Queen Mary University of London

Main Examination period 2023 – May/June – Semester B

MTH6139 / MTH6139P: Time Series

Duration: 2 hours

The exam is intended to be completed within **2 hours**. However, you will have a period of **4 hours** to complete the exam and submit your solutions.

For actuarial students only: This module also counts towards IFoA exemptions. For your submission to be eligible, **you must submit within the first 3 hours.**

You should attempt ALL questions. Marks available are shown next to the questions.

All work should be **handwritten** and should **include your student number**. Only one attempt is allowed – **once you have submitted your work, it is final**.

In completing this assessment:

- You may use books and notes.
- You may use calculators and computers, but you must show your working for any calculations you do.
- You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.
- You must not seek or obtain help from anyone else.

When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;

Examiners: J. Griffin, D. S. Coad

Question 1 [21 marks].

(a) Let $\{X_t\}$ be a weakly stationary time series. The $(2q+1)$-point moving average of this series is given by

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j}.$$

Show that W_t is weakly stationary. **[8]**

The observed values of $\{X_t\}$ are $(3, 4, 2, B+2, C+1, 7, 9)$ for $t = 1, \ldots, 7$, where B and C are the second-to-last and last digits of your ID number, respectively.

(b) Given filter coefficients

$$a_0 = \frac{2}{5}, a_{-1} = a_1 = \frac{2}{5}, a_{-2} = a_2 = -\frac{1}{10},$$

calculate the output of the two-sided linear filter applied to $\{X_t\}$ at the value $t = 4$. **[3]**

(c) Find the value of each of the following quantities for $t = 5$:

$$\nabla^2 X_t;\ \nabla_2 X_t;\ \text{and}\ \nabla_2 \nabla^2 X_t.$$ **[6]**

(d) If $\{X_t\}$ is assumed to have a trend m_t, calculate the estimated value of the trend $\widehat{m}_t$ at $t = 5$ using exponential smoothing with smoothing parameter $\alpha = 0.6$. **[4]**

Question 2 [17 marks]. An MA(1) process to model $\{X_t\}$ is of the form

$$X_t = Z_t + \theta Z_{t-1},$$

where $\{Z_t\}$ is a white noise process $WN(0, \sigma^2)$. Consider the two sets of parameters $\theta = 3, \sigma = 2$ and $\theta = 1/3, \sigma = 6$.

(a) Show that they have the same autocovariance function for all lag values. **[4]**

(b) Explain how you can choose between these processes, including what the criterion means for relating $\{X_t\}$ and $\{Z_t\}$. State which set of parameters you would choose. Calculate the autocorrelation function for this set of parameters. **[6]**

(c) The time series denoted by $\{X_t\}$, $t = 1, \ldots, n$, with $n = 50 + C$, where C is the last digit of your ID number, follows the model chosen in part (b). Calculate the value of

$$Var\left(\bar{X}\right),$$

where $\bar{X}$ is the sample mean. **[5]**

(d) Suppose that the data in the preceding model are normally distributed, and we would like to test whether the population mean is zero. If the data are stored in `x` in R, why would the following command generally give incorrect results?

```
t.test(x)
```

[2]

Question 3 [23 marks]. Consider the AR(2) process

$$X_t = 0.7\,X_{t-1} - 0.1\,X_{t-2} + Z_t,$$

where $\{Z_t\}$ is a white noise process with mean 0 and variance 4.

(a) Is this model causal? Explain your answer. **[4]**

(b) Its autocorrelation function obeys the linear difference equation

$$\rho(h) - 0.7\,\rho(h-1) + 0.1\,\rho(h-2) = 0,\ h \geq 1.$$

Using this equation, find the values of $\rho(0)$ and $\rho(1)$. **[3]**

(c) The general solution of the linear difference equation in (b) is of the form

$$\rho(h) = k_1 r_1^{-h} + k_2 r_2^{-h},\ h \geq 1,$$

where r_1 and r_2 are the roots of the characteristic polynomial from the AR(2) process above, and k_1 and k_2 are constants. Find $\rho(h)$ by computing the values for these quantities, with initial condition $k_1 + k_2 = 1$. **[8]**

(d) Find the partial autocorrelation function for all lag values $h \geq 1$ for this model. **[3]**

(e) Suppose that we have data known to follow the preceding model, $(X_1, \ldots, X_5) = (6, -3, -1, A, B)$, where A and B are the third-to-last and second-to-last digits of your ID number, respectively. Calculate the best linear predictor forecasts for the values of X_6 and X_7. **[5]**

Question 4 [12 marks]. For each of the following processes, describe the expected behaviour of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots:

(a) White noise; **[3]**

(b) MA(3); **[3]**

(c) AR(2); **[3]**

(d) ARMA(3,1). **[3]**

Question 5 [27 marks]. The time series $\{X_t\}, t = 1, \ldots, 110$, represents the annual mean temperature at a certain location. The data are stored as `mtemp` in R. We want to fit an ARIMA model to the data.

(a) Before fitting a model, the following command is run:

```
plot(diff(mtemp))
```

Explain what the purpose of this command would be, and what you would look for in the resulting graph. **[3]**

(b) The following commands are then run:

```
model1 = arima(mtemp, order=c(2,1,1))
model1
```

Part of the output is:

```
Coefficients:
         ar1      ar2      ma1
      0.1951  -0.1613  -0.5505
s.e.  0.1896   0.1104   0.1802
sigma^2 estimated as 0.01369
```

Write down the fitted model in terms of $\{X_t\}$ and a white noise process $\{Z_t\}$, including the numerical values of the parameters. In order to calculate the likelihood, what assumption is made about the distribution of each Z_t? **[7]**

(c) Another model is fitted using this command:

```
model2 = arima(mtemp, order=c(2,1,0))
```

Suppose that the log-likelihoods for `model1` and `model2` are $100 + 10 \times B + 5.5$ and $100 + 10 \times C + 7.4$, respectively. Here, B and C are the second-to-last and last digits of your ID number, respectively. Calculate the AIC and BIC for both models. Based on these, which model is preferred? **[8]**

(d) The command below is run:

```
Box.test(model1$residuals, lag=20, type="Ljung-Box", fitdf=3)
```

with output as follows:

```
X-squared = 22.497, df = 17, p-value = 0.16635
```

What is the purpose of this code, and what does the output tell us? **[4]**

(e) If the data were collected monthly instead of annually, what change to the model structure of part (b) would be likely to be needed. Write out the modified model equation (with symbols, not specific numerical parameter values). **[5]**

End of Paper.