Long-Term Forecasting of COMEX Gold Futures Using LSTM Models

Jingyi Bai
Department of Statistics
University of Michigan
Ann Arbor, U.S.
kohaku@umich.edu
https://github.com/bai-umich/Stats507_Fall2025

***Abstract*—This study investigates whether historical daily price and volume data of COMEX gold futures can be used to forecast future gold prices. We collect data from 2004 to 2025 and preprocess it by removing missing values and zero-volume days, then scale the features and build a sliding-window dataset. We train a Long Short-Term Memory (LSTM) neural network on 30-day input sequences to predict the next day's closing price. The model is evaluated on training, validation, and test datasets. Results show that during relatively stable periods the model can follow actual price trends, but its accuracy decreases when forecasting far into the future or during volatile periods. The findings suggest that while simple historical-data models like LSTM can capture short- to medium-term price patterns, they have limited capacity in predicting long-term fluctuations. We discuss implications for gold price forecasting and propose directions for future work, including integrating additional market indicators or sentiment data to improve robustness.**

I. Introduction

Gold is a globally important financial asset and is often considered a safe-haven investment during periods of economic uncertainty. Its price fluctuates due to economic conditions, geopolitical events, and market sentiment. Understanding and forecasting gold price movements can support investors, traders, and policymakers in making informed decisions and managing risk effectively.

The motivation for this project comes from the observation that, while sudden events such as financial crises or geopolitical conflicts are unpredictable, historical price data often contain patterns that can be leveraged for forecasting. From Figure 1, we can see that gold prices have been continuously rising from 2004 to 2025. According to John J. Murphy in Technical Analysis of the Financial Markets, there are three basic assumptions in technical analysis: 1) market behavior reflects all information, 2) prices move in trends, and 3) history tends to repeat itself. Based on these principles, even though daily market shocks are unpredictable, overall price movements reflect market valuation. This motivates the development of a long-term predictive model using only historical price data.

This project focuses on analyzing COMEX gold futures prices from 2004 to 2025, including daily open, high, low, close prices, and trading volume, with the goal of building a predictive model that can forecast long-term price trends. The main objective is to develop a systematic, data-driven framework using historical numerical data to evaluate the predictive power of past price movements.

Several recent studies have advanced the understanding of gold price prediction. Zou and Hou (2024) explored the relationship between the CBOE Volatility Index (VIX) and COMEX gold futures. They found that the VIX has predictive power for gold prices, confirming gold's role as a safe-haven asset during market turbulence. Moghar and Hamiche (2020) demonstrated that Long Short-Term Memory (LSTM) neural networks can effectively forecast stock prices by capturing trends from historical data, and that prediction accuracy improves with increased training. Liu, Zhang, and Peng (2024) combined textual analysis from online news with BiLSTM models to forecast Chinese gold futures. Their study shows that integrating textual indicators with historical prices significantly improves short-term prediction accuracy, highlighting the value of advanced machine learning models in financial forecasting.

Building on these insights, this project applies LSTM-based deep learning models to COMEX gold futures. Unlike studies incorporating external data such as VIX or news sentiment, this work focuses on purely numerical historical price and volume data. By doing so, it evaluates whether long-term patterns in price history alone can provide useful forecasts, and compares model performance across training, validation, and test datasets. This approach contributes to understanding the predictive potential of historical price dynamics and the practical application of deep learning for long-term gold price prediction.

II. Method

A. Problem Formulation

The objective of this study is to predict the next-day closing price of COMEX gold futures using historical daily data. Let the dataset consist of $N = 8,230$ observations, where each observation is denoted as $(\mathbf{x}_t, y_t)$. Here, $\mathbf{x}_t = [\text{open}_t, \text{high}_t, \text{low}_t, \text{close}_t, \text{volume}_t]$ represents the input features for day t, and $y_t = \text{close}_{t+1}$ is the target output, the closing price of the following day. A sliding window of size $w = 30$ days is applied to construct sequences for the model:

$$\mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_{i+w-1}], \tag{1}$$

$$y_i = y_{i+w-1}, \quad i = 1, 2, \ldots, N - w. \tag{2}$$

This transformation generates temporally structured input suitable for sequential modeling.

B. Data Preprocessing

The raw dataset consists of daily COMEX gold futures data, including open, high, low, close, and trading volume. Standard preprocessing steps were applied prior to model training. All rows containing missing values were removed, and observations with zero trading volume were excluded to avoid inactive or erroneous trading days. After cleaning, the dataset was re-indexed to ensure a continuous time series. All numerical features were subsequently scaled to the $[0, 1]$ range using Min–Max normalization:

$$x^{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{3}$$

This processed dataset was then used to construct supervised learning samples for model development.

C. Dataset Splitting

The processed data sequences were divided into training, validation, and test sets in a 70% / 15% / 15% ratio.

This split ensures that model evaluation is performed on unseen data while preserving temporal order.

D. Model Formulation

A Long Short-Term Memory (LSTM) network was employed to capture temporal dependencies in the sequential data. The LSTM cell computes hidden states and cell states as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{4}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad \tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad h_t = o_t \odot \tanh(C_t), \tag{6}$$

where x_t is the input at time t, h_t is the hidden state, C_t is the cell state, and σ denotes the sigmoid activation. The final hidden state h_w is fed into a fully connected layer to produce the predicted closing price:

$$\hat{y} = W_{\text{out}} h_w + b_{\text{out}}. \tag{7}$$

E. Training Procedure

The model was trained using the Mean Squared Error (MSE) loss function:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \tag{8}$$

and optimized using the Adam optimizer. Training was conducted for 20 epochs with validation monitoring to prevent overfitting. Batch size was set to 32, and gradient updates were performed at each batch iteration.

F. Evaluation Metrics

We assess predictive accuracy using four standard regression metrics: MSE, RMSE, MAE, and MAPE. MSE penalizes larger errors more heavily, making it useful for detecting models that occasionally produce large deviations. RMSE, as the square root of MSE, expresses this penalty in the original scale of the target variable, improving interpretability. MAE measures the average absolute deviation and is less sensitive to outliers, providing a stable indicator of overall error magnitude. MAPE captures the average percentage error, allowing performance to be interpreted relative to the true value; this is particularly useful in financial settings where proportional deviations matter.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}, \tag{9}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|, \quad \text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|. \tag{10}$$

These metrics together offer a balanced view of absolute, squared, and relative errors in model performance.

III. Results

This section presents the data pipeline, model training process, numerical evaluation, and interpretation of forecasting performance. The dataset consists of daily COMEX gold futures from January 2004 to October 2025. After removing missing values and dropping rows where trading volume equals zero, a total of 5487 observations remain. All features were scaled to the $[0, 1]$ range before modeling. To construct supervised learning samples, a sliding window of 30 days was used to generate input–target pairs.

The dataset was divided into 70% training data, 15% validation data, and 15% test data. The LSTM model was trained for 30 epochs. Training curves (Figure 2) show that both training and validation losses decrease smoothly, and no significant divergence is observed, indicating that the model learned meaningful temporal patterns without severe overfitting.

Model evaluation was performed on the test set by converting predictions back to the original USD price scale. The prediction trajectory (Figure 3) shows that the model performs well from 2022 through 2024, closely following the upward movements and local fluctuations of the true price series. However, predictions beyond this period exhibit higher variance and deviate more noticeably from actual prices, reflecting the difficulty of capturing long-horizon dynamics in volatile financial markets.

Table I reports the complete set of error metrics. The training set achieves the lowest error, with $\text{MSE} = 249.72$ and $\text{RMSE} \approx 15.80$. Validation error is slightly higher, with $\text{MSE} = 405.56$ and $\text{RMSE} \approx 20.14$. The test error is substantially larger, with $\text{MSE} = 5596.51$ and $\text{RMSE} \approx 74.81$, indicating reduced generalization capability when forecasting unseen future prices. Despite the high absolute errors, relative

performance remains reasonable, as the test set yields a MAPE of 1.68%, suggesting that predictions remain accurate in proportional terms.

Overall, these results indicate that the LSTM model successfully captures short- and medium-term temporal dependencies in gold price movements. The model produces reliable forecasts during stable or gradually trending periods but struggles with long-range extrapolation and sudden market shifts. This behavior aligns with common limitations of deep sequence models in financial prediction tasks.

TABLE I
PERFORMANCE METRICS ACROSS DATA SPLITS

Dataset	MSE	RMSE	MAE	MAPE (%)
Train	249.72	15.80	10.43	0.25
Validation	405.56	20.14	14.13	0.33
Test	5596.51	74.81	48.75	1.68

IV. CONCLUSION

In this work, we developed a data-driven forecasting model for COMEX gold futures using only historical daily price and volume data. We applied careful data cleaning, sliding-window preprocessing, feature scaling, and a Long Short-Term Memory network to model temporal dependencies. Our experiments—on training, validation, and test splits—demonstrate that the model is reasonably effective in capturing short- to medium-term price trends under stable market conditions. However, when predicting prices over longer horizons or during volatile phases, prediction errors increase substantially.

These results highlight both the potential and limitations of using pure historical data for price forecasting. On one hand, deep learning models such as LSTM can exploit temporal patterns without manual feature engineering. On the other hand, their predictive power declines as market dynamics deviate from historical patterns, possibly due to changing macroeconomic conditions, regime shifts, or unexpected events.

For future work, incorporating additional information — such as macroeconomic indicators, volatility indices, or news sentiment — could help improve prediction accuracy and robustness. Alternatively, adopting ensemble methods or hybrid models that combine statistical and machine-learning approaches may yield more stable forecasts. Overall, this study provides a foundation for gold price forecasting based on historical data and outlines practical limitations and paths for improvement.

REFERENCES

[1] M. Zou, "Study the relationship between VIX and COMEX gold futures price," pp. 82–87, Jul. 2024, doi: https://doi.org/10.1145/3690001.3690025.

[2] A. Moghar and M. Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network," Procedia Computer Science, vol. 170, pp. 1168–1173, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.049.

[3] Y. Liu, Y. Zhang, and X. Peng, "Textual analysis and gold futures price forecasting: Evidence from the Chinese market," Finance Research Letters, vol. 69, p. 106116, Nov. 2024, doi: https://doi.org/10.1016/j.frl.2024.106116.

[4] J. J. Murphy, Technical analysis of the financial markets : a comprehensive guide to trading methods and applications. New York: New York Institute Of Finance, 1999.

[5] "ZombitX64/xauusd-gold-price-historical-data-2004-2025 at main," Huggingface.co, Oct. 05, 2025. https://huggingface.co/datasets/ZombitX64/xauusd-gold-price-historical-data-2004-2025/tree/main (accessed Dec. 04, 2025).

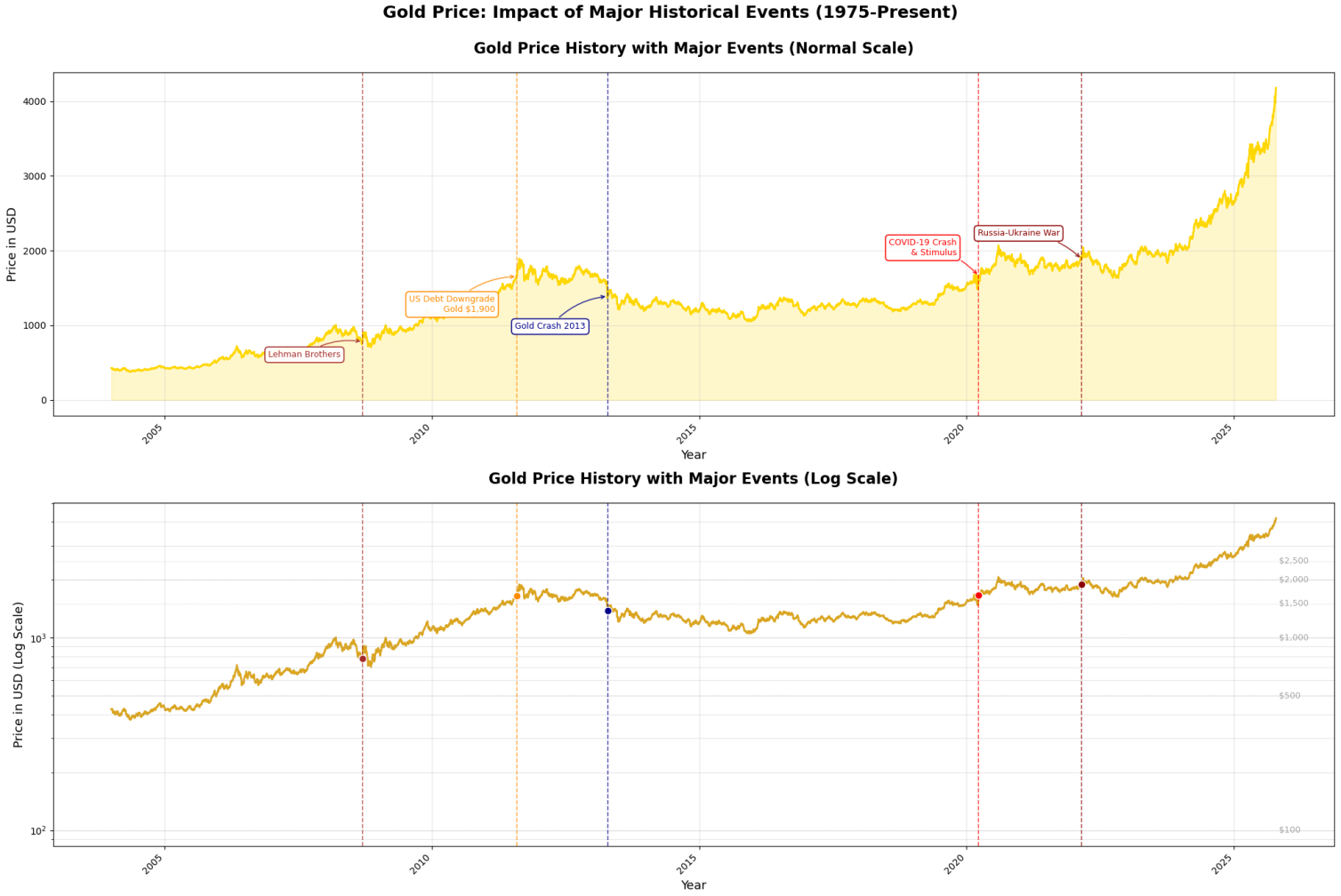

Fig. 1. Gold Price: 1975-2025

Fig. 2. LSTM: Training vs. validation Loss

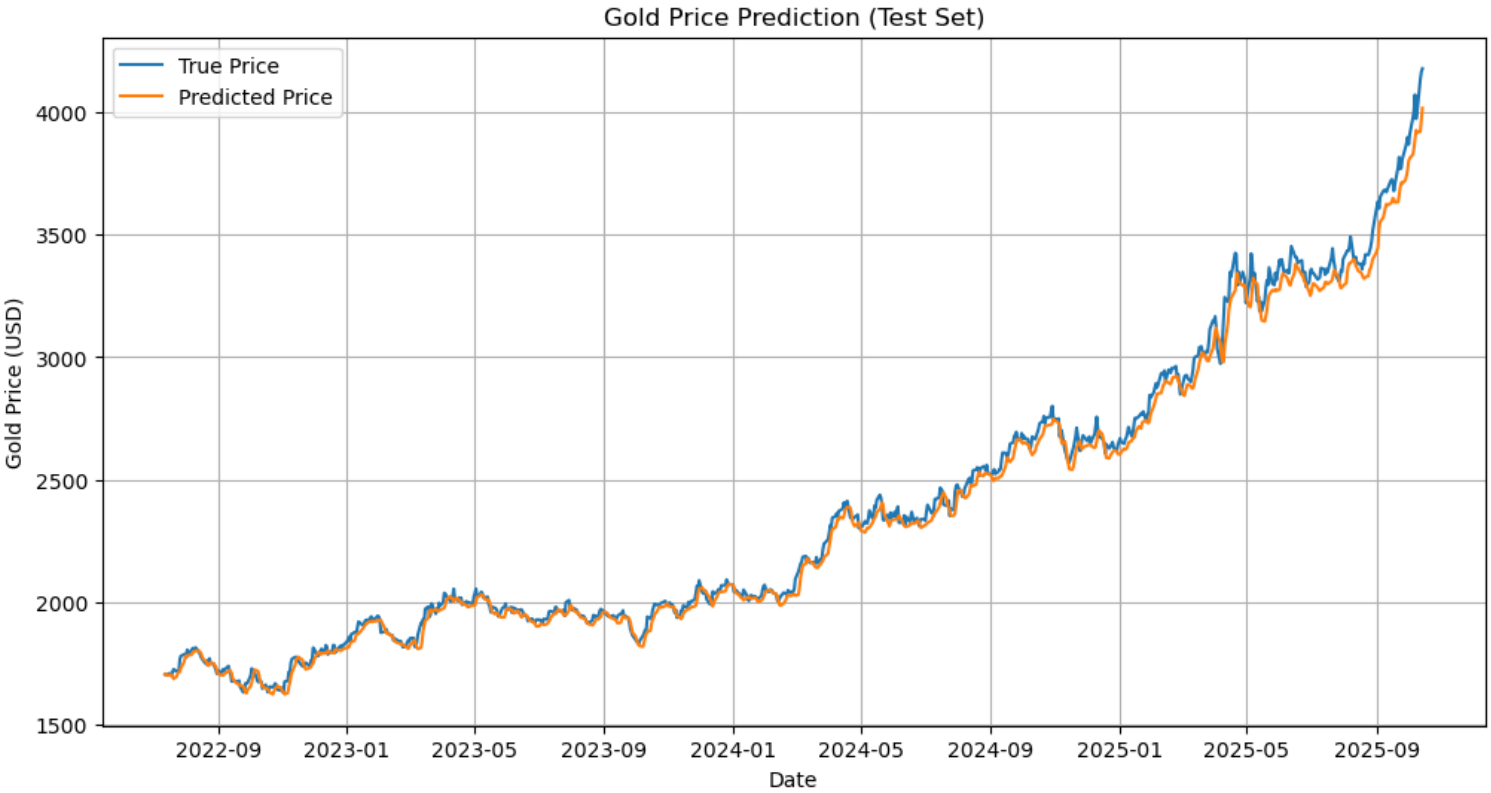

Fig. 3. LSTM: Prediction vs. True values of Test Set