Quantitative Analysis of Finance I

ECON90033

WEEK 2

LINEAR REGRESSION MODELS

THE CAPITAL ASSET PRICING MODEL

Reference:

HMPY: § 3.1-3.4

Dr László Kónya
January 2023

THE CAPITAL ASSET PRICING MODEL (*CAPM*)

- Last week we considered the one-period simple return and logarithmic return (log-return),

$$R_t = \frac{P_t}{P_{t-1}} - 1 \quad \text{and} \quad r_t = \ln \frac{P_t}{P_{t-1}} = \Delta \ln P_t$$

which compare the (log) price of one share of an asset (e.g., stock or bond) at time t to the (log) price at time $t-1$, disregarding any possible intermediate cash flows, like dividend and coupon payments.

The distribution of a company's earnings to its shareholders

The annual interest paid on a bond, expressed as a percentage of the face value.

Unlike simple return, the total return of an investment captures both the price changes (capital gains or losses) and the income that it generates.

→ Assuming that an asset pays dividend D_t sometime between time $t-1$ and time t, its total net return is

$$R_t^* = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} + \frac{D_t}{P_{t-1}}$$

simple net return + dividend yield

- What factors determine the total net return of an asset?

  The total net return is uncertain at time $t-1$, it can be observed only at time t. An investor can only rely on the conditional expected return, $E\,[R_t^* \mid \Omega_{t-1}]$, where Ω_{t-1} is the set of all available information at time $t-1$.

  In addition to the actual value of the expected return, another important feature of an asset is the risk associated with it.

  Assuming that $R_t^* \sim N\,(R^*\,;\,\sigma^2)$, risk is typically measured by σ.

  The typical investor is risk averse, that is given R^*, he/she prefers lower risk to higher risk. Hence, if the risk on an investment is expected to be relatively high, it must be compensated by higher-than-average return.

This compensation for risk is called excess return or risk premium, and it is the difference between the expected return on a given asset, R^*, and the expected return on a risk-free ($\sigma_f = 0$) asset, R_f,

expect return

$$\boxed{R^* - R_f}$$

Risk-free return

(Note: To be consistent with the week 1 notes, we use different notations than the prescribed text, namely $R^* = r_t$ and $R_f = r_{ft}$).

- If an investor holds a portfolio of assets, the expected return on the portfolio (R_p) is the weighted average of the expected returns (R_1, R_2).

⟶ Assuming that there are only two types of assets in the portfolio

$$\boxed{R_p = wR_1 + (1-w)R_2}$$

where w is the weight of the first asset, measured as its proportion in the portfolio.

The variance of this portfolio (σ_p^2) is

Variance
↳ a measure of dispersion i.e. how far NO.s spread from avg value

$$\boxed{\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\sigma_{12}}$$

where σ_1, σ_2 are the standard deviations of R_1 and R_2, respectively, and σ_{12} is the covariance between R_1 and R_2.

By definition,

$$\sigma_{12} = \rho_{12}\sigma_1\sigma_2$$ where ρ_{12} is the correlation coefficient between R_1, R_2.

$$\longrightarrow \quad \sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho_{12}\sigma_1\sigma_2$$

$\longrightarrow$ Given w, σ_1, σ_2, the variance of the portfolio is an increasing function of ρ_{12}.

If $\rho_{12} \geq 0$, σ_p^2 takes its largest value when $\rho_{12} = 1$, i.e., the returns on the two assets are perfectly correlated, and σ_p^2 takes its smallest value when $\rho_{12} = 0$, i.e., the returns are perfectly uncorrelated.

$\rho_{12} < 0$ is also possible. In fact, in the most extreme and unlikely case of $\rho_{12} = -1$, $w = 0.5$, $\sigma_1 = \sigma_2$, $\sigma_p^2 = 0$, i.e., the portfolio is risk-free.

- Based on the previous formulas, the expected return and variance of a portfolio consisting of asset i and the risk-free asset f are

$$R_p = wR_i + (1-w)R_f$$

$$\sigma_p^2 = w^2\sigma_i^2 + (1-w)^2\sigma_f^2 + 2w(1-w)\rho_{if}\sigma_i\sigma_f = w^2\sigma_i^2 \quad \longleftarrow \quad \sigma_f = 0$$

$$w = \frac{\sigma_p}{\sigma_i}$$

$$R_p = wR_i + (1-w)R_f = R_f + w(R_i - R_f) = R_f + \frac{\sigma_p}{\sigma_i}(R_i - R_f)$$

$$R_i - R_f = \frac{\sigma_i}{\sigma_p}(R_p - R_f)$$

Risk premium equal ratio of correlation of of asset against market, multiplied by the mkt RP

Suppose now that portfolio *p* is the portfolio of the entire market, *m*.

Risk prem.
↳ Asset return minus risk-free return

$$\underbrace{R_i - R_f} = \frac{\sigma_i}{\sigma_m}(R_m - R_f)$$

Capital Asset Pricing Model (CAPM)

This is a purely deterministic economic model.

To obtain a statistical model, let's augment it with an intercept (α) and a stochastic error term (ε_i), denote the slope parameter as β, and attach a time index (t) to the variables.

∴ Risk premium of asset depends on it correlation to market RP?

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

This simple linear regression model of the relationship between the excess return on asset *i* ($y_t = R_{it} - R_{ft}$) and the excess return on the market ($x_t = R_{mt} - R_{ft}$) is the empirical version of the CAPM.

If x_t and y_t have a joint normal distribution, the parameters are

$$\boxed{\beta = \frac{\sigma_{xy}}{\sigma_x^2}} \quad \text{and} \quad \boxed{\alpha = E(y) - \beta E(x)}$$

β = systematic risk of asset

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.

Based on it, an individual stock or a portfolio is classified as

Parameter value tells you bus. char.

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)

The intercept 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.

Time series linear regression

- The CAPM is supposed to satisfy assumptions about the available sample data, the underlying population regression model, and the conditional distributions of the ε error terms ($\varepsilon_t : \varepsilon \mid x_t$, $t = 1, 2, \ldots, T$).

TSLR1: The model is estimated from a random sample of $T > 2$ statistically independent but identically distributed pairs of observations that satisfy the population regression model,

$$\boxed{y_t = \alpha + \beta x_t + \varepsilon_t \quad , \quad t = 1, 2, \ldots, T}$$

TSLR2: Each random error has zero conditional expected value, i.e.,

$$\boxed{E(\varepsilon_t) = 0}$$

TSLR3: The conditional variance of the random error is constant, i.e.,

$$\boxed{Var(\varepsilon_t) = \sigma^2} \longrightarrow \varepsilon_t \text{ is homoskedastic.}$$

TSLR4: Conditional on the independent variable, the random errors in any two different time periods ($t \neq t'$) are uncorrelated, i.e.

$$\boxed{E(\varepsilon_t \varepsilon_{t'}) = 0} \longrightarrow$$

ε_t is serially uncorrelated; or, in other words, it is not autocorrelated.

TSLR5: In the sample (and thus in the population as well), the independent variable is not constant.

$\longrightarrow$ There is not perfect multicollinearity, so it is possible to estimate the model from the sample at hand.

If 5 ass are satisfied then OLS is the best linear estimate possible

- Gauss-Markov Theorem: *BLUE*

  Under assumptions *TSLR1-TSLR5*, the OLS estimators of the regression parameters are the best (i.e., has the smallest variance) in the class of all linear unbiased estimators.

Note: The assumptions were stated for simple linear regression models, like CAPM, but they are supposed to be met by multiple linear regression models as well, with some minor modifications. For example, when $k > 1$, *TSLR5* also excludes any perfect linear relationship among the independent variables (including the constant term). no. of x's

- In addition to assumptions *TSLR1-TSLR5*, it is customary to make three further assumptions to facilitate statistical inference about the population regression model.

TSLR6: The random errors are uncorrelated with the independent variable, i.e. (combined with *TSLR2*),

$$Cov(\varepsilon_t, x_t) = E(\varepsilon_t x_t) = 0$$

If errors are correlated w/ calculator then you can't seperate of impact of x on y from impact of ε on y

TSLR7: The random errors are normally distributed, i.e. (combined with *TSLR2* and *TSLR3*),

$$\varepsilon_t \sim N(0, \sigma^2)$$

TSLR8: The random errors are stationary and weakly dependent.

Without going into details at this stage, stationarity means that the probability distribution of ε_t is stable over time,

while weak dependence means that as *s* increases without bound, the correlation between ε_t and ε_{t-s} approaches zero.

- Granted that *TSLR2* and *TSLR6* are satisfied, the CAPM can be manipulated as follows.

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

$$(R_{it} - R_{ft})^2 = (\alpha + \beta(R_{mt} - R_{ft}))^2 + \varepsilon_t^2 + 2(\alpha + \beta(R_{mt} - R_{ft}))\varepsilon_t$$

$$E\left[(R_{it} - R_{ft})^2\right] = E\left[(\alpha + \beta(R_{mt} - R_{ft}))^2\right] + E\left(\varepsilon_t^2\right) + \underbrace{E\left[2(\alpha + \beta(R_{mt} - R_{ft}))\varepsilon_t\right]}_{2E(\alpha) + 2\beta E((R_{mt} - R_{ft})\varepsilon_t) = 0}$$

$$\longrightarrow \underbrace{E\left[\left(R_{it} - R_{ft}\right)^2\right]}_{\text{Total risk}} = \underbrace{E\left[(\alpha + \beta(R_{mt} - R_{ft}))^2\right]}_{\text{Systematic or non-diversifiable risk}} + \underbrace{E\left(\varepsilon_t^2\right)}_{\text{Idiosyncratic or diversifiable risk}}$$

The systematic and idiosyncratic risks depend on the unknown α and β parameters, but once CAPM is estimated, we can get the residuals and

$$E\left(\varepsilon_t^2\right) \approx \frac{1}{n-2}\sum_{t=1}^{T} e_t^2 = s_\varepsilon^2$$

← Estimate of the squared standard error of regression

Moreover, this decomposition is similar to the decomposition of the total sum of squares (*SST*),

$$SST = SSR + SSE \longrightarrow 1 = \frac{SSR}{SST} + \frac{SSE}{SST}$$

R^2: coefficient of determination

Consequently,

$$\frac{E\left[(\alpha + \beta(R_{mt} - R_{ft}))^2\right]}{E\left[\left(R_{it} - R_{ft}\right)^2\right]} \approx R^2 \quad \text{and} \quad \frac{E\left(\varepsilon_t^2\right)}{E\left[\left(R_{it} - R_{ft}\right)^2\right]} \approx 1 - R^2$$

Hence, R^2 is an estimate of the proportion of the total risk that is systematic, and $1 - R^2$ represents the proportion of the total risk that is idiosyncratic.

Ex 1:

Monthly data from July 1963 to March 2023 downloaded from Ken French's professional website on

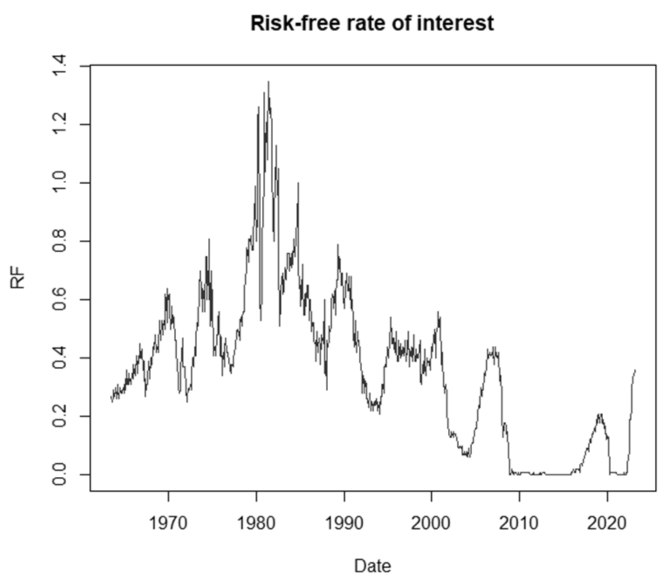
RF: risk-free rate of interest (1-month US Treasury Bill rate);

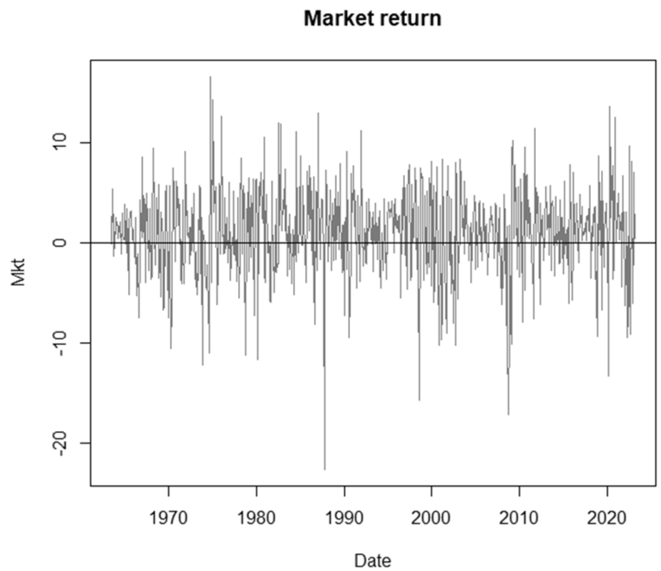
Mkt: value-weighted average stock market return to all firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ;

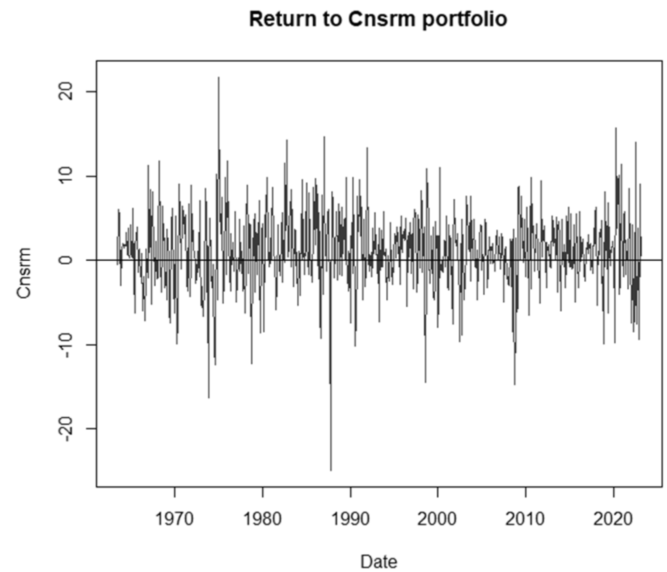
Cnsrm: return to an industry portfolio that includes consumer durables and nondurables, wholesale, retail, and some services.

(http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

```
plot.ts(RF, xlab = "Date", ylab = "RF",
   main = "Risk-free rate of interest", col = "red")
plot.ts(Mkt, xlab = "Date", ylab = "Mkt",
   main = "Market return", col = "aquamarine4")
plot.ts(Cnsrm, xlab = "Date", ylab = "Cnsrm",
   main = "Return to Cnsrm portfolio", col = “brown4")
```







The large number of observations (T = 717) makes these plots a bit messy. Yet, it seems that the returns move somewhat similarly.

Excess returns:

ER.Mkt = ts(Mkt - RF, frequency = 12,
start = c(1963,7), end = c(2023,3))
ER.Cnsmr = ts(Cnsmr - RF, frequency = 12,
start = c(1963,7), end = c(2023,3))

cor(ER.Mkt, ER.Cnsmr)

```
[1] 0.9107641
```

This sample Pearson correlation coefficient indicates that the market excess return (*Mkt*) and the excess return to the consumer portfolio (*Cnsmr*) are strongly and positively correlated with each other.

a) Calculate the common descriptive statistics for the two excess return series and briefly comment on them.

define terms for Anki

```
library(pastecs)
round(stat.desc(cbind(ER.Mkt, ER.Cnsmr),
    basic = FALSE, desc = TRUE, norm = TRUE), 3)
```

	ER.Mkt	ER.Cnsmr
median	0.920	0.850
mean	0.556	0.634
SE.mean	0.168	0.172
CI.mean.0.95	0.330	0.337
var	20.221	21.176
std.dev	4.497	4.602
coef.var	8.083	7.255
skewness	-0.499	-0.297
skew.2SE	-2.735	-1.624
kurtosis	1.724	2.401
kurt.2SE	4.729	6.586
normtest.W	0.980	0.979
normtest.p	0.000	0.000

Note: Students are supposed to be familiar with the statistics on this printout.

A few observations:

The consumer portfolio tends to outperform the market (sample mean), but it is more volatile than the market (sample standard deviation).

Yet, compared to the respective means, *Mkt* is more volatile than *Cnsmr* (sample coefficient of variation).

Both returns are skewed to the left (sample skewness), and have relatively more unusually small or large (i.e., extreme) values than a normal distribution (sample excess kurtosis > 0, i.e., leptokurtic).

None of the returns seems normally distributed (skew.2SE, kurt.2SE, and the Shapiro-Wilk test for normality).

b) Estimate the CAPM for the consumer portfolio and comment on the results.

m.Cnsmr = lm(ER.Cnsmr ~ ER.Mkt)
summary(m.Cnsmr)

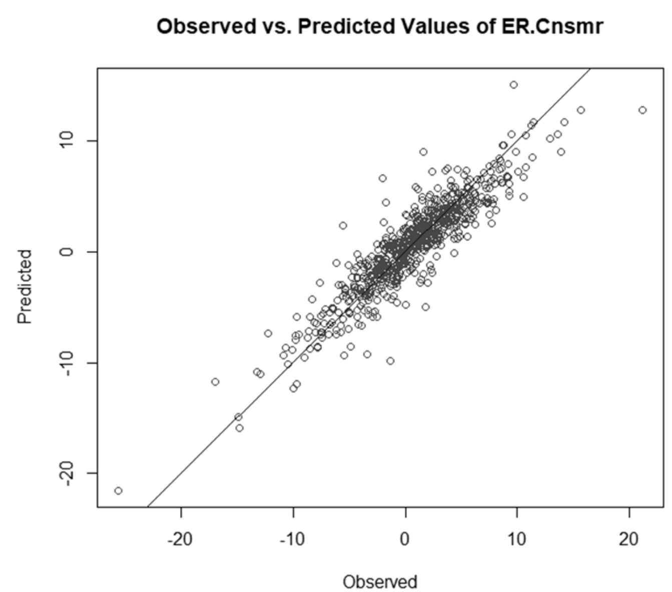
```
Call:
lm(formula = ER.Cnsmr ~ ER.Mkt)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.6577 -1.1059 -0.0551  1.0258  8.5854 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.11572    0.07156   1.617    0.106    
ER.Mkt       0.93201    0.01580  58.977   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.902 on 715 degrees of freedom
Multiple R-squared:  0.8295,	Adjusted R-squared:  0.8293 
F-statistic:  3478 on 1 and 715 DF,  p-value: < 2.2e-16
```

ER.Cnsmr_hat = ts(predict(m.Cnsmr),
frequency = 12, start = c(1963,7),
end = c(2023,3))
plot(ER.Cnsmr, ER.Cnsmr_hat,
xlab = "Observed", ylab = "Predicted",
main = "Observed vs. Predicted
Values of ER.Cnsmr",
pch = 1, col = "darkgreen", cex = 1)
abline(a = 0, b = 1, col = "blue")



The sample coefficient of determination (R^2) is about 0.83, suggesting that this simple linear regression accounts for about 83% of the variations in *ER.Cnsrm* in the sample at hand.

The scatter plot of the observed versus predicted values also shows that this regression fits to the data reasonably well.

The F-test for the overall significance of the regression model serves to test

$$H_0: \beta = 0 \text{ against } H_A: \beta \neq 0, \text{ or equivalently, } H_0: \rho^2 = 0 \text{ against } H_A: \rho^2 > 0,$$

where ρ^2 is the population coefficient of determination. INTERPRET TEXT

The p-value of this test is practically zero, hence we can safely reject the null hypothesis and conclude at any reasonable significance level that (i) this regression model is significant, and (ii) R^2 is significantly positive.

For simple linear regression models this F-test is equivalent to a two-tail t-test for the slope with zero hypothetical parameter value, so we can also conclude that the consumer portfolio has a significant β - risk. → Interpretation of this detail is expected

The significant slope estimate suggests that a one percentage point (pp) increase of the excess return to the market is expected to be accompanied by an about 0.932 pp rise of the excess return to the consumer portfolio.

The p-value of the t-test on the intercept with H_0: $\alpha = 0$ and H_A: $\alpha \neq 0$ is 0.106, so we maintain the null hypothesis even at the 10% significance level and conclude that the consumer portfolio has an insignificant α - risk.

Although in general it is not recommended to interpret insignificant coefficients, for the sake of illustration, the intercept estimate implies that when the return to the market is zero, the excess return to the consumer portfolio is about 0.116 pp.

c) Obtain and test the residuals to see whether *TSLR1*, *TSLR3*, *TSLR4* and *TSLR7* are likely satisfied.

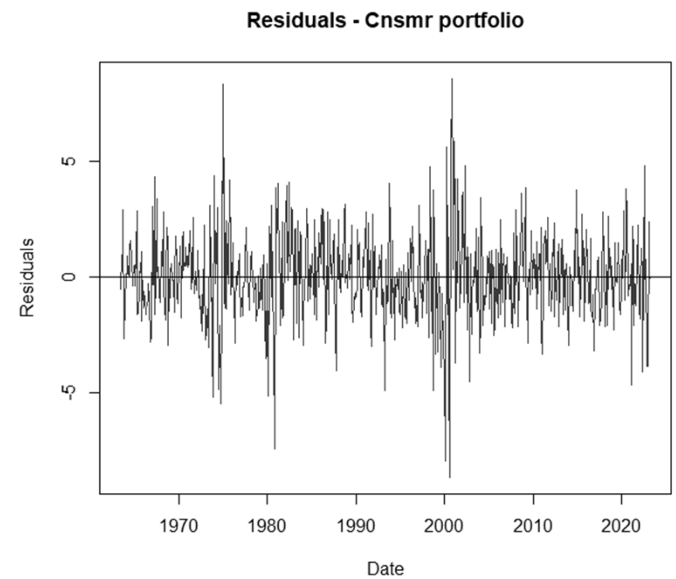
Include tests in linear regression

Anki Qs

$$e_t = y_t - \hat{y}_t$$

m.Cnsmr_res = ts(residuals(m.Cnsmr),
frequency = 12, start = c(1963,7), end = c(2023,3))

plot.ts(m.Cnsmr_res, xlab = "Date", ylab = "Residuals",
main = "Residuals - Cnsmr portfolio",
col = "violetred4")
abline(h = 0)



The residuals seem to have some cycle and their variance appears to vary over the sample.

White (*W*) test for heteroskedasticity (*studentized Breusch-Pagan test* in *R*)

H_0: homoskedasticity vs. H_A: heteroskedasticity

library(lmtest)
bptest(m.Cnsmr, ~ ER.Mkt + I(ER.Mkt^2))

```
        studentized Breusch-Pagan test

data:  m.Cnsmr
BP = 52.004, df = 2, p-value = 5.099e-12
```

The *p*-value is practically zero, so H_0 can be rejected at any reasonable significance level.

95% = 5%
never use >10%

Breusch-Godfrey (*BG*) LM test for autocorrelation in the error term:

H_0: no autocorrelation up to order 6 vs. H_A: some 1-6 order autocorrelation

library(lmtest)
bgtest(m.Cnsmr, order = 6, type = "Chisq")

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

data:  m.Cnsmr
LM test = 19.625, df = 6, p-value = 0.003229
```

p-value < 0.01, so H_0 can be rejected even at the 1% level.

Jarque-Bera (*JB*) test for normality of the error term:

H_0: normal distribution vs. H_A: non-normal distribution

library(tseries)
jarque.bera.test(m.Cnsmr_res)

```
	Jarque Bera Test

data:  m.Cnsmr_res
X-squared = 165.17, df = 2, p-value < 2.2e-16
```

The *p*-value is practically zero, so H_0 can be rejected at any reasonable significance level.

There is one more important and frequently used diagnostic in financial econometrics that is not related directly to the *TSLR1* – *TSLR8* assumptions.

It serves to test for a specific form of heteroskedasticity, called autoregressive conditional heteroskedasticity (*ARCH* – to be discussed later in this course).

Lagrange Multiplier (*LM*) test for *ARCH* errors:

This test can be used to find out whether the squared error terms (ε_t^2) are autocorrelated. Assuming, for example, that there is at most first order autocorrelation, the test regression is

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \xi_t$$

and the hypotheses are

H_0: $\alpha_1 = 0$ vs. H_A: $\alpha_1 \neq 0$

library(FinTS)
ArchTest(m.Cnsmr_res, lags = 1)

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

data:  m.Cnsmr_res
Chi-squared = 37.941, df = 1, p-value = 7.292e-10
```

The *p*-value is practically zero, so H_0 can be rejected at any reasonable significance level.

Ramsey's Regression Specification Error Test (*RESET*):

This test can be used to detect general functional form misspecification, i.e.,

H_0: correct functional form vs. H_A: incorrect functional form

resettest(m.Cnsmr, power = 3, type = "fitted")

```
        RESET test

data:  m.Cnsmr
RESET = 2.6623, df1 = 1, df2 = 714, p-value = 0.1032
```

p-value > 0.10, so H_0 cannot be rejected not even at the 10% level.

d) Every test but the *RESET* test rejected the respective null hypothesis at the 1% significance level.

Focusing on heteroskedasticity and autocorrelation at this stage, this means that the OLS estimators of the regression parameters are not the best linear estimators, meaning that they do not have the smallest variance (i.e., they are not efficient) anymore.

A possible remedy for this inefficiency of the OLS estimators is provided by the Newey-West heteroskedasticity and autocorrelation consistent (*HAC*) standard errors.

Usual

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11572    0.07156   1.617    0.106
ER.Mkt       0.93201    0.01580  58.977   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

library(sandwich)
coeftest(m.Cnsmr,
vcov = vcovHAC(m.Cnsmr, lag = 0, prewhite = TRUE))

```
t test of coefficients:
```

HAC

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.115722   0.084431  1.3706   0.1709
ER.Mkt      0.932007   0.024538 37.9822   <2e-16 ***
```

The point estimates are the same on the two printouts (apart from the different numbers of decimals),

but the *HAC* standard errors and *t* values are different from the ordinary standard errors and *t* values.

Yet, the *t*-test results do not change.

e) Is the *Cnsmr* portfolio conservative? HOW TO DO ONE SIDED TESTS

Recall from slide #7 that a portfolio is classified conservative if its β-risk is exclusively between 0 and 1.

P-value

Hence, to answer this question, we need to perform two *t*-tests on the slope coefficient with the following hypotheses:

This print out is for a two-tail test - we want 1 tail tests.

$$H_{01}: \beta = 0 \text{ vs. } H_{A1}: \beta > 0 \text{ and } H_{02}: \beta = 1 \text{ vs. } H_{A2}: \beta < 1.$$

→ If either is false you cannot reject null - need both to reject

1. check if slope sign matches alt. hyp.
2. Divide P-value by 2

Based on the regression printout from slide #16,

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11572    0.07156   1.617    0.106
ER.Mkt       0.93201    0.01580  58.977   <2e-16 ***
```

the first *t*-test can be performed as follows.

The reported *p*-values are for two-tail *t*-tests with zero hypothesized parameter values. To perform a one-tail *t*-test, we need to (i) check whether the sign of the *t* value matches the alternative hypothesis, and if it does, (ii) compare half of the reported *p*-value to the preselected significance level.

Both of these criteria are met, so we reject H_{01} and conclude that the β-risk of the *Cnsmr* portfolio is significantly positive.

The second t-test can be performed as a general F-test.

It can be used to test the validity of *any linear equality* restriction(s) on the regression coefficients, like H_0: $\beta = 1$ vs. H_A: $\beta \neq 1$.

library(car)
linearHypothesis(m.Cnsmr,"ER.Mkt = 1")

```
Linear hypothesis test

Hypothesis:
ER.Mkt = 1

Model 1: restricted model
Model 2: ER.Cnsmr ~ ER.Mkt

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    716 2652.2
2    715 2585.2  1    66.935 18.512 1.923e-05 ***
```

This test compares the fit of the original, i.e., *unrestricted,* model (Model 2) to that of a model *restricted* by H_0 (Model 1) to see whether the restriction is binding in the sense that it deteriorates the fit to the data.

When H_0 involves a single restriction, like this time, the F-test is equivalent to a two-tail t-test, so we can use a similar decision rule as in the first t-test.

Namely, since (i) β-hat = 0.932 < 1 and (ii) p-value = $Pr(>F)$ / 2 is practically zero, we reject H_{02} and conclude that the β-risk of the *Cnsmr* portfolio is significantly smaller than one.

The two t-tests imply that the *Cnsmr* portfolio is conservative.

- The original CAPM was expanded by Eugene Fama and Kenneth French in 1992 to a three-factor model and in 2014 to a five-factor model.

  In addition to the risk premium on the market, the three-factor CAPM incorporates *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).

  The five-factor CAPM also incorporates *profitability*, which refers to the concept that companies reporting higher future earnings have higher returns in the stock market,

  and *investment*, which is the difference between the returns on conservative investment portfolios and aggressive investment portfolios.

These extra factors can be captured by

SMB (Small Minus Big): the average return on small stock portfolios minus the average return on big stock portfolios;

HML (High Minus Low): the average return on value portfolios minus the average return on growth portfolios;

RMW (Robust Minus Weak): the average return on robust operating profitability portfolios minus the average return on weak operating profitability portfolios;

CMA (Conservative Minus Aggressive): the average return on conservative investment portfolios minus the average return on aggressive investment portfolios.

$\longrightarrow$ Fama-French three-factor CAPM:

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

Fama-French five-factor CAPM:

$$\begin{aligned} R_{it} - R_{ft} &= \alpha + \beta_1 (R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t \\ &+ \beta_4 RMW_t + \beta_5 CMA_t + \varepsilon_i \end{aligned}$$

Ex 2:

Using data from Ken French's professional website, estimate the Fama-French five-factor CAPM for the *Cnsrm* portfolio.

m5.Cnsmr = lm(ER.Cnsmr ~ ER.Mkt
+ SMB + HML + RMW + CMA)
summary(m5.Cnsmr)

```
Call:
lm(formula = ER.Cnsmr ~ ER.Mkt + SMB + HML + RMW + CMA)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.3712 -1.0201 -0.0612  1.0357  7.1249 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.08648    0.06630  -1.304 0.192522    
ER.Mkt       0.96976    0.01586  61.155  < 2e-16 ***
SMB          0.12046    0.02312   5.210 2.47e-07 ***
HML         -0.02255    0.02962  -0.761 0.446708    
RMW          0.41516    0.03123  13.293  < 2e-16 ***
CMA          0.15541    0.04519   3.439 0.000617 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.696 on 711 degrees of freedom
Multiple R-squared:  0.8651,	Adjusted R-squared:  0.8642 
F-statistic: 912.1 on 5 and 711 DF,  p-value: < 2.2e-16
```

Compared to the single-factor CAPM in Ex 1, the adjusted coefficient of determination increased from 0.83 to 0.86,

and with the exception of *HML*, the slope estimates of the new factors are strongly significant, in fact, significantly positive.

The β-risk estimate (i.e., the slope estimate of *ER.Mkt*) remained practically the same.

Since three of the four new factors are significant individually, they are expected to be jointly significant as well. Still, for the sake of illustration, let's perform a general *F*-test on them.

$\longrightarrow$ H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ against H_A: at least one of β_2, β_3, β_4, β_5 is different from zero

This is a composite null hypothesis as it involves more than one, actually four, parameter restrictions.

library(car)
linearHypothesis(m5.Cnsmr,
c("SMB = 0", "HML = 0",
"RMW = 0", "CMA = 0"))

```
Linear hypothesis test

Hypothesis:
SMB = 0
HML = 0
RMW = 0
CMA = 0

Model 1: restricted model
Model 2: ER.Cnsmr ~ ER.Mkt + SMB + HML + RMW + CMA

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    715 2585.2
2    711 2045.0  4     540.2 46.953 < 2.2e-16 ***
```

As expected, the group of the four new regressors is strongly significant.

Hence, the single-factor CAPM seems to be incorrectly specified, it might suffer from omitted variables.

Yet, this is not an issue this time since the β-risk estimate did not change.

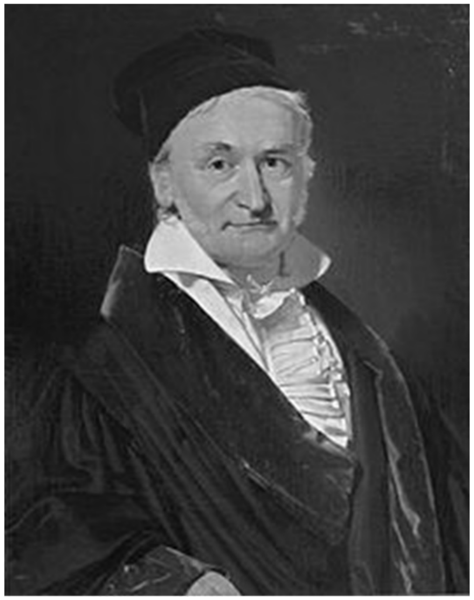
Without showing the details, apart from the *BG* test, the diagnostic tests indicate that this five-factor CAPM has the same problems as the single factor CAPM, and the *HAC* standard errors do no alter the *t*-test results.

WHAT SHOULD YOU KNOW?

- Simple and multiple linear regression – estimation, interpretation, hypothesis testing, diagnostics (heteroskedasticity, autocorrelation, normality, autoregressive conditional heteroskedasticity, functional form)
- Single-factor and multiple-factor capital asset pricing models

BOARD OF FAME

Johann Carl Friedrich Gauss (1777-1855):



German mathematician, geodesist, and physicist

Director of the astronomical observatory at the University of Göttingen

Algebra, number theory, least squares method, differential geometry and topology

Andrey Andreyevich Markov (1856-1922):

Russian mathematician

Professor at St. Petersburg University

Member of the Russian Academy of Sciences

Differential and integral calculus, Markov chain

Eugene Francis Fama (1939-):

American economist

Professor of Finance at the University of Chicago Booth School of Business

Nobel Memorial Prize in Economic Sciences in 2013

Efficient market hypothesis, Fama–French CAPM