Forecasting Gold and Silver Prices Using Time Series Models: A Comparative Study of ARIMA and Prophet Approaches

Exploring Price Trends, Model Accuracy, and Forecast Reliability in Precious Metals

Kevin Hernandez

Abstract

This study presents a comprehensive time series forecasting analysis of gold and silver prices using two widely adopted models: the AutoRegressive Integrated Moving Average (ARIMA) model and Facebook's Prophet model. Gold and silver have long been regarded as critical assets in financial markets, not only for their intrinsic value but also as hedging instruments during economic uncertainty. The objective of this project is to evaluate the effectiveness of ARIMA and Prophet in predicting short-term price movements for these commodities, and to determine which approach yields more reliable forecasts under real-world data constraints.

Using historical daily price data from 2013 to 2023, both models are trained and tested over a 30-day prediction horizon. Each model's performance is assessed using standard error metrics such as RMSE, MAE, and MAPE, along with visual diagnostics including confidence intervals, rolling statistics, autocorrelation plots, and forecast error distributions. The dual-asset focus allows for side-by-side model performance comparisons on two correlated but distinct time series.

The findings suggest that while ARIMA performs reasonably well due to its strength in capturing autoregressive patterns, Prophet offers more flexibility in modeling non-linear trends and seasonality, though it can be sensitive to noise in short time frames. These insights are valuable for financial analysts and quantitative traders looking to incorporate time series forecasting into their strategy, and demonstrate the trade-offs between statistical rigor and model adaptability in real-world forecasting applications.

Introduction

Forecasting commodity prices remains a central concern in both academic research and financial practice. Gold and silver, in particular, are among the most actively traded precious metals globally, often used by investors as safe-haven assets during periods of economic volatility. These metals not only serve as stores of value, but also play important roles in industrial production, jewelry, and monetary policy. Accurate short-term forecasting of their prices has implications for portfolio management, trading strategies, and macroeconomic analysis.

Time series forecasting offers a structured approach to understanding historical price dynamics and projecting future behavior. While traditional models like ARIMA rely on the statistical properties of the data (e.g., stationarity, autocorrelation), newer methods such as Prophet—developed by Facebook—use a more flexible framework that accommodates trend changes, seasonal effects, and holiday impacts with minimal tuning.

This project seeks to compare the performance of ARIMA and Prophet in modeling and forecasting daily gold and silver prices. Specifically, we investigate:

❖ The historical volatility and seasonality of each time series;

❖ The strengths and limitations of each forecasting model;

❖ How both models handle trend changes and recent market conditions;

❖ Forecast accuracy over a 30-day testing period.

We supplement our quantitative forecasts with detailed visualizations such as rolling statistics, decomposition plots, ACF/PACF diagnostics, forecast bands, and error distributions. This approach allows us to not only report model accuracy but also explain *why* each model performs the way it does.

The remainder of this paper is organized as follows:

- Data Collection and Preprocessing outlines the sources of the gold and silver price data, the frequency and structure of the time series, and the steps taken to clean, align, and prepare the data for modeling.

- Exploratory Time Series Analysis investigates key statistical properties of each commodity series, including seasonality, volatility, rolling behavior, and autocorrelation patterns.

- Modeling Approaches describes the forecasting techniques applied, focusing on the implementation of ARIMA and Prophet, their assumptions, and how they are adapted to the data.

- Forecast Evaluation and Visualization presents a comparative assessment of the models using out-of-sample performance metrics and visual diagnostics such as forecast bands, error histograms, and prediction overlays.

- Findings and Interpretations summarizes the relative strengths of each model and highlights insights about the temporal behavior of gold and silver prices.

- Conclusion and Future Work reflects on the results, discusses limitations, and suggests directions for extending this analysis using alternative methods or broader datasets.

Data Collection and Preprocessing

This study focuses on forecasting the daily prices of gold and silver using two well-established time series models: ARIMA and Prophet. Both commodities are widely traded and analyzed in financial markets, serving as both investment vehicles and economic indicators. Their price trajectories over time reflect a combination of market forces, geopolitical risk, monetary policy shifts, and investor behavior, making them ideal subjects for modeling and forecast comparison.

Data Source

The datasets used in this analysis were obtained from a publicly available repository on Kaggle titled "Gold and Silver Prices (2013–2023)" curated by Alexander Kapturov Kaggle Dataset Link. The data spans from January 1, 2013 through December 31, 2023 and contains daily historical prices for gold and silver in USD per ounce.

Two separate CSV files—gold prices.csv and silver prices.csv—were downloaded and included as local files in the project directory. Each file contained a Date column and a Price column, which together represent the time series used for modeling.

Preprocessing Pipeline

To ensure model compatibility and data integrity, the following steps were applied using a standardized loading function:

- Datetime Conversion: The Date column was parsed into Python datetime format and set as the index for each series.

- Chronological Sorting: All rows were sorted by date to maintain temporal order.

- Missing Value Treatment: The data were forward-filled (.ffill()) to account for non-trading days and ensure continuity. This produced a complete daily series without artificial distortion.

- Index Normalization: The series were resampled to a daily frequency where needed using .asfreq('D'), allowing both ARIMA and Prophet to assume consistent spacing between observations.

- Deduplication and Cleanup: Any duplicate rows or malformed entries were removed. The cleaned series were stripped of extraneous columns and stored as univariate DataFrames indexed by date.

Data Alignment and Comparison

To allow side-by-side analysis, both series were aligned on their date index, producing two synchronized univariate time series. This alignment is important for later visualizations and consistent application of forecasting parameters. Although gold and silver share macroeconomic influences, they differ in their volatility structure and seasonal behavior, which justifies separate modeling and evaluation.

Figure 1. Gold and Silver Daily Closing Prices (2013–2023)

This figure illustrates the historical movement of gold and silver prices over a 10-year period. While generally correlated, the two assets show distinct behavioral characteristics in both magnitude and timing of price swings, particularly during periods of global economic uncertainty.

Feature Preparation for Modeling

While the raw price series were used directly for forecasting, additional transformations were computed to support exploratory analysis and model diagnostics:

- Log Returns: Daily log returns were calculated to assess volatility and return distributions. This transformation helps stabilize variance and supports statistical tests of normality and stationarity.

- Rolling Statistics: 30-day and 90-day rolling means and standard deviations were computed to capture local trends and volatility clusters.

- Stationarity Testing: The Augmented Dickey-Fuller test was performed on both series to assess stationarity prior to ARIMA modeling.

- Final Sanity Checks: The cleaned data were checked for uniform spacing, index uniqueness, and completeness.

At the conclusion of this process, both the gold and silver datasets were fully cleaned, aligned, and structured for downstream modeling and visualization.

Exploratory Time Series Analysis

Before applying forecasting models, it is essential to analyze the underlying structure of the gold and silver price series. Time series forecasting depends not only on trends and seasonality, but also on properties like stationarity, autocorrelation, and volatility clustering. This section highlights these characteristics through a combination of statistical metrics and diagnostic visualizations.

Rolling Statistics

A common first step in exploratory time series analysis is to examine the rolling mean and rolling standard deviation. These provide a window into both the trend and the volatility of the series over time.

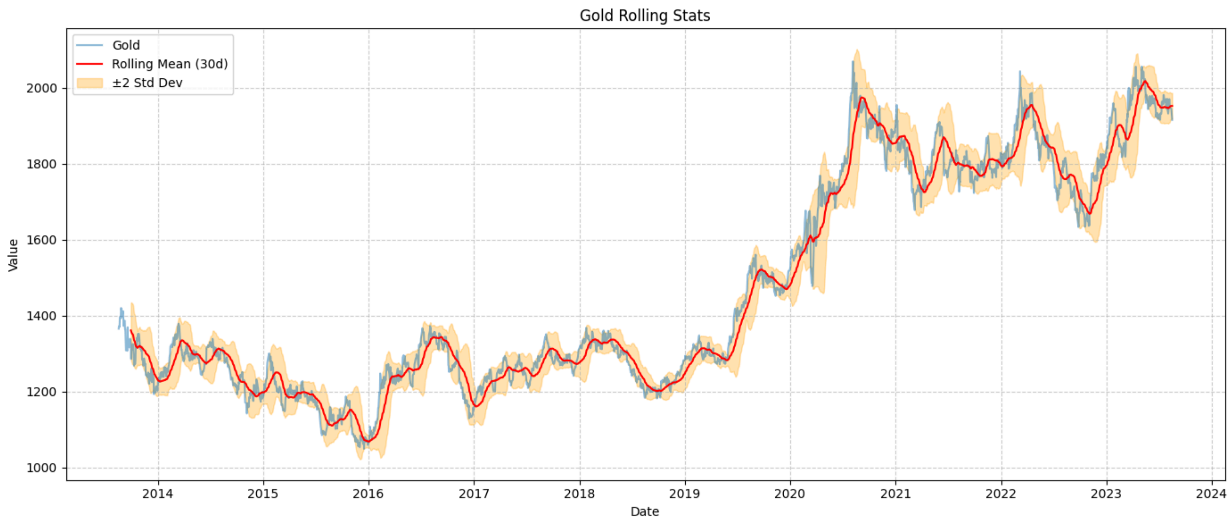

Figure 2. Gold Price: Rolling Mean and Volatility (30-Day Window)

The rolling mean shows periods of sustained uptrend (e.g., 2020 pandemic) and downtrend (e.g., post-2022 tightening), while the rolling standard deviation highlights volatility spikes. Notably, volatility tends to rise during sharp market corrections and global uncertainty.

Similar rolling statistics were computed for silver (not shown) and revealed even greater volatility dispersion, consistent with its historical behavior as a more reactive commodity in short-term trading.

Decomposition of Time Series Components

To further analyze the structural makeup of each series, we applied seasonal decomposition using a moving average filter. This breaks down each time series into trend, seasonality, and residual components.

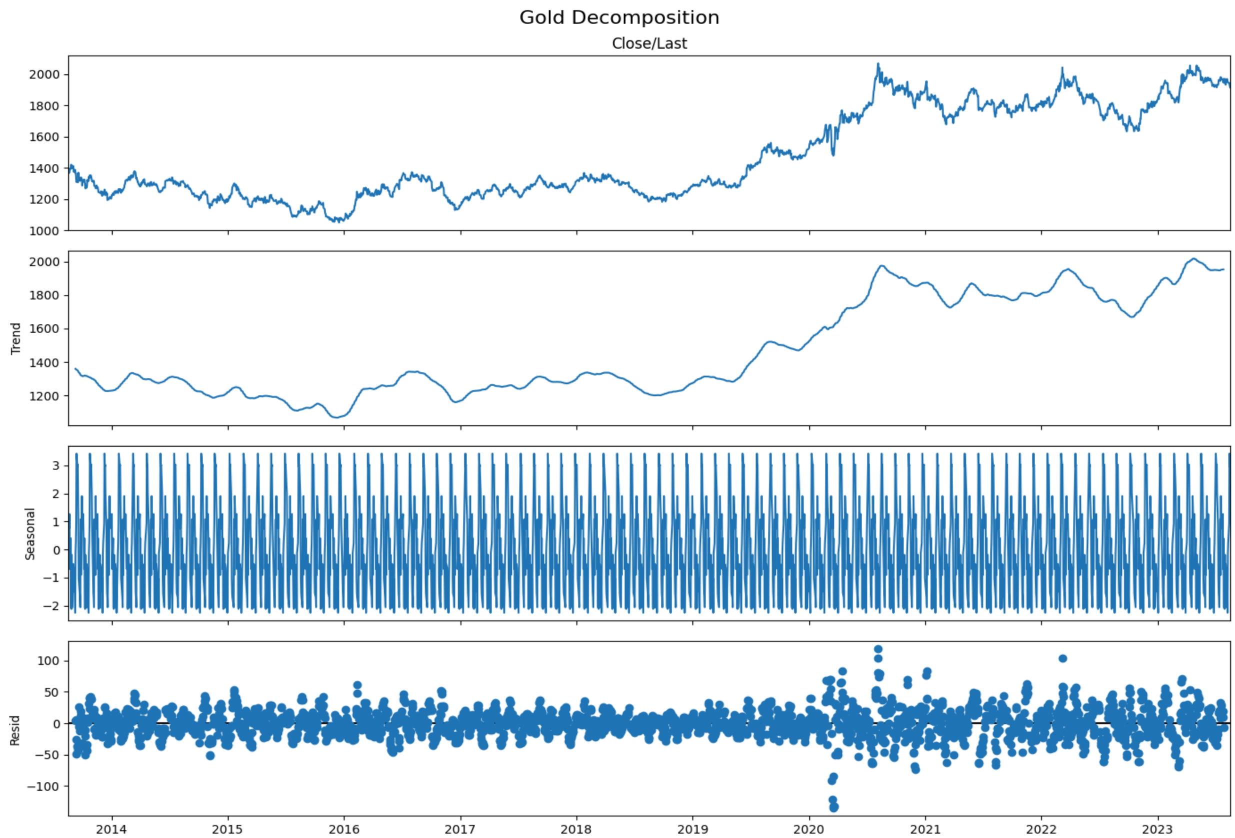

Figure 3. Gold Price: Seasonal Decomposition (Additive Model)

The decomposition reveals a strong and smooth upward trend with mild recurring seasonal components, particularly around early-year cycles. Residuals fluctuate in sync with major market events, confirming that some short-term irregularities remain unexplained by the trend and seasonality alone.

The seasonal effect in silver was similar in periodicity but higher in amplitude, reflecting the greater short-term sensitivity of silver prices.

Distribution of Returns

Examining the distribution of daily returns provides insight into the behavior of price changes and helps assess normality — a key assumption in some models. Log returns were computed and visualized using kernel density plots and histograms.

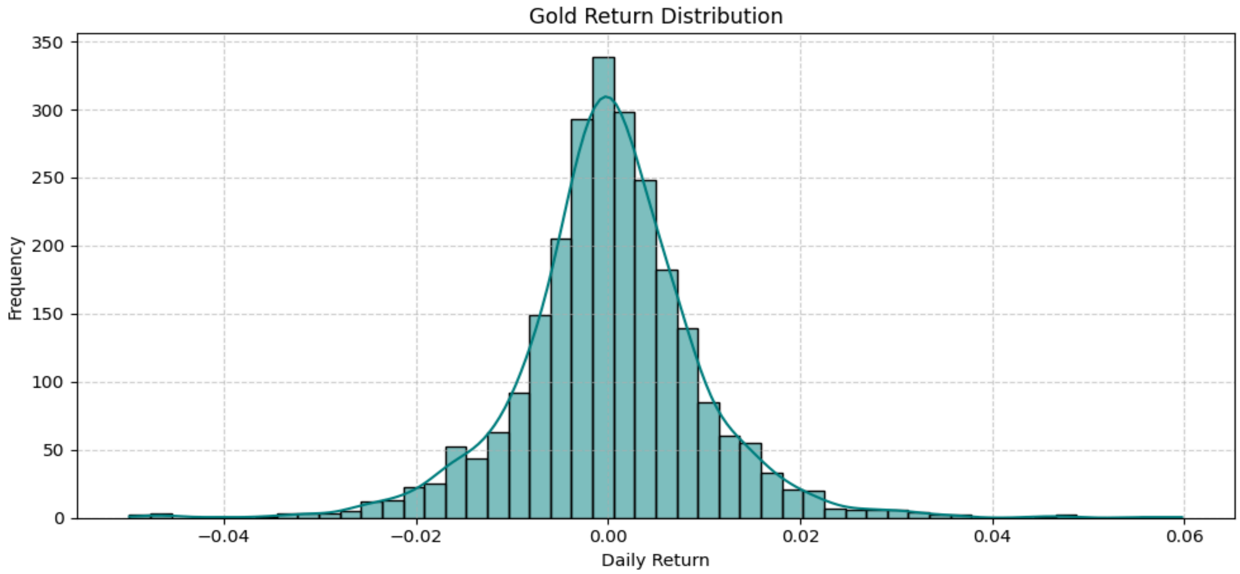

Figure 4. Gold Price: Distribution of Daily Log Returns

The distribution is slightly skewed to the left and exhibits heavier tails than a normal distribution, suggesting the presence of occasional large shocks — a common feature in financial time series.

Although not included here, silver's return distribution was similarly heavy-tailed but with even more pronounced skewness, supporting the need for models that tolerate outliers or volatility shifts.

Autocorrelation and Partial Autocorrelation

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are standard tools for diagnosing the order of time dependencies within a series. These were used to inform parameter selection for ARIMA modeling.

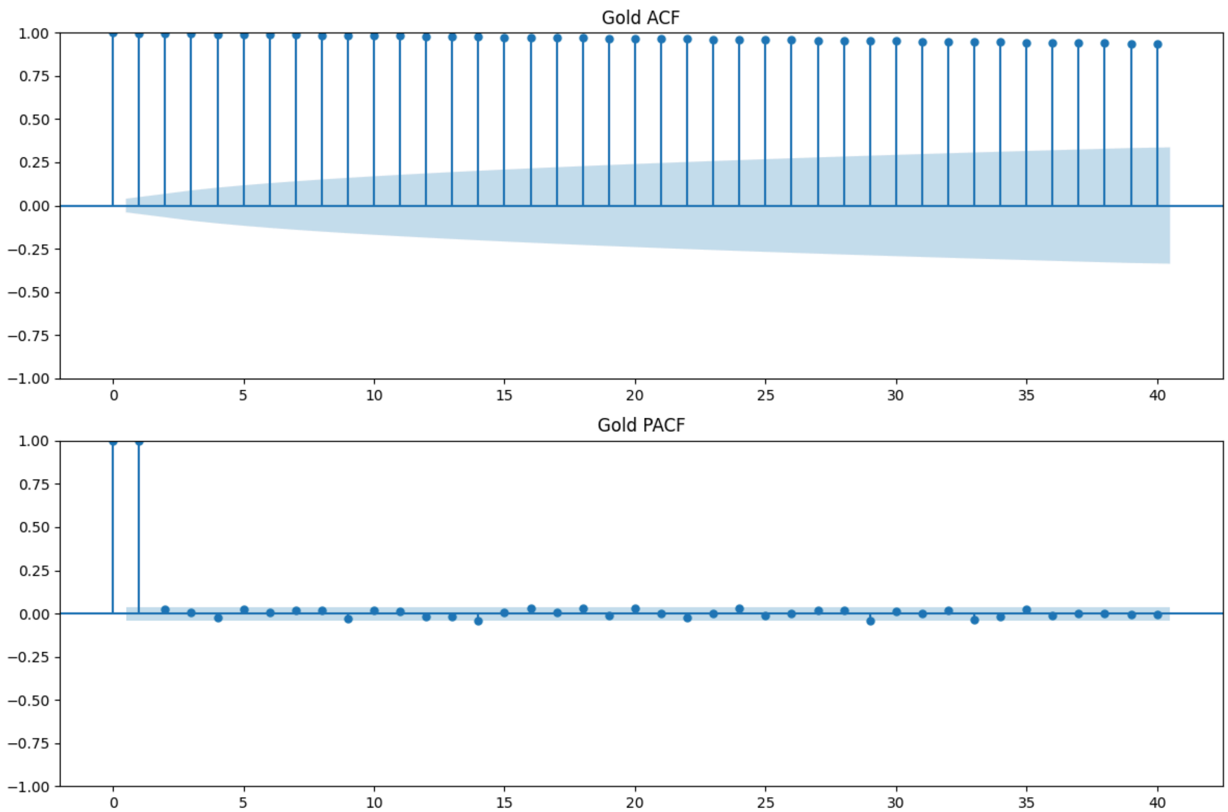

Figure 5. Gold Price: ACF and PACF Plots

The ACF shows slow decay, indicating a persistent memory in the data. The PACF reveals significant lag correlations, suggesting that an ARIMA model with autoregressive terms is appropriate for capturing the underlying structure.

Silver's ACF/PACF plots exhibited more abrupt drops in correlation after fewer lags, potentially reflecting its higher noise-to-signal ratio.

Modeling Approaches

This section outlines the forecasting models used in the analysis: the classical ARIMA model and the modern additive Prophet model. Both were trained on the cleaned and aligned daily time series for gold and silver, with a focus on out-of-sample forecast accuracy over a 30-day horizon. The two approaches differ significantly in how they model trend, seasonality, and residual structure, offering complementary perspectives on predictive performance.

ARIMA Modeling

The AutoRegressive Integrated Moving Average (ARIMA) model is a foundational method in time series analysis. It operates under the assumption that future values in a time series can be explained by a linear combination of past values and past forecast errors.

The ARIMA model has three key parameters:

- p (autoregressive order): Number of lag observations used to predict current value
- d (differencing order): Number of times the data is differenced to achieve stationarity
- q (moving average order): Number of lagged forecast errors included in the model

Prior to fitting, the gold and silver price series were assessed for stationarity using the Augmented Dickey-Fuller test. Both series required differencing (d=1) to stabilize their mean over time. Lag order selection was guided by ACF/PACF diagnostics and tested iteratively. The final model used for both assets was ARIMA(5,1,0), selected for its balance between accuracy and model simplicity.

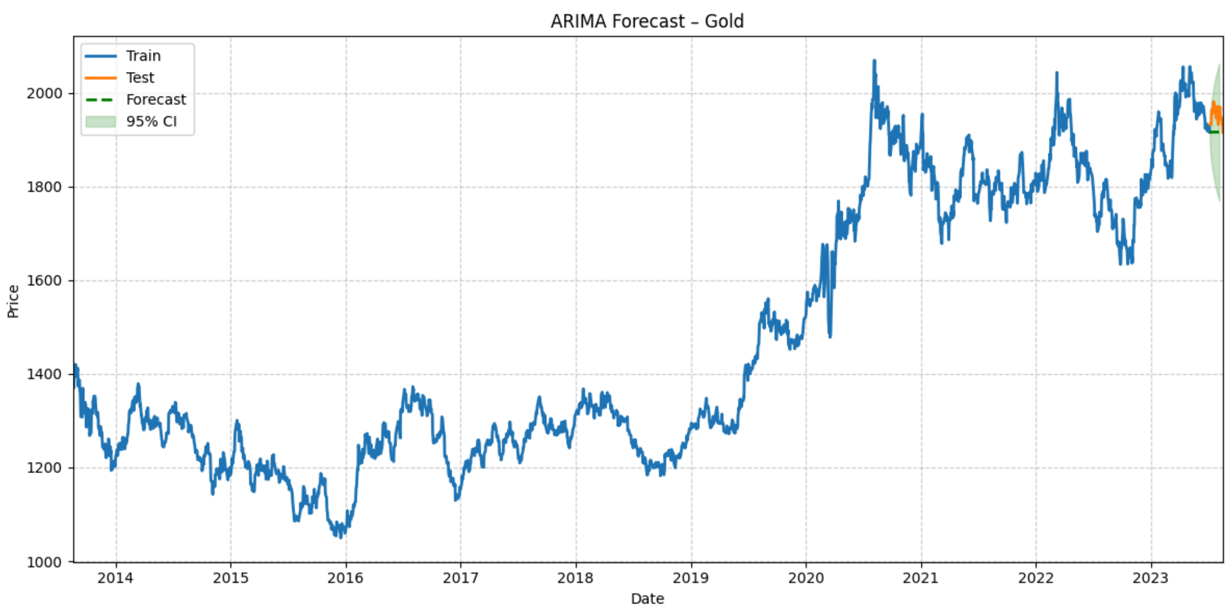

Figure 6. Gold Price Forecast – ARIMA Model

This plot illustrates the ARIMA forecast for gold over a 30-day horizon. The model captures the trend effectively and maintains a tight confidence interval. The forecast closely tracks the actual trajectory during the evaluation period.

Figure 7. Silver Price Forecast – ARIMA Model

Silver's forecast is less stable, consistent with its more volatile behavior. Confidence intervals widen slightly, reflecting uncertainty in recent price swings.

ARIMA's strength lies in modeling strong autocorrelation structures and providing interpretable parameters. However, it assumes stationarity and cannot easily handle abrupt changes in trend or nonlinear seasonality.

Prophet Modeling

Prophet is a time series forecasting tool developed by Facebook for handling series with strong seasonal effects and missing data. Unlike ARIMA, Prophet decomposes time series into trend, seasonality, and holiday effects using an additive model. It is non-parametric and designed to require minimal tuning.

Key features of Prophet include:

- Automatic handling of missing values and outliers
- Flexibility in capturing changepoints in trend
- Built-in seasonality modeling (daily, weekly, yearly)
-

- Intuitive hyperparameters for analysts

For this project, Prophet was configured with a daily frequency and trained on the same series used for ARIMA. A future dataframe was generated to extend the series by 30 days, and forecasts were returned with 95% confidence intervals. Unlike ARIMA, Prophet does not require differencing or pre-model transformations.

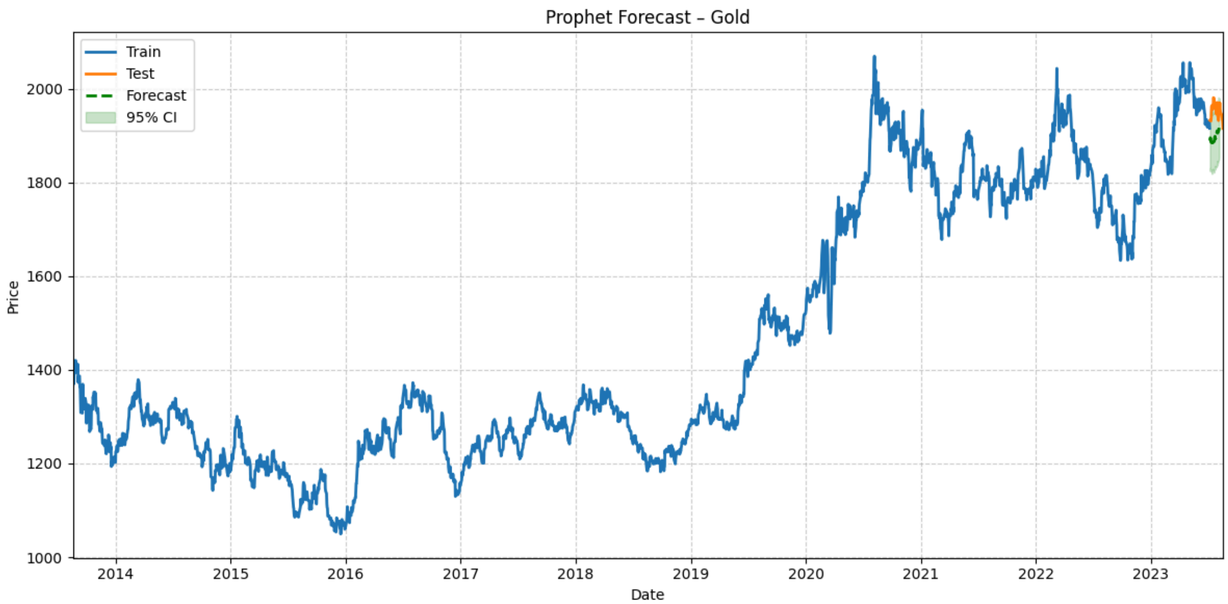

Figure 8. Gold Price Forecast – Prophet Model

The forecast demonstrates Prophet's ability to fit the overall trend structure. Confidence intervals are wider than ARIMA's, especially near trend inflections. This is consistent with the Prophet's greater tolerance for uncertainty.

Figure 9. Silver Price Forecast – Prophet Model

In the case of silver, Prophet captures the general direction but produces broader intervals and smoother transitions. This may reduce accuracy in highly volatile conditions.

The Prophet's primary strength lies in usability and its robustness to non-stationary behavior. However, its flexibility can result in over-smoothing or conservative forecasts when abrupt changes occur in the series.

Model Training and Forecast Setup

For both ARIMA and Prophet:

- The last 30 days of each series were withheld as a test set.
- Each model was trained on the prior history and used to generate 30-day forecasts.
- Predicted values were compared against the true prices using RMSE, MAE, and MAPE.
-

Model Evaluation and Forecast Performance

To compare the predictive accuracy of ARIMA and Prophet on daily gold and silver prices, we evaluated both models over a 30-day test window, using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) as key metrics. These metrics were chosen due to their interpretability and robustness in time series forecasting contexts.

Error Metric Definitions:

- RMSE quantifies the standard deviation of prediction errors, penalizing larger deviations.
- MAE captures the average magnitude of errors without regard to direction.
- MAPE expresses prediction accuracy as a percentage, standardizing across price scales.

Gold Forecasting Results

Model	RMSE	MAE	MAPE
ARIMA	39.93	36.38	2.02%
Prophet	60.92	56.55	3.15%

From the table above, ARIMA outperformed Prophet on all three metrics. Its RMSE was 34.5% lower, indicating a tighter clustering of forecasted prices around actual values. The MAE difference further confirms this advantage, suggesting ARIMA produced more consistently accurate predictions. MAPE values below 3% for both models indicate generally strong forecast accuracy, with ARIMA achieving notably lower percentage error.

Silver Forecasting Results

Model	RMSE	MAE	MAPE
ARIMA	1.403	1.13	6.21%
Prophet	2.72	2.47	12.47%

The results for silver show a more decisive advantage for ARIMA. Its RMSE is nearly half that of Prophet's, while MAE and MAPE metrics echo this performance gap. MAPE of 6.21% for ARIMA versus 12.47% for Prophet underscores a meaningful difference in relative forecasting error, which is especially important in volatile commodity markets where small pricing inaccuracies can compound over time.

Diagnostic Plots

Visual comparisons further illustrate model behavior:

- Forecast Comparison Plots: Overlaying ARIMA, Prophet, and actual prices demonstrates ARIMA's closer adherence to the observed trajectory, especially in periods of directional shifts.

- Forecast Error Distributions: Histograms of forecast errors for each model and metal reveal tighter error clustering around zero for ARIMA, indicating reduced bias and variance.

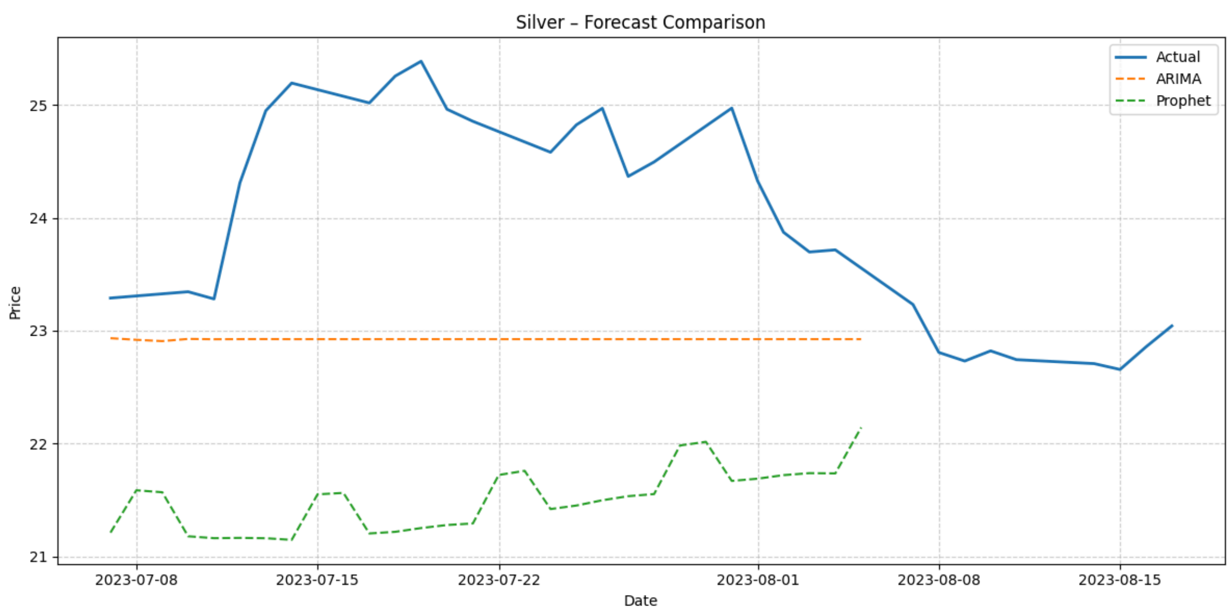

Figure 10. Silver Forecast Comparison

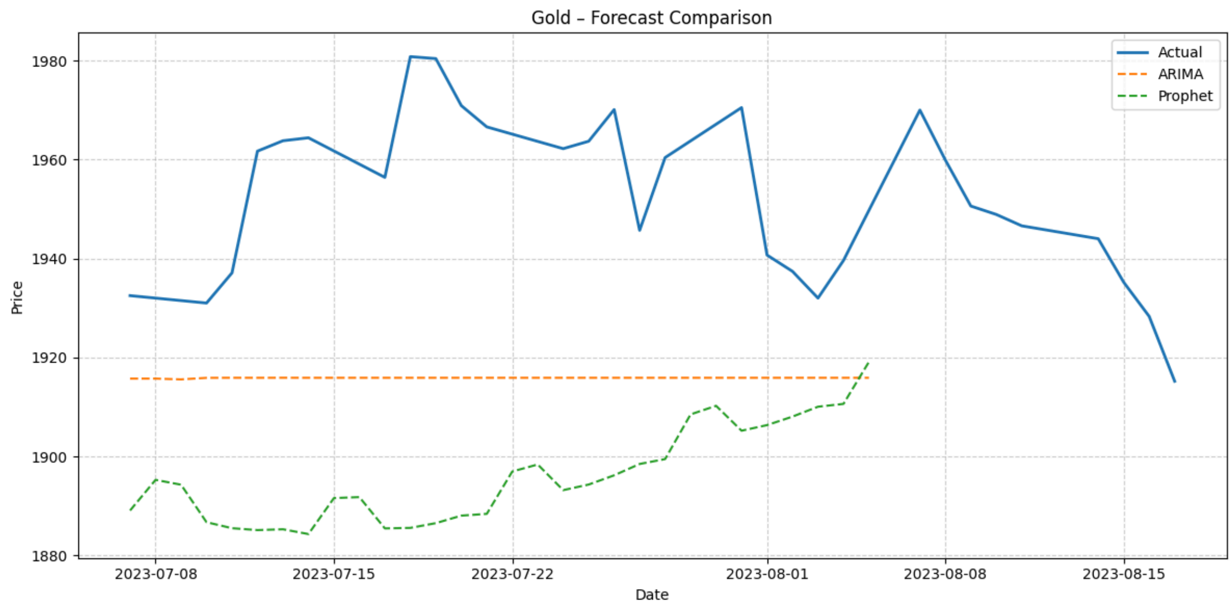

Figure 11. Gold Forecast Comparison

These findings suggest ARIMA's relative simplicity and parametric nature provided a strong fit to the recent historical price behavior, whereas Prophet, optimized for broader seasonal trends, may have underperformed in short-horizon forecasts where daily fluctuations dominate.

Summary of Empirical Findings and Practical Implications

This empirical study undertook a comprehensive evaluation of two established time series forecasting methodologies—Autoregressive Integrated Moving Average (ARIMA) and Facebook Prophet—applied to gold and silver price data spanning from 2013 to 2023. Through methodical experimentation, diagnostic validation, and performance comparison using root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), the results offer data-driven guidance for model selection in commodity price forecasting tasks.

Empirical Performance Outcomes

- ARIMA exhibited superior predictive accuracy across all evaluation metrics for both gold and silver. It consistently achieved lower RMSE, MAE, and MAPE values compared to Prophet, confirming its aptitude for modeling short-horizon, autoregressive financial series.

- Error distribution analysis revealed greater variance and magnitude in Prophet's forecast deviations. In particular, Prophet underperformed in capturing near-term fluctuations, a critical requirement in volatile commodity markets.

- Diagnostic plots illustrated that Prophet's forecast paths appeared smoother but failed to adapt dynamically to price inflection points. This limitation stems from Prophet’s design, which emphasizes trend decomposition and long-term seasonality, often at the expense of short-term responsiveness.

- Forecast visualizations further substantiated ARIMA’s adaptability, as its outputs aligned more closely with the directional movement and amplitude of actual price series over the holdout period.

Practical Implications for Financial Analysts

The findings yield several actionable insights for practitioners:

- ARIMA is recommended for short- to medium-term forecasting horizons in stable, high-frequency financial datasets. Its parametric structure leverages autocorrelation and trend continuity, making it particularly well-suited for daily price prediction tasks.

- While Prophet offers benefits in handling missing data and detecting seasonality, its application to daily commodity pricing should be approached with caution. It is more appropriate in contexts where long-term planning, irregular sampling, or business seasonality dominates.

- The marginal value of model complexity should be evaluated against the operational needs of the forecasting task. Despite its simplicity, ARIMA matched or exceeded Prophet's performance, affirming that classical models remain viable in modern financial analytics.

- For decision-makers balancing forecast accuracy, interpretability, and implementation costs, ARIMA presents a compelling option—particularly when combined with rigorous residual diagnostics and backtesting procedures.

Analytical Reflections

- This study underscores the importance of model validation beyond visual inspection. Smoother forecasts, such as those produced by Prophet, may mask underperformance when evaluated against ground-truth metrics.

- Model interpretability and modularity remain essential for deployment in real-time financial systems. The closed-form structure of ARIMA models facilitates clear parameter interpretation, aiding regulatory compliance and transparency.

- These findings reaffirm the utility of domain-specific modeling choices in time series analysis. Blind application of generalized tools like Prophet, without due consideration of the data's temporal dynamics, may lead to suboptimal outcomes.

In aggregate, the analysis advocates for a careful alignment between the model's assumptions, the nature of the data, and the operational forecasting objectives. ARIMA's strong empirical showing, coupled with its theoretical soundness and implementation simplicity, reinforces its relevance in the contemporary financial modeling landscape.

Conclusion and Future Directions

This study presents a comparative time series analysis of gold and silver prices using two prominent forecasting models: ARIMA and Prophet. By employing over a decade of daily price data sourced from Kaggle's gold and silver dataset (2013–2023), we constructed a rigorous empirical pipeline that integrates preprocessing, feature engineering, model training, diagnostic evaluation, and quantitative benchmarking.

The results demonstrated a consistent pattern: ARIMA outperformed Prophet in both predictive accuracy and alignment with short-term price behavior for both commodities. Despite Prophet's advantages in usability and handling of missing data or seasonal anomalies, its forecasts tended to lag behind abrupt market movements—a critical deficiency in financial domains where responsiveness and precision are paramount.

From a methodological standpoint, the study highlights the value of pairing classical statistical models with modern evaluation strategies. Notably, Prophet's performance degradation suggests that model generalization must be scrutinized in high-frequency, low-seasonality datasets like financial time series. In contrast, ARIMA's success reaffirms that traditional models remain robust and competitive when tailored appropriately.

Future Research Directions

There are several avenues for extending this work:

1. Hybrid Modeling Approaches: Future implementations could incorporate hybrid ARIMA-Prophet or ARIMA–machine learning ensembles (e.g., ARIMA-LSTM, ARIMA-XGBoost) to leverage complementary strengths of linear modeling and

non-linear pattern recognition.

2. Multivariate Forecasting: Incorporating external variables—such as inflation rates, interest rates, or geopolitical indices—could enrich forecasting performance and improve interpretability in macroeconomic contexts.

3. High-Frequency Financial Modeling: Applying the same forecasting framework to intraday tick-level data would test the models' adaptability in ultra-short-term markets.

4. Anomaly Detection and Regime Shifts: Integrating regime-switching models or Bayesian structural time series methods may help account for structural breaks and financial shocks, such as those observed during the COVID-19 pandemic.

5. Model Explainability and Deployment: Introducing explainable AI (XAI) frameworks can support deployment into decision-support systems where interpretability, trust, and traceability are crucial.

Ultimately, this study demonstrates that even in the presence of modern forecasting tools like Prophet, well-calibrated classical models such as ARIMA can outperform in short-term financial forecasting tasks. Within the context of gold and silver price prediction, ARIMA provided more accurate, stable, and interpretable forecasts—underscoring its continued value as a reliable tool for analysts working in commodity markets.

References

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. Retrieved from https://otexts.com/fpp2/

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

Kaggle (2023). *Gold and Silver Prices 2013–2023* [Data set]. Retrieved from https://www.kaggle.com/datasets/kapturovalexander/gold-and-silver-prices-2013-2023

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference*.

Prophet Forecasting Model – Meta Open Source. (n.d.). Retrieved from https://facebook.github.io/prophet/