# S8 Time Series Analysis
# Coursework 1

Handout: Friday 2 November 2018.
Deadline: Hand in at the Maths Undergraduate Office before Thursday 15 November 2018, 4pm.

NB: Stationarity by itself always means 'second-order' stationarity. $\{\epsilon_t\}$ denotes white noise with mean zero and variance $\sigma_\epsilon^2$. Assume $\{\epsilon_t\}$ is Gaussian/normal here. $\Delta$ represents the difference operator $(1-B)$, where $B$ is the back-shift operator.

You may use any results from the notes you wish, but you must properly cite them.

You may hand write your solutions should you wish, however, figures must obviously be produced computationally. Please attach the code you used to your solutions.

## Question 1 (6 marks)

This question will be concerned with the time series model

$$X_t = \mu_t + Y_t \tag{1}$$

where $\mu_t$ is a polynomial trend of order $d-1$ $(d \geq 1)$, and $Y_t$ is a zero mean AR$(p)$ process of the form

$$Y_t = \sum_{j=1}^{p} \phi_{j,p} Y_{t-j} + \epsilon_t.$$

(i) Show the de-trended process $\{X_t^{(d)}\}$, where $X_t^{(d)} \equiv \Delta^d X_t$, is ARMA$(p,d)$. Determine whether $\{X_t^{(d)}\}$ is invertible.

We will now focus on the case where $\mu_t = \alpha + \beta t$ (i.e. it is a linear trend), and $Y_t$ is an AR$(1)$ process of the form

$$Y_t = \phi_{1,1} Y_{t-1} + \epsilon_t$$

where $|\phi_{1,1}| < 1$.

(ii) You have shown in part (i) that $\{X_t^{(2)}\}$ will be ARMA(1,2). Explain why $\{X_t^{(2)}\}$ is stationary, and write it in the form of a General Linear Process, providing expressions for the coefficients $\{g_k, k = 0, 1, ...\}$ in terms of $\phi_{1,1}$.

(iii) Find expressions for the first three terms of the autocovariance sequence ($s_0$, $s_1$ and $s_2$) for $\{X_t^{(2)}\}$. [HINT: All three can be written in terms of $\phi_{1,1}$ without the need for infinite sums.]

## Questions 2 (8 marks)

We will now look at simulating the zero mean stationary AR(1) process

$$Y_t = 0.5Y_{t-1} + \epsilon_t, \tag{2}$$

with $\sigma_\epsilon^2 = 1$.

In order to simulate $Y_1$ (and hence $Y_2$, $Y_3$, etc.), we need to know $Y_0$. It was shown in Examples Sheet 1 that if you set $Y_0 = 0$, the process that follows is non-stationary. In theory, the only boundary conditions that can be set are at $-\infty$ or $+\infty$. This means that for simulation purposes, if one sets the first value to be zero, we have to let the process run for a period of time before it *approximately* reaches the stationary state that we expect under the theory. This period of time is known as *burn-in*. Once we have had enough burn-in time that we are content the process is close to being stationary, we set that to be the first time point $Y_1$ and throw away all the data before that point. Note: the only way to achieve an exactly stationary state under this method is to have an infinitely long burn-in time. We don't have that long!

In this question we will study how quickly we converge towards the stationary process when increasing burn-in time.

To do this, we will look at how closely the *empirical* covariance matrix $\hat{\Sigma}$ for the vector $[Y_1, Y_2, Y_3]$ estimated from simulations matches the theoretical form

$$\Sigma = \begin{bmatrix} s_0 & s_1 & s_2 \\ s_1 & s_0 & s_1 \\ s_2 & s_1 & s_0 \end{bmatrix}$$

assumed under stationarity.

(i) For the AR(1) model in Equation (2), determine the matrix $\Sigma$.

The simulation procedure then proceeds as follows

- For a burn-in period of $N_b$ points, set the first value to be zero and simulate $N_b + 3$ elements using the formula in Equation (2), i.e. simulate the vector $[Y_{-N_b+1}, Y_{-N_b+2}, ..., Y_0, Y_1, Y_2, Y_3]$.
- Discard the first $N_b$ elements to give the vector of the first three elements of your process after the burn-in time, i.e. $\boldsymbol{Y} = [Y_1, Y_2, Y_3]$.

(ii) Using your preferred programming language from the one of MATLAB, R or Python (I strongly suggest MATLAB), undertake the following:

- For a burn-in period of $N_b = 1$, using the simulation procedure outlined above, simulate $M = 1 \times 10^6$ independent and identically distributed sample vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_M$, storing them in a single $M \times 3$ array.
- Compute the empirical covariance matrix $\hat{\Sigma}$ from $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_M$.
- Compute $\|\Sigma - \hat{\Sigma}\|_F$, where $\|\cdot\|_F$ is the Frobenius matrix norm (this will be a measure of the distance between the empirical covariance matrix computed from your simulations and the theoretical $\Sigma$.)
- Repeat this for all 20 values of $N_b$ in $N_b = 1, 2, ..., 20$.

Do not report all of the output from the above! Please just plot $\|\Sigma - \hat{\Sigma}\|_F$ as a function of $N_b$.

Try setting $N_b = 1000$. What is $\|\Sigma - \hat{\Sigma}\|_F$ in this instance. Comment on all your results.

(iii) Now try different starting values. Comment on your results.

## Question 3 (6 marks)

For a sequence of observations $X_1$, $X_2$, ..., $X_N$ from some zero mean stationary process $\{X_t\}$, we propose the autocovariance estimator

$$\hat{s}_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} X_t X_{t+\tau}. \tag{3}$$

In this question we will explore the properties of this estimator. To do so, we will work with the process

$$X_t = 1 + 0.02t + Y_t,$$

where $\{Y_t\}$ is the AR(1) process given in Equation (2).

(i) In your preferred programming language, write a function that performs the difference operator on a time series. Take care with the end points. The output time series will inevitably be shorter than the input time series. Your answer to this question will be the code you have written.

(ii) Undertake the following:

- Using a burn-in time that you are happy with, simulate a realization of length $N = 50$ (i.e. for times $t = 1, 2, ..., 50$) of the process $\{X_t\}$.
- Fully de-trend $\{X_t\}$ using the function you wrote in (i).
- Estimate $s_0$, $s_1$, and $s_2$ of the de-trended process using the estimator in (3).
- Repeat this for $M = 10000$ independent realizations of the de-trended process. Compute the empirical variance and bias of the estimators for $s_0$,$s_1$ and $s_2$. For the empirical bias, you will need to use your answer to Question 1.

Repeat this procedure for realizations of length $N = 100$ up to $N = 1000$ in steps of 50 (i.e. 100, 150, 200, ..., 1000). Present and comment on your results.