Financial Econometrics Rauli Susmel
FINA 7354/4397-A

Financial Econometrics: First Midterm - SOLUTIONS

You want to study the effect of changes in Consumer Sentiment and changes in gold prices on the stock returns of UNP (UNP). You have data on CAT stock prices, Euro exchange rates against the USD (**USD_EUR**), WTI crude oil prices (**Crude_WTI_Oil**), gold prices (**Gold**), the Michigan University Consumer Sentiment Index (**Cons_sent**) and the Fama-French 5 factors: **Mkt_RF** (Market excess returns), **SMB** (size), **HML** (book-to-market), **CMA** (style), and **RMW** (profitability). The data covers the period 1973:January – 2023: July, for a total of 606 observations (***T*** **= 606**). (Recall that when you compute returns or log changes, you lose one observation.)

First, I read the data and define variables:

```
SFX_da <-
read.csv("http://www.bauer.uh.edu/rsusmel/4397/Stocks_FX_1973.csv",head=TRUE,sep=",")
names(SFX_da)
x_dat <- SFX_da$Date
x_cat <- SFX_da$CAT
x_S <- SFX_da$USD_EUR
x_wti <- SFX_da$Crude_WTI
x_gold <- SFX_da$Gold
x_cs <- SFX_da$Cons_sent
x_ip <- SFX_da$IP
x_Mkt_RF<- SFX_da$Mkt_RF
x_SMB <- SFX_da$SMB
x_HML <- SFX_da$HML
x_CMA <- SFX_da$CMA
x_RMW <- SFX_da$RMW
x_RF <- SFX_da$RF

T <- length(x_slb)
e <- log(x_S[-1]/x_S[-T])
lr_wti <- log(x_wti[-1]/x_wti[-T])
lr_cs <- log(x_cs[-1]/x_cs[-T])
lr_slb <- log(x_slb[-1]/x_slb[-T])
lr_unp <- log(x_unp[-1]/x_unp[-T])
lr_gold <- log(x_gold[-1]/x_gold[-T])
lr_oil <- log(x_wti[-1]/x_wti[-T])
lr_ip <- log(x_ip[-1]/x_ip[-T])
x0 <- matrix(1,T-1,1)
Mkt_RF <- x_Mkt_RF[-1]/100
SMB <- x_SMB[-1]/100
HML <- x_HML[-1]/100
CMA <- x_CMA[-1]/100
```

RMW <- x_RMW[-1]/100
RF <- x_RF[-1]/100

unp_x <- lr_unp - RF   # UNP excess returns

1. (20 points) To answer this question, define in R **log changes** of oil prices, *oil,* and of gold prices*, gold.*

a. Suppose the sample mean and sample standard deviation of *oil* are equal to 0.0049 and 0.094, respectively. (You can compute both statistics using R.) Test if the mean of *gold* is equal to zero.

```
> m1 <- 0.0049                  ## Mean
> sd <- 0.094                   ## SD
T <- 611
> se_m1 <- sd/sqrt(T)
> t <- (m1-0)/se_m1
> t
[1] 1.293431
```

$\Rightarrow$ |t-test| > **1.96** $\Rightarrow$ Cannot reject H_0: Mean(*oil*) = 0..

b. Suppose the sample skewness and sample kurtosis of *oil* are equal to 0.61 and 19.84, respectively. (You can compute both statistics using R.) Test if it follows a Normal distribution.

```
> b1 <- 0.61
> b2 <-  19.84
>
> JB <- (b1^2+(b2-3)^2/4)*T/6
> JB
[1] 7255.64
```

$\Rightarrow$ JB test > **5.99** $\Rightarrow$ Reject H_0. The distribubion is not Normal .

c. Using a bootstrap with B=1,000, calculate a 95% C.I. for the correlation between *oil* and the Fama-French market factor, Mkt_RF. Is the correlation equal to zero?

```
> dat_c <- data.frame(lr_oil, Mkt_RF)
> library(boot)
> # function to obtain cor from the data
> cor_xy <- function(data, i) {
+   d <-data[i,]
+   return(cor(d$lr_oil,d$Mkt_RF))
+ }
> boot.samps <- boot(data=dat_c, statistic=cor_xy, R=1000)
> boot.ci(boot.samps, type="perc")

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.samps, type = "perc")

Intervals :
Level     Percentile
95%   (-0.0566,  0.1792 )
Calculations and Intervals on Original Scale
```

<u>Conclusion</u>: Since zero is in the 95% C.I., we cannot reject H_0: corr(*oil*, Mkt_RF) = 0.

2. (35 points) You model CAT excess returns (log CAT returns minus risk-free rate), r_i, as a function of *oil* (log changes of oil prices, as defined in Question 1)**, log changes** of Cons_sent, *CS,* and the first 3 Fama-French factors: Mkt_RF, SMB, and HML. Then, your model becomes a 5-factor model:

$$r_i = \beta_0 + \beta_1 \text{ Mkt\_RF}_i + \beta_2 \text{ SMB}_i + \beta_3 \text{ HML}_i + \beta_4 \, CS_i + \beta_5 \, oil_i + \varepsilon_i \qquad (*)$$

a. Report the regression.

```
> fit_ff5 <- lm(y ~ Mkt_RF + SMB + HML + lr_cs + lr_gold)     # Model
> summary(fit_ff5)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001360   0.002445  -0.556   0.5783
Mkt_RF       1.016252   0.055618  18.272   <2e-16 ***
SMB         -0.276321   0.083566  -3.307   0.0010 **
HML          0.199323   0.079780   2.498   0.0127 *
lr_cs       -0.005196   0.048638  -0.107   0.9150
lr_oil       0.023785   0.025861   0.920   0.3581
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05939 on 605 degrees of freedom
Multiple R-squared:  0.3656,   Adjusted R-squared:  0.3604
F-statistic: 69.75 on 5 and 605 DF,  p-value: < 2.2e-16
```

b. What are the drivers of CAT excess returns in the regression?
Market excess returns (Mkt_RF), SMB, and HML.

c. Report and interpret the R^2 and the F-goodness of fit test.
R^2: **0.3656**. Interpretation: 36.56% of the variation of UNP excess returns is explained by the variation of the five explanatory variables (factors).
F-statistic: **69.75**. Given the p-value, (**2.2e-16**), the five explanatory variables are jointly significant

d. Interpret coefficient β_1.
b_1 = 1.0163. Interpretation: If market excess returns increases by 1%, CAT excess returns should increase by 1.0163%.

e. Does the regression suffer from outliers? Report the proportion of leverage observations and standardized residuals that are significant according to the "Rules of Thumb" presented in class. Interpret your results. (Hint: for leverage, use olsrr package.)

```
> x_resid <- residuals(fit_ff5)
> ## 2.e Check for outliers (using olsrr package)
> x_resid <- residuals(fit_ff5)
> x_stand_resid <- x_resid/sd(x_resid) # standardized residuals
> sum(x_stand_resid > 2)/T            # Rule of thumg count (5% s OK)
```

```
[1] 0.0212766
> library(olsrr)                          # need to install package olsrr
> x_lev <- ols_leverage(fit_ibm_ff3)      # leverage residuals
> sum(x_lev > (2*k+2)/T)/T               # Rule of thumb count (5% is OK)
[1] 0.05728314
```

Conclusion: Low proportion of outliers (less than 5%), as measured by standardized residuals. High leverage observations are a bit higher, 5.73%, than the usual 5% threshold. According to rules of thumb, at the 5% level, there is evidence of outliers, as measured by high leverage observations.

f. Does the regression suffer from multicollinearity problems? Report VIF and Condition Index number. Use the "rules of thumb" presented in class to interpret your results. (Hint: You can use olsrr package.)

```
> ols_vif_tol(fit_ibm_ff3)
  Variables Tolerance      VIF
1    Mkt_RF 0.8888476 1.125052
2       SMB 0.9322602 1.072662
3       HML 0.9493857 1.053313

> ols_eigen_cindex(fit_ibm_ff3)
  Eigenvalue Condition Index  intercept      Mkt_RF        SMB      HML
1  1.3457300        1.000000 0.05495615 0.326658156 0.22100420 0.070303
2  1.1103423        1.100907 0.35353536 0.006678000 0.03900969 0.457141
3  0.9069222        1.218131 0.46242518 0.009218542 0.43253698 0.139981
4  0.6370055        1.453474 0.12908331 0.657445302 0.30744914 0.332573
```

Conclusion: According to Rules of Thumb, all regressors have a VIF<5, which indicates no multicollinearity. Same result for the Rules of Thumb for the Condition Index, where all the variables have a Condition Index lower than 10. No evidence of multicollinearity in the model.

g. According to the estimated 5-factor model, did UNP over-perform or under-perform? Estimate the over/under performance.

```
> b_ff5 <- fit_ff5$coefficients
> mean_x <- c(mean(Mkt_RF), mean(SMB), mean(HML), mean(lr_cs), mean(lr_
gold))
> exp_ret <- t(b_ff5[2:6])%*% mean_x
>
> # over/underperfromance of CAT
> mean(y)  - exp_ret
             [,1]
[1,] -0.001375493
```

Alternative answer: The constant (alpha) is negative & significant $\Rightarrow$ UNP underperformed by alpha = -**0.136%.**

3. (30 points) Continuation.

a. Test if H_0: $\beta_1 = 1$ vs H_1: $\beta_1 \neq 1$. Is CAT as risky as the market? Explain.

```
> t_beta_1 <- (summary(fit_ff5)$coefficients[2,1] - 1)/summary(fit_ff5)
$coefficients[2,2]
> t_beta_1
[1] 0.2922
```

|t-test | < **1.96** $\Rightarrow$ Cannot reject H_0: $\beta_1 = 1$.

b. Test the CAPM –i.e., H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$– against your 5-factor model. Is the CAPM a good model for DIS? Interpret test result.

```
> library(car)
> linearHypothesis(fit_ff5, c("SMB = 0","HML = 0", "lr_cs = 0", "lr_oil
= 0"), test="F")
Linear hypothesis test

Hypothesis:
SMB = 0
HML = 0
lr_cs = 0
lr_oil = 0

Model 1: restricted model
Model 2: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    609 2.1963
2    605 2.1336  4  0.062674 4.4429 0.001515 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: At the 5% level, we strongly reject H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

c. Test H_0: $\beta_4 = 0.5$ and $\beta_5 = 0.2$ vs H_1: $\beta_5 \neq 0.5$ and/or $\beta_5 \neq 0.2$.

```
> linearHypothesis(fit_ff5, c("lr_cs = 0.5","lr_oil = 0.2"), test="F")
Linear hypothesis test

Hypothesis:
lr_cs = 0.5
lr_oil = 0.2

Model 1: restricted model
Model 2: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    607 2.6596
2    605 2.1336  2   0.52596 74.569 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
---
```

Conclusion: At the 5% level, we strongly reject H_0: $\beta_4 = 0.3$ and $\beta_5 = 0.2$.

d. Using a Wald test, if Profitability factor, *RMW,* is missing from your regression. What are the implications of your test result?

```
library(lmtest)
fit_ff5_RMW <- lm(y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil + RMW)     #No
w, U Model
> waldtest(fit_ff5_RMW, fit_ff5)
Wald test

Model 1: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil + RMW
Model 2: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil
  Res.Df Df      F   Pr(>F)
1    604
2    605 -1 10.645 0.001166 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: At the 5% level, the Wald test strongly rejects the restricted model without RMW. That is, RMW is missing from model.

e. An observer says that given that UNP is a big transportation company, Industrial Production is missing from the model. Using a Wald test, check if log changes in IP are missing from your regression. Is the observer correct?

```
> fit_ff5_IP <- lm(y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil + lr_ip)  #No
w, U Model
> waldtest(fit_ff5_IP, fit_ff5)
Wald test

Model 1: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil + lr_ip
Model 2: y ~ Mkt_RF + SMB + HML + lr_cs + lr_oil
  Res.Df Df     F Pr(>F)
1    604
2    605 -1 0.419 0.5177
```

Conclusion: At the 5% level, we cannot reject the H_0 that changes in IP are not missing from the model. (We cannot reject the restricted model, without lr_ip.)

e. Date the beginning of the 2008 Financial Crisis in September 2008 (observation **429**). Test with a Chow test if the 2008 Financial Crisis caused a structural break in your model.

```
> library(strucchange)
> sctest(unp_x ~ Mkt_RF + SMB + HML + lr_cs + lr_oil, type = "Chow", po
int = x_break)

        Chow test

data:  unp_x ~ Mkt_RF + SMB + HML + lr_cs + lr_oil
F = 1.5206, p-value = 0.1688
```

Conclusion: At the 5% level, we cannot reject H_0: *no structural break.*

4. (15 points) True of False (Provide a very brief statement justifying your answer. No justification, no points.)

a. If in the CLM, the errors, ε, are normally distributed, then **b**, the OLS estimator of $\boldsymbol{\beta}$**.** is also normally distributed. ε. **True**. Conditioning on **X**, **b** inherits the distribution of the errors.

b. High leverage observations ("outliers far from the mean") can have no effect on a regression. **True**. It is possible, as long as it aligns on the line of the other points.

c. Imposing a false restriction causes inefficiency in the estimation. **False**. Any restriction increases the precision of estimators..

d. R^2 can never decrease when we add variables in a regression. **True**. RSS can never increase with higher *k*.

e. It is impossible to build a confidence interval if the distribution of the data is not normal.
False. We can use asymptotic theory or a bootstrap.