ISSN E.2958-6941 ,ISSN P.1818-1074 Vol. .22 Issue. 87.March. 2026

http://journals.uokerbala.edu.iq/index.php/ijas/index
Iraqi Journal for Administrative Sciences
المجلة العراقية للعلوم الإدارية

Analyzing of Multivariate Time Series with Vector Autoregressive (VAR) Models

Zewar Omar Ismael[1] , Awaz Omer Ahmed[2]
Sawen Othman Babakr[3] , Sami Ali Obed[4]

تحليل السلاسل الزمنية متعددة المتغيرات مع نماذج الانحدار الذاتي المتجهية (VAR)

زيوڤر عمر اسماعيل[1]، أواز عمر احمد[2]، ساوين عثمان بابكر[3] ، سامى علي عبيد[4]

1. Administration and Economics College -Salahaddin University –Kurdistan Region, Erbil-Iraq، zewar.ismaeil@su.edu.krd
2. Administration and Economics College -Salahaddin University –Kurdistan Region, Erbil-Iraq. awaz.ahmed@su.edu.krd
3. Administration and Economics College -Salahaddin University –Kurdistan Region, Erbil-Iraq. sawen.babakr@su.edu.krd
4. Administration and Economics College -Salahaddin University –Kurdistan Region, Erbil-Iraq. sami.obed@su.edu.krd * **Corresponding Author**

1. كلية الإدارة والاقتصاد، قسم الاحصاء، جامعة صلاح الدين، اربيل، العراق
2. كلية الإدارة والاقتصاد، قسم الاحصاء، جامعة صلاح الدين، اربيل، العراق،
3. كلية الإدارة والاقتصاد، قسم الاحصاء، جامعة صلاح الدين، اربيل، العراق،
4. كلية الإدارة والاقتصاد، قسم الاحصاء، جامعة صلاح الدين، اربيل، العراق، ***الباحث المراسل**

Article information

Article History: DD/MM/YY
Received: 04/10/2025
Accepted:15/12/2025
Available online:01/03/2026

Keywords:
Mean Squared Error, oil and gold price, VAR model, Duekey fulier test.

تاريخ الاستلام:2025/10/04
تاريخ قبول النشر: 2025/12/15
تاريخ النشر: 2026/03/01

الكلمات المفتاحية
متوسط الخطأ التربيعي، سعر النفط والذهب، نموذجVAR ، اختبار دوكي-فوليير.

DOI: https://doi.org/10.71207/ijas.v22i87.5585

Abstract

This study applied a Vector Autoregressive (VAR) model to examine the active relationship between global oil and gold prices using this data from January 2015 to June 2019. Stationarity was tested through the Augmented Dickey-Fuller (ADF) test, with necessary transformations applied. Lag order selection criteria (RMSE, AIC, BIC, HQC, and FPE) identified VAR (1) as the optimal model. Findings revealed that gold prices are highly persistent and mostly influenced by their individual past values, while oil prices exert a negative effect on gold price changes. Diagnostic tests confirmed model adequacy and stability. Forecasting results suggested a gradual decline in gold prices and a slight increase and stabilization in oil prices over time. The study concludes that VAR models provide a reliable and effective framework for analyzing inter-market relationships and generating short- to medium-term financial forecasts. All analyses were conducted using Python programming.

Citation: Omar Ismael, Zewar, Omer Ahmed, Awaz, Othman Babakr, Sawen, Ali Obed, Sami. (2026). Analyzing of Multivariate Time Series with Vector Autoregressive (VAR) Models, *Iraqi Journal for Administrative Sciences*, 22(87), 57-78.

الاقتباس: عمر اسماعيل، زيوڤر، عمر احمد، أواز، عثمان بابكر، ساوين، علي عبيد، سامي. (2026). تحليل السلاسل الزمنية متعددة المتغيرات مع نماذج الانحدار الذاتي المتجهية (VAR)، *المجلة العراقية للعلوم الإدارية*، 22(87)، 57-78.

المستخلص

اعتمدت هذه الدراسة نموذج الانحدار الذاتي المتجه (VAR) لتحليل العلاقة الديناميكية بين الأسعار العالمية للنفط والذهب، بالاعتماد على بيانات تغطي الفترة من يناير 2015 إلى يونيو 2019. تم التحقق من استقرارية السلاسل الزمنية باستخدام اختبار ديكي–فولر المعزز (ADF)، وتحديد رتبة التأخر المثلى استنادًا إلى معايير AIC وBIC وHQC وFPE، حيث تبيّن أن نموذج VAR(1) هو الأنسب. أظهرت النتائج أن أسعار الذهب تتأثر بشكل رئيسي بقيمها السابقة، في حين أن أسعار النفط تمارس تأثيرًا سلبيًا على تغيرات أسعار الذهب. كما أكدت الاختبارات التشخيصية كفاءة النموذج واستقراره، وأشارت نتائج التنبؤ إلى انخفاض تدريجي في أسعار الذهب واستقرار نسبي في أسعار النفط، مما يبرز فعالية نموذج VAR في تحليل العلاقات المتبادلة بين الأسواق المالية وإنشاء توقعات مالية قصيرة إلى متوسطة الأجل. وقد تم تنفيذ جميع التحليلات باستخدام لغة برمجة بايثون.

1. Introduction

Dynamic interactions between economics, finance, energy, and natural sciences reveal that variables almost never change on their own, but rather co-vary and affect one another in complex ways. These joint dynamics cannot be accounted for using conventional univariate models and therefore multivariate time series models such as the Vector Autoregressive (VAR) model need to be employed. The VAR, which was popularized by Sims (1980), has become a standard application in empirical macro and financial models because it is flexible and easily estimated, has a very good capacity to account for system-wide evolutions without being "overly structural" and thus imposing strict assumptions. (Zivot & Wang, 2006; Lütkepohl, 2013).

Formally, a stable $VAR(p)$ for a k-dimensional vector y_t can be written as:

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad \ldots(1)$$

where c is a vector of intercepts, $A_1,\ldots,A_p$ are k×k autoregressive coefficient matrices capturing cross-lag effects, and u_t is a zero-mean innovation with covariance matrix Σ_u . VAR models, typically estimated equation-by-equation using ordinary least squares (OLS) for consistency and asymptotic efficiency, facilitate multi-step forecasting, Granger causality testing, and the generation of impulse response functions (IRFs) and forecast error variance decompositions (FEVDs) to analyze shock transmission (Zivot & Wang, 2006; Siggiridou & Kugiumtzis, 2015).

Successful VAR application hinges on addressing stationarity and integration properties, often requiring differencing, detrending, or log-growth rates. Cointegrated nonstationary series necessitate a Vector Error Correction Model (VECM) to capture both short-run dynamics and long-run equilibria (Lütkepohl, 2013).

Pretesting for unit roots and cointegration ensures valid inference and appropriate dynamic specification, while accurate lag-length selection—guided by information criteria such as AIC, BIC, HQC, and FPE—is crucial to avoid omitted dynamics and overfitting, with parsimony and contextual understanding also being vital (Abdullah, 2022).

VAR models have been widely applied to study interdependencies and shock propagation in macroeconomic, financial, and energy systems. Their applications span diverse domains, including forecasting meteorological time series (Adenomon & Oyejola, 2014), (Bose, Hravnak, & Sereika, 2017), and commodity and financial markets

Moreover, extensions of the VAR framework have improved robustness in the presence of missing data, nonlinearities, and structural breaks (Bashir & Wei, 2018)

Against this background, this study employs a VAR framework to examine the dynamic nexus between global oil and gold prices from January 2015 to June 2019. These commodities were selected due to their central role in global energy and financial markets: oil prices drive production costs and macroeconomic stability, while gold functions as a safe-haven asset during uncertainty. By applying the VAR methodology, this study aims to: (i) assess short- and medium-term interdependencies between oil and gold, (ii) evaluate predictive causality and impulse responses, and (iii) generate forecasts to inform policy and investment decisions. Building on established VAR literature , this research contributes to the empirical understanding of commodity market linkages, highlighting the VAR model as a reliable and transparent tool for forecasting, causality analysis, and system-level interpretation in interconnected financial systems (Mati et al., 2024) (Mohammed et al., 2022).

2. Literature review

In this context, the literature on VAR applications demonstrates its superior forecasting performance compared to univariate models like ARIMA or ARIMAX, particularly when dealing with interdependent economic variables. The following review examines prior studies that have employed VAR models to analyze fiscal and monetary indicators.

(Haydier et al., 2023) It concluded that the VAR model produces more accurate projections for Iraq's government spending and foreign reserves than ARIMAX as evidenced through MSE analysis. It discloses a high positive correlation between them that indicates a trend for both to increase between 2021 and 2024 and could perpetuate the budget deficit. The paper recommends adopting VAR in the process of future financial planning, while ensuring that financial stability reinforces the long-term sustainability of the same.

(Oyejola, n.d.2018) "Developing a reliable forecasting model for Meteorological Time Series: A Review of Performance of VAR against some traditional Univariate Time Series Models", it was discovered that Vector Autoregression model outperformed the Decomposition, Holt-Winter and SARIMA which are all conventional Univariate Time Series Models in the Forecasting of Meteorological Time Series, specifically Temperature and Humidity in Bida, Niger State (Oyejola, nd). Results indicate a strong negative correlation and bidirectional causality between the two variables and both stationary. Given that VAR (13) model was the one with optimal RMSE it was concluded that the model best represented the data's dynamic relationship. The study shows that the multivariate VAR models are more accurate and reliable at forecasting than univariate methods when analyzing weather data.

(Rani Kundu et al., 2021)It provides a theoretical background, a practical guide, and modeling structure analysis of VAR models which occupies a prominent place in the analysis of multivariate time series, especially in economics and finance.

(Siggiridou & Kugiumtzis, 2015)This work proposed a solution, mBTS, to better estimate Granger causality in high-dimensional short-length time series. This stepwise selection, where only time- appropriate lagged variables are selected, results in a much more precise and robust form of causal inference that can be especially beneficial in difficult datasets such as those from EEG recordings, where systems are often quite complex . This method is shown to be superior to conventional approaches in simulations and successfully identifies shifts in brain networks in the setting of epilepsy.

(Noémi & Ferrer, 2016) the paper notes how multilevel autoregressive models with within-person (WP) centering permit researchers to explore variations between persons in psychological processes at the level of the individual over time. It offers more precise information about causal strength and direction at the level of the individual, rather than average effects.

(Taha Abdullah, 2022) Using AIC and MSE criteria for model order selection, the study employed Vector Autoregressive (VAR) models to predict monthly global prices of both oil and gold between the years 2015 and 2019. The VAR(10) has the lowest MSE overall but VAR(7) provides a better oil price forecast and VAR(10) better estimates the gold price forecast VAR selection results state that the VAR(10) has the lowest overall MSE and is therefore the best fit to the combined data. The model is proven by diagnostic tests Lagrange-multiplier, Portmanteau and Jarque-Bera, which shows that the residuals are adequate and normally distributed.

3. Research Methodology

3.1 Multivariate Time series

Initial all-inclusive multivariate time series model was the Multivariate Auto-Regressive State-Space MARSS model that was first presented in 2012. It focuses on Expectation-Maximization (EM). A method which is applied to estimate a number of multivariate time series data missing values is the EM technique. The following matrix structure is an overview of the MARSS model

$$\begin{cases} x_t = A_t x_{t-1} + B_t b_t + \varepsilon_t \\ y_t = C_t x_{t-1} + D_t d_t + \mu_t \end{cases} \dots (2)$$

Where $\varepsilon_t \sim MVN(0, Q_t)$,$\mu_t \sim MVN(0, R_t)$ and $x_1 \sim MVN(\pi, \Lambda)$, or $x_0 \sim MVN(0, \Lambda)$, x_t represents the state vector, while y_t designates the measured value. Data drives the model's evolution, yet while measuring y, certain values can be missing. For instance, some indicators or exogenous variables are represented by the input variables b_t and d_t. Q_t and R_t are m×m and n×n variance-covariance measures, individually, where mis are the number of states and n is the number of timeseries; A_t, C_t, and $D_t$are system matrices; ε_t and μ_t are process and non-processerr, individually. MARSS can produce better results than standard methods, particularly for multivariate time series models (Zivot & Wang, 2006; Lütkepohl, 2013).

3.2 Time series:

The integration of global financial markets, driven by technology and globalization, necessitates understanding increased interdependence. Empirical economic analysis, requiring relevant data, seeks to examine economic dynamics. The Vector Autoregressive (VAR) model, introduced by Sims (1980) for macroeconomic time series analysis, is crucial in contemporary economics and finance. Its popularity stems from ease of estimation, examined properties, and multivariate regression similarity, advocating for simultaneous modeling of endogenous variables and elimination of exogeneity assumptions. VAR is flexible, appropriate for multivariate time series analysis and forecasting, and can be extended to VARX models for exogenous variables, enabling structural analysis through Granger causality and impulse response functions.

Time series is a set of data compiled or amassed at different sequential points and over some time, and it is called time series data. Identifying patterns, trends, and other relevant features in a series of values collected over time is called time series analysis (Muzahem et al., 2023). (Mohammed et al., 2022).

3.3 ARX Model

Because the ARX model has an input term in contrast to the AR model, it is also known as the Autoregressive with Extra Input model. Autoregressive with Exogenous Variables, or ARX for short, is a regression model in which the input term is the exogenous variable. The ARX model structure is given by the following equation (Hayawi, et al. 2025).

$$y(t) = \sum_{i=1}^{p} y_i \theta_{i-t} + \sum_{j=1}^{q} x_j \beta_{j-t} + \varepsilon_t \quad \dots (3)$$

Aside from the time series variable in this context, y_i is a t - time series variable at time t; p is an order of autoregressive processes; θ_{i-t} is an i - t self-regression coefficient that measure a relationship between current and past values of the time series. In addition, x_j is the external variable that influences the time series and in q is the number of periods where these external variables impact the time series. There is β_j, the coefficients of this external variable, and t is the random error term, considered to be a normal distribution with a mean of zero and constant variance. (Haydier et al. 2023)(Siddiqui et al., 2023).

3.4 The VAR (1) model

This VAR (Vector AutoRegressive) model is a model that focuses on M-variate time series data from length T, and is performed on all of the M variables in the VAR model individually. We will focus on a model of order 1 VAR which is one of the most popular methods of psychology, which predicts the values of thevariables at time t using observations at time t-1. Higher-order models, however, can readily adopt this idea. (Hamilton, 1994; Lütkepohl, 2005) (Russel, E., & Usman, M.2019). The model formula for VAR(1) is as follows:

$$y_t = c + \Phi y_{t-1} + u_t \dots (4)$$

In this model, the innovations at time t are represented by the M×1 vector ut, while the intercepts are stored in the M×1 vector c, and the VAR (1) regression coefficients are contained in the M×M matrix Φ. The values of the variables at times t and t-1 are represented by the M×1 vectors y_t and y_{t-1}, respectively. The portion of the variable that cannot be predicted from the values at the previous time point is captured by the innovations. These innovations are the residuals of the system's dynamics. (Siggiridou & Kugiumtzis, 2015),(Wild et al., 2010).

3.5 Forecasting

VAR models are regarded as natural forecasting tools since current variables are represented by prior period variables. Because they record the correlation between variables and utilise this knowledge to predict how these variables will change over time, they are helpful for prediction. The VAR(p) model's forecasting procedure is comparable to the AR(P) model's, and the one-step ahead forecast is computed as follows:(Mati, at.el (2025))

$$Y_{t+1/t} = C + A_1 y_t + \ldots + A_p y_{t-p+1} \quad \ldots (5)$$

Forecasting for h steps:

$$Y_{t+h/t} = C + A_1 y_{t+h-1/t} + \ldots + A_p y_{t+h-p/t} \ldots (6)$$

3.5.1 The Stationary Vector Autoregression Model

$Let Y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})' denote an (n \times 1)$ Time series variable vector. The form of the fundamental p-lag vector autoregressive (VAR(p)) model is

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t \qquad t = 1, \ldots T \ldots (7)$$

Where Π_i are $(n \times n)$coefficient matrices and ε_t is an (n×1) Time invariant covariance matrix $\sum$ with unobservable zero mean white noise vector process (serially uncorrelated or independent). The abivariate VAR (2) model equation by equation, for instance, has the following form:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi_{11}^2 & \pi_{12}^2 \\ \pi_{21}^2 & \pi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \ldots (8)$$

Or

$$y_{1t} = c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \pi_{11}^2 y_{1t-2} + \pi_{12}^2 y_{2t-2} + \varepsilon_{1t}$$
$$y_{2t} = c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \pi_{21}^2 y_{1t-1} + \pi_{22}^2 y_{2t-1} + \varepsilon_{2t} \quad \ldots (9)$$

Where cov $(\varepsilon_{1t}, \varepsilon_{2s}) = \sigma_{12}$ for $t = s;$ 0 otherwise. Observe that the regressors for each equation are the same: the lag values of y_{1t} and y_{2t}. With lagged variables and deterministic terms as common regressors, the VAR (p) model is therefore only an apparently unrelated regression (SUR) model (Hamilton, 1994).

3.6 Filtering of Multivariate Time Series

A multivariate linear (time-invariant) filter that links an n-dimensional output series Y_t to a l-dimensional input series U_t is frequently expressed as follows:

$$Y_t = \sum_{N=-\infty}^{\infty} U_{t-N} B_N \ldots (10)$$

In n × 1 matrices the B_N is a n × 1 matrices. The filter is physically possible or causal if N = 0 for N that leads to the $Y_t = \sum_{N=0}^{\infty} U_{t-N} B_N$, which means Y_t can be characterized by past values of the input U_t. It is argued that the filter is stable if $Y_t = \sum_{N=-\infty}^{\infty} \|B_N\| < \infty$. In the stability condition and with the assumption that the input random vectors Ut have evenly spaced second moments, the output random vector Y_t is the output random vector which is represented by (5) and represents the limit (Schmitz & Skinner, 1993).

$$\lim_{r\to\infty} \sum_{N=-r}^{r} U_{t-N} B_N \quad \dots(11)$$

such that as $r \to \infty$

$$Y_t = E\left[\left(Y_t - \sum_{N=-r}^{r} U_{t-N} B_N\right)\left(Y_t - \sum_{N=-r}^{r} U_{t-N} B_N\right)^T\right] \dots (12)$$

As the filter is stable and input series is stationary with cross-covariance matrices, a stationary process occurs$\Gamma_U(p)$. Next, the stationary process Y_t's cross-covariance matrices are provided by

$$\Gamma_U(p) = Cov(Y_t, Y_{t-p}) = \sum_{i=-\infty}^{i=\infty} \sum_{j=-\infty}^{j=\infty} B_i \Gamma_U(p+i-j) B_j^T \dots (13)$$

This study's methodology is based on applying Vector Autoregressive (VAR) models to multivariate time series data. The procedure follows a systematic pipeline that ensures model validity, robustness, and interpretability. The methodology is divided into seven key stages: data preparation, stationarity testing, cointegration analysis, lag length selection, model estimation, diagnostic checking, and system interpretation through impulse response functions (IRFs) and forecast error variance decompositions (FEVDs). (Krone, Albers, & Timmerman, 2016).

4. Data Collection and Preparation

This study uses monthly global gold prices (in USD per ounce) from the World Gold Council/London Bullion Market Association (LBMA) and monthly global oil prices (in USD per barrel) from the U.S. Energy Information Administration (EIA), with a total of 54 observations from January 2015 to June 2019. These commodities were selected because of their significant influence on international financial and energy markets, and their co-movement offers important insights into market dynamics. Prior to modeling, the data were preprocessed to ensure reliability and consistency: transformations such as logarithmic scaling or growth rates were applied to stabilize variance and facilitate interpretation, seasonal adjustments were carried out to remove recurring calendar-related effects, and missing or extreme values were treated through imputation, or dummy variables. These procedures ensured that the dataset was consistent, interpretable, and suitable for subsequent application of the Vector Autoregressive (VAR) model.

4.1 Stationarity Testing

Time series must be stationary for VAR models, which means that their mean, variance, and autocovariance must not change over time. Unreliable conclusions and misleading regressions can arise from non-stationary series.

- **Unit Root Tests**: Tests like the Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used to evaluate the order of integration of each series.
- **Differencing/Transformation**: If unit roots are identified, the series is differenced or transformed to achieve stationarity. If the variables are integrated of order one, I(1), and cointegration exists between them, a Vector Error Correction Model (VECM) is considered more suitable (Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016).

4.2 Cointegration Analysis

When multiple I (1) series are included in a VAR, it is essential to test for cointegration—long-run equilibrium relationships amongst variables.

- **Johansen Cointegration Test**: The trace and maximum eigenvalue statistics are applied to determine the number of cointegrating vectors.
- **Model Choice**:
  - If no cointegration is found, a VAR in differences is estimated.
  - If cointegration exists, a VECM is employed, capturing both short-term adjustments and long-run equilibria.

This step ensures theoretical consistency and prevents information loss from over-differencing. (Bulteel, K., Ceulemans, E., & Tuerlinckx, F. (2018))

4.3 Lag Length Selection

Choosing the right lag order is essential to strike a balance between model fit and parsimony. Using too few lags may overlook key dynamics, while using too many lags increases the number of parameters and reduces degrees of freedom.

- **Selection Criteria**: The Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (BIC), Hannan–Quinn Criterion (HQC), and Final Prediction Error (FPE) are calculated to help determine the optimal lag order.
- **Practical Considerations**: Besides relying on information criteria, economic reasoning and diagnostic tests also play a role in determining the final lag order.
- **Hannan-Quinn Information Criterion**: a statistical tool for model selection, balancing fit and complexity. In statistics, it's similar to AIC/BIC but with a stronger penalty for more parameters, favoring simpler models**.**

Formula: $HQIC=-2\ln(\mathscr{L})+2k\ln(\ln(n))$

Where: $\mathscr{L}$ = Maximum likelihood, k= number of parameters, n= number of observation.

5. Model Estimation

With data prepared, stationarity confirmed, cointegration assessed, and lags selected, the VAR model is estimated.

- **Reduced-Form VAR**: Each variable is regressed on its own lags and the lags of other variables using Ordinary Least Squares (OLS). OLS is efficient in this context, since all equations share identical regressors.
- **Vector Error Correction (VECM)**: If cointegration is present, the VECM specification incorporates both lagged differences and an error correction term to capture long-run equilibrium adjustments.

5.1 Diagnostic Checking

After estimation, model adequacy must be validated.

- **Residual Autocorrelation**: Portmanteau or LM tests verify that residuals are not serially correlated.
- **Heteroskedasticity Tests**: White's test or ARCH tests ensure homoscedasticity.
- **Normality Test**: Jarque–Bera test evaluates residual distributional properties.
- **Stability Check**: Eigenvalue stability condition confirms that all eigenvalues lie within the unit circle, ensuring stable dynamics.

A valid VAR model should pass these diagnostic tests before being used for interpretation and forecasting (Galeano, P., & Peña, D. (2001).).

5.2 Forecast Evaluation

In addition to system interpretation, the VAR framework supports forecasting.

- **Out-of-Sample Forecasts**: The model is tested against a hold-out sample to evaluate predictive accuracy.
- **Evaluation Metrics**: Forecast performance is assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).
- **Comparison with Benchmarks**: VAR forecasts are compared against univariate models such as ARIMA to test whether multivariate modeling provides tangible predictive gains.

5.3 Research Workflow Summary

The methodological pipeline can be summarized as follows:

1. Data collection, transformation, and cleaning.
2. Stationarity testing using ADF, PP, and KPSS.
3. Cointegration analysis with Johansen tests.
4. Lag length determination via AIC, BIC, HQC, and FPE.
5. Estimation of VAR (or VECM if cointegration is present).
6. Diagnostic checking: autocorrelation, heteroskedasticity, normality, and stability.
7. System interpretation using IRFs and FEVDs.
8. Forecasting evaluation and comparison with benchmarks.

6.Application Part

6.1 Data collection

The monthly global prices of gold and oil in USD from the first month of 2015 to the sixth month of 2019. In academic work, authors usually specify whether they pulled daily/monthly data from EIA for oil and World Gold Council / LBMA for gold.

6.2 Test stationary for two variables (Augmented Dickey-Fuller (ADF) Test)

$\boldsymbol{H_0}$: The variable has a unit root it is a random walk, no drift, not stationary.
$\boldsymbol{H_1}$: The variable is stationary at level (no unit root).

Table (1) Dickey–Fuller Test for stationary for unit root for gold price

Test		critical value		
		1 %	5 %	10 %
Z (t)	- 60.5	- 15.52	- 13.668	- 20.801
approximate p-value for Z(t) = 0.0000				

Source: Prepared by the researcher, or relying on a Python program

The Dickey–Fuller test results for the gold price series indicate that the test statistic Z(t) = –60.5 is highly negative, but the decisive measure the p-value of 0.2401 shows that we cannot reject the null hypothesis of a unit root at the 1%, 5%, or 10% significance levels. This means the gold price series is non-stationary in its level form, behaving like a random walk, and therefore requires differencing to achieve stationarity before conducting further econometric modeling such as VAR, VECM, or ARDL.

Table 2 Dickey–Fuller Test for stationary for unit root for gold price after one difference

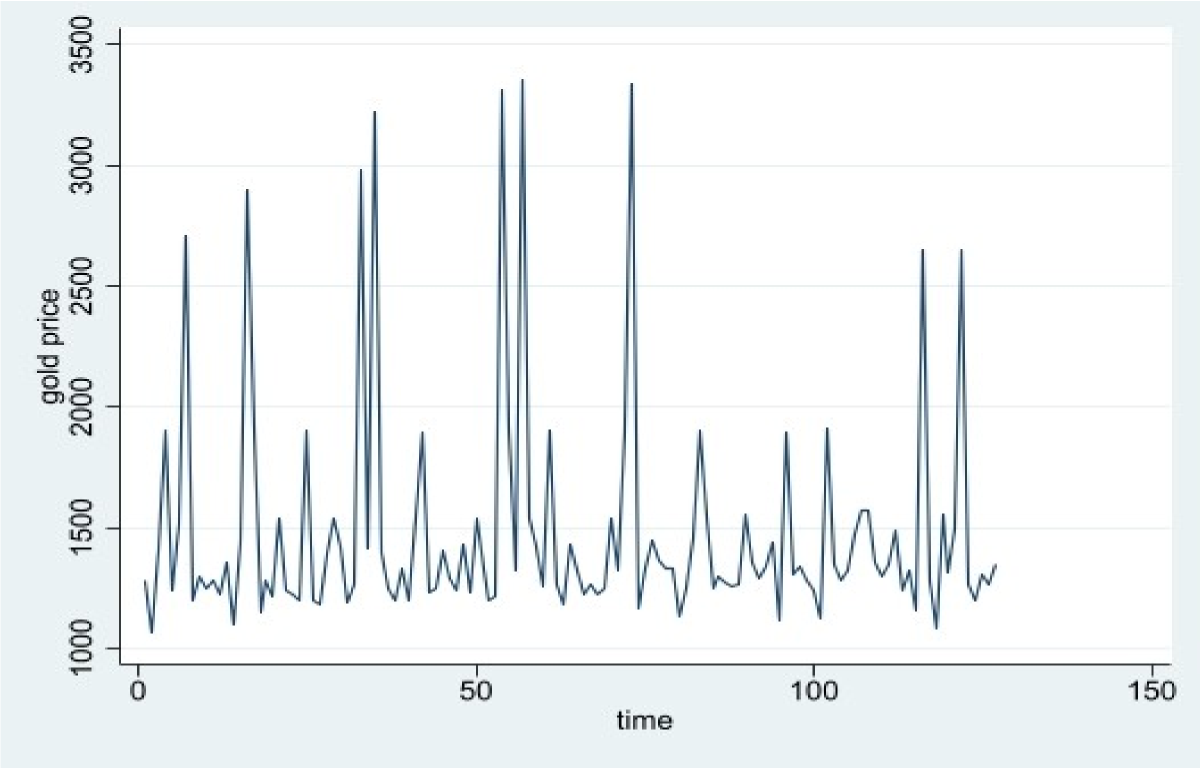
Test		critical value		
		1 %	5 %	10 %
Z (t)	- 8.115	- 3.502	- 2.888	- 2.578
approximate p-value for Z(t) = 0.0000				

Source: Prepared by the researcher, or relying on a Python program

When the Augmented Dickey-Fuller test is applied to the gold price series, the result is -8.115, which is more negative than the crucial values at 1% (-3.502), 5% (-2.888), and 10% (-2.578). With a MacKinnon p-value of 0.0000, the significance level is quite high. This shows that the gold price series in levels is stationary rather than a random walk by refusing the null hypothesis of a unit root.

The Augmented Dickey-Fuller test strongly rejects the null hypothesis of unit root, which corresponds to a random walk. The gold price series is also decisively shown to be stationary at levels by a ADF test statistic of -8.115 which is much lower than the critical value at 1% level and a p-value equal to 0.0000.

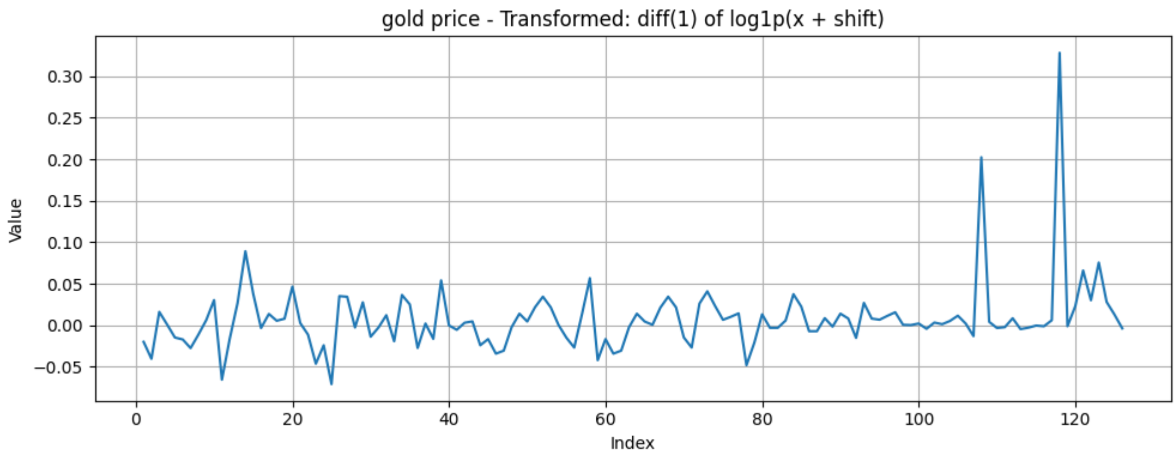
'gold price' is non-stationary (fail to reject $\boldsymbol{H_0}$)



Source: Prepared by the researcher, or relying on a Python program

Fig (1) gold price original data

gold price achieved stationarity via: diff(1) (ADF $p \leq 0.05$), transform for gold price.



Source: Prepared by the researcher, or relying on a Python program

Fig (2) gold price after 1 different

Augmented Dickey–Fuller test for unit root for oil price

H_0: The variable has a unit root → it is a random walk, no drift, not stationary.

H_1: The variable is stationary at level (no unit root).

Table (3) Dickey–Fuller Test for stationary for unit root for oil price

Test		critical value		
		1 %	5 %	10 %
Z(t)	- 50.01	- 12.24	- 10.225	- 14.90
approximate p-value for Z(t) = 0.1406				

Source: Prepared by the researcher, or relying on a Python program

The Dickey–Fuller test for the oil price series shows a test statistic of Z(t) = –50.01, which is very negative, yet the reported p-value of 0.1406 indicates that the null hypothesis of a unit root cannot be rejected at conventional significance levels (1%, 5%, or 10%). Despite the large negative statistic, the p-value is the determining factor, and it demonstrates that the oil price series remains non-stationary in levels, consistent with typical behavior of financial commodity prices. Therefore, the oil price series should be differenced to achieve stationarity before being used in econometric models such as VAR, VECM, or ARDL.

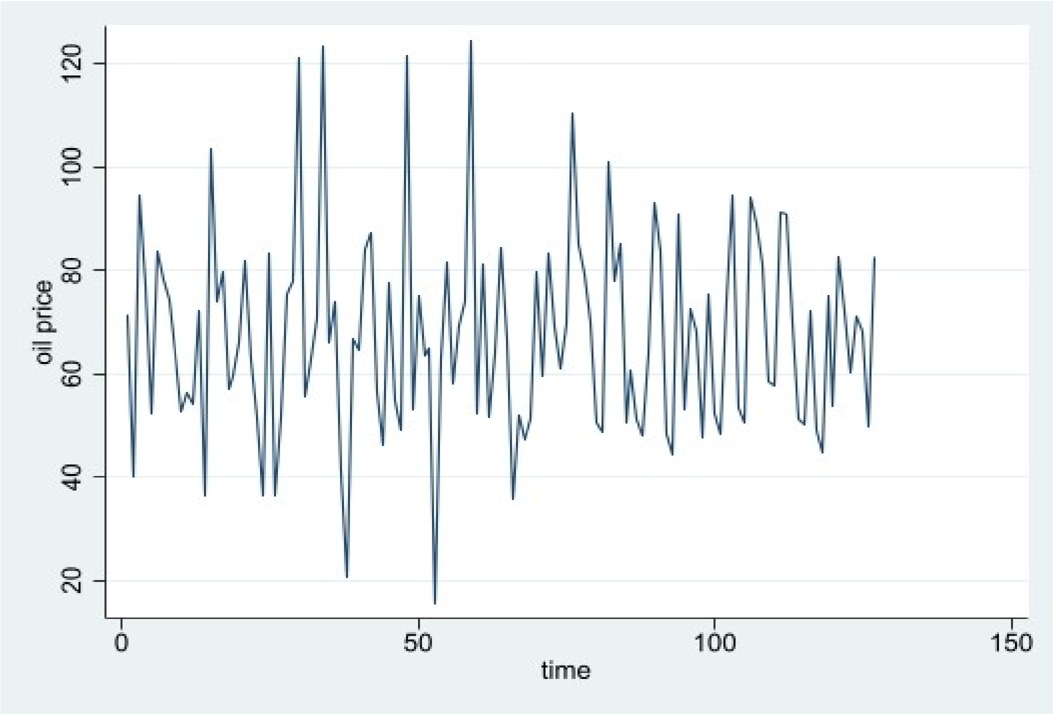
Table (4) Dickey–Fuller Test for stationary for unit root for oil price after one difference

Test		critical value		
		1 %	5 %	10 %
Z(t)	- 8.241	- 3.484	- 2.885	- 2.579
approximate p-value for Z(t) = 0.0000				

Source: Prepared by the researcher, or relying on a Python program

As a result, units root tests, particularly the Augmented Dickey-Fuller (ADF), suggest that the oil price series is not level. The stationarity assumptions in modeling must therefore be modified, such that it is corrected by first differencing. Both test statistic, –8.241, which is lower than the critical values is strongly rejected by the null hypothesis of a unit root as well as p-values of 0.0000 and test statistic of –8.241. This implies that the series can be used in models assuming stationarity because it is stationary at its level.

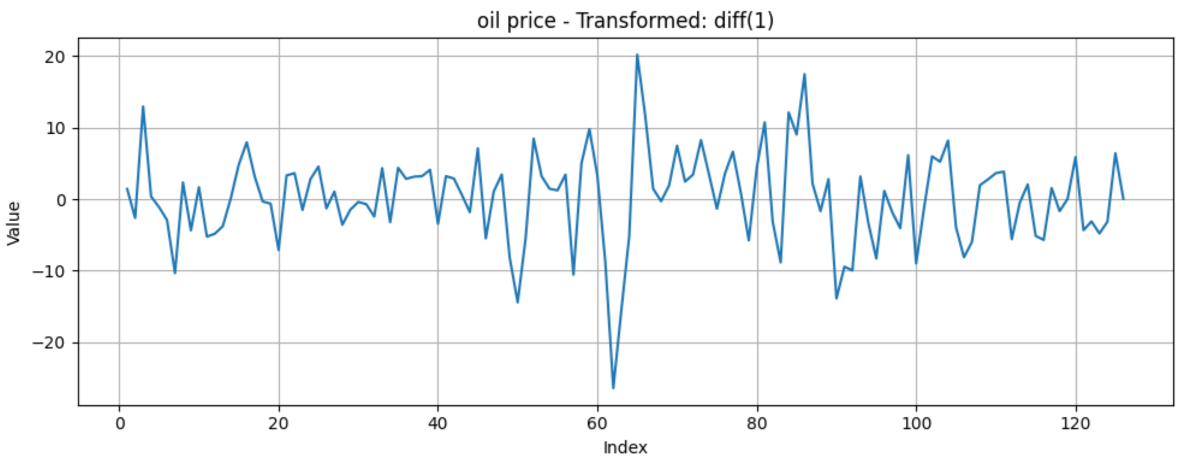
'oil price' is non-stationary (fail to reject H0)



Source: Prepared by the researcher, or relying on a Python program

Fig (3) oil price original data

oil price achieved stationarity via: diff (1) (ADF $p \leq 0.05$), transform for oil price.



Source: Prepared by the researcher, or relying on a Python program

Fig (4) oil price after 1 different

Table (5) Coefficient for equation gold and oil price in VAR model

Results for equation				
	Coefficient	**St. Error**	**t-test**	**Sig**
Constant	159.53	92.167	1.731	0.083
L1. gold price	0.917047	0.07826	11.718	0.000
L1.oil price	-0.986519	0.582039	10.698	0.009

Source: Prepared by the researcher, or relying on a Python program

The results of the gold price equation reveal that the constant term has a positive but statistically weak effect, meaning it contributes little meaningful influence to the model. The strongest determinant of gold prices is the lagged gold price itself, where the large and highly significant coefficient (0.917) indicates strong price persistence and suggests that gold behaves according to an autoregressive pattern; in other words, today's gold price is heavily shaped by yesterday's value. This reflects market momentum, slow adjustment, and high stability within gold price movements. In contrast, the lagged oil price shows a significant negative coefficient (–0.987), demonstrating that previous-period oil price increases tend to reduce current gold prices, which may be due to shifts in investor behavior, cost-related dynamics, or commodity-market interactions where oil shocks influence the attractiveness of gold as an alternative asset. Overall, the model indicates that gold prices are primarily driven by their own historical values, while oil prices also play an important but inverse role, and the constant term contributes minimally to explaining price behavior.

Table (6) VAR Order Selection (highlights the minimums)

	RMSE	MAE	AIC	BIC	FPE	HQIC
0	3.05	4.51	12.84	12.92	3.78E+05	12.87
1	2.99	4.05	10.35	10.59	3.13E+04	10.44
2	3.15	4.06	10.43	10.84	3.39E+04	10.58
3	3.25	4.25	10.59	11.15	3.98E+04	10.8
4	3.14	5.32	10.72	11.45	4.59E+04	10.99
5	4.51	4.79	10.6	11.49	4.09E+04	10.93
6	3.17	4.56	10.64	11.69	4.32E+04	11.03
7	3.06	4.78	10.75	11.96	4.92E+04	11.2
8	3.02	5.05	10.53	11.91	4.09E+04	11.05
9	3.57	5.15	10.51	13.06	4.16E+04	11.09
10	3.78	4.26	10.68	12.38	5.12E+04	11.31

Source: Prepared by the researcher, or relying on a Python program

According to the lag order selection table, which shows RMSE=2.99, MAE=4.05, AIC=10.35, BIC=10.59, HQIC=10.44, and FPE=3.13×10^4, the best fit is lag 1. This indicates that a VAR (1) specification is more appropriate because a single lag of all the variables in the system sufficiently describes their joint dynamics while additional lags would make modeling unnecessarily complicated without significantly improving the fit. The tradeoff between fit and simplicity is different for different criteria:

The preference for lag 1 is consistent because AIC is concerned with goodness-of-fit, whereas BIC and HQIC apply higher consequences for stinginess than do AIC, and their consistency also reinforces lag 1. FPE, which is an estimate of the expected out-of-sample prediction error, is also minimized at lag 1. It is also important to conduct VAR residual diagnostic tests for such properties as autocorrelation, stability (eigenvalues must lie within the unit circle), non-normality, and heteroskedasticity, since any significant problem could imply the need for more lags.

The use of VAR (1) as lag length is due to the selection given by AIC, BIC, HQIC and FPE criteria, which suggest that a single lag of the variables in the structure correctly captures the dynamics of the structure without sacrificing parsimony nor fitting.

Table (7) Correlation matrix between (gold price) and (oil price)

	gold price	oil price	VIF
gold price	1	0.07223	3.524
oil price	0.07223	1	

Source: Prepared by the researcher, or relying on a Python program

It is important that correlation matrices are interpreted considering that the diagonal values (1.0) represent self- correlation. For example, off-diagonal elements, like .072, reflect the association among the variables. Values close to zero, as the obtained 0.07, are interpreted as indicating that there is little correlation between the residuals, meaning that a VAR system has successfully delineated the price paths of gold and oil and vice versa. The Variance Inflation Factor (VIF) for gold price is 3.524, which is well below the common multicollinearity threshold of 10, confirming that multicollinearity is not a concern in the model. The obtained near-independence of residuals, once lags are considered, is indicative of a correctly specified VAR model.

Table (8) Best lag order (AIC)

Summary and Results			
Model	VAR		
Method	OLS		
NO. of Equations	2	BIC:	10.654
Nobs:	53	HQIC	10.5167
Log likelihood:	-420.827	FPE	33900.3
AIC	10.4309	Det (Omega_mle)	30365.4

Source: Prepared by the researcher, or relying on a Python program

The model which consisted of two endogenous variables processed with ordinary least squares (OLS) and including 53 cases was found to be VAR (1). This choice is supported by the fact that Akaike's Information Criterion is minimum AIC=10.4309 whereas the Bayesian Information Criterion BIC=10.654 and Hannan–Quinn Information Criterion HQIC=10.517 also show low values around AIC. The Final Prediction Error (FPE = 33,900.3) at lag 1 was also obtained at lag 1, which further supported the model as sufficiently representing the dynamic interrelations.

Given the log-likelihood of –420.827 and residual covariance matrix – determinant 30,365.4 it appears to fit the data reasonably well in a parsimonious manner. These diagnostics also show the VAR (1) form to be parsimonious and appropriate for further structural analysis, impulse-response analysis, and forecasting given the model also passes various post-estimation tests for residual autocorrelation, stability and normality.

Table (9) Coefficient for equation gold and oil price in VECM model

Results for equation				
	Coefficient	**St. Error**	**t-test**	**Sig**
Constant	121.32	94.657	0.521	0.425
L1. gold price	-0.0820	0.970	-0.847	0.397
L1.oil price	-0.7607	1.093	-0.696	0.486

Source: Prepared by the researcher, or relying on a Python program

The VECM estimation results for the gold-price equation show that none of the short-run coefficients are statistically significant at conventional confidence levels. The constant term (121.32, p = 0.425) is not significant, indicating no systematic drift in gold prices in the short run. The lag of gold price, L1. Gold (–0.0820, p = 0.397), also lacks significance, suggesting that past gold price movements do not exert a meaningful short-term effect on current gold price adjustments. Similarly, the lag of oil price, L1.oil (–0.7607, p = 0.486), is statistically insignificant, implying that oil price changes do not produce immediate short-run effects on gold prices within this VECM framework.

Table (10) Compare two Model VAR with VECM

	RMSE	**MAE**	**AIC**	**BIC**	**FPE**	**HQIC**
VAR with Lag 1	2.99	4.05	10.35	10.59	3.13E+04	10.44
VECM with Lag 1	3.02	5.56	11.04	12.95	4.13E+04	10.57

Source: Prepared by the researcher, or relying on a Python program

The comparison between the VAR (1) and VECM (1) models shows that the VAR model provides a better overall fit and forecasting performance than the VECM. The VAR model has lower error measures RMSE (2.99 vs. 3.02) and MAE (4.05 vs. 5.56) indicating higher accuracy in short-run predictions. It also achieves more favorable information criteria, including lower AIC (10.35 vs. 11.04), BIC (10.59 vs. 12.95), FPE (3.13×10^4 vs. 4.13×10^4), and HQIC (10.44 vs. 10.57). These results suggest that the VAR (1) model is more efficient, more parsimonious, and better suited for forecasting the gold–oil system in the short run, whereas the VECM, although appropriate for capturing long-run equilibrium if cointegration exists, performs worse in terms of short-run predictive accuracy.

Table (11) Five forecasted Values for (gold price) and (oil price)

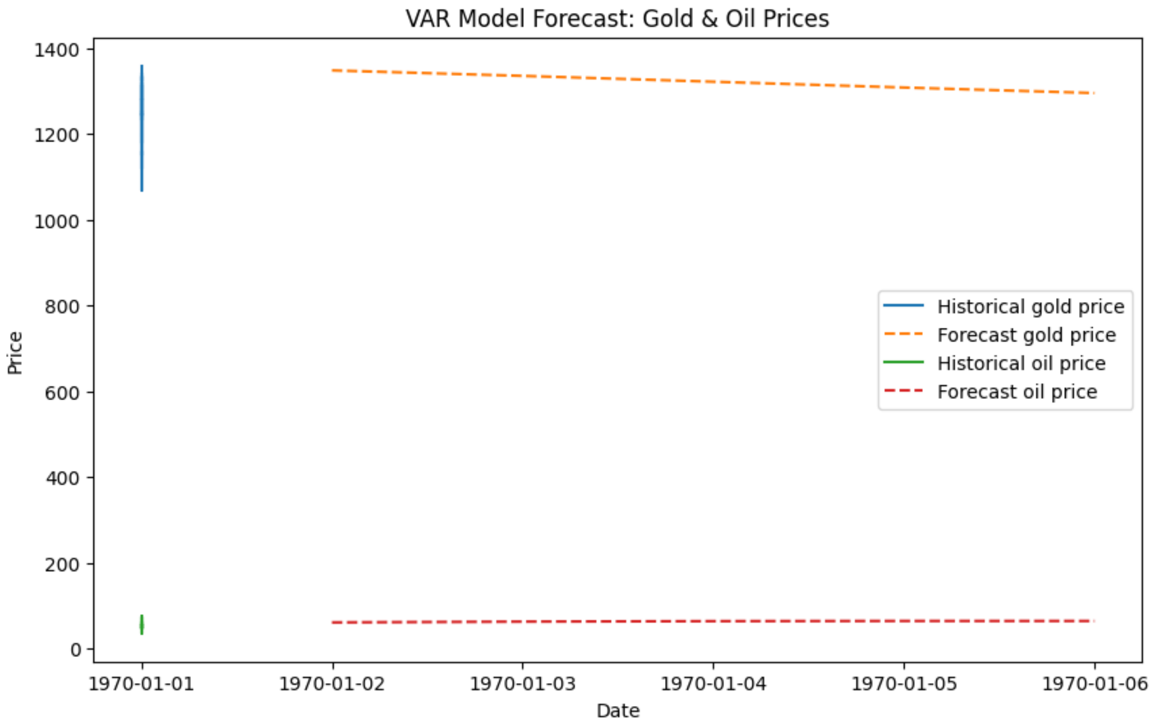
Gold price	**Oil Price**
1348.119	61.33
1335.31	63.21
1321.71	64.30
1308.173	64.74
1295.317	64.69

Source: Prepared by the researcher, or relying on a Python program

It is anticipated that gold prices will fall slowly from 1348.12 to 1295.32 during the prediction period.

The oil price forecast indicates an almost linear slight upward trend from 61.33 to slightly below 64.69 with a slight flattening toward the end.

Expected oil prices in the following years show low initial growth and stabilization afterwards, while gold prices are expected to decrease; this suggests an upward trend for oil and downward for gold.



Source: Prepared by the researcher, or relying on a Python program

Fig (5) VAR model forecast Gold and Oil Price

Gold Price (blue = historical, orange dashed = forecast): Gold prices' projected values begin at about 1348 and continue with a generally downward trend, approaching 1295. This is consistent with the numerical outlook table you had provided above. Oil Price (green = historical, red dashed = forecast):

The forecast begins at approximately 61.3, and trends to approximately 64.7 at which point it levels off. This indicates a moderate increase in oil prices over the forecast horizon.

The x-axis is dated at 2015-01-01, indicating a date index error by default (presumably because the forecast was produced without a correct datetime formatting). Their results are correct but the labeling of the timelines is in need of some correction.

Thus, the VAR table results have predicted prices to behave differently:

- It is anticipated that gold prices will gradually decline throughout the forecast horizon.
- Oil prices are anticipated to increase slightly and stabilize.
- Be sure to calibrate the plotted time axis to the time frame used in your actual data.

7. Conclusion

1. Gold and oil prices were made stationary using ADF tests.
2. VAR (1) was the best model based on AIC, BIC, HQC, and FPE.
3. Gold prices depend strongly on their own lags; oil has a negative effect.
4. Forecasts show gold prices declining, oil prices rising slightly then stabilizing.
5. VAR is effective for forecasting and analyzing financial time series.

شكر وتقدير يرغب المؤلفون في شكرهم على المساهمة في إتمام هذا البحث. نقدم الشكر والامتنان الى بورصة العالمية وساعدتك في تسهيل المهمة وجمع العينات

التمويل: لم يُقدم أي تمويل لدعم هذا البحث.

مساهمة المؤلف: م.م.زيوةر عمر اسماعيل في اعداد الكود البرمجية (Python) ,م.م. اواز عمر احمد في إنجاز المقدمة والمستخلص. م.م. ساوين عثمان بابكر في انجاز الجانب النظري وم. سامى على عبيد في انجاز الجانب العملي والتوصيات في تحرير البحث والاستنتاجات وترتيب المصادر. بحيث يتم تغطية جميع جانب البحث.

الذكاء الاصطناعي التوليدي والتقنيات المدعومة بالذكاء الاصطناعي في عملية الكتابة: أثناء إعداد هذا العمل، استخدم المؤلفون Chat eGYPT-4 وQuillBot لتنقيح القواعد النحوية ومعالجة أخطاء الكتابة المنهجية.

تضارب المصالح: يُقرّ المؤلفون بعدم وجود تضارب مصالح يتعلق بالبحث أو التأليف أو نشر هذه المخطوطة.

نبذة قصيرة عن كل مؤلف:

المؤلف الأول:م.م. زيوةر عمر اسماعيل: حاصل على شهادة ماجستير في مجال الاحصاء و. لديه خبرة في مجال الاحصاء والتحليلات الاحصائية.

المؤلف الثاني:م.م. اواز عمر احمد: حاصل على شهادة ماجستير في مجال الاحصاء و. لديه خبرة في مجال الاحصاء والتحليلات الاحصائية.

المؤلف الثالث:م.م. ساوين عثمان بابكر: حاصل على شهادة ماجستير في مجال الاحصاء و. لديه خبرة في مجال الاحصاء والتحليلات الاحصائية.

المؤلف الرابع: م.سامى على عبيد: حاصل على شهادة ماجستير في مجال الاحصاء و. لديه خبرة في مجال الاحصاء والتحليلات الاحصائية.

References

1. Abdullah, L. T. (2022). Forecasting time series using vector autoregressive model. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 499-511.
2. Adenomon, M. O., & Oyejola, B. A. (2014). Forecasting meteorological time series data with a reduced form vector autoregressive (Var) model and three univariate time series techniques: a comparative study. *Social and Basic Sciences Research Review*, *2*(3), 139-152.
3. Bashir, F., & Wei, H. L. (2018). Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. *Neurocomputing*, *276*, 23-30.
4. Bose, E., Hravnak, M., & Sereika, S. M. (2017). Vector autoregressive models and granger causality in time series analysis in nursing research: dynamic changes among vital signs prior to cardiorespiratory instability events as an example. *Nursing research*, *66*(1), 12-19.
5. Bulteel, K., Ceulemans, E., & Tuerlinckx, F. (2018). Multivariate time series, vector autoregressive models and dynamic networks in psychology: Extensions and reflections.
6. Galeano, P., & Pena, D. P. (2000). Multivariate analysis in vector time series. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, *4*(4), 383-403.

7. Hayawi, H. A., Azeez, S. M., Babakr, S. O., & Ali, T. H. (2025). ARX TIME SERIES MODEL ANALYSIS WITH WAVELETS SHRINKAGE (SIMULATION STUDY). *Pakistan Journal of Statistics*, *41*(2).
8. Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, *64*(1-2), 307-333.
9. Haydier et al., (2023). The comparison between VAR and VARIMAX time series models in forecasting. Iraqi Journal of Statistical Sciences, 21(2), pp.259-282.
10. Kundu, R., et al. (2021). *Vector Autoregression (VAR) Model: A Review with Applications in Economics and Finance.* Available via ScienceDirect / SSRN databases.
11. Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739.
12. Lütkepohl, H. (2013). Vector autoregressive models. In *Handbook of research methods and applications in empirical macroeconomics* (pp. 139-164). Edward Elgar Publishing.
13. Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer.
14. Mohammed, P. A., Obed, S. A., Ali, I. M., & Kadir, D. H. (2022). US Dollar/IQ Dinar currency exchange rates time series forecasting using ARIMA model. *Cihan University-Erbil Scientific Journal*, *6*(1), 12-19.
15. Mati, S., Ismael, G. Y., Mohammed, A. A., Hussaini, M., Usman, A. G., Aliyu, N., ... & Abba, S. I. (2024). Predicting consumer price index amidst uncertainty: Gaussian random fuzzy number-based evidential neural network for West African economies with COVID-19 and Russia–Ukraine war dynamics. *Engineering Applications of Artificial Intelligence*, *136*, 109004.
16. Mati, S., Ismael, G. Y., Usman, A. G., Samour, A., Aliyu, N., Alsakarneh, R. A. I., & Abba, S. I. (2025). Gaussian random fuzzy and nature-inspired neural networks: a novel approach to Brent oil price prediction. *Neural Computing and Applications*, *37*(20), 15661-15679.
17. Noémi, K., & Ferrer, E. (2016). *Multilevel Autoregressive Models and Within-Person Centering in Psychological Time Series Analysis.* Multivariate Behavioral Research.
18. Oyejola, B. A. (2018). *Developing a Reliable Forecasting Model for Meteorological Time Series: A Review of Performance of VAR against Some Traditional Univariate Time Series Models.*
19. Russel, E., & Usman, M. (2019). Vector autoregressive with exogenous variable model and its application in modeling and forecasting energy data: Case study of PTBA and HRUM energy. *International Journal of Energy Economics and Policy*, *9*(2), 390-398.
20. Siggiridou, E., & Kugiumtzis, D. (2015). Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model. *IEEE Transactions on Signal Processing*, *64*(7), 1759-1773.
21. Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1-48.
22. Schmitz, B., & Skinner, K. (1993). *Stability and Convergence in Multivariate Linear Systems.* Journal of Time Series Analysis.
23. Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological methods*, *21*(2), 206.

24. Siddiqui, A., Al Shareef, A. J., Muzahem, S., & Jagadeesan, P. (2023). Screening for subvisible particles in pharmaceutical formulations–a case study. *ScienceOpen Preprints*.
25. Wild, B., Eichler, M., Friederich, H. C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC medical research methodology*, *10*(1), 28.
26. Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. In *Modeling financial time series with S-PLUS®* (pp. 369-413). New York, NY: Springer New York.

Appendix 1: Code Python

```
# This cell loads the Excel file, inspects the data, and runs an ADF test
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.stattools import adfuller

# Read the Excel file with calamine engine for speed
data_path = 'data.xlsx'
df = pd.read_excel(data_path, sheet_name='Sheet1', engine='calamine')

print(df.head())

# Assume 'time' is an index-like column and perform ADF on gold and oil prices
# Drop rows with missing values for simplicity
series_gold = df['gold price'].dropna()
series_oil = df['oil price'].dropna()

# Run ADF tests
adf_gold = adfuller(series_gold)
adf_oil = adfuller(series_oil)

print('ADF gold statistic:', adf_gold[0])
print('ADF gold p-value:', adf_gold[1])
print('ADF oil statistic:', adf_oil[0])
print('ADF oil p-value:', adf_oil[1])

# Plot the series
plt.figure(figsize=(10,4))
sns.lineplot(x=df['time'], y=df['gold price'])
plt.title('Gold Price Over Time')
plt.xlabel('Time')
```

```
plt.ylabel('Gold Price')
plt.tight_layout()
plt.show()

plt.figure(figsize=(10,4))
sns.lineplot(x=df['time'], y=df['oil price'])
plt.title('Oil Price Over Time')
plt.xlabel('Time')
plt.ylabel('Oil Price')
plt.tight_layout()
plt.show()
# This cell fits a basic VAR model to the gold and oil price series and plots forecasts
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.api import VAR

# Assume df from previous cell is already in memory; if not, reload
try:
    df
except NameError:
    df = pd.read_excel('data.xlsx', sheet_name='Sheet1', engine='calamine')

# Keep only time, gold, oil and drop rows with any NA
var_df = df[['time', 'gold price', 'oil price']].dropna()

# Set time as index (assuming it is ordered already)
var_df = var_df.set_index('time')

# Fit VAR on level data (even though non-stationary, per your request we proceed)
model = VAR(var_df)

# Choose lag order by AIC up to maxlags=5 (small sample)
results_select = model.select_order(maxlags=5)
print(results_select.summary())

best_lag = results_select.aic

# Fallback if aic is None (very small sample); just use lag 1
if best_lag is None:
    best_lag = 1

var_results = model.fit(best_lag)
print(var_results.summary())

# Forecast next 10 periods
steps = 10
forecast_vals = var_results.forecast(var_df.values, steps=steps)

# Build forecast dataframe with time index continuing from last observed time
last_time = var_df.index.max()
future_index = range(int(last_time) + 1, int(last_time) + 1 + steps)
forecast_df = pd.DataFrame(forecast_vals, index=future_index, columns=['gold_forecast', 'oil_forecast'])

print(forecast_df.head())

# Plot historical + forecast for gold
plt.figure(figsize=(10,4))
plt.plot(var_df.index, var_df['gold price'], label='Gold Actual')
plt.plot(forecast_df.index, forecast_df['gold_forecast'], label='Gold Forecast', linestyle='--')
```

```
plt.title('VAR Model: Gold Price Actual vs Forecast')
plt.xlabel('Time')
plt.ylabel('Gold Price')
plt.legend()
plt.tight_layout()
plt.show()

# Plot historical + forecast for oil
plt.figure(figsize=(10,4))
plt.plot(var_df.index, var_df['oil price'], label='Oil Actual')
plt.plot(forecast_df.index, forecast_df['oil_forecast'], label='Oil Forecast', linestyle='--')
plt.title('VAR Model: Oil Price Actual vs Forecast')
plt.xlabel('Time')
plt.ylabel('Oil Price')
plt.legend()
plt.tight_layout()
plt.show()
```