Computer Assignment4—Fama-Macbeth Regression

YiTao Hu

03/02/2020

Executive Summary

To estimate the market coefficients on three exposures, namely Market Cap, three-year Price Momentum, and Market Perception (meastured by 1 over Price), I performed a Cross-sectional Regression on the annual returns of 2974 stocks from 1973 to 2018 and a pooled linear Regression with NeweyWest standard errors adjusting for heteroskedasticity and error correlation. The cross-sectional regression results indicate that all the coefficients are statistically significant, while the pooled regression with Newey West standard error implies that the coefficient on the Momentum variable is not significant. Both of the regression show a negative correlation between stock gross From the results above, I concluded that inverse price, which reflect Market Perception,and Market Cap variable have significant explanatory power of the gross stock returns, while mid-term Momentum factor has insignificant explanory power. Futher, I would argue that people's sentiments have substantial impacts on returns for short term but not for mid-term. In other words, I would argue that the market is efficiency for mid or long term, but behaves inefficient during the short term.

Assumptions and Methodologies

For this study, I collected fiscal year end price and market value for all stocks from the CRSP database. Although this collection method may cause time mismatch of returns because companies have different fiscal year ending date, I would argue this impact would be ignorable given the long horizon and large number of samples.Besides Market Cap, the returns and other two variables are computed using the formula below:

$$Return_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

$$Mom_t = \frac{P_t - P_{t-3}}{P_{t-3}}$$

$$P_{inv_t} = \frac{1}{P_t}$$

After computing corresponded independent variable, I performed the first step of cross-sectional regression:

For each time step t from 1 to 18, perform the following multi-variant regression:

$$R_{t,i} = b_{0,t} + b_{cap,t}Cap_{t,i} + b_{Mom,t}Mom_{t,i} + b_{Pinv,t} * P_{invt,i} + e_{t,i}$$

After running the regression for each year, I calculated the coefficient estimates by taking the sample arithmetic mean and corresponded standard errors by taking sample standard deviation:

$$\hat{\beta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_t, \sigma^2(\hat{\beta}) = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{\beta}_t - \hat{\beta})^2$$

To test the statistical significance, I computed the t statistics by dividing the coefficient estimate by corresponded standard errorrs.

$$t - stat = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})}$$

Note one key assumption in cross-sectional regression here is that, although the regression error term $\epsilon_{t,i}$ can be correlated and heteroskedastic across different observations (in fact Fama-MacBeth type regression was designed to adjust this heteroskedasticity), we assume that they are independently identically distributed across time. Otherwise, our standard errors computed with such assumption are underestimated.

For the pooled regression with Newey West standard errors, I melted the panel data into a large data frame with each time step and company as one specific observations. Afterward, run one single regression on all the data.

$$R_i = b_0 + b_{cap,t}Cap_i + b_{Mom}Mom_i + b_{Pinv} * P_{inv,i} + e_i$$

In this case, we do not need to assume i.i.d. residuals because Newey West robust standard errors adjust residual correlation and heteroskedasticity across both time and observations.

Results and Interpretations

From those two Regression outputs, we observe that estimated coefficients on all variables are closely statistically significant at 5% level in cross-sectional regression, while Momentum coefficient is insignificant in pooled regression. I would argue that this discrepancy comes from adjustment for residual correlation and heteroskedasticity across time. Because cross-sectional regression assumes time i.i.d. residuals, it tends to underestimate the standard errors. For this reason, I would consider the pooled Regression results as a more robust estimate of the coefficients. The fact that Mid-term price momentum factor does not have sufficient explanatory power is somewhat intuitive. Although the investors may be irrational in the short run and thus make investment affected by stocks with strong short-term momentum, they would adjust their expectations in the long run. As a result, mid-term price momentum would already be adjusted in the current stock price, and therefore has no explanatory power over concurrent stock returns.

In addition, we also observed negative estimated coefficients on Price inverse and Market Cap variables. Negative relation between stock returns and stock market cap confirms the argument in Fama-French three factor model that size risk is one systematic risk requiring additional risk premium. As a result, small cap stocks tend to outperform large cap stocks.

As for price inverse variable, I intepreate it as the market perceived future return of a particular stock. If the current price level is high, the perceived futuren return of this stock will be low, and vice versa. The negative estimated coefficient on market perception implies that investors sentiment does pose a negative impacts on the stock return, which partly confirmed the argument of behavioral finance. Besides, the negative sign of the estimated coefficient implies the existence of short-term momentum, because it proves that stock returns does not follow a mean-reverse pattern,but further research is still needed.

Code and Tables

```r
#import data and library
library(readr)
library(dplyr)
library(reshape2)
library(lmtest)
library(sandwich)
HW3_Data <- read_csv("~/Desktop/UCLAMFE/Corporate Finance/week5/HW3_Data.csv",
    col_types = cols(GVKEY = col_skip(),
        conm = col_skip(), consol = col_skip(),
        costat = col_skip(), curcd = col_skip(),
        datadate = col_skip(), datafmt = col_skip(),
        indfmt = col_skip(), popsrc = col_skip()))
#drop all the nas
HW3_Data=na.omit(HW3_Data)
#compute Return and MoMentum Factor
Prices=dcast(data = HW3_Data,formula = fyear~tic,value.var = 'prcc_f',fun.aggregate = sum)[,-1]
#Compute Price_inv
Prices_inv=1/Prices
Prices_inv=Prices_inv[-1:-3,]
Prices_inv=t(Prices_inv)
```

```r
#compute MktCap
MktCap=dcast(data = HW3_Data,formula = fyear~tic,value.var = 'mkvalt',fun.aggregate = sum)[,-1]
MktCap=MktCap[-1:-3,]
MktCap=t(MktCap)
#Compute Returns
Returns=(Prices[-1,]-Prices[-nrow(Prices),])/Prices[-nrow(Prices),]
Returns=Returns[-1:-2,]
Returns=t(Returns)
#Compute
MoM=(Prices[-1:-3,]-Prices[-(nrow(Prices)-2):-nrow(Prices),])/Prices[-(nrow(Prices)-2):-nrow(Prices),]
MoM=t(MoM)
#perform 1st step of Fama-MacBeth Reg
#initialize the beta matrix
betas=data.frame(matrix(0,nrow = ncol(Returns),ncol = 4))
colnames(betas)=c('b0','b_cap','b_Mom','b_Pinv')


for (t in 1:ncol(Returns)){
  #clean abnormal data for each regression
  Reg_data_t=cbind(Returns[,t],MktCap[,t],MoM[,t],Prices_inv[,t])
  colnames(Reg_data_t)=c('Return','Cap','Mom','Pinv')
  #get rid of nas
  Reg_data_t=na.omit(Reg_data_t)
  Reg_data_t=data.frame(Reg_data_t)
  #get rid of 0,-1,inf
  Reg_data_t=filter(Reg_data_t,Return!=0,Cap!=0,Mom!=0,Pinv!=0,Return!=-1,Cap!=-1,Mom!=-1,Pinv!=-1,Retu
#Run ts Reg
  {if(nrow(Reg_data_t)==0){betas[t,]=0}
  else{
  Model=lm(formula = Return~Cap+Mom+Pinv,data = Reg_data_t)
  betas[t,]=Model$coefficients}
  }
}
#gather the coeff and compute sderr
 betas_output=list()
 betas_output$coeff=colMeans(betas)
 betas_output$sderr=apply(betas,MARGIN = 2,FUN = sd)/((nrow(betas))^0.5)
 betas_output$t_stats=betas_output$coeff/betas_output$sderr
#show final output
 names(betas_output)='Cross-sectional Regression output'
 betas_output
```

```
## $`Cross-sectional Regression output`
##            b0         b_cap         b_Mom        b_Pinv
##  1.632451e-01 -1.014071e-06  2.685801e-01 -2.958201e-02
##
## $<NA>
##           b0        b_cap        b_Mom       b_Pinv
## 6.499491e-02 4.414509e-07 7.070982e-02 1.642558e-02
##
## $<NA>
##        b0     b_cap     b_Mom    b_Pinv
##  2.511660 -2.297133  3.798342 -1.800972
```

```r
#melten and combine all the data in a pool
melten_Returns=melt(Returns,id.vars = colnames(Returns))
melten_Pinv=melt(Prices_inv,id.vars = colnames(Prices_inv))
melten_Mom=melt(MoM,id.vars = colnames(MoM))
melten_Cap=melt(MktCap,id.vars = colnames(MktCap))
melten_Reg=cbind(melten_Returns,melten_Pinv,melten_Mom,melten_Cap)
melten_Reg=melten_Reg[,c(1,2,3,6,9,12)]
colnames(melten_Reg)=c('tic','year','Return','Pinv','Mom','Cap')
#get rid of abnormal data
melten_Reg=na.omit(melten_Reg)
#get rid of 0,-1,inf
melten_Reg=filter(melten_Reg,Return!=0,Cap!=0,Mom!=0,Pinv!=0,Return!=-1,Cap!=-1,Mom!=-1,Pinv!=-1,Return
Model_pooled=lm(formula =Return~Pinv+Mom+Cap,data = melten_Reg)
#output of the pooled reg using NW sd
pool_reg_ouput=coeftest(vcov. = NeweyWest(x = Model_pooled,lag = 3),x = Model_pooled,adjust=T)
print('Pooled Regression output with Newy West Robust standard errors')
```

```
## [1] "Pooled Regression output with Newy West Robust standard errors"
```

```r
pool_reg_ouput
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)  2.2577e-01  1.1010e-02 20.5055 < 2.2e-16 ***
## Pinv        -3.1651e-02  7.5792e-03 -4.1760 2.973e-05 ***
## Mom          1.1725e-02  1.1252e-02  1.0421    0.2974
## Cap         -7.7611e-07  1.2008e-07 -6.4632 1.036e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```