Assignment 2 - STAT603 S2 2024

Enter your name - DO NOT REMOVE THE QUOTES

2024-09-21

The purpose of this assignment is to assess your analytical and computing skills on the material covered.

Total Possible Marks: 100 marks, which contribute 25% towards your final grade in this paper.

Deadline: 23:59 October 20 (Sunday), 2024

Submission: The assignment must be submitted as a soft copy in a single .pdf file on Canvas. Your filename must include 1) your last name, 2) your first name, and 3) your student id, e.g., if John White submits his assignment, his .pdf file must be named "White_John_123456789".

Report/Assignment: Your assignment must be self-contained, i.e., you need to embed your `R` code in your answers. See example in the box below:

Page Limit: Maximum number of pages is 20 including graphs and `R` code.

Data for Questions: automobiles - QUESTIONS 1-3, you will use the file `automobile.csv`.

- QUESTION 4, you will use the global_economy and aus_livestock datasets (from the fpp3 R package).
- Question 5: NO data required.

Note: All data should be converted into time series using `tsibble` functions in R.

R: All computing tasks must be done using R (you can use RStudio as your editor).

Plagiarism: If this is the case for your assignment, your case will be referred to an appropriate university's office. Exceptional Circumstances: If your performance and/or your ability to complete this assignment by the due date is seriously affected by exceptional circumstances beyond your control (e.g., injury or illness) you may apply for special consideration (SCA), WITH supporting evidence. To apply for special consideration, you must complete the special consideration form via Canvas (STAT603 Home Page).

Lateness penalties: Late assignments without and approved extension (or SCA) will be subject to a deduction of one grade (e.g., from C+ to C) out of your total mark for each 24-hr period, or part thereof, for up to a maximum of 3 DAYS. This means that the last day to apply for an SCA is 23rd October, 2024. Assignments over three days late (without an SCA) will not be marked and you will receive an DNC (Did Not Complete) for this assessment.

Question 1. ETS {Total: 20 marks}

HINT: Read Sections 8.1 - 8.7 of the online book.

First, copy the `automobile.csv` file from CANVAS and read it into `R`.

(a) Plot the series and discuss the main features, including stationarity {2 marks}.

(b) Forecast the next two years using:

(i) Simple exponential smoothing,
(ii) Holt's linear trend, and
(iii) Holt's damped trend.

Plot the series and the forecasts. Based solely on the time-series plots, discuss the adequacy of each method to forecast from this series. **Explain your answer** {2 marks - forecasts + 3 marks **discussion**; TOTAL = 5 marks}.

(c) Repeat Q1(b), but using Holt-Winters' seasonal methods. Discuss whether additive or multiplicative seasonality is necessary. Explain your answer {2 marks - forecasts + 3 marks **discussion**; TOTAL = 5 marks}.

(d) Compare the mean squared error (MSE) and the mean absolute error (MAE) of the one-step-ahead, four-steps-ahead and six-steps-ahead forecasts from methods discussed in (b)-(c) above. Report your results neatly and clearly (suggestion: use a Table).

Which method has the highest accuracy? Does this selection depend on the number of pre-specified (steps-ahead) forecasts? Explain your answer {2 marks - results clearly presented + 3 marks **discussion**; TOTAL = 5 marks}.

(e) Briefly discuss the potential mistake/error we may unintentionally introduce in the discussion when comparing models (b) - (c) using the MSE and MAE {3 marks}.

Question 2. Stationarity {Total: 20 marks}

HINT: Read Sections 9.1, 9.2,and 9.5 of the online book.

Note: Question 2 continues to use the automobile data.

(a) Plot the autocorrelation function (ACF) and the partial ACF.

(i) Briefly discuss the stationarity of the series based on the ACF.

- Does your answer here conform with your answer in Question 1(a)?

(ii) Should the series be differenced in order to obtain a stationary series?

- Explain your answer. {3 marks ACF/PACF + 5 marks discussion; TOTAL = 8 marks}

(b) Find an appropriate Box-Cox transformation and order of differencing to obtain stationary data.

- Note: Justify your choices (even if no Box-Cox transformation is required) {2 marks - working + 5 marks **discussion/rationale** (a) + 5 marks **discussion/rationale** (b) ; TOTAL = 12 marks}.

Question 3. ARIMA modelling {Total: 30 marks}

HINT: Read Sections 9.1, 9.5 and 9.7 of the online book.

Question 3 continues to use the automobile data.

(a) By studying the appropriate graphs of the series in R, propose an appropriate ARIMA(p, d, q) or ARIMA(p, d, q)(P, D, Q) structure to model the series. Justify your answer. Plots/Figures can be included as part of your answer {3 marks working + 5 marks rationale; TOTAL = 8 marks}.

(b) Should a constant be included in the model? Justify your answer {2 marks}. *HINT: Read sub-section 'Understanding constants in R', from Section 9.7. Note that, to fit ARIMA models with and without a constant (respectively), we use the syntax:*

```
fit1 <- DATA %>% model('m1' = ARIMA( Y ~ 1....)
fit1 <- DATA %>% model('m1' = ARIMA( Y ~ 0....)
```

(c) Fit the ARIMA model proposed in Q3(a) using R functions and examine the residuals. Is the proposed model satisfactory? **Justify your answer** {3 marks - working (code) + 5 marks - **explanation**; TOTAL = 8 marks}.

(d) Let `ARIMA()` choose an ARIMA model for this data. Does `ARIMA()` return the same model as the one you chose in 3(a)? If not, which model do you think suits best? (Explain your answer) (3 marks - working and code + 5 marks - rationale; TOTAL = 8 marks}.

(e) Which method do you think is best between ETS and ARIMA to forecast from this series (compare **Q1** and **Q3** results)? Explain your answer {4 marks - reasons/rationale}.

Question 4. Seasonality and the `accuracy()` function {Total: 10 marks}

(a) Fit a Holt's Linear (with no damping parameter), a Holt-Winters additive and a Holt-Winters multiplicative model to the New Zealand consumer price index (CPI) from the data set `global_economy`. Then, compare their in-sample accuracy with the function `accuracy()`.

The output will return a `NaN` for the Holt-Winters models. Why is this happening? Write down a short paragraph (2 - 3 sentences) discussing this question {2 marks code + 3 marks discussion; TOTAL = 5 marks}.

Run the following code to extract the data (you will need to write code to estimate the models, and to check the in-sample accuracy measure with the function `accuracy()`):

```
library("fpp3")
mydata <- global_economy |>
  filter(Country == "New Zealand")
```

(b) Now, repeat this analysis with the number of pigs slaughtered in Victoria, available in the dataset `aus_livestock`. Did you observe any warnings (or `NaN`) from Holt-Winters? If not - Why [as opposed to the output in Q5(a)]? **Briefly explain your answer - compare to 4(a).**

Code to extract and plot the data.

```r
myseries <- aus_livestock |>
  filter(Animal == "Pigs", State == "Victoria")
myseries |> autoplot(Count)
```

You will need to write code to fit the required model and to check the in{sample accuracy measure with the function `accuracy()`. {2 marks code + 3 marks discussion; TOTAL = 5 marks}.

Question 5. Forecasting Theory {Total: 20 marks; 5 marks each}

(a) In general, prediction intervals from the ARIMA models increase as the forecast horizon increases. *HINT: Read Section 9.8 of the online book.*

- **TRUE/FALSE Explain your answer**

(b) The AICc cannot be used to compare between ARIMA and ETS models. *HINT: Read Section 9.10 of the online book; especially the section comparing ETS() and ARIMA() on non-seasonal data.*

- **TRUE/FALSE Explain your answer**

(c) Time series cross-validation can be used to compare between ARIMA and ETs models. *HINT: Read Section 9.8 of the online book.*

- **TRUE/FALSE Explain your answer**

(d) Read Section 9.10, subsection Comparing ETS() and ARIMA() on seasonal data. In this section, after a deep analysis, the ETS model was selected for forecasting based on its forecasting performance from the test set.

- **TRUE/FALSE Explain your answer**

— **END OF ASSIGNMENT 2 (STAT2603 2024) —