

Md Uddin – Final Project

Abstract

In this project we will analyze a multi-asset portfolio consisting of stocks (SPY), inverse bonds (TNX), oil (USO) and gold (GLD) looking at year-to-date (YTD) returns. We will perform summary statistical analysis, calculate portfolio returns, perform multi-regression analysis and lastly analyze YTD price action through K-means clustering.

However, as this is not a traditional portfolio, we first look at the different components:

Stocks are represented by SPY - The SPDR S&P 500 trust exchange-traded fund which trades on the NYSE. This fund is the largest and oldest ETF in the world.

As interest rates rise, bond prices go down, as thus, we consider an inverse exposure to bonds through TNX - CBOE Ten Year Treasury Note Yield Index . TNX is a proxy for tradeable exposure such as TBF - ProShares Short 20+ Yr Treasury.

Oil has taken a center stage in portfolio construction as economies globally start to reopen. For exposure in oil, we use USO. USO invests primarily in listed crude oil futures contracts and other oil-related contracts and may invest in forwards and swap contracts. It is important to note that USO uses oil futures, whereby over the long term, the negative roll yields add up.

In this inflationary environment, gold is historically used as a hedge. For exposure to gold, we use GLD. For a few years, the fund was the second-largest exchange-traded fund in the world, and it was briefly the largest.

Loading Data

```
Gold_Oil <- read_csv(  
  file = "Gold _ Oil.csv",  
  col_types = cols(  
    Date = col_date(format = "%m/%d/%y")) |>  
  clean_names()
```

```
Stocks_Bonds <- read_csv(  
  file = "Stocks _ Bonds.csv",  
  col_types = cols(  
    Date = col_date(format = "%m/%d/%y")) |>  
  clean_names()
```

```
Stocks_Bonds_Gold_Oil <- merge(  
  Stocks_Bonds, Gold_Oil, by = "Date", all = TRUE)
```

```

  x = Stocks_Bonds, y = Gold_Oil, by = 'date'
)
head(Stocks_Bonds_Gold_Oil)

##           date spy_equities tnx_10_year_treasury_yield gld_gold uso_oil
## 1 2022-01-03      472.2423                1.628    168.33   54.83
## 2 2022-01-04      472.0841                1.668    169.57   55.59
## 3 2022-01-05      463.0191                1.705    169.06   55.43
## 4 2022-01-06      462.5841                1.733    166.99   56.89
## 5 2022-01-07      460.7553                1.771    167.75   56.67
## 6 2022-01-10      460.1819                1.780    168.26   56.16

```

Here we have merged both files “Stocks_Bonds” and “Gold_Oil” into one data set by date.

Summary Analysis

```

summary_stat = Stocks_Bonds_Gold_Oil |>
  select(-date) |>
  summarise(
    across(
      .fns = list(
        min = min,
        mean = mean,
        max = max,
        sd = sd
      ), .names = "{.col}={.fn}"
    ) |> t()
summary_stat <- data.frame(summary_stat) |>
  as_tibble(rownames = 'var') |>
  separate(
    col = var, into = c("Variable", "Statistic"),
    sep = '=',
    extra = 'warn'
  ) |>
  pivot_wider(
    names_from = Statistic,
    values_from = summary_stat)
summary_stat

## # A tibble: 4 × 5
##   Variable      min    mean    max    sd
##   <chr>      <dbl>  <dbl>  <dbl>  <dbl>
## 1 spy_equities    357.   408.   472.   27.8
## 2 tnx_10_year_treasury_yield  1.63    2.91    4.23    0.712
## 3 gld_gold       151.   168.   192.    9.04
## 4 uso_oil        54.8   73.2   92.0    7.48

```

Percent of change difference between the min and max price YTD for SPY is 32%, for TNX is 159%, for GLD is 27% and for USO is 67%. SPY which offers a risk premium had a standard deviation of 27.8 and TNX has a standard deviation of .712.

YTD Returns

```
Stocks_Bonds_Gold_Oil_xts <- xts(
  Stocks_Bonds_Gold_Oil[,2:5],
  order.by = Stocks_Bonds_Gold_Oil[,1])

returns_data = data.frame()
for (i in names(Stocks_Bonds_Gold_Oil_xts)){
  returns_data[1, i] = periodReturn(
    x = Stocks_Bonds_Gold_Oil_xts[,i], period = 'yearly')
}
returns_data

##   spy_equities tnx_10_year_treasury_yield    gld_gold    uso_oil
## 1   -0.1672072          1.191032 -0.007544728  0.1462702
```

YTD SPY is down 16%, TNX is up 119%, GLD is down .7% and USO is up 14.6%. Higher inflation led to higher interest rates which have slowed company growth as capital is now more expensive. Thus, stocks are down for the year. Gold is considered an inflation hedge, and while it is not positive for the year, it has greatly outperformed stocks. Oil is up for the year led by greater demand with economies opening globally and restricted supply in the backdrop of Russia – Ukraine conflict.

60 / 40 Portfolio Analysis

```
(1+returns_data[1,1])*6000 #SPY

##           [,1]
## [1,] 4996.757

(1+returns_data[1,2])*4000#TNX

##           [,1]
## [1,] 8764.128
```

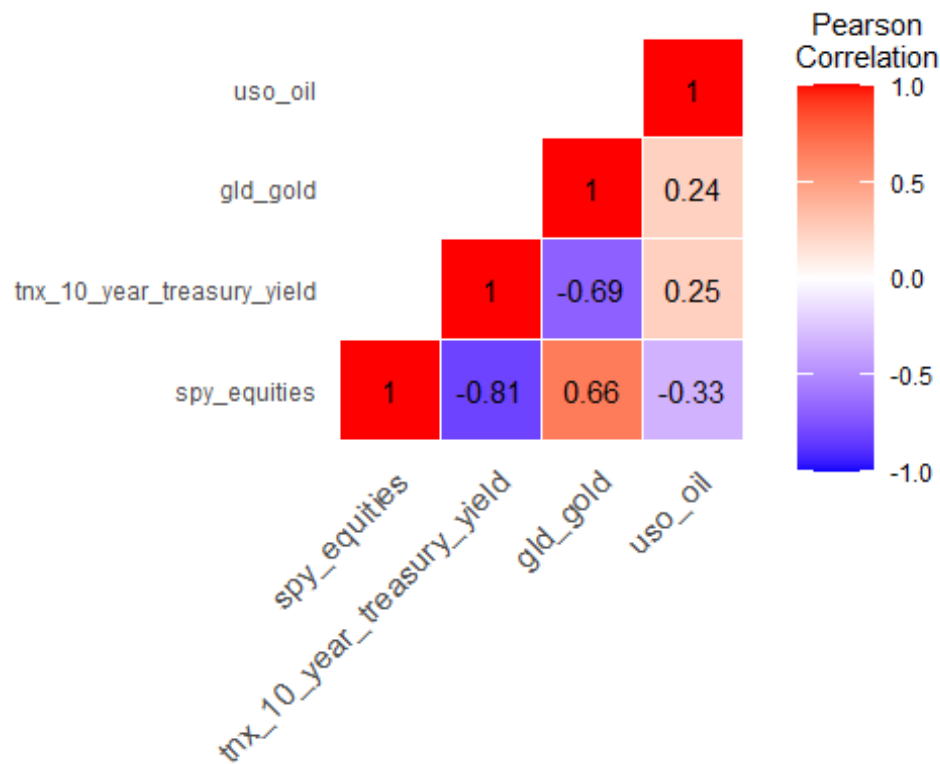
It is common to construct a portfolio that is 60% equities and 40% bonds. However, as rates have been increasing this year, what if you constructed a portfolio of 60% stocks and 40% inverse bonds – for which we will use TNX as a proxy. Here with a 60/40 portfolio with \$6000 in SPY and \$4000 in TNX (proxy only) YTD you would have a total balance of \$13760.7.

Correlation Matrix

```
cormat <- round(cor(Stocks_Bonds_Gold_Oil_xts),2)

get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

upper_tri <- get_upper_tri(cormat)
melted_cormat <- melt(upper_tri, na.rm = TRUE)
ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
  geom_tile(color = "white")+
  scale_fill_gradient2(
    low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name="Pearson\nCorrelation") +
  theme_minimal() +
  geom_text(
    aes(Var2, Var1, label = value),
    color = "black", size = 4) +
  theme(
    axis.text.x = element_text(
      angle = 45, vjust = 1, size = 12, hjust = 1),
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.grid.major = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank(),
    axis.ticks = element_blank(),
    legend.direction = "vertical")+
  guides(fill = guide_colorbar(
    barwidth = 2, barheight = 10,
    title.position = "top", title.hjust = 0.5)) +
  coord_fixed()
```



In constructing a portfolio, we want our components to be diverse. Analyzing the matrix above we can see that SPY & TNX has a strong negative correlation of -.81. This means that stocks and bonds this year would have a positive correlation. Thus, a portfolio of stocks and bonds will be down for the year in both components. Stocks and gold (GLD) have a positive correlation. Stocks and oil (USO) this year had a negative correlation. Higher inflation which have held down stocks YTD have been in part caused by higher oil prices.

Multi Regression Analysis

```
multiple_reg <- lm(
  formula = spy_equities~tnx_10_year_treasury_yield+gld_gold+uso_oil,
  data = Stocks_Bonds_Gold_Oil_xts)
summary(multiple_reg)

##
## Call:
## lm(formula = spy_equities ~ tn timer_10_year_treasury_yield + gld_gold +
##     uso_oil, data = Stocks_Bonds_Gold_Oil_xts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.673 -11.807   2.675  10.974  29.710
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          315.2026      28.7689  10.956 < 2e-16 ***
## tn timer_10_year_treasury_yield -16.4283      2.1870  -7.512 1.24e-12 ***
## gld_gold             1.3781      0.1719   8.018 5.19e-14 ***
## uso_oil              -1.2432      0.1551  -8.017 5.21e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.98 on 233 degrees of freedom
## Multiple R-squared:  0.7502, Adjusted R-squared:  0.747
## F-statistic: 233.2 on 3 and 233 DF,  p-value: < 2.2e-16
```

Here our equation is:

$$SPY = 315.2 - 16.4 (TNX) + 1.37(GLD) - 1.24 (USO)$$

Looking at the adjusted R-squared we also that the 3 components of TNX, GLD & USO describe our MEDV (SPY) with 74.7% accuracy.

K-means Clustering for Price Action Analysis

In our K-means clustering the normalized stock prices are clustered into 6 cluster points with each dates.

```
Stocks_Bonds_Gold_Oil_2 <- Stocks_Bonds_Gold_Oil
rownames(Stocks_Bonds_Gold_Oil_2) <- Stocks_Bonds_Gold_Oil_2$date
Stocks_Bonds_Gold_Oil_2$date <- NULL

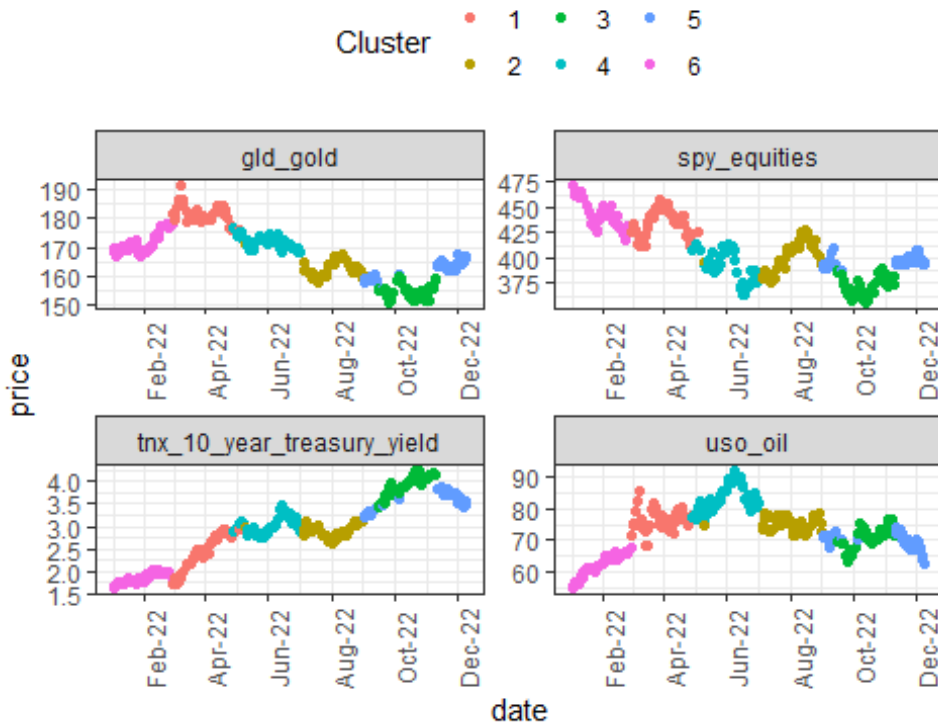
Stocks_Bonds_Gold_Oil_normalised <- sapply(Stocks_Bonds_Gold_Oil_2, scale)
rownames(Stocks_Bonds_Gold_Oil_normalised) <- rownames(Stocks_Bonds_Gold_Oil_2)

km <- kmeans(Stocks_Bonds_Gold_Oil_normalised, 6)

Stocks_Bonds_Gold_Oil_3 <- Stocks_Bonds_Gold_Oil
Stocks_Bonds_Gold_Oil_3$Cluster <- km$cluster
Stocks_Bonds_Gold_Oil_3$Cluster <- factor(Stocks_Bonds_Gold_Oil_3$Cluster)

Stocks_Bonds_Gold_Oil_3 %>%
  pivot_longer(
    cols = spy_equities:uso_oil,
    names_to = 'index', values_to = 'price'
  ) %>%
  ggplot(aes(x = date, y = price, color = Cluster)) +
  geom_point() +
  scale_x_date(date_breaks = "2 month", date_labels = "%b-%y") +
  facet_wrap(~index, scales = 'free') +
```

```
theme_bw() +
theme(legend.position = 'top', axis.text.x = element_text(angle = 90))
```



Analyzing our cluster chart above we can aptly compare price movements throughout the year for our securities. We can determine trends such as SPY YTD has been in a downward trend while TNX have been in an upward trend. It is interesting to note that both SPY and TNX seems to reverse their longer-term YTD trend in October. This may be because inflation expectations began to shift in October. USO peaked around July when inflation also peaked and have been coming down since – although still higher YTD. Previously looking at YTD returns we know gold is flat for the year, however, price movement has been anything but flat. After peaking in April, GLD has been in a downward run after reversing course in October.

Conclusion

In this project we analyzed 4 potential components of a non-traditional portfolio given the current macroeconomic backdrop including stocks (SPY), bond yields (TNX), oil (USO) and gold (GLD).

In our summary statistical analysis, we determined that YTD, SPY has a standard deviation of 27.8 and TNX has a standard deviation of .712. Herein, considering YTD returns, SPY is

down 16.7% and TNX is up 119%.

In our correlation analysis we determined that YTD SPY and TNX have a correlation of -.81 and SPY and GLD have a correlation of .66.

In performing our multi regression analysis we determined the formula for our MEDV (SPY) is $315.2 - 16.4 (TNX) + 1.37(GLD) - 1.24 (USO)$.

Lastly, using our K-means clustering analysis we were able to better understand YTD price movement of all our 4 components.

The markets this year have been a wild ride with a backdrop of high inflation, global instability and rising rates. However, we know that macro conditions change and with it the markets. Herein, while this year have not been great for stocks and bonds, you may have hedged with an inverse bond product as yields rose. In this, next year the market will change and analyzing different securities and their price action we can develop a diverse portfolio in consideration of mean-variance and correlation to weather any storms ahead!