STRATHCLYDE BUSINESS SCHOOL

DEPARTMENT OF MANAGEMENT SCIENCE

CASE STUDY REPORT

Predicting Future Prices of Gold

MS930 – Becoming an Effective Business Analyst

Group 4

Student names	**Registration numbers**
Keshav Krishnan	202358313
Kun-Teh Chou	202389756
Nguyen Minh Chau	202357159
Sarin Pornkasem	202385425

Section 1. Variable selection and Data preparation

1.1 Variable selection

In our pursuit to construct a predictive model for gold prices, the following factors have been identified as key features in our analysis that can exhibit potential impact on gold prices:

(1) Crude Oil Price: The price of crude oil is a crucial economic variable. The energy-intensive nature of gold mining and its impact on production costs suggest a plausible correlation with gold prices.

(2) Futures Silver Prices: As a precious metal often associated with gold in investment portfolios, tracking silver futures provides insights into broader market sentiments and demand for gold.

(3) Futures Copper Prices: Reflecting industrial demand and economic activities, copper prices can signal shifts in market dynamics that may influence gold prices.

(4) S&P 500 Index: The S&P 500 is a key equity index, widely considered as an indicator of overall US stock market. When the stock market faces uncertainty or downturns, investors often seek safe-haven assets like gold to preserve wealth, which could impact the prices.

(5) US Dollar Index: Given that gold is globally priced in U.S. dollars, fluctuations in the US Dollar Index can impact the attractiveness of gold as an investment.

(6) USD-CNY Exchange Rate: Given China's role in the global gold market, changes in this exchange rate can indicate shifts in trade relationships, influencing the demand and prices for gold.

(7) VIX (Volatility) Index: The VIX Index, representing market volatility, can be indicative of investor sentiment. Elevated volatility often prompts a shift towards safe-haven assets like gold, making the VIX a potential leading indicator for gold price movements.

1.2 Data source

The data used is taken from official and reliable websites, particularly LBMA Gold Prices (PM) are from World Gold Council; Brent Crude Oil Prices are from U.S. Energy Information Administration (EIA); Futures Silver and Copper Prices, S&P 500 Index, US Dollar Index, and USD-CNY Exchange Rate are from Investing UK (uk.investing.com); VIX Index (Close prices) is from Chicago Board Options Exchange (CBOE).

The data ranges from 02/01/ 2004 to 08/01/ 2024, totalling 5,258 days.

1.3 Data Handling/Cleaning

Our datasets were merged, aligned together using shared dates. From here there were many missing values. As these were so numerous, we used imputation to fill in the missing values with a rolling median; datapoints were generated based on the median of the values around them.

Section 2. Key findings

The key findings are presented based on three analytics questions.

Q1: What is correlated to this metal's future price – & ideally what is not correlated?

Pearson Correlation was utilised to find out the relationship between variables. The result coefficients are displayed in the table below.

	Oil Price	S&P 500	Futures Silver Prices	Futures Copper Prices	US Dollar Index	USD-CNY Exchange Rate	VIX Index
Gold Price	.379**	.715**	.800**	.675**	.366**	-.694*	.135**

** Correlation is significant at 0.01 level (2-tailed)
* Correlation is significant at 0.05 level (2-tailed)

Apparently, all variables correlate with Gold Prices; however, only **S&P 500 Index** (0.715), **Futures Silver Prices** (0.8), **Futures Copper Prices** (0.675) and **USD-CNY Exchange Rate** (-0.694) have a strong correlation with Gold Prices. Therefore, only these four factors are included in the prediction model.

Q2: How to predict whether prices will go 'up or down' next month – & how accurate is that prediction?

A linear regression model was trained by fitting the explanatory variable data to the existing historical gold prices. Subsequently, this trained model was applied to generate daily predictions for gold prices. To provide a more accurate prediction, the formula considers the predictors above-mentioned and the moving average of historical gold prices, specifically 3-point and 9-point moving averages of gold prices, denoted as MA(3) and MA(9) respectively. Subsequently, the prediction of gold price is not just affected by the external data but also by its own past trends.

Based on the linear regression model result, the following forecasting formula was generated:

$$\textbf{Gold Price} = 17.45 + 1.19 * \text{MA}(3) - 0.20 * \text{MA}(9) + 0.001 * \text{S\&P 500 Index}$$

$$+0.53 * \text{Futures Silver} - 0.67 * \text{Futures Copper} - 1.98 * \text{USD CNY Exchange Rate}$$

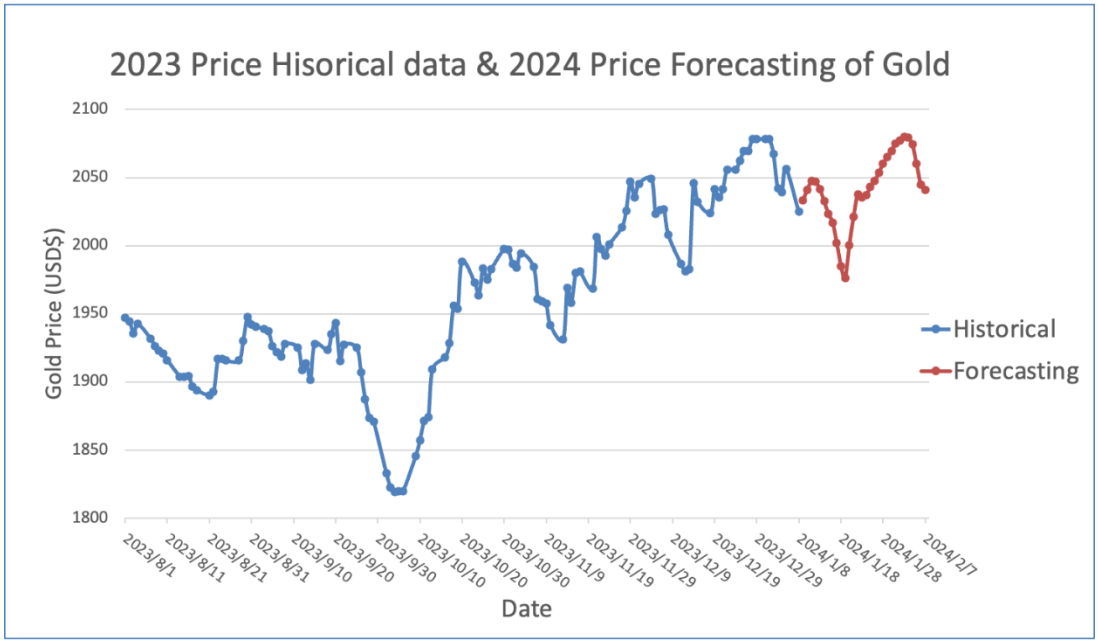
This predictor was adapted to output either "up" or "down" to suit Foresight's specifications. The predicted price for 8 February 2024 is as below:

Date	08/02/2024
Signal	Up
Predicted Gold Price	2040.857831

To find out how accurate this method is, each day of October to December 2023 was predicted and compared with the actual historical data. From this, we found that the model presented a 55% accuracy in correctly predicting the trend of 30 days after the future gold price. Although the accuracy still has room to be improved, the gold price was fluctuating repeatedly every day, and the variation was small, which means that every little change will affect its actual trend. And also, within the limited time of this project, this model puts more focus on the future actual price, which will be mentioned more in the next section.

Q3: What formula/algorithm is used to predict actual prices next month – & how accurate is that formula?

This same linear regression model and forecasting formula were used to predict the gold price for the next 30 days, which is presented in the graph below.



To estimate the expected error in our model, we attempted to predict the entirety of the last month. The root mean absolute error (RMSE) was $65.38 per troy ounce, indicating that the predictions, on average, were $65.38 away from the actual values. Additionally, the mean absolute percentage error (MAPE) was 2.72%, showing the average percentage by which the predictions deviated from the actual gold prices.

Overall, the forecast for the next 30 days indicates a positive trend, foreseeing a notable drop to $1976.16 per troy ounce in the second week of the prediction period. This reflects the inherent nature of gold prices, characterised by significant daily fluctuations but an overall upward trend.

References

QuantInsti. (2021). *Gold Price Prediction Using Machine Learning In Python*. Available at: (Accessed 11 January 2024)

Appendix

1. Glossary

1.1 Multiple Linear Regression: a statistical modeling technique used to analyse the relationship between a dependent variable and two or more independent variables. The model assumes a linear relationship, meaning it seeks to find the best-fitting linear equation to predict the dependent variable based on the given independent variables.

1.2 Root Mean Square Error (RMSE): measures the average difference between the predicted values and the actual values. It is calculated by taking the square root of the mean of the squared differences between predicted and actual values. It penalises larger errors more heavily, which means it is sensitive to outliers and reflects accuracy better. Additionally, RMSE is in the same units as the target variable (gold prices), allowing for a more intuitive interpretation of forecasting errors.

1.3 Mean Absolute Percentage Error (MAPE): measures the average percentage difference between predicted and actual values, providing insights into relative performance of a model.

2. Correlation heat map