Introduction

Gold is an easy marker of a number of economic factors. Gold's current performance can be an indicator of the health of the economy, if citizens are feeling optimistic or pessimistic about the current or future economic state, and more. Due to this, if the sign of gold, positive or negative, is able to be predicted then this would provide traders with a wealth of knowledge to create their trading strategies.

I will look at predicting positive signs of gold using a large dataset of financial variables. These predictors include returns from a number of exchange rates around the world, United States Treasury bonds, and stock market indexes. In addition, the lagged returns, or returns from the previous week, for all of these variables are also included. If we can successfully predict the direction of gold changes using any of these, then we will also have the ability to trade securities with great profitability.

To try and predict positive changes in gold prices, I will apply a linear nominal logistic model. After that, I will run a non-linear random forest model. The sign of gold response variable is binary with one equaling gold is greater than zero, or positive, and zero is less than zero, or negative. Logistic models were created to predict categorical data while random forest can be used for regression or classification.

Analysis and Model Comparison

To start, I will create a holdback column for cross-validation purposes. I will split the data into 60 percent training data, 20 percent validation data, and the last 20 percent will be testing data. Then, I will run a nominal logistic model with sign of gold as the response variable. This model was created to predict binary variables such as this one. While this linear model is easier to interpret than non-linear models, it is unable to model predictors that are highly complex or non-linear.

Due to this, I will then run a second model called random, or bootstrap, forest with random seed 123. This model can handle both categorical and continuous data. It will add many layers of chance by randomly choosing a set of variables as well as a subset of the data to create uncorrelated trees. This is different from a normal decision tree which looks at the full dataset and has correlated trees. These added layers of randomness eliminate the disadvantages to decision trees, except that the results are harder to interpret than linear models because there are no parameter estimates.

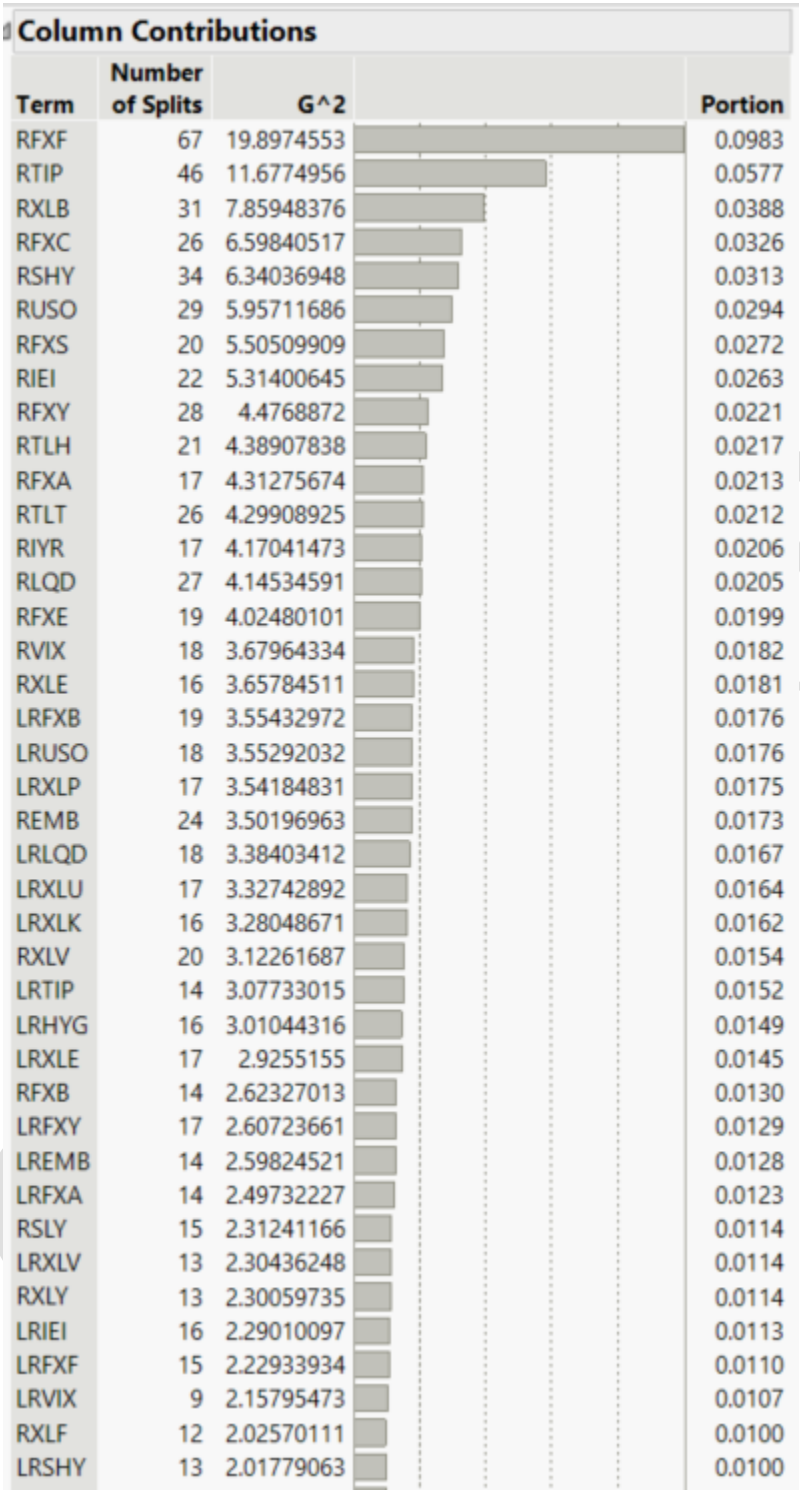
This algorithm-based ensemble model will take two parameters, the number of variables in the random subset of data at each node and the maximum number of trees in the forest, and use them to estimate the validation data. The model will stop building when the validation data and the r squared value reach their maximum values. After this is complete, the random forest will take an average of all the predictions.

**Measures of Fit for Sign of RGLD**

| Holdback | Creator | .2 .4 .6 .8 | Entropy RSquare | Generalized RSquare | Mean -Log p | RMSE | Mean Abs Dev | Misclassification Rate | N | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Fit Nominal Logistic | | 0.3904 | 0.5556 | 0.4181 | 0.3635 | 0.2674 | 0.1792 | 307 | 0.8899 |
| 0 | Bootstrap Forest | | 0.5629 | 0.7209 | 0.3 | 0.2738 | 0.2503 | 0.0292 | 308 | 0.9988 |
| 1 | Fit Nominal Logistic | | -0.028 | -0.052 | 0.7122 | 0.4829 | 0.3791 | 0.3077 | 104 | 0.7174 |
| 1 | Bootstrap Forest | | 0.2176 | 0.3472 | 0.5422 | 0.4258 | 0.3898 | 0.2500 | 104 | 0.8128 |
| 2 | Fit Nominal Logistic | | 0.0112 | 0.0205 | 0.685 | 0.4758 | 0.3637 | 0.3398 | 103 | 0.7426 |
| 2 | Bootstrap Forest | | 0.1911 | 0.3102 | 0.5604 | 0.4358 | 0.4045 | 0.3107 | 103 | 0.8057 |

Comparing the two models conducted on the positive sign of gold, the chosen model is the bootstrap forest. This model has higher r squared values, a lower misclassification rate, and a higher AUC. This means that this random forest model is able to correctly classify more than the linear nominal logistic model. The bootstrap forest obtained these better results and stopped building once it reached 39 trees.

Interpretation



Column Contributions

| Term | Number of Splits | G^2 | Portion |
|---|---|---|---|
| RFXF | 67 | 19.8974553 | 0.0983 |
| RTIP | 46 | 11.6774956 | 0.0577 |
| RXLB | 31 | 7.85948376 | 0.0388 |
| RFXC | 26 | 6.59840517 | 0.0326 |
| RSHY | 34 | 6.34036948 | 0.0313 |
| RUSO | 29 | 5.95711686 | 0.0294 |
| RFXS | 20 | 5.50509909 | 0.0272 |
| RIEI | 22 | 5.31400645 | 0.0263 |
| RFXY | 28 | 4.4768872 | 0.0221 |
| RTLH | 21 | 4.38907838 | 0.0217 |
| RFXA | 17 | 4.31275674 | 0.0213 |
| RTLT | 26 | 4.29908925 | 0.0212 |
| RIYR | 17 | 4.17041473 | 0.0206 |
| RLQD | 27 | 4.14534591 | 0.0205 |
| RFXE | 19 | 4.02480101 | 0.0199 |
| RVIX | 18 | 3.67964334 | 0.0182 |
| RXLE | 16 | 3.65784511 | 0.0181 |
| LRFXB | 19 | 3.55432972 | 0.0176 |
| LRUSO | 18 | 3.55292032 | 0.0176 |
| LRXLP | 17 | 3.54184831 | 0.0175 |
| REMB | 24 | 3.50196963 | 0.0173 |
| LRLQD | 18 | 3.38403412 | 0.0167 |
| LRXLU | 17 | 3.32742892 | 0.0164 |
| LRXLK | 16 | 3.28048671 | 0.0162 |
| RXLV | 20 | 3.12261687 | 0.0154 |
| LRTIP | 14 | 3.07733015 | 0.0152 |
| LRHYG | 16 | 3.01044316 | 0.0149 |
| LRXLE | 17 | 2.9255155 | 0.0145 |
| RFXB | 14 | 2.62327013 | 0.0130 |
| LRFXY | 17 | 2.60723661 | 0.0129 |
| LREMB | 14 | 2.59824521 | 0.0128 |
| LRFXA | 14 | 2.49732227 | 0.0123 |
| RSLY | 15 | 2.31241166 | 0.0114 |
| LRXLV | 13 | 2.30436248 | 0.0114 |
| RXLY | 13 | 2.30059735 | 0.0114 |
| LRIEI | 16 | 2.29010097 | 0.0113 |
| LRFXF | 15 | 2.22933934 | 0.0110 |
| LRVIX | 9 | 2.15795473 | 0.0107 |
| RXLF | 12 | 2.02570111 | 0.0100 |
| LRSHY | 13 | 2.01779063 | 0.0100 |

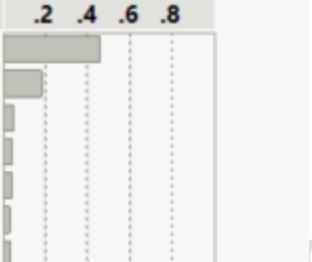
Using JMP's column contributions, which was specifically created for the bootstrap forest, the most important variable in this model is RFXF, or the return on the exchange rate for the Swiss franc. This contributes to ten percent of the changes in the sign of gold. Next is inflation with six

percent, then the material sector index with four percent. After that is the Canadian dollar, one to three-year U.S. Treasury bonds, oil prices, the Swedish krona, and three to seven-year U.S. Treasury bonds all contributing to the fluctuations by three percent each. These top eight most important variables together influence changes in the sign of gold by 35 percent.

**Variable Importance: Independent Uniform Inputs**

**Summary Report**

| Column | Main Effect | Total Effect |
|---|---|---|
| RFXF | 0.39 | 0.457 |
| RTIP | 0.135 | 0.189 |
| RIEI | 0.029 | 0.053 |
| RIYR | 0.023 | 0.046 |
| RUSO | 0.017 | 0.04 |
| RFXS | 0.015 | 0.036 |
| RFXC | 0.014 | 0.036 |
| RFXE | 0.014 | 0.034 |
| RXLB | 0.011 | 0.029 |
| RFXA | 0.011 | 0.025 |
| RSHY | 0.01 | 0.023 |
| LRUSO | 0.009 | 0.021 |
| RLQD | 0.009 | 0.02 |
| RTLH | 0.009 | 0.019 |
| RTLT | 0.007 | 0.018 |
| LRXLP | 0.006 | 0.017 |
| LRXLU | 0.008 | 0.017 |
| LRVIX | 0.006 | 0.014 |
| RXLV | 0.005 | 0.012 |
| RFXY | 0.005 | 0.011 |
| LREMB | 0.004 | 0.011 |
| LRXLK | 0.004 | 0.011 |
| LRFXB | 0.005 | 0.011 |
| LRHYG | 0.004 | 0.011 |
| RXLE | 0.004 | 0.011 |
| LRXLV | 0.004 | 0.01 |
| RXLF | 0.004 | 0.01 |
| RXLK | 0.004 | 0.01 |
| RFXB | 0.005 | 0.01 |

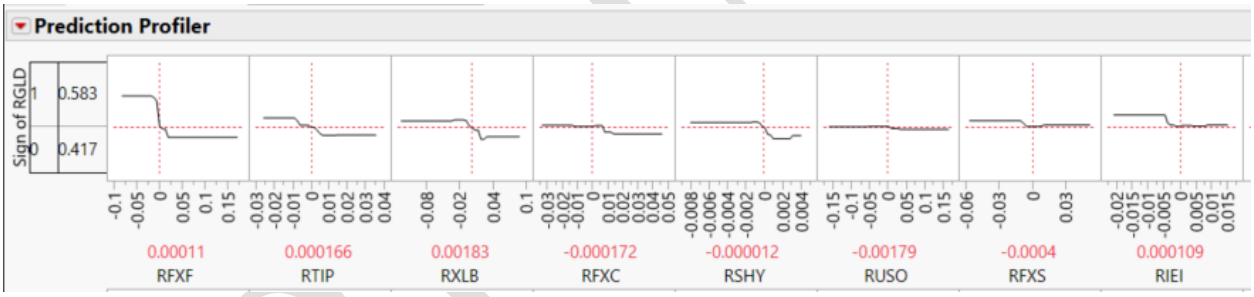



`

The variable importance feature in JMP paints a different picture than the column contributions specifically made for random forest. While the Swiss franc and inflation are still the top two most important variables, they contribute to a greater percentage of the fluctuations in the sign of gold. The Swiss franc explains 46 percent of the changes, or 22 percent more, while inflation explains three percent more. After these two variables, both agree on the importance of four of the next six variables, but not in the same order. Both agree that three to seven-year U.S

treasury bonds, oil prices, Swedish krona, and Canadian dollars are important variables to predict the changes in the sign of gold. Unlike column contributions, this variable importance feature adds in the real estate sector index and the European Union euro in the top eight most important predictors over the material sector index and one to three-year U.S. Treasury bonds. These top eight combined make up 90 percent of the fluctuations in the sign of gold versus 35 percent from the column contributions top eight.

Returns on gold predicted by a linear stepwise model had three important variables. The Swiss franc was the most important explaining 60 percent of fluctuations in the returns. Inflation was second with 20 percent and the Canadian dollar exchange rate was third with 19 percent. All three of these named the return on the Swiss franc exchange rate as the most important variable to predict returns on gold or the sign of gold. Swiss franc's ability to explain gold fluctuations ranged from ten percent to 60, but it was number one across the board. Inflation was also second place in all three ranging from six to 20 percent in its capability to predict changes in gold returns or the sign of gold.




The positive sign of gold has a positive correlation with the Swiss franc, inflation, the material sector index, and three to seven-year U.S. Treasury bonds. The Swiss franc and treasury bonds are similar to gold in that they are seen as safe havens in uncertain economic times. Therefore, it makes sense that all three would move in the same direction as each other. Inflation works hand-in-hand with these economic factors. When inflation rises, not only does paper money become less valuable increasing gold's value, but also the Federal Reserve will put a stop to the rising inflation that is causing risk and uncertainty in the economy. They will do this by raising interest rates, usually a significant amount, which is bad news for the stock market. The material sector index covers acquiring, developing, and processing raw materials, such as gold. If the stock index for this industry is doing well, then the sign of gold is also going in a positive direction.

The positive sign of gold has a negative correlation with four of the top eight predictors. If oil prices, one to three-year U.S Treasury bonds, the return on the Canadian dollar and the Swedish krona exchange rates are negative then the sign of gold will be positive. Decreasing oil prices is a sign of a bad economy. If the economy is in a depression, citizens are unable to afford cars, trips, or other various items that need oil and with less demand, then oil prices drop. Therefore, when oil prices drop, the sign of gold moves in a positive direction because it's a mark of a bad economy which drives investments in safe havens like gold.

The relationship between the sign of gold and one to three-year U.S. Treasury bonds is similar to its relationship with oil. While both bonds and gold are seen as safe havens to go to during indicators of a recession, short time frame bonds such as these of a one to three-year time span would go up if investors felt the economy would be making a quick turnaround. These investors would want that money back when the economy improves. For this reason, the positive sign of gold is more likely when three to seven-year bonds increase and one to three-year bonds decrease because investors feel they need to ride out a long-term recession in safe haven investments.

Canada has one of the largest oil reserves in the world. For the same reason a decline in oil prices increase the chance of gold signs being positive, so does the exchange rate on the Canadian dollar. The Swedish krona is seen as an indicator of international trade. If the Swedish krona is going down, then there is overall uncertainty in the global economy and risk for recessions worldwide. With this doubt in the economy, investors would rush to gold, which would push its sign in a positive direction.

If the Swiss franc is going up 0.00011, inflation 0.000166, the material sector index 0.00183, and three to seven-year U.S Treasury bonds 0.000109 while the Canadian dollar is going down -0.000172, one to three-year U.S Treasury bonds -0.000012, oil -0.00179, and Swedish Krona -0.0004 then the sign of gold has a 58.3 percent probability of being positive and a 41.7 percent chance of being negative. If any of these variables increased by a larger amount, then the probability of a positive sign of gold would be even more likely.

No matter the model, or if returns on gold or the sign of gold is being predicted, returns on the Swiss franc exchange rate is the most important variable. After that, inflation is a key predictor. The Canadian dollar exchange rate appears in varying importance levels and with different correlations, but it is always a factor. Beyond these three predictors, depending on the model and

the response variable, a range of other elements can play a part in predictions. Overall, all of them play a role because of their responses to the health of the economy, which in turn drives investment decisions.