Problem Background

In this lab we are going to start with a one-factor CAPM model, and then extend it to the three-factor Fama-French model.

We will use the data set for Stocks, FX and Bonds from 2004-2005.

Data for the Fama-French factors are avilable Prof. Kenneth French's website.

Where RF is the risk-free rate and Mkt.RF, SMB and HML are the Fama-French factors.

```
# Stock/Bond/FX data.
stocks <- as.data.table(read.csv(paste0(data.dir, "Stock_FX_Bond_2004_to_2005.csv"),
                                 header=T))
stocks$Date <- as.Date(stocks$Date, format = "%d-%b-%y")
stocks_subset <- stocks[, .(Date, GM_AC, F_AC, UTX_AC, MRK_AC)]
stocks_diff <- data.table(Date = stocks_subset[-1]$Date,
                          100 * apply(log(stocks_subset[, .(GM_AC, F_AC, UTX_AC, MRK_AC)]), 2,

# Fama-French data.
FF_data <- as.data.table(read.table(paste0(data.dir, "FamaFrenchDaily.txt"),
                                    header=T))
FF_data$Date <- as.Date(as.character(FF_data$date), format = "%Y%m%d")
FF_data$date <- NULL

# Combine into one data.table.
consolidated.data <- merge(stocks_diff, FF_data, on = c("Date"))
capm.data <- consolidated.data[, .(Date,
                                   GM = GM_AC - RF,
                                   Ford = F_AC - RF,
                                   UTX = UTX_AC - RF,
                                   Merck = MRK_AC - RF,
                                   Mkt.RF, SMB, HML, RF)]


fit <- lm(as.matrix(cbind(GM, Ford, UTX, Merck))
                    ~ Mkt.RF + SMB + HML, data = capm.data)
```

1.) Which one(s) of the 4 stocks are "value" stocks, according to this model? How can you tell?

Table 1: Fama-French Model

Factor	GM	Ford	UTX	Merck
Alpha	-0.25114	-0.19508	0.00000	-0.05983
Mkt.RF	1.38891	1.35115	1.02862	0.70927
SMB	-0.25044	-0.01570	-0.29268	-0.41740
HML	0.60056	0.34122	-0.00096	-0.95592

A "value" stock is one with a positive HML (high minus low) coefficent, which would be **GM** and **Ford**. **Merck** looks like a *"growth"* stock, and UTX is somewhere in between a "growth" and "value" stock.

2.) Fama-French model fits which stock best? Worst?

```
lm.stats <- summary(fit)

dt.stats <- data.table(stat = c("R-Squared", "Adj. R-Squared"),
            GM = c(lm.stats$`Response GM`$r.squared,
                   lm.stats$`Response GM`$adj.r.squared),
            Ford = c(lm.stats$`Response Ford`$r.squared,
                     lm.stats$`Response Ford`$adj.r.squared),
            UTX = c(lm.stats$`Response UTX`$r.squared,
                    lm.stats$`Response UTX`$adj.r.squared),
            Merck = c(lm.stats$`Response Merck`$r.squared,
                      lm.stats$`Response Merck`$adj.r.squared))
dt.stats[, 2:5] <- dt.stats[, 2:5] * 100
pretty_kable(dt.stats, "Model Fit Statistics", dig = 2)
```

Table 2: Model Fit Statistics

stat	GM	Ford	UTX	Merck
R-Squared	17.33	26.40	35.16	6.60
Adj. R-Squared	16.83	25.95	34.77	6.04

The Fama-French model fits **UTX** the best, explaining roughtly 35% of the variance of the returns, and **Merck** the worst, explaining about 6.6% of the return variance.

3.) For UTX, which factor is most important besides the market return?

For **UTX** the most important factor outside the market return is SMB, or small minus big, at *-30%*. A strong negative relationship with SMB means it behaves, unsurprisingly, like a large cap stock.

4.) Use the Fama-French model to predict the next day return of UTX.

Use the first 250 days (**the training set**) to fit the model. Note that this model will be different from the one above. Use this model to make 1-day forecasts for the rest of the data (**the test set**).

```
utx.data <- capm.data[, .(Date, UTX, Mkt.RF, SMB, HML)]
train.size <- 250
test.size <- nrow(utx.data) - train.size

train.data <- utx.data[1:train.size,]
test.data <- utx.data[train.size+1:test.size,]

stopifnot(nrow(train.data) == train.size & nrow(test.data) == test.size)

summary(utx.fit <- lm(UTX ~ Mkt.RF + SMB + HML,
                      data = train.data))
```

```
Call:
lm(formula = UTX ~ Mkt.RF + SMB + HML, data = train.data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2594 -0.5651  0.0086  0.5294  3.2582

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01061    0.05900  -0.180   0.8575
Mkt.RF       0.91550    0.10106   9.059   <2e-16 ***
SMB         -0.13925    0.14173  -0.983   0.3268
HML          0.29929    0.16424   1.822   0.0696 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9266 on 246 degrees of freedom
Multiple R-squared:  0.2998,    Adjusted R-squared:  0.2913
F-statistic: 35.11 on 3 and 246 DF,  p-value: < 2.2e-16
```

- What is the percentage of days in the test set where the sign of the forecast return agrees with that of the actual return? (sign=-1,0,1)

```
# Use the 1 day lagged prediction.
# i.e, the sign of 1/3/05 close is using the FF data from 12/31/04.
results <- data.table(Actual = test.data[-1]$UTX,
                      Pred = predict(utx.fit, newdata = head(test.data, -1)))
results[, correct := sign(Actual) == sign(Pred)]
results$Date <- test.data[-1]$Date

pct.correct <- (nrow(results[correct == T]) / nrow(results)) * 100
```

Using the 3-factor model, we can correctly predict the sign of the next day return: **46.83%**

```
suppressWarnings(print({
   ggplot(results, aes(x = Date)) +
      geom_line(aes(y = Actual)) +
      geom_line(aes(y = ifelse(correct == T, Actual, NA)), col = "darkgreen") +
      geom_line(aes(y = ifelse(correct == F, Actual, NA)), col = "darkred") +
      labs(title = "Fama-French 3-Factor Prediction")
}))
```