Data Task

Peichen Li

Data Task

Data Preparation

First we take a look at the dataset we are interested in.

```
setwd("C:/Users/90596/Desktop")
df = read_csv("F-F_Research_Data_5_Factors_2x3_daily.CSV")
glimpse(df)
```

```
#> Rows: 14,434
#> Columns: 7
#> $ Date     <dbl> 19630701, 19630702, 19630703, 19630705, 19630708, 19630709...
#> $ `Mkt-RF` <dbl> -0.67, 0.79, 0.63, 0.40, -0.63, 0.45, -0.18, -0.16, -0.12,...
#> $ SMB      <dbl> 0.00, -0.27, -0.17, 0.08, 0.04, 0.00, 0.21, 0.14, 0.02, 0....
#> $ HML      <dbl> -0.32, 0.27, -0.09, -0.28, -0.18, 0.10, 0.01, -0.30, -0.11...
#> $ RMW      <dbl> -0.01, -0.07, 0.17, 0.08, -0.29, 0.14, 0.06, -0.06, 0.12, ...
#> $ CMA      <dbl> 0.15, -0.19, -0.33, -0.33, 0.13, -0.04, -0.07, 0.05, 0.04,...
#> $ RF       <dbl> 0.012, 0.012, 0.012, 0.012, 0.012, 0.012, 0.012, 0.012, 0....
```

Now we need to do some data cleaning

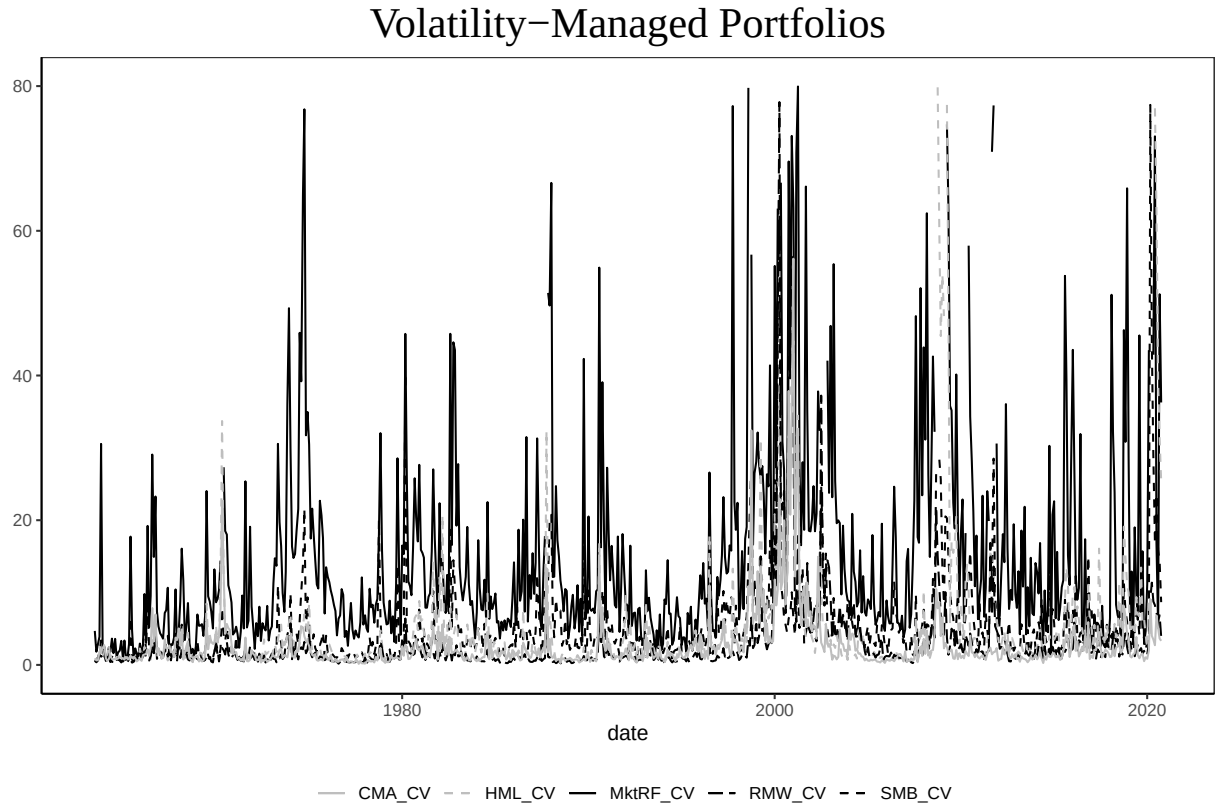
```
# Transform Date column from integer type to Date object
df$Date = ymd(df$Date)
# We calculate the portfolio's conditional variance by equation (2)
df = df %>%
  group_by(year(Date), month(Date)) %>%
  mutate( MktRF_CV = sum( ( `Mkt-RF` - mean(`Mkt-RF`) )^2 ),
          SMB_CV = sum( ( SMB - mean(SMB) )^2),
          HML_CV = sum( ( HML - mean(HML) )^2),
          RMW_CV = sum( ( RMW - mean(RMW) )^2),
          CMA_CV = sum( ( CMA - mean(CMA) )^2) ) %>%
  slice(1) %>%
  ungroup() %>%
  select( Date, MktRF_CV, SMB_CV, HML_CV, RMW_CV, CMA_CV )


# Get the data of the monthly volatility of each factor
df %>% write_csv("volatility_by_factor_monthly.csv")
```

Question 1

```
ggplot(df,aes(x=Date)) +
  geom_line(aes(y=MktRF_CV, linetype="MktRF_CV", color="MktRF_CV")) +
  geom_line(aes(y=SMB_CV, linetype="SMB_CV", color="SMB_CV")) +
```

```
  geom_line(aes(y=HML_CV, linetype="HML_CV", color="HML_CV")) +
  geom_line(aes(y=RMW_CV, linetype="RMW_CV", color="RMW_CV")) +
  geom_line(aes(y=CMA_CV, linetype="CMA_CV", color="CMA_CV")) +
  scale_x_date(date_breaks = "20 years", date_labels = "%Y") +
  ylim(0, 80) +
  labs(x="date", y="", title="Volatility-Managed Portfolios") +
  scale_linetype_manual(name="",values = c("MktRF_CV"="solid",
                                           "SMB_CV"="dashed",
                                           "HML_CV"="dashed",
                                           "RMW_CV"="longdash",
                                           "CMA_CV"="solid")) +
  scale_color_manual(name="",values = c("MktRF_CV"="black",
                                        "SMB_CV"="black",
                                        "HML_CV"="gray",
                                        "RMW_CV"="black",
                                        "CMA_CV"="gray")) +
  theme_bw() +
  theme(legend.position="bottom",
        axis.line = element_line(colour = "black"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5, size = 22, family="serif"))
```



Question 2

```
## Question 2
# Calculate c using the full sample of daily data
df = read_csv("F-F_Research_Data_5_Factors_2x3_daily.CSV")
c1 <- sd(df$`Mkt-RF`)
c2 <- sd(df$SMB)
c3 <- sd(df$HML)
c4 <- sd(df$RMW)
c5 <- sd(df$CMA)

df$Date = ymd(df$Date)
# Calculate the variance of each factor's excess returns within each month
df = df %>%
  group_by(year(Date), month(Date)) %>%
  mutate( MktRF_CV = sum( ( `Mkt-RF` - mean(`Mkt-RF`) )^2 ),
          SMB_CV = sum( ( SMB - mean(SMB) )^2),
          HML_CV = sum( ( HML - mean(HML) )^2),
          RMW_CV = sum( ( RMW - mean(RMW) )^2),
          CMA_CV = sum( ( CMA - mean(CMA) )^2) ) %>%
  slice(1) %>%
  ungroup() %>%
  mutate(Year = year(Date), Month = month(Date))
# Import the monthly factor excess returns as df1
df1 = read_csv("F-F_Research_Data_5_Factors_2x3.CSV")
df1 = df1 %>%
  mutate(Year = as.numeric(substr(Date, 1, 4)), Month =
           as.numeric(substr(Date, 5, 6)))
# Merge the data sets
df2 = full_join(df,df1,by=c("Year","Month")) %>%
  select(Year, Month, MktRF_CV, SMB_CV, HML_CV, RMW_CV, CMA_CV,
         MktRF=`Mkt-RF.y`, SMB=SMB.y, HML=HML.y, RMW=RMW.y, CMA=CMA.y) %>%
# Then calculate the volatility managed factors by merging
# the monthly realized factor variances with the monthly factor
# excess returns to calculate equation (1)
  mutate(MP_MktRF = c1 / lag(MktRF_CV) * MktRF,
         MP_SMB = c2 / lag(SMB_CV) * SMB,
         MP_HML = c3 / lag(HML_CV) * HML,
         MP_RMW = c4 / lag(RMW_CV) * RMW,
         MP_CMA = c5 / lag(CMA_CV) * CMA) %>%
# Remove the first move since its previous month's variance is undefined
  slice(-1)

# Write to csv
df2 %>% write_csv("equation_1_data.csv")

# Now Using the volatility managed factors and the original factors,
# run the regression specified in equation 3 for each of the factors

# Notice that this is a simple linear regression model,
# with the the volatility managed factors being the dependent variable and the

# original factors being the independent variable for each of the factor
```

```r
attach(df2)

m1 <- lm(MP_MktRF ~ I(MktRF*12))
m2 <- lm(MP_SMB ~ I(SMB*12))
m3 <- lm(MP_HML ~ I(HML*12))
m4 <- lm(MP_RMW ~ I(RMW*12))
m5 <- lm(MP_CMA ~ I(CMA*12))

options(digits=2)
stargazer(m1, m2, m3, m4,m5, type="latex",
          dep.var.labels=c("MktVo","SMBVo","HMLVo","RMWVo","CMAVo"),
          covariate.labels=c("MktRF","SMB","HML", "RMW","CMA"),
          out="models.txt",
          header = FALSE)
```

Table 1:

	Dependent variable:				
	MktVo	SMBVo	HMLVo	RMWVo	CMAVo
	(1)	(2)	(3)	(4)	(5)
MktRF	0.007***				
	(0.0003)				
SMB		0.014***			
		(0.001)			
HML			0.015***		
			(0.001)		
RMW				0.015***	
				(0.001)	
CMA					0.016***
					(0.001)
Constant	0.024	0.002	0.035	0.062***	0.007
	(0.017)	(0.020)	(0.023)	(0.020)	(0.016)
Observations	687	687	687	687	687
R^2	0.430	0.490	0.430	0.350	0.470
Adjusted R^2	0.430	0.490	0.430	0.350	0.470
Residual Std. Error (df = 685)	0.430	0.540	0.600	0.520	0.400
F Statistic (df = 1; 685)	524.000***	658.000***	513.000***	369.000***	602.000***
Note:					*p<0.1; **p<0.05; ***p<0.01

```r
detach(df2)

# Note that this table is note nicely formatted, and then needs to be
# improved in several ways:
# 1) Rename the intercept (constant) as Alpha and rename the observations to
# capital N
```

```
# 2) Report  RMSE(root-mean-square-deviation)
# 3) Can use linear regression with robust std errors
```

References

- RA Data Task (2020), NYU Stern Professor Robert Richmond.
- MOREIRA, A. and MUIR, T., 2017. Volatility-Managed Portfolios. The Journal of Finance, 72(4), pp.1611-1644.