

FINANCIAL STATISTICS

2023-04-21

Time:	14.00 - 19.00
Place:	TBD
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. Preferably, you do it in plain R, rather than in the RStudio. See a separate help file for R-commands available on the homepage of the exam. Below are examples of possible instructions if you decide to make separate uploads:

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

If an R-package you want to use is missing from the provided list, indicate clearly in your exam report the name of the package and for what purpose you use it. 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 an answer how to solve it.

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

You are asked to analyze **DAILY** stock data **for the last year** using R and submit a written assignment in an essay form. Your essay should contain key R-commands, the output from R-Studio as well as clear interpretation of results. Submissions in .docx and .pdf formats are ok but the .pdf format is preferred. In case no key R commands are clear in the text of your answers, the final grade will be reduced.

You are also asked to attach the whole R code, preferably converted into .pdf format for the reference purposes. It is not expected to be checked.

- **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: 19:00 (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” may 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 (both questions) or solve one of the problems from Part II in a satisfactory way. Satisfactory answer to Part II (both parts) may compensate for lacking in e-g of Part I
- A: answer Part I in full with minor deficiencies. Answer Part II (both questions) 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.

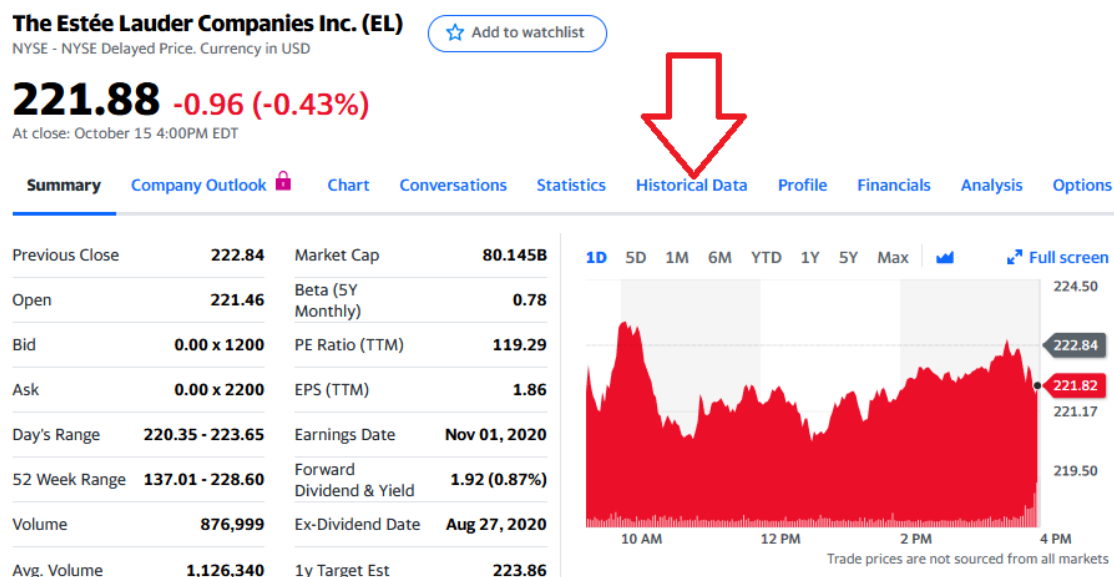
GOOD LUCK

Part I, Data

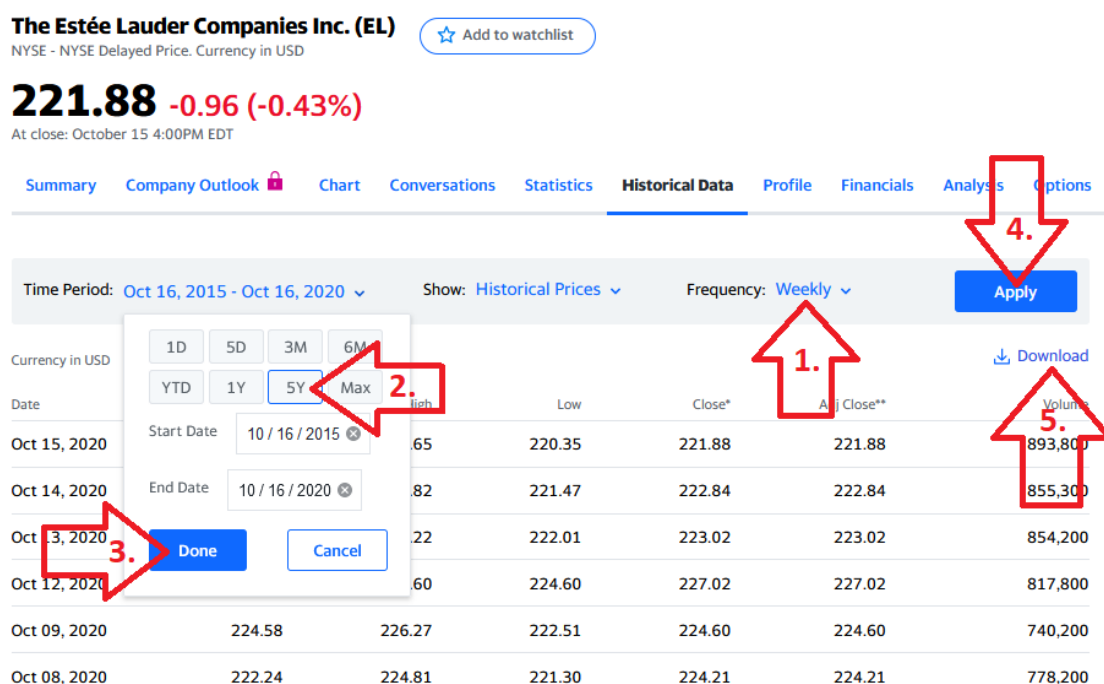
Below, you will find a selection of companies from the S&P 500. Your data corresponds to the number from the anonymous code, assigned to you for this exam. For instance, if you have the anonymous code 311-0001-XXX, your data is “WMT’s”, Walmart Inc. Follow these instructions to download your stock:

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 “Daily”; Click Time Period, choose “1Y”, and click “Done”; Click “Apply”; finally, click “Download.” Make sure that you have one year of daily data, or slightly more (252 observations or so). If your stock does not have 1 year of data, choose “Max.” In this case, you will of course have less than 252 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 when available. Otherwise, state clearly what data you use.

Code	Symbol	Name
0001	WMT	Walmart Inc.
0002	KO	The Coca-Cola Company
0003	ERIC	Telefonaktiebolaget LM Ericsson
0004	GLD	SPDR Gold Shares
0005	Telia.st	Telia Company AB
0006	^VIX	CBOE Volatility Index
0007	GOOGL	Alphabet Inc. Class A
0008	NVDA	Nvidia Corporation
0009	BTC-USD	Bitcoin USD
0010	SLV	iShares Silver Trust
0011	AZN	AstraZeneca PLC
0012	AAPL	Apple Inc.
0013	WMT	Walmart Inc.
0014	KO	The Coca-Cola Company
0015	ERIC	Telefonaktiebolaget LM Ericsson
0016	GLD	SPDR Gold Shares
0017	Telia.st	Telia Company AB
0018	^VIX	CBOE Volatility Index
0019	GOOGL	Alphabet Inc. Class A
0020	NVDA	Nvidia Corporation
0021	BTC-USD	Bitcoin USD
0022	SLV	iShares Silver Trust
0023	AZN	AstraZeneca PLC
0024	AAPL	Apple Inc.
0025	WMT	Walmart Inc.
0026	KO	The Coca-Cola Company
0027	ERIC	Telefonaktiebolaget LM Ericsson
0028	GLD	SPDR Gold Shares
0029	Telia.st	Telia Company AB
0030	^VIX	CBOE Volatility Index
0031	GOOGL	Alphabet Inc. Class A
0032	NVDA	Nvidia Corporation
0033	BTC-USD	Bitcoin USD
0034	SLV	iShares Silver Trust
0035	AZN	AstraZeneca PLC
0036	AAPL	Apple Inc.

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? Test for the trend using simple linear regression. You can choose different time intervals to “prove” that the time series has extensive time periods of trends, if any.

For steps b, c, and d, leave out **the last 10 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. In case you have to choose between mildly (visualize the time series in question) non-stationary time series and “white noise” style time series, choose the former in favor of the latter.
- (c) Plot the ACF and PACF plots. Choose the number of lags appropriately and justify your choice. Explain what the plots show. Specifically state and interpret the value of the **first (1st)** bar in the PACF plot and explain how it has been used in the analysis. Return to this question when you have chosen “the best” model. Make specific references to the models you selected for the step (d). Use these plots as the basis for your choices in (d).
- (d) Whatever the (P)ACF plots you got, 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. You have to have a detailed description of your selection process: why selected by you models are better than other potential candidates. Provide output, analysis and the AIC scores for each “potentially good” 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 **10 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. Specifically discuss the values of the RMSE for the selected models.
- (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. Visualize the ARCH effect for your model.
- (g) Pick “the best” model from question (e). Write down the model explicitly and discuss the coefficients. Discuss connection to the (P)ACF plots. Provide residual analysis of the model and report your conclusions. Further, provide overall conclusions for your analysis.

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

1. Let A denote the event: "a randomly selected student passes the Financial Statistics exam." Consider the logistic regression model, where $p(A|x)$ is the probability of passing the exam, given x number of hours studied per week

$$\log\left(\frac{p(A|x)}{1-p(A|x)}\right) = -4.5 + 3 * 0.YY * x,$$

where "YY" is exactly the number you used in Part I for selecting your individual stock. For instance, if you have your anonymous code 311-0011-XXX, the right-hand part of the equation above will be " $-4.5+3*0.11*x=-4.5+0.33*x$ ".

- (a) Find the odds-ratio of the variable x
- (b) Find the estimated number of hours one needs to study to have at least 50% probability of passing the exam
- (c) Find the probability that a randomly chosen student passes the exam, given that she studies 35 hours per week.

Moreover, describe the expected outcome if one does not study at all. If you have time and desire (optional) suggest your own logistic regression that in your view would fit the expected outcome better. Motivate your choice.

2. Provide detailed calculations for the following questions (note that the correct answer without calculations is not accepted as a solution):

- a) Derive the autocorrelation function for the following MA(1) process $Z_t = \varepsilon_t + 1.4\varepsilon_{t-1}$. Assume that Z_t is stationary.
- b) Derive the autocorrelation function for the following MA(2) process $Z_t = \varepsilon_t + 1.4\varepsilon_{t-1} - 0.6\varepsilon_{t-2}$. Assume that Z_t is stationary.
