Financial Econometrics

Assumption and Biases

Darya Yuferova

In this exercise, you investigate the Carhart four-factor model. The model suggest that the exposures of a stock/portfolio to the market, high-minus-low (HML), small-minus-big (SMB), and momentum (WML) factors explain the return of a stock.

Load *Data_for_Carhart_test.Rdata* which includes average industry returns and factor loadings (b = market beta, s = exposure to SMB, h = exposure to HML, w = exposure to momentum) from 2010 to 2016.

Please conduct following analysis:

1. Load and inspect the data
2. Winsorize the data at the 95%. Please create new variables.
3. Make a scatter plot with market beta on the x-axis and return on the y-axis (using winsorized variables). General asset pricing intuition suggests the higher the beta, the higher the return. Does the data support this intuition?
4. Run three regression models explaining return with:
   - The market beta using winsorized variables
   - The Carhart model using the raw variables
   - The Carhart model using the winsorized variables
5. Compute heteroskedasticity robust standard errors for all models. Use *coeftest(..., vcov = hccm)* in which ... indicates the output from *lm(...)*
6. Investigate multicollinearity in model 2 using the Variance Inflation Factor
7. Make a nice stargazer output that uses the heteroskedasticity robust standard errors
8. [Optional] Make a scatter plot with the winsorized returns on the x-axis and the predicted returns from model 3 on the y-axis. Include a 45 degree line. Does the model explain the data well?

The main take away: being able to implement heteroskedasticity robust standard errors, assessing the importance of outliers, and investigating multicollinearity.

```
######################
#      Solution      #
######################

# Packages
require(DescTools)
require(car)
require(lmtest)
require(stargazer)

# set working directory
setwd("C:/Users/s13163/Dropbox/FIE401/data/data_labs/")

# load the data
load("Data_for_Cahart_test.Rdata")

# rename
cahart <- Data_for_Cahart_test
rm(Data_for_Cahart_test)

# investigate data
str(cahart)
```

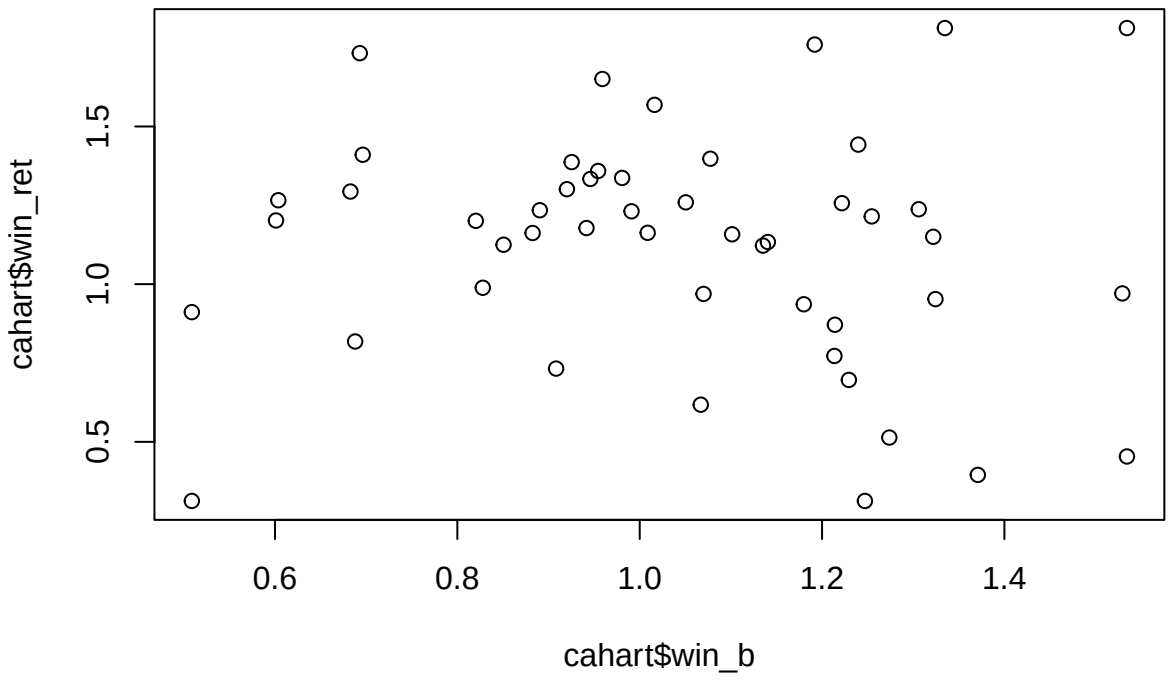
```
## 'data.frame':    49 obs. of  6 variables:
##  $ Industry: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ret     : num  0.696 1.266 1.294 1.202 1.651 ...
##  $ b       : num  1.229 0.604 0.683 0.601 0.959 ...
##  $ s       : num  -0.228 -0.342 -0.607 -0.45 -1.107 ...
##  $ h       : num  -0.203 -0.1067 -0.1976 -0.1609 -0.0733 ...
##  $ w       : num  -0.247 0.134 0.301 0.234 0.317 ...
```

```
summary(cahart)#Create summary statistics
```

```
##     Industry       ret                b                   s
##  Min.   : 1   Min.   :-1.4669   Min.   :-0.0000293   Min.   :-1.1069
##  1st Qu.:13   1st Qu.: 0.9362   1st Qu.: 0.8904692   1st Qu.:-0.1067
##  Median :25   Median : 1.1779   Median : 1.0505369   Median : 0.1530
##  Mean   :25   Mean   : 1.0969   Mean   : 1.0319837   Mean   : 0.1935
##  3rd Qu.:37   3rd Qu.: 1.3336   3rd Qu.: 1.2293805   3rd Qu.: 0.4997
##  Max.   :49   Max.   : 2.2329   Max.   : 1.6630170   Max.   : 1.5680
##        h                   w
##  Min.   :-0.660120   Min.   :-1.313328
##  1st Qu.:-0.160865   1st Qu.:-0.168166
##  Median :-0.020733   Median :-0.002366
##  Mean   : 0.003257   Mean   :-0.064437
##  3rd Qu.: 0.194137   3rd Qu.: 0.130594
##  Max.   : 0.693469   Max.   : 0.426008
```

```
# Winsorize variables
cahart$win_b <- Winsorize(cahart$b,val=quantile(cahart$b,probs=c(0.025, 0.975)))
cahart$win_s <- Winsorize(cahart$s, val=quantile(cahart$s,probs=c(0.025, 0.975)))
cahart$win_h <- Winsorize(cahart$h, val=quantile(cahart$h,probs=c(0.025, 0.975)))
cahart$win_w <- Winsorize(cahart$w, val=quantile(cahart$w,probs=c(0.025, 0.975)))
cahart$win_ret <- Winsorize(cahart$ret, val=quantile(cahart$ret,probs=c(0.025, 0.975)))
```

```r
# Plot return versus market beta
plot(cahart$win_b, cahart$win_ret )
```




```r
# Regression models

fit1 <- lm(win_ret ~ win_b, data = cahart )
fit2 <- lm(ret ~ b + s + h + w, data = cahart)
fit3 <- lm(win_ret ~ win_b + win_s + win_h + win_w, data = cahart)

# Compute heteroskedasticity robust SEs

coeftest(fit1) #not robust to heteroskedasticity
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  1.24744    0.22191  5.6213 1.003e-06 ***
## win_b       -0.11789    0.20711 -0.5692    0.5719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coeftest(fit1, vcov = hccm) #robust to heteroskedasticity
```

```
##
## t test of coefficients:
##
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  1.24744    0.29098  4.2871 8.918e-05 ***
## win_b       -0.11789    0.28505 -0.4136    0.6811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# store heteroskedasticity robust SEs
se1 <- coeftest(fit1, vcov = hccm)[,2]
se2 <- coeftest(fit2, vcov = hccm)[,2]
se3 <- coeftest(fit3, vcov = hccm)[,2]

# Check for imperfect multicollinearity for model 2
vif(fit2)
```

```
##        b        s        h        w
## 1.292079 1.540492 1.219207 1.364531
```

```
# correlation
round(cor(cahart[,7:10]), 2)
```

```
##       win_b win_s win_h win_w
## win_b  1.00  0.42  0.34 -0.20
## win_s  0.42  1.00  0.25 -0.48
## win_h  0.34  0.25  1.00 -0.01
## win_w -0.20 -0.48 -0.01  1.00
```

```
# regression output
stargazer(list(fit1, fit2, fit3),
          se = list(se1, se2, se3),
          type="text", keep.stat=c("n","rsq","adj.rsq"),
          report=('vc*t'))
```

```
##
## ================================================
##                         Dependent variable:
##                     -----------------------------
##                      win_ret      ret      win_ret
##                        (1)        (2)        (3)
## ------------------------------------------------
## win_b                 -0.118                0.116
##                     t = -0.414            t = 0.584
##
## b                                0.130
##                                t = 0.381
##
## s                                0.078
##                                t = 0.737
##
## h                                -0.355
##                                t = -1.530
##
## w                                1.276***
##                                t = 2.997
##
## win_s                                       0.137
##                                           t = 1.294
##
```

```
## win_h                                -0.245
##                                    t = -1.419
##
## win_w                                0.945***
##                                    t = 6.879
##
## Constant       1.247***   1.031***   1.035***
##               t = 4.287  t = 2.887  t = 5.246
##
## ------------------------------------------------
## Observations      49         49         49
## R2              0.007      0.654      0.565
## Adjusted R2     -0.014     0.623      0.525
## ================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```r
# plot actual return versus fitted return
cahart$ret_hat <- fitted(fit3) # extract fitted values
plot(cahart$ret_hat, cahart$win_ret) #Create scatter plot for fitted values and average returns
abline(coef=c(0,1)) #Add 45-degree line
```