Financial Econometrics Rauli Susmel
FINA 7354/4397

First Midterm - SOLUTIONS

• 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$Date
x_slb <- SFX_da$SLB
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_HS <- SFX_da$House_starts
x_Com <- SFX_da$Commodities
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_cat <- log(x_cat[-1]/x_cat[-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])
lr_hs <- log(x_HS[-1]/x_HS[-T])
lr_com <- log(x_Com[-1]/x_Com[-T])
x0 <- matrix(1,T-1,1)
Mkt_RF <- x_Mkt_RF[-1]/100
SMB <- x_SMB[-1]/100
```

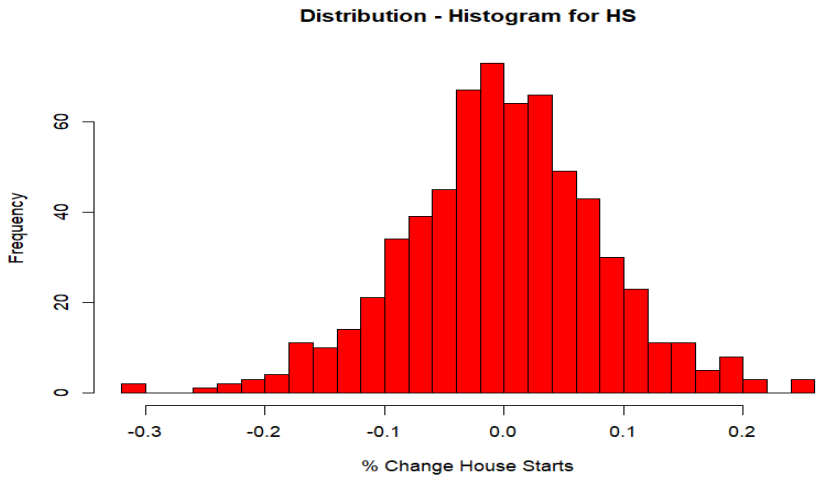
```
HML <- x_HML[-1]/100
CMA <- x_CMA[-1]/100
RMW <- x_RMW[-1]/100
RF <- x_RF[-1]/100

cat_x <- lr_cat - RF                    # CAT excess returns
```

1. (**25 points**) To answer this question you need to define in R **log changes** for House starts, ***HS*** *and* CAT (Caterpillar) excess returns (log CAT returns minus risk-free rate), r_i.

a. Do a histogram for ***HS***.

```
x <- lr_hs
n_breaks <- 40
h <- hist(x, breaks=n_breaks, col="red", xlab="% Change House Starts", main = "Distribution - Histogram for HS")
```



b. Report the mean, standard deviation, skewness (γ_1) and kurtosis (γ_2).

```
> T <- length(x)
> m1 <- sum(x)/T                    ## Mean
> m1
[1] -0.001081553
> m2 <- sum((x-m1)^2)/T             ## Var
> sd <- sqrt(m2)                    ## SD
> sd
[1] 0.08290469
> 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] -0.08193708
```

```
> b2 <- (m4/m2^2)                    ## Sample Kurtosis
> b2
[1] 3.652233
```

c. Using an appropriate test, test if ***HS*** is normally distributed (write down H_0 and H_1, compute test and your decision).
H_0(normal): $\gamma_1 = 0$ and $\gamma_2 = 3$,
H_1 (not normal): $\gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$.

```
> JB <- (b1^2+(b2-3)^2/4)*T/6
> JB
[1] 12.09804
> qchisq(.95, df = 2)          # chi-squared (df=2) value at 5% level
[1] 5.991465
> p_val <- 1 - pchisq(JB, df = 2)       # p-value of LM_test
> p_val
[1] 0.002360173
```

Conclusion: Since JB > **5.99** (= $\chi^2_{2,.05}$), we reject H_0. That is, HS are not normally distributed.

d. Using an appropriate test, test if the mean of ***HS*** is 1% (write down H_0 and H_1, calculate test and your decision).

```
> se_m1 <- sd/sqrt(T)
> t <- (m1-0.01)/se_m1
> t
[1] -3.386796
```

Conclusion: Since |t| > **1.96**, we reject H_0. That is, HS mean is not 1% (actually, less than 1%).

e. Test if the annual rate of growth of the ***HS*** is -1%. (For the t-test, we need to annualized the mean and SD)

```
> se_a1 <- sd * sqrt(12)/sqrt(T/12)
> ta <- (m1*12 + 0.01)/se_a1
> ta
[1] -0.07586209
```

Conclusion: Since |t| < **1.96**, we cannot reject H_0.

f. Using an empirical bootstrap, estimate a 95% C.I for the correlation between ***HS*** and r_i. Based on your C.I., is the correlation 0?

```
> library(boot)
> 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="basic")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

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

Intervals :
Level      Basic
95%   (-0.0591,  0.1628 )
Calculations and Intervals on Original Scale
```

Conclusion: The 95% CI goes from -0.0591 to 0.1628, which includes 0. Thus, we cannot reject that the correlation is 0.

2. (**25 points**) You model CAT (Caterpillar) excess returns (log CAT returns minus risk-free rate), r_i, as a function of *HS* and the first 3 Fama-French factors: Mkt_RF $(r_{m,t} - r_f)$, SMB_t, and HML_t. Then, your model becomes a 4-factor model:

$$r_i = \beta_0 + \beta_1 (r_{m,t} - r_f) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 HS_t + \varepsilon_t \qquad (*)$$

a. Report the regression.

```
> y <- cat_x
> fit_ff4 <- lm(y ~ Mkt_RF + SMB + HML + lr_hs)                # Model
> ## 2.a Report Regression
> summary(fit_ff4)

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

Residuals:
     Min       1Q   Median       3Q      Max
-0.272754 -0.039856 -0.001259  0.039313  0.268203

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.005316   0.002725  -1.951   0.0515 .
Mkt_RF       1.246681   0.061976  20.116  < 2e-16 ***
SMB          0.023332   0.092335   0.253   0.8006
HML          0.518486   0.089213   5.812 9.77e-09 ***
lr_hs       -0.033776   0.032418  -1.042   0.2979
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06781 on 637 degrees of freedom
Multiple R-squared:  0.4077,        Adjusted R-squared:  0.404
F-statistic: 109.6 on 4 and 637 DF,  p-value: < 2.2e-16
```

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

c. Report and interpret the R^2 and the F-goodness of fit test.
$R^2 = 0.4077$
Interpretation: The variability of the explanatory variables explains 40.77% of the variability of CAT excess returns.
F-stats = **109.6**, with p-value: < 2.2e-16
Interpretation: The four explanatory variables are jointly significant to explains CAT excess returns.

d. Interpret coefficient β_1.
A 1% increase in Market excess returns, increases CAT excess returns **by 1.25%**. The size of the coefficients points towards a riskier stock than the market.

e. Given the estimated coefficient for HML_t, is CAT a value or a growth company?
Given that the sign of $\beta_{3=HML1}$ is positive, CAT is a high book-to-market stock -i.e., a value stock.

f. Does the regression suffer from multicollinearity? Compute VIF and Condition Index.

```
> ols_vif_tol(fit_ibm_ff3)
  Variables Tolerance      VIF
1    Mkt_RF 0.8881575 1.125926
2       SMB 0.9335445 1.071186
3       HML 0.9423446 1.061183
>
> ols_eigen_cindex(fit_ibm_ff3)
  Eigenvalue Condition Index intercept  Mkt_RF   SMB       HML
1  1.3292285       1.000000 0.06241848 0.341676673 0.21131302 0.06401859
2  1.1199683       1.089424 0.30256016 0.009977154 0.05147953 0.47823822
```

3 0.9289208 **1.196218** 0.48105022 0.009402296 0.43732132 0.09084582
4 0.6218824 **1.461994** 0.15397113 0.638943877 0.29988612 0.36689738

Conclusion: For all regressors, we have VIF < 5 & Condition Index < 10 => No evidence of multicollinearity.

g. Does the regression suffer from outliers? Check if the number of standardized residuals and Cooks D statistic violate the Rule of Thumb presented in class.

```
> x_resid <- residuals(fit_ff4)
> x_stand_resid <- x_resid/sd(x_resid)          # standardized residuals
> sum(abs(x_stand_resid )> 2)/T                 # Rule of thumg count (5% count is OK)-
proportion of "outliers"
[1] 0.0560747
> library(olsrr)                                # need to install package olsrr
> cooksd <- cooks.distance(fit_mod)
> sum(cooksd > 4/T)/T                           # Rule of thumb count (5% count is OK)
[1] 0.0482866
```

Conclusion: Low proportion of outliers (around 5%), as measured by standardized residuals and Cook's D. According to rules of thumb, at the 5% level, there is evidence of outliers, as measured by high leverage observations. Overall, not very strong evidence.

h. According to the Model (*), did CAT outperform or underperform?

```
> b_ff4 <- fit_ff4$coefficients
> mean_x <- c(mean(Mkt_RF), mean(SMB), mean(HML), mean(lr_cs))
> exp_ret <- t(b_ff4[2:5])%*% mean_x
> # over/underperfromance of CAT
> mean(y)  - exp_ret
            [,1]
[1,] -0.005302243
```

Conclusion: Underperformance, close to **0.53%**. Is it significant? We look at the significance of the constant in the regression. The p-value is very, very close to **1.96**. Thus, I would consider it significant.

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_ff4)$coefficients[2,1] - 1)/summary(fit_ff4)$coefficients[2,2]
> t_beta_1
[1] 3.980278
```

Conclusion: Since |**t-test = 3.98 | > 1.96** $\Rightarrow$ Reject : H_0: $\beta_1 = 1$. Since the coefficient is gr eater than 1, CAT is riskier that the market.t is, H_0: $\beta_1 = 1$.

b. Test if SMB_t and HML_t are needed to explain CAT excess returns (write down H_0: and H_1).

```
> library(car)
> linearHypothesis(fit_ff4, c("SMB = 0","HML = 0"), test="F")

Linear hypothesis test:
SMB = 0
HML = 0

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

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    639 3.0871
2    637 2.9290  2   0.15817 17.199 5.311e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: Since p-value is smaller than 5%, we reject H_0: $\beta_3 = \beta_4 = 0$.

c. Test H_0: $\beta_3 = \beta_4 = 0.1$ vs H_1: $\beta_3 \neq 0.1$ and/or $\beta_4 \neq 0.1$.

```
> linearHypothesis(fit_ff4, c("HML = 0.1","lr_hs = 0.1"), test="F")

Linear hypothesis test:
HML = 0.1
lr_hs = 0.1

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

  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    639 3.109
2    637 2.929  2   0.18002 19.575 5.619e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: Since the p-value is smaller than 5%, at the 5% level, we reject H_0: $\beta_3 = \beta_4 = 0.1$.

d. Using an LM test, test if **log changes in Gold prices,** *gold,* are missing from your model. What are the implications of your test result?

```
> library(lmtest)
> fit_ff4_RMW <- lm(y ~ Mkt_RF + SMB + HML + lr_hs + lr_gold)          #Now, U Mo
del
> waldtest(fit_ff4_gold, fit_ff4)
Wald test

Model 1: y ~ Mkt_RF + SMB + HML + lr_hs + lr_gold
Model 2: y ~ Mkt_RF + SMB + HML + lr_hs
  Res.Df Df      F   Pr(>F)
1    636
2    637 -1 10.796 0.001073 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: Since the p-value is smaller than 5%, at the 5% level, the Wald test rejects the restricted model -i.e., with no *gold*. That is, we have evidence that changes in gold prices are a driver of CAT excess returns.

e. Given that CAT sells machines that are used for agriculture, a researcher speculates that **log changes in Commodity prices** affect CAT excess returns. Test if the researcher's speculation is right.

```
> fit_ff4_IP <- lm(y ~ Mkt_RF + SMB + HML + lr_hs + lr_com)  #Now, U Model
> waldtest(fit_ff4_IP, fit_ff4)
Wald test

Model 1: y ~ Mkt_RF + SMB + HML + lr_hs + lr_com
Model 2: y ~ Mkt_RF + SMB + HML + lr_hs
  Res.Df Df      F Pr(>F)
1    636
2    637 -1 0.1727 0.6779
```

Conclusion: Since the p-value is bigger than 5%, we cannot reject the restricted model -i.e., we have no evidence that changes in commodity prices are a driver of CAT excess returns.

f. 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 (*).

```
> x_dat[429]
[1] "9/30/2008"
> x_break <- 428                                   # We lost one observation when
taking log changes
> library(strucchange)
```

```
> sctest(cat_x ~ Mkt_RF + SMB + HML + lr_hs, type = "Chow", point = x_break)

        Chow test

data:  cat_x ~ Mkt_RF + SMB + HML + lr_hs
F = 2.1971, p-value = 0.05306
```

Conclusion: At the 5% level, the Chow test cannot rejects the null hypothesis of No structural break. However, the p-value is very close to 5%. A "marginal" decision here. r

g. Overall, is the 4-factor model (*) a good model to explain CAT excess returns? Briefly explain.
Since we have evidence that *gold* is missing, the model is mis-specified. Morever, as shown in the previous question (2.h), the negative constant, which points out to underperformance, is very, very close to be significant. Overall, the model is not a good model.

4. (15 points) True of False (Provide a very brief statement justifying your answer. No justification, no points.)
a. False. The inverse of $(\boldsymbol{X}'\boldsymbol{X})$ becomes imprecise with multicollinearity.
b. False. Strictly speaking this is not true. However, it is recommended that T is not "too small."
c. False. It is possible to preferred a biased estimator if the bias is not very big and it has a very small variance.
d. False. As shown in class, even the most irrelevant regressors can increase R^2.
e. True. When the errors are not normal -i.e., (**A5**) does not apply-, the formulation of the F-test changes to $J * F$, which asymptotically follows a χ^2_J.