How to predict the Gold Price

Predictive Data Analysis

Gold, as the most valuable metal, is also a very stable financial investment product for people to avoid economic risks while maintaining the absolute value of their possessions. To have a better economic return from this investment, the time to buy the Gold is crucial. A better prediction method of Gold Price is therefore very critical. As many financial institutes and websites claimed, the Gold Price has certain correlation with other economic data, oil prices, silver, and other mining stock. On *Sunshine Profits* (www.sunshineprofits.com/gold-silver/dictionary/gold-sp/), authority tools for effective gold&silver investment declares the debatable relationship between the S&P 500 Index and Gold Price. On *Macrotrends* (www.macrotrends.net/1334/gold-prices-vs-oil-prices-historical-correlation), there is an article discussing *Gold Prices vs. Oil Prices - Historical Relationship.*

Therefore, I get interested in finding out whether there is a correlation between the Gold Price and any other index and which correlation can help to make the best prediction. I chose a dataset contained comprehensive Gold Price, S&P 500 Index, Silver Price, US Oil Price, and EUR/USD from January 2008 to May 2018 on a daily base. I downloaded this dataset from Kaggle, an open datasets website, and it was uploaded by Debdatta Chatterjee 7 months ago. The limitation of this dataset is that the original source of the data of the variables are not mentioned, which weakens its validity.

First of all, I run the code of "*ggplot*" to show all data points in the chart of Gold Price (GLD) VS. US Oil Price (USO), regarding the US Oil Price as the independent variable and the Gold Price as the dependent variable. Then I

draw the best fit line to see whether there is a proper one to indicate their negative correlation (Figure. 1). However, look at the points located on the chart, the Gold Price is lowest when "USO =45", and increases in either direction from it. This means the Gold Price increases as the US Oil Price decreases when "USO<45", and vice versa when "USO<45". As the best fit line shows, the correlation between these two variables is slightly negative, and the adjusted R-squared is 0.03431. Then I try to predict a Gold price with this best fit line, and "USO=30" leads to a correlated Gold Price as 121. I run "*broom*" to get the standard error, 0.480. When "USO=60", the predictive Gold Price is 116, with the standard error 0.839. Looked at the chart, it's also clearly showed that the predicted Gold Price is higher than the actual Gold Price at most times and have the largest standard error during "USO=40-60". These results show the two variable is weakly correlated.

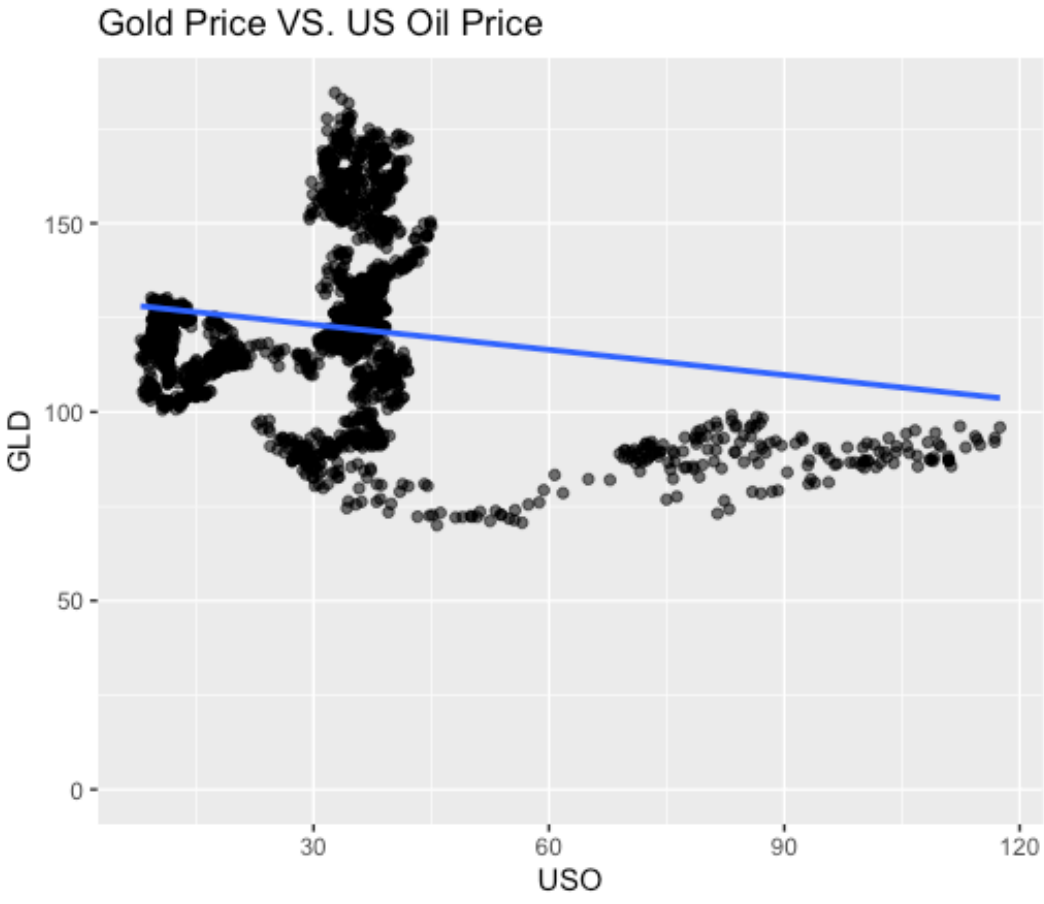

Figure. 1

With the same procedure, I got he chart of Gold Price VS. S&P 500 Index (SPX), while the S&P 500 Index is the independent variable, and the Gold Price is the dependent variable. Then I draw the best fit line to see whether there is a proper one to indicate their slightly positive correlation (Figure. 2), which means the Gold Price increases as the S&P 500 Index increases. The adjusted R-squared of it is 0.001999. Next, I use "SPX=1500" to get a predicted Gold Price, which is 122, while the standard error is 0.507. I tried "SPX=2500" to get a predicted Gold Price, 125, and the standard error is 0.929.

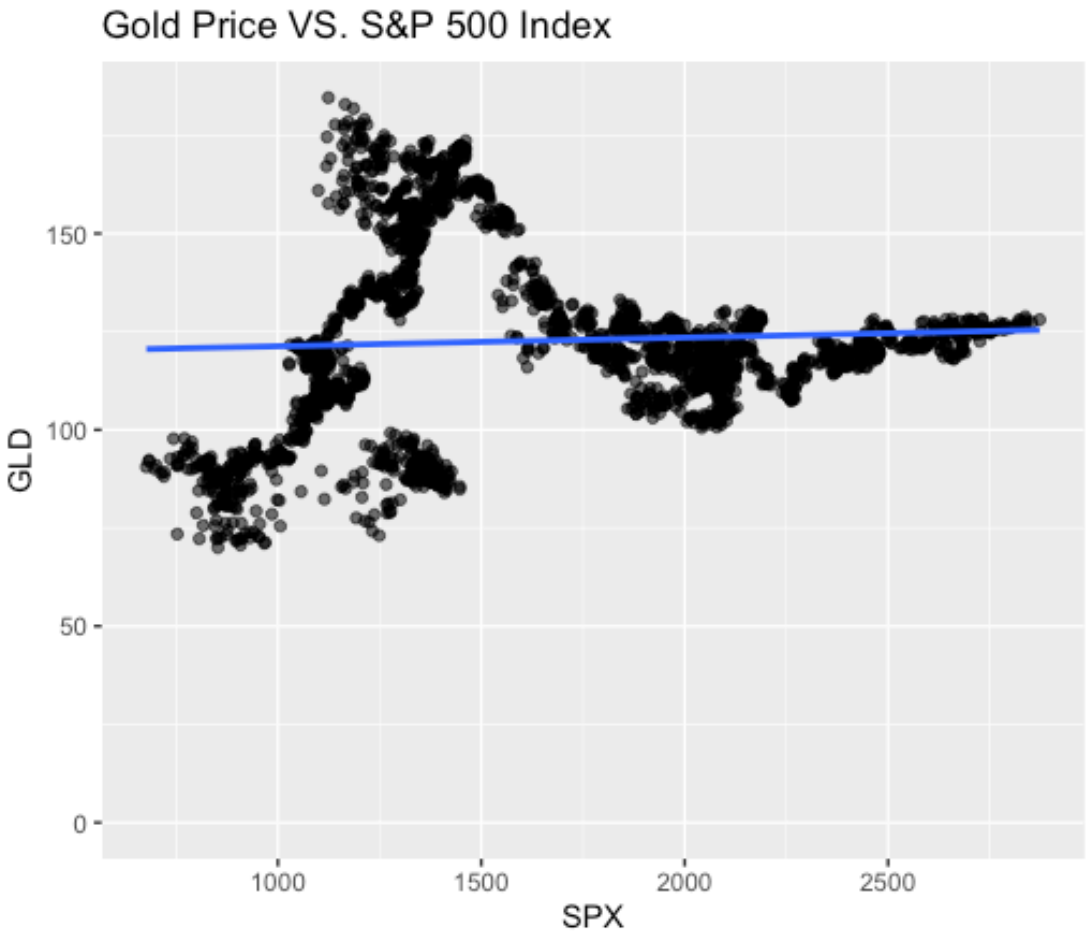

Figure. 2

At last, I want to find out the relationship between the Gold Price and the Silver Price (SLV). I got the chart of Gold Price VS. Silver Price, while the Silver Price is the independent variable and the Gold Price is the dependent variable. The best fit line clearly shows their positive correlation (Figure. 3), which means the Gold Price increases when the Silver Price increases. The adjusted R-squared this time is 0.7509. I use "SLV=10" to get a predicted Gold Price, which is 94, while the standard error is 0.422. Then, I tried "SLV=30" to get a predicted Gold Price, 151, and the standard error is 0.417.

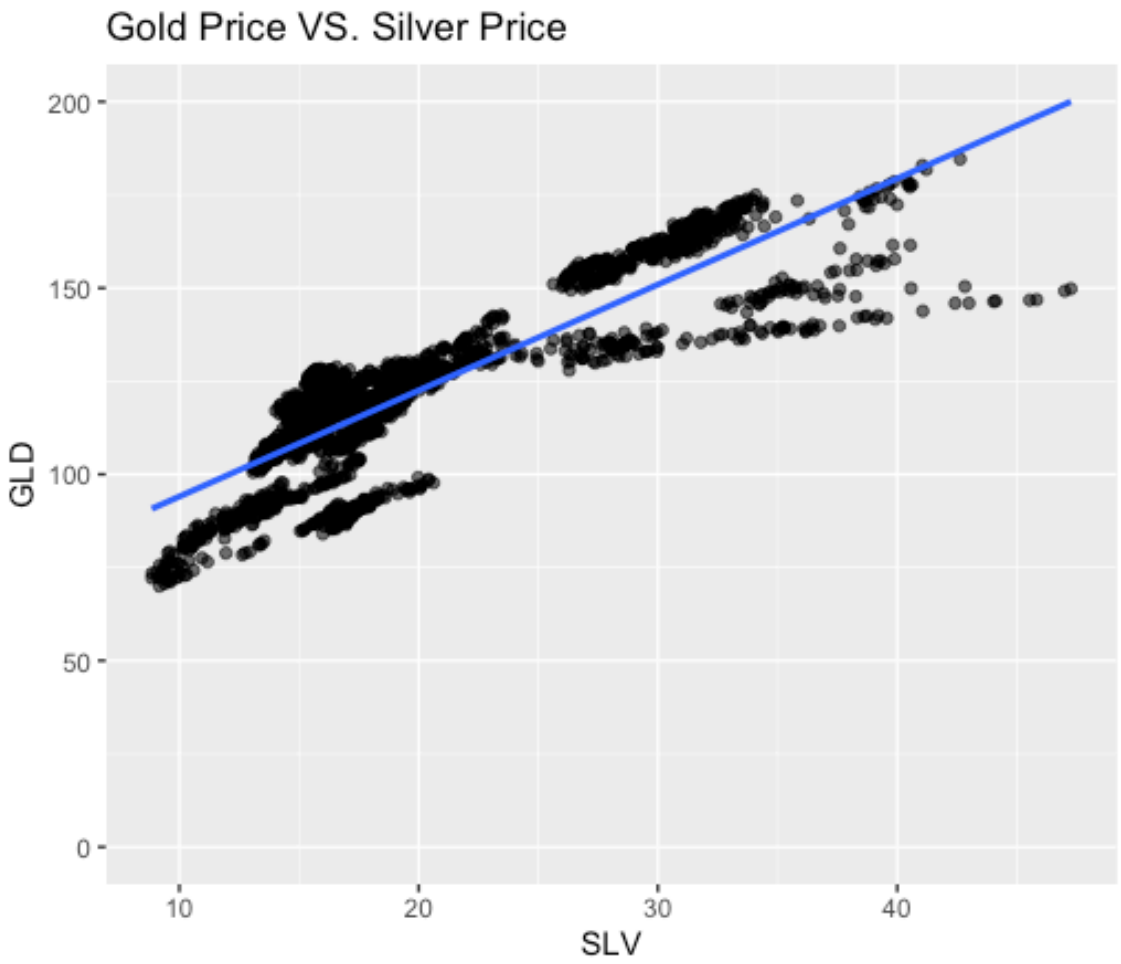

Figure. 3

Among those three correlations, the Sliver Price gives the best prediction on the Gold Price overall, with the lowest standard error and the adjusted R-squared closest to 1. The S&P 500 Index and the US Oil Price might provide the weaker prediction on the tendency of the Gold Price with higher standard errors and the adjusted R-squared close to 0.

Reference

1. Sunshine Profits, www.sunshineprofits.com/gold-silver/dictionary/gold-sp/.
2. Macrotrends, www.macrotrends.net/1334/gold-prices-vs-oil-prices-historical-correlation.

Appendix

```
library(lubridate)
library(dplyr)
library(tidyverse)
library(broom)

gld<-read.csv("Desktop/gld_price_data.csv",header=TRUE,check.names =FALSE)
head(gld)
tail(gld)
#Gold Price VS. USO
ggplot(gld,aes(x=USO, y=GLD))+  geom_jitter(alpha=.6)+  expand_limits(y=0)+
stat_smooth(method = "lm",se=FALSE)+ labs(title = "Gold Price VS. US Oil Price")
lm_gld<-lm(GLD~USO,data=gld)
summary(lm_gld)
new_gld <- data.frame("USO" = 30)
predict(lm_gld, newdata=new_gld)
mygld<- broom::augment(lm_gld, newdata=new_gld)
```

```
mygld
#Gold Price VS. S&P 500 Index
ggplot(gld,aes(x=SPX, y=GLD))+geom_jitter(alpha=.6)+expand_limits(y=0)+
 stat_smooth(method = "lm",se=FALSE)+ labs(title = "Gold Price VS. S&P 500 Index")
lm_gld<-lm(GLD~SPX,data=gld)
summary(lm_gld)
new_gld <- data.frame("SPX" = 2500)
predict(lm_gld, newdata=new_gld)
mygld<- broom::augment(lm_gld, newdata=new_gld)
mygld
#Gold Price VS. Silver Price
ggplot(gld,aes(x=SLV, y=GLD))+geom_jitter(alpha=.6)+expand_limits(y=0)+
 stat_smooth(method = "lm",se=FALSE)+ labs(title = "Gold Price VS. Silver Price")
lm_gld<-lm(GLD~SLV,data=gld)
summary(lm_gld)
new_gld <- data.frame("SLV" = 30)
predict(lm_gld, newdata=new_gld)
mygld<- broom::augment(lm_gld, newdata=new_gld)
mygld
```