STAT1043 (2024/25)	**Statistical Techniques and Time Series**	**Contribution: 50% of course**
Course Leader: Dr Karolos Korkas	**Time Series Analysis Coursework**	**Deadline Date:** 6th December 2024
This coursework will be marked anonymously YOU MUST NOT PUT ANY INDICATION OF YOUR IDENTITY IN YOUR SUBMISSION		
This coursework should take an average student who is up-to-date with tutorial work approximately 15 hours Feedback and grades are normally made available within 15 working days of the coursework deadline		
Learning Outcomes: 1 Demonstrate knowledge and understanding of some of the commonly used statistical techniques to analyse time series data. 2 Apply multiple linear regression models to data, carry out the required statistical analysis and reflect on results. 3 Make a choice of suitable statistical methods to analyse time series data by choosing appropriate models. 4 Understand autoregressive (AR), moving average (MA) and ARMA processes and apply them to a range of time series data, using a statistical package, interpret computer outputs of their analysis and report on findings.		

Plagiarism is presenting somebody else's work as your own. It includes copying information directly from the Web or books without referencing the material; submitting joint coursework as an individual effort; copying another student's coursework; stealing coursework from another student and submitting it as your own work. Suspected plagiarism will be investigated and if found to have occurred will be dealt with according to the procedures set down by the University. Please see your student handbook for further details of what is / isn't plagiarism.

All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.

Your work will be submitted for plagiarism checking. Any attempt to bypass our plagiarism detection systems will be treated as a severe Assessment Offence.

Coursework Submission Requirements

- ***An electronic copy of your work for this coursework must be fully uploaded on the Deadline Date of***
  6th December 2024 (23:30)
  using the link on the coursework Moodle page for STAT1043.
- ***For this coursework you must submit a single PDF document. In general, any text in the document must not be an image (i.e. must not be scanned) and would normally be generated from other documents (e.g. MS Office using "Save As .. PDF"). An exception to this is handwritten mathematical notation, but when scanning do ensure the file size is not excessive.***
- ***There are limits on the file size (see the relevant course Moodle page).***
- ***Make sure that any files you upload are virus-free and not protected by a password or corrupted otherwise they will be treated as null submissions.***
- ***Your work will not be printed in colour. Please ensure that any pages with colour are acceptable when printed in Black and White.***
- ***You must NOT submit a paper copy of this coursework.***
- ***All coursework must be submitted as above. Under no circumstances can they be accepted by academic staff***

The University website has details of the current Coursework Regulations, including details of penalties for late submission, procedures for Extenuating Circumstances, and penalties for Assessment Offences. See http://www2.gre.ac.uk/current-students/regs

---

- *Detailed Specification*
  Coursework must be completed individually. Students are allowed to work in groups in order to produce results and discuss implications of their findings
- *Deliverables*
  Single PDF document. No Excel file or R files (e.g. .R) should be uploaded.
- *Grading Criteria*

• Marks will be distributed as follows:	
0-30	An unsatisfactory submission that demonstrates little understanding of the subject matter.
31-39	A substandard submission which has no critical appraisal of the numerical methods and concepts of scientific methodology.
40-49	A satisfactory submission with a basic understanding of the models used and basic descriptions with no proofs
50-59	A submission which includes a reasonable understanding of the underlying models, has some proofs and basic discussion.

60-69	A good submission which includes a good understanding of the numerical methods and concepts of scientific algorithms. Elaboration and discussion of the underlying theory, implications of tests, results and underlying models.
70+	A very good submission that clearly demonstrates the relative merits of the underlying models and implications of their use. Detailed discussion of the results, underlying theory and relating the dynamics of the underlying time series to those implied by the models. Equations and proofs are given where necessary with critical derivation and discussion.
80+	An excellent submission which demonstrates a clear understanding of the underlying models and implications of their use. Clear derivation and presentation of all necessary formulas used and implications of their dynamics relevant to the dynamics of the time series under investigation. Excellent discussion of the underlying theory of discrete time stochastic processes, methodology and theory of financial markets.

STAT-1043 Statistical Techniques and Time Series

Deadline: Please upload your report before 23:30.
6th December 2024

Submission: online via Moodle.

Deliverables – **PDF Processed Analysis Report** incorporating **R outputs** where necessary, such as tables, plots, etc.
In detail:

- Do not upload the R file
- Do not upload direct copies of **R** or code chunks – convert to table/ plot
- **No handwritten** notes should be attached on the report.
- You should use Microsoft Equation editor (or the built-in equation editor).
- You will need to find the best way of displaying your results (Tables, Plots) using the PAGE restriction in place
- No scanned/image output (will end up in subtracting marks)
- Your final report to be uploaded must be converted into a PDF

What you are expected of: Show all underlying theory behind every computation you make, derive all equations, and explain in detail all formulas, background theory and implications of each theoretical assumption.

Restrictions:
A maximum of 4 pages (and no more will be accepted)

Topic:
Estimation and Forecasting tasks.

Data:
Yahoo finance using R package `quantmod` to get your data or R package `fpp` to get other non-financial time series data.

Report:
Use headings to identify each section in your report. Necessary graphs and tables must be copied from your spreadsheet into the Report (always use "Paste as Picture" to avoid linking). Do NOT copy any time series list into your report e.g. dates and prices. State the mathematical definitions of any quantities that you describe.

R coding:
You can find some nice examples on the links below:

You will need the following libraries installed:

```
library(stats)
```

```
library(quantmod)
library(fpp)
library(forecast)
```

Excel Spreadsheet:
No Excel spreadsheet should be submitted. Your report should contain all necessary information for a complete understanding of your analysis.

Mark Weighting:
See within.

Assessment: You must provide evidence of understanding. This implies that you are required to discuss the relevant theory behind any of your calculations and implementation. To achieve full marks, you will need to devise the appropriate amount of discussion behind the theory and formulas used and methodology as applied to the relevant time series data. ALL EQUATIONS SHOULD BE GIVEN AND DISCUSSION AT AN APPROPRIATE LEVEL MUST BE GIVEN.

Coursework tasks

Download data of any stock or financial index or any other economic time series data you prefer (e.g. GDP of an emerging economy or inflation data) using `library(quantmod)` or `library(fpp)`. Any other non-economic/financial time series data is acceptable.

For each of your answers below you will need to:

- Show the 1-liner code responsible for the output
- Write the respective mathematical equation e.g., for taking diff and/or for the ARMA model fit
- Explain any hypothesis testing: (i) What is the underlying hypothesis test null and alternative hypothesis, (ii) comment on the p-values
- Display plots, with caption and explanation
- No screenshots of results are allowed

[Total: 100 marks]

On the series:

(i) Obtain the time series using the correct code syntax.

[5 marks]

(ii) Based on what we discussed in the lectures, choose any method you like to de-trend the series and plot the de-trended series y_t. By visualizing the detrended plot assess whether the series appears to be stationary and if there is seasonality in your series.

[10 marks]

(iii) Create the ACF and PACF functions of y_t in R and explain how you can deduce if the time series you chose is a unit root process or not.

[10 marks]

(iv) Using a statistical test from the lectures check if y_t is serially auto-correlated, describe the test-hypothesis and report/comment on the result. Compare your result against ACF and PACF from the question above.

[10 marks]

(v) Check y_t for stationarity using the correct test statistic and comment on the output. Describe whether it is consistent with your conclusions from the ACF and PACF plots.

[10 marks]

(vi) Perform a normality test of your choice on y_t and report the output. Explain what the result means for your analysis.

[10 marks]

(vii) Fit an ARMA model and determine the best lag order; show the 1-liner codes for output. Are the ACF/PACF plots you created above good indicators of the best lag order selected by *auto.arima?*

[15 marks]

(viii) Report the coefficients for the chosen ARMA model, show the respective equation given these coefficients and show how you will forecast 1 step ahead.

[15 marks]

(ix) The residuals from an ARMA fit require that:

a. The residuals have zero mean $E[e_t] = 0$
b. Have a finite and constant variance $Var[e_t] = \sigma^2$
c. Have zero autocovariance $E[e_t e_\tau] = 0$

Based on what we covered in the lectures, do the residuals from your fitted model satisfy the three conditions above.

[15 marks]