GOLD VS S&P 500:

TRENDS AND RELATIONSHIP

Project by :

Sanjay Man Tamrakar
STA 502
Department of Statistics
Miami University

Submitted to :

Dr. Stephen Wright
Department of Statistics
Miami University

Introduction

Stock markets are pretty much unpredictable and complicated with its voluminous amount of data. In order to understand the behavior and tendencies of the market, it is imperative that we have a look at the historical data which helps us in analyzing the past for forecasting the future trends. Before we make any efforts at analyzing the data, we need to first retrieve the necessary raw data and then clean, extract, transform, and combine them as per the requirements of our analysis. After the necessary data are extracted from the various sources, we can perform various statistical analysis to provide concrete and significant information. In this project, we explore two essential parts of the investment market, namely, Gold futures and S&P 500, and perform statistical analysis to look at the trends and relationship between them with the help of Statistical Analysis Software (SAS).

According to Investopedia, "Futures are financial contracts obligating the buyer to purchase an asset or the seller to sell an asset, such as a physical commodity or a financial instrument, at a predetermined future date and price."[3] Gold futures are the instruments for diversifying the risks associated with the gold price fluctuations. Gold is considered as a safe investment by a large number of people as it helps in hedging the risk, but the physical possession of gold is expensive as additional costs such as transport,storage protection and trade adds up. So, we need contracts and deriatives in order to make our lives easier and reduce costs. Gold futures are actively traded every single day, hence they are more accurate and provides a lot of information about gold, which also helps to predict the spot price of the gold.

According to Investopedia, "The Standard & Poor's 500 Index (S&P 500) is an index of 500 stocks seen as a leading indicator of U.S. equities and a reflection of the performance of the large cap universe, made up of companies selected by economists."[4] The S&P 500 is considered as one of the best portrayal of the US stock market. The index includes 500 leading companies and repesents approximately 80% coverage of available market capitalization. It is one of the essential benchmarks to see how the overall economy of the market is perfoming on a daily basis.

There is a perception that the gold price fluctuates negatively with respect to the S&P 500 index price. As the S&P 500 index increases, the price of gold, and consequentially, the price of Gold futures are expected to decrease in a similar fashion. Gold is used as a security; hence, when the market is not doing good, people are motivated to invest in gold so as to hedge the risk, thus increasing the price of the gold. In the same way, when the market is doing good, people tend to sell the gold in order to grab the opportunities present in the current market thus leading to the decrease in price of the gold. To observe if there are any such trends, we need to investigate the historical data for the Gold futures and S&P 500 index.

Description of Data

The data obtained for this project are the historical data for the Gold futures and S&P 500 index. Historical data are basically the past data, which are used for forcasting the future data or trends. For our analysis, the daily data from January, 2006 to December, 2015 are extracted from the website into the SAS system. The data for Gold futures is obtained from

http://www.investing.com/commodities/gold-historical-data [2]. The data are in the form of a table which contains 7 variables – Date, Price of the day, Open price of the day, Highest price of the day, Lowest price of the day, Volumes traded, and Percentage Changes. Volume traded is the number of shares exchanged in the maket in a particular day. We need to specify the time frame for which we want to perform our analysis in the box shown in Figure 1. For our time frame of 10 years, we have 2546 observations for Gold futures. The first 5 observations of Gold futures from the webpage are shown in Figure 1.

Since we want to extract the data from the webpage, we need to look at the HTML (Hyper Text Markup Language) codes by viewing the source of the website (right click and select view page source). Our first observation is "Dec 31, 2015", which is in line 1513 of the HTML file. The variables are in different lines and the values of the variables are bounded by <td> and </td> , which are the opening and closing commands for a standard cell in an HTML table. We need to extract those values between the commands into the SAS system. A portion of the HTML source file is shown in Figure 2.

Similarly, the data for S&P 500 are obtained from http://www.investing.com/indices/us-spx-500-historical-data [5]. The structure of the data are pretty much similar to that of Gold futures except that the S&P 500 data contains 2514 observations. We also noticed that we don't have volumes for the S&P 500.

There are some missing data in the tables where volumes are missing. Since no analysis were based on the volume we do not need to figure out the missing values. In other instances, where volume is the basis of our analysis, we need to figure out the missing values first, then do the inference as missing values can very much change the inference.

Methods

Data Import Process

Since the data should be extracted from the HTML source file, we need to first find out where our first observation starts from in the HTML file. We need to read the downloaded HTML file into the SAS using the INFILE statement in the DATA step. We can know the location of our first observation by first extracting all the information from the source file as a string and look for 'Dec 31, 2015', which is our first observation in the table, after which we need to use 'SYMPUT' to extract the value of the first observation and allocate it to the macro variable. Thereafter, we need another DATA step to read the file again in order to extract the data into the SAS data set as shown in Figure 3.

Data Cleaning

We can see from Figure 3 that the Date and Volume are character variables, which makes it difficult to perform numerical analysis. Therefore, we need to convert them into numerical variables. We first convert the date into a "yymmdd" format, which takes the value of SAS date format, so that we can sort them in ascending order as shown in Figure 4. In addition, we use COMPRESS to extract just the numerical part and multiply by 1000 to get the required volume.

Some of the values in the variable volume of Gold futures are missing, which may be because of very small or negligible volume. S&P 500 has the variable volume, but do not contain any values; hence, we will not be considering the volume for our analysis. Last but not the least, we need to merge the two different data sets into a single data set in SAS as shown in Figure 5 so that we can make our comparisons and deduce relevant conclusions.

Line Chart

Line chart demonstrates the information as a series of data points connected by straight line segments. We draw a line chart for the price of Gold futures and S&P 500 with respect to the date and check if we detect some trends. We further draw line chart for the price of Gold futures and the ratios of price of Gold futures and S&P 500.

Correlation

Correlation is a statistical technique that can show how strongly pairs of variables are related. One of the common measure of correlation in statistics is the Pearson Correlation. The full name is the Pearson Product Moment Correlation or PPMC. It shows the linear relationship between two variables. The value comes out to be of range -1 to 1, in which 1 shows that we have the highest positive correlation. 0 indicates that we do not have any correlation whereas -1 indicates that we have the highest negative correlation.

Ratio

The price of Gold futures to the price of S&P 500 is computed and multiplied by 1000. Multiplication by 1000 was necessary because the ratio alone would be too small to show side by side in a single chart with Gold futures and S&P 500. Since the prices of Gold futures and S&P 500 are in the range of 1000, multiplication by 1000 would allow us to view the trends in the historical data. The ratio between the prices were calculated because we could not find much relationship between Gold futures and S&P 500. We think that it should be opoosite but in the long run both seems to rise and have a positive correlation. Similarly, the price of S&P 500 to the price of Gold futures is computed. We can then draw the graphs for the ratios and see if we get some significant results.

Macros

After importing data and analysing the data, we try to make our code more useful to the users by implementing macros in the SAS code. The LET statement on the top helps us to specify the file location, extension, and filenames. If we have similar data for different stocks, indices, futures or options we can just specify the different values in the LET statements and all the line charts and correlation are calculated accordingly without having to repeat the process over and over again if the data have the same structure. Macros in the code are designed so that similar data can be easily brought into the program and no change or a very minimal change is needed in order for the code to be used. In the program, we used macros at two places – one to bring in two specific data sets without repeating the process again, and another to plot three line charts using the single PROC SGPLOT to get the desired output.

Results

From Figure 6, we observed that the price of both Gold futures and S&P 500 increased from the year 2006 to 2015. We saw that in 2009-2013, both the Gold futures and S&P 500 have increased, while from the midst of 2013, S&P 500 seemed to rise but there was a downfall of Gold futures. We also found from Table 1 that there was a very little correlation between Gold futures and S&P 500, as the correlation was found to be 0.15829. Looking at these two results, we have a conjecture that there might be a possibility of a negative correlation between Gold futures and S&P 500.

To find out more about the inverse relation, we also found the Pearson Correlation Coefficient. The correlation between price of S&P 500 and (Ratio of Gold futures to S&P 500)*1000 was found to be -0.50167 whereas the correlation between price of S&P 500 and (Ratio of S&P 500 to Gold futures)*1000 was 0.42782 as shown in Table 1. Likewise, we found out that the correlation between price of Gold futures and (Ratio of Gold futures to S&P 500)*1000 was 0.76042 whereas the correlation between Gold futures and (Ratio of S&P 500 to Gold futures)*1000 was -0.76841.

Figure 7 clearly shows the negative correlation between Gold futures and (Ratio of S&P 500 to Gold futures)*1000. Similarly, Figure 8 shows the positive correlation between Gold futures and (Ratio of Gold futures to S&P 500)*1000. From these results, we have a reason to believe that since gold is used as a security, when the market or the S&P 500 index rises, it should fall and vice-versa.

Discussion

From the line chart in Figure 6, we find that we don't have a big relationship between the price of Gold futures and the price of S&P 500 in the long run, even though we might think them as having an inverse relationship. To further investigate the case, the correlation was calculated, but it seemed to be positively correlated. Since both of these measures did not provide much information, we calculated the ratio of the prices of Gold futures and S&P 500 and found out that it was informative as compared to just comparing with the prices of each other. When we found out the ratios, we observed from the line chart that the ratios of price of Gold futures and S&P 500 are inversely related to the price of S&P 500 and directly related to the price of Gold futures. We can also see that the Pearson Coefficient increases substantially while using the ratios.

Hence, there seemed to be an inverse relationship between gold futures and S&P 500 when we look at the data for the recent 2-3 years, but in the long run, we found that both of these rise, which is because of inflation. We cannot infer from this alone that they are positively correlated. Therefore, the range of time should be wide enough to incorporate various phases of the association between the prices since we found out that the prices sometimes moved along the same way, sometimes opposite, and sometimes even parallel if we look at a certain part of the historical data. Moreover, it would be vital to look at the ratios of their prices before making inferences since we found out that the ratios were more interesting than the prices alone and gave us vital informations when we considered it in our analysis.

References

1. Bailer, A. John. Statistical Programming in SAS. Cary, NC: SAS Institute, 2010. Print.
2. "Gold Price - Investing.com." *Investing.com*. N.p., n.d. Web. 30 Nov. 2016.
3. Root."Futures." Investopedia. N.p., 2003. Web. 09 Dec. 2016.
4. Root."Standard & Poor's 500 Index - S&P 500." Investopedia. N.p., 2014. Web. 09 Dec. 2016.
5. "S&P 500 Index - Investing.com." *Investing.com*. N.p., n.d. Web. 30 Nov. 2016.
6. SAS Institute. 1999. SAS/STAT User's Guide, Version 8. Cary, NC, USA.
7. "SAS/STAT(R) 9.3 User's Guide." *SAS/STAT(R) 9.3 User's Guide*. N.p., n.d. Web. 30 Nov. 2016.

Appendices

Tables

SNP500 VS (Ratio of price of GoldFutures to SNP500)times 1000

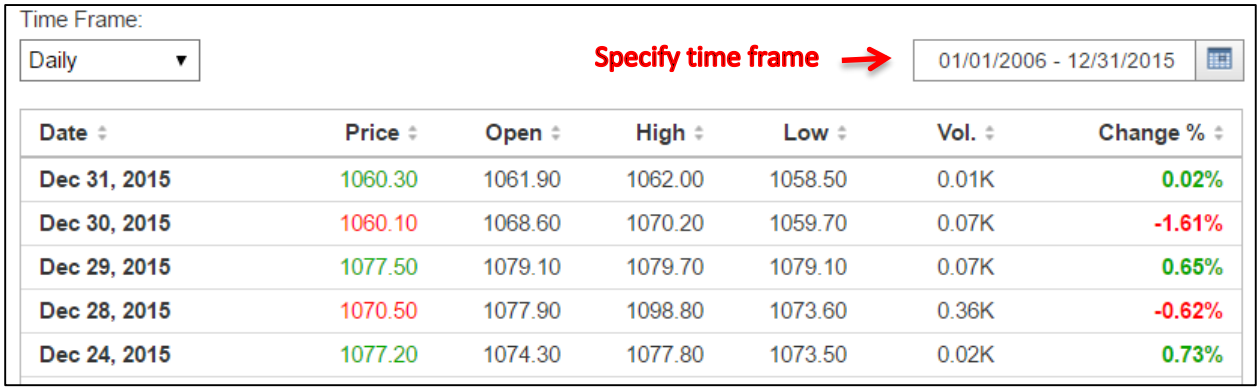
The CORR Procedure

3 With Variables:	SNP500toGoldFutures_Ratio GoldFuturestoSNP500_Ratio Price_GoldFutures
2 Variables:	Price_SNP500 Price_GoldFutures

Pearson Correlation Coefficients, N = 2514		
	Price_SNP500	**Price_GoldFutures**
SNP500toGoldFutures_Ratio (Ratio of price of SNP500 to GoldFutures)times 1000	0.42782	-0.76841
GoldFuturestoSNP500_Ratio (Ratio of price of GoldFutures to SNP500)times 1000	-0.50167	0.76042
Price_GoldFutures Price of GoldFutures ($)	0.15829	1.00000

Table 1: Pearson Correlation Coefficients between Gold Futures, S&P 500 and their Ratios

Figures



Date	Price	Open	High	Low	Vol.	Change %
Dec 31, 2015	1060.30	1061.90	1062.00	1058.50	0.01K	0.02%
Dec 30, 2015	1060.10	1068.60	1070.20	1059.70	0.07K	-1.61%
Dec 29, 2015	1077.50	1079.10	1079.70	1079.10	0.07K	0.65%
Dec 28, 2015	1070.50	1077.90	1098.80	1073.60	0.36K	-0.62%
Dec 24, 2015	1077.20	1074.30	1077.80	1073.50	0.02K	0.73%

Figure 1: Gold Historical Data from the webpage

```
1512                <tr>
1513            <td data-real-value="1451520000" class="first left bold noWrap">Dec 31, 2015</td>
1514            <td class="greenFont">1060.30</td>
1515            <td>1061.90</td>
1516            <td>1062.00</td>
1517            <td>1058.50</td>
1518            <td data-real-value="10">0.01K</td>            <td class="bold greenFont">0.02%</td>
1519        </tr>
1520                <tr>
1521            <td data-real-value="1451433600" class="first left bold noWrap">Dec 30, 2015</td>
1522            <td class="redFont">1060.10</td>
1523            <td>1068.60</td>
1524            <td>1070.20</td>
1525            <td>1059.70</td>
1526            <td data-real-value="71">0.07K</td>            <td class="bold redFont">-1.61%</td>
1527        </tr>
1528                <tr>
1529            <td data-real-value="1451347200" class="first left bold noWrap">Dec 29, 2015</td>
```

Figure 2: HTML source code sample for gold futures

VIEWTABLE: Work.Goldfutures1

	Date	Price of GoldFutures ($)	Open Price of GoldFutures ($)	High price of GoldFutures ($)	Low price of GoldFutures ($)	Volume of GoldFutures	Percentage change of GoldFutures
1	Dec 31, 2015	1060.3	1061.9	1062	1058.5	0.01K	0.02%
2	Dec 30, 2015	1060.1	1068.6	1070.2	1059.7	0.07K	-1.61%
3	Dec 29, 2015	1077.5	1079.1	1079.7	1079.1	0.07K	0.65%
4	Dec 28, 2015	1070.5	1077.9	1098.8	1073.6	0.36K	-0.62%
5	Dec 24, 2015	1077.2	1074.3	1077.8	1073.5	0.02K	0.73%

Figure 3: First 5 observations in the SAS data set

VIEWTABLE: Work.Goldfutures2

	Price of GoldFutures ($)	Open Price of GoldFutures ($)	High price of GoldFutures ($)	Low price of GoldFutures ($)	Volume of GoldFutures	Percentage change of GoldFutures	Date
1	530.7	518.6	528.5	518.6	10	2.33%	06-01-03
2	533.9	533.6	533.6	533.5	10	0.60%	06-01-04
3	526.3	529	529	526	10	-1.42%	06-01-05
4	539.7	539.7	539.7	539.7	.	2.55%	06-01-06
5	549.1	549.1	549.1	549.1	.	1.74%	06-01-09

Figure 4: SAS data set after converting volume and date in appropriate format

VIEWTABLE: Work.Combined

	Price of GoldFutures ($)	Open Price of GoldFutures ($)	High price of GoldFutures ($)	Low price of GoldFutures ($)	Volume of GoldFutures	Percentage change of GoldFutures	Date	Price of SNP500 ($)	Open Price of SNP500 ($)	High price of SNP500 ($)	Low price of SNP500 ($)	Percentage change of SNP500	(Ratio of price of GoldFutures to SNP500)times 1000	(Ratio of price of SNP500 to GoldFutures)times 1000
1	530.7	518.6	528.5	518.6	10	2.33%	06-01-03	.	.	.	.		.	.
2	533.9	533.6	533.6	533.5	10	0.60%	06-01-04	1273	1269	1275	1268	0.37%	419.3	2385.2
3	526.3	529	529	526	10	-1.42%	06-01-05	1273	1273	1277	1270	0.00%	413.3	2419.7
4	539.7	539.7	539.7	539.7	.	2.55%	06-01-06	1285	1273	1286	1273	0.94%	419.9	2381.8
5	549.1	549.1	549.1	549.1	.	1.74%	06-01-09	1290	1285	1291	1285	0.37%	425.6	2349.6

Figure 5: Combined data set of Gold Futures and S&P 500

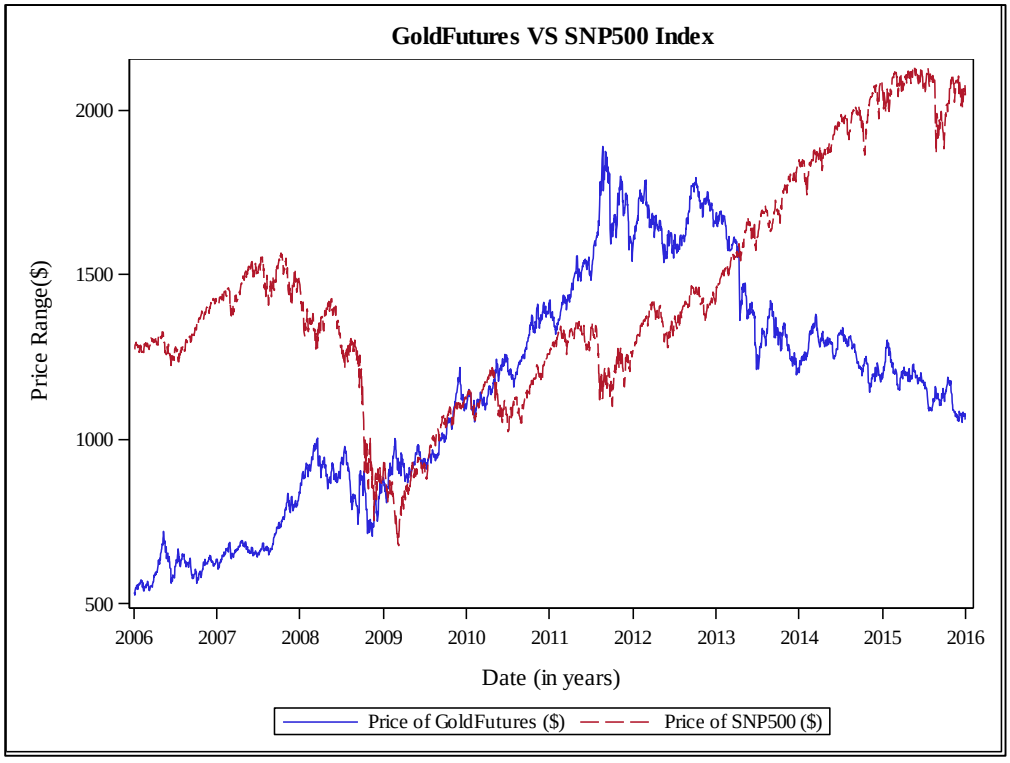


Figure 6: Line Chart of Gold Futures VS S&P 500 Index

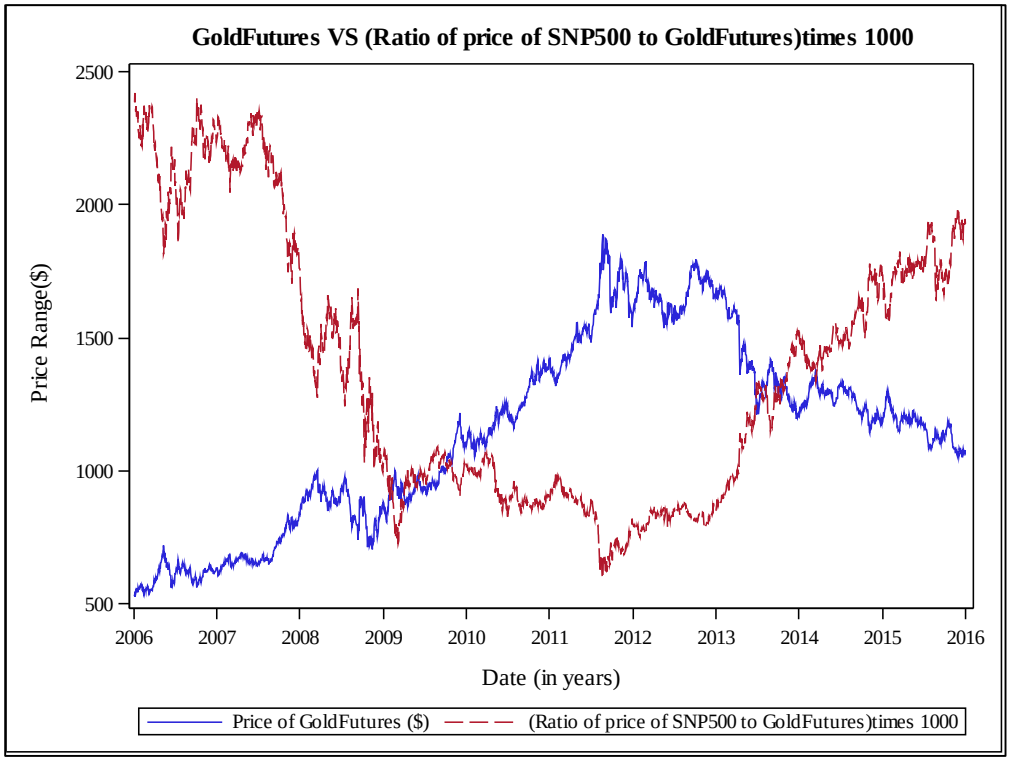


Figure 7: Line Chart of Gold Futures VS Ratio

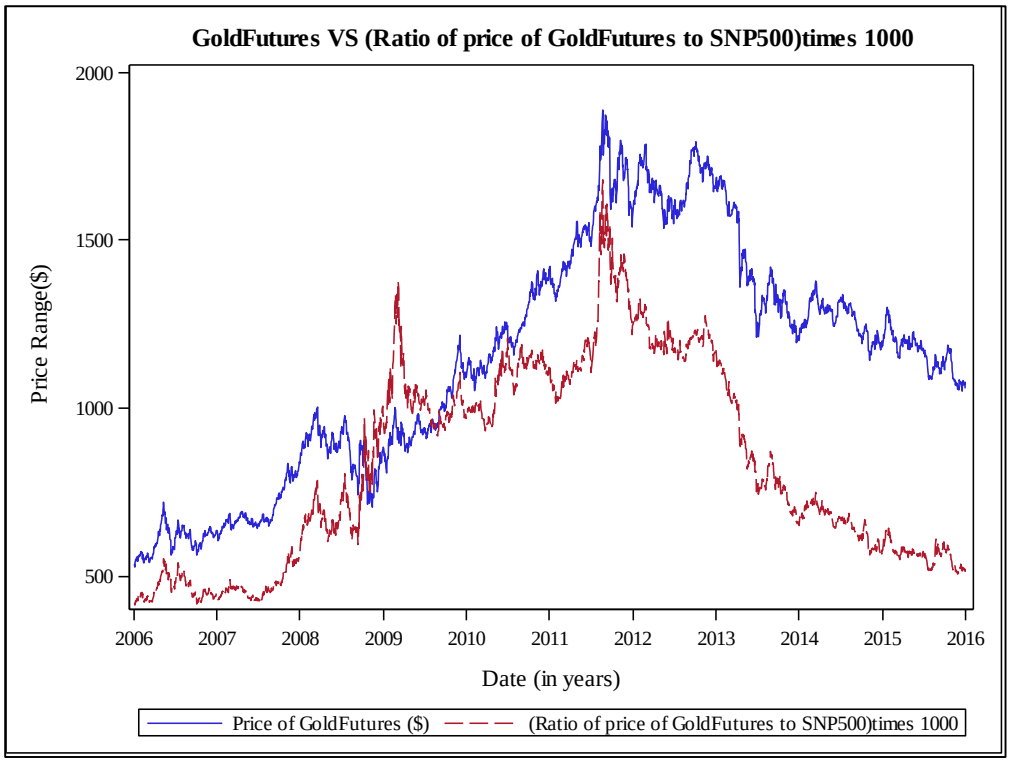


Figure 8: Line Chart of Gold Futures VS Ratio

Relevant Codes

```
/*%LET - creates a macro variable and assigns value to it*/
%LET file_location = C:\Users\sanjay\Desktop\final project;
%LET ext = .html;
%LET first = GoldFutures;
%LET second = SNP500;
%LET num = 1;
%LET num2 = 2;

/* %MACRO - begins a macro definition*/
%MACRO project(dataname);

 /* DATA step obtains and prepares the data*/
 data &dataname;

      /* INFILE - uses the data assigned to it*/
      infile "&file_location\&dataname&ext"
      delimiter = '>';

      /* INPUT - indicates what to expect on each line of data*/
      input string $ 32767. @@;
      n = _N_;

      /* INDEX - searches a character expression for a string of characters,
      and returns the position of the string's first character for the first
      occurrence of the string */
      a = index(string, "Dec 31, 2015");

            /* IF/ELSE helps us to perform an experiment conditionally*/
            if a = 0 then b = a;
            else call symput('c', n); /*CALL SYMPUT - assigns value produced
                                        in data step to macro-variable*/

 run;

 data &dataname&num;
      infile "&file_location\&dataname&ext"
            firstobs = &c end = eof
            delimiter = '<';
      input @'bold noWrap">' Old_date_&dataname $ 12.
            @'Font">' Price_&dataname
            @'<td>' Open_&dataname
            @'<td>' High_&dataname
            @'<td>' Low_&dataname
            @'">' Vol_&dataname$
            @'">' Pct_Change_&dataname$;

      /* LABEL - changes what is shown by many procs in place of variable
      names*/
      label old_date_&dataname = "Date"
            Price_&dataname = "Price of &dataname ($)"
            Open_&dataname = "Open Price of &dataname ($)"
            High_&dataname = "High price of &dataname ($)"
            Low_&dataname = "Low price of &dataname ($)"
            Vol_&dataname = "Volume of &dataname"
```

```
            Pct_change_&dataname = "Percentage change of &dataname";
 run;

 data &dataname&num2;

      /* SET - reads the observation from the SAS dataset*/
      set &dataname&num;

      /* INPUT - convert a character value to a numeric value*/
      /* ANYDTDTE - informat that extracts the date part from the derived
      value*/
      Date = input (old_date_&dataname, anydtdte12.);
      format Date yymmdd8.;
      vol_&dataname = compress(Vol_&dataname, 'K') *1000;
      drop old_date_&dataname;
 run;

/* PROC SORT - sorts the data */
 proc sort data=&dataname&num2;
      by Date;
 run;

/* %MEND - ends a macro definition*/
%MEND project;
%project(&first);
%project(&second);

data combined;
      /* MERGE - joins observations from two or more SAS data sets into a
      single observation*/
      merge &first&num2 &second&num2;
      by date;
      &first.to&second._Ratio = (price_&first/price_&second)*1000;
      &second.to&first._Ratio = (price_&second/price_&first)*1000;
      label &first.to&second._Ratio = "(Ratio of price of &first to
                                      &second)times 1000";
      label &second.to&first._Ratio = "(Ratio of price of &second to
                                      &first)times 1000";

      /* DROP - excludes variables from output SAS data sets*/
      drop vol_&second;
run;

/* ODS RTF - send output generated below to this file (RTF = Word document)*/
ods rtf bodytitle file = "C:\Users\sanjay\Desktop\final project\final.rtf";
%MACRO chart1(tie=, by=);

 /* PROC SGPLOT - draw graphs using a Statistical Graphics plotting
 procedure*/
 proc sgplot data = combined;
      title &tie;
      series x = date y = price_&first;
      series x = date y = &by;
      xaxis label = "Date (in years)";
      yaxis label = " Price Range($)";
 run;
```

```
%MEND chart1;
%chart1(tie="&first VS &second Index", by= price_&second) /* This code
produces Figure 6*/
%chart1(tie = "&first VS (Ratio of price of &second to &first)times 1000 ",
by = &second.to&first._Ratio) /* This code produces Figure 7*/
%chart1(tie = "&first VS (Ratio of price of &first to &second)times 1000 ",by
= &first.to&second._Ratio) /* This code produces Figure 8*/

/* PROC CORR - computes Pearson correlation coefficients*/
proc corr data = combined     /* This code produces Table 1*/

      /* NOMISS - excludes observations with missing analysis values from the
      analysis*/
      /* NOPROB - Suppresses p-values*/
      /* NOSIMPLE - Suppresses descriptive statistics*/
      outp = combined1 nomiss noprob nosimple;
      var  price_&second price_&first;
      with &second.to&first._Ratio &first.to&second._Ratio price_&first;
run;

/*close the file in SAS so we can open it in Word*/
ods rtf close;
```