Eksis: Jurnal Ilmiah Ekonomi dan Bisnis, 11(2), November 2020, 151-161
ISSN 2580-6882 (Online), ISSN 2087-5304 (Print), DOI 10.33087/eksis.v11i2.209

Multivariate Time Series in Macroeconomics

Ahmadi and R. Adisetiawan
Faculty of Economics, Batanghari University
Email correspondence: ahmadi.feunbari@gmail.com, r.adisetiawan@yahoo.co.id

Abstrak. Emas merupakan salah satu komoditas primadona dan alternatif investasi yang sangat populer. Harga emas diduga dipengaruhi oleh beberapa faktor lain seperti US Dolar, harga minyak, tingkat inflasi, dan bursa saham sehingga pemodelan harga emas tidak hanya dipengaruhi oleh nilainya sendiri. Penelitian ini dilakukan untuk menentukan model peramalan terbaik serta mengetahui faktor-faktor apa yang berpengaruh terhadap harga emas. Penelitian ini memodelkan harga emas secara multivariat dan mengkaji ulang pemodelan univariat yang akan dipergunakan sebagai model pembanding dari pemodelan multivariat. Pemodelan univariat dilakukan menggunakan model ARIMA dimana hasil pemodelan menyatakan bahwa fluktuasi harga emas sebagai white noise. Pemodelan harga emas secara multivariat dilakukan menggunakan Vector Error Correction Model dengan harga emas, minyak, indeks US Dolar dan Dow Jones, serta tingkat inflasi sebagai prediktor. Hasil penelitian menunjukkan bahwa model VECM telah mampu memodelkan harga emas dengan baik dan semua faktor yang diteliti berpengaruh pada harga emas. US Dolar dan harga minyak berkorelasi negatif dengan harga emas, sementara tingkat inflasi berkorelasi positif dengan harga emas. Indeks Dow Jones berkorelasi positif dengan harga emas hanya pada 2 periode.

Kata Kunci: multivariat time series, makroekonomi

Abstract. Gold is one of the most popular commodities and investment alternatives. Gold prices are thought to be influenced by several other factors such as the US Dollar, oil price, inflation rate, and stock exchange so that gold price modeling is not only influenced by its own value. This research was conducted to determine the best forecasting model and to find out what factors influence the price of gold. This research modeled the price of gold in a multivariate and reviewed the univariate modeling that will be used as a comparison model of multivariate modeling. Univariate modeling is done using ARIMA model where the modeling results state that gold price fluctuations as white noise. Multivariate gold price modeling is done using Vector Error Correction Model with gold, oil, US Dollar and Dow Jones indices, and inflation rate as predictors. The results showed that the VECM model has been able to model the gold price well and all the factors studied influenced the gold price. The US dollar and oil prices are negatively correlated with gold prices, while the inflation rate is positively correlated with gold prices. The Dow Jones index was positively correlated with gold prices in just two periods.

Keywords: multivariate, US Dollar, oil price, inflation rate, stock exchange, gold price

INTRODUCTION

Gold investment can be done in various forms such as gold bullion, gold jewelry, gold coins, gold certificates, gold company shares, and so on. Many people even almost every country in the world keeps its wealth in the form of gold. The United States in 2015 held a gold wealth of 8,133.5 tons, followed by Germany storing 3,381 tons of gold, Italy 2,451.8 tons, France 2,335.7 tons, and China depositing 1,797.5 tons of gold (Holmes, 2016). People's tendency to keep their wealth in gold is due to the long-term increase in gold prices and gold's ability to cope with inflationary turmoil (Kusuma, 2015). However, this trend does not necessarily make gold a top choice in investing. Predictions or forecasts of gold exchange rates in the future are taken into consideration before making a decision to invest.

Shankari and Manimaran (2015) found the fact that the price of gold follows the form of randam walk where this research was conducted on the price of gold specifically in India. Both studies stated that the price of gold studied could not be modeled using only the gold data itself. The price of a commodity is undeniably influenced by other factors besides its own historical value such as the price of substitution goods, inflation, economic conditions, politics, and so on. Sepanek (2014) stated that the gold price is influenced by several factors including the US Dollar and inflation rate. Another factor that contributes to the gold price is economic and political conditions (Friend of Pawnshop, 2016). Gilroy (2014) stated that there is a reverse relationship between the price of gold and world oil. The influence of these factors can be modeled scientifically using multivariates that will be discussed in this study. Previous research has modeled the price of gold univariately which only uses the price of gold in modeling. The assessment of the influence of other variables can be done by doing multivariate modeling. Multivariate modeling traces and models the influence of other variables on the price of gold so that it is believed to be able to provide more accurate results.

This research modeled the price of gold in a multivariate and reviewed gold price modeling using univariate modeling that will be used as a comparison and

modeling multivariate. Multivariate modeling is done with the aim of obtaining a more accurate model and knowing the magnitude of the influence of other variables on the price of gold. Multivariate modeling of gold prices is done by including influential factors to get a better model. Factors to be analyzed in this study are the US Dollar, inflation rate, Dow Jones Index, as well as world crude oil prices.

Library Overview

Model Building Strategy

Finding the right model to model time series data is not easy. A strategy in building models has been put forward very well (Box & Jenkin, 1976). The strategy is divided into three main stages, namely: model specifications, model fittings, model diagnosis (Chan & Cyrer, 2008).

Stationarity

Stationarity is a very important assumption in time series data modeling. The basis of stationerity is that the laws of probability governing the conduct of the process do not change with time. A process (Y_t) said strictly stationar, if the joint distribution of Y_{t1}, Y_{t2},… Y_{tn} same as the joint distribution of Y_{t1-k}, Y_{t2-k},… Y_{tn-k} on all time options t_1, t_2,… t_n and time $lag\ k$. If n = 1, then the univariate distribution of Y_t equals Y_{t-k} in each t and k to $E(Y_t) = E(Y_{t-k})$ on each t and k so that the mean function is constant over time. The variance of Y_t has the same value as the variance of Y_{t-k} on each t and k so that the variance is said to be constant against time. The Y_t is said to be weakly stationary when the mean function is constant to time and covariance (Y_t, Y_{t-k}) = covariance (Y_0, Y_k) on each t and k.

Autocorrelation and Partial Autocorrelation

Autocorrelation is defined as a correlation between past and future values of a time series data that shows the linear relationship between time-separated k-lag timesheet data. The autocorrelation function is formulated as follows:

$$\rho_k = \frac{\sum_{t=1}^{N-k}(Y_i-\bar{Y})(Y_{i+k}-\bar{Y})}{\sum_{i=1}^{N-k}(Y_t-\bar{Y})^2} \quad (1)$$

Partial autocorrelation measures the impact of one autocorrelation on the other autocorrelation. Partial autocorrelation of $k - lag$ is the autocorrelation of periods i and $i + k$ where the periods in which they are ignored.

Augmented Dickey Fuller Test

Augmented Dickey Fuller Test is a type of testing to test the stationary of a time series. This test is done by regressing the first difference of a time series against the first lag of the time series and $k - lag$ of the first difference of a time series. The regression equation can be seen in (2).

$$Y_t - Y_{t-1} = \alpha Y_{t-1} + \emptyset_1(Y_{t-1} - Y_{t-2}) + \ldots + \emptyset_k(Y_{t-k} - Y_{t-k-1}) + e_t$$

$$\Delta Y_t = \alpha Y_{t-1} + \emptyset_1 Y_{t-1} + \ldots + \emptyset_k Y_{t-k} + e_t \quad (2)$$

The initial hypothesis on this test is α = 0 which means that the Y_t not stationer. The altenative hypothesis in this test is α < 0 which means Y_t stationary (Chan & Cryer, 2008).

White Noise

A very important example of stationary process is white noise process which is defined as a series of random variables that are independent and distributed randomly identically (e_t). White noise is usually assumed to have mean 0 and variance σ_e^2. White noise is often used to produce many useful processes (Chan & Cryer, 2008).

Information Criteria

Information criteria have proven effective in choosing a statistical model. There are several criteria proposed in the time series literature where all the criteria in principle are based on likelihood and contain two components. The first component deals with the goodness of fit of a model on the data, while the second component gives a greater penalty to more complex models. Some types of information criteria are as follows:

$$AIC = Ln|\widehat{\sum}e| + \frac{2}{T}(p + q) \quad (3)$$

$$BIC = Ln|\widehat{\sum}e| + \frac{ln(T)}{T}(p + q) \quad (4)$$

$$HQ = Ln|\widehat{\sum}e| + \frac{2\ln[\ln(T)}{T}(p + q) \quad (5)$$

$$\widehat{\sum}e = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{e}_t\,\hat{e}_t, \quad (6)$$

where T is sample size, p and q are the number of parameters, $\widehat{\sum}e$ is the estimated maximum likelihood of $\sum e$ (matric covariance error (e_t) and $\hat{e}_t\ \hat{e}_t$, is the residual and residual tranpose respectively on t. AIC is Akaike Information Criterion that stated Akaike (1973). BIC is Bayesian Information Criterion put forward Scwarz (1978) and HQ put forward Hannan & Quinn (1979) (Tsay, 2013).

Univariate Time Series

Moving Average Processes

Moving Average Processes is represented by $MA(q)$ where q is a parameter that determines the lag length of the model, and is formulated as follows:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad (7)$$

where e_t is the error term at the time t. this terminology arises from the fact that Y_t is obtained by charging 1, $-\theta_1$, $-\theta_2$, ..., $-\theta_q$ on the variable e_t, e_{t-1}, e_{t-2}, …, e_{t-1} and move and load e_{t+1}, e_t, e_{t-1}, e_{t-2}, …, e_{t-q+1} to get a value of Y_{t+1} and so on. This model was first introduced Slutsky (1927) and Wold (1938), Chan & Cryer, (2008).

Autoregressive Processes

Autoregressive Processes as the name implies refers to regressions performed on itself. Specifically, a $p-order$ of autoregressive process (Y_t) meets the equation:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots + \emptyset_p Y_{t-p} + e_t \quad (8)$$

The value of Y_t value against the combination of the previous value p is added with an "innovation" of e_t that combines everything new at the time of t that is not covered from the value in the past. Value e_t independent of $Y_{t-1}, Y_{t-2}, Y_{t-3}$ and so on. This model was introduced Yule (1926) and noted as $AR(p)$ (Chan & Cryer, 2008).

Mixed Autoregressive Moving Average

Mixed Autoregressive Moving Average Model is used if it is assumed that a series consists of part autoregressive and partial moving average. The model is formulated as follows:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + .. + \emptyset_p Y_{t-p} + e_t - \emptyset_1 e_{t-1} - \emptyset_2 e_{t-2} - .. - \emptyset_q e_{t-q} \quad (9)$$

where Y_t called mixed autoregressive moving average process and this model is noted as $ARMA(p,q)$. one of the assumptions that must be fulfilled by this model is that the series is stationer where the series has a constant mean and variance to time and the series is invertible. $ARMA$ was then developed to solve series problems with nonstationary mean called Integrated Autoregressive Moving Average ($ARIMA$) (Chan & Cryer, 2008).

Integrated Autoregressive Moving Average (ARIMA)

$ARIMA$ is a model to model d difference where $W_t = \triangle^d Y_t$ is an *ARMA* process that stationer W_t follow the model $ARMA(p,q)$ so that Y_t is a process $ARIMA(p,d,q)$. Parameter d commonly used in practice is 1 or at most 2. $ARIMA(p,1,q)$ will be formulated as follows: (Chan & Cryer, 2008):

$$W_t = \emptyset_1 W_{t-1} + \emptyset_2 W_{t-2} + .. + \emptyset_p W_{t-p} + e_t - \emptyset_1 e_{t-1} - \emptyset_2 e_{t-2} - .. - \emptyset_q e_{t-q} \quad (10)$$

Multivariate Time Series

Multivariate time series analysis reviews more than 1 variable (time series) and reviews and models the influence of each variable. Multivariate time series analysis is part of multivariate statistical analysis, but specifically deals with dependent data. The multivariate forecasting models that will be used in this study are vector autoregressive and vector error correction model.

Granger Causality

Granger Causality is a causality formulation presented by Granger (1969). The idea of Granger Causality is that if variable Y can be predicted better by using historical data variables Y and X compared to historical data Y only, then X is called Granger-Causes against Y. The second definition of causality expressed by Granger is that if $\sigma^2(X|U) < \sigma^2(X|U-Y)$ (variance of X predicted using universal data, U is smaller than the predicted X variance using all information except Y), so it can be said that Y Granger-Causes X. Feedback system occurs when variable X Granger-Causes Y and variable Y Granger-Causes X, noted $X \leftrightarrow Y$. The Granger Causality test equation can be seen in the equation (11).

$$Y_t = \sum_{j=1}^{m} \alpha_j Y_{t-j} + \sum_{i=1}^{n} \beta_i Y_{t-1} + D_t + \varepsilon_t \quad (11)$$

D_t is deterministics, ε_t is random error term, α_j is the coefficient of Y_t and β_i is the coefficient of X_t. The initial hypothesis of the Granger-Causality test was $\beta_i = \beta_2 = \ldots = \beta_n = 0$ (non-causality). The equation (11) is called the unrestricted model where there is a lagged value of another variable (X), while the restricted model is a model where there is only a lagged value of the dependent variable (Y). the effectiveness of granger testing is measured by minimizing Mean-Square Error, this test can be done on stationer data (Tsay, 2013).

Toda and Yamamoto Test

Toda and Yamamoto developed an alternative test of the Granger Causality test known as the Toda and Yamamoto Augmented Granger Causality (1969). This test can be used on data that is not stationer at the integration level regardless of $l(m)$ as well as on co-integrated data or not. This test is based on the following equation:

$X_t = \sum_{i=1}^{h+d} \beta_i X_{t-1} + \sum_{\mathrm{j}=1}^{k+d} \alpha_{\mathrm{j}} Y_{t-\mathrm{j}} + \varepsilon_t$ (12)

The equation (12) d represents the maximum order of integration, while h and k are the optimal lag length obtained from the initial hypothesis information criteria of this test is H_0: $\sum_{i=1}^{h+d} \alpha_{\mathrm{j}} = 0$ or Y not Granger Causes X. (Josheski & Bardarova, 2013)

Vector Autoregressive Model

This model is the most commonly used multivariate time series model especially in econometrics, because it is relatively easy to estimate and its properties have been widely studied in various literatures. This model is noted as $VAR(p)$ and is modeled as follows:

$Y_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + e_t$ (13)

$\emptyset_0$ is a k-dimensional vector and $\emptyset_i$ is a $k\ x\ k$ matrix for $i > 0$, $\emptyset_p \neq 0$, e_t is an identical and independent array of positive random vectors with the mean 0 and covariant $\sum e$. This model can be done if the time series data to be used is stationer (Tsay, 2013).

Cointegration

Time series data that is not stationer can contain cointegration. Non stationer data that does not contain cointegration can be modeled using VAR by first making a difference. Non stationer time series data containing cointegration cannot be modeled using VAR, as it will produce spurious regression. The data can be modeled using the Vector Error Correction Model.

$\beta' Y_t = \beta_1 Y_{it} + \ldots + \beta_n Y_{nt} \sim I(0)$ (14)

$Y_t = (y_{it}, \ldots, y_{nt})$' presented $(n\ x\ 1)$ an integrated vector in the first order $(l(1))$. Y_t is called integrating when there are $(n\ x\ 1)$ vectors $\beta = (\beta_1, \ldots, \beta_n)$' and linear combinations of Y_t and β stationer or $I(0)$ vectors. Vector β is also called cointegrating vector. If $(n\ x\ 1)$ the Y_t vector is integrated, there will be $0 < r < n$ rank of cointegration. Testing of r value can be done using Johansen procedure (Wang & Zivot, 2006).

Vector Error Correction Model

Time series data that are non stationer and integrated can be modeled using Vector Error Correction Model as in the equation (15).

$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$ (15)

ΔY_t is a vector $(n\ x\ 1)$ difference of a time series data, Π is a matrix $(n\ x\ n)$ which is the result of multiplication between matrix $(n\ x\ r)$ adjustment coefficient (α) with matrix $(r\ x\ n)$ vector integration β. (Wang & Zivot, 2006).

Johansen Procedure

Johansen simulates a likelihood ratio (LR) statistics to determine the rank of the Π matrix, this test is based on the estimated value of the eigen value of the Π matrix. The initial hypothesis of this test is that there is no cointegration $(r = 0)$ in the data and the alternative hypothesis of this test is that there is cointegration in the data $(r > 0)$. The statistical value calculation of this test is calculated using the equation (16) and will then be compared with critical value. Statistical values smaller than critical value state that the initial hypothesis is accepted.

$LR = -T \sum_{r_0+1}^{n} ln(1 - \hat{\lambda}_i)$ (16)

Rejection of the initial hypothesis indicates that there is a cointegration $(r > 0)$, so the test is continued by testing whether the value $r = 1$ and so on until the initial hypothesis is rejected (Wang & Zivot, 2006).

Diagnostic Test

Diagnostic test is a test to see if the model built is valid so that it can be used to do a forecast. A model is said to be valid if the residuals of the model are independent, identical, and normally distributed. Diagnostic test in multivariate time series model is done by doing Portmanteau test, $ARCH - LM$ test, Jarque-Bera Test. Portmanteau testing has an initial residual hypothesis not serially correlated (independent). $ARCH - LM$ test aims to see if residuals have been identical to the initial residual hypothesis has been identical. Jarque-Bera test is a normality test with normal distributed residual initial hypothesis with no skewness and has a kurtosis value of 3.

Impulse Response Function

Impulse Response Function (IRF) is an approach to looking at relationships between variables. Impulse response is a dynamic function to track the effect of a shock that occurs on a variable against other variables. IRF calculates the amount of change that will be experienced by all variables in $t > 0$ if there is a shock change in 1 variable at the time of $t = 0$. The assumption in IRF calculation is $E(Z_t) = 0$, because mean has no effect on the movement of the Y_t by shock. Mathematically, the purpose of irf is to find out the effect of y_{it} changes on $y_{t+\mathrm{j}}$ for $j > 0$, where each other element does not change so that it can be assumed that $t = 0$, $y_t = 0$ for $t \leq 0$, and $\alpha_0 = (1, 0, 0, \ldots)$'. The value y_t, $t > 0$ can be searched using the $MA(\theta)$ representation of the $VAR(p)$ model (Tsay, 2013).

$$\alpha_0 = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, y_1 = \theta_1 \alpha_0 = \begin{bmatrix} \theta_{1.11} \\ \theta_{1.21} \\ \cdot \\ \cdot \\ \theta_{1.k1} \end{bmatrix}, y_2 = \theta_2 \alpha_0 = \begin{bmatrix} \theta_{2.11} \\ \theta_{2.21} \\ \cdot \\ \cdot \\ \theta_{2.k1} \end{bmatrix} \quad (17)$$

METHOD

The literature study process contains study activities independently in order to find information that supports research through books and journals. This process aims to find an overview of the research to be done. This process will also be the basis for determining the research gap. This process aims to find gaps to be achieved in research. Gap search is done by conducting research on related research and identifying things explored from previous researches. The process of collecting data in this research was carried out by taking the data needed in this research, namely historical gold data as well as factors that affect gold prices such as stocks, oil, US Dollar, and inflation rate. Data in the form of monthly data in the last few years. The available data is then processed so that it can be used in this research. Data analysis in this study, among others: univariate time series modeling, diagnostic test, multivariate time series modeling, measuring the performance of forecasting, and drawing conclusions.

RESULTS AND DISCUSSIONS

Multivariate Time Series Modeling

Multivariate time series data modeling is a modeling that uses data from several other variables to predict the value of a variable. Multivariate modeling is done because in real life, everything must have a relationship with something else. Models of multivariate modeling used in research are Vector Autoregressive Model and Vector Error Correction Model where the selection of methods is adjusted to the characteristics of the data. Gold price multivariate modeling is gold price modeling with gold prices and other variables that are influential as predictors.

The study will try to use the U.S. Dollar index, Dow Jones Index, oil prices, and inflation rates in modeling gold prices. The variables that will be used to model the price of gold are only variables that are tested to affect the price of gold. The test will be conducted using Granger Causality test. Granger Causality testing aims to find out if the historical value of a variable can help forecasting from other variable values. The selection of variables to be tested on the modeling of a variable (in this study gold) can not be done carelessly, but rather based on the strong suspicion that the variables have an influence on the value of gold. The variables to be tested to model gold include the US Dollar index, the Dow Jones index, crude oil prices, and the inflation rate in the United States. The US dollar is believed to be a target commodity for investors as opposed to gold. The slump in the value of the US Dollar will make investors switch to buying gold so that the price of gold increases (Seputar Forex, 2016). The Dow Jones index is one of the largest stock market indices in America and the world that measures the performance of industrial components in the American stock market (Seputar Forex, 2018). Stock indexes can represent market conditions influenced by economic, political, social and so on factors. World oil prices are closely related to inflation (Jannah, 2015) where rising inflation rates will encourage investors to buy gold to maintain their purchasing power.

Granger Causality testing of these variables was conducted using the Toda and Yamamoto methods. This test is chosen because it can be used to test data that is not stationary, and can also be applied to data that integrates. This test uses Vector Autoregressive models to help prove the existence of causality where the lag to be used is $(p + q)$. Parameter p is obtained from information criteria such as $AIC, BIC,$ etc. Parameter m is the highest value of the number of differences of all variables until it is stationary. Parameter m obtained by doing ADF test on all variables to be tested. ADF test results can be seen in Table 1.

Table 1
ADF Test Results

ADF	Gold	Dow Jones	US Dollar	Crude Oil	Inflation
p-value level	0,9928	0,1466	0,1869	0,3345	0,0242
p value first dif	0,0010	0,0039	0,0031	0,0070	0,0082

Source: processed data

ADF test results show that all variables are not stationary in the initial condition ($p - value \geq 0.05$) so it is necessary to make a difference in order for the data to be stationary. The test results showed that all data had been stationary after a difference of 1 time ($p - value < 0.05$). The number of differences in each variable until the variable is stationary is the same as 1 so that the value of m is 1. The p parameter used is obtained from the information criteria by following the Akaike Information Criterion method. The AIC information criteria advise to use lag 2 so that parameter p in causality testing is 2. Granger Causality testing is then done using $VAR(p + m)$ namely $VAR(3)$, this test has an initial hypothesis that the variables have no effect on gold and alternative hypothesis that the variables have an effect (granger

causes) on the price of gold. The results of this test can be seen in Table 2.

Table 2
Granger Causality Test Results

	Dow Jones	US Dollar	Crude Oil	US Inflation
p-value	0,0000	0,0204	0,0000	0,0903

Source: processed data

The test results showed that the US. Dollar index, Dow Jones index, crude oil prices, and inflation rates in America were the cause of gold prices. The test results showed that modeling by adding these variables could not make the forecasting of gold prices more accurate. The result of this test is also variable price of crude oil does not need to be discarded and can be used to model the price of gold. The next step is to determine which model to use for multivariate modeling. Modeling using Vector Autoregressive (VAR) can be used if the data is stationary. Based on ADF testing, it has been known that the variables to be modeled are not stationary, so modeling the data directly using var model can not be done, because it will cause the appearance of spurious regression. Data that is not stationary can be modeled using VAR after making the first difference if the data does not integrate. Data that is not stationary and integrating can be modeled using the Vector Error Correction Model ($VECM$).

Cointegration testing in this study was conducted using Johansen Cointegration Test. This test aims to determine the cointegration of data and to know the number of rank (r) that indicates the number of cointegrating vectors. The maximum rank amount is $n-1$ where n is the number of variables (in this study 5). The initial hypothesis in this test is $r = 0$, where there is no cointegration and the alternative hypothesis is cointegration ($r > 0$). Rejection of the initial hypothesis ($r = 0$) will be followed by testing to determine rank (number of integrating vectors). Johansen Cointegration Test results can be seen in Table 3.

Table 3
Johansen Cointegration Test Results

	Gold	Dow	Us Dollar	Crude	Inf
Gold(-1)	0.641338	0.859537	0.013709	-0.019250	-0.002188
	(0.48366)	(12.5051)	(0.00999)	(0.05591)	(0.00109)
	[1.32600]	[0.06873]	[1.37160]	[-0.34430]	[-2.01509]
Gold(-2)	0.058002	-3.578280	-0.009172	-0.056225	-3.34E-05
	(0.50362)	(13.0210)	(0.01041)	(0.05822)	(0.00113)
	[0.11517]	[-0.27481]	[-0.88128]	[-0.96579]	[-0.02953]
Dow(-1)	0.008738	0.597798	-0.000351	0.003854	0.000238
	(0.02393)	(0.61873)	(0.00049)	(0.00277)	(5.4E-05)
	[0.36515]	[0.96618]	[-0.70947]	[1.39318]	[4.43752]
Dow(-2)	0.005940	-0.051684	0.000259	-0.000409	0.000104
	(0.02934)	(0.75868)	(0.00061)	(0.00339)	(6.6E-05)
	[0.20244]	[-0.06812]	[0.42656]	[-0.12063]	[1.58229]
Us Dollar(-1)	1.572502	-278.3788	1.084734	-0.987364	-0.127057
	(19.7678)	(511.095)	(0.40850)	(2.28507)	(0.04438)
	[0.07955]	[-0.54467]	[2.65543]	[-0.43209]	[-2.86317]
Us Dollar(-2)	5.553516	494.5552	-0.367132	0.730210	0.052935
	(16.5595)	(428.145)	(0.34220)	(1.91420)	(0.03717)
	[0.33537]	[1.15511]	[-1.07287]	[0.38147]	[1.42397]
Crude(-1)	-6.633698	-30.50235	0.199664	-0.068445	-0.011992
	(5.17221)	(133.727)	(0.10688)	(0.59788)	(0.01161)
	[-1.28257]	[-0.22809]	[1.86808]	[-0.11448]	[-1.03279]
Crude(-2)	1.355610	53.78706	-0.119601	0.213968	0.002688
	(3.62380)	(93.6932)	(0.07488)	(0.41890)	(0.00813)
	[0.37409]	[0.57408]	[-1.59714]	[0.51079]	[0.33045]
Inf(-1)	47.51925	-125.2854	-1.788011	-0.522933	0.342213
	(159.523)	(4124.46)	(3.29650)	(18.4402)	(0.35811)
	[0.29788]	[-0.03038]	[-0.54240]	[-0.02836]	[0.95561]
Inf(-2)	-102.4838	-1406.683	3.728958	-12.51190	-0.259277

	(94.9133)	(2453.98)	(1.96136)	(10.9716)	(0.21307)
	[-1.07976]	[-0.57323]	[1.90121]	[-1.14039]	[-1.21687]
C	-225.9341	-3767.345	15.23085	112.2978	3.624194
	(1870.26)	(48355.5)	(38.6484)	(216.194)	(4.19851)
	[-0.12080]	[-0.07791]	[0.39409]	[0.51943]	[0.86321]
R-squared	0.952758	0.556026	0.770965	0.838728	0.976804
Adj. R-squared	0.905515	0.112052	0.541930	0.677457	0.953607
Sum sq. resids	36481.65	24387206	15.57876	487.4804	0.183849
S.E. equation	60.40004	1561.640	1.248149	6.981980	0.135591
F-statistic	20.16745	1.252383	3.366144	5.200715	42.11003
Log likelihood	-108.1282	-176.4307	-26.66226	-62.81735	19.95303
Akaike AIC	11.34554	17.85054	3.586882	7.030224	-0.852670
Schwarz SC	11.89267	18.39767	4.134013	7.577354	-0.305539
Mean dependent	1559.410	26094.42	96.97948	48.79048	1.585714
S.D. dependent	196.4972	1657.247	1.844168	12.29375	0.629512
Determinant Residual Covariance			4.66E+08		
Log Likelihood (d.f. adjusted)			-358.5735		
Akaike Information Criteria			39.38795		
Schwarz Criteria			42.12361		

Source: processed data

The initial hypothesis is rejected if the test value is greater than critical value at a significance level of 5%. Test results showed that the data was integrated because the initial hypothesis was rejected. The test results showed that the rank count is 2 which means there are 2 cointegrating vectors that will be used to model the data using $VECM$. The rank count of 2 means a linear combination of 2 vectors integrating with stationary variables. The next step is to do modeling using $VECM$. Estimation of parameters in modeling is obtained using the help of R software.

$$\begin{bmatrix} \Delta Gold_t \\ \Delta Dj_t \\ \Delta Crude_t \\ \Delta\$_t \\ \Delta I_t \end{bmatrix} = C + \alpha\beta \begin{bmatrix} Gold_{t-1} \\ Dj_{t-1} \\ Crude_{t-1} \\ \$_{t-1} \\ I_{t-1} \end{bmatrix} + \Gamma_1 \begin{bmatrix} \Delta Gold_{t-1} \\ \Delta Dj_{t-1} \\ \Delta Crude_{t-1} \\ \Delta\$_{t-1} \\ \Delta I_{t-1} \end{bmatrix} + \Gamma_2 \begin{bmatrix} \Delta Gold_{t-2} \\ \Delta Dj_{t-2} \\ \Delta Crude_{t-2} \\ \Delta\$_{t-2} \\ \Delta I_{t-2} \end{bmatrix}$$

Gold lowered the price of gold, DJs eased the Dow Jones index, Crude lowered crude oil prices, $ dissed the U.S. Dollar index, and I lowered the inflation rate in America. C is a constant vector that represents the mean value, α is a matrix $(5\ x\ 2)$ that represents the adjustment coefficient, β is a matrix $(2\ x\ 5)$ thatots the integrating vector. Vector integration is a long-term equilibrium relationship between variables and the adjustment coefficient is the speed of correction to error in order to achieve equilibrium at a time. Γ_1 and Γ_2 are matrices $(5\ x\ 5)$ containing the autoregressive coefficients of the first and second lags. Testing of residuals and model forecasting is carried out by first transforming the model into a VAR form. Diagnostic Tests are carried out to ascertain whether the model formed is valid and can be used to perform forecasting.

$$\begin{bmatrix} Gold_t \\ Dj_t \\ Crude_t \\ \$_t \\ I_t \end{bmatrix} = C + A_1 \begin{bmatrix} Gold_{t-1} \\ Dj_{t-1} \\ Crude_{t-1} \\ \$_{t-1} \\ I_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} Gold_{t-2} \\ Dj_{t-2} \\ Crude_{t-2} \\ \$_{t-2} \\ I_{t-2} \end{bmatrix} + \Gamma_2 \begin{bmatrix} Gold_{t-3} \\ Dj_{t-3} \\ Crude_{t-3} \\ \$_{t-3} \\ I_{t-3} \end{bmatrix}$$

The result of converting $VECM$ model into VAR with model where $VAR(3)$ where C is $(n\ x\ 1)$ constant vector that specify mean while A_1, A_2, A_3 and is a matrix $(5\ x\ 5)$ containing autoregressive coefficient at the first, second, and third lag. The next step is to perform diagnostic test consisting of serial correlation test, heteroscedasticity test, and normality test on residual model. Diagnostic tests are required as validation to ensure that errors from the model follow IIDN assumptions. The test results of serial correlation can be seen in Table 4.

Table 4
Serial Correlation Test Results

F-statistic	0.414595	Probability	0.868480
Obs*R-squared	4.259212	Probability	0.749484

Source: processed data

Serial correlation testing aims to determine if residuals from the model are serially correlated. The hypothesis on this test is:
H_0: Residuals are not serially correlated
H_1: Residuals correlate serially
Test results showed that the initial hypothesis was received because the $p - value > 0.05$; which means that residual models do not correlate until the 16th lag. Testing Results of heteroscedasticity test can be seen in Table 5.

Table 4
Testing Heteroscedasticity

F-statistic	1.554269	Probability	0.269627
Obs*R-squared	16.81716	Probability	0.266053

Source: processed data

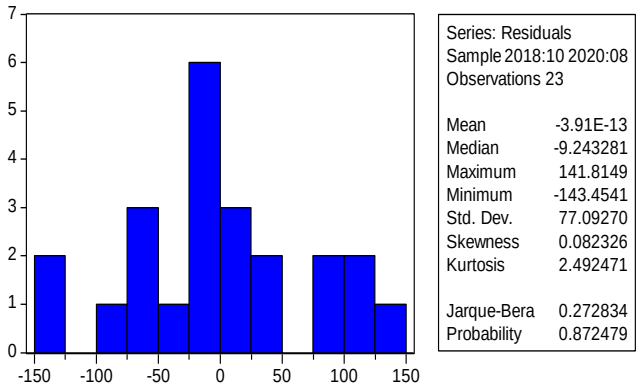


Figure 1
Residual Normality Test Results

Heteroscedasticity testing aims to determine whether the residuals produced by the model are identical or not. Residuals that are identical are called homoscedasticity, while non-identic residuals are called heteroscedasticity. The hypothesis on this test is:
H_0: Residual identic (homoscedasticity)
H_1: Residual is not identical (heteroscedasticity)
The test results showed that residuals were identical due to the initial hypothesis received ($p - value > 5\%$). The next test is normality testing on the residual model where the results of the test can be seen in Figure 1. Testing normality on residuals using Jarque Bera Test with the following test hypothesis:
H_0: Normal distributed residuals
H_1: Residual is not normally distributed
Normality test results showed that residual distribution was normal (initial hypothesis accepted). There are 2 other types of statistical tests on JB normality testing, namely skewness and kurtosis. A normal distribution will have a skewness value of 0 and a kurtosis of 3. The initial hypothesis on skewness testing is the skewness $lue = 0$, while the initial hypothesis of kurtosis testing is the kurtosis $value = 3$. The test results showed that residuals from the model had a skewness value of zero, kurtosis of 3, and normal distributed residuals. Diagnostic test results show that the residuals of the model have been valid so that the model is feasible for forecasting. The results of model forecasting can be seen in Figure 2.

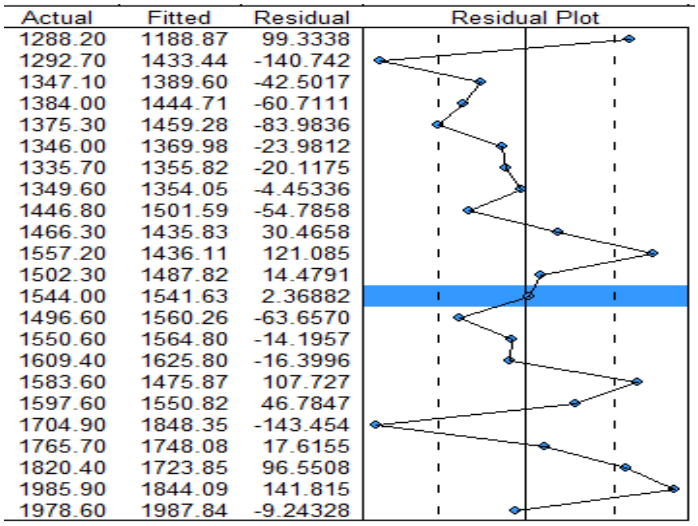


Figure 2
Model Forecasting Results

Models that have been built using data for the period 2018.10 – 2020.08. Gold price forecast in Figure 2 which is the forecast for the period 2018.10 – 2020.08 which will be compared to the actual price of gold. The forecast results can be seen in the Fitted column, while the actual gold price data can be seen in the Actual column. The forecast results show a trend of increasing gold prices until August 2020, while the actual price of gold has increased in a long run. The results of forecasting using the model that has been built show that the actual value of gold price is at 95% Confidence Interval model forecasting results. $RMSE$ of forecast 74.2 or 6.1% of the actual gold price mean. These results show that the forecasting model has been well constructed and shows fairly accurate results. The graph of gold price forecasting results can be seen in Figure 3.

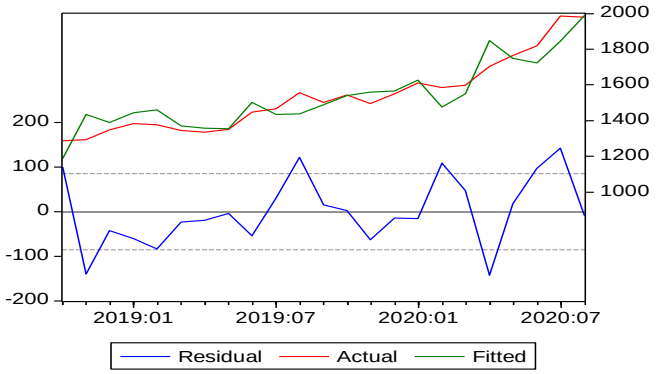


Figure 3
Gold Price Forecasting

The results of forecasting using the model that has been built are shown through the red line in Figure 3. The actual price of gold is indicated through the blue line,

while the lower limit and upper limit of gold price forecasting are indicated through gray and orange lines. Forecasting charts show that the forecasting results are at confidence intervals.

Impulse Response Function

Granger causality test states that Dow Jones variables, US Dollar, US Inflation, and Crude Oil help in predicting gold price, but how big and direction of influence each variable is unknown. The influence of each element can be known using Impulse Response Function (IRF) analysis. Impulse Response Function calculates the amount of change that will be experienced by all variables if there is a change/shock on one variable at the time of calculation results impulse response function for 10 lag forward with gold price as a response can be seen in Figure 4.

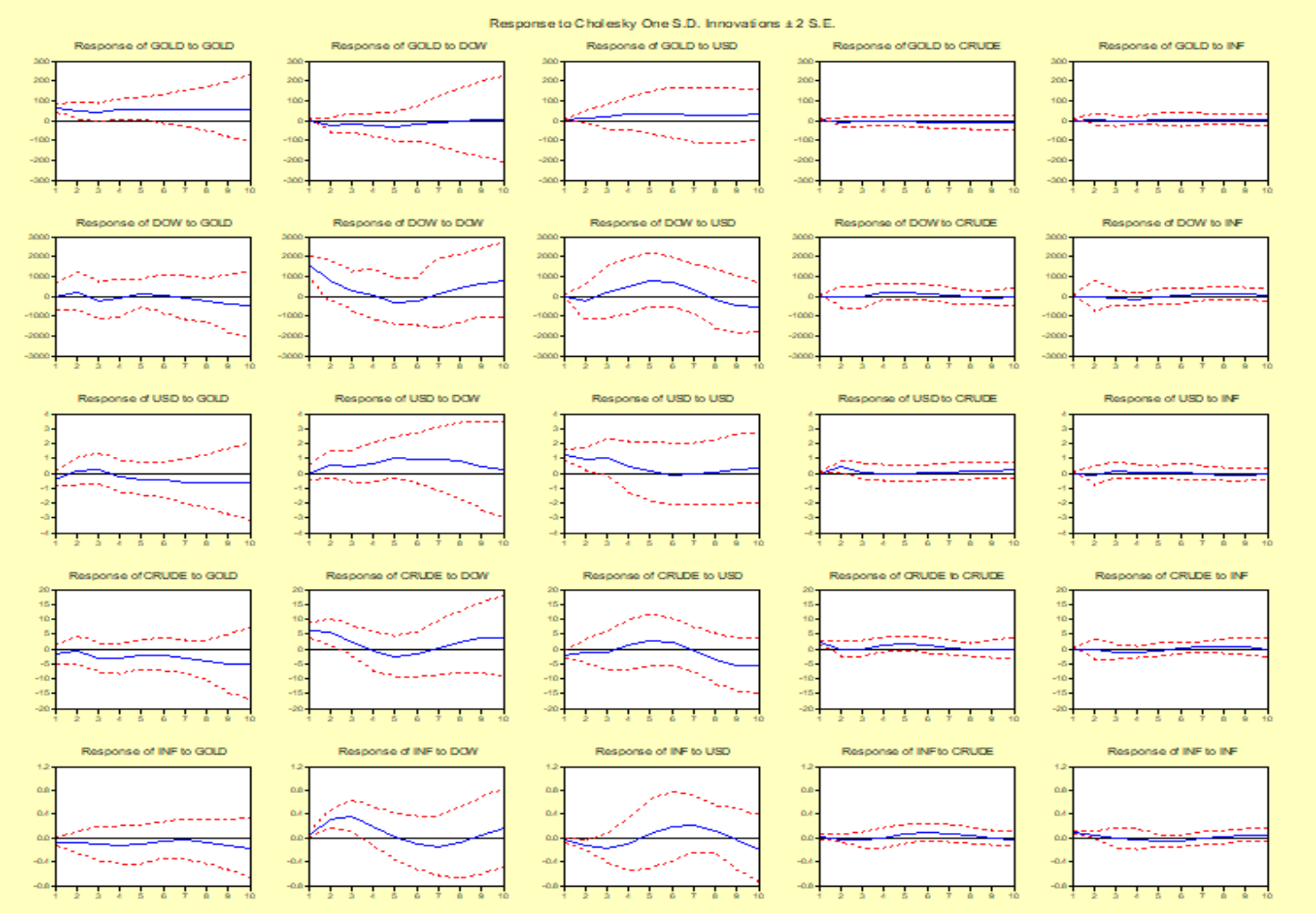


Figure 4
Impulse Response Function

Figure 4 shows the gold price response in the next 10 periods at each increase of 1 point by other variables and themselves. IRF calculations show that the current US dollar index increase of 1 point will be followed by a decline in the gold price of $6.24 per troy ounce in the next 1 month, $10.75 in the two months from now, and a decrease of $12.3 in the 10 months from now. An increase in the current inflation rate of 1% would increase the price of gold by $20.2 in the month from now and $52.8 in the next 10 months. An increase in the price of oil by 1 dollar at this time will lower the gold price by 1.73 dollars next month and 3.43 dollars 10 months from now. A 1-point increase in the Dow Jones index will cause gold prices to rise in the first and second months in the future, but will lower the price of gold after 2 months.

The results of analysis through impulse response function showed that the US Dollar index and Crude Oil price have a negative correlation to gold price, while the inflation rate has a positive correlation with gold price. The Dow Jones index has a positive correlation with gold in the first and second periods in the future, and has a negative correlation there after. The negative correlation

between the US Dollar and gold indicates that the weakening US dollar will encourage economic actors to keep their money in gold in order to maintain purchasing power. The positive correlation between gold prices and inflation indicates that the value of gold is not eroded by inflation, thus supporting the statement of gold investment called able to withstand inflationary turmoil (Kusuma, 2015).

CONCLUSION

Univariate gold price modeling shows that the gold price follows the $ARIMA(0,1,0)$ model which means that fluctuations in gold prices are white noise. This fact states that the price of gold cannot be modeled using only its own historical data. The fact of this research is supported botha (1980), as well as Shankari & Manimaran (2015). Multivariate modeling to model gold prices is done using Vector Error Correction Model ($VECM$) because the variables used are nonstationary and integrate. The results showed that gold prices were influenced by the US Dollar index, Dow Jones index, crude oil prices, and inflation rates. The US Dollar index and crude oil prices have a negative correlation to gold prices, while the inflation rate is positively correlated with the gold price. The Dow Jones index is positively correlated with gold prices only at the first and second lags.

REFERENCES

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., *International Symposium on Information Theory*, 267-281.

Box, George E. P. and Jenkins, Gwilym M., (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Cryer, J. D., & Chan, K. S. (2008). T*ime Series Analysis with Applications in R* (2nd Ed.). New York: Springer. http://dx.doi.org/10.1007/978-0-387-75959-3

Gilroy, Annie. (2014), Why gold and the US dollar have an inverse relationship, *yahoo finance*, Retrieved https://finance.yahoo.com/news/why-gold-us-dollar-inverse-170016117.html

Granger, C.W.J., (1969), Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, 37(3), 424-438

Hannan, E. J., & Quinn, B. G., (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society*, 41(2), 190-195

Holmes, F., (2016), Top 10 Countries With The Largest Gold Reserves, *Forbes*, Retrieved https://www.forbes.com/sites/greatspeculations/2016/05/26/top-10-countries-with-the-largest-gold-reserves/#7c55b6cd730a

Jannah, K. M. (2015). Harga Minyak Dunia Selamatkan Inflasi, *Okefinance* Retrieved https://economy.okezone.com/read/2015/02/17/20/1107161/harga-minyak-dunia-selamatkan-inflasi

Josheski, Dushko & Bardarova, Snezana, (2013). Causal relationship between wages and prices in R. Macedonia: VECM analysis, *EconStor Preprints*, 77910, ZBW Leibniz Information Centre for Economics.

Kusuma, D. M. (2015), Ini Keuntungan Investasi Emas Batangan dalam 2 Tahun. *detikfinance*, Retrieved https://finance.detik.com/berita-ekonomi-bisnis/d-3028781/ini-keuntungan-investasi-emas-batangan-dalam-2-tahun

Manimaran, S., Shankari, S. (2015), An Empirical Investigation of Random Walks in Gold Price Movement, *Global Journal of Commerce & Management Prespective*, 4(2), 16-21 Retrieved https://www.longdom.org/articles/an-empirical-investigation-of-random-walks-in-gold-price-movement.pdf

Schwarz, G., (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, 6(2), 461-464.

Seputar Forex, (2018), Pengaruh Dow Jones di Pasar Forex, *Seputar forex*, Retrieved https://www.seputarforex.com/belajar/forex/pengaruh-dow-jones-terhadap-forex/

Sahabat Pengadaian, (2016), 5 Faktor yang Memengaruhi Harga Emas Naik dan Turun, *Sahabatpengadaian*, Retrieved https://sahabatpegadaian.com/emas/5-faktor-harga-emas-naik-dan-turun

Seputar Forex, (2016), Apa Saja yang Mempengaruhi Pergerakan Gold?, *Seputar forex*, Retrieved https://www.seputarforex.com/tanya-jawab-forex/apa-saja-yang-mempengaruhi-pergerakan-gold-109978

Sepanek, E., (2014), How These 10 Factors Regularly Influence Gold Princes, *Scottsdale*, Retrieved https://www.sbcgold.com/blog/10-factors-regularly-influence-gold-prices/

Slutzky, E., (1937), The Summation of Random Causes as the Source of Cyclic Processes, *Econometrica*, 5(2), 105-146

Tsay, Ruey S., (2013), *Multivariate Time Series Analysis With R and Financial Applications*, 1st Edition, Wiley

Wold, H., (1938), *A Study in the Analysis of Stationary Time Series*. 2nd Edition, Almqristand Witsett, Stockhol.

Yule, G. U. (1926), Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in

Sampling and the Nature of Time Series, *Journal of the Royal Statistical Social Sciences*, 89, 1-64.
Zivot, Eric., and Wang, Jishui, (2006), *Modeling Financial Time Series with S-PLUS*, 2nd Edition, Springer