Exploring portfolio returns

A four-factor model analysis of different risk sources

0. Data preparation

```
library(tidyverse)
library(ggfortify)
library(fpp2)
library(lmtest)
library(sandwich)
library(car)
library(readxl)
library(quantreg)
library(moments)
library(strucchange)
library(urca)
library(fable)
library(hts)
library(knitr)
library(extrafont)
library(ggplot2)
library(jtools)
library(huxtable)
library(officer)
library(flextable)
library(openxlsx)

ds <- read_excel("First_Assignment_dataset.xlsx")

#date
ds$date = seq(as.Date(0, origin="1980-10-01"),length=363, by="1 month")

#converting returns in excess returns
ds = ds %>% mutate(XS01 = S1-rf,
                   XS02 = S2-rf,
                   XS03 = S3-rf,
                   XS04 = S4-rf,
                   XS05 = S5-rf,
                   XS06 = S6-rf,
                   XS07 = S7-rf,
                   XS08 = S8-rf,
                   XS09 = S9-rf,
                   XS10 = S10-rf)

#making dataset in long format
ds_long = ds %>% pivot_longer(XS01:XS10)
```

1. Data Visualization and Summary Statistics

Figure 1: Excess returns by portfolio

```
# excess return e outliers
theme_set(theme_minimal(base_family = "serif"))

xs01_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS01), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS01, 0.25)-1.5*IQR(ds$XS01), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS01, 0.75)+1.5*IQR(ds$XS01), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS01),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)



xs02_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS02), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS02, 0.25)-1.5*IQR(ds$XS02), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS02, 0.75)+1.5*IQR(ds$XS02), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS02),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)

xs03_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS03), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS03, 0.25)-1.5*IQR(ds$XS03), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS03, 0.75)+1.5*IQR(ds$XS03), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS03),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
```

```
        xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)

xs04_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS04), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS04, 0.25)-1.5*IQR(ds$XS04), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS04, 0.75)+1.5*IQR(ds$XS04), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS04),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
        xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)

xs05_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS05), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS05, 0.25)-1.5*IQR(ds$XS05), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS05, 0.75)+1.5*IQR(ds$XS05), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS05),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
        xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
        xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
        ymin = -Inf, ymax = Inf,
        fill = "#014F86",alpha=0.2)

xs06_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS06), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS06, 0.25)-1.5*IQR(ds$XS06), linetype = 2, col = 'brown')+
```

```
  geom_hline(yintercept = quantile(ds$XS06, 0.75)+1.5*IQR(ds$XS06), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS06),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)

xs07_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS07), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS07, 0.25)-1.5*IQR(ds$XS07), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS07, 0.75)+1.5*IQR(ds$XS07), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS07),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)

xs08_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS01), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS01, 0.25)-1.5*IQR(ds$XS01), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS01, 0.75)+1.5*IQR(ds$XS01), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS08),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)
```

```r
xs09_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS09), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS09, 0.25)-1.5*IQR(ds$XS09), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS09, 0.75)+1.5*IQR(ds$XS09), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS09),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)
xs10_plot = ds %>% ggplot()+
  geom_hline(yintercept = mean(ds$XS10), col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS10, 0.25)-1.5*IQR(ds$XS10), linetype = 2, col = 'brown')+
  geom_hline(yintercept = quantile(ds$XS10, 0.75)+1.5*IQR(ds$XS10), linetype = 2, col = 'brown')+
  geom_line(aes(date, XS10),lwd = 0.2)+
  xlab(NULL)+
  annotate(geom = "rect",
         xmin = as.Date('1981-03-01'), xmax = as.Date('1981-11-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('1986-11-01'), xmax = as.Date('1988-01-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)+
  annotate(geom = "rect",
         xmin = as.Date('2008-07-01'), xmax = as.Date('2009-06-01'),
         ymin = -Inf, ymax = Inf,
         fill = "#014F86",alpha=0.2)

library(gridExtra)
library(grid)
grid.arrange(xs01_plot,
             xs06_plot,
             xs02_plot,
             xs07_plot,
             xs03_plot,
             xs08_plot,
             xs04_plot,
             xs09_plot,
             xs05_plot,
             xs10_plot,
             nrow = 5)
```

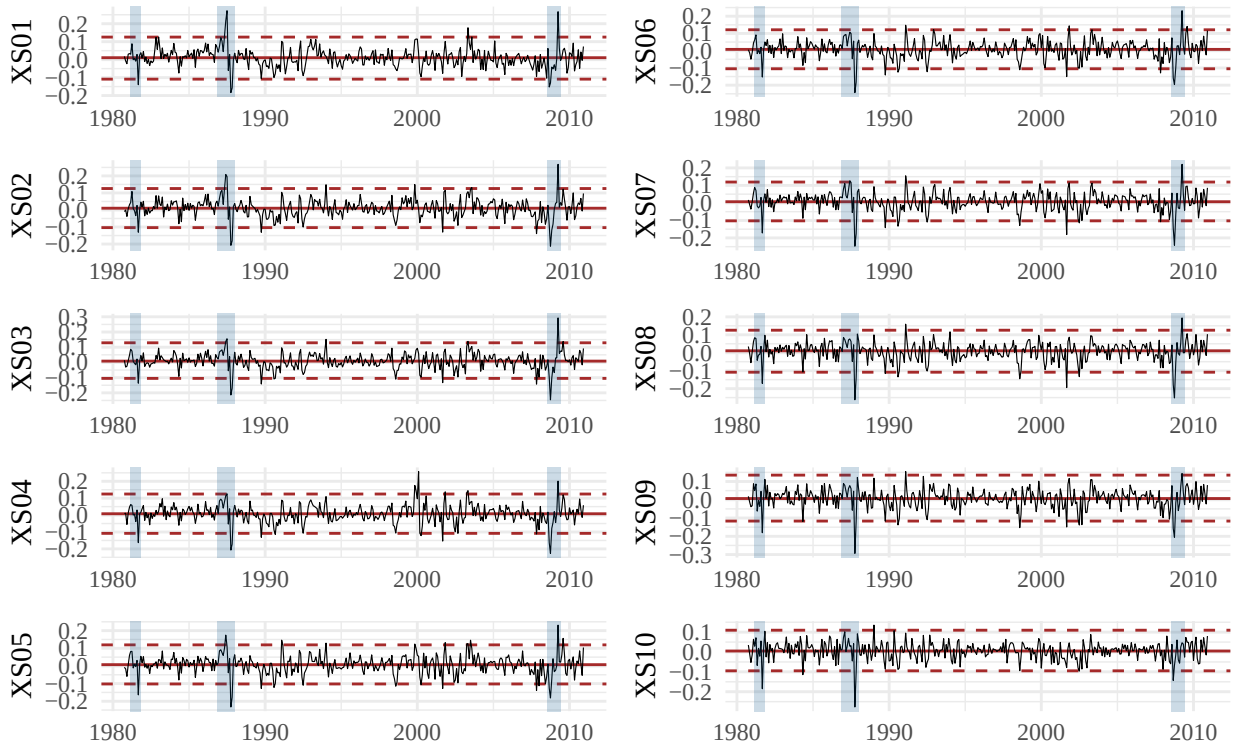


Table 1: Summary statistics of the excess returns

```
# summary
library(moments)
summary_tab = as.data.frame(ds_long %>%
                              group_by(name) %>%
                              summarise(mean=mean(value) %>% round(3),
                                        var=var(value)%>% round(3),
                                        st_dev=sd(value)%>% round(3),
                                        min=min(value)%>% round(3),
                                        q25=quantile(value, 0.25)%>% round(3),
                                        q50=quantile(value, 0.5)%>% round(3),
                                        q75=quantile(value, 0.75)%>% round(3),
                                        max=max(value)%>% round(3),
                                        skewness=skewness(value)%>% round(3),
                                        kurtosis=kurtosis(value)%>% round(3)) %>%
                              mutate("Sharpe ratio" = (c((mean(ds$S1)-mean(ds$rf))/sd(ds$S1),
                                                         (mean(ds$S2)-mean(ds$rf))/sd(ds$S2),
                                                         (mean(ds$S3)-mean(ds$rf))/sd(ds$S3),
                                                         (mean(ds$S4)-mean(ds$rf))/sd(ds$S4),
                                                         (mean(ds$S5)-mean(ds$rf))/sd(ds$S5),
                                                         (mean(ds$S6)-mean(ds$rf))/sd(ds$S6),
                                                         (mean(ds$S7)-mean(ds$rf))/sd(ds$S7),
                                                         (mean(ds$S8)-mean(ds$rf))/sd(ds$S8),
                                                         (mean(ds$S9)-mean(ds$rf))/sd(ds$S9),
                                                         (mean(ds$S10)-mean(ds$rf))/sd(ds$S10)))%>% roun
kable(summary_tab, caption = 'Summary statistics of the excess returns', align = 'l')
```

Table 1: Summary statistics of the excess returns

name	mean	var	st_dev	min	q25	q50	q75	max	skewness	kurtosis	Sharpe ratio
XS01	0.011	0.003	0.054	-0.184	-0.021	0.010	0.038	0.273	0.484	6.435	0.201
XS02	0.010	0.003	0.055	-0.213	-0.017	0.011	0.040	0.268	-0.039	5.949	0.191
XS03	0.009	0.003	0.054	-0.246	-0.017	0.011	0.041	0.293	-0.269	7.071	0.163
XS04	0.008	0.003	0.055	-0.227	-0.021	0.009	0.037	0.258	-0.195	6.018	0.144
XS05	0.007	0.003	0.055	-0.233	-0.019	0.010	0.036	0.231	-0.293	5.358	0.132
XS06	0.008	0.003	0.055	-0.243	-0.020	0.011	0.036	0.232	-0.514	5.603	0.140
XS07	0.007	0.003	0.054	-0.247	-0.019	0.010	0.036	0.221	-0.771	6.213	0.124
XS08	0.007	0.003	0.054	-0.262	-0.019	0.014	0.039	0.194	-1.016	6.564	0.122
XS09	0.007	0.003	0.055	-0.294	-0.022	0.014	0.041	0.155	-0.933	6.053	0.122
XS10	0.005	0.002	0.045	-0.277	-0.018	0.010	0.033	0.136	-1.037	7.318	0.111

Figure 2: Distribution of excess returns by portfolio

```
# boxplots
theme_set(theme_minimal(base_family = "serif"))
ds_long %>% ggplot()+
  geom_boxplot(aes(name, value, fill = name), show.legend = F) +
  scale_fill_manual(name="color", values=c("#012A4A", "#013A63", "#01497C", "#014F86", "#2A6F97", "#2C7
  theme(plot.title = element_text(hjust = 0.5, size = 12))+
  xlab("Portfolios")+
  ylab("Excess returns")
```

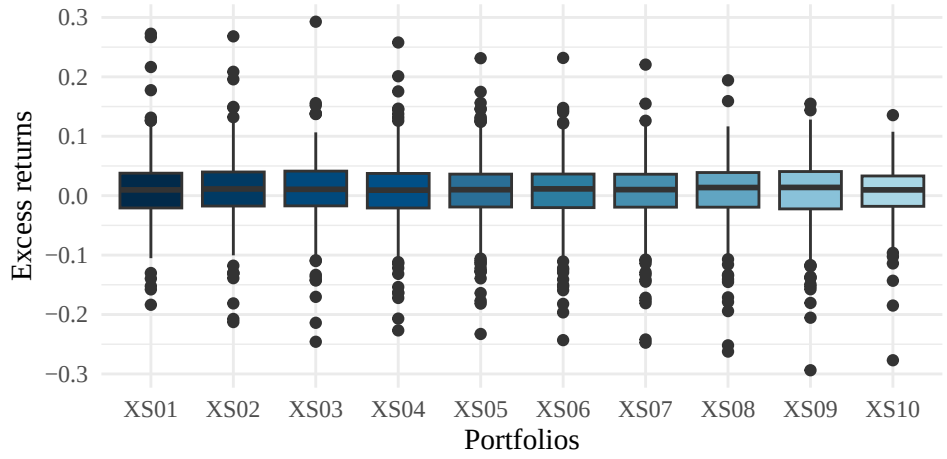


Figure 3: Correlations between portfolios' excess returns

```
# correlation between the portfolios
library(GGally)
theme_set(theme_minimal(base_family = 'serif'))
ds %>%
  dplyr::select(XS01:XS10) %>%
  ggpairs(upper = list(continuous = wrap('cor',size = 2.5)),
          lower = list(continuous = wrap("points", pch = 1, size=0.8)))+
  theme(axis.text = element_text(size = 5),
```

```
        text = element_text(size = 5),
        strip.text = element_text(size = 8))
```

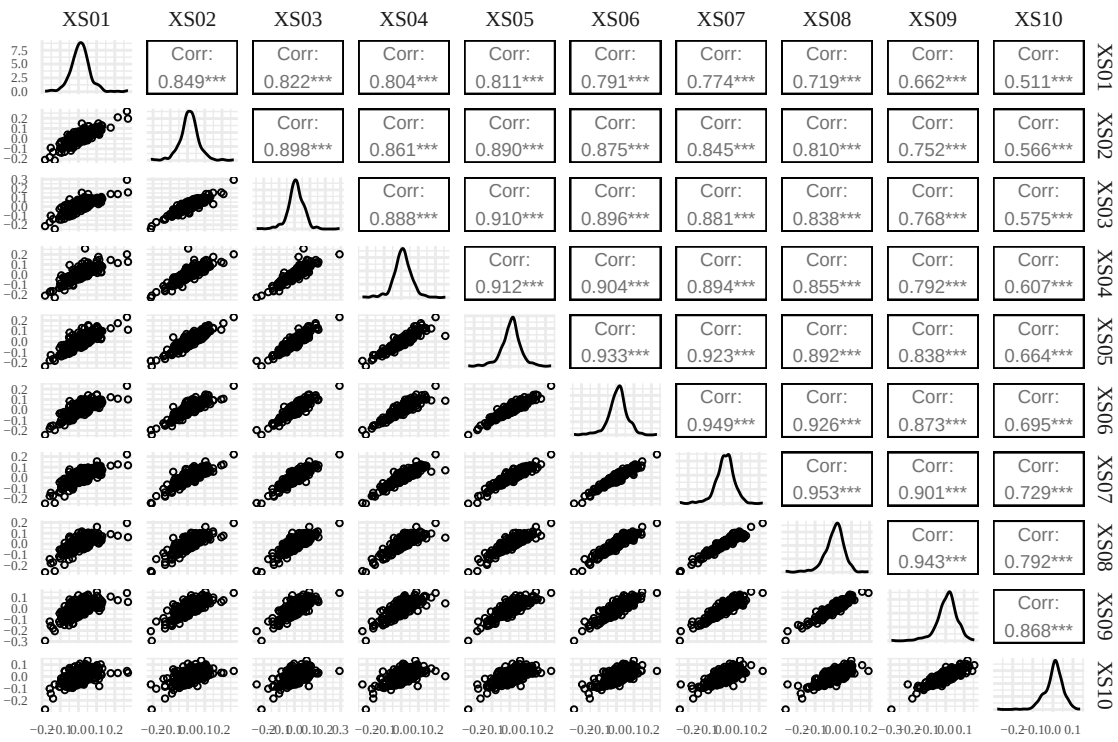


Figure 4: Correlations between factors

```
ds %>%
  select(c(umd, hml , smb, rmrf)) %>%
  cor() %>%
  ggcorrplot::ggcorrplot(hc.order = FALSE,
             type = "full",
             lab = TRUE,
             digits = 2,
             lab_size = 2.2) +
  theme(text = element_text(family = "serif"))
```

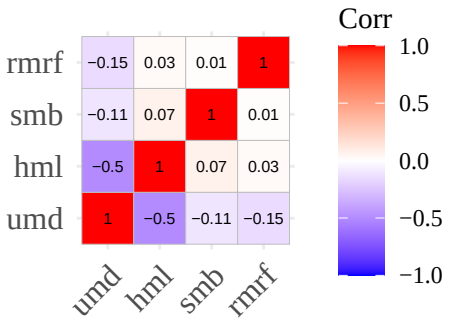
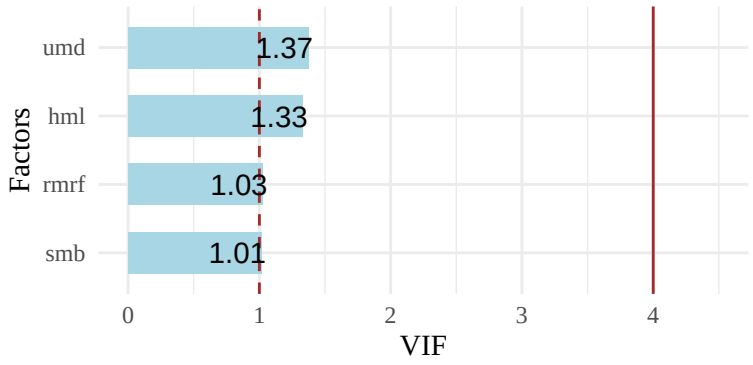


Figure 5: Variance inflation factor

```
vif_tab = vif(lm(XS01~umd+hml+smb+rmrf, data=ds)) %>%
  as.data.frame()

vif_tab = vif_tab %>%
  rename(vif = '.') %>%
  mutate(factors = row.names(vif_tab))

theme_set(theme_minimal(base_family = 'serif'))
vif_tab %>%
  mutate(factors = reorder(factors,vif), decreasing = F) %>%
  ggplot()+
  geom_col(aes(y = factors, x = vif), fill = '#A9D6E5', width = 0.6)+
  theme(plot.title = element_text(hjust = 0.5, size = 12)) +
  xlim(0, 4.5)+
  ylab("Factors")+
  xlab("VIF")+
  geom_vline(xintercept = 4, color = "brown", linetype = "solid") +
  geom_vline(xintercept = 1, color = "brown", linetype = "dashed") +
  geom_text(aes(y = factors, x = vif, label = round(vif,2)), nudge_x=-0.18)
```



2. Multifactor Regressions

```
ptf.ts = ts(ds[,8:18], start = c(1980,10), frequency = 12)

ptf_lm = lm(cbind(XS01,XS02,XS03,XS04,XS05,XS06,XS07,XS08,XS09,XS10)~rmrf+smb+hml+umd, data = ds)

#interpreting coefficient, t-values, p-values

# storing slope coefficients
coefficients.ds = data.frame(ptf_lm$coefficients)
coefficients.ds = t(coefficients.ds)

#coefficients with sderr, t-values and p-values robust to heteroskedasticity
#this code is in order to obtain dataframe with everything we need:
#we have to export it and then we will save the estimates of the coefficients
#and use those as regressors on the 10 ptfs
coefficients_robust.ds = coeftest(ptf_lm, vcovHC = vcovHC(ptf_lm))
estimates = as.vector(coefficients_robust.ds[,1])
std_err = as.vector(coefficients_robust.ds[,2])
t_values = as.vector(coefficients_robust.ds[,3])
p_values = as.vector(coefficients_robust.ds[,4])

XS_vec = paste0("XS", c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5),
                rep(6,5),rep(7,5),rep(8,5),rep(9,5),rep(10,5)))
xs_values = paste0(XS_vec, c(": Intercept",
                  ": rmrf",
                  ": smb",
                  ": hml",
                  ": umd"))

coefficients_robust.ds_final = data.frame(ptf_variables = xs_values,
                                          coefficient = estimates,
                                          std_error = std_err,
                                          t_value = t_values,
                                          p_value = p_values)
# getting r2
s = summary(ptf_lm)
r2_XS1 = s$`Response XS1`$r.squared
adj_r2_XS1 = s$`Response XS1`$adj.r.squared
r2_XS2 = s$`Response XS2`$r.squared
adj_r2_XS2 = s$`Response XS2`$adj.r.squared
r2_XS3 = s$`Response XS3`$r.squared
adj_r2_XS3 = s$`Response XS3`$adj.r.squared
r2_XS4 = s$`Response XS4`$r.squared
adj_r2_XS4 = s$`Response XS4`$adj.r.squared
r2_XS5 = s$`Response XS5`$r.squared
adj_r2_XS5 = s$`Response XS5`$adj.r.squared
r2_XS6 = s$`Response XS6`$r.squared
adj_r2_XS6 = s$`Response XS6`$adj.r.squared
r2_XS7 = s$`Response XS7`$r.squared
adj_r2_XS7 = s$`Response XS7`$adj.r.squared
r2_XS8 = s$`Response XS8`$r.squared
adj_r2_XS8 = s$`Response XS8`$adj.r.squared
r2_XS9 = s$`Response XS9`$r.squared
```

```
adj_r2_XS9 = s$`Response XS9`$adj.r.squared
r2_XS10 = s$`Response XS10`$r.squared
adj_r2_XS10 = s$`Response XS10`$adj.r.squared

ptfs_r2 = c(rep(r2_XS1, 5), rep(r2_XS2, 5), rep(r2_XS3, 5), rep(r2_XS4, 5), rep(r2_XS5, 5),
    rep(r2_XS6, 5), rep(r2_XS7, 5), rep(r2_XS8, 5), rep(r2_XS9, 5), rep(r2_XS10, 5))

ptfs_adj_r2 = c(rep(adj_r2_XS1, 5), rep(adj_r2_XS2, 5), rep(adj_r2_XS3, 5), rep(adj_r2_XS4, 5),
                rep(adj_r2_XS5, 5), rep(adj_r2_XS6, 5), rep(adj_r2_XS7, 5), rep(adj_r2_XS8, 5),
                rep(adj_r2_XS9, 5), rep(adj_r2_XS10, 5))

coefficients_robust.ds_final$r2 = ptfs_r2
coefficients_robust.ds_final$adj_r2 = ptfs_adj_r2
```

```
# separated regression
ptf_lm1 = lm(XS01~rmrf+smb+hml+umd, data = ds)
ptf_lm2 = lm(XS02~rmrf+smb+hml+umd, data = ds)
ptf_lm3 = lm(XS03~rmrf+smb+hml+umd, data = ds)
ptf_lm4 = lm(XS04~rmrf+smb+hml+umd, data = ds)
ptf_lm5 = lm(XS05~rmrf+smb+hml+umd, data = ds)
ptf_lm6 = lm(XS06~rmrf+smb+hml+umd, data = ds)
ptf_lm7 = lm(XS07~rmrf+smb+hml+umd, data = ds)
ptf_lm8 = lm(XS08~rmrf+smb+hml+umd, data = ds)
ptf_lm9 = lm(XS09~rmrf+smb+hml+umd, data = ds)
ptf_lm10 = lm(XS10~rmrf+smb+hml+umd, data = ds)

library(stargazer)
```

Table 2: Multifactor Regressions Results (part b)

	Dependent variable:				
	XS01	XS02	XS03	XS04	XS05
rmrf	0.670^{***}	0.747^{***}	0.746^{***}	0.808^{***}	0.864^{***}
	(0.039)	(0.034)	(0.029)	(0.026)	(0.024)
smb	0.833^{***}	0.896^{***}	0.929^{***}	1.005^{***}	0.904^{***}
	(0.054)	(0.046)	(0.040)	(0.036)	(0.033)
hml	0.104^{*}	0.055	0.073^{*}	-0.106^{***}	0.041
	(0.056)	(0.048)	(0.041)	(0.037)	(0.035)
umd	0.095^{*}	-0.030	-0.002	0.087^{**}	0.054^{*}
	(0.051)	(0.043)	(0.038)	(0.034)	(0.032)
Constant	0.005^{***}	0.006^{***}	0.004^{**}	0.002^{*}	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Observations	363	363	363	363	363
R^2	0.602	0.719	0.778	0.830	0.850
Adjusted R^2	0.598	0.716	0.776	0.828	0.848
Residual Std. Error (df = 358)	0.034	0.029	0.025	0.023	0.021
F Statistic (df = 4; 358)	135.549^{***}	229.535^{***}	314.474^{***}	436.290^{***}	505.626^{***}
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 3: Multifactor Regressions Results (part b)

	Dependent variable:				
	XS06	XS07	XS08	XS09	XS10
rmrf	0.894***	0.925***	0.996***	1.076***	0.955***
	(0.021)	(0.019)	(0.017)	(0.018)	(0.010)
smb	0.883***	0.845***	0.725***	0.451***	−0.150***
	(0.029)	(0.026)	(0.023)	(0.024)	(0.014)
hml	0.021	−0.016	−0.034	−0.009	−0.008
	(0.030)	(0.027)	(0.024)	(0.025)	(0.014)
umd	0.014	0.009	0.046**	−0.057**	0.047***
	(0.027)	(0.025)	(0.022)	(0.023)	(0.013)
Constant	0.002*	0.001	0.0005	0.001	−0.00002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)
Observations	363	363	363	363	363
R^2	0.888	0.907	0.926	0.922	0.963
Adjusted R^2	0.886	0.906	0.925	0.921	0.962
Residual Std. Error (df = 358)	0.018	0.017	0.015	0.016	0.009
F Statistic (df = 4; 358)	706.798***	876.378***	1,117.527***	1,055.308***	2,321.720***
Note:					*p<0.1; **p<0.05; ***p<0.01

Figure 6: Regression coefficients plot (XS01-XS05)

```
plot_summs(ptf_lm1, ptf_lm2, ptf_lm3, ptf_lm4, ptf_lm5,
  robust = TRUE,
  model.names = c(str_c("XS0", 1:5)),
  point.shape = FALSE,
  point.size = 4,
  colors = c("#012A4A", "#01497C", "#2A6F97", "#61A5C2", "#A9D6E5"))+
  theme(text = element_text(family = "serif", colour="black"),
        legend.title = element_text(family = "serif", colour = "black"),
        axis.title = element_text(family = "serif", colour="black"),
        axis.text.x = element_text(family = "serif", colour = "black"),
        axis.text.y = element_text(family = "serif", colour="black")
  )
```

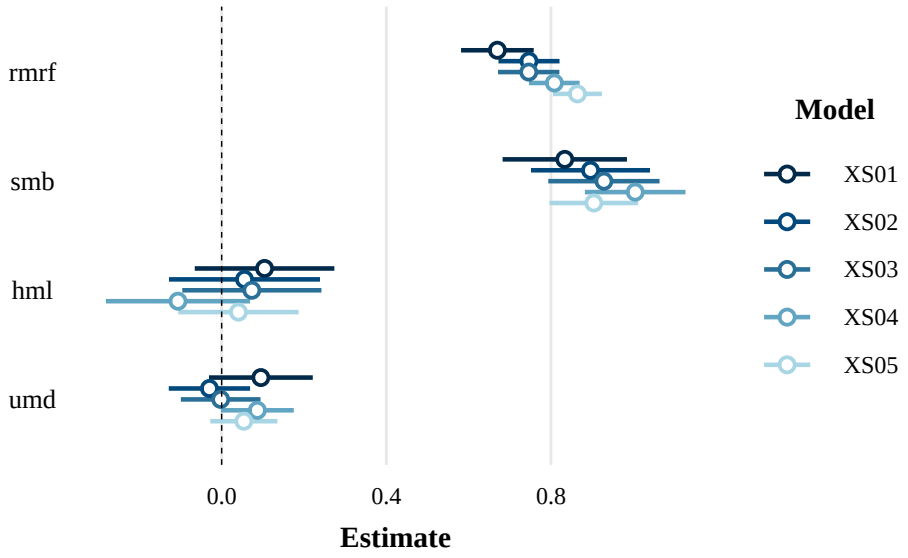


Figure 7: Regression coefficients plot (XS06-XS10)

```
plot_summs(ptf_lm6,
           ptf_lm7, ptf_lm8,
           ptf_lm9, ptf_lm10,
           robust = TRUE,
           point.shape = F,
           point.size = 4,
           colors = c("#012A4A", "#01497C", "#2A6F97", "#61A5C2", "#A9D6E5"),
           model.names = c(str_c("XS0", 6:9), "XS10"))+
  theme(text = element_text(family = "serif", colour="black"),
        legend.title = element_text(family = "serif", colour = "black"),
        axis.title = element_text(family = "serif", colour="black"),
        axis.text.x = element_text(family = "serif", colour = "black"),
        axis.text.y = element_text(family = "serif", colour="black")
  )
```

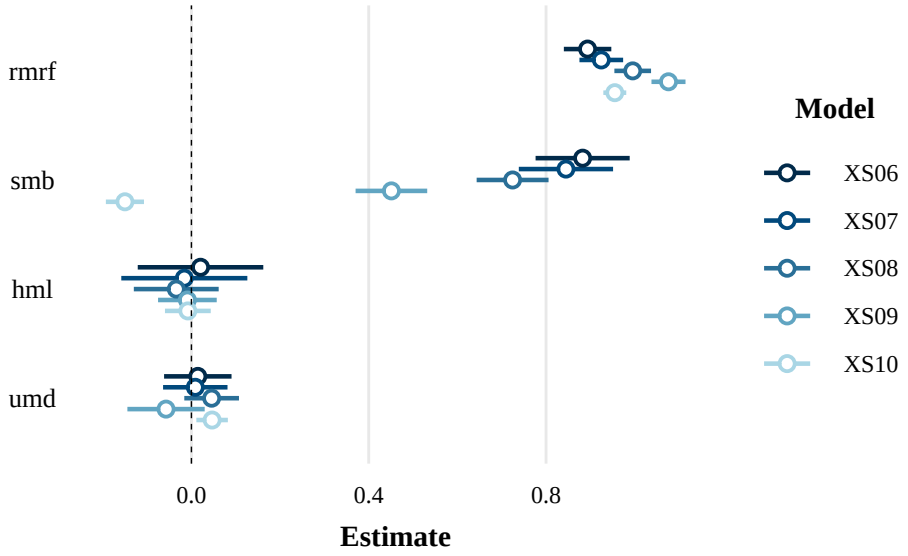


Table 4: Partial R-squared

```
XS01_lm = lm(XS01~rmrf+smb+hml+umd, data = ds)
XS02_lm = lm(XS02~rmrf+smb+hml+umd, data = ds)
XS03_lm = lm(XS03~rmrf+smb+hml+umd, data = ds)
XS04_lm = lm(XS04~rmrf+smb+hml+umd, data = ds)
XS05_lm = lm(XS05~rmrf+smb+hml+umd, data = ds)
XS06_lm = lm(XS06~rmrf+smb+hml+umd, data = ds)
XS07_lm = lm(XS07~rmrf+smb+hml+umd, data = ds)
XS08_lm = lm(XS08~rmrf+smb+hml+umd, data = ds)
XS09_lm = lm(XS09~rmrf+smb+hml+umd, data = ds)
XS10_lm = lm(XS10~rmrf+smb+hml+umd, data = ds)

XS01_lm_noRMRF = lm(XS01~smb+hml+umd, data = ds)
XS02_lm_noRMRF = lm(XS02~smb+hml+umd, data = ds)
XS03_lm_noRMRF = lm(XS03~smb+hml+umd, data = ds)
XS04_lm_noRMRF = lm(XS04~smb+hml+umd, data = ds)
XS05_lm_noRMRF = lm(XS05~smb+hml+umd, data = ds)
XS06_lm_noRMRF = lm(XS06~smb+hml+umd, data = ds)
XS07_lm_noRMRF = lm(XS07~smb+hml+umd, data = ds)
XS08_lm_noRMRF = lm(XS08~smb+hml+umd, data = ds)
XS09_lm_noRMRF = lm(XS09~smb+hml+umd, data = ds)
XS10_lm_noRMRF = lm(XS10~smb+hml+umd, data = ds)

XS01_lm_noSMB = lm(XS01~rmrf+hml+umd, data = ds)
XS02_lm_noSMB = lm(XS02~rmrf+hml+umd, data = ds)
XS03_lm_noSMB = lm(XS03~rmrf+hml+umd, data = ds)
XS04_lm_noSMB = lm(XS04~rmrf+hml+umd, data = ds)
XS05_lm_noSMB = lm(XS05~rmrf+hml+umd, data = ds)
XS06_lm_noSMB = lm(XS06~rmrf+hml+umd, data = ds)
XS07_lm_noSMB = lm(XS07~rmrf+hml+umd, data = ds)
XS08_lm_noSMB = lm(XS08~rmrf+hml+umd, data = ds)
XS09_lm_noSMB = lm(XS09~rmrf+hml+umd, data = ds)
XS10_lm_noSMB = lm(XS10~rmrf+hml+umd, data = ds)

XS01_lm_noHML = lm(XS01~smb+rmrf+umd, data = ds)
XS02_lm_noHML = lm(XS02~smb+rmrf+umd, data = ds)
XS03_lm_noHML = lm(XS03~smb+rmrf+umd, data = ds)
XS04_lm_noHML = lm(XS04~smb+rmrf+umd, data = ds)
XS05_lm_noHML = lm(XS05~smb+rmrf+umd, data = ds)
XS06_lm_noHML = lm(XS06~smb+rmrf+umd, data = ds)
XS07_lm_noHML = lm(XS07~smb+rmrf+umd, data = ds)
XS08_lm_noHML = lm(XS08~smb+rmrf+umd, data = ds)
XS09_lm_noHML = lm(XS09~smb+rmrf+umd, data = ds)
XS10_lm_noHML = lm(XS10~smb+rmrf+umd, data = ds)


XS01_lm_noUMD = lm(XS01~smb+hml+rmrf, data = ds)
XS02_lm_noUMD = lm(XS02~smb+hml+rmrf, data = ds)
XS03_lm_noUMD = lm(XS03~smb+hml+rmrf, data = ds)
XS04_lm_noUMD = lm(XS04~smb+hml+rmrf, data = ds)
XS05_lm_noUMD = lm(XS05~smb+hml+rmrf, data = ds)
XS06_lm_noUMD = lm(XS06~smb+hml+rmrf, data = ds)
XS07_lm_noUMD = lm(XS07~smb+hml+rmrf, data = ds)
```

```
XS08_lm_noUMD = lm(XS08~smb+hml+rmrf, data = ds)
XS09_lm_noUMD = lm(XS09~smb+hml+rmrf, data = ds)
XS10_lm_noUMD = lm(XS10~smb+hml+rmrf, data = ds)

partial_R2 = matrix(data = NA, nrow = 10, ncol = 4)

rownames(partial_R2) = c("XS01", "XS02", "XS03", "XS04", "XS05",
                         "XS06", "XS07", "XS08", "XS09", "XS10")
colnames(partial_R2) = c("rmrf", "smb", "hml", "umd")

library("asbio")
XS01_lm = lm(XS01~rmrf+smb+hml+umd, data = ds)
XS02_lm = lm(XS02~rmrf+smb+hml+umd, data = ds)
XS03_lm = lm(XS03~rmrf+smb+hml+umd, data = ds)
XS04_lm = lm(XS04~rmrf+smb+hml+umd, data = ds)
XS05_lm = lm(XS05~rmrf+smb+hml+umd, data = ds)
XS06_lm = lm(XS06~rmrf+smb+hml+umd, data = ds)
XS07_lm = lm(XS07~rmrf+smb+hml+umd, data = ds)
XS08_lm = lm(XS08~rmrf+smb+hml+umd, data = ds)
XS09_lm = lm(XS09~rmrf+smb+hml+umd, data = ds)
XS10_lm = lm(XS10~rmrf+smb+hml+umd, data = ds)

XS01_lm_noRMRF = lm(XS01~smb+hml+umd, data = ds)
XS02_lm_noRMRF = lm(XS02~smb+hml+umd, data = ds)
XS03_lm_noRMRF = lm(XS03~smb+hml+umd, data = ds)
XS04_lm_noRMRF = lm(XS04~smb+hml+umd, data = ds)
XS05_lm_noRMRF = lm(XS05~smb+hml+umd, data = ds)
XS06_lm_noRMRF = lm(XS06~smb+hml+umd, data = ds)
XS07_lm_noRMRF = lm(XS07~smb+hml+umd, data = ds)
XS08_lm_noRMRF = lm(XS08~smb+hml+umd, data = ds)
XS09_lm_noRMRF = lm(XS09~smb+hml+umd, data = ds)
XS10_lm_noRMRF = lm(XS10~smb+hml+umd, data = ds)


parial.R2_XS01_noRMRF = partial.R2(XS01_lm_noRMRF, XS01_lm)
parial.R2_XS02_noRMRF = partial.R2(XS02_lm_noRMRF, XS02_lm)
parial.R2_XS03_noRMRF = partial.R2(XS03_lm_noRMRF, XS03_lm)
parial.R2_XS04_noRMRF = partial.R2(XS04_lm_noRMRF, XS04_lm)
parial.R2_XS05_noRMRF = partial.R2(XS05_lm_noRMRF, XS05_lm)
parial.R2_XS06_noRMRF = partial.R2(XS06_lm_noRMRF, XS06_lm)
parial.R2_XS07_noRMRF = partial.R2(XS07_lm_noRMRF, XS07_lm)
parial.R2_XS08_noRMRF = partial.R2(XS08_lm_noRMRF, XS08_lm)
parial.R2_XS09_noRMRF = partial.R2(XS09_lm_noRMRF, XS09_lm)
parial.R2_XS10_noRMRF = partial.R2(XS10_lm_noRMRF, XS10_lm)

XS01_lm_noSMB = lm(XS01~rmrf+hml+umd, data = ds)
XS02_lm_noSMB = lm(XS02~rmrf+hml+umd, data = ds)
XS03_lm_noSMB = lm(XS03~rmrf+hml+umd, data = ds)
XS04_lm_noSMB = lm(XS04~rmrf+hml+umd, data = ds)
XS05_lm_noSMB = lm(XS05~rmrf+hml+umd, data = ds)
XS06_lm_noSMB = lm(XS06~rmrf+hml+umd, data = ds)
XS07_lm_noSMB = lm(XS07~rmrf+hml+umd, data = ds)
XS08_lm_noSMB = lm(XS08~rmrf+hml+umd, data = ds)
```

```
XS09_lm_noSMB = lm(XS09~rmrf+hml+umd, data = ds)
XS10_lm_noSMB = lm(XS10~rmrf+hml+umd, data = ds)


parial.R2_XS01_noSMB = partial.R2(XS01_lm_noSMB, XS01_lm)
parial.R2_XS02_noSMB = partial.R2(XS02_lm_noSMB, XS02_lm)
parial.R2_XS03_noSMB = partial.R2(XS03_lm_noSMB, XS03_lm)
parial.R2_XS04_noSMB = partial.R2(XS04_lm_noSMB, XS04_lm)
parial.R2_XS05_noSMB = partial.R2(XS05_lm_noSMB, XS05_lm)
parial.R2_XS06_noSMB = partial.R2(XS06_lm_noSMB, XS06_lm)
parial.R2_XS07_noSMB = partial.R2(XS07_lm_noSMB, XS07_lm)
parial.R2_XS08_noSMB = partial.R2(XS08_lm_noSMB, XS08_lm)
parial.R2_XS09_noSMB = partial.R2(XS09_lm_noSMB, XS09_lm)
parial.R2_XS10_noSMB = partial.R2(XS10_lm_noSMB, XS10_lm)

XS01_lm_noHML = lm(XS01~smb+rmrf+umd, data = ds)
XS02_lm_noHML = lm(XS02~smb+rmrf+umd, data = ds)
XS03_lm_noHML = lm(XS03~smb+rmrf+umd, data = ds)
XS04_lm_noHML = lm(XS04~smb+rmrf+umd, data = ds)
XS05_lm_noHML = lm(XS05~smb+rmrf+umd, data = ds)
XS06_lm_noHML = lm(XS06~smb+rmrf+umd, data = ds)
XS07_lm_noHML = lm(XS07~smb+rmrf+umd, data = ds)
XS08_lm_noHML = lm(XS08~smb+rmrf+umd, data = ds)
XS09_lm_noHML = lm(XS09~smb+rmrf+umd, data = ds)
XS10_lm_noHML = lm(XS10~smb+rmrf+umd, data = ds)

parial.R2_XS01_noHML = partial.R2(XS01_lm_noHML, XS01_lm)
parial.R2_XS02_noHML = partial.R2(XS02_lm_noHML, XS02_lm)
parial.R2_XS03_noHML = partial.R2(XS03_lm_noHML, XS03_lm)
parial.R2_XS04_noHML = partial.R2(XS04_lm_noHML, XS04_lm)
parial.R2_XS05_noHML = partial.R2(XS05_lm_noHML, XS05_lm)
parial.R2_XS06_noHML = partial.R2(XS06_lm_noHML, XS06_lm)
parial.R2_XS07_noHML = partial.R2(XS07_lm_noHML, XS07_lm)
parial.R2_XS08_noHML = partial.R2(XS08_lm_noHML, XS08_lm)
parial.R2_XS09_noHML = partial.R2(XS09_lm_noHML, XS09_lm)
parial.R2_XS10_noHML = partial.R2(XS10_lm_noHML, XS10_lm)

XS01_lm_noUMD = lm(XS01~smb+hml+rmrf, data = ds)
XS02_lm_noUMD = lm(XS02~smb+hml+rmrf, data = ds)
XS03_lm_noUMD = lm(XS03~smb+hml+rmrf, data = ds)
XS04_lm_noUMD = lm(XS04~smb+hml+rmrf, data = ds)
XS05_lm_noUMD = lm(XS05~smb+hml+rmrf, data = ds)
XS06_lm_noUMD = lm(XS06~smb+hml+rmrf, data = ds)
XS07_lm_noUMD = lm(XS07~smb+hml+rmrf, data = ds)
XS08_lm_noUMD = lm(XS08~smb+hml+rmrf, data = ds)
XS09_lm_noUMD = lm(XS09~smb+hml+rmrf, data = ds)
XS10_lm_noUMD = lm(XS10~smb+hml+rmrf, data = ds)

parial.R2_XS01_noUMD = partial.R2(XS01_lm_noUMD, XS01_lm)
parial.R2_XS02_noUMD = partial.R2(XS02_lm_noUMD, XS02_lm)
parial.R2_XS03_noUMD = partial.R2(XS03_lm_noUMD, XS03_lm)
parial.R2_XS04_noUMD = partial.R2(XS04_lm_noUMD, XS04_lm)
parial.R2_XS05_noUMD = partial.R2(XS05_lm_noUMD, XS05_lm)
```

```
parial.R2_XS06_noUMD = partial.R2(XS06_lm_noUMD, XS06_lm)
parial.R2_XS07_noUMD = partial.R2(XS07_lm_noUMD, XS07_lm)
parial.R2_XS08_noUMD = partial.R2(XS08_lm_noUMD, XS08_lm)
parial.R2_XS09_noUMD = partial.R2(XS09_lm_noUMD, XS09_lm)
parial.R2_XS10_noUMD = partial.R2(XS10_lm_noUMD, XS10_lm)

partial_R2 = matrix(data = NA, nrow = 10, ncol = 4)

rownames(partial_R2) = c("XS01", "XS02", "XS03", "XS04", "XS05",
                         "XS06", "XS07", "XS08", "XS09", "XS10")
colnames(partial_R2) = c("rmrf", "smb", "hml", "umd")

partial_R2[,1] = c(parial.R2_XS01_noRMRF,
                   parial.R2_XS02_noRMRF,
                   parial.R2_XS03_noRMRF,
                   parial.R2_XS04_noRMRF,
                   parial.R2_XS05_noRMRF,
                   parial.R2_XS06_noRMRF,
                   parial.R2_XS07_noRMRF,
                   parial.R2_XS08_noRMRF,
                   parial.R2_XS09_noRMRF,
                   parial.R2_XS10_noRMRF)

partial_R2[,2] = c(parial.R2_XS01_noSMB,
                   parial.R2_XS02_noSMB,
                   parial.R2_XS03_noSMB,
                   parial.R2_XS04_noSMB,
                   parial.R2_XS05_noSMB,
                   parial.R2_XS06_noSMB,
                   parial.R2_XS07_noSMB,
                   parial.R2_XS08_noSMB,
                   parial.R2_XS09_noSMB,
                   parial.R2_XS10_noSMB)

partial_R2[,3] = c(parial.R2_XS01_noHML,
                   parial.R2_XS02_noHML,
                   parial.R2_XS03_noHML,
                   parial.R2_XS04_noHML,
                   parial.R2_XS05_noHML,
                   parial.R2_XS06_noHML,
                   parial.R2_XS07_noHML,
                   parial.R2_XS08_noHML,
                   parial.R2_XS09_noHML,
                   parial.R2_XS10_noHML)

partial_R2[,4] = c(parial.R2_XS01_noUMD,
                   parial.R2_XS02_noUMD,
                   parial.R2_XS03_noUMD,
                   parial.R2_XS04_noUMD,
                   parial.R2_XS05_noUMD,
                   parial.R2_XS06_noUMD,
                   parial.R2_XS07_noUMD,
                   parial.R2_XS08_noUMD,
```

```
                    parial.R2_XS09_noUMD,
                    parial.R2_XS10_noUMD)

kable(partial_R2, caption = 'Partial R-squared', align = 'l', digits=3)
```

Table 4: Partial R-squared

	rmrf	smb	hml	umd
XS01	0.450	0.404	0.010	0.010
XS02	0.581	0.516	0.004	0.001
XS03	0.649	0.605	0.009	0.000
XS04	0.726	0.687	0.022	0.018
XS05	0.778	0.672	0.004	0.008
XS06	0.835	0.725	0.001	0.001
XS07	0.869	0.747	0.001	0.000
XS08	0.906	0.731	0.006	0.012
XS09	0.912	0.492	0.000	0.017
XS10	0.962	0.250	0.001	0.035

3. Cross-Section Regressions

```
#cross section
x_sect_reg.df = t(ptf_lm$coefficients) %>% as.data.frame()

#creating 10 mean returns
ds_with_means = ds %>% mutate(across(S1:S10, ~mean(.)))
s_means = ds_with_means %>% dplyr::select(S1:S10) %>% slice_head() %>% t()

x_sect_reg.df = x_sect_reg.df %>% mutate(mean_returns = s_means)

# first model
model <- lm(mean_returns ~ rmrf + smb, data = x_sect_reg.df)
model_summary <- summary(model)

# coefficients and p-values
coefficients <- model_summary$coefficients
p_values <- coefficients[, 4]

# table
results_table <- data.frame(
  Coefficient = rownames(coefficients),
  Estimate = coefficients[, 1],
  Std_Error = coefficients[, 2],
  t_value = coefficients[, 3],
  p_value = p_values,
  r2=model_summary$r.squared,
  adj_r2=model_summary$adj.r.squared
)

# mod 2
model2 = lm(mean_returns~rmrf+hml, data = x_sect_reg.df)
model2_summary <- summary(model2)
```

```r
# coefficients and p-values
coefficients <- model2_summary$coefficients
p_values <- coefficients[, 4]

# table
results_table2 <- data.frame(
  Coefficient = rownames(coefficients),
  Estimate = coefficients[, 1],
  Std_Error = coefficients[, 2],
  t_value = coefficients[, 3],
  p_value = p_values,
  r2=model2_summary$r.squared,
  adj_r2=model2_summary$adj.r.squared
)

# mod 3
model3 = lm(mean_returns~rmrf+umd, data = x_sect_reg.df)
model3_summary <- summary(model3)

# coefficients and p-values
coefficients <- model3_summary$coefficients
p_values <- coefficients[, 4]

# table
results_table3 <- data.frame(
  Coefficient = rownames(coefficients),
  Estimate = coefficients[, 1],
  Std_Error = coefficients[, 2],
  t_value = coefficients[, 3],
  p_value = p_values,
  r2=model3_summary$r.squared,
  adj_r2=model3_summary$adj.r.squared
)

final_result=rbind(results_table, results_table2,results_table3) %>% select(Estimate:adj_r2)
```

Table 5: Cross-section regression results

```r
stargazer(model, model2, model3,
          title="Cross-section Regressions Results",
          type="latex",
          header=F,
          covariate.labels = c('coeff rmrf','coeff smb','coeff hml','coeff umd','constant'),
          #style="all", #to include in the output t-stats and p-values, but table too big
          model.numbers = F,
          no.space = T, font.size = "small")
```

Table 5: Cross-section Regressions Results

	Dependent variable:		
	mean_returns		
coeff rmrf	-0.010^{**}	-0.011^{**}	-0.014^{***}
	(0.003)	(0.003)	(0.003)
coeff smb	0.001		
	(0.001)		
coeff hml		0.006	
		(0.007)	
coeff umd			-0.011
			(0.007)
constant	0.022^{***}	0.023^{***}	0.026^{***}
	(0.003)	(0.003)	(0.002)
Observations	10	10	10
R^2	0.754	0.740	0.789
Adjusted R^2	0.684	0.665	0.729
Residual Std. Error (df = 7)	0.001	0.001	0.001
F Statistic (df = 2; 7)	10.728^{***}	9.944^{***}	13.081^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01