

# 程序代写代做 CS编程辅导

## UNSW ECON2209 Assessment



Project

2024

At the start of an R session, remember to type `library(fpp3)` in the R Studio Console. This will then load (most of) the R packages you will need, including some data sets.

### Details:

- Total value: 25 marks.
- Submission is due on **Friday of Week 9 (12 April) by 4pm.**
- A submission link is on the Moodle site under Assessments.
- Submit your answer document in PDF format. Your file name should follow this naming convention:  
**CP\_your first name\_zID\_your last name\_ECON2209.pdf**  
For example: CP\_John\_z1234567\_Smith\_ECON2209.pdf
- You get one opportunity to submit your file. Make sure that you submit the file that you intend to submit.
- Your submitted answers should include the R code that you used.
- **Format:** No longer than 20 pages, including code, figures, tables and any appendices. Do not include a separate title page. At least 11 point font should be used, with adequate margins for comments. Any extra pages will not be marked.
- **Use of AI tools such as ChatGPT are prohibited. In cases where use is detected, it will result in 0 marks for “Interpretation of the results, arguments used and conclusions drawn”. It may also result a referral to the Academic Integrity Committee.**
- Use the methods and R packages taught in this course. **Failure to do so will result in a mark of 0 for Suitability of Methods.**
- There are videos in the section “Support Videos” on the course Moodle site that you should watch before preparing your answers:
  - How to Answer Questions in this Course
  - How to Export figures that are readable
- This project requires you to analyse time series data. The series will differ between students.
- Unless approval for an extension is given on medical grounds (supported by a medical certificate submitted through the Special Consideration process) there will be an immediate late penalty of 5% from 4:01pm on 12 April, followed by additional penalties of 5% per calendar day or part thereof. Submissions will not be accepted after 5 days (120 hours) of the original deadline.

**Marking for this Project:** Marks are awarded by overall achievement against the following criteria:

(a) Suitability of methods. **10 marks**

- 0 marks: Little or no attempt, **or use of methods and R packages not taught in this course.**
- 2 marks: Inappropriate methods used or methods inappropriately implemented.
- 5 mark: An attempt to answer the question using methods that are appropriate and appropriately implemented.
- 7 marks: A reasonable answer to questions that generally follow the provided solutions.
- 8.5 marks: Systematic and sensible discussion of all results.
- 10 marks: More detailed discussion than asked for.



(b) Interpretation of the results, arguments used and conclusions drawn. **10 marks**

- 0 marks: Little or no attempt, **or use of AI detected.**
- 2 marks: Little attempt to discuss the results or a poor understanding of the results found.
- 5 marks: An attempt has been made to understand and explain all the results.
- 7 marks: Systematic and sensible discussion of all results.
- 8.5 marks: Discussion of the results seem correct and insightful.
- 10 marks: Insightful discussion beyond what might reasonably be expected, possibly drawing on external references and other research.

(c) Presentation: Appropriate style of graphs, tables, reporting and clarity of writing. **5 marks**

- 0 marks: Little or no attempt.
- 1 marks: Difficult to follow what has been done. Small font making graphs and tables hard to read. Lack of clear writing.
- 2 marks: Presentation of results falls short of the standard in the provided solutions for tutorial exercises and problem sets.
- 3 marks: Presentation of results consistent with the standard in the provided solutions for tutorial exercises and problem sets.
- 4 marks: More polished presentation.
- 5 marks: Professional style report. Tables can still be in R output format - reformatting not required.

**Maximum marks: 25**

Note that criteria (b) and (c) together comprise 60% of the overall mark for the project.

## Select the data series that you will analyse

In this project you will use data from the Australian Bureau of Statistics (ABS). Specifically, you will use data on components of the [Consumer Price Index](#): ABS Catalogue 6401.0, Table 9. CPI: Group, Sub-group and Expenditure Class, Index Numbers by Capital City.

The data series you will use is in the form of a price index. CPI indexes are currently based in financial year 2011. The level of the quarterly values average to 100 for this financial year (i.e. the index values for quarters 2011 Q3 to 2012 Q2 equal 100 for each series).

We can download the data from the ABS website, or we can use the R package `readabs` to read it.

```
library(readabs)
cpidata_full <- read_abs("6401.0", tables = "9", check_local=FALSE) %>%
  mutate(Quarter = yearquarter(date)) %>%
  as_tsibble(
    index = Quarter,
    key = c(series_id)
  )
```

We will drop several data series in the full data set that are either not very interesting or very tricky to forecast. To do this, we will use the package `tidyverse`. You will need to install this package if you do not already have it installed. Once you have installed `tidyverse`, use the following commands:

```
library(tidyverse)
cpidata <- cpidata_full %>%
  filter(!str_detect(`series`, "All groups")) %>%
  filter(!str_detect(`series`, "Furn")) %>%
  filter(!str_detect(`series`, "Insurance")) %>%
  filter(!str_detect(`series`, "Financial services")) %>%
  filter(!str_detect(`series`, "Deposit")) %>%
  filter(!str_detect(`series`, "Health"))
```

You must use the following method for selecting your data series.

Use the seven digits of your UNSW student ID to get the data series that you will analyse in this project, as in the following example for the case when your student ID is z1234567:

```
set.seed(1234567)
myseries <- cpidata %>%
  filter(`series_id` == sample(`series_id`, 1)) %>%
  filter(!is.na(value))
```

Note while `sample()` takes a random sample, using the same “seed” through `set.seed()` will result in the same series being selected each time you run the code on the same computer.

Make a note of the ID of your series, in case you run into computer problems and need to retrieve the series manually:

```
list(myseries$series_id[1])
```

Note that different data series can have different lengths. These are the official data, so these are the data you will use, even if your series has a different length from those of your classmates.

## Tasks

First, plot your data using the following code, without changing anything.

```
myylab <- substr(myseries$series[1], 1, 6)
myxlab <- "Quar
mytitle <- past
substr(myseries$series[1], 1, 6) + substr(myseries$series[1], 7, 10)

myseries %>%
  autoplot(varying = "year", theme = theme_minimal()) +
  labs(y = myylab, x = myxlab, title = mytitle)
```

The `substr()` commands take parts of the series description for use as the y-axis label and the figure title. Note that you can use `myylab`, `myxlab` and `mytitle` wherever relevant in other figures in this Project.

- Based on just this plot, discuss characteristics of the series.
- Decide if a transformation of your data is required. Explain your decision. If a transformation is needed, then use it throughout the rest of this Project.
- Create a training dataset (denoted as `myseries_tr`) consisting of observations before 2010. Visually check that the data were split appropriately by plotting the training and test data sets in the same figure.
- Fit an ETS model to your training data using the default ETS() command. Describe the model chosen and comment on the residuals, using the standard plots (i.e. `gg_tsresiduals()`) and a Ljung-Box test.
- Produce forecasts for the test data, and plot these along with the data series from 2000. Include and comment on the prediction intervals.
- Compare and comment on the accuracy of the model on the training data relative to the accuracy on the test data.
- In preparation for ARIMA modelling, use the visual inspection of plots to find the appropriate order of differencing needed for stationary data. Then use statistical tests to check your choices.
- Select an appropriate ARIMA model. Explain your choice. Comment on the residuals, using the standard plots (i.e. `gg_tsresiduals()`) and a Ljung-Box test.
- Using the training data set as before, try an STL decomposition followed by ARIMA on the seasonally adjusted data; that is, an STL-ARIMA model. Using the test data set, compare the accuracy of the forecast performance with the ETS model you obtained earlier. Plot forecasts from both models on the same figure, along with the actual data from 2000 onwards. Include and comment on the prediction intervals.

Be sure to label all your figures.