\*\*\* This document includes information on the steps taken to complete team project 1. \*\*\*

Benjamin, Bruce, and Gary

* We have decided to analyze data which is relevant to the beginning of economic downturns. The goal is to find observable signals that can be used for trading or monetizing on the prediction of a potential downturn.
* The first steps taken were to set up group responsibilities, create questions that we wanted to answer, and decide which data sets would be utilized for analysis.
* The team decided that each section of the project would be worked on together. We would focus on individual parts of each “next-step” subjects, reconvene to share answers, and come to agreed upon answers and conclusions for each section.
* We decided that “quandl.com” would be the source of all financial data. A kudos must be given to Benjamin Aubry for suggesting the data site.
* A meeting was held that involved everyone scanning the website, “quandl.com”, for data that would be meaningful and impactful in our analysis.
* After reviewing the data site, it was concluded that we will analyze the following:
  1. Differences between corporate bond yields with ratings of AAA, BBB, and B ratings, vs the 5-year Treasury Bond yield. These ratings represent the highest credit quality, average investment grade credit quality, and a very low-grade credit quality vs the highest credit quality issued by the United States Treasury, respectively.
  2. The S&P 500’s average dividend yield vs the difference between the US 10-year Treasury Bond yield and the US 2-year Treasury Bond yield.
  3. Gold 1-month futures vs copper 1-month futures.
  4. Comparisons between each calculated metric mentioned above and the S&P 500 1-month futures.
* A shared github repository was created to complete the project.
* The team decided that we wanted to analyze data after 1985 for each realized economic downturn.
  1. 1987 (Black Monday)
  2. 1999/2000 (Dot-com bubble)
  3. 2001 (September 11th)
  4. 2008 (The Great Recession)
  5. 2020 (Coronavirus)

\*Some data is not available for the specified dates, so the next earliest date available is used for comparison when necessary.

* The team collected data and began merging appropriate datapoints for analysis.
* Each team member then took a portion of the project to complete, while everyone shared input on each other’s respective parts.
  1. Ben Aubry focused on the differences between corporate bond yields with ratings of AAA, BBB, and B ratings, vs the 5-year Treasury Bond yield.
     + These were his general steps taken
       - Imported data from CSVs (the Bond Index files)
       - Performed multiple merging as each index was a separated file
       - Merged on “Date” field
       - Renamed the columns with new names (names from initial CSVs were not very indicative)
       - Sorted by date
       - Converted dates from “string” to “date” using the datetime library
       - Imported the Treasury data (US Treasury Bond data) and turned it into a a Pandas dataframe.
       - Sorted the treasury data by date
       - Merged the Treasury with the Bond data to create a larger dataframe.
       - Plotted the yields to have a look at the data from the data frame
       - Created a subset dataframe to calculate spreads against the 5Y Treasury bond data –
       - Created a “Returns” dataframe (with the daily returns for all indices) - Spread Return dataframe
       - Imported S&P500 data and calculated daily returns on the S&P500
       - Merged the Bond spread data frame with the S&P500 Returns data frame
       - Created a subset data frame just for the LTCM time period (1997-1999) and plotted S&P500 vs Spread (B-Bond index – 5Y Treasury)
       - Plotted the daily returns for both S&P500 and Spread Index
       - Calculated a rolling correlation between the S&P500 and Bond Index returns – and plotted the correlation. This is where was able to see some potential insights from the data
       - Moved forward to creating a subset data frame for the Internet bubble (this was created from the large data frames created earlier)
       - Performed similar calculations as above (i.e. Rolling correlations)
       - Created a subset for data from 2017-2020 – performed the same analysis as above but added rolling standard deviations and rolling mean for signals.
       - Plotted the chart (line plots)
  2. Gary Fisher focused on the S&P 500’s average dividend yield vs the difference between the US 10-year Treasury Bond yield and the US 2-year Treasury Bond yield.
     + These were his general steps taken
       - Imported data from CSVs
       - Merged the data using Outer method
       - Sorted the data by Index
       - Changed date type to "datetime" for further sorting by date range
       - Cleaned data to start at earliest common start date for calculations
       - Calculated Daily Dividend Yield from the monthly value
       - Calculated daily percent returns
       - Inserted daily percent returns into data frame
       - Checked for NaN values, and shape of data frame
       - Dropped NaN rows from data frame
       - Sorted data frame by date range
       - Calculated total returns for date range
       - Created Master data frame and dropped infinite numbers
       - Created initial chart of variables used for charting
       - Created function for scatter plots
       - Created scatterplots of different variables for correlation comparisons
       - Created daily return chart for primary variables used
       - Created scatterplot of primary variables for correlation comparison
       - Created primary line plots showing full range of data used in charts
       - Created multiples line plots from different date ranges to show individual instances of findings from data
  3. Bruce Mark focused on Gold 1-month futures vs copper 1-month futures.
     + These were his general steps taken
       - Import csv’s, review data structure and for n/a’s.
       - Sort the SP df by date
       - Calculate the daily return for S&P and add it to the df.
       - Create separate DF’s for different timeframes and chart the S&P settle price for these time periods. Each time period was approximately 1 year prior to the event:
         * 87 crash
         * LTCM
         * Internet Bubble
         * 9/11
         * 2008 Financial crisis
         * 2012 sell – off
         * China Black Monday sell-off
         * Covid virus sell-off
       - Merge the copper (CME future) and Gold (CME future) df’s by date
       - Consolidate columns in df to date and settle price for both futures.
       - Calculate the correlation of the copper and gold settle prices
       - Calculate the copper gold ratio and add it to the df (copper settle / gold settle)
       - Create the DF’s for different timeframes and chart the ratio for these periods.
         * 87 crash
         * LTCM
         * Internet Bubble
         * 9/11
         * 2008 Financial crisis
         * 2012 sell – off
         * China Black Monday sell-off
         * Covid virus sell-off
       - Merge the copper gold df with the S&P df
       - Plot the S&P settle price for each time period above against the Copper Gold ratio
       - Calculate the daily change of the copper gold ratio and plot against the S&P settle price for each time period.
       - Calculate the correlation of the S&P settle price against the Copper Gold ratio
       - Plot the daily return of the S&P settle price against the daily % change in the copper gold ratio.
       - Calculate the correlations of copper gold ratio against the S&P settle price for each time frame.