

### **DATA2001 Assignment 3 (Weight: 25%)**

**Due: 25 October 2024 @ 3:00pm**

The aim of this assignment is to gain practical experience in analysing timeseries data and using it for forecasting. You must complete this task in Python using a Jupyter notebook. You will need to submit a single Jupyter notebook (.ipynb file) via Blackboard.

#### **Context:**

Unemployment is one of the critical economic indicators that reflects the health of an economy and the well-being of its labour force. It is commonly used by the policymakers and economists to assess labour market conditions and guide decisions related to fiscal and monetary policies.

In this assignment, the goal is to utilise Queensland Labour Force data to explore historical trends and build a robust forecasting model to predict the future unemployment rates. This analysis will provide insights into the evolution of unemployment in Queensland and its potential trajectory over the coming months, offering valuable information to stakeholders, including government officials and economists.

**Datasets:** In this assignment, you will examine Queensland labour force data and relevant side data on inflation and interest rates, to make predictions of its future movements using timeseries forecasting techniques. Specifically, you will use three series to build and evaluate your forecasters:

**1. Unemployment data**, collected by the Australian Bureau of Statistics (ABS), and is available at:

<https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release#data-downloads>, Table 6. *Labour force status by Sex, Queensland - Trend, Seasonally adjusted and Original*.

The file you download comprises several series. For your assignment, the series of interest is **Series ID A84423620T**, titled "Unemployment rate ; Persons ; Percent **Original**".

**2. Inflation data**, also collected by the Australian Bureau of Statistics (ABS), available at:

<https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/consumer-price-index-australia/jun-quarter-2024#data-downloads>, TABLES 1 and 2. CPI: All Groups, Index Numbers and Percentage Changes

Again, the file you download comprises several series, and here you should use **Series ID A2325846C**, titled "Index Numbers ; All groups CPI ; **Australia** ; Index Numbers **Original**".

**3. Interest rates**, published by the Reserve Bank of Australia (RBA):

<https://www.rba.gov.au/statistics/tables/xls/f01d.xlsx>

Specifically, use the **Series ID FIRMMCRTD**, titled "Cash Rate Target on date, Daily, Original".

### Task description:

The submitted notebook should address the following 6 tasks (see marking grid for mark allocation):

1. Data Preparation:
  - a. Compile all the data for the period 2011-2023 into one series using the “pandas” library and set up an index for the entire dataset in an appropriate way for timeseries analysis.
  - b. The RBA interest rate data has a different resolution to the unemployment and inflation data. Explain and briefly justify the adjust to the resolution of the RBA data that you use.
2. Exploratory Data Analysis (EDA): Visualise the data timeseries, and comment on the patterns you can observe with respect to the features discussed in the lectures. Include visualisations appropriate for uncertainty and correlation where appropriate.
3. STR decomposition: Focus now on the unemployment timeseries.
  - a. Split the data into **training** and **testing** series, selecting the testing series to be the last three years of the data (2021-2023).
  - b. Manually step through the STR decomposition process on the training data, as described in the course material. Visualise and interpret each of the components of the STR decomposition for unemployment. (*Hint: You may wish to validate the output of your manual process against an automated modelling approach.*)
4. Timeseries models:
  - a. Fit an ARIMA model for the trend-cycle component of your STR decomposition of the training data and interpret the estimated model parameters.
  - b. Using the STR components that you estimated in tasks 3b and 4a, produce forecasts of Queensland unemployment for the test data series. Include the uncertainty in the forecasts and visualise the predictions.
5. Pure forecasters: now consider your choice of neural network-based methods:
  - a. Select an appropriate pure forecasting method to predict the trend component of the unemployment training data.
  - b. Using the seasonal component that you estimated in task 3b and the pure forecaster from 5a, produce forecasts of Queensland unemployment for the test data series. Include the uncertainty in the forecasts and visualise the predictions.
6. Using side data: how might different interest rate projections affect the forecasts?
  - a. Specifically, generate a series of values corresponding to a hypothetical reduction of interest rates by 0.5% for the next six months beyond the end of the test data, and a second series of values corresponding to interest rates remaining at their current level for the same duration.
  - b. Use each projection of the side information to generate two forecasts of the unemployment timeseries for six months beyond the end of the test data. Use any model that you wish, e.g. ARIMAX or neural network. Visualise the two forecasts, including uncertainties in them, and discuss the difference between the two timeseries predictions.