and D. A. Pierce. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65: 1509–1526.

Brooks, C., 2014. Introductory Econometrics for Finance. Second edition, Cambridge University Press.

Ciner, C., Gurdgiev, C., & Lucey, B. M. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates. International Review of Financial Analysis, 29, 202–211.

CNBC (2022). *The Federal Reserve just raised interest rates 75 basis points. Here's what it means for the Main Street economy*. https://www.cnbc.com/2022/07/27/heres-what-the-fed-interest-rate-hike-means-for-main-street.html. (2022-08-10)

Dichtl, Hubert. (2019). Forecasting Excess Returns of the Gold Market: Can We Learn from Stock Market Predictions?*. Journal of Commodity Markets. 19.

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

Enders, W. (2015) Applied Econometric Time Series, 4th Edition, Wiley, Hoboken, NJ, 7-70.

Engel, C., & West, K. D. (2005). Exchange rates and fundamentals. Journal of Political Economy, 113(3), 485–517.

Ghazali, M. F., Lean, H. -H., & Bahari, Z. (2013). Is gold a hedge or a safe haven? An empirical evidence of gold and stocks in Malaysia. International Journal of Business and Society, 14(3), 428–443.

Gupta, R., Hammoudeh, S., Kim, W. J., & Simo-Kengne, B. D. (2014). Forecasting China's foreign exchange reserves using dynamic model averaging: The role of macroeconomic fundamentals, financial stress and economic uncertainty. North American Journal of Economics and Finance, 28, 170–189.

Ismail, Z., Yahya, A., & Shabri, A. (2009). Forecasting gold prices using multiple linear regression method. American Journal of Applied Sciences, 6(8), 1509–1514.

J.H, Stock & M.W, Watson. (2002). Forecasting Using Principal Components From a Large Number of Predictors. Journal of the American Statistical Association. 97. 1167-1179.

Jolliffe, I. T. 2002. Principal Component Analysis. 2nd ed. New York: Springer.

Kaiser, H. F. 1974. An index of factor simplicity. Psychometrika 39: 31–36.

Kiraz, Ş & Üstün, Esra. (2020). COVID-19 and force majeure clauses: an examination of arbitral tribunal’s awards. Uniform Law Review. 25.

Koop, Gary & Korobilis, Dimitris. (2009). UK Macroeconomic Forecasting with Many Predictors: Which Models Forecast Best and When Do They Do So?. Economic Modelling. 28.

Ljung, G. M., and G. E. P. Box. 1978. On a measure of lack of fit in time series models. Biometrika 65: 297–303.

O'Connor, F.A., Lucey, B.M., Batten, J.A., Baur, D.G., 2015. The financial economics of gold – a survey. Int. Rev. Financ. Anal. 41, 186–205.

Pierdzioch, C., Risse, M., & Rohloff, S. (2014a). On the efficiency of the gold market: Results of a real-time forecasting approach. International Review of Financial Analysis, 32, 95–108.

Pierdzioch, C., Risse, M., & Rohloff, S. (2014b). The international business cycle and gold-price fluctuations. The Quarterly Review of Economics and Finance, 54(2), 292–305.

Pindyck, R. S., & Rotemberg, J. J. (1990). The excess co-movement of commodity prices. The Economic Journal, 100, 1173–1189.

Shafiee, S., & Topal, E. (2009). When will fossil fuel reserves be diminished? Energy Policy, 37(1), 181–189.

Shi, Shuping & Hurn, Stan & Phillips, Peter. (2020). Causal Change Detection in Possibly Integrated Systems: Revisiting the Money–Income Relationship*. Journal of Financial Econometrics. 18. 158-180.

Sjaastad, L. J. (2008). The price of gold and the exchange rates: Once again. Resources Policy, 33, 118–124.

Stock, James & Watson, Mark. (2005). Implications of Dynamic Factor Models for VAR Analysis.

Swanson, N. R. 1998. Money and Output Viewed through a Rolling Window. Journal of Monetary Economics 41: 455–474.