**Financial Informatics HW2: Forecast future return (with regression analysis)**

2013310443 최재필

* **Data used:** 
  + S&P500 data imported from Quantmod library (2015.01.01 ~ 2018.11.13)
  + Daily Fama-French 3 Factors data downloaded from Kenneth R. French blog. (1926.07.01 ~ 2018.09.28)

1. **Workflow (each step is thoroughly explained in the R script.)**
   1. Download FF3 data from the website and import current S&P500 data from Quantmod. (Both are daily data)
   2. Set up the investment universe like before. Select top 100 stocks with the highest annual returns. This is my portfolio.
   3. Calculate daily returns of the equally weighted portfolio.
   4. Merge daily portfolio return table with FF3 data.
   5. Standardize the scale of FF3 data with the daily return. That is, converting yearly return of Rf and Rm-Rf to daily return for proper analysis. (You can’t add or subtract different scales)
   6. Run linear regression on the final data to forecast return. Give lag on daily return incrementally and interpret the result.
2. **Result of the regression.** 
   1. 1 day lag

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.0009933 0.0005062 1.962 0.050 .

Rm\_Rf\_daily 3.0406385 0.1579523 19.250 <2e-16 \*\*\*

SMB -0.0010558 0.0010347 -1.020 0.308

HML -0.0094495 0.0009781 -9.661 <2e-16 \*\*\*

---

Residual standard error: 0.01551 on 939 degrees of freedom

Multiple R-squared: 0.3297, Adjusted R-squared: 0.3276

F-statistic: 153.9 on 3 and 939 DF, p-value: < 2.2e-16

* + 1. P value of the coefficients show that [Rm-Rf](market risk premium) and [HML](High Book to Market Value – Low Book to Market Value) are statistically significant.
    2. β1: 1 %p increase (i.e, 1 unit of Rm-Rf\_daily) of market risk premium increases 3%p of portfolio risk premium.
    3. β2: 1%p increase in the difference between small market cap stock and big market cap stock (i.e, 1 unit increase in SMB) decreases 0.001%p of portfolio risk premium, virtually making it neutral.
    4. Adjusted R-Square value is 0.3276, meaning that this regression explains roughly 32% of portfolio risk premium’s movement.
    5. The regression’s P-value is near 0, indicating that there is no multicollinearity issue. The model is valid.
  1. 2 day lag

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.0017012 0.0006189 2.749 0.0061 \*\*

Rm\_Rf\_daily -0.1505631 0.1935079 -0.778 0.4367

SMB 0.0009393 0.0012651 0.742 0.4580

HML -0.0007224 0.0011954 -0.604 0.5458

---

Residual standard error: 0.01894 on 937 degrees of freedom

Multiple R-squared: 0.001609, Adjusted R-squared: -0.001587

F-statistic: 0.5034 on 3 and 937 DF, p-value: 0.68

* + 1. P value of the coefficients show that no factor is statistically significant as P value of each coefficient is far above 0.05. The model is not valid.
    2. The model indicators gets worse as I increase the length of the lag.

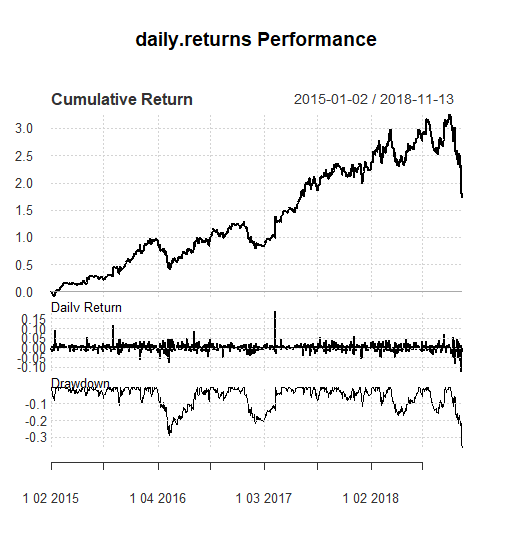
1. **Reviewing the work.** 
   1. What were the limitations?
      1. Hindsight (survivorship) bias. The companies in current S&P500 list are the ‘survivors’, who managed to stay in the list without going out of business. This might explain why the return of the portfolio was too good to be real.   
          

Figure 1. Cumulative return of S&P500 top 100 firms selected at Dec 31, 2015

* + 1. FF3 factors were not enough to explain the alpha of the return. It only explains about 30% of the return. (Additional variables were needed.)
    2. Forecasting return with regression analysis is not enough because it doesn’t take account of other factors like maximum drawdown.
  1. Implications
     1. The lag bigger than a day is long enough to decay market signals. Because the market is efficient and everybody knows FF3 factors, the arbitrage opportunities are quickly exhausted.