

Practical 3: Time Series and Forecasting

What you need to do

Before starting the Minitab practical, it would be a good idea (if you've not already done so) to bookmark the webpage containing all of the information for Semester 2 of ACE2013. The easiest way to do this is to go to the course webpage directly:

www.mas.ncl.ac.uk/~nlf8/teaching/ace2013

To add this page to your favourites, select Favourites → Add to Favourites from the menu bar at the top. You will now be able to get to this page easily by selecting Favourites and selecting the page from the list.

In this practical you will be using Minitab to perform time series analyses similar to those shown in lectures over the past two weeks. You should work through *all* the questions given below. Notice that there is no assessed work in this practical; however, Practical 4 will assume you have worked through this sheet completely and will have a substantial assessed element. **I would strongly advise that you have Section 2.6 of your lecture notes to hand while working through these questions.**

Log into a PC and open Minitab by clicking on Start → All programs → Statistical Software → Minitab 16 Statistical Software English. Now work through the following questions.

Click on the link for the Backpackers dataset, available to download from the course webpage. Once opened, you will see that there are data in the first four columns of this worksheet.

- Column 1 shows Y , the number of backpackers visiting the *Hot Rock Hostel* in Uluru, Australia
- Column 2 indicates the “season” – 1 being the first quarter (January–March), 2 being the second quarter (April–June), 3 being the 3rd quarter (July–September) and 4 being the last quarter (October–December)
- Column 3 indicates the year
- Column 4 is a time index from 1 to 50

Look at the worksheet, and then answer the following questions.

Description

1. Produce a time series plot for these data, making sure you use appropriate labels for the axes – particularly the time axis.
2. Use the space below to write down any interesting features you might notice in the plot.

Trend

3. Use Minitab to find the series of moving averages Y^* by clicking on **Stat–Time Series–Moving Average**. Make sure Minitab stores the resulting moving averages in a column of the worksheet, and re-label this column “MA”.
4. Now you should use Minitab to fit a simple linear regression model to the series of moving averages, again making sure the fitted values are stored in a column of the worksheet (re-label this column “Trend”). In the space below, write down the resulting regression equation:

$$\begin{aligned} Y^* &= \beta_0 + \beta_1 T + \epsilon, & \epsilon &\sim N(0, \sigma^2) \\ Y^* &= \end{aligned}$$

Is the slope term β_1 significant here?

5. Now re-produce the time series plot from question 1, but overlay the moving averages (in the column you should have labelled “MA”) and the trend line (the fitted values for this are in the column you should have labelled “Trend”). You can do this by selecting **Graph–Time Series Plot–Multiple**, and then entering the original dataset, the moving averages and the trend in the **Series** box.

Seasonal variation

6. You should now use Minitab to obtain the seasonal effects (use **Stat–Time Series–Decomposition**). Make a note of these below:

$$\begin{aligned} \hat{s}_{Q1} &= \\ \hat{s}_{Q2} &= \\ \hat{s}_{Q3} &= \\ \hat{s}_{Q4} &= \end{aligned}$$

As I showed you in class, make sure you get Minitab to store the seasonal effects in a column, and re-label that column “Season”.

Forecasting

7. You should now use the regression equation obtained in question 4, and the seasonal effects in question 6, to forecast the number of backpackers visiting the *Hot Rock Hostel* in

(a) October–December 2009;

(b) April–June 2010.

(c) What is wrong with your forecast in part (b) above?

Obtaining the residual series

8. You should now obtain the residual series for the backpackers dataset. click on **Editor** and select **Enable Commands**. Now go to the Session window (the white window above the worksheet) and at the **Minitab** command prompt (**MTB >**) type

```
MTB > LET C10 = C1-Trend-Season
```

Label column **C10** as “Residuals”.

9. Now examine the residual series by producing a time series plot for the values in column **C10**. Write down anything of interest in the space below.

Autocorrelation

10. You should now check the residual series for autocorrelation. Click on **Stat–Time Series–Lag**. Enter the residual series (which should be column **C10**) in the **Series** box and you should **Store lags** in column **C11**. Then click on **Stat–Basic Statistics–Correlation** and enter the residuals (column **C10**) and the lagged data (column **C11**) in the **Variables** box. In the space below write down the lag 1 autocorrelation coefficient and its p -value. Is there significant autocorrelation here?

$r_1 =$
 $p - \text{value} =$

11. Use Minitab to produce a *correlogram* to check for autocorrelation at other lags by clicking on Stat–Time Series–Autocorrelation. Enter the residuals dataset (column C10) in the Series box and click OK. Check that the first autocorrelation in the plot matches up with the calculated value for r_1 in question 10. Are there any higher–order dependencies beyond lag 1?

Fitting an AR(1) model

12. You should now fit an AR(1) model to the series of residuals by clicking on Stat–Time Series–ARIMA; enter the residual series (column C10) in Series and enter 1 in the Autoregressive (Nonseasonal) box and 0 in the other two boxes. Tick the box that says Include constant term in model and then click OK. Write down the estimated model in the space below:

$$Y_t = \mu + \alpha(Y_{t-1} - \mu)$$
$$Y_t =$$

Is the autoregressive parameter α significant in the model? Explain.

13. If you scroll down to the bottom of your worksheet you should find that the last residual (the last value in column C10) has the value -2.134 .
- (a) Use this value to Forecast the next residual at time–point 51, i.e. use the fitted model above to find Y_{51} given that $Y_{50} = -2.134$;
 - (b) Use the linear regression equation you wrote down for the trend (question 4) to obtain the fitted trend value for time–point 51.
 - (c) Now add the trend from part (b) above, as well as the seasonal effect for the 3rd quarter obtained in question 6, to the forecasted residual in part (a) to obtain a full forecast for the number of backpackers the *Hot Rock Hostel* can expect to see in July–September 2009.