ECON90033 – QUANTITATIVE ANALYSIS OF FINANCE I

TUTORIAL 3

Download the t3e1 and t3e2 Excel data file from the subject website and save it to your computer or USB flash drive. Read this handout and complete the tutorial exercises before your tutorial class so that you can ask for help during the tutorial if necessary.

The Capital Asset Pricing Model

The basic empirical capital asset pricing model (CAPM) is a simple linear regression model of the relationship between the excess return on asset *i* and the excess return on the market,

$$R_{it} - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \varepsilon_t$$

where R_{it} is the one-period simple return on asset *i*, R_{mt} is the one-period simple return on the market portfolio, R_{ft} is the one-period simple return on a risk-free asset, $R_{it} - R_{ft}$ is the excess return on asset *i*, $R_{mt} - R_{ft}$ is the excess return on the market, and ε_t is a white noise stochastic error term[1], all in in time *t.* As regards the parameters, the β slope parameter is the β- risk of asset *i*, a measure of exposure of the returns on asset *i* to movements in the market, relative to a risk-free asset, and the α parameter is the α- risk of asset *i*, the abnormal return to asset *i* in addition to the asset's exposure to the excess return on the market.

Based on the β- risk, an individual stock or a portfolio can be classified as

Aggressive:	$\beta > 1$	(e.g., technology stocks)
Benchmark:	$\beta = 1$	(e.g., S&P 500 in the US)
Conservative:	$0 < \beta < 1$	(e.g., blue chip stocks)
Uncorrelated:	$\beta = 0$	(risk-free stocks like treasury bonds)
Imperfect Hedge:	$-1 < \beta < 0$	(e.g., gold, cash)
Perfect Hedge:	$\beta = -1$	(an ideal but 'non-existing hedge)

Exercise 1 (HMPY, p. 74, Ex 1)

The *t3e1.xlsx Excel* file contains monthly observations for the period April 1990 to July 2004 on the equity prices of Exxon (*Exxon*), General Electric (*GE*), IBM (*IBM*), Microsoft (*Msoft*), and Walmart (*Wmart*), together with the price of Gold (*Gold*), the S&P 500 index (*SP500*), and a short term interest rate measured by the annual rate of the US 3-month Treasury Bill (*Tbill*).[2]

---

[1] $\{\varepsilon_t\}$ is a white noise if each element has an identical, independent, and mean-zero distribution.
[2] This data set is *t2e1.xlsx* with two more variables, *SP500* and *Tbill*.

a) Compute the monthly simple excess returns on Exxon. Use the S&P 500 as a proxy for the market and the US 3-month Treasury Bill as a proxy for the simple return of the risk-free asset.

Launch *RStudio*, create a new project and script, and name them *t3e1*. Import the data from the *t3e1* Excel file, attach it to your project, and save it as *t3e1.RData*.

attach(t3e1)

In Exercise 1 of Tutorial 2 we already calculated the simple returns for Exxon. We can follow the same steps to obtain the simple returns for S&P 500 as well.

Exxon = ts(Exxon, start = c(1990, 4), end = c(2004, 7), frequency = 12)
R.Exxon = Exxon / lag(Exxon,-1) - 1

SP500 = ts(SP500, start = c(1990, 4), end = c(2004, 7), frequency = 12)
R.SP500 = SP500 / lag(SP500,-1) – 1

As regards the simple returns on the risk-free asset, it is important to recall that *Tbill* is an annual rate, so it hast to be divided by 12 so as to get the monthly risk-free rate.

Tbill = ts(Tbill, start = c(1990, 4), end = c(2004, 7), frequency = 12)
R.Tbill = Tbill / 12

From the simple returns we can obtain the excess returns:

ER.Exxon = R.Exxon – R.Tbill
ER.SP500 = R.SP500 – R.Tbill

b) Estimate the CAPM simple linear regression for Exxon and briefly comment on the results.

A linear regression can be set up and estimated in R with the lm() function, which has the following general format:[3]

lm(formula = y ~ x1 + x2 + …)

where y is the dependent variable and x1, x2 etc. denote the independent variables.[4] *By default, the model has an implied intercept term, and the "formula =" part of the specification can be omitted for the sake of brevity.*

[3] This specification of the *lm()* function assumes that regression is going to be performed on the active data set. Otherwise, *lm()* has to be augmented with the *data = [data source]* argument.

[4] This is the most frequently used specification of the formula. The general syntax is *formula = y ~ terms*, where *terms* is the set of independent variables. It can be given in the form (i) *first+second*, meaning that all the terms in *first* together with all the terms in *second* with duplicates removed; (ii) *first:second*, meaning the set of terms obtained by taking the interactions of all terms in *first* with all terms in *second*; or (iii) *first*second*, meaning the

To obtain the sample regression model for Exxon, execute the following commands:

CAPM.Exxon = lm(ER.Exxon ~ ER.SP500)
summary(CAPM.Exxon)

You should get the following printout:

```
Call:
lm(formula = ER.Exxon ~ ER.SP500)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.098339 -0.024114 -0.001793  0.020650  0.143391 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.01265    0.00302   4.188 4.53e-05 ***
ER.SP500     0.50900    0.07075   7.195 1.95e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03927 on 169 degrees of freedom
Multiple R-squared:  0.2345,	Adjusted R-squared:  0.2299 
F-statistic: 51.76 on 1 and 169 DF,  p-value: 1.946e-11
```

It shows the executed *lm()* command and some location statistics of the residuals, the regression coefficients with the corresponding standard errors, *t*-ratios and *p*-values, and some useful summary measures of the goodness of fit, like the standard error of regression, the unadjusted and adjusted R^2 statistics, and the details of the *F*-test of overall significance.[5]

Starting with the sample coefficient of determination, R^2 is about 0.23, suggesting that in the sample at hand this simple linear regression accounts for only about 23% of the variations of the excess returns on the Exxon stocks.

The *p*-value of the *F*-test of the overall significance of this simple linear regression model is practically zero, so we can safely reject the null hypothesis and conclude at any reasonable significance level that (i) this regression model is significant, (ii) the slope estimate is significant(ly different from zero)[6], and (iii) R^2 is significantly positive.

Turning to the estimates of the coefficients, the sample regression equation is

union of *first+second* and *first:second*.

[5] If any of these terms sound unfamiliar to you, look it up in an introductory business statistics book or in section 3.4 of the prescribed textbook.

[6] For simple linear regression models the *F*-test of overall significance is equivalent to a two-tail *t*-test for the slope with zero hypothetical parameter value.

$$\widehat{R_{it} - R_{ft}} = \hat{\alpha} + \hat{\beta}(R_{mt} - R_{ft}) = 0.01265 + 0.50900(R_{mt} - R_{ft})$$

The point estimate of the β- risk of Exxon is 0.50900, meaning that Exxon follows the market to some extent, but its movements on average are only about half (51%) of the movements in the whole market. The estimate of β- risk is between 0 and 1, so Exxon is a conservative stock.

The point estimate of the α- risk of Exxon is 0.01265. It means that the excess return on Exxon stocks exceeds the excess return on the market portfolio by 0.01265.

From the signs of the point estimates, which are the same as those of the *t*-ratios, and the one-tail *p*-values of the *t*-tests, which are halves of the reported two-tail *p*-values (*Pr(>|t|)* on the printout), we can conclude at any reasonable significance level that both point estimates are not only significantly different from zero, but they are significantly positive. Hence, Exxon has a significant β- risk.

An investor might also wish to know whether the β- risk of Exxon is significantly different from 1 and thus it does not track the market one-to-one. In this case the hypotheses are

$$H_0 : \beta = 1 \quad , \quad H_A : \beta \neq 1$$

and the appropriate test is again a two-tail *t*-test, but this time the hypothesised parameter value is 1. The observed test statistic from the point estimate and the standard error is

$$t_{obs} = \frac{\hat{\beta} - \beta_{H_0}}{s_{\hat{\beta}}} = \frac{0.50900 - 1}{0.07075} = -6.9399$$

It could be compared to the *t*-critical value for *df* = 169 and the preferred significance level, but it is more convenient to perform a general *F*-test with *R*.

The general *F*-test is a generalization of the *F*-test of overall significance, and it can be applied to verify any set of linear equality restrictions on the parameters of a population regression model. In principle, it is a three-step procedure:

i. Estimate the unrestricted model and get the sum of squares due to error, *SSE*.
ii. Impose the restrictions on the unrestricted model, estimate the restricted model and get again the sum of squares due to error, SSE_r.
iii. Perform an *F*-test based on the following test statistic:

$$F = \frac{(SSE_r - SSE)/m}{SSE/(n-k-1)} = \frac{n-k-1}{m}\frac{SSE_r - SSE}{SSE}$$

where *k* is the number of independent variables in the unrestricted model and *m* is the number of linear restrictions imposed on the unrestricted model.

In R, following lm, the general F-test can be performed with the linearHypothesis function of the car package:

linearHypothesis(model, hypothesis.matrix, test = c("F", "Chisq"))

where model is the previously fitted model object, hypothesis.matrix is either the matrix of the linear combinations of the coefficients (by rows) under the null hypothesis or a character vector that specifies the null hypothesis in symbolic form, i.e., by referring to the slope parameters by the names of the corresponding independent variables, and test is a character that specifies whether to compute the finite sample F statistic with approximate F distribution (this is the default option) or the large sample Chi-squared statistic with asymptotic Chi-squared distribution.

In this case, after having installed the *car* package, we need to execute

library(car)
linearHypothesis(model = CAPM.Exxon, c("ER.SP500 = 1"))

to obtain the following printout:

```
Linear hypothesis test

Hypothesis:
ER.SP500 = 1

Model 1: restricted model
Model 2: ER.Exxon ~ ER.SP500

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    170 0.33494
2    169 0.26065  1  0.074288 48.167 8.003e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The observed *F*-test statistic is 48.167, which is the square of the observed *t*-statistic on the previous page. The reported *p*-value is practically zero, so we can reject the null hypothesis and conclude that Exxon stock does not track the whole market one-to-one.

Before moving on to part (c) of this exercise, recall from the tutorial on last week that for small returns

$$r_t = \ln(1 + R_t) \approx R_t$$

i.e., the log return is approximately equal to the simple return. This implies that the

$$R_{it} - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \varepsilon_t$$

CAPM has an alternative version:

$$r_{it} - R_{ft} = \alpha + \beta(r_{mt} - R_{ft}) + \varepsilon_t$$

Estimate this model for Exxon by executing the following commands:

er.Exxon = diff(log(Exxon), 1) - R.Tbill
er.SP500 = diff(log(SP500), 1) - R.Tbill
CAPM.Exxon_v2 = lm(er.Exxon ~ er.SP500)
summary(CAPM.Exxon_v2)

You should get this printout:

```
Call:
lm(formula = er.Exxon ~ er.SP500)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.096897 -0.023558 -0.000765  0.020722  0.130575 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.012018   0.002945    4.08 6.92e-05 ***
er.SP500    0.501768   0.068830    7.29 1.14e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03838 on 169 degrees of freedom
Multiple R-squared:  0.2392,    Adjusted R-squared:  0.2347 
F-statistic: 53.14 on 1 and 169 DF,  p-value: 1.138e-11
```

If you compare this printout to the one on page 3, you can see that the two printouts are not the same, but the differences are qualitatively negligible.

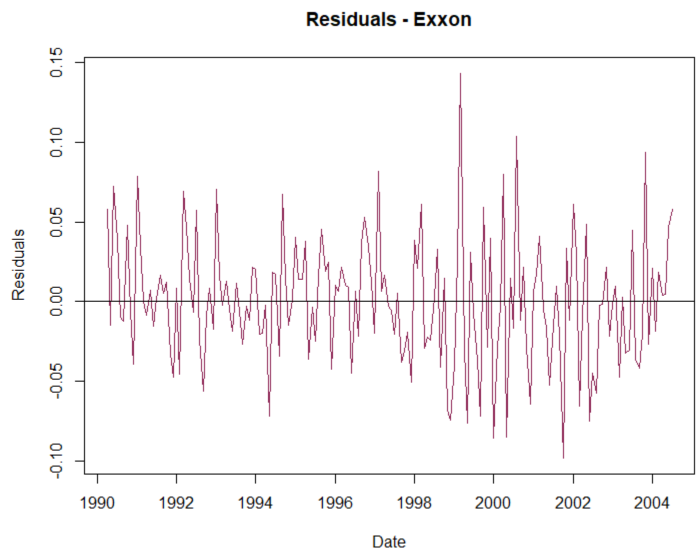
c) Consider again the original CAPM and conduct a residual analysis by (i) plotting the residuals and performing the (ii) White (*W*) test for heteroskedasticity, (iii) Breusch-Godfrey (*BG*) LM test for residual autocorrelation, (iv) Jarque-Bera (*JB*) test for normality, (v) Lagrange Multiplier (*LM*) test for *ARCH* errors, and (vi) Ramsey's Regression Specification Error Test (*RESET*).[7]

We can follow the steps in Ex 1 of the lectures on week 2.

CAPM.Exxon.res = ts(residuals(CAPM.Exxon),
start = c(1990, 4), end = c(2004, 7), frequency = 12)
plot.ts(CAPM.Exxon.res, xlab = "Date", ylab = "Residuals",
main = "Residuals - Exxon", col = "violetred4")
abline(h = 0)

[7] LK: If these tests sound unfamiliar, you can review them in Heij, C. et al. (2004) or in a similar book.

return the following residual plot:



The residuals fluctuate around zero[8]. They do not exhibit any obvious patter, except that their variance appears to me a bit larger in the second half of the sample period.

In the *W* test for heteroskedasticity, H_0: homoskedasticity and H_A: heteroskedasticity. The

library(lmtest)
bptest(CAPM.Exxon, ~ ER.SP500 + I(ER.SP500^2))

commands return,

```
        studentized Breusch-Pagan test

data:  CAPM.Exxon
BP = 1.3857, df = 2, p-value = 0.5001
```

The *p*-value is about 0.5, so H_0 cannot be rejected at the usual significance levels. Hence, the White test does not support my visual impression based on the residual plot.

To perform the *BG* test for H_0: no autocorrelation up to order 12 in the error term against H_A: some 1-12 order autocorrelation, run

bgtest(CAPM.Exxon, order = 12, type = "Chisq")

[8] Of course, this is not surprising, since OLS residuals from regressions with a *y*-intercept term always have zero mean.

It returns,

```
        Breusch-Godfrey test for serial correlation of order up to 12

data:  CAPM.Exxon
LM test = 12.684, df = 12, p-value = 0.3924
```

The p-value is about 0.39, so H_0 cannot be rejected at the usual significance levels, implying that the stochastic error terms, $\{\varepsilon_t\}$, might be serially uncorrelated.

The next test is the *JB* test for normality of the error term. H_0: normal distribution is tested against H_A: non-normal distribution by performing the

library(tseries)
jarque.bera.test(CAPM.Exxon.res)

commands. They return,

```
        Jarque Bera Test

data:  CAPM.Exxon.res
X-squared = 5.2452, df = 2, p-value = 0.07261
```

The p-value is about 0.073, so at the 5% significance level we maintain H_0 and conclude that the stochastic error terms might be normally distributed.

In the simplest version of the *LM* test for *ARCH* errors, H_0: no first-order autocorrelation in the squared stochastic error terms is tested against H_A: first-order autocorrelation in the squared stochastic error terms. The

library(FinTS)
ArchTest(CAPM.Exxon.res, lags = 1)

commands return,

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  CAPM.Exxon.res
Chi-squared = 0.1444, df = 1, p-value = 0.7039
```

The p-value is about 0.70, too large again to reject H_0 at the usual significance levels, so there is no evidence of *ARCH* errors.[9]

[9] If you repeat this test for *ARCH* errors of orders 1-12, i.e., run *ArchTest(CAPM.Exxon.res, lags = 12)*, the conclusion does not change.

Finally, the *RESET* test for H_0: correct functional form against H_A: incorrect functional form, can be performed by executing

resettest(CAPM.Exxon, power = 3, type = "fitted")

It returns the following printout:

```
        RESET test

data:  CAPM.Exxon
RESET = 0.08224, df1 = 1, df2 = 168, p-value = 0.7746
```

The *p*-value is about 0.77, too large again to reject H_0 at the usual significance levels, so the functional form of the model might adequate.

All things considered, this CAPM for Exxon satisfies the assumptions we discussed in the week 2 lectures.

d) Repeat part (b) for IBM this time.

The following commands

IBM = ts(IBM, start = c(1990, 4), end = c(2004, 7), frequency = 12)
R.IBM = IBM / lag(IBM,-1) - 1
ER.IBM = R.IBM - R.Tbill

CAPM.IBM = lm(ER.IBM ~ ER.SP500)
summary(CAPM.IBM)

produce the printout shown on the next page.

The sample coefficient of determination, R^2 is about 0.30, suggesting that in the sample at hand this simple linear regression accounts for about 30% of the variations of the excess returns on the IBM stocks.

The *p*-value of the *F*-test of the overall significance is zero, so this regression model is also significant.

The sample regression equation is

$$\widehat{R_{it} - R_{ft}} = 0.007587 + 1.237888(R_{mt} - R_{ft})$$

The point estimate of the β- risk of IBM is 1.23788. It is positive and since (half of) its reported *p*-value is practically zero, it is significantly positive. In addition, this point estimate is greater than one, so IBM follows the market and on average its movements are 24% larger than the movements of the whole market. Therefore, IBM is an aggressive stock.

The point estimate of the α-risk of IBM is 0.007587 and it could be interpreted in the usual way. Note, however, that its reported *p*-value is 0.219, so this point estimate is insignificant(ly different from zero), even at the 20% significance level. For this reason, it is better to refrain from interpreting it.

```
Call:
lm(formula = ER.IBM ~ ER.SP500)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.281564 -0.046724 -0.001361  0.043256  0.268241 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.007587   0.006146   1.234    0.219    
ER.SP500    1.237888   0.143946   8.600 5.24e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07991 on 169 degrees of freedom
Multiple R-squared:  0.3044,	Adjusted R-squared:  0.3003 
F-statistic: 73.95 on 1 and 169 DF,  p-value: 5.244e-15
```

Next, execute

linearHypothesis(model = CAPM.IBM, c("ER.SP500 = 1"))

to perform an *F*-test for

$H_0 : \beta = 1 \quad , \quad H_A : \beta \neq 1$

The printout is

```
Linear hypothesis test

Hypothesis:
ER.SP500 = 1

Model 1: restricted model
Model 2: ER.IBM ~ ER.SP500

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    170 1.0965                            
2    169 1.0790  1  0.017438 2.7311 0.1003
```

The reported *p*-value is 0.1003, so we cannot reject the null hypothesis, not even at the 10% significance level. This means that the IBM stock might track the whole market one-to-one.[10]

If that is indeed the case, the CAPM reduces to

$$R_{it} - R_{ft} = \alpha + (R_{mt} - R_{ft}) + \varepsilon_t \quad \rightarrow \quad R_{it} = \alpha + R_{mt} + \varepsilon_t$$

which is referred to as the market model.

Finally, we can again estimate the alternative version of CAPM by executing

```
er.IBM = diff(log(IBM), 1) - R.Tbill
CAPM.IBM_v2 = lm(er.IBM ~ er.SP500)
summary(CAPM.IBM_v2)
```

The new printout is below.

```
Call:
lm(formula = er.IBM ~ er.SP500)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.31986 -0.04193  0.00096  0.04605  0.23138 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.004340   0.006045   0.718    0.474    
er.SP500    1.205004   0.141260   8.530 7.97e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07876 on 169 degrees of freedom
Multiple R-squared:  0.301,	Adjusted R-squared:  0.2968 
F-statistic: 72.77 on 1 and 169 DF,  p-value: 7.974e-15
```

It confirms that a CAPM based on log returns is not identical to a CAPM based on simple returns, but it provides a reasonable approximation in most applications as the returns are usually relatively small.

Save your *R* code and quit *RStudio*.

---

[10] Please note the wording of this conclusion, in particular the word *might*. In statistical tests based on the classical approach, a rejected null hypothesis is a far stronger outcome than a maintained null hypothesis. At the conventional significance levels, when the null is rejected, the alternative is likely true, so we are willing to take a small risk of Type I error and accept the alternative. However, when there is not enough evidence against the null and thus it is not rejected, there is still a relatively large chance that it is actually false, so we maintain it without actually accepting it.

Fama-French Three-Factor CAPM

The Fama-French three-factor CAPM is a generalization of the basic CAPM as in addition to the risk premium on the market, it also considers size risk, which represents the return spread between big market capitalization stocks and small market capitalization stocks, and value risk, which is the difference between the returns on value stocks (sold below their actual value) and growth stocks (have above-average revenue and earnings growth potential).

Accordingly, the population regression model takes the form

$$R_{it} - R_{ft} = \alpha + \beta_1 (R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_i$$

where SMB_t is ‘Small Minus Big’, i.e., the average return on small stock portfolios minus the average return on big stock portfolios, and HML_t is ‘High Minus Low’, i.e., the average return on value portfolios minus the average return on growth portfolios.

Exercise 2 (HMPY, p. 75, Ex 2)

The *t3e2.xlsx Excel* file contains monthly Fama-French data[11] for market (*MKT*), risk-free (*RF*), size risk (*SMB*), and value risk (*HML*) for the period January 1927 to December 2013. The return on the market was constructed as the value-weighted return of all firms in the CRSP (Center for Research in Security Prices) data base that are incorporated in the United States and listed on the NYSE, AMEX, or NASDAQ, and the risk-free rate is the 1-month US treasury bill rate expressed as a monthly rate. The file also contains the monthly returns on a US energy portfolio (*ENERGY*).

a) Estimate the Fama-French three factor model for the Energy portfolio and comment on the results.

   Launch *RStudio*, create a new project and script, and name them *t3e2*. Import the data from the *t3e2* Excel file, attach it to your project, and save it as *t3e2.RData*.

   ```
   attach(t3e2)
   ```

   Create time series objects by executing

   ```
   MKT = ts(MKT, start = c(1927, 1), end = c(2013, 12), frequency = 12)
   RF = ts(RF, start = c(1927, 1), end = c(2013, 12), frequency = 12)
   SMB = ts(SMB, start = c(1927, 1), end = c(2013, 12), frequency = 12)
   HML = ts(HML, start = c(1927, 1), end = c(2013, 12), frequency = 12)
   ENERGY = ts(ENERGY, start = c(1927, 1), end = c(2013, 12), frequency = 12)
   ```

[11] Source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

The model is estimated by executing

ER.ENERGY = ENERGY - RF
ER.MKT = MKT - RF
FF3 = lm(ER.ENERGY ~ ER.MKT + SMB + HML)
summary(FF3)

These commands produce the following output:

```
Call:
lm(formula = ER.ENERGY ~ ER.MKT + SMB + HML)

Residuals:
    Min      1Q  Median      3Q     Max 
-17.667  -2.085  -0.086   2.098  16.327 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.19015    0.11620   1.636    0.102    
ER.MKT       0.86911    0.02289  37.969  < 2e-16 ***
SMB         -0.21592    0.03771  -5.726 1.35e-08 ***
HML          0.21320    0.03351   6.363 2.96e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.712 on 1040 degrees of freedom
Multiple R-squared:  0.6208,	Adjusted R-squared:  0.6197 
F-statistic: 567.4 on 3 and 1040 DF,  p-value: < 2.2e-16
```

The adjusted coefficient of determination, Adj-R^2 is about 0.62, suggesting that in the sample at hand this multiple linear regression accounts for about 62% of the variations of the excess returns on the Energy portfolio.

The *p*-value of the *F*-test of the overall significance is zero, so this regression model is significant.

The sample regression equation is

$$\widehat{R_{it} - R_{ft}} = 0.19015 + 0.86911(R_{mt} - R_{ft}) - 0.215921 SMB_t + 0.21320 HML_t$$

The reported *p*-values of the three slope estimates are all practically zero, so the first and the third slope estimates are significantly positive, while the second is significantly negative.

The point estimate of the β- risk is 0.86911. It is positive and smaller than one, so the Energy portfolio is conservative. It follows the market but, its movements on average are about 13% smaller than the movements of the whole market, assuming that SMB_t and HML_t are kept constant.[12]

[12] The meaning of a regression slope parameter or estimate is conditional on all other independent variables in

The slope estimate of the size risk is -0.21592, so when the difference between the average returns on small market capitalization stocks and large market capitalization stocks increases by one, the excess return on the Energy portfolio is expected to decrease by 0.21592, assuming that $R_{mt} - R_{ft}$ and HML_t are kept constant.

The slope estimate of the value risk is 0.21200, so when the difference between the average returns on value stocks and growth stocks increases by one, the excess return on the Energy portfolio is expected to increase by 0.21200, assuming that $R_{mt} - R_{ft}$ and SMB_t are kept constant.

Before moving on to part (b), it is worth to mention that it is possible to replace the four commands we used to estimate the model and to print the results with a single command:

summary(lm(I(ENERGY - RF) ~ I(MKT - RF) + SMB + HML))

In this command the

I() function is used to avoid confusions when a regression formulae involves arithmetic expressions. For example, "x + y" on the right-hand side of a regression formula means that x and y are two separate independent variables, while I(x+y) is a single independent variable, z = x+y.

If you execute this command and compare the new printout to the previous one, you can see that they are indeed identical.

b) If $\beta_2 = \beta_3 = 0$, the three-factor CAPM collapses to the basic, i.e., one-factor, CAPM. Test these restrictions and if they are supported, estimate the basic CAPM and compare the two estimates of the β- risk.

Perform a general F-test for H_0: $\beta_2 = \beta_3 = 0$ against H_A: $\beta_2 \neq 0$ or $\beta_3 \neq 0$ or both by executing

library(car)
linearHypothesis(model = FF3, c("SMB = 0", "HML = 0"))

The printout is on the next page. The p-value is practically zero, so we can safely reject the null hypothesis and conclude that the Fama-French 3-factor model better explains the movements in the excess returns on the Energy portfolio than the basic CAPM.

---

the model, and it shows the average change in the dependent variable in response to a one unit increase in the corresponding independent variable, while the other independent variables in the model are kept constant. This is the so-called ceteris paribus (all other things being equal) condition of regression analysis.

```
Hypothesis:
SMB = 0
HML = 0

Model 1: restricted model
Model 2: ER.ENERGY ~ ER.MKT + SMB + HML

  Res.Df   RSS Df Sum of Sq      F    Pr(>F)    
1   1042 15290                                  
2   1040 14329  2    960.44 34.853 2.234e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

c) In the Fama-French model in part (b) the point estimates of β_2 and β_3 are -0.21592 and 0.21320, respectively, suggesting that the effects of the size risk and the value risk on the excess return of the Energy portfolio have opposite signs but similar absolute values. Test the restriction $\beta_2 + \beta_3 = 0$. Depending upon the outcome of the test re-estimate the Fama-French three factor model.

Perform a general F-test for H_0: $\beta_2 + \beta_3 = 0$ against H_A: $\beta_2 + \beta_3 \neq 0$ by executing

linearHypothesis(model = FF3, c("SMB + HML = 0"))

You should get

```
Linear hypothesis test

Hypothesis:
SMB  + HML = 0

Model 1: restricted model
Model 2: ER.ENERGY ~ ER.MKT + SMB + HML

  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   1041 14329                           
2   1040 14329  1  0.042453 0.0031 0.9557
```

The p-value is 0.9557, far too large to reject the null hypothesis at any reasonable significance level. This means that the Fama-French 3-factor model can be replaced with the following restricted version:

$$R_{it} - R_{ft} = \alpha + \beta_1 (R_{mt} - R_{ft}) + \beta_2 (SMB_t - HML_t) + \varepsilon_i$$

Estimate this new model by executing

FF3_r = lm(ER.ENERGY ~ ER.MKT + I(SMB - HML))
summary(FF3_r)

The new printout is

```
Call:
lm(formula = ER.ENERGY ~ ER.MKT + I(SMB - HML))

Residuals:
     Min       1Q   Median       3Q      Max 
-17.7067  -2.0924  -0.0819   2.1044  16.3263 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.18962    0.11574   1.638    0.102    
ER.MKT        0.86864    0.02123  40.923   <2e-16 ***
I(SMB - HML) -0.21439    0.02567  -8.353   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.71 on 1041 degrees of freedom
Multiple R-squared:  0.6208,	Adjusted R-squared:   0.62 
F-statistic:   852 on 2 and 1041 DF,  p-value: < 2.2e-16
```

Comparing this printout to the one on page 13, you can see that they are practically the sane in terms of goodness of fit and the slope estimate of the excess return on the market portfolio. This means that the $\beta_2 + \beta_3 = 0$ restriction is indeed not binding, just as implied by the general *F*-test.

Save your *R* code and quit *RStudio*.

In part (f) of Exercise 2 of the last tutorial you were asked to develop a scatter diagram of the logarithm of US equity prices (*P*) and the logarithm of US dividends (*D*) from January 1900 to September 2016 to see whether there is likely to be a linear relationship between them, as predicted by the present value model. In the next exercise you are going to estimate this model.

Exercise 3 (HMPY, p. 75, Ex 3)

a) Launch *RStudio*, create a new project and script, and name them *t3e3*. Import the data from the *t2e2* Excel file, save it as *t3e3.RData* and attach it to your project.[13]

   attach(t2e2)

   Create *ts* objects from *PRICE* and *DIVIDEND* and estimate the present value model

   $$\ln P_t = \alpha + \beta \ln D_t + \varepsilon_t$$

   and write out the estimated model.

[13] Although the data set is saved as *t3e3.RData*, in the project it is still coded as *t2e2*.

Execute

PRICE = ts(PRICE, start = c(1900, 1), end = c(2016, 9), frequency = 12)
DIVIDEND = ts(DIVIDEND, start = c(1900, 1), end = c(2016, 9), frequency = 12)

pvm = lm(log(PRICE) ~ log(DIVIDEND))
summary(pvm)

to get

```
Call:
lm(formula = log(PRICE) ~ log(DIVIDEND))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.00487 -0.19920  0.01812  0.20341  0.77718 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.078607   0.008951   343.9   <2e-16 ***
log(DIVIDEND) 1.230732   0.005287   232.8   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2928 on 1399 degrees of freedom
Multiple R-squared:  0.9748,	Adjusted R-squared:  0.9748 
F-statistic: 5.418e+04 on 1 and 1399 DF,  p-value: < 2.2e-16
```

The sample regression equation is

$$\widehat{\ln P_t} = \hat{\alpha} + \hat{\beta} \ln D_t = 3.078607 + 1.230732 \ln D_t$$

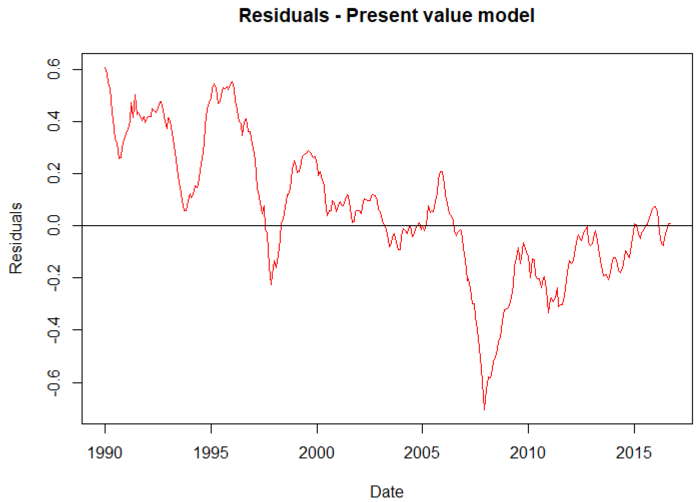
It fits to the data extremely well (R^2 = 0.9748) and significant (*F*-test).

b) Examine the properties of the estimated model by performing the following diagnostic tests:

   i. Plot the ordinary least squares residuals and interpret their time series patterns.
   ii. Test for autocorrelation of orders 1 to 6.
   iii. Test for second order *ARCH* errors.
   iv. Test for normality.

Develop the residual plot by executing

pvm.res = ts(residuals(pvm), start = c(1990, 1), end = c(2016, 9), frequency = 12)
plot.ts(pvm.res, xlab = "Date", ylab = "Residuals",
main = "Residuals - Present value model", col = "red")
abline(h = 0)



There are long runs (sequences) of positive residuals and long runs of negative residuals. The fact that the residuals change sign infrequently is a clear indication of positive first order autocorrelation in the residuals.

To see whether there is indeed significant autocorrelation of orders 1-6 in the residuals, we can perform the *BG* test for H_0: no residual autocorrelation up to order 6 against H_A: some 1-6 order residual autocorrelation by executing

library(lmtest)
bgtest(CAPM.Exxon, order = 12, type = "Chisq")

```
        Breusch-Godfrey test for serial correlation of order up to 6

data:  pvm
LM test = 1367.3, df = 6, p-value < 2.2e-16
```

The *p*-value is practically zero, so we can reject the null hypothesis at any reasonable significance level and conclude that there is evidence of autocorrelation of orders 1-6.

The *LM* test for second order *ARCH* errors has the following hypotheses:

H_0: no first-order and second-order autocorrelation in the squared stochastic error terms,
H_A: first-order or/and second order autocorrelation in the squared stochastic error terms.

It is performed by executing the following commands:

library(FinTS)
ArchTest(pvm.res, lags = 2)

They return,

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  pvm.res
Chi-squared = 303.87, df = 2, p-value < 2.2e-16
```

The *p*-value is practically zero, so we reject H_0 and conclude that there is evidence of first or/and second order *ARCH* errors.

The last test is the *JB* test for normality of the residuals. H_0: normal distribution is tested against H_A: non-normal distribution by performing the

library(tseries)
jarque.bera.test(pvm.res)

commands. They return,

```
        Jarque Bera Test

data:  pvm.res
X-squared = 0.6999, df = 2, p-value = 0.7047
```

The *p*-value is about 0.70. Hence, H_0 is maintained at any reasonable significance level, so the stochastic error terms might be normally distributed.[14]

c) In part (a) we estimated the present value model

$$\ln P_t = \alpha + \beta \ln D_t + \varepsilon_t$$

If you review page 12 of Tutorial 2, you can see that this specification came from the alleged linear relationship between the logarithms of P_t and D_t:

$$\ln P_t = -\ln \delta + \ln D_t$$

where δ_t is the discount factor at time *t* and the slope parameter is $\beta = 1$. Test this restriction and interpret the result. If the restriction is satisfied, re-estimate the present value model subject to this restriction.

Perform a general *F*-test for H_0: $\beta = 1$ against H_A: $\beta \neq 1$ by executing

[14] Be careful with the conclusion. In statistical tests based on the classical approach, which considers the Type I error more harmful (costly) than the Type II error, a non-rejected null hypothesis is a very weak evidence in favour of the null hypothesis. Therefore, it would be a mistake to conclude this time that "there is evidence the error term is normally distributed" or something similar.

library(car)
linearHypothesis(model = FF3, c("SMB + HML = 0"))

You should get

```
Linear hypothesis test

Hypothesis:
log(DIVIDEND) = 1

Model 1: restricted model
Model 2: log(PRICE) ~ log(DIVIDEND)

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1400 283.17
2   1399 119.93  1    163.24 1904.2 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The *p*-value is practically zero, so we reject H_0 and conclude that the slope parameter is different from one. This means that the restricted or pure form of the present value is not supported by the data, so we stick to its unrestricted form.

d) Estimate the discount factor used to compute the present value of the dividend stream.

The point estimate of the discount factor can be obtained from the *y*-intercept estimate of the present value model, since

$$\alpha = -\ln \delta \quad \rightarrow \quad \delta = e^{-\alpha}$$

Thus,

$$\hat{\delta} = e^{-\hat{\alpha}} = e^{-3.078607} = 0.0460$$

so the estimate of the implied discount factor is about 4.6%.

Save your *R* code and quit *RStudio*.