Financial Econometrics — Rauli Susmel
FINA 4397/7354

Homework 2 - Solutions

1.1 (**Tests of Hypothesis**). Download the Shiller dataset (Shiller_2020data.csv) from my homepage. You have stock prices (P), dividends (D), earning (E), consumer prices (CPI) and long interest rates (Long_i). Regress log stock returns, r_i, against log earning changes, $earn_i$, inflation rate (in log changes), Inf_i, and interest rates, int_i (need to subtract one observation):

$$r_i = \beta_0 + \beta_1\, earn_i + \beta_2\, Inf_i + \beta_3\, int_i + \varepsilon_i$$

a. Report the regression

```
SH_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/Shiller_data.csv", head=TRUE, sep=",")
x_date <- SH_da$Date
x_P <- SH_da$P
x_D <- SH_da$D
x_E <- SH_da$E
T <- length(x_S)
cpi_us <- SH_da$CPI
i_us <- SH_da$Long_i

T <- length(x_P)
lr_p <- log(x_P[-1]/x_P[-T])
lr_d <- log(x_D[-1]/x_D[-T])
lr_e <- log(x_E[-1]/x_E[-T])
inf <- log(cpi_us[-1]/cpi_us[-T])
int <- (i_us[-1])/100

T <- length(lr_p)

### Report Regression
fit_mod <- lm (lr_p ~ lr_e + inf + int)

> summary(fit_mod)
Call:
lm(formula = lr_p ~ lr_e + inf + int)

Residuals:
     Min       1Q   Median       3Q      Max
-0.31212 -0.01876  0.00355  0.02338  0.40981

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.003925   0.002062   1.904   0.0571 .
lr_e         0.129385   0.025484   5.077 4.22e-07 ***
inf          0.362472   0.092137   3.934 8.67e-05 ***
int         -0.028214   0.041046  -0.687   0.4919
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04015 on 1822 degrees of freedom
Multiple R-squared:  0.02444,  Adjusted R-squared:  0.02283
```

```
F-statistic: 15.21 on 3 and 1822 DF,  p-value: 8.833e-10
```

b. The model explains **2.44%** of the variability of stock returns. More precisely, the variation of the independent variables explain 2.44% of the variation of stock returns.

c. If log changes in earnings increase by 1%, stock returns increase by **.12941%.**

d. Test with a goodness of fit test H_0: $\beta_1 = \beta_2 = \beta_3 = 0$.
F-statistic: 15.21 with p-value: **8.833e-10**

<u>Conclusion</u>: Reject H_0 at 5% level.

f. Test with an F-test H_0: $\beta_1 = \beta_3 = 0$.

```
e_u <- fit_mod$residuals
RSS_u <- sum(e_u^2)                          # Unrestricted RSS

fit_r  <- lm (lr_p ~ inf )                   # Restricted model
e_r <- fit_r$residuals
RSS_r <- sum(e_r^2)                          # Restricted RSS

F_test_1 <- ((RSS_r - RSS_u)/2)/(RSS_u/(T-4)) # F-test
>F_test_1
[1] 13.16431

p_val <- 1 - pf(F_test_1, df1 = 2, df2= T-4)   # p-value of F-test
>p_val
 [1] 2.107306e-06
```

$\Rightarrow$ reject H_0 at 5% level.

<u>Conclusion</u>: We strongly reject H_0 at 5% level.

<u>Note</u>: You could have tested . H_0: $\beta_1 = \beta_3 = 0$ using the R package car

```
library(car)
linearHypothesis(fit_mod, c("lr_e = 0","int = 0"), test="F")        # F: exact test
```

g. Test with a Wald test H_0: $\beta_2 = 0.5$ and $\beta_3 = -0.1$

```
library(car)
>linearHypothesis(fit_mod, c("inf = 0.5","int = -0.1"), test="F")  # F: exact test

Linear hypothesis test

Hypothesis:
inf = 0.5
int = - 0.1

Model 1: restricted model
Model 2: lr_p ~ lr_e + inf + int

  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1   1824 2.9441
2   1822 2.9364  2 0.0077386 2.4009 0.09093 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Given that stock returns are not normal distributed, I also use the asymptotic version of the F-test. In car, I set test='Chisq"

```
> linearHypothesis(fit_mod, c("inf = 0.5","int = -0.1"), test="Chisq")

Hypothesis:
inf = 0.5
int = - 0.1

Model 1: restricted model
Model 2: lr_p ~ lr_e + inf + int

  Res.Df    RSS Df Sum of Sq  Chisq Pr(>Chisq)
1   1824 2.9441
2   1822 2.9364  2 0.0077386 4.8017    0.09064 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: Both versions of the test (exact F & asymptotic W) cannot reject H_0 at 5% level.

f. Check if the model shows structural change. Perform a Chow test with T_{SB}= October 1973

```
> x_date[1234]                          # Check Tequila effect date
[1] 1973.1
x_break <- 1233                         # We lost one observation when taking log changes
y <- lr_p
 x <- cbind(lr_e, inf, int)
T <- nrow(x)
k <- ncol(x)
T0 <- 1                                 # Initial observation of Regime 1
T1 <- x_break                           # Structural break time (end of Regime 1)
T2 <-  T1 + 1                           # Initial observatin of Regime 2

# Restricted RSS (From Whole sample)
e_ct <- fit_mod$residuals
RSS_ct <- sum(e_ct^2)                   # RSS_R

# Unrestricted Estimation (Two periods)
x_ct_1 <- x[T0:T1,]
y_ct_1 <- y[T0:T1]
x_ct_2 <- x[T2:T,]
y_ct_2 <- y[T2:T]
k1 <- ncol(x_ct_1)
k2 <- ncol(x_ct_2)

fit_mod_1 <- lm(y_ct_1 ~ x_ct_1)        #OLS regression from 1 to T1 (Regime 1)
e_ct_1 <- fit_mod_1$residuals
RSS_ct_1 <-  sum(e_ct_1^2)

fit_mod_2 <- lm(y_ct_2 ~ x_ct_2)        #OLS regression from T1+1 to T (Regime 2)
```

```
e_ct_2 <- fit_mod_2$residuals
RSS_ct_2 <-  sum(e_ct_2^2)

RSS_ct_tot <- RSS_ct_1 + RSS_ct_2                # RSS_U

T_ct <- length(y_ct)
J <- k2 + k1 - k
F_test <- ((RSS_ct - RSS_ct_tot)/J)/(RSS_ct_tot/(T_ct - k1 - k2))# F test
> F_test
[1] 3.055063
p_val <- 1 - pf(F_test, df1=J, df2=(T_ct - k1 - k2))
> p_val
[1] 0.02742764
```

⇒ reject H_0 at 5% level

<u>Conclusion</u>: Evidence of Structural break at 5% level.

1.2 (**Bootstrapping**). Bootstrap the t-statistics in the above regression, with B=1,000.

```
fit_mod <- lm (lr_p ~ lr_e + inf + int)
sh_ols <- summary(fit_mod)$coef[,1]              #extracting OLS estimated parameter from lm (1st row in table)

dat_Sh <- data.frame(lr_p, lr_e, inf, int)       # create R dataframe to use in lmboot
sim_size <- 1000
```

R Note: R will mention that it's recommended to habe B=sim_size > T. You can set B = 2000

```
library(lmboot)
sh_b <- paired.boot(lr_p ~ lr_e + inf + int, data=dat_Sh, B=sim_size)      #paired bootstrap in lm
```

a. Report the mean and the bias in your estimation for each parameter.

```
## a-1. Mean values for b (bootstrap estimates of b
> mean(sh_b$bootEstParam[,1])          # print mean of bootstrap samples for constant
[1] 0.003973828
> mean(sh_b$bootEstParam[,2])  # print mean of bootstrap samples for lr_e
[1] 0.1274647
> mean(sh_b$bootEstParam[,3])  # print mean of bootstrap samples for inf
[1] 0.3816334
> mean(sh_b$bootEstParam[,4])  # print mean of bootstrap samples for int
[1] -0.02762141

# You can get from lmboot the OLS estimated parameters
sh_b$origEstParam                      # OLS ("Original") estimated parameters

## a-2. bootstrap bias (OLS estimate - Bootstrap estimated)
> sh_b$origEstParam[1] - mean(sh_b$bootEstParam[,1])
[1] -1.628643e-06
```

```
> sh_b$origEstParam[2] - mean(sh_b$bootEstParam[,2])
[1] -0.003842865
> sh_b$origEstParam[3] - mean(sh_b$bootEstParam[,3])
[1] -0.003120396
> sh_b$origEstParam[4] - mean(sh_b$bootEstParam[,4])
[1] -0.0007803871
```

Conclusion: The bias is small –i.e., the mean of the parameters estimated using the bootstrap is similar to the OLS. We feel good about the OLS estimation.

b. Build a 95% C.I. for β_2.

```
> ## b. bootstrap 95% CI for beta_2 (inf) parameter (percentile method)
> quantile(sh_b$bootEstParam[,3], probs=c(.025, .975))
     2.5%     97.5%
0.2002979 0.5622499
```

Conclusion: β_2 is positive (C.I. does not include 0).

1. 3 (**Non-nested Tests**)

a. Estimate two Fama-French 3-factor model for GE returns: One with Mkt_RF, SMB and HML (Model 1) and the other with Mkt_RF, CMA and RMW.

```
SFX_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/Stocks_FX_1973.csv",head=TRUE,sep=",")
x_ge <- SFX_da$GE
x_xom <- SFX_da$XOM
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_ibm)
lr_ge <- log(x_ge[-1]/x_ge[-T])
lr_xom <- log(x_xom[-1]/x_xom[-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
ge_x <- lr_ge - RF                                   # GE excess returns

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

Call:
```

lm(formula = y ~ Mkt_RF + SMB + HML)

Residuals:

```
     Min        1Q    Median       3Q      Max
-0.314703 -0.030023 -0.001069  0.031406  0.211469
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.007995   0.002240  -3.569 0.000387 ***
Mkt_RF       1.246405   0.050496  24.683  < 2e-16 ***
SMB         -0.253154   0.075703  -3.344 0.000877 ***
HML          0.419352   0.072351   5.796  1.1e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05427 on 602 degrees of freedom
Multiple R-squared:  0.5068,	Adjusted R-squared:  0.5043 
F-statistic: 206.2 on 3 and 602 DF,  p-value: < 2.2e-16
```

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

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

Residuals:

```
     Min        1Q    Median       3Q      Max
-0.311685 -0.029477 -0.001219  0.031069  0.211848
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.008604   0.002357  -3.651 0.000284 ***
Mkt_RF       1.215267   0.053414  22.752  < 2e-16 ***
CMA          0.373678   0.117773   3.173 0.001586 ** 
RMW          0.115554   0.099077   1.166 0.243956    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05564 on 602 degrees of freedom
Multiple R-squared:  0.4816,	Adjusted R-squared:  0.479 
F-statistic: 186.4 on 3 and 602 DF,  p-value: < 2.2e-16
```

b. Use a J-test to select a model

J-test using the jtest in the lmtest package:

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

Model 1: y ~ Mkt_RF + SMB + HML
Model 2: y ~ Mkt_RF + CMA + RMW
                Estimate Std. Error t value  Pr(>|t|)    
M1 + fitted(M2) -0.32956    0.39050  -0.844     0.399    
M2 + fitted(M1)  1.10864    0.19698   5.628 2.798e-08 ***
```

⇒ Model 1 explains Model 2.

Conclusion: J-test favors Model 1.

c. Perform an encompassing test to select or favor a model
 fit_enc <- lm(y ~ Mkt_RF + SMB + HML + CMA + RMW)
> summary(fit_enc)

Call:
lm(formula = y ~ Mkt_RF + SMB + HML + CMA + RMW)

Residuals:
```
      Min        1Q    Median        3Q       Max
-0.315495 -0.029857 -0.001139  0.031638  0.212462
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.007454   0.002311  -3.225 0.001330 **
Mkt_RF       1.233135   0.053307  23.133  < 2e-16 ***
SMB         -0.275463   0.081172  -3.394 0.000735 ***
HML          0.469216   0.098254   4.776 2.26e-06 ***
CMA         -0.098644   0.154500  -0.638 0.523409
RMW         -0.084095   0.105013  -0.801 0.423565
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05432 on 600 degrees of freedom
Multiple R-squared:  0.5075,	Adjusted R-squared:  0.5034
F-statistic: 123.7 on 5 and 600 DF,  p-value: < 2.2e-16
```

Encompassing test using the encomptest in the lmtest package:

library(lmtest)
> encomptest(fit1,fit2)
Encompassing test

```
Model 1: y ~ Mkt_RF + SMB + HML
Model 2: y ~ Mkt_RF + CMA + RMW
Model E: y ~ Mkt_RF + SMB + HML + CMA + RMW
   Res.    Df  Df       F    Pr(>F)
M1 vs. ME  600 -2  0.4734  0.6231     ⇒ F-test cannot reject restrictions implied by M1.
M2 vs. ME  600 -2 15.8145 2.027e-07  ⇒ F-test rejects restrictions implied by M2.
```

Conclusion: In the encompassing model, CMA and RMW are not significant. Individual tests favors Model 1.

Overall Conclusion: Model 1 is the one we would use to draw inferences and do forecasting.

1.4 (Structural Change)
a. Explain the term 'parameter structural stability'?
Parameters are constant over the whole sample.

b. The question refers to the existence of two regimes.

$r_t = \alpha_1 + \beta_1 r_{mt} + \varepsilon t$ Regime 1 (pre-Oct 1987)

$r_t = \alpha_2 + \beta_2 r_{mt} + \varepsilon t$ Regime 2 (post-Oct 1987)

Then, Unrestricted model: Two regimes (4 parameters: α_1, α_2, β_1, & β_2)
Restricted model: One regime (2 parameters: α, & β)

c. What are the null and alternative hypotheses that are being tested here, in terms of α and β?
H_0 (no structural break): $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$.
H_0 (no structural break): $\alpha_1 \neq \alpha_2$ and/or $\beta_1 \neq \beta_2$

d. Perform the test. What is your conclusion?

```
RSS_R = 0.189
RSS1 = 0.079
RSS2 = 0.082
RSS_U <- RSS1 + RSS2

T = 180
k_U = 4
k_R = 2
J <- k_U - k_R
F <- ((RSS_R - RSS_U)/J)/(RSS_U/(T-k_U))
> F
[1] 15.30435

> qf(.95, df1=J, df2=(T - k_U))
[1] 3.047307
p_val <- 1 - pf(F_test, df1=J, df2=(T - k_U))
>  p_val
[1] 7.447977e-07
```

⇒ p-value < .05 reject H_0 (no structural break) at 5%.

Conclusion: Strong evidence in favor of structural change.

1.5 (Theory Review)
Check Notes.