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 Caterpillar (**CAT**). 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.)

1. (20 points) To answer this question, you need to define in R **log changes** of gold prices, *gold.*
a. Suppose the sample mean and sample standard deviation of *gold* are equal to 0.0055 and 0.057, respectively. (You can compute both statistics using R.) Test if the mean of *gold* is equal to zero.

```
> x_dat <- SFX_da$Date
> x_cat <- SFX_da$CAT
> x_xom <- SFX_da$XOM
> x_wti <- SFX_da$Crude_WTI_Oil
> x_cs <- SFX_da$Cons_sent
> x_gold <- SFX_da$Gold
> 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_SMB)
> 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_cat <- log(x_cat[-1]/x_cat[-T])
> lr_xom <- log(x_xom[-1]/x_xom[-T])
> lr_gold <- log(x_gold[-1]/x_gold[-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
> cat_x <- lr_cat - RF                  # GE excess returns
> xom_x <- lr_xom - RF                  # SLB excess returns
> x <- lr_gold
> T <- length(x)
> m1 <- sum(x)/T                        ## Mean
> m1
[1] 0.005536038
> m2 <- sum((x-m1)^2)/T                 ## Var
> sd <- sqrt(m2)                        ## SD
> sd
[1] 0.05706022

> se_m1 <- sd/sqrt(T)
```

```
> t <- m1/se_m1
> t
[1] 2.388372
```
⇒ |t-test| > **1.96** ⇒ Reject H_0: Mean(*gold*) = 0..

b. Suppose the sample skewness and sample kurtosis of *gold* are equal to 0.461 and 7.447, respectively. (You can compute both statistics using R.) Test if *gold* follows a Normal distribution.

```
> m3 <- sum((x-m1)^3)/T                ## For numerator of S
> m4 <- sum((x-m1)^4)/T                ## For numerator of K
> b1 <- m3/m2^(3/2)          ## Sample Skewness
> b1
[1] 0.4605973
> b2 <- (m4/m2^2)                      ## Sample Kurtosis
> b2
[1] 7.447329

> JB <- (b1^2+(b2-3)^2/4)*T/6
> JB
[1] 520.8402
```
⇒ JB test > **5.99** ⇒ Reject H_0: *gold* follows a Normal .

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

```
> dat_c <- data.frame(lr_gold, Mkt_RF)
> library(boot)
> # function to obtain cor from the data
> cor_xy <- function(data, i) {
+   d <-data[i,]
+   return(cor(d$lr_gold,d$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%   (-0.0729,  0.1151 )
Calculations and Intervals on Original Scale
```

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

2. (25 points) You model CAT excess returns (log CAT returns minus risk-free rate), r_i, as a function of *gold* (log changes of gold 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:

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

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|)
               Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006721   0.002738  -2.455   0.01437 *
Mkt_RF       1.226003   0.061909  19.803   < 2e-16 ***
SMB         -0.018486   0.093053  -0.199   0.84259
HML          0.536923   0.088355   6.077 2.18e-09 ***
lr_cs        0.020000   0.054270   0.369   0.71261
lr_gold      0.129857   0.047084   2.758   0.00599 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06595 on 600 degrees of freedom
Multiple R-squared:  0.4219,   Adjusted R-squared:  0.4171
F-statistic: 87.59 on 5 and 600 DF,  p-value: < 2.2e-16
```

b. What are the drivers of CAT excess returns in the regression?

Market excess returns (Mkt_RF), HML and *gold*.

c. Report and interpret the R^2 and the F-goodness of fit test.

R^2: **0.4219**. Interpretation: 42% of the variation of the caterpillar excess returns is explained by the variation of the five explanatory variables (factors).

F-statistic: **87.59**. Given the p-value, (**2.2e-16**), the five explanatory variables are jointly significant

d. Interpret coefficient β_1.

b_1 = 1.226003. Interpretation: If market excess returns increases by 1%, CAT excess returns should increase by 1.226%.

e. According to the estimated 5-factor model above, did CAT 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
> exp_ret
            [,1]
[1,] 0.009582753
> exp_ret1 <- sum(b_ff5[2:6]*mean_x)
> # over/underperfromance b_ff4[1]
> mean(y)  - exp_ret
             [,1]
[1,] -0.006828306
```

Alternative answer: The constant (alpha) is negative & significant $\Rightarrow$ CAT underperformed by alpha = **-0.68%.**

3. (25 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_ff4)$coefficients[2,1] - 1)/summary(fit_ff4)
$coefficients[2,2]
```

```
> t_beta_1
[1] -3.603066
```

|t-test |> **1.96** $\Rightarrow$ 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.

```
> library(car)
> linearHypothesis(fit_ff4, c("SMB = 0","HML = 0", "lr_cs = 0", "lr_gol
d = 0"), test="F")
Linear hypothesis test

Hypothesis:
SMB = 0
HML = 0
lr_cs = 0
lr_gold = 0

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

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    604 2.8037
2    600 2.6164  4   0.18732 10.739 2.061e-08 ***
---
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: $\beta_3 = 0.3$ and $\beta_5 = 0.2$ vs H_1: $\beta_3 \neq 0.3$ and/or $\beta_5 \neq 0.2$.

```
> linearHypothesis(fit_ff5, c("HML = 0.3","lr_gold = 0.2"), test="F")
Linear hypothesis test

Hypothesis:
HML = 0.3
lr_gold = 0.2

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

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    602 2.6598
2    600 2.6164  2  0.043418 4.9785 0.007172 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
---
```

Conclusion: At the 5% level, we strongly reject H_0: $\beta_3 = 0.3$ and $\beta_5 = 0.2$.

d. Using a Wald test, test if log changes in WTI crude oil prices, *p,* are missing from your regression. What are the implications of your test result?

```
> library(lmtest)
> fit_ff5_WTI <- lm(y ~ Mkt_RF + SMB + HML + lr_cs + lr_wti + lr_gold)
        #Now, U Model
> waldtest(fit_ff5_WTI, fit_ff5)
Wald test

Model 1: y ~ Mkt_RF + SMB + HML + lr_cs + lr_wti + lr_gold
Model 2: y ~ Mkt_RF + SMB + HML + lr_cs + lr_gold
  Res.Df Df      F  Pr(>F)
1    599
2    600 -1 2.9867 0.08447 .
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: At the 5% level, log changes in WTI do not seem to be missing from model.

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.

```
> F_test <- ((RSS_ct - RSS_ct_tot)/J)/(RSS_ct_tot/(T_ct - k1 - k2))  # F
test
> F_test
[1] 3.597402
> qf(.95, df1=J, df2=(T - k1 - k2))          # F_value from Table at (1
-alpha)=0.95
[1] 2.229142
> p_val <- 1 - pf(F_test, df1=J, df2=(T_ct - k1 - k2))
> p_val
[1] 0.003255293
```

Conclusion: At the 5% level, we reject H_0: *no structural break.*

4. (15 points) You have two competing 4-factor models, Model 1 & Model 2.
a. Fit both models for CAT excess returns, report which variables are driving CAT excess returns in each model:

$$r_i = \beta_0 + \beta_1 \text{ Mkt\_RF}_i + \beta_2 \text{ SMB}_i + \beta_3 \text{ HML}_i + \beta_4\, CS_i + \varepsilon_i \qquad (*) \qquad \text{Model 1}$$

$$r_i = \beta_0 + \beta_1 \text{ Mkt\_RF}_i + \beta_2 \text{ CMA}_i + \beta_3 \text{ RMW}_i + \beta_4\, gold_i + \varepsilon_i \quad (*) \qquad \text{Model 2}$$

```
> fit1 <- lm(y ~ Mkt_RF + SMB + HML + lr_cs + lr_wti)        # Model 1
> summary(fit1)

Call:
lm(formula = y ~ Mkt_RF + SMB + HML + lr_cs + lr_wti)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.225252 -0.040091 -0.000285  0.038413  0.280195 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.005988   0.002742  -2.184   0.0293 *  
Mkt_RF       1.232355   0.062468  19.728  < 2e-16 ***
SMB         -0.011521   0.093804  -0.123   0.9023    
HML          0.537294   0.089458   6.006 3.29e-09 ***
lr_cs        0.008616   0.054612   0.158   0.8747    
lr_wti      -0.038283   0.029011  -1.320   0.1875    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06635 on 600 degrees of freedom
Multiple R-squared:  0.4144,  Adjusted R-squared:  0.4095 
F-statistic: 84.91 on 5 and 600 DF,  p-value: < 2.2e-16

> fit2 <- lm(y ~ Mkt_RF + CMA + RMW + lr_gold)       # Model 2
> summary(fit2)
```

```
Call:
lm(formula = y ~ Mkt_RF + CMA + RMW + lr_gold)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.25544 -0.03843 -0.00029  0.04058  0.28676 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.009081   0.002809  -3.233  0.00129 ** 
Mkt_RF       1.298463   0.063382  20.486  < 2e-16 ***
CMA          0.727008   0.139784   5.201 2.72e-07 ***
RMW          0.370033   0.117780   3.142  0.00176 ** 
lr_gold      0.112635   0.047125   2.390  0.01715 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06601 on 601 degrees of freedom
Multiple R-squared:  0.4194,	Adjusted R-squared:  0.4155 
F-statistic: 108.5 on 4 and 601 DF,  p-value: < 2.2e-16
```

b. Use a J-test to select a model.

```
> jtest(fit1,fit2)
J test

Model 1: y ~ Mkt_RF + SMB + HML + lr_cs + lr_wti
Model 2: y ~ Mkt_RF + CMA + RMW + lr_gold
                Estimate Std. Error t value  Pr(>|t|)    
M1 + fitted(M2)  0.73697    0.18858  3.9079 0.0001037 ***
M2 + fitted(M1)  0.74582    0.21896  3.4062 0.0007028 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: Based on the J-test we select Model 2, the one with the strongest rejection.

5. (15 points) True of False (Provide a very brief statement justifying your answer. No justification, no points.)

a. The variance of **b**, the OLS estimator, decreases with the variance of the errors, ε.

True – check the formula for Var[**b**].

b. We use a bootstrap only when we do not believe the data is normally distributed.

False – Not only when the data is not normal, there are other reasons, for example, when we do not believe on the asymptotic distribution

c. Imposing a false restriction causes inefficiency, but no biases in OLS estimates.

False. - Any restriction reduces variance –i.e., increase "efficiency." A false restriction can cause bias, for example, with omitted variables.

d. RSS always decreases when we add irrelevant variables in a regression.

True – RSS always increase when we add more regressors.

e. A Wald-test can be used to test for missing regressors in a regression.

True – We have used a Wald test it in question 3 to test for missing *p* in our model.