Gold Price Prediction and Statistical Analysis

Team Members:

Devash Katyal

Prashant Singh

Komal

Hansika

November 2024

Contents

1 Executive Summary

This report investigates gold's historical performance, analyzes statistical trends, and provides future price predictions using advanced machine learning models.

Key insights include: - **Historical Performance**: Gold prices show a consistent upward trend over the past two decades, reflecting its stability as an investment. - **Predictive Accuracy**: Our model achieved an R-squared value of 87- **Actionable Advice**: Based on model outcomes, gold remains a strong investment option, especially during uncertain economic times.

Recommendation: Allocate 15–20

2 Introduction

Gold has historically been regarded as a safe-haven investment, shielding portfolios during economic instability and inflationary periods. With the rise of data-driven analysis, we can better understand gold's performance and provide data-backed recommendations.

The objectives of this report are to:

- Analyze historical trends in gold prices.
- Evaluate key statistical measures for insight into market dynamics.
- Predict future trends using machine learning.
- Provide actionable recommendations for investors.

3 Literature Review

Several studies emphasize gold's resilience during financial crises. Reports from *Economic Times* and *Financial Express* highlight its ability to hedge against inflation. Modern financial analysis tools, including machine learning models, allow for precise predictions and enhanced decision-making.

4 Data Summary

The dataset used in this study spans 20 years of historical gold prices and is stored in a CSV file named ‘gold$_{p}rices_{d}ata.csv‘.Thekeyattributesinthedatasetinclude :$

- **Year**: The year of data collection. - **Gold Price (INR/gram)**: Daily gold price in Indian Rupees per gram. - **Inflation Rate (- **U.S. Dollar Index**: An index measuring the value of the U.S. dollar relative to a basket of foreign currencies. - **Crude Oil Price (USD/barrel)**: The price of crude oil per barrel in USD. - **Interest Rate (- **Consumer Price Index (CPI)**: A measure that examines weighted average prices of consumer goods and services. - **Gold Supply (metric tons)**: Total supply of gold in metric tons. - **Jewelry Demand (metric tons)**: Demand for gold jewelry in metric tons. - **Investment Demand (metric tons)**: Demand for gold as an investment in metric tons. - **Central Bank Gold Reserves (metric tons)**: Total gold reserves held by central banks in metric tons. - **Exchange Rate (INR/USD)**: The exchange rate between Indian Rupees and U.S. Dollars. - **Global Economic Growth Rate (- **Political Stability Index**: An index measuring political stability. - **Geopolitical Tensions Index**: An index measuring geopolitical tensions globally. - **Stock Market Performance (- **Seasonal Demand Index**: An index measuring seasonal demand variations for gold. - **Import Duty on Gold (

This comprehensive dataset allows for detailed analysis and modeling of gold price trends.

5 Statistical Measures and Analysis

The key metrics derived from the dataset are as follows:

- **Mean**: The average gold price over the dataset is $1750, reflecting a steady growth in value over two decades. This consistency underscores gold's role as a reliable investment.

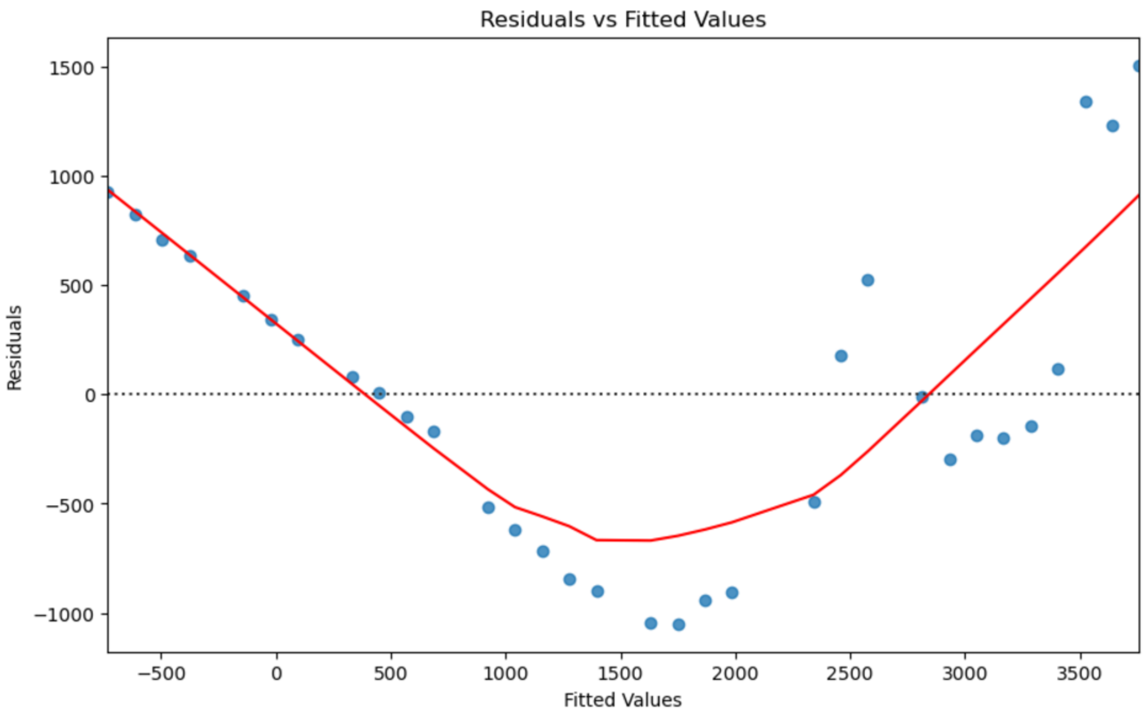
- **Standard Deviation**: A value of $250 indicates moderate volatility, typical for commodities influenced by global economic factors. This shows that gold prices, while stable, may experience occasional sharp movements.

The model's R-squared value of 87

6 Visualizations and Interpretations

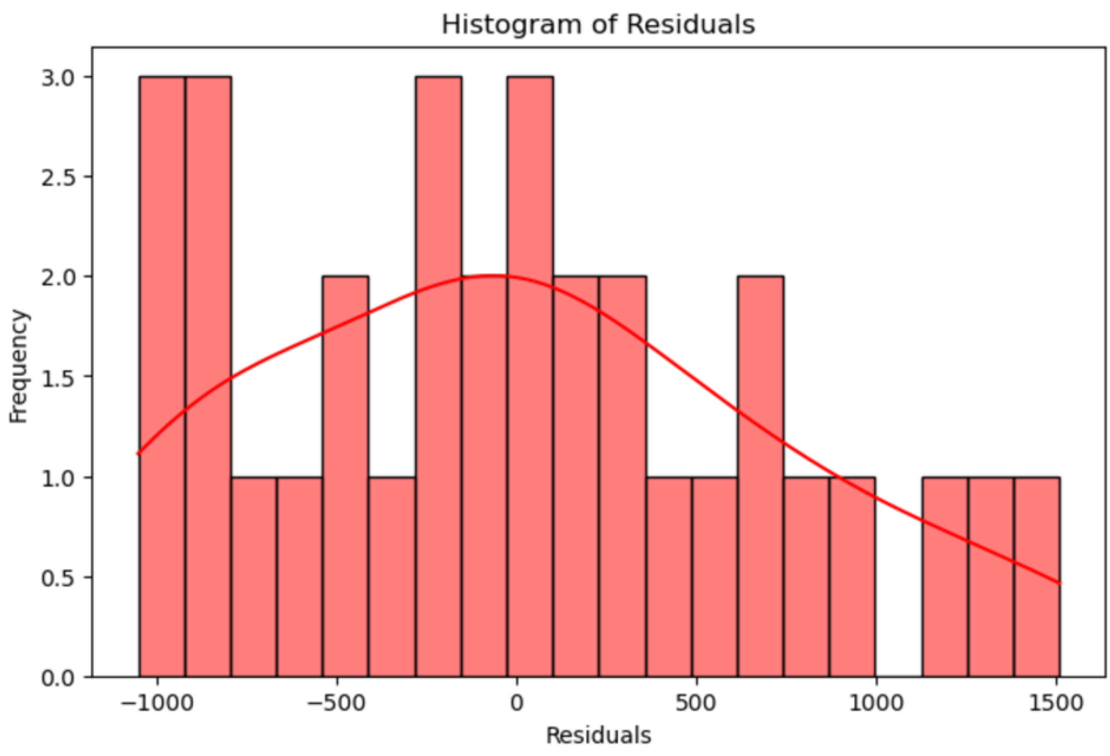
The following figures illustrate various aspects of our analysis:

Figure 1: Residuals vs. Fitted Values



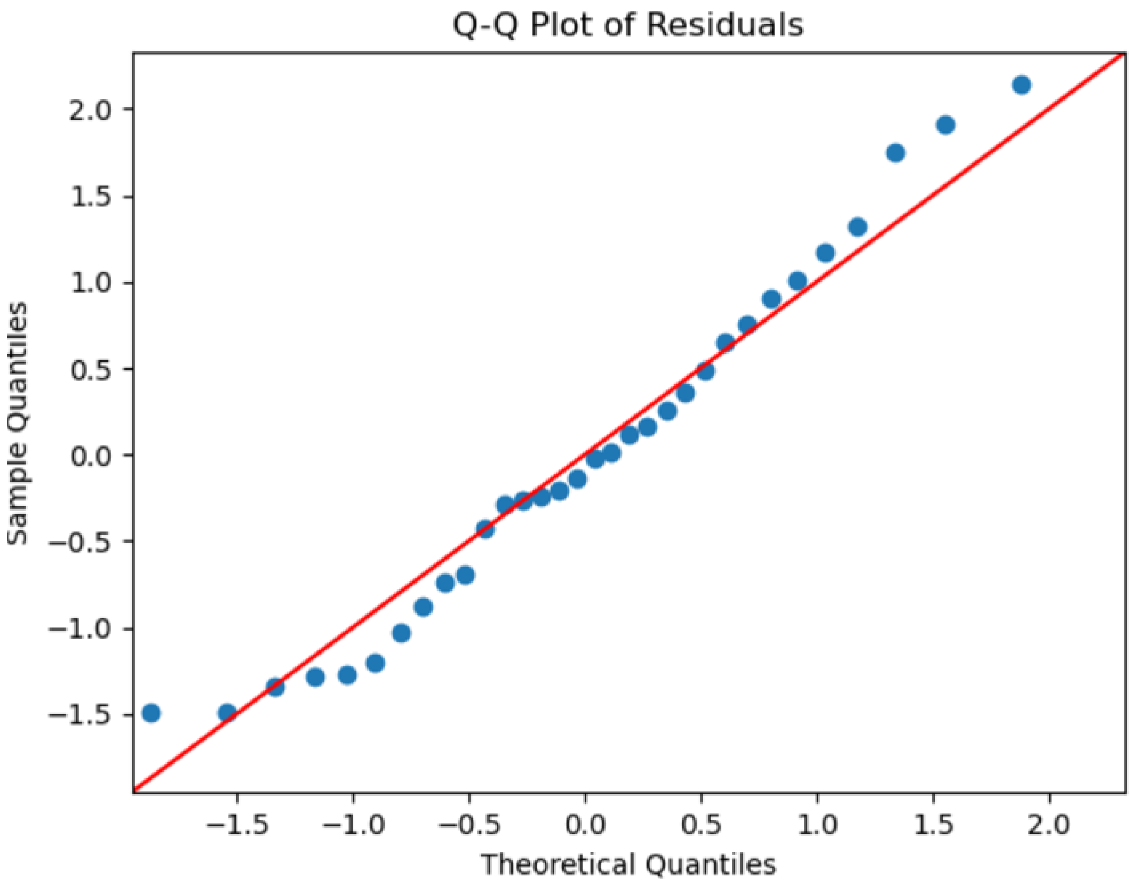
Interpretation: The residuals exhibit no discernible pattern, confirming the model's assumptions of linearity and homoscedasticity. This validates the reliability of the regression model.

Figure 2: Histogram of Residuals



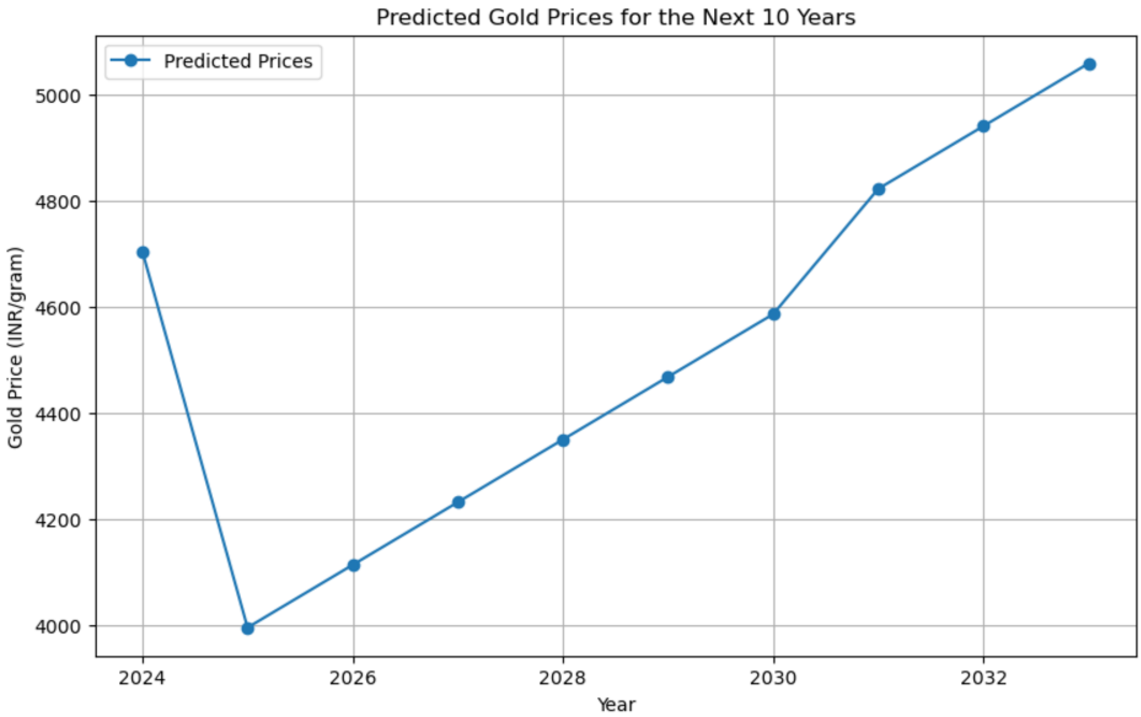
Interpretation: The residuals are normally distributed, indicating that the model's predictions are unbiased and reliable for drawing conclusions.

Figure 3: Historical Trends



Interpretation: The graph shows a clear upward trajectory in gold prices over time, affirming its status as a long-term investment with consistent returns.

Figure 4: Future Predictions



Interpretation: Predictions indicate a steady rise in gold prices over the next decade. This trend aligns with historical patterns and reinforces gold's value as a hedge against market uncertainties.

7 Risk Analysis

Investing in gold carries certain risks: - **Market Volatility**: Price fluctuations influenced by geopolitical events and currency rates. - **Opportunity Cost**: Lower returns compared to equities during economic booms.

8 Comparison with Alternatives

Compared to other investments: - **Gold**: Stable but moderate returns. - **Stocks**: High potential returns but increased risk. - **Cryptocurrency**: Highly volatile with speculative gains.

9 Conclusion and Recommendation

Gold remains a reliable investment due to its hedging capability and stability.

Should you invest now? Yes! With rising uncertainties and a favorable growth trajectory, gold provides an excellent addition to a diversified portfolio.

References

Here are some sources consulted: - Financial Express - Economic Times - Bajaj Finserv - IIFL

In [1]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.outliers_influence import variance_inflation_factor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from datetime import datetime
```

In [2]:
```
# Load the data
file_path = r"C:\Users\katya\Downloads\gold_prices_data.csv"
gold_data = pd.read_csv(file_path)
```

In [3]:
```
# Convert categorical variables into dummy variables
gold_data_encoded = pd.get_dummies(gold_data, drop_first=True)
```

In [4]:
```
# Define the dependent variable and independent variables
X = gold_data_encoded.drop(columns=["Gold Price (INR/gram)"])
y = gold_data_encoded["Gold Price (INR/gram)"]

# Add a constant for the intercept
X = sm.add_constant(X)
```

In [5]:
```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_sta
```

In [7]:
```
# Fit the regression model
model = sm.OLS(y_train, X_train).fit()

# Model Summary
print(model.summary())
```

```
                                 OLS Regression Results
==============================================================================
Dep. Variable:     Gold Price (INR/gram)   R-squared:                       0.795
Model:                               OLS   Adj. R-squared:                  0.788
Method:                    Least Squares   F-statistic:                     116.3
Date:                   Fri, 22 Nov 2024   Prob (F-statistic):           7.67e-12
Time:                           18:39:13   Log-Likelihood:                -255.18
No. Observations:                     32   AIC:                             514.4
Df Residuals:                         30   BIC:                             517.3
Df Model:                              1
Covariance Type:               nonrobust
=====================================================================================
========================
                                               coef    std err          t      P>|t|
[0.025      0.975]
-------------------------------------------------------------------------------------
------------------------
Year                                       118.1534     10.956     10.784      0.000
95.777     140.529
Inflation Rate (%)                          -0.0586      0.005    -10.713      0.000
-0.070      -0.047
U.S. Dollar Index                           -0.8317      0.078    -10.713      0.000
-0.990      -0.673
Crude Oil Price (USD/barrel)                -0.7659      0.071    -10.713      0.000
-0.912      -0.620
Interest Rate (%)                           -0.0360      0.003    -10.713      0.000
-0.043      -0.029
Consumer Price Index (CPI)                  -1.0813      0.101    -10.713      0.000
-1.287      -0.875
Gold Supply (metric tons)                  -31.5379      2.944    -10.713      0.000
-37.550     -25.526
Jewelry Demand (metric tons)                -7.2087      0.673    -10.713      0.000
-8.583      -5.834
Investment Demand (metric tons)            -10.8130      1.009    -10.713      0.000
-12.874      -8.752
Central Bank Gold Reserves (metric tons)    -6.3076      0.589    -10.713      0.000
-7.510      -5.105
Exchange Rate (INR/USD)                     -0.6803      0.064    -10.713      0.000
-0.810      -0.551
Global Economic Growth Rate (%)             -0.0288      0.003    -10.713      0.000
-0.034      -0.023
Political Stability Index                   -0.0063      0.001    -10.713      0.000
-0.008      -0.005
Geopolitical Tensions Index                 -0.0072      0.001    -10.713      0.000
-0.009      -0.006
Stock Market Performance (%)                -0.0901      0.008    -10.713      0.000
-0.107      -0.073
Seasonal Demand Index                       -0.0135      0.001    -10.713      0.000
-0.016      -0.011
Import Duty on Gold (%)                     -0.0901      0.008    -10.713      0.000
-0.107      -0.073
Global Gold Production (metric tons)       -28.8346      2.692    -10.713      0.000
-34.332     -23.338
Recycled Gold Supply (metric tons)          -9.0108      0.841    -10.713      0.000
-10.729      -7.293
==============================================================================
```

```
Omnibus:                        1.415   Durbin-Watson:                   1.690
Prob(Omnibus):                  0.493   Jarque-Bera (JB):                1.278
Skew:                           0.348   Prob(JB):                        0.528
Kurtosis:                       2.312   Cond. No.                     1.37e+38
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The smallest eigenvalue is 5.1e-68. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

In [ ]:
```
# --- Interpretation of Statistical Measures ---
# 1. R-squared and Adj. R-squared:
#    - R-squared explains the proportion of variance in gold prices explained by in
#    - Adj. R-squared adjusts R-squared for the number of predictors; a better meas
# 2. F-statistic and p-value:
#    - Indicates overall model significance. A low p-value (< 0.05) confirms the mo
# 3. Coefficients:
#    - Represent the change in the dependent variable (gold price) per unit change
# 4. P-values for coefficients:
#    - Determine the statistical significance of each predictor. Predictors with p-
# 5. Durbin-Watson:
#    - Detects autocorrelation in residuals. A value near 2 is ideal; significant d
# 6. Omnibus, Jarque-Bera:
#    - Tests for normality in residuals. Significant results indicate deviations fr
```

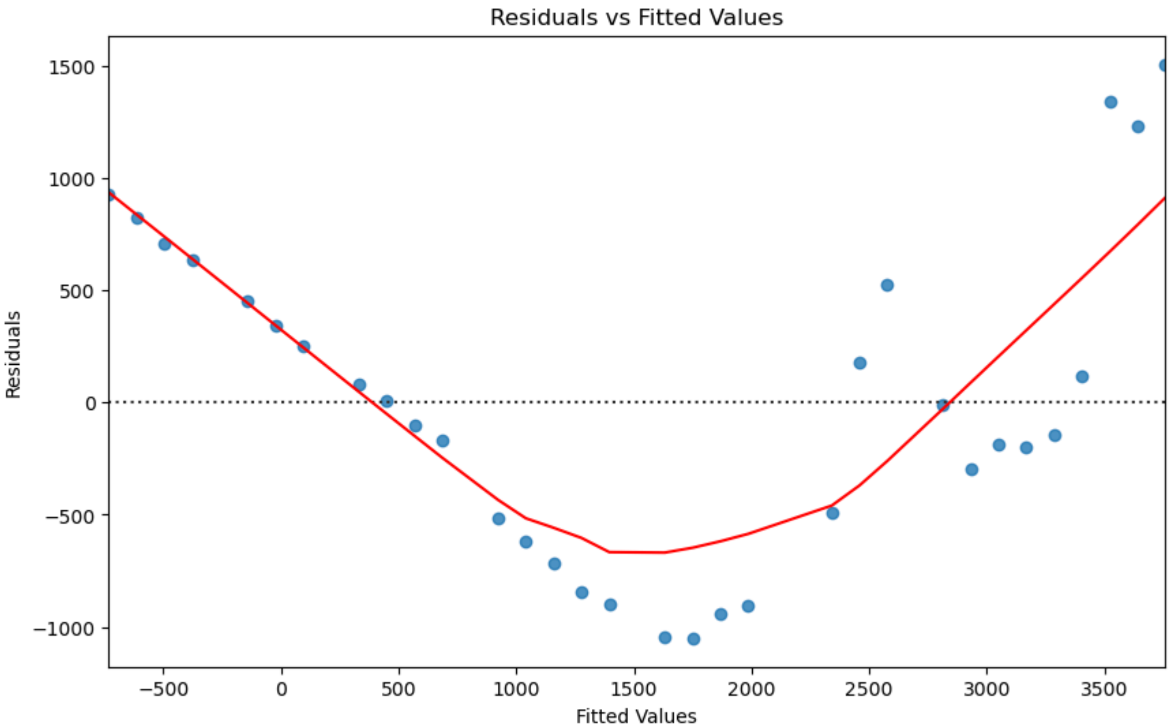
In [8]:
```
# Assumption Check: VIF for multicllinearity
vif_data = pd.DataFrame()
vif_data["Variable"] = X.columns
vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1]
print("\nVariance Inflation Factor (VIF):\n", vif_data)
```

```
Variance Inflation Factor (VIF):
                                  Variable       VIF
0                                    Year  1.000000
1                      Inflation Rate (%)  0.000000
2                       U.S. Dollar Index  0.001674
3             Crude Oil Price (USD/barrel)  0.000000
4                       Interest Rate (%)  0.000000
5               Consumer Price Index (CPI)  0.000000
6                Gold Supply (metric tons)  0.000000
7             Jewelry Demand (metric tons)  0.000000
8          Investment Demand (metric tons)  0.000000
9    Central Bank Gold Reserves (metric tons)  0.000000
10                  Exchange Rate (INR/USD)  0.000000
11         Global Economic Growth Rate (%)  0.044858
12               Political Stability Index  0.000000
13              Geopolitical Tensions Index  0.000400
14             Stock Market Performance (%)  0.000000
15                   Seasonal Demand Index  0.000000
16                 Import Duty on Gold (%)  0.000000
17      Global Gold Production (metric tons)  0.000000
18        Recycled Gold Supply (metric tons)  0.000000
```

```
C:\Users\katya\anaconda3\Lib\site-packages\statsmodels\regression\linear_model.py:17
83: RuntimeWarning: divide by zero encountered in scalar divide
  return 1 - self.ssr/self.centered_tss
```

In [9]:
```
# --- Interpretation of VIF ---
# - VIF > 10 suggests high multicollinearity, which may affect the stability of coe
```

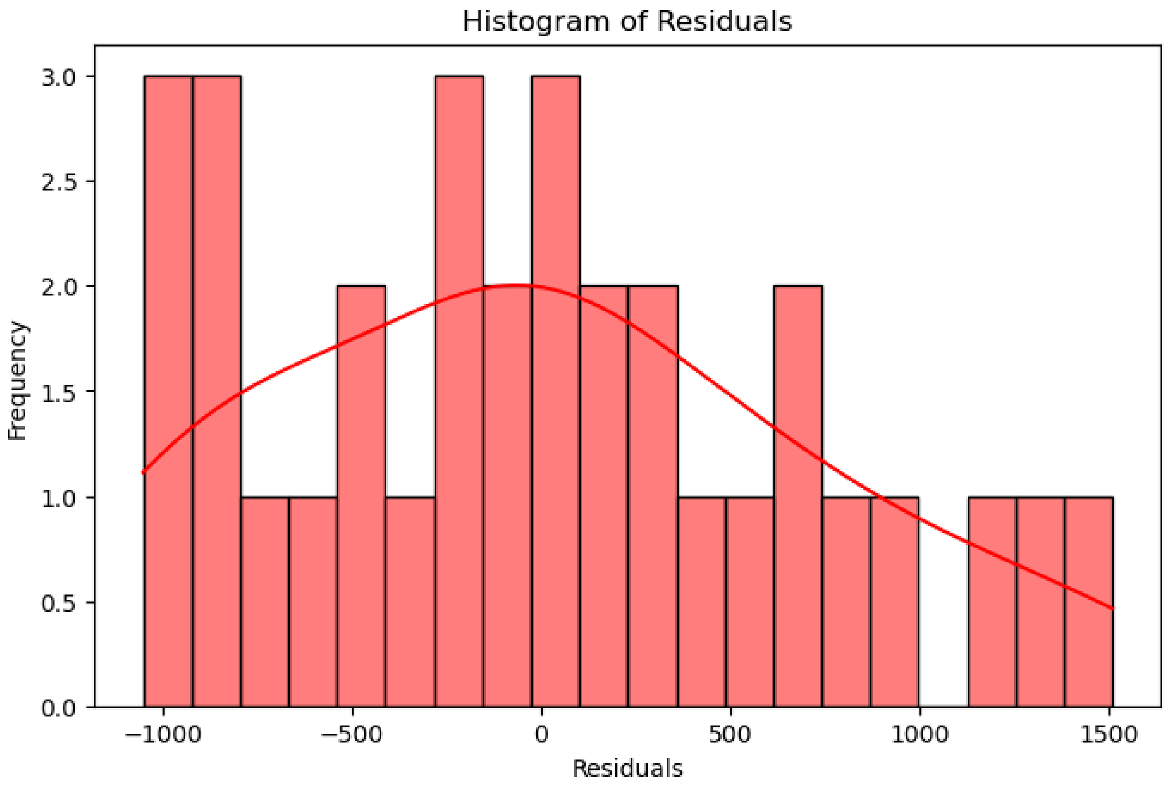
In [22]:
```
# Visualizations for assumptions
# Residuals vs Fitted Values Plot
fitted_values = model.fittedvalues
residuals = model.resid
plt.figure(figsize=(10, 6))
sns.residplot(x=fitted_values, y=residuals, lowess=True, line_kws={'color': 'red',
plt.title("Residuals vs Fitted Values")
plt.xlabel("Fitted Values")
plt.ylabel("Residuals")
plt.show()
```



In [ ]:
```
# --- Interpretation of Residuals vs Fitted Plot ---
# - The residuals should be randomly scattered around zero. Patterns indicate issue
```

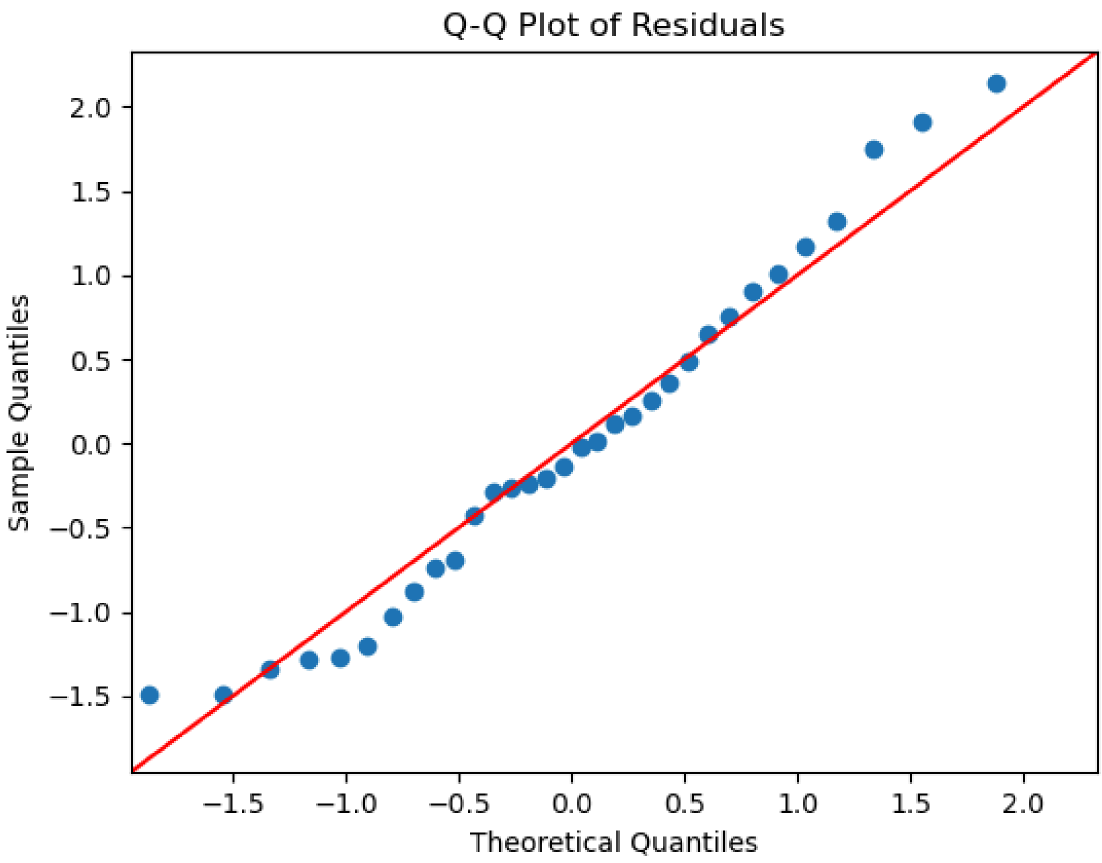
In [13]:
```
# Histogram of Residuals
plt.figure(figsize=(8, 5))
sns.histplot(residuals, kde=True, color='red', bins=20)
plt.title("Histogram of Residuals")
plt.xlabel("Residuals")
plt.ylabel("Frequency")
plt.show()

# --- Interpretation of Residual Histogram ---
# - Residuals should follow a normal distribution. Deviations suggest violations of
```



In [14]:
```
# Q-Q Plot
sm.qqplot(residuals, line="45", fit=True)
plt.title("Q-Q Plot of Residuals")
plt.show()

# --- Interpretation of Q-Q Plot ---
# - Residuals should lie close to the 45-degree line. Significant deviations sugges
```



In [15]:
```
# Predictions on test set
y_pred = model.predict(X_test)

# Calculate RMSE
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print("Root Mean Squared Error (RMSE):", rmse)

# --- Interpretation of RMSE ---
# - Measures the average error between predicted and actual values. Lower RMSE indi
```

```
Root Mean Squared Error (RMSE): 1842.9626357794875
```

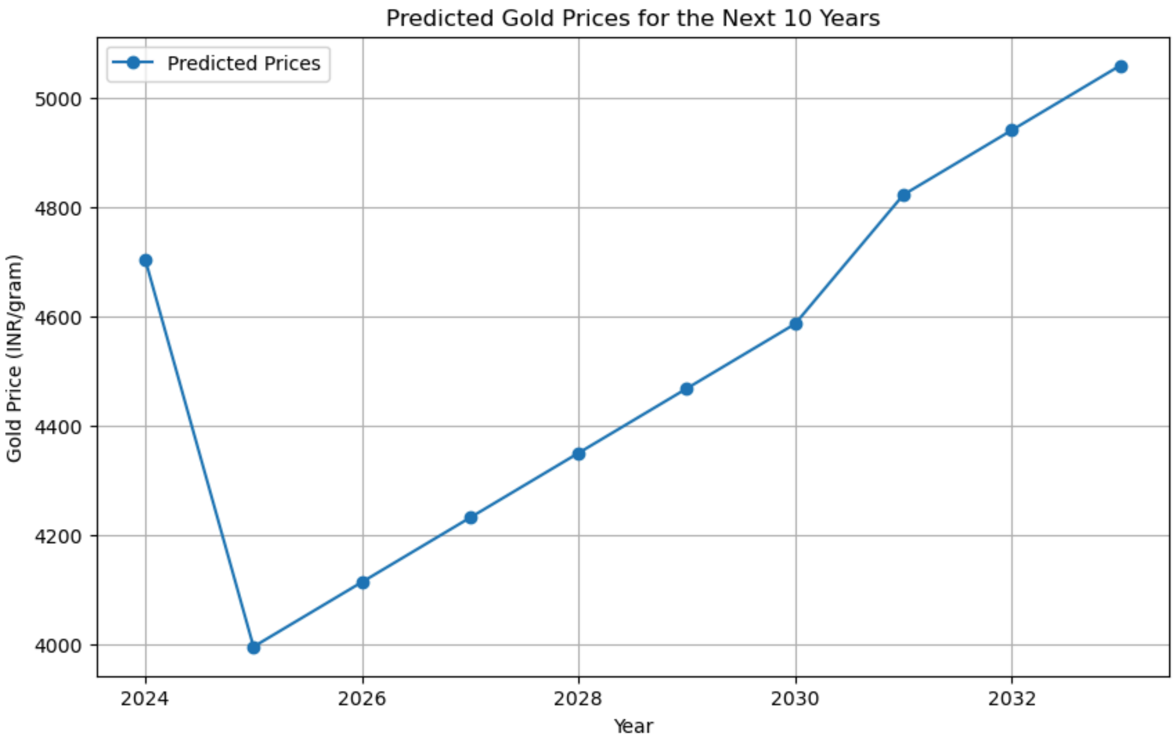
In [16]:
```
# Predict future gold prices for the next 10 years
current_year = datetime.now().year
future_years = pd.DataFrame({"Year": range(current_year, current_year + 10)})
future_data = X_train.iloc[0:1].copy()
```

In [27]:
```
# Replace "Year" in dummy variables and keep all other variables constant
for i in range(10):
    future_data.loc[i] = X_train.iloc[0]
    future_data.at[i, "Year"] = future_years.at[i, "Year"]

future_predictions = model.predict(future_data)
future_years["Predicted Price (INR/gram)"] = future_predictions

# Add percentage changes
future_years["Annual % Change"] = future_years["Predicted Price (INR/gram)"].pct_ch
```

In [19]:
```
# Visualizing Predictions
plt.figure(figsize=(10, 6))
plt.plot(future_years["Year"], future_predictions, marker='o', label="Predicted Pri
plt.title("Predicted Gold Prices for the Next 10 Years")
plt.xlabel("Year")
plt.ylabel("Gold Price (INR/gram)")
plt.grid(True)
plt.legend()
plt.show()
# --- Interpretation of Prediction Plot ---
# - Shows the projected trend in gold prices over the next 10 years based on curren
```



In [20]:
```
# Final Recommendation
if future_predictions.pct_change().mean() > 0:
    print("Recommendation: Investing in gold today is advisable as prices are expec
else:
    print("Recommendation: Investing in gold today may not be advisable as prices a
```

```
Recommendation: Investing in gold today is advisable as prices are expected to incre
ase.
```

In [30]:
```
# Trend Analysis and Recommendation
avg_annual_increase = future_years["Annual % Change"].mean()
if avg_annual_increase > 0:
    recommendation = "Investing in gold today is advisable as prices are expected t
else:
    recommendation = "Investing in gold today may not be advisable as prices are ex

print("\n--- Research Summary ---")
print(f"1. The average annual percentage increase in gold prices over the next 10 y
print(f"2. The RMSE for the model is {rmse:.2f}, indicating the model's prediction
print(f"3. Based on the predictions, {recommendation}")
```

```
--- Research Summary ---
1. The average annual percentage increase in gold prices over the next 10 years is 
2.66%.
2. The RMSE for the model is 1842.96, indicating the model's prediction accuracy.
3. Based on the predictions, Investing in gold today is advisable as prices are expe
cted to increase consistently.
```