

## FINANCIAL STATISTICS

2022-12-02

---

**Time:** 08.00 - 13.00  
**Place:** TBD: **Data Hall 6**, Albano Hus 1  
**Approved aid:** available at the homepage of the exam; calculator

**Questions about the content of the exam:** Examiner plans to stop by the exam venue at the start of the exam and after about 1.5-2h. Otherwise, check with the personal that supervises the exam.

You are expected to upload R-packages you need and link the corresponding libraries. There is a separate file provided on the homepage of the exam. Below are examples of possible instructions:

```
install.packages("forecast")  
library(forecast)  
install.packages("tseries")  
library(tseries)  
install.packages("lmtest")  
library(lmtest)  
install.packages("FinTS")  
FinTS::ArchTest()
```

If some package you want to use is missing from the provided list, indicate clearly what package and for what purpose you use. No IT help will be provided for this step, except for if R-Studio does not work as expected.

If you ask a question during the exam, you will get a clarifying answer about the content of the exam question, not how to solve it.

- **Part I (Time Series, R-Studio based)**

You are asked to analyze **WEEKLY** stock data using R and submit a written assignment in essay form. Your essay should contain output from R as well as clear interpretation of results. You are also asked to attach R code for the outputs you use to motivate your answers. In case no R code is attached, the final grade will be reduced.

- **Part II (Exercises)**

You are expected to provide detailed solutions to the asked questions and write them clearly on the paper provided at the exam. Submit your solutions on time. Answers without solutions are not accepted.

**Reminder (both parts): all the sheets you are submitting should contain your anonymous code and be ordered the way you want your answers are read. If labelling of the pages is not properly done, we cannot guarantee that the pages that are not marked properly will be graded.**

**Upload Deadline: 13:00** (no submissions after the deadline will be graded)

- **Grading, minimum requirements:**

- E: fully answer “Part I” a-d. Satisfactory answers in “Part II” or remaining of “Part I” can compensate for somewhat lacking but not wrong answer in d) of Part I
- C: fully answer Part I a-d, answer e-g in a satisfactory way; make a fair attempt on Part II. Satisfactory answer to Part II may compensate for lacking in e-g of Part I
- A: answer Part I in full with minor deficiencies. Answer Part II in a satisfactory way

These are preliminary requirements. The examiner might adjust the scale upward, if necessary.

- NOTE! Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement for a higher grade.
- If you are not satisfied with your grade, you may contact the examiner for further instructions.

**GOOD LUCK**

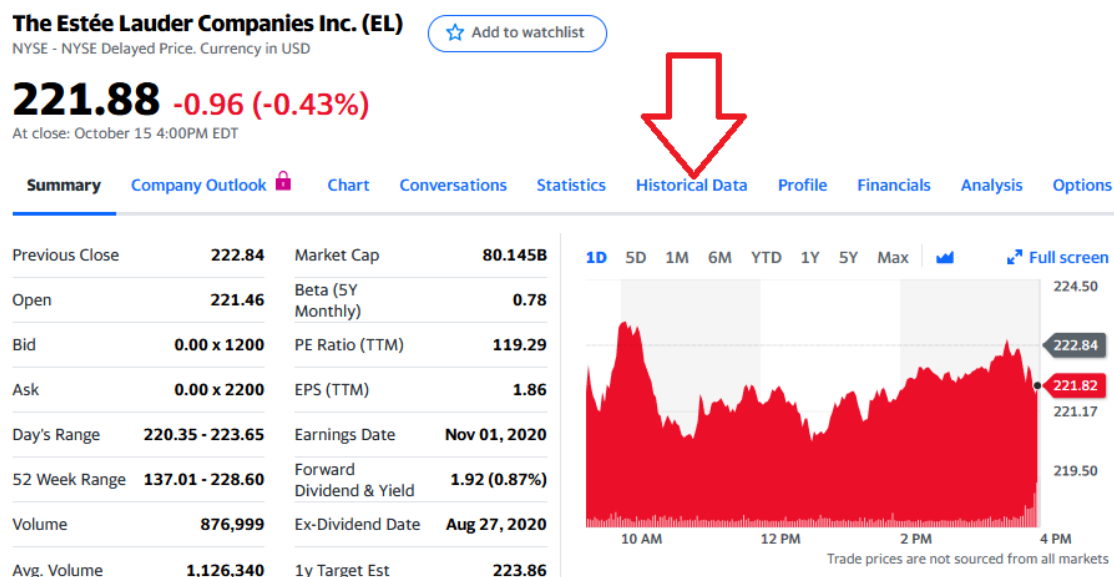
---

### **Part I, Data**

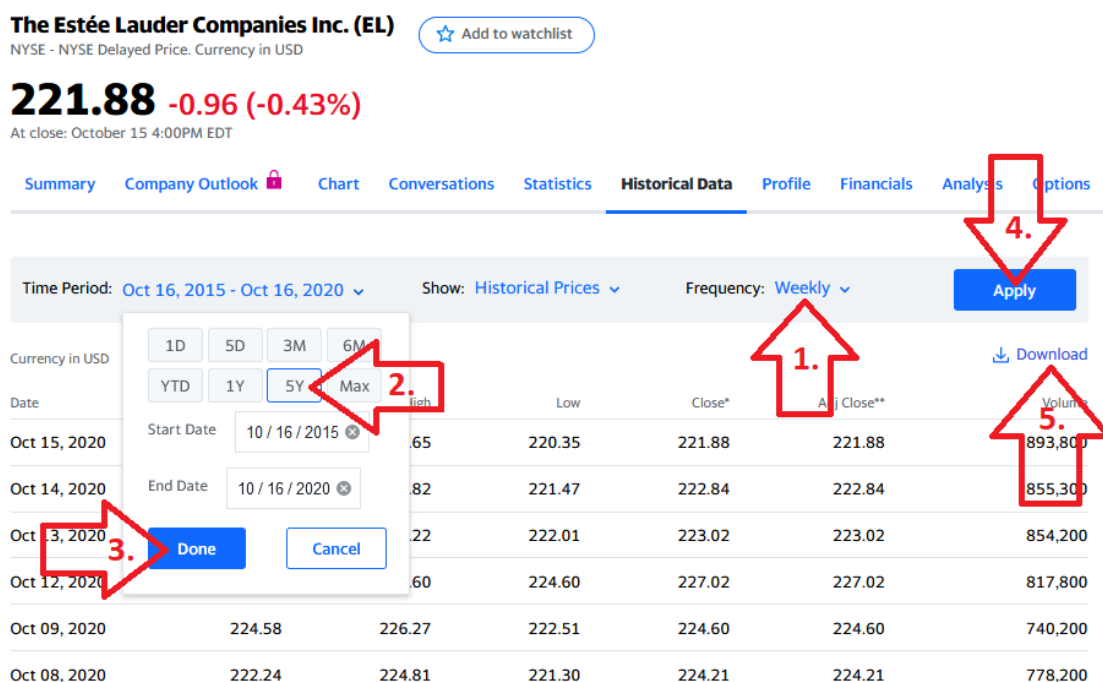
Below, you will find a selection of companies from the S&P 500. Your data corresponds to the number from the anonymous code, assigned to you for this exam. For instance, if you have the anonymous code 311-0001-XXX, your data is “MACY’s”. Follow these instructions to download your stock:

1. Go to <https://finance.yahoo.com/>
2. Enter the “Symbol” for your stock into the Yahoo! Finance search field and press enter.  
*For example, if your stock is The Estee Lauder Companies Inc. enter “EL” into the search field and press enter.*

- Click historical data:



- Change Frequency to “**Weekly**”; Click Time Period, choose “**5Y**”, and click “**Done**”; Click “**Apply**”; finally, click “**Download**.” Make sure that you have five years of weekly data, or slightly more (263 observations or so). If your stock does not have 5 years of data, choose “**Max**.” In this case, you will of course have less than 263 observations.



- You may read the data into R any way you want; the following step is just a suggestion. Open a new .R-file and paste in the following code:

```
EL <- read.csv("EL.csv")
```

This assumes that you have copied the downloaded stock data file to your work directory. Remember that you can find out what your work directory is with the command

```
getwd()
```

You can list the files in your work directory with

### **list.files()**

If you are having trouble loading the file, first make sure that your file is in the work directory. Save the .R-file and run it. Then verify that the data is correctly read into R.

Use Adj.Close price.

Code	Symbol	Name
0001	M	Macy's Inc.
0002	WMT	Walmart Inc.
0003	SNAP	Snap Inc.
0004	AAPL	Apple Inc.
0005	KO	The Coca-Cola Company
0006	Telia.st	Telia Company AB
0007	SHOP	Shopify Inc.
0008	ERIC	Telefonaktiebolaget LM Ericsson
0009	CL=F	Crude Oil
0010	SBUX	Starbucks Corporation
0011	^VIX	CBOE Volatility Index
0012	GOOGL	Alphabet Inc. Class A
0013	M	Macy's Inc.
0014	WMT	Walmart Inc.
0015	SNAP	Snap Inc.
0016	AAPL	Apple Inc.
0017	KO	The Coca-Cola Company
0018	Telia.st	Telia Company AB
0019	SHOP	Shopify Inc.
0020	ERIC	Telefonaktiebolaget LM Ericsson
0021	CL=F	Crude Oil
0022	SBUX	Starbucks Corporation
0023	^VIX	CBOE Volatility Index
0024	GOOGL	Alphabet Inc. Class A
0025	M	Macy's Inc.
0026	WMT	Walmart Inc.
0027	SNAP	Snap Inc.
0028	AAPL	Apple Inc.
0029	KO	The Coca-Cola Company
0030	Telia.st	Telia Company AB
0031	SHOP	Shopify Inc.
0032	ERIC	Telefonaktiebolaget LM Ericsson
0033	CL=F	Crude Oil
0034	SBUX	Starbucks Corporation
0035	^VIX	CBOE Volatility Index
0036	GOOGL	Alphabet Inc. Class A

The goal of the assignment in Part I is to perform the analysis of your assigned stock in **R** and then write a short essay answering the questions below. Make one headline in the essay for each part that you answer, e.g. “a. Introduction and Trend” and “b. Stationarity.” It is important that you include plots/tables that are outputs from R as well as key R-commands you use.

- (a) Describe the time series with an appropriate plot/diagram. Comment on the diagram and provide summary statistics. Is there a visible trend? You do not need to consider seasonal effects this time.

For steps b, c, and d, leave out **the last 5 observations** that will be referred to as “the testing set”.

- (b) Stationarity. Use a formal test for stationarity. If it is clear from the diagram that your data is non-stationary, you may apply an appropriate transformation but perform testing first. State the hypotheses and interpret the outcome of the test. Use 5% significance level. If the data does not seem to be stationary even after transformation, attempt another transformation and repeat this step. Make sure that you have the figure containing the stationary time series.
- (c) Plot the ACF and PACF plots. Explain what the plots show. Specifically state and interpret the value of the **second (2nd)** bar in the PACF plot and explain how it has been used in the analysis. Make specific references to the models you selected for the step (d). Use these plots as the basis for your choices in (d).
- (d) Choose at least four the most promising ARIMA models that are top candidates to fit the data (leaving out the testing set). One of them should be an ARIMA(0,d,0): indicate what “d” you have chosen and why. Provide output, analysis and the AIC scores for each model. Interpret your findings. Which model seems most promising? Should any of your models be disregarded completely? Why?
- (e) Choose the two “best” models from (d). Use the last **5 observations**, i.e. your testing set, to calculate RMSE. The calculations should be explicit and easy to follow. Compare the two and interpret the difference. Moreover, visualize both predictions and discuss the figures. Make clear and well-motivated conclusions.
- (f) Pick “the best” model from question (e). Test that model for ARCH effect. State the hypothesis and explain the outcome of your test. Explain what the presence of ARCH effect would mean for your results in (d), even if you did not find any. Discuss briefly why estimating variance is important in financial forecasting.
- (g) Pick “the best” model from question (e). Write the model explicitly and discuss the parameters. Provide residual analysis of the model and report your conclusions. Further, provide overall conclusions for your analysis.

**Part II (One “Pen-and-Paper-Answer” question: one cannot use R-Studio to solve it. Write your answers on a paper.)**

**Question 1.**

As part of their business research, a dating app offers "premium membership" for the price of 149 kr per year, to a randomly selected sample of 2800 of their active members. The business research team collects data and decide on the following logistic regression model:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3.$$

$y$  = the log-odds of buying premium membership.

$x_1$  = male (1 = yes, 0 = no).

$x_2$  = age.

$x_3$  = average hours spent per week on the app.

Their output can be found here on the next page:

Call:

```
glm(formula = premium ~ male + age + hours, family =  
    binomial(link = "logit"), data = website)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.2573	-0.2745	-0.2202	-0.1826	2.9582

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-4.62665	0.54158	-8.543	< 2e-16 ***
male	0.56685	0.21932	2.585	0.00975 **
age	0.01276	0.02017	0.633	0.52695
hours	0.34772	0.03329	10.446	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 934.12 on 2799 degrees of freedom  
Residual deviance: 830.49 on 2796 degrees of freedom  
AIC: 838.49

Number of Fisher Scoring iterations: 6

- Test at the 1% level whether the parameter for "age" is significantly different from zero, given the present model. Would you typically keep this variable in the model?
- Suppose that true odds-ratio of the variable "age" is one. What would the interpretation of that odds-ratio that be?
- Calculate and interpret the odds-ratio for the variable "male." What does the double star next to the  $p$ -value mean?
- The average person in the sample is 25 years old and spends 2 hours per week on the app. Given the present model, find the estimated probability that a person who is female, 25 years old, and spends 2 hours per week on the app buys premium membership.
- Find the odds ratio of the variable 'hours'. Use this and your answer in (d) to find the change in probability if we increase hours spent per week by one compared to the person in (c), all else being held constant.