Gold Futures Price Analysis

Historical Insights and Future Forecasts

Soumyajoy Kundu

M.Sc Data Science
Chennai Mathematical Institute

Time Series Analysis
November 13, 2024

What are Gold Futures?

Lock in Your Gold Price Today,
Buy It in the Future !!!

Gold Price vs Gold Futures

	Gold Price (Spot Price)	Gold Futures
When?	Immediate	In the future (specific date)
Purpose?	Buy/sell gold now	Speculate or hedge against future prices

Problem Statement

Problem Statement

This study aims to analyze the fluctuations in Gold Futures prices. By focusing on accurate univariate forecasting, the study seeks to capture the price trends and volatility of gold futures, thereby aiding traders, policymakers, and investors in making informed decisions amidst market uncertainties.

Objectives

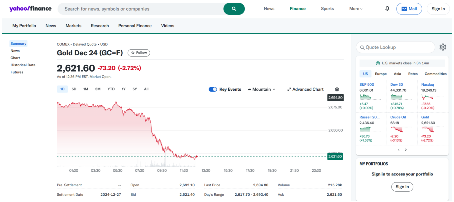
- Analyze Gold Futures data to identify trends and price volatility.
- Compare the forecasting performance of statistical and machine learning models and also evaluating them.
- Provide insights for market participants, including traders, investors, and policymakers, on improving Gold Futures price forecasting and decision-making strategies.

Data

- Aug. 30, 2000 – Dec. 29, 2023
- Size of Data: 5854
- Frequency: **Daily**

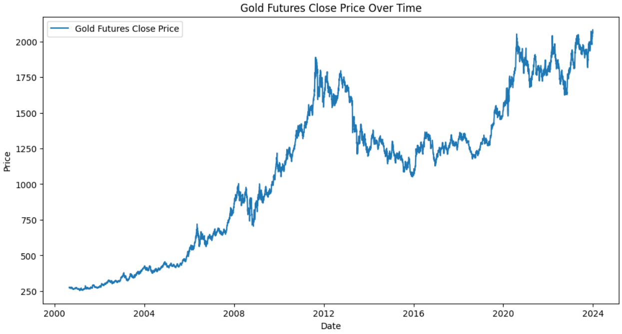
c

Price	Adj Close	Close	High	Low	Open	Volume
Ticker	GC=F	GC=F	GC=F	GC=F	GC=F	GC=F
Date						
2023-12-22 00:00:00+00:00	2057.100098	2057.100098	2068.699951	2052.199951	2055.699951	202
2023-12-26 00:00:00+00:00	2058.199951	2058.199951	2060.800049	2054.199951	2060.000000	64
2023-12-27 00:00:00+00:00	2081.899902	2081.899902	2064.800049	2064.800049	2067.300049	586
2023-12-28 00:00:00+00:00	2073.899902	2073.899902	2087.300049	2066.500000	2081.600098	338
2023-12-29 00:00:00+00:00	2062.399902	2062.399902	2068.899902	2062.100098	2068.000000	47

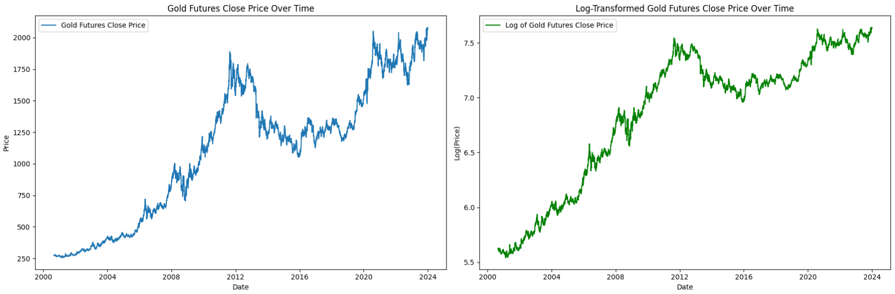


Why close_price?

- Reflects Daily Market Sentiment
- Minimizes Intraday Noise
- Foundation for Technical Analysis

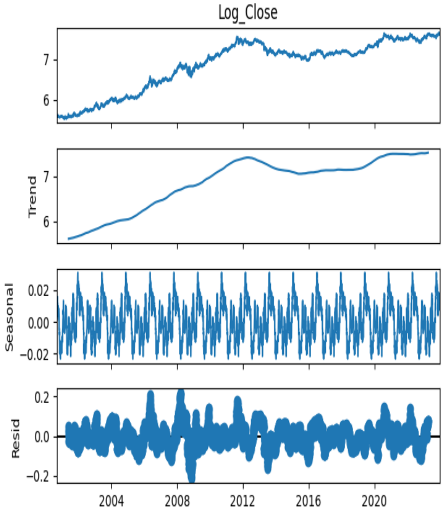


Log Transformation



- **Stabilize Variance**
  - Making the series more stationary, which is a key assumption for many time series models.
- **Handle Exponential Growth**
  - A log transformation has linearize this trend, making it easier to model.

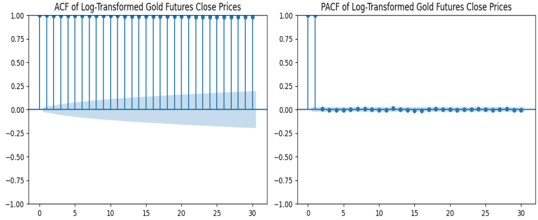
Seasonal Decomposition



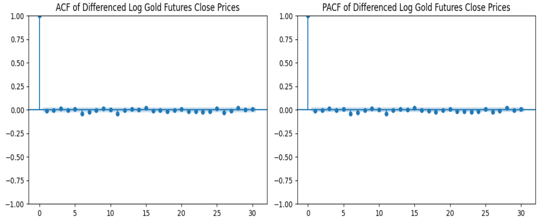
- `Original`
  - overall increasing trend with fluctuations over time
- `Trend`
  - captures the long-term upward movement
  - steady increase in gold prices @ inflation
- `Seasonal`
  - minor regular cycles corresponds to market behaviors
- `Residual`
  - Stationary but Heteroskedastic

ACF & PACF Plots

Before Differencing



After Differencing



`ACF Plot`

- Significant positive lags at all levels, gradually tapering off
- Indicates a strong persistence or trend in the data.
- The slow decay suggests that the series is non-stationary,

`PACF Plot`

- A significant spike at lag 1 and becomes insignificant afterward.

What is the nature of the series?

Random Walk

Tests for Stationary

Stationarity

- Critical Assumption to ensure robustness
- Statistical properties of a series do not change over time.

Variable	**I(0)**	**I(1)**
log `close_price`	$\times$	$\checkmark$

ADF Test

- Checks presence of unit root.

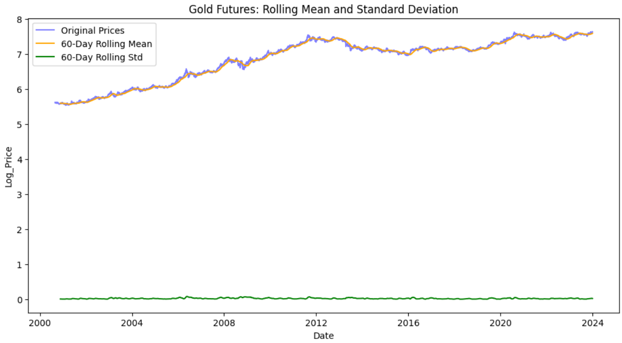
H_0 : Non-Stationary vs. H_1 : Stationary

KPSS Test

- Complementary to the ADF test.
- Checks stationarity around a deterministic trend.

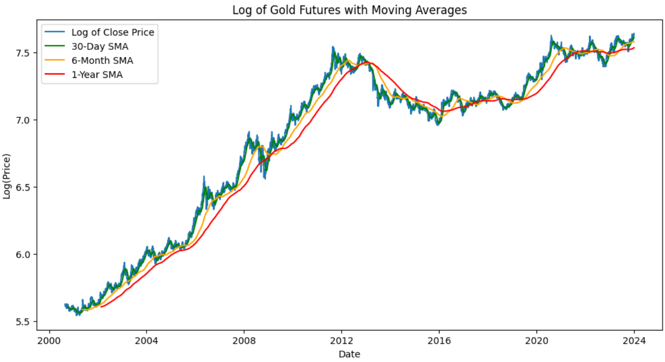
H_0 : Stationary vs. H_1 : Non-Stationary

Trends, Stability, and Volatility



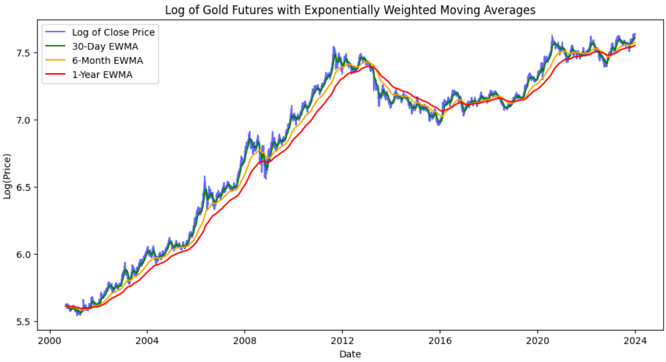
- `Trend`
  - Long-term increase in value.
- `Volatility` – *Low* and *Stable* Rolling S.D
  - suggesting consistent and low volatility in gold prices over time.
- `Stability` – Rolling Mean *closely follows* orginal prices.
  - stable mean reversion with limited short-term fluctuations.

Smoothing – k-period SMA



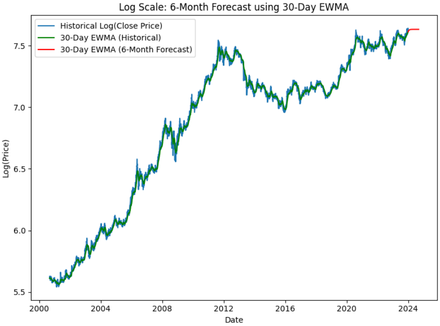
SMA Period	RMSE	MAE
30-Day ✓	0.0319	0.0245
6-Month	0.0775	0.0608
1-Year	0.1202	0.0973

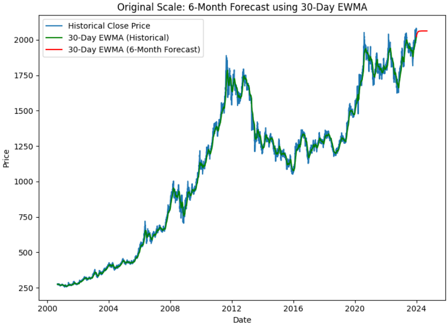
Smoothing – k-period EWMA



EWMA Period	RMSE	MAE
30-Day ✓	0.0270	0.0206
6-Month	0.0683	0.0532
1-Year	0.1083	0.0879

30 Day EWMA – Forecasting





Holt Winters Exponential Smoothing

- Extension to Holt's method that finally allows for the capturing of a seasonal component.
- Triple exponential smoothing.
- **Assumption!**

  The time series has a Level, Trend and Seasonal component.

Forecast Equation

A: Additive T: Trend M: Multiplicative S: Seasonality

ATAS

$$\boxed{F_{t+k} = L_t + (k * T_t) + S_{t+k-M}}$$

ATMS

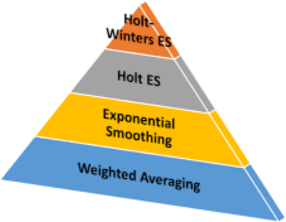
$$\boxed{F_{t+k} = [L_t + (k * T_t)] * S_{t+k-M}}$$

MTAS

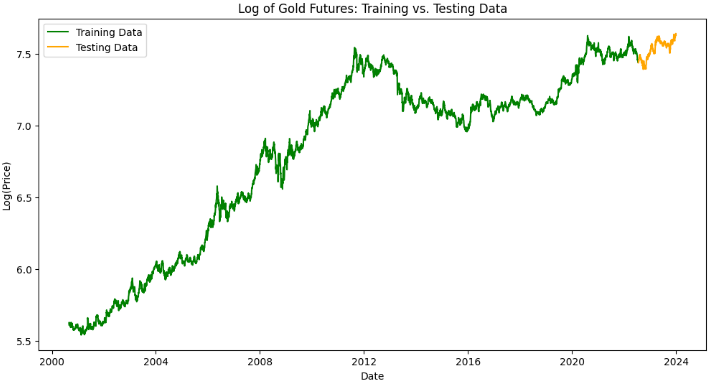
$$\boxed{F_{t+k} = [L_t * (T_t)^k] + S_{t+k-M}}$$

MTMS

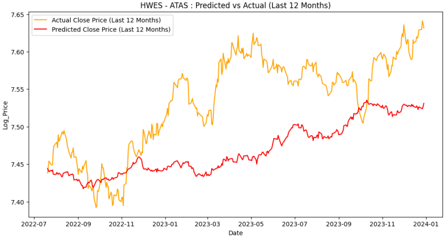
$$\boxed{F_{t+k} = [L_t * (T_t)^k] * S_{t+k-M}}$$

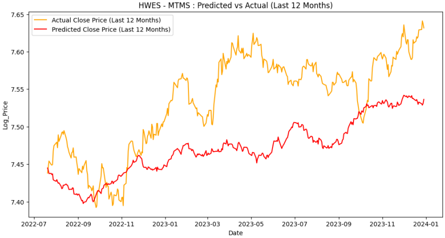


HWES – Splitting of Data



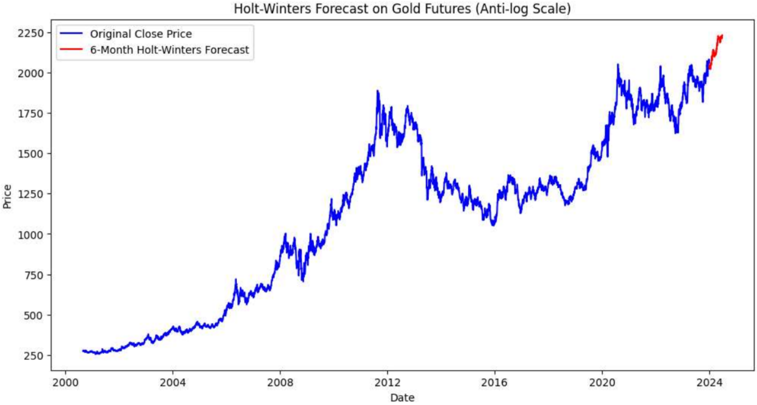
HWES – Prediction & Evaluation





Model	RMSE (Log-Scale)	RMSE (Original Scale)
MTMS ✓	0.0757	140.79
ATAS	0.0802	148.67

HWES – Forecasting on 6 months



ARIMA

Auto-Regressive Integrated Moving Average – ARIMA

- **AR(p)** – look back in time and analyze the previous values
- **I(d)** – differencing steps applied to the data to make it stationary.
- **MA(q)** – the past and current values of residuals

Mathematical Form

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-B)^d Y_t = (1+\theta_1 B+\theta_2 B^2+\ldots+\theta_q B^q)\epsilon_t$$

Assumptions

- Linearity
- Stationarity (after differencing)
- No Seasonality
- $\epsilon_i \sim WN(0, \sigma^2)\ \forall\, i$

ARIMA(0,1,0) – Best Model

- $(p, d, q) \in \{0, 1, 2\} \times \{0, 1, 2\} \times \{0, 1, 2\}$ – Search Space
- Optimal Parameters: $(p, d, q) = (0, 1, 0)$
- Lowest AIC = -33859.189

Mathematical Form

$$(1 - B)Y_t = \epsilon_t$$

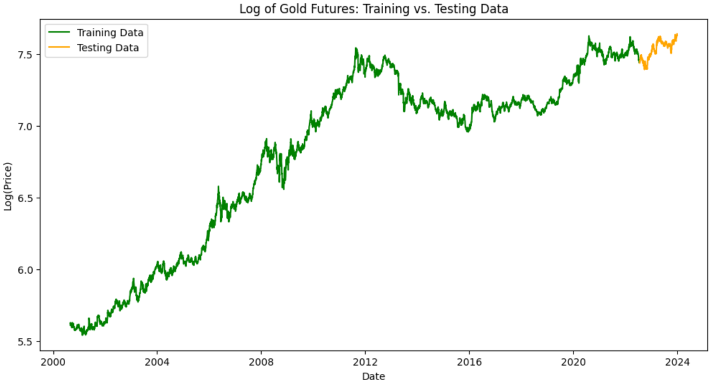
$$Y_t = Y_{t-1} + \epsilon_t$$

where, $\epsilon_i \sim WN(0, \sigma^2) \; \forall \, i$

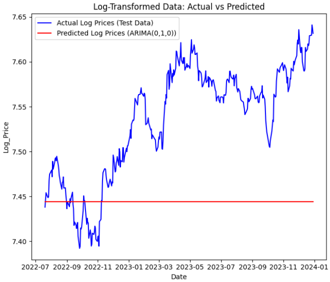
Random Walk

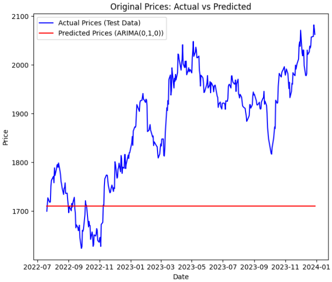
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   GC=F   No. Observations:                 5489
Model:                 ARIMA(0, 1, 0)   Log Likelihood               16930.595
Date:                Mon, 11 Nov 2024   AIC                         -33859.189
Time:                        20:29:49   BIC                         -33852.579
Sample:                             0   HQIC                        -33856.884
                               - 5489
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2         0.0001   1.21e-06    100.940      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.63   Jarque-Bera (JB):              6919.56
Prob(Q):                              0.43   Prob(JB):                         0.00
Heteroskedasticity (H):               0.80   Skew:                            -0.30
Prob(H) (two-sided):                  0.00   Kurtosis:                         8.47
===================================================================================
```

ARIMA(0,1,0) – Splitting of Data



ARIMA(0,1,0) – Prediction & Evaluation

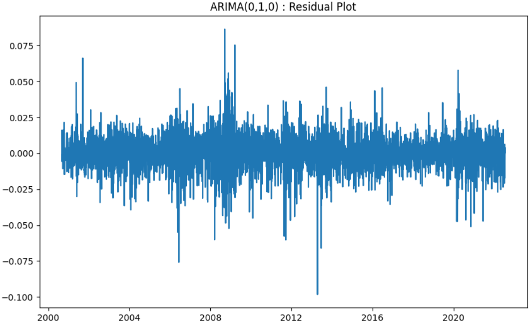




Metric	**Log Data**	**Original Data**
MAE	0.0962	175.2195
RMSE	0.1099	201.5852

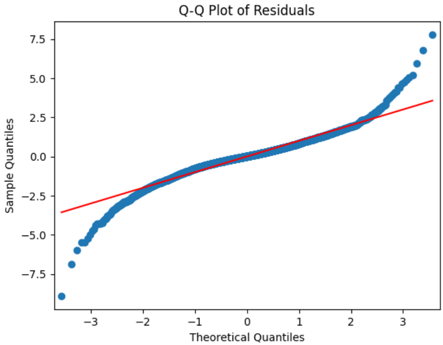
- Very Low AIC – Underfitting
- Assumes each future value is a continuation of the last observed value

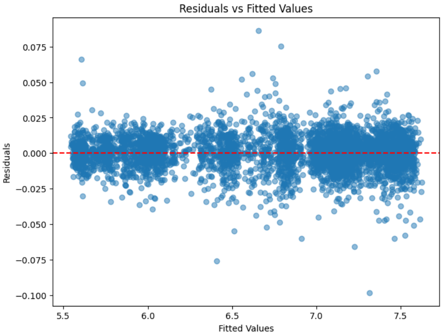
ARIMA(0,1,0) – Residual Diagnostics



Tests	p-value	Decision	Remark
Jarque - Bera Test	0.00	Reject H_0	Normality ×
Ljung - Box Test	0.43	Accept H_0	No Autocorrelation ✓
Berush - Pagan - Godfrey Test	0.18	Accept H_0	Homoscedasticity ✓

ARIMA(0,1,0) – Residual Diagnostics

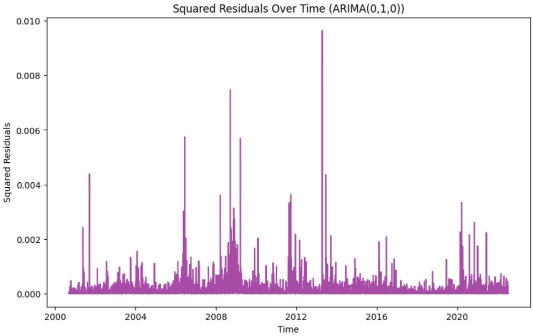




- Deviation in Tails
- Good fit near center
- Fat or Heavy Tails in the distribution

- Centered around zero
- No Apparent Trend
- Spread is Relatively Consistent

ARIMA(0,1,0) – Squared Residuals



- Magnifying variance, which revealed underlying volatility clustering
- Squared residuals exhibits patterns or clusters – heteroscedasticity (time-varying volatility)

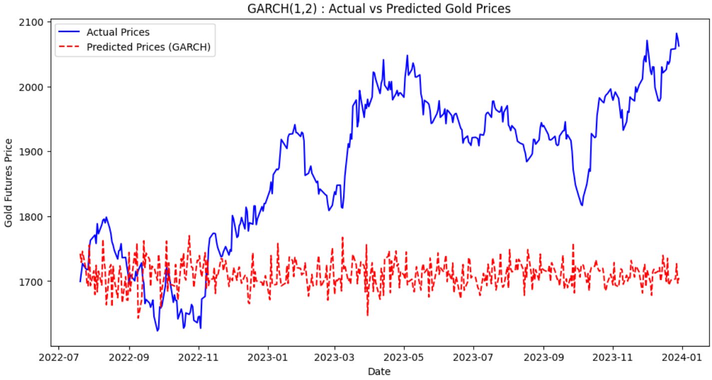
Let's fit a GARCH model to better capture the volatility.

GARCH(1,2) – Predictions & Evaluation

$$y_t = \mu + \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2$$

where $\epsilon_t = \sigma_t z_t$ and $z_t \sim \mathcal{N}(0, 1)$

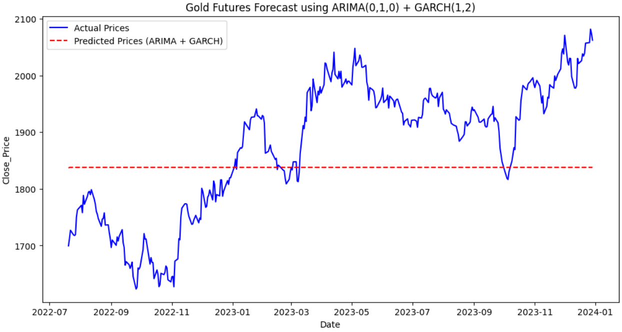


ARIMA(0,1,0) + GARCH(1,2)

$$y_t = y_{t-1} + \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2$$

where $\epsilon_t = \sigma_t z_t$ and $z_t \sim \mathcal{N}(0, 1)$

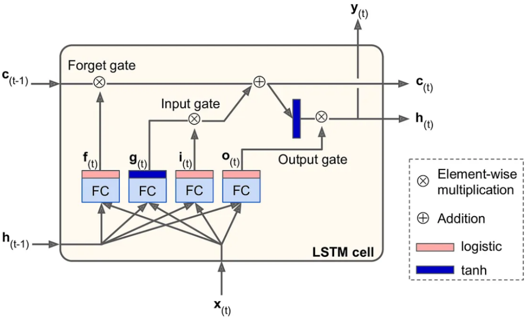


Metric	Value
RMSE	121.92
MAE	108.11

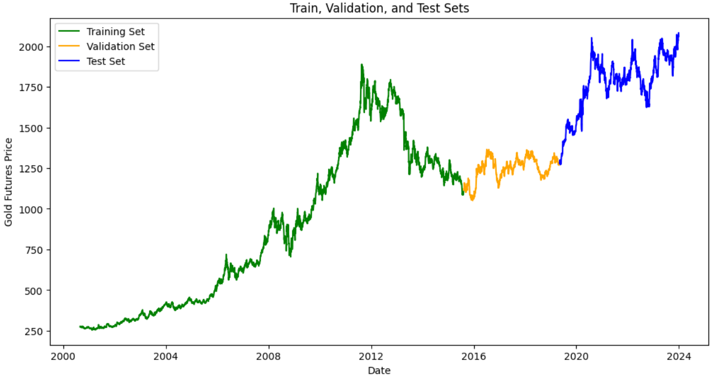
LSTM Network

Long Short Term Memory

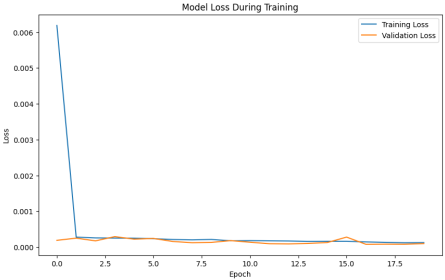
- Variant of RNN.
- Designed to handle sequence data and capture long-term dependencies.
- 2 LSTM(50) layers and 1 Dense(25) Layer

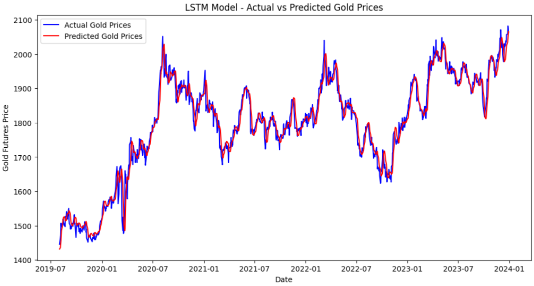


LSTM – Splitting of Data



LSTM – Training & Predictions





- Train Loss : 1.4740e-04
- Valid Loss : 1.8853e-04

Metric	Value
RMSE	26.40
MAE	19.21

LSTM – Forecasting

Summary

Model	RMSE
Log Scale	
SMA_30D	0.0319
EWMA_30D ✓	0.0270
HWES_MTMS	0.0757
ARIMA(0,1,0)	0.1099
ARIMA(0,1,0)+GARCH(1,2)	0.0656
Original Scale	
LSTM_2_(50)_1_(25) ✓	26.40

- All log-scale models have RMSE on original scale above 50.

Thank You

soumyajoy.mds2023@cmi.ac.in