STOCKHOLM UNIVERSITY
Department of Statistics
Autumn 2024, period A-B

Andriy Andreev (examiner)

FINANCIAL STATISTICS

2024-10-29

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

Questions about the content of the exam: 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 "Word" help file that contains all commands to install packages that are easy to copy/paste. Remember to use 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 **WEEKLY** 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 no key R commands are clear in the text of your answers,

the final grade will be drastically reduced, possibly to "F". In case there is no output or interpretation of the output, zero credit is given.
Please attach your complete R-code as a separate file, preferably converted into the .pdf format 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 (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 are graded)

- **Grading, minimum requirements:**
  - E: 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"
  - C: 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
  - A: 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 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 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 "nvda_us-w", Nvidia Corp. Upload available .csv file as explained during lectures/labs.

Code	Symbol	Name
0001	nvda_us-w	Nvidia Corp
0002	ag_us_w	First Majestic Silver Corp, AG
0003	amzn_us_w	Amazon.com, Inc
0004	coin_us_w	Coinbase Global, Inc
0005	djt_us_w	Trump Media & Technology Group Corp
0006	ionq_us_w	IonQ, Inc
0007	nok_us_w	Nokia Oyj
0008	tsla_us_w	Tesla Inc
0009	tsm_us_w	Tesla Inc
0010	uber_us_w	Uber Technologies, Inc.
0011	nvda_us-w	Nvidia Corp
0012	ag_us_w	First Majestic Silver Corp, AG
0013	amzn_us_w	Amazon.com, Inc
0014	coin_us_w	Coinbase Global, Inc
0015	djt_us_w	Trump Media & Technology Group Corp
0016	ionq_us_w	IonQ, Inc
0017	nok_us_w	Nokia Oyj
0018	tsla_us_w	Tesla Inc
0019	tsm_us_w	Tesla Inc
0020	uber_us_w	Uber Technologies, Inc.
0021	nvda_us-w	Nvidia Corp
0022	ag_us_w	First Majestic Silver Corp, AG
0023	amzn_us_w	Amazon.com, Inc
0024	coin_us_w	Coinbase Global, Inc
0025	djt_us_w	Trump Media & Technology Group Corp
0026	ionq_us_w	IonQ, Inc
0027	nok_us_w	Nokia Oyj
0028	tsla_us_w	Tesla Inc
0029	tsm_us_w	Tesla Inc
0030	uber_us_w	Uber Technologies, Inc.
0031	nvda_us-w	Nvidia Corp
0032	ag_us_w	First Majestic Silver Corp, AG
0033	amzn_us_w	Amazon.com, Inc
0034	coin_us_w	Coinbase Global, Inc
0035	djt_us_w	Trump Media & Technology Group Corp
0036	ionq_us_w	IonQ, Inc
0037	nok_us_w	Nokia Oyj
0038	tsla_us_w	Tesla Inc
0039	tsm_us_w	Tesla Inc
0040	uber_us_w	Uber Technologies, Inc.
0041	nvda_us-w	Nvidia Corp
0042	ag_us_w	First Majestic Silver Corp, AG
0043	amzn_us_w	Amazon.com, Inc
0044	coin_us_w	Coinbase Global, Inc

0045	djt_us_w	Trump Media & Technology Group Corp
0046	ionq_us_w	IonQ, Inc
0047	nok_us_w	Nokia Oyj
0048	tsla_us_w	Tesla Inc
0049	tsm_us_w	Tesla Inc
0050	uber_us_w	Uber Technologies, Inc.
0051	nvda_us-w	Nvidia Corp
0052	ag_us_w	First Majestic Silver Corp, AG
0053	amzn_us_w	Amazon.com, Inc
0054	coin_us_w	Coinbase Global, Inc
0055	djt_us_w	Trump Media & Technology Group Corp
0056	ionq_us_w	IonQ, Inc
0057	nok_us_w	Nokia Oyj
0058	tsla_us_w	Tesla Inc
0059	tsm_us_w	Tesla Inc
0060	uber_us_w	Uber Technologies, Inc.

The goal of the assignment in Part I is to analyze assigned time series in R and then write an 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." Preferably, use a new page when you answer a new question. It is important that you include plots/tables that are outputs from R and the key R-commands.

(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 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 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 to use it in the analysis. Compare with the same bar in the ACF. What is the difference? For the A-B grade, remember to 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 get, choose (by performing 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: 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 checked models be disregarded completely? Why?

(e) Choose 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 "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. 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.

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

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

In this part, choose to solve two out of the three below stated problems. If you solve all three problems, the two best answers form the base for grading.

1. Let us assume that we consider an ARCH-style time series X_t without a drift, i.e. $\boldsymbol{X_t = X_{t-1} + \varepsilon_t}$ with $\mathbf{Var}(\boldsymbol{\varepsilon_t}) = \boldsymbol{h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$; further, let us assume that $\boldsymbol{\alpha_0 = 0.0006}$, $\boldsymbol{\alpha_1 = 0.09}$ and the error terms to be normally distributed.

Calculate the standard deviations $\sqrt{\boldsymbol{h_3}}$, $\sqrt{\boldsymbol{h_4}}$ for the time stamps *t=3*, *t=4* and the probability $\mathbf{P}(\boldsymbol{X_4 > 11})$ if we observe that $\boldsymbol{X_1 = 9.71}$, $\boldsymbol{X_2 = 9.70}$, $\boldsymbol{X_3 = 9.90}$. Further, assume that $\boldsymbol{\alpha_0 = 0.0006}$, $\boldsymbol{\alpha_1 = 0}$ and calculate $\mathbf{P}(\boldsymbol{X_4 > 11})$.

What is the difference between two probabilities? Why? Provide "financial intuition" to the difference of the probabilities.

---

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 $Z_t = \varepsilon_t - \varepsilon_{t-1}$. Assume that Z_t is stationary.

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

---

3. 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

$$\log\left(\frac{p(A|x)}{1-p(A|x)}\right) = -4 + 0.YY * x,$$

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 + 0.11 * x".

(a) Find the odds-ratio of the variable x and provide a clear interpretation of your findings. Moreover, describe the expected exam outcome if one does not study at all, assuming that the above stated model is a good model to link time spent by student preparing and the final grade.
(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 25 hours per week.

If you have time and desire (optional) suggest your own logistic regression model that in your view would fit the expected outcome better. Motivate your choice.

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