STOCKHOLM UNIVERSITY
Department of Statistics
Spring 2025, period A-B

Andriy Andreev (examiner)

FINANCIAL STATISTICS
2025-03-18

Time:	14.00 - 19.00
Place:	**TBD**
Approved aid:	available for download at the SEB-imbedded homepage of the exam; calculator

Questions about the content of the exam: The examiner (or the TA) plans to stop by the exam venue at the start and one more time during the exam.

To upload the R-packages use plain R. If you try to upload the packages in RStudio, your PC might become irresponsive and you need to restart the RStudio or/and computer. There is ".txt" help file that contains a function to install most of needed packages. It is easy to copy/paste. There is a separate help file for R-commands available on the homepage of the exam. After the packages are uploaded using the nearest mirror site (usually, Umeå), you can link the libraries using RStudio. FinTS package is wise to unpack only at the end, when one tests for the ARCH effect. If you need to upload some additional packages, do so using plain R. 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 is available for this step.

Remember to save your progress in RStudio frequently in case you need to re-start RStudio/PC.

Below are the 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 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. No IT help is available during the exam.

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

  Your task is to analyze **DAILY** stock data provided in .csv format using RStudio and submit a written report in an essay form. Your essay should contain the 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, the key R commands are not clearly incorporated in the

text of your answers, the final grade will be drastically reduced. In case there is no output or interpretation of the output, zero credit is given.

Please <u>attach your complete R-code</u> as a separate file for the reference purposes. The file that contains the R-code is not graded and rarely looked into for the grading purposes. All important information should be present in your report.

- **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. The answers (correct or not) without solutions receive zero credit for grading purposes.

Reminder: the first page of your answers for Part1 and ALL the sheets you are submitting as solutions for Part2 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, one cannot guarantee that the pages that are not marked properly will be graded.

Upload Deadline: 19:00 (no submissions after the deadline are graded)

- **Grading, minimum requirements:**

    - <u>E:</u> fully answer "Part I" a-d and make a fair attempt on remaining of "Part I" & at least one of the questions in "Part II". Satisfactory answers in "Part II" or remaining of "Part I" may compensate for somewhat lacking but nor wrong answers in a)-d) of "Part I"
    - <u>C:</u> fully answer Part I a-d, answer e-g in a satisfactory way; make a fair attempt on Part II (two questions) or solve one of the problems from Part II in a satisfactory way. Satisfactory answer to Part II (two questions) may compensate for lacking in e-g of Part I
    - <u>A:</u> answer Part I in full with minor deficiencies. Answer Part II (two 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 <u>cannot</u> 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 time series data sets. 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 "0f08_uk_d", Kongberg Gruppen. Upload available .csv file as explained during lectures/labs.

Code	Symbol	Name
0001	0f08_uk_d	Kongsberg Gruppen
0002	0gwl_uk_d	Saab AB
0003	0iw5_uk_d	Thales
0004	0ni1_uk_d	Rheinmetall AG
0005	ba_uk_d	BAE Systems
0006	lmt_us_d	Lockheed Martin Corporation
0007	panw_us_d	Palo Alto Networks
0008	rtx_us_d	Raytheon Technologies
0009	tsla_us_d	Tesla
0010	crwd_us_d	Crowd Strike
0011	0f08_uk_d	Kongsberg Gruppen
0012	0gwl_uk_d	Saab AB
0013	0iw5_uk_d	Thales
0014	0ni1_uk_d	Rheinmetall AG
0015	ba_uk_d	BAE Systems
0016	lmt_us_d	Lockheed Martin Corporation
0017	panw_us_d	Palo Alto Networks
0018	rtx_us_d	Raytheon Technologies
0019	tsla_us_d	Tesla
0020	crwd_us_d	Crowd Strike
0021	0f08_uk_d	Kongsberg Gruppen
0022	0gwl_uk_d	Saab AB
0023	0iw5_uk_d	Thales
0024	0ni1_uk_d	Rheinmetall AG
0025	ba_uk_d	BAE Systems
0026	lmt_us_d	Lockheed Martin Corporation
0027	panw_us_d	Palo Alto Networks
0028	rtx_us_d	Raytheon Technologies
0029	tsla_us_d	Tesla
0030	crwd_us_d	Crowd Strike
0031	0f08_uk_d	Kongsberg Gruppen
0032	0gwl_uk_d	Saab AB
0033	0iw5_uk_d	Thales
0034	0ni1_uk_d	Rheinmetall AG
0035	ba_uk_d	BAE Systems
0036	lmt_us_d	Lockheed Martin Corporation
0037	panw_us_d	Palo Alto Networks
0038	rtx_us_d	Raytheon Technologies
0039	tsla_us_d	Tesla
0040	crwd_us_d	Crowd Strike
0041	0f08_uk_d	Kongsberg Gruppen
0042	0gwl_uk_d	Saab AB
0043	0iw5_uk_d	Thales

0044	0ni1_uk_d	Rheinmetall AG
0045	ba_uk_d	BAE Systems
0046	lmt_us_d	Lockheed Martin Corporation
0047	panw_us_d	Palo Alto Networks
0048	rtx_us_d	Raytheon Technologies
0049	tsla_us_d	Tesla
0050	crwd_us_d	Crowd Strike
0051	0f08_uk_d	Kongsberg Gruppen
0052	0gwl_uk_d	Saab AB
0053	0iw5_uk_d	Thales
0054	0ni1_uk_d	Rheinmetall AG
0055	ba_uk_d	BAE Systems
0056	lmt_us_d	Lockheed Martin Corporation
0057	panw_us_d	Palo Alto Networks
0058	rtx_us_d	Raytheon Technologies
0059	tsla_us_d	Tesla
0060	crwd_us_d	Crowd Strike
0061	0f08_uk_d	Kongsberg Gruppen
0062	0gwl_uk_d	Saab AB
0063	0iw5_uk_d	Thales
0064	0ni1_uk_d	Rheinmetall AG
0065	ba_uk_d	BAE Systems
0066	lmt_us_d	Lockheed Martin Corporation
0067	panw_us_d	Palo Alto Networks
0068	rtx_us_d	Raytheon Technologies
0069	tsla_us_d	Tesla
0070	crwd_us_d	Crowd Strike
0071	0f08_uk_d	Kongsberg Gruppen
0072	0gwl_uk_d	Saab AB
0073	0iw5_uk_d	Thales
0074	0ni1_uk_d	Rheinmetall AG
0075	ba_uk_d	BAE Systems
0076	lmt_us_d	Lockheed Martin Corporation
0077	panw_us_d	Palo Alto Networks
0078	rtx_us_d	Raytheon Technologies
0079	tsla_us_d	Tesla
0080	crwd_us_d	Crowd Strike
0081	0f08_uk_d	Kongsberg Gruppen
0082	0gwl_uk_d	Saab AB
0083	0iw5_uk_d	Thales
0084	0ni1_uk_d	Rheinmetall AG
0085	ba_uk_d	BAE Systems

The goal of the assignment in Part I is to analyze assigned time series in RStudio and then write an essay answering the questions below. Make one headline in the essay for each part that you answer, e.g. "a) Describe the time series and test for trend" and "b) Stationarity." Preferably, use a new page when you answer a new question. It is important that you include plots/tables that are outputs preceded by the key R-commands. Use "Close" column from the data. In case you need to convert the character format of the date in the first column, you might want to use the following (or similar) command: Ticker$Date←as.Date(Ticker$Date, format = "%d / %m / %Y")

(a) <u>Describe the time series and test for trend</u> with an appropriate plot/diagram. Comment on the diagram and provide summary statistics. What properties of "stationarity" look like not fulfilled? 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 7 observations that will be referred to as "the testing set".

(b) <u>Stationarity</u>. Use a formal test for stationarity and for detecting a trend in the time series. If it is clear from the diagram that your data is non-stationary, you may apply an appropriate transformation but perform both tests first. State the hypotheses and interpret the outcome of each test. Use 5% significance level. If the data does not seem to be stationary even after transformation, attempt another transformation and repeat the stationarity testing. Make sure that you have the figure containing the original and 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 the PACF graphs. Choose the number of lags appropriately and justify your choice. Explain what the plots show. Specifically state and interpret the values of the second (2nd) bar in both ACF and PACF plots. Give an outline (no exact calculations required) of how you would calculate them if no R-Studio is available. Compare the two values. What is the difference? Discuss. For the A-B grade, remember to return to these plots 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 get, choose (by performing a proper statistical analysis) at least four the most promising ARIMA models that are the top candidates to fit the data (leaving out the testing set). One of the remaining models 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. Explain why selected by you models are better than other potential candidates. Provide output, analysis and calculate the AIC scores for each "potentially good" model. Interpret your findings. Which model seems most promising? Should any of the checked models be disregarded completely? Why?

(e) Select the two "best" models from (d). Use your testing set to calculate RMSE. The calculations should be explicit and easy to follow. Compare the two and interpret the difference. If needed, perform an exploratory residual analysis to select the best model. 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 and make conclusions on the quality of the models. Finally, use Holt method to have a benchmark model. Compare

the quality predictions of the selected ARIMA models and the Holt method. Discuss the differences.

(f) Pick "the best" ARIMA model from part (e). Test it for the "ARCH effect". State the hypothesis and explain the outcome of your test. Explain what presence of ARCH effect would mean for your results in (d), even if you do not find any. Discuss briefly, why estimating variance is important in financial forecasting and for the particular stock you got for this exam. Visualize the ARCH effect for your model.

(g) Pick "the best" model from question (e). Write down the model for the initial time series 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 quality of your statistical forecast.

------------------------------------   END OF PART I   -----------------------------------

Part II ("Pen-and-Paper-Answer" questions: one cannot provide R-Studio based solutions. Write your detailed answers on paper)

1. The sinking of Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1,502 of the 2,228 passengers and crew. Using the titanic data, we fit the following multiple logistic regression:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

z= log-odds of surviving (0=Dead, 1=Alive)
x1= Class (1=first class, 0=otherwise)
x2= Gender (1=Male, 0=Female)
x3= Age

The output can be found here:

Survive ~ Class + Gender + Age				
	Estimate	Std. Error	Z-value	P-value
Constant	1.50	0.25	6.093	1.11e-09
Class	2.01	0.25	8.111	5.03e-16
Gender	-2.55	0.20	-12.629	< 2e-16
Age	-0.03	0.007	-4.058	4.95e-05

a) Calculate the survival probability of Jack and Rose.(Jack: 3rd class, male,20 years old, Rose: 1st class, female, 19 years old)

b) Calculate the 95% confidence interval of Jack and Rose's survival probability

For a "fair" solution, one is expected to solve part a) and have a clear outline of the strategy to solve part b).

2. Provide detailed calculations for the following questions. Note that the correct answer without calculations is not accepted as a solution and will be marked "empty+". For a "fair" attempt grade, one has to have a "fair" attempt on both parts of the exercise:

   a) Derive the autocorrelation function for the following MA(1) process $Z_t = \varepsilon_t - 3\varepsilon_{t-1}$. Assume that Z_t is stationary.

   b) Derive the autocorrelation function for the following MA(2) process $Z_t = \varepsilon_t - 3\varepsilon_{t-1} + \varepsilon_{t-2}$. Assume that Z_t is stationary.

   For a "fair" grade, one is expected to solve part a) and clearly show how to calculate part b). For a "satisfactory" grade, one is expected to answer both questions a) and b) with minor mistakes.

-------------- **GOOD LUCK** ------------------