Multivariate Forecasting of GAU/IDR Gold Prices for the 2021-2026 Period Using a Hybrid Prophet-LSTM Model Based on Macroeconomic Indicators

Safira Khoirulanisa Salsabila[1*], Nur Achmey Selgi Harwanti[2]
[1]Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
[2]Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: fsalsafira@students.unnes.ac.id

Received: 17 June 2026
Revised: 18 August 2026
Accepted: 24 August 2026

ABSTRACT – **The domestic gold price (GAU/IDR) from January 2021 to April 2026 shows a long-term upward trend with non-stationary and volatile characteristics influenced by global and domestic macroeconomic conditions. This study aims to analyze the historical characteristics of the GAU/IDR price and evaluate the performance of the Hybrid Prophet-LSTM multivariate model for gold price forecasting. This model employs a residual learning approach, in which Prophet captures trends and temporal patterns while incorporating seven exogenous variables (namely GAU/USD, DXY, the Federal Funds Rate, the BI Rate, Indonesian inflation, crude oil prices, and USD/IDR), while LSTM learns nonlinear patterns in Prophet's residuals. Daily data from January 2021 to April 2026 were used, with a 90-day conditional out-of-sample evaluation from January 31 to April 30, 2026. The Hybrid Prophet-LSTM model achieved an RMSE of IDR 28,478, an MAE of IDR 22,270, a MAPE of 0.8476%, and an R^2 of 0.943405, outperforming both the multivariate Prophet model and the multivariate LSTM model. These results indicate that residual learning improves the accuracy of GAU/IDR price forecasts based on the evaluated data and the testing period.**

Keywords – **GAU/IDR, Hybrid Prophet-LSTM, Macroeconomic Indicators, Multivariate Forecasting, Residual Learning.**

I. INTRODUCTION

Gold is an investment instrument that serves as a store of value and a hedge, especially during times of economic uncertainty [1]. Gold prices can be influenced by changes in inflation, financial market conditions, monetary policy, and geopolitical uncertainty. From 2024 through early 2026, global economic uncertainty, changes in monetary policy, exchange rate fluctuations, geopolitical tensions, and rising demand for safe-haven assets will all influence gold price movements [2]. In Indonesia, these dynamics are reflected in domestic gold prices denominated in rupiah (GAU/IDR), which show a long-term upward trend accompanied by volatility during the period from January 2021 to April 2026. These characteristics highlight the importance of forecasting GAU/IDR to understand domestic gold price dynamics and provide quantitative information that can support data-driven decision-making.

Movements in the GAU/IDR are not only linked to global gold prices but also to global and domestic macroeconomic conditions. Domestic gold prices are directly linked to global gold prices (GAU/USD) and the USD/IDR exchange rate, while other indicators can influence their movements through monetary mechanisms, exchange rates, inflation, and commodity markets [3]. Therefore, this study uses seven exogenous macroeconomic variables, namely the global gold price (GAU/USD), the U.S. Dollar Index (DXY), the Federal Funds Rate, the Bank Indonesia (BI) interest rate, Indonesian inflation, crude oil prices, and the USD/IDR exchange rate. The use of these seven variables forms a multivariate forecasting framework that takes global and domestic economic factors into account when modeling movements in GAU/IDR.

Forecasting gold prices presents challenges because financial time series data can exhibit nonstationarity, volatility, nonlinear relationships, and temporal dependence. These characteristics mean that a single forecasting method alone may not be able to capture all the patterns present in the data. Prophet is capable of modeling trends, seasonality, and long-term temporal patterns, and allows for the inclusion of additional predictor variables [4]. However, Prophet has limitations in capturing complex nonlinear patterns. In contrast, Long Short-Term Memory (LSTM) is designed to learn nonlinear relationships and long-term dependencies in sequential data, but using LSTM on its own may not necessarily be able to optimally represent certain trend structures [5]. These differences in characteristics form the basis for combining Prophet and LSTM in a hybrid forecasting approach.

Various previous studies have applied statistical and machine learning methods to gold price forecasting, including LSTM, Prophet, the CNN-LSTM-GRU model, and deep learning-based hybrid models. Other studies have also integrated economic, financial, and sentiment information into gold price modeling. In addition, the Hybrid Prophet-LSTM approach, which is based on residual learning, has been applied to other forecasting fields. Nevertheless, there remains room for development in forecasting domestic gold prices (GAU/IDR). Several previous studies have focused on global gold prices such as XAU/USD, while applications to GAU/IDR remain limited. Only a few studies have simultaneously integrated seven macroeconomic indicators, a residual learning approach, and a mechanism to maintain level continuity between historical data and operational forecast results. Thus, the contribution of this study lies not in the creation of an entirely new Prophet-LSTM architecture, but rather in its application to GAU/IDR forecasting through multivariate macroeconomic information, residual learning, and level adjustment during the forecasting phase.

To address these issues, this study applies a Hybrid Prophet-LSTM model with a multivariate Prophet component and a residual LSTM based on residual learning. Prophet is used as the main component to capture trends and temporal patterns by incorporating seven macroeconomic indicators as additional regressors. The residuals generated by Prophet are then modeled using LSTM to capture nonlinear patterns that cannot yet be explained by Prophet. The final forecast from this hybrid model is formed by combining the Prophet forecast with the residual correction generated by the LSTM. For the 30-day operational forecast through May 2026, a threshold adjustment is applied to align the initial forecast level with the most recent actual observation. These adjustments are used solely to maintain level continuity during the transition to the forecasting period and are not used to improve or evaluate the model's accuracy outside the sample.

Based on this framework, this study aims to analyze the historical characteristics of the GAU/IDR gold price and its relationship with seven exogenous macroeconomic indicators for the period from January 2021 to April 2026, apply and evaluate the Hybrid Prophet-LSTM model (which incorporates a multivariate Prophet component and a residual LSTM based on residual learning), and compare its performance with that of the multivariate Prophet model and the multivariate LSTM model using RMSE, MAE, MAPE, and R^2. Additionally, this study evaluates the transition of 30-day operational forecast results using threshold adjustments. Through this approach, this study examines the ability of residual correction using LSTM to improve the performance of the multivariate Prophet model in forecasting domestic gold prices (GAU/IDR).

II. LITERATURE REVIEW

A. Gold Prices and Macroeconomic Indicators

Gold is viewed not merely as a precious metal commodity, but as a financial instrument with two main theoretical foundations: the safe-haven theory and the theory of gold's relationship with macroeconomic variables. A safe haven is defined as an asset that maintains or increases its value when financial markets experience systemic shocks, while gold's ability to serve as a hedge tends to be more consistent than its role as a safe haven, which is conditional in nature [6]. The domestic gold price in rupiah (GAU/IDR) is derived from the global gold price (GAU/USD) using the USD/IDR exchange rate; mathematically, it is expressed as follows:

$$P\left(\frac{GAU}{IDR}\right) = P(GAU/USD) \times E(USD/IDR) \qquad (1)$$

where P(GAU/IDR) is the domestic price of gold per gram, P(GAU/USD) is the global price of gold in U.S. dollars, and E(USD/IDR) is the exchange rate of the rupiah against the U.S. dollar. Consequently, the volatility of GAU/IDR is determined simultaneously by fluctuations in the global gold price and movements in the rupiah exchange rate, resulting in two sources of uncertainty that interact with one another. In addition to the exchange rate, interest rates (the Federal Funds Rate and the BI Rate) influence the price of gold through opportunity costs and inflation expectations; inflation affects purchasing power and gold's role as a hedge; while crude oil prices are related to production costs and macroeconomic pressures in a nonlinear and regime-dependent manner.

B. Characteristics of Financial Time Series Data

Daily gold price data exhibit distinctive statistical characteristics: time-varying volatility (heteroscedasticity), non-stationarity, non-linearity, and temporal dependence/long memory. A time series is said to be stationary if its mean, variance, and autocovariance remain constant over time; financial asset price data are generally non-stationary because it follows a unit root process, which can be tested using the Augmented Dickey-Fuller (ADF) test [7]. The relationship between gold prices and macroeconomic variables is also not always linear; there is regime switching that cannot be fully captured by linear models such as VAR/ARIMA without structural modifications. Therefore, neural networks such as LSTM are needed as universal approximators capable of learning nonlinear relationships without pre-specifying the functional form [8].

C. Multivariate Forecasting

Time series forecasting can be performed using a univariate approach, which bases predictions solely on the historical values of the variable itself, or a multivariate approach, which explicitly includes exogenous variables as additional predictors [9]. A multivariate approach is equivalent to relaxing the efficient-market hypothesis and acknowledging that macroeconomic fundamentals have predictive content regarding gold prices [10]. This study uses seven exogenous variables as additional regressors in the Prophet model, while the LSTM in the hybrid model is used univariately to examine the residuals generated by Prophet.

D. Prophet Model

Prophet is an additive decomposition-based forecasting model that estimates nonlinear trends and seasonal fluctuations while accommodating the effects of holidays [11]. The general formula is [12]:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \qquad (2)$$

with $y(t)$ is the time series value at time t, $g(t)$ is a trend component, $s(t)$ is a seasonal component, $h(t)$ is the holiday effect, and $\varepsilon(t)$ is a normally distributed error term. The piecewise linear trend component is expressed as:

$$g(t) = (k + a(t)^T\delta) \cdot t + (m + a(t)^T\gamma) \qquad (3)$$

where k is the baseline growth rate, δ is the growth rate adjustment vector at each changepoint, m is the offset, and $a(t)$ is the active changepoint indicator vector. The seasonal component is modeled using a Fourier series [12]:

$$s(t) = \sum_{n=1}^{N}\left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right) \qquad (4)$$

where P is the seasonal period and N is the Fourier order. Prophet's advantages include its ability to easily handle missing values, its robustness to outliers, and the interpretability of its decomposition components; its limitation is the assumption of linear interactions among components, which prevents it from capturing complex nonlinear relationships, a gap that is filled by the LSTM component in the hybrid architecture. [13].

E. Residual Learning and the Prophet-LSTM Hybrid Model

Residual learning uses shortcut connections to overcome the limitations of the base model. Let $y(t)$ be the actual value and $\hat{y}_{Prophet(t)}$ be Prophet's prediction; the residual is defined as [14]:

$$r(t) = y(t) - \hat{y}_{Prophet(t)} \quad (5)$$

The LSTM component is then trained to learn the residual mapping function for historical features:

$$r(t) \approx f_{LSTM}\left(X_{(t-L):t}\right) \quad (6)$$

where $X_{(t-L):t}$ is the feature matrix over a historical window of length L steps. The final prediction of the hybrid model is the sum of the two:

$$\hat{y}_{hybrid(t)} = \hat{y}_{Prophet(t)} + f_{\mathrm{LSTM}}\left(X_{(t-L):t}\right) \quad (7)$$

The Prophet-LSTM hybrid architecture follows the sequential stacking ensemble paradigm: Prophet acts as a level-1 base learner that captures deterministic components (trends, seasonality, and regressor effects), while LSTM acts as a level-2 meta-learner that learns from the residual signal. From the perspective of the bias-variance tradeoff, Prophet has high structural bias but low variance, whereas LSTM has the potential to reduce bias at the cost of higher variance; training the LSTM solely on the residuals (rather than the full data) effectively separates the tasks of the two models [13].

F. Long Short-Term Memory (LSTM)

LSTM is a variant of the Recurrent Neural Network (RNN) designed to address the vanishing gradient and exploding gradient problems in conventional RNNs through a three-gate mechanism (forget, input, and output gates) [15]. The mathematical equation for one time step is as follows.

$$f_t = \sigma\left(W_f\left[h_{(t-1)}, x_t\right] + b_f\right) \quad (8)$$

$$i_t = \sigma\left(W_i\left[h_{(t-1)}, x_t\right] + b_i\right) \quad (9)$$

$$\tilde{C}_t = \tanh\left(W_c\left[h_{(t-1)}, x_t\right] + b_c\right) \quad (10)$$

$$C_t = f_t \odot C_{(t-1)} + i_t \odot \tilde{C}_t \quad (11)$$

$$o_t = \sigma\left(W_o\left[h_{(t-1)}, x_t\right] + b_o\right) \quad (12)$$

$$h_t = o_t \odot \tanh(C_t) \quad (13)$$

where x_t is the input vector at time t, $h_{(t-1)}$ is the previous hidden state, $C_{(t-1)}$ is the previous cell state, W and b are the learned parameters, σ is the sigmoid function, and ⊙ denotes element-wise multiplication (Hadamard product). The forget gate determines the proportion of old information that is retained; when this gate is close to 1, the gradient flows through the constant error carousel (CEC) with little attenuation:

$$\frac{\partial C_t}{\partial C_{(t-1)}} = f_t \approx 1 \quad (14)$$

As a result, LSTM is capable of learning dependencies spanning hundreds of time steps, which is relevant for gold price data that exhibits long-memory patterns.

G. Recursive Forecasting and Forecasting Continuity

In multi-step operational forecasting, future residuals cannot be observed directly and must be estimated recursively through the trained LSTM network across the forecast horizon. However, transitioning from historical observations to multi-step recursive projections often introduces a level discrepancy or vertical jump at the forecast boundary. To mitigate this discontinuity, a boundary-condition-based additive correction is applied. The offset δ is calculated as the difference between the latest actual observation y_T and the model's prediction at the same time step $\hat{y}_T$, which is then uniformly added to all future projections:

$$\hat{y}_{corrected}(T+h) = \hat{y}(T+h) + \delta, \quad \delta = y_T - \hat{y}_T \quad (15)$$

Applying the offset δ ensures a smooth transition from the historical period to the future projection without altering the relative patterns of the model's predicted outcomes.

H. Model Evaluation Criteria

The performance of the forecasting models was evaluated using the following four quantitative metrics.

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (17)$$

$$MAPE = \left(\frac{100}{n}\right)\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (18)$$

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] \quad (19)$$

where y_i represents the actual gold price at time index i, $\hat{y}_i$ is the corresponding forecasted value, $\bar{y}$ denotes the mean of the actual gold prices over the evaluation period, and n is the total number of out-of-sample observations ($n = 90$).

III. METHODOLOGY

A. Data Source

The data used consists of secondary data in the form of daily domestic gold prices (GAU/IDR) and seven macroeconomic variables for the period January 1, 2021-April 30, 2026, obtained from investing.com, Bank Indonesia

(BI), Federal Reserve Economic Data (FRED), the U.S. Energy Information Administration (EIA), the Central Statistics Agency (BPS), and Macrotrends. All data series, which had varying original frequencies, were combined into a daily index framework using forward fill, resulting in 1,946 daily observations with no missing values.

B. Research Data

This study uses one target variable and seven exogenous variables, as presented in Table 1.

Table 1 Research Data

Variable	Description
GAU/IDR	Domestic gold price in Indonesia (IDR/gram) - target variable
GAU/USD	Global gold price (USD/gram)
USD/IDR	Exchange rate of the U.S. dollar against the rupiah
DXY	U.S. Dollar Index
Federal Funds Rate	U.S. central bank policy interest rate
BI Rate	Bank Indonesia's policy interest rate
Indonesian Inflation	Indonesia's monthly inflation rate
Crude Oil	Global crude oil prices (USD/barrel)

C. Data Distribution

To prevent data leakage, the dataset was divided chronologically using an out-of-sample (OOS) evaluation scheme. The training data covers January 1, 2021–December 31, 2025 (1,826 days; 93.83%), while the test period covers January 1-April 30, 2026 (120 days; 6.17%). A historical input window consisting of the last 120 observations from the training data (September–December 2025) was used as a lookback buffer to form the first input sequence at the beginning of the test period. For model evaluation, the first 30 days of the test period were used as the initial input window, and performance was evaluated on the subsequent 90 days, from January 31 to April 30, 2026. This 90-day period was used consistently across all three models to ensure a fair comparison.

D. Analysis Stages

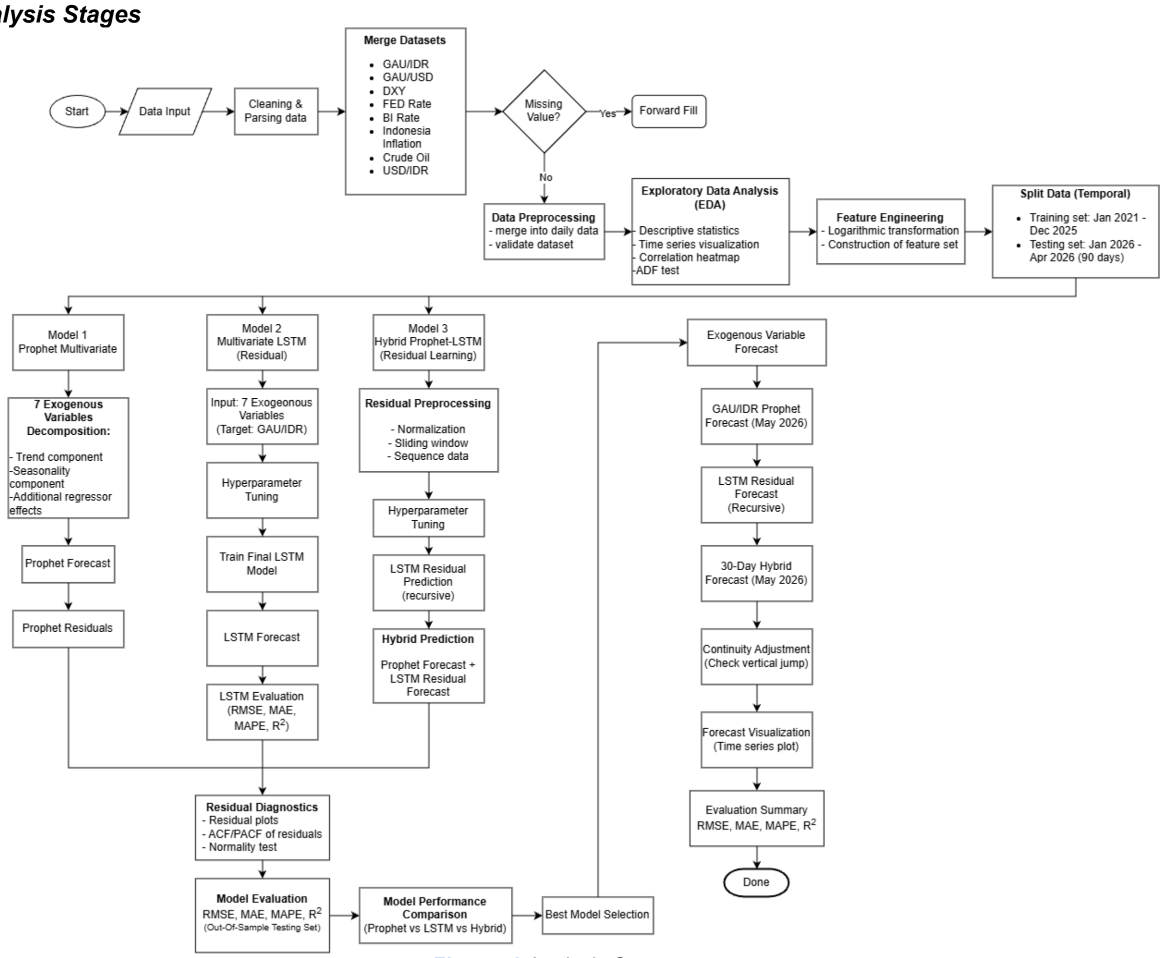

Figure 1 Analysis Stages

The data analysis procedures and stages in this study are carried out systematically through the following main steps, corresponding to the workflow in Figure 1:

1. Data Ingestion and Preprocessing: Merging multi-source datasets (GAU/IDR, GAU/USD, DXY, Federal Funds Rate, BI Rate, Indonesian inflation, Crude Oil, and USD/IDR), handling missing values via forward fill, and validating data completeness.
2. Exploratory Data Analysis (EDA) and Stationarity Testing: Conducting descriptive statistical analysis, time series visualization, correlation analysis, and Augmented Dickey-Fuller (ADF) tests to assess unit roots.
3. Feature Engineering: Applying natural logarithmic transformations to large-scale monetary variables and constructing the feature matrix for modeling.
4. Temporal Data Splitting: Partitioning the dataset chronologically into a training set (January 2021 to December 2025) and an out-of-sample test period (January 2026 to April 2026, comprising 120 days), with model evaluation conducted on the final 90 days after the initial 30-day input window.
5. Model 1 Execution (Multivariate Prophet): Decomposing the series into trend, seasonality, and seven additional exogenous regressors to generate baseline forecasts and extract residual series.
6. Model 2 Execution (Multivariate LSTM Benchmark): Training a standalone multivariate LSTM using the 7 exogenous features and GAU/IDR targets with hyperparameter tuning as a comparative baseline.
7. Model 3 Execution (Hybrid Prophet-LSTM): Preprocessing Prophet residuals (normalization, sliding-window sequence creation), optimizing network hyperparameters, and recursively generating residual predictions to construct the final hybrid forecast.
8. Model Evaluation and Residual Diagnostics: Comparing the out-of-sample performance across all three models using RMSE, MAE, MAPE, and R^2, supplemented by residual diagnostics (normality tests, ACF/PACF plots, and stationarity checks).
9. Operational Forecasting and Continuity Adjustment: Generating a 30-day forward operational forecast (May 1 to May 30, 2026) using exogenously projected features, recursive residual correction, and continuity adjustments to ensure vertical level alignment.

It should be emphasized that the evaluation in step eight is conditional (oracle): the actual values of the seven exogenous variables for the test period are available and used directly as inputs to Prophet, unlike the operational forecasting in step nine, which requires forecasting the exogenous variables first.

E. Multivariate Prophet Model Specification

The Prophet model is specified with trend and seasonality components, as well as additional exogenous regressors:

$$y(t) = g(t) + s(t) + \sum_{i=1}^{K} \beta_i X_i(t) + \varepsilon(t) \quad (20)$$

where $K = 7$ is the number of exogenous macroeconomic indicators, $X_i(t)$ represents the i-th regressor, β_i is its corresponding coefficient, and $\varepsilon(t)$ is the error term. The model's hyperparameters are presented in Table 2.

Table 2 Hyperparameters of the multivariate Prophet model

Hyperparameter	Value
changepoint_**prior**_scale	0.10
seasonality_prior_scale	5.0
yearly_seasonality	True
weekly_seasonality	True
daily_seasonality	False
interval_width	0.95
number of additional regressors	7

F. LSTM Model Specifications

A multivariate LSTM was built as the control model (baseline), using the GAU/IDR log target history and seven exogenous variables that had been preprocessed and differentiated, then normalized using MinMaxScaler(0,1), a 14-day lookback, an LSTM(32) architecture → Dropout(0.3) → Dense(16, ReLU) → Dense(1), the Adam optimizer (learning rate 0.0005), the MSE loss function, and EarlyStopping and ReduceLROnPlateau as callbacks. The residual LSTM in the hybrid architecture was trained on Prophet residuals (log scale) normalized with MinMaxScaler(-1,1), a 30-day lookback, an LSTM(16) → Dropout(0.4) → Dense(1) architecture, L2 regularization = 0.01, the Adam optimizer (learning rate 0.001), and EarlyStopping (patience 10). The hybrid prediction for the evaluation period is formed as:

$$\hat{y}_t^H = \hat{y}_t^P + \lambda \cdot \hat{e}_{LSTM,t}, \text{ with } \lambda = 0.5 \quad (21)$$

The value of λ = 0.5 was set as a fixed parameter based on the researchers' judgment (rather than through hyperparameter optimization) so that the LSTM correction would act as a smoothing mechanism rather than the primary source of the prediction. For the operational forecast for May 2026, a boundary-level adjustment was used: the raw forecast for the first day was calculated; the offset δ was calculated as the logarithmic difference between the last actual price (anchor) and that raw forecast; and this offset was then added as a constant to all 30 projection points so that the transition would be smooth without altering the shape of the model's output trajectory.

IV. RESULTS AND DISCUSSIONS

A. Descriptive Analysis and Data Exploration

The research dataset comprises 1,946 daily observations covering the period from January 1, 2021, to April 30, 2026, with no missing values or duplicate entries. Descriptive statistics for the research variables are presented in Table 3.

Table 3 Descriptive statistics for the research variables

Variable	Mean	Std	Min	Max	Skewness
GAU/IDR (IDR)	1,228,469.56	524,725.76	777,214.00	2,967,408.00	1.4784
GAU/USD	78.21	28.76	52.14	173.46	1.5186
DXY	100.83	5.30	89.44	114.11	-0.2918
Federal Funds Rate	3.26	2.05	0.06	5.33	-0.6495
BI Rate	5.02	1.08	3.50	6.25	-0.5099
Indonesian Inflation	2.79	1.35	-0.09	5.95	0.6769
Crude Oil	80.23	13.70	51.09	127.98	0.8416
USD/IDR	15,586.19	848.41	13,880.00	17,310.00	-0.1367

The GAU/IDR gold price showed a wide range, from IDR 777,214 per gram in early 2021 to nearly IDR 2,967,408 per gram toward the end of the observation period, reflecting a combination of rising global gold prices and the depreciation of the rupiah. Pearson correlation analysis shows that GAU/USD ($r = +0.998$) and USD/IDR ($r = +0.805$) have the strongest positive linear relationships with GAU/IDR, consistent with Equation (1); The BI Rate and the Federal Funds Rate are weakly positively correlated ($r \approx +0.35$); whereas Crude Oil, inflation, and the DXY are very weakly or negatively correlated.

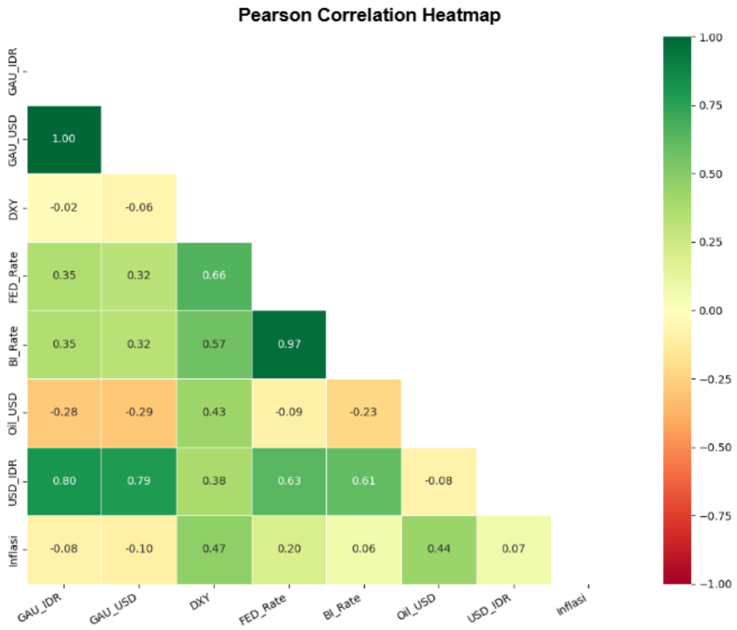

Figure 2 Pearson Correlation Heatmap

The ADF stationarity test at the level indicates that all variables, including the log of GAU/IDR, are nonstationary ($p > 0.05$); after first-order differentiation, all variables become stationary after first-order differentiation ($p < 0.05$), as presented in Table 4. These results confirm the nonstationary nature of the gold price series, consistent with the general characteristics of financial data. Prophet handles trend nonstationarity internally through piecewise-linear trend components, so the model is trained directly on the log GAU/IDR without explicit differentiation; the ADF test above is used solely as part of data exploration.

Table 4 Summary of the Augmented Dickey-Fuller (ADF) Test Results

Variable	Level (p-value)	First-Order Differentiation (p-value)	Conclusion
log GAU/IDR	0.998258	0.000000	Stationary after differentiation
log GAU/USD	0.996035	0.000000	Stationary after differentiation
DXY	0.216183	0.000000	Stationary after differentiation
Federal Funds Rate	0.330826	0.000000	Stationary after differentiation
BI Rate	0.685598	0.000000	Stationary after differentiation
log Crude Oil	0.161253	0.000000	Stationary after differentiation
log USD/IDR	0.797597	0.000000	Stationary after differentiation
Indonesian Inflation	0.475705	0.000000	Stationary after differentiation

B. Interpretation of Macroeconomic Variables

Based on Figure 2 and the price transmission mechanism in Equation (1), the seven exogenous variables contribute unevenly to the formation of the GAU/IDR gold price. GAU/USD is the most dominant variable because it is definitively the main component determining the domestic gold price, as every movement in the global gold price is passed on almost proportionally to the domestic price before being converted using the exchange rate. USD/IDR ranks second as the variable that determines the extent to which global gold prices are translated into rupiah; a weakening rupiah will consistently increase the GAU/IDR price even if global gold prices remain relatively stable.

The Federal Funds Rate and BI Rate show a weak positive correlation ($r \approx +0.35$) with GAU/IDR, slightly contradicting classical opportunity cost theory, which predicts a negative relationship: as interest rates rise, gold (as a non-yielding asset) becomes less attractive. This apparent contradiction can be explained by a confounding effect: the rise in the Federal Funds Rate during the 2022-2023 period coincided with a period of high inflation and geopolitical uncertainty, which actually drove demand for gold as a safe haven; thus, the effect of interest rates was overshadowed by the more dominant safe-haven effect during that period. Indonesian inflation shows a very weak correlation (nearing zero) with GAU/IDR, indicating that gold's role as a hedge against domestic inflation does not appear strong in the 2021-2026 data, likely because Indonesian inflation remained relatively under control throughout that period and thus did not serve as a primary driver of gold demand. Crude oil and the DXY also exhibit a weak correlation, indicating that the transmission channels of oil prices and the dollar index on GAU/IDR are indirect (via inflation or the exchange rate), resulting in a weak direct linear correlation.

Overall, the primary determinants of GAU/IDR movements are the GAU/USD and USD/IDR pairs, while other macroeconomic variables act as accompanying factors whose influence is more conditional and not always linear. This provides the rationale for incorporating the LSTM component to capture the nonlinear interactions among variables that are not captured by Prophet's linear regression.

C. Performance Comparison of Three Models

The three models were evaluated over the aligned test period from January 31 to April 30, 2026 (90 days, aligned test). This evaluation is conditional (oracle): the actual values of the seven exogenous variables during the test period are known and used directly as inputs for Prophet; therefore, the metrics in this section describe the models' ability to predict the GAU/IDR when the realized values of the exogenous variables are available, rather than their pure operational forecasting accuracy without that information. The results of the evaluation comparison are presented in Table 5.

Table 5 Comparison of out-of-sample model evaluations (N = 90 days)

Model	RMSE (IDR)	MAE (IDR)	MAPE (%)	R^2
Multivariate Prophet	33.860	27.932	1.0655%	0.919995
Multivariate LSTM	49.769	36.210	1.3761%	0.827147
Hybrid Prophet-LSTM	28.478	22.270	0.8476%	0.943405

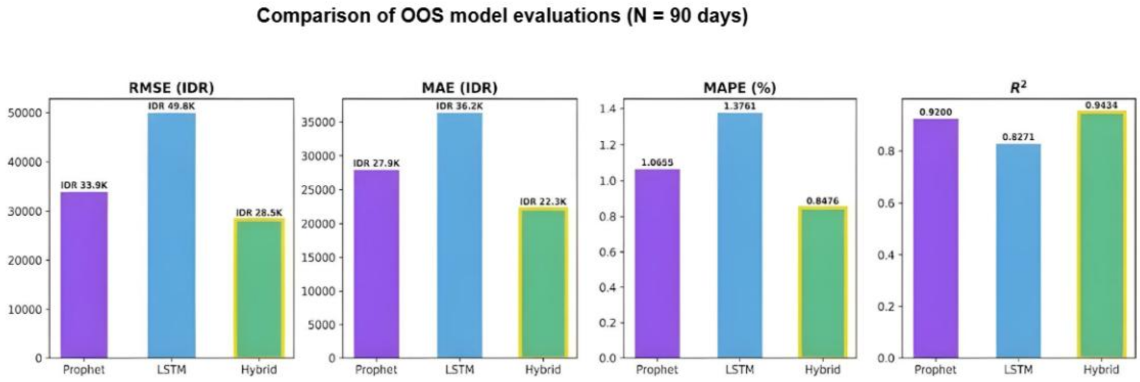

Figure 3 Comparison of OOS model evaluations (N = 90 days)

The Hybrid Prophet-LSTM model produces lower errors (RMSE, MAE, MAPE) and a higher R^2 compared to both the multivariate Prophet and the multivariate LSTM. Residual learning via LSTM has been shown to reduce prediction errors, as evidenced by a decrease in RMSE from IDR 33,860 (Prophet alone) to IDR 28,478 (hybrid), representing a reduction of approximately 15.90%, and a decrease in MAPE from 1.0655% to 0.8476%. This improvement occurs because Prophet, as an additive decomposition model, focuses on modeling trends, seasonality, and the effects of exogenous variables, leaving some unexplained nonlinear patterns in the residuals. The LSTM component then learns these residual patterns and generates additional corrections to Prophet's predictions, thereby reducing the total error without altering the previously modeled trend and seasonal structures.

Although the Multivariate LSTM uses the same target and the same seven exogenous variables as Prophet, it still produces the highest error (RMSE IDR 49,769; MAPE 1.3761%) and the lowest R^2 (0.827147) among the three models. These results indicate that the mere availability of exogenous variables does not necessarily lead to better performance than Prophet; the Multivariate LSTM must learn trends, seasonality, and multivariate relationships simultaneously from the differentiated time series without an explicit decomposition, as Prophet does, making it more susceptible to estimation errors at the test horizon. Conversely, the best performance was achieved when the LSTM was narrowly focused on learning the Prophet residuals in the hybrid model, rather than on directly predicting the target, confirming that the LSTM's role as a residual learner yields better performance than when used as a standalone predictive model in this case. Since the implementation notebook does not include formal statistical tests of accuracy differences between models (e.g., the Diebold-Mariano test), these differences are presented as reductions in error on the available test sample, rather than as statistically significant differences.

D. Residual Analysis of the Hybrid Model

The final residual statistics (in IDR/gram) for the test period are presented in Table 6.

Table 6 Final residual statistics for the test period

Model	Mean Residual (IDR)	Std Residual (IDR)	Jarque-Bera Test (p-value)	Normality
Multivariate Prophet	-15.027	30.342	0.1839	Normal
Multivariate LSTM	-4.085	49.601	0.0910	Normal
Hybrid Prophet-LSTM	-9.352	29.899	0.3438	Normal

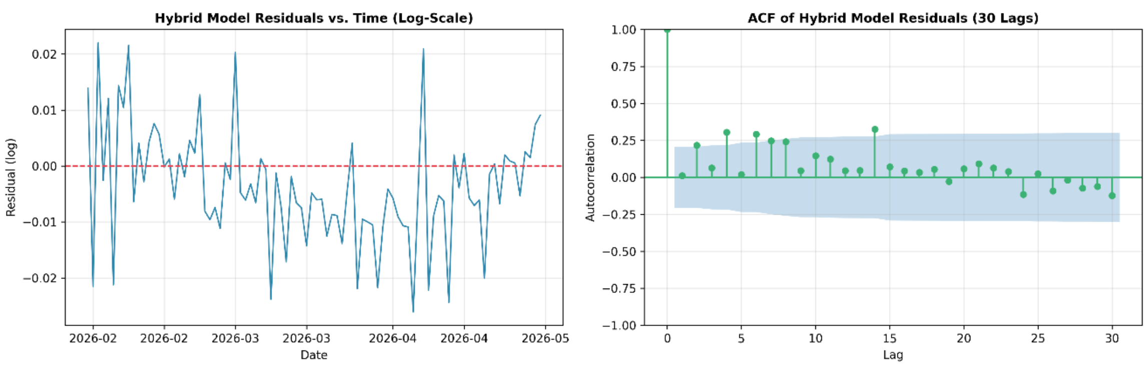

Figure 4 Plot of the ACF of the Hybrid Prophet-LSTM residuals (30 lag)

On a log scale, the final residuals of the hybrid model have a mean of −0.003615 and a standard deviation of 0.010217. The ADF test on the centered residuals yielded a statistic of −1.2981 with a p-value of 0.6300; thus, the null hypothesis of the presence of a unit root cannot be rejected, indicating that the residuals of the hybrid model have not been proven to be stationary. This finding is important because normality is not the same as the white-noise property, since white noise also requires stationarity and the absence of significant autocorrelation. Thus, the claim that the hybrid model's residuals are white noise cannot be upheld, indicating that there is still structure in the residuals that has not been fully captured by the LSTM component and that this provides a potential direction for future model development in future research.

E. 30-Day Operational Forecast (May 2026)

The 30-day operational forecast (May 1-30, 2026) uses a hybrid model following a two-stage procedure because the actual values of the seven exogenous variables for May 2026 are not yet available. In the first stage, the seven exogenous variables are forecasted using separate univariate Prophet models, yielding relatively stable projections around their historical closing levels without drastic changes. In the second stage, the main Prophet model produced a pure GAU/IDR log projection (before residual correction) ranging from IDR 2,722,347 to IDR 2,756,752 per gram, and future residuals were estimated recursively by the residual LSTM.

The raw hybrid forecast for the first day (May 1, 2026) is IDR 2,746,442/gram, while the anchor (actual price on April 30, 2026) is IDR 2,573,810/gram, resulting in a rebase offset $\delta = -0.064919$ (log scale) that is uniformly applied to all 30 projection points based on Equation (15). After the adjustment, the forecast for the first day is aligned with the last observed price of IDR 2,573,810/gram. Therefore, the boundary adjustment ensures level continuity by construction. This alignment should not be interpreted as evidence of forecasting accuracy, as the first forecast value is explicitly anchored to the last observed value. Table 7 presents the first seven days of the projection as an illustration.

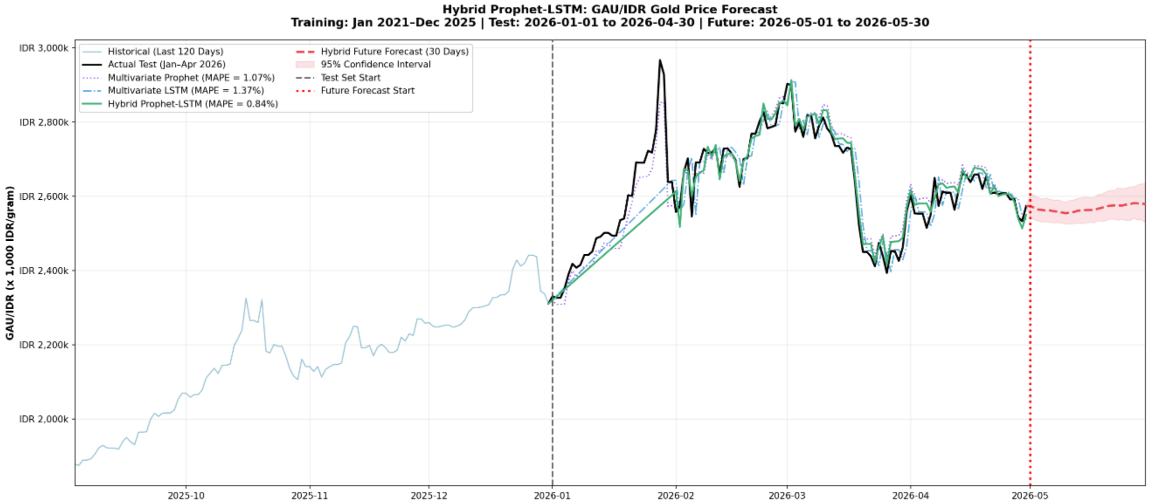

Figure 5 Visualization of Forecasting Continuity Lines

Table 7 GAU/IDR hybrid forecast for the first seven days of May 2026

Date	Forecast (IDR)	Lower Bound (IDR)	Upper Bound (IDR)	Prophet Only (IDR)
01-05-2026	2,573,810	2,542,294	2,606,633	2,734,821
02-05-2026	2,569,529	2,537,381	2,602,539	2,730,742
03-05-2026	2,564,723	2,534,197	2,597,742	2,726,520
04-05-2026	2,563,704	2,531,655	2,593,082	2,726,390
05-05-2026	2,562,623	2,532,175	2,596,270	2,725,495
06-05-2026	2,560,453	2,531,405	2,592,933	2,727,035
07-05-2026	2,559,172	2,529,121	2,590,609	2,725,568

Overall, the projection range for May 2026 is between a minimum of IDR 2,555,128/gram and a maximum of IDR 2,580,735/gram, with a final value of IDR 2,577,984/gram on May 30, 2026, compared to the anchor value of IDR 2,573,810/gram on April 30, 2026. The relatively narrow projection range (approximately −0.73% to +0.27% relative to the anchor value) indicates a consolidation-oriented forecast; the model does not suggest a strong upward trend or a sharp decline, but rather predicts relatively stable price movements around the level seen at the end of April 2026.

F. Discussion and Limitations

Although the proposed hybrid model shows potential for good performance, there are several crucial aspects related to the methodology and data characteristics that need to be considered. Based on the results of the analysis and testing, the following are some points for discussion and limitations of this study:

1. The model performance evaluation in Table 5 uses the actual values of exogenous variables during the testing period (conditional/oracle), and thus differs from the operational forecasting conditions for May 2026, in which exogenous variables must first be estimated.
2. The damping factor $\lambda = 0.5$ was set as a fixed value based on the researchers' judgment, rather than through hyperparameter optimization or separate data validation.
3. The comparison of performance across models has not been supported by formal statistical tests (such as the Diebold-Mariano test), so the observed differences cannot yet be interpreted as statistically significant.
4. The final residuals of the hybrid model do not meet the stationarity assumption based on the ADF test, indicating that there are still patterns that could potentially be modeled further.
5. The evaluation was conducted over a single testing period without a rolling-origin evaluation at multiple cutoff points; therefore, the results cannot yet be generalized to various market conditions.

Given these limitations, the results of this study are not intended to serve as the sole basis for investment decisions, but rather as preliminary empirical evidence that the residual-learning-based Hybrid Prophet-LSTM approach has the potential to produce lower prediction errors than single models on the data and test periods used.

V. CONCLUSIONS AND SUGGESTIONS

A. *Conclusions*

1. The historical characteristics of the GAU/IDR gold price for the period January 2021-April 2026 show a long-term upward trend from IDR 777,214 to IDR 2,967,408 per gram, with a right-skewed distribution (skewness 1.4784). GAU/USD ($r = +0.998$) and USD/IDR ($r = +0.805$) exhibit the strongest positive linear relationships with GAU/IDR, consistent with the price transmission equation $\mathrm{P(GAU/IDR)} = \mathrm{P(GAU/USD)} \times \mathrm{E(USD/IDR)}$. All variables are non-stationary at the level and become stationary after first-order differentiation, confirming the characteristics of financial time series data as being trending and volatile.
2. The Hybrid Prophet-LSTM model, which combines a multivariate Prophet component and a residual LSTM, proved effective in forecasting the GAU/IDR gold price, yielding an RMSE of IDR 28,478, an MAE of IDR 22,270, a MAPE of 0.8476%, and an R^2 of 0.943405 during the test period, demonstrating better performance than the multivariate Prophet model (RMSE IDR 33,860; MAPE 1.0655%; R^2 0.919995) and the multivariate LSTM as the comparison model (RMSE IDR 49,769; MAPE 1.3761%; R^2 0.827147). Residual learning by the LSTM reduced the RMSE by approximately 15.90% compared to Prophet alone.
3. The application of boundary-level adjustment aligned the first forecast value with the last observed price through a rebase offset applied uniformly across all 30 projection points, ensuring a continuous transition between historical data and forecast results. This adjustment was applied as a boundary condition and should not be interpreted as an indicator of forecasting accuracy. The May 2026 forecast results show a price consolidation pattern within a range of -0.73% to $+0.27\%$ relative to the anchor value, with no indication of a sharp upward or downward trend.

B. *Suggestions*

Future research is recommended to apply rolling-origin cross-validation with multiple cutoff points and to supplement the comparison between models with formal statistical tests such as the Diebold-Mariano test. Given that the residuals of the hybrid model have not been proven to be stationary ($p = 0.6300$), future research could model the remaining residual structure, for example, using ARIMA or GARCH. It is recommended that the damping factor λ be determined through hyperparameter tuning on separate validation data, and prediction intervals can be developed using simulation-based approaches such as residual bootstrapping or ensemble models. Investors or researchers adopting this model should note that the evaluation is conditional (oracle-based) and that the May 2026 forecast results are best interpreted as a projection of short-term price stability, rather than an investment momentum signal.

ACKNOWLEDGEMENT

The author expresses sincere gratitude to the academic supervisors for their invaluable guidance and constructive insights throughout this research, and to the Head of the Study Program for the continuous academic support during the author's studies. Special thanks are also extended to the "Lulus sih Lulus" circle for their unwavering encouragement, camaraderie, and companionship through every joy and challenge from the beginning to the finish line, with the hope that this friendship endures for years to come. Finally, appreciation goes to everyone who contributed to the completion of this paper.

REFERENCES

[1] A. Halim, D. P. Ompusunggu, and T. R. Simbolon, "Peran Emas sebagai Safe Haven di Tengah Melemahnya Rupiah dan Turunnya Daya Beli Masyarakat," vol. 9, no. 1, pp. 1–12, 2026.

[2] D. Magfiroh and K. Komarudin, "Melambungnya Harga Emas di Tengah Ketegangan Geopolitik dan Tekanan Inflasi Global," *Lentera Multidiscip. Stud.*, vol. 3, no. 3, pp. 270–276, 2025, doi: 10.57096/lentera.v3i3.165.

[3] M. Hafiz, Y. Anisa, A. Gani, and M. Malik, "Pemodelan Keterkaitan Harga Emas Dunia dan Saham Antam Menggunakan Vector Auto Regression," *JISTech (Journal Islam. Sci. Technol. JISTech*, vol. 10, no. 1, pp. 56–63, 2025, [Online]. Available: http://jurnal.uinsu.ac.id/index.php/jistech

[4] A. F. Gaib, A. Rasyid, and J. D. Giu, "Pengembangan Model Time Series untuk Peramalan Harga Beras Medium Eceran di Provinsi Gorontalo," vol. 9, no. 2, 2026, doi: 10.31004/jutin.v9i2.48897.

[5] S. Albahli, "LSTM vs. Prophet: Achieving Superior Accuracy in Dynamic Electricity Demand Forecasting," *Energies*, vol. 18, no. 2, 2025, doi: 10.3390/en18020278.

[6] Bivort and Corentin, "What characterizes a safe haven asset and how does it perform across market cycles," 2023.

[7] B. L. Marvillia, "Pemodelan Dan Peramalan Penutupan Harga Saham Pt. Telkom Dengan Metode Arch - Garch," *J. Unesa*, pp. 1–7, 2013

[8] Z. Liu, Z. Zhu, J. Gao, and C. Xu, "Forecast Methods for Time Series Data: A Survey," *IEEE Access*, vol. 9, pp. 91896–91912, 2021, doi: 10.1109/ACCESS.2021.3091162.

[9] C. K. Ekowati, S. S. Garak, A. Madu, I. H. E. Rimo, and F. A. Halim, "Time series data analysis to forecast the percentage of on-time graduates at Nusa Cendana university," *Edelweiss Appl. Sci. Technol.*, vol. 8, no. 6, pp. 260–278, 2024, doi: 10.55214/25768484.v8i6.2051.

[10] M. M. Ataey and A. K. Azizi, "Forecasting Gold Price Volatility Using Econometric and Machine-Learning Models," *J. Soc. Sci. Humanit.*, vol. 3, no. 1, pp. 173–188, 2026, doi: 10.62810/jssh.v3i1.231.

[11] G. Saefani and Teti Sofia Yanti, "Model Prophet dengan Random Search untuk Peramalan Jumlah Kunjungan Wisatawan Mancanegara," *J. Ris. Stat.*, pp. 111–120, 2025, doi: 10.29313/jrs.v5i2.8278.

[12] S. J. Taylor and B. Letham, "Forecasting at Scale," *Am. Stat.*, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.

[13] M. Milli, "Designing a residual-enhanced hybrid Prophet–LSTM framework for urban air pollution forecasting in Beijing," *Sci. Rep.*, vol. 15, no. 1, pp. 1–24, 2025, doi: 10.1038/s41598-025-27510-y.

[14] J. Yang, X. Zhang, W. Chen, and F. Rong, "Prophet – CEEMDAN – ARBiLSTM-Based Model for Short-Term Load Forecasting," pp. 1–16, 2024.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," vol. 9, no. 8, pp. 1–32, 1997.