

Introduction

Data are frequently recorded at regular time intervals, e.g.

- daily stock market indices
- monthly rate of inflation
- annual profit figures

In this Chapter we think about how to **display** and **model** such data.

We will consider how to detect **trends** and **seasonal effects** and then use these to make **forecasts**.

As well as review the methods covered in MAS1403, we will also consider a class of time series **models** known as **autoregressive moving average** models.

Introduction

Why is this topic useful?

Making forecasts allows organisations to make better decisions and to plan more efficiently. For instance, reliable forecasts

- enable a retail outlet to anticipate demand
- allow hospitals to plan staffing levels
- allow manufacturers to keep appropriate levels of inventory

Time series and time series plots

A **time series** is a collection of observations made sequentially in time.

When observations are made continuously, the time series is said to be **continuous**.

When observations are taken only at specific time points, the time series is said to be **discrete**.

In this course we consider only discrete time series, where the observations are taken at equal intervals.

Time series and time series plots

The first step in the analysis of time series is usually to plot the data against time, in a **time series plot**.

Suppose we have the following four-monthly sales figures for *Turner's Hangover Cure* as described in Practical 2 (in thousands of pounds):

	Jan–Apr	May–Aug	Sep–Dec
2006	8	10	13
2007	10	11	14
2008	10	11	15
2009	11	13	16

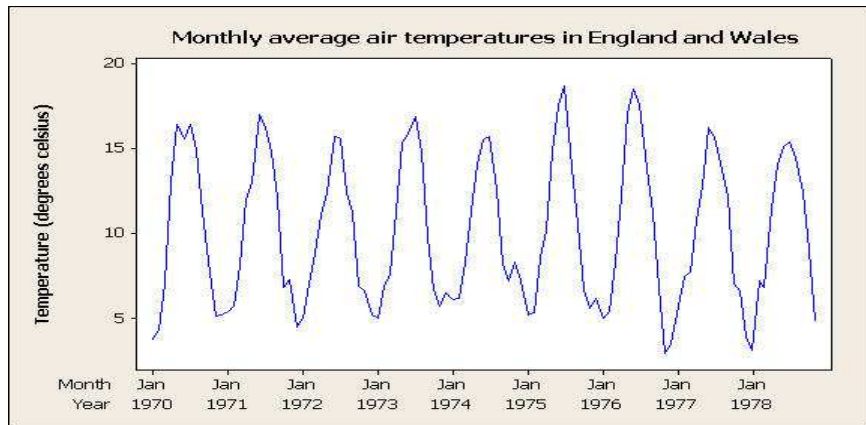
Minitab

Describing time series

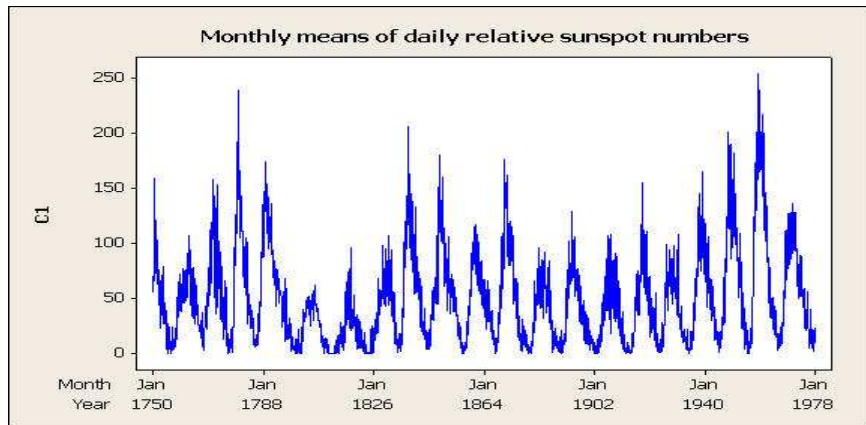
So what can we say about the sales figures for *Turner's Hangover Cure*?

- Obvious **positive trend** – the mean level of sales is increasing through time
- **Cyclic variation**, or **Seasonality** – with sales being highest in September–December (Fresher's week/Christmas parties?!) and lowest in January–April
- No obvious outliers

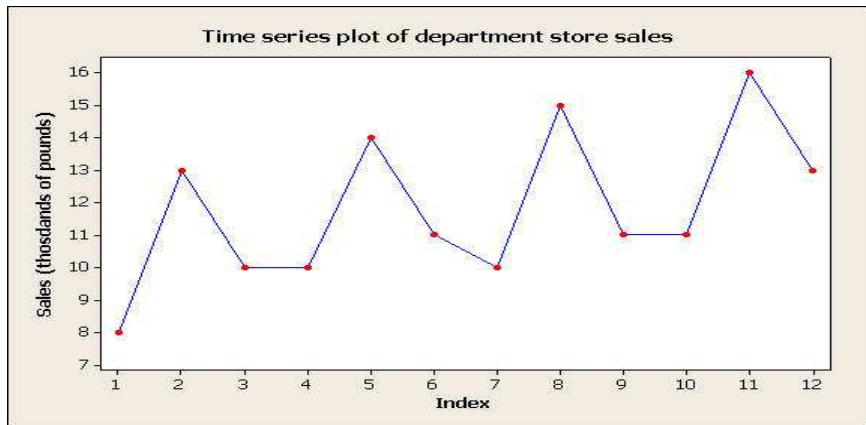
Describing time series



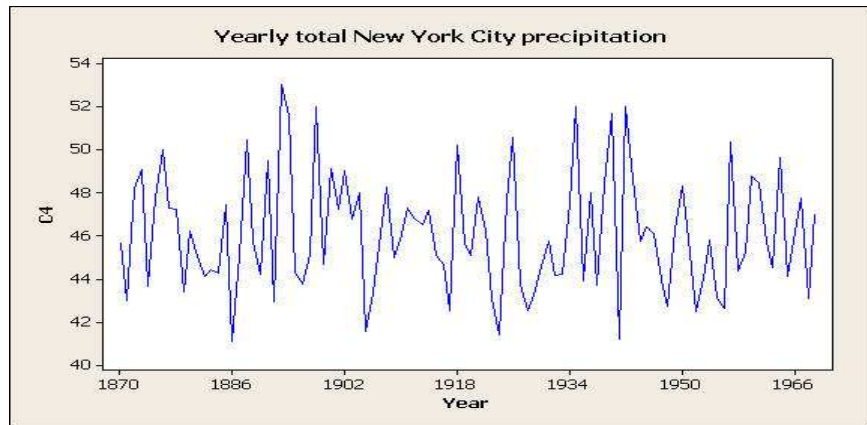
Describing time series



Describing time series



Describing time series



Isolating the trend: MAS1403 review

There are several methods we could use for isolating the trend.

The method we will study is based on the notion of *moving averages*.

To calculate a moving average, we simply average over the cycle around an observation.

Moving averages

For example, for *Turner's* sales figures, we have three “seasons” (Jan–Apr, May–Aug and Sep–Dec) and so a full cycle consists of three observations.

Thus, the first moving average is

$$\frac{8 + 10 + 13}{3} = 10.33,$$

and this is centred around **time point 2**.

Similarly, the second moving average is

$$\frac{10 + 13 + 10}{3} = 11,$$

and this is centred around **time point 3**.

Moving averages

The rest can be calculated in a similar fashion (table 2.1):

	Moving averages		
	Jan–Apr	May–Aug	Sep–Dec
2006	*	10.33	11.00
2007	11.33	11.67	11.67
2008	11.67	12.00	12.33
2009	13.00	13.33	*

Minitab

Quarterly and monthly data

In MAS1403 we considered the calculation of moving averages when the cycle length was a **convenient number**.

In the last example, the cycle length was 3; taking the average over every triple **centres** the average around the middle observation.

Let $Y_1, Y_2, \dots, Y_n$ be our time series of interest, and so y_t , $t = 1, \dots, n$ are the *observed* values at time t .

Then the **three-point moving average** at time t is given by

$$y_t^* = \frac{y_{t-1} + y_t + y_{t+1}}{3},$$

and this is **centred** around time point t .

Quarterly data

Suppose we have 3-monthly (quarterly) data, e.g.

2007	1
	2
	3
	4
2008	1
	2
	3
	4

Now simple averaging over a cycle around an observation cannot be used – this would not be centred on an integer value of t .

Quarterly data

For example, suppose we take $t = (2007, 4)$. Why not average over Q2, Q3 and Q4 of 2007 and Q1 of 2008?

- This would *not* give an estimate of the trend at $t = (2007, 4)$
- This would give us an estimate of the trend somewhere between $t = (2007, 3)$ and $t = (2007, 4)$

Idea! Average over five quarters! No, because...

- ...this would span $(2007, 2)$, $(2007, 3)$, $(2007, 4)$, $(2008, 1)$, $(2008, 2)$...
- ... and so Q2 would have twice as much weight!

Quarterly data

Thus, we use the following formula as an estimate for the moving average at time t :

$$y_t^* = \frac{y_{t-2} + 2(y_{t-1} + y_t + y_{t+1}) + y_{t+2}}{8}.$$

- This centres the moving average around time point t
- This gives equal weight to all quarters!

Example: British Airways passenger figures

Table 2.2 shows the quarterly passenger figures (rounded, in Millions) for British Airways between 2006–2008 (inclusive).

Calculate the series of *quarterly moving averages* and enter your results in the correct cells of table 2.3.

	Q1 (Jan–Mar)	Q2 (Apr–Jun)	Q3 (Jul–Sep)	Q4 (Oct–Dec)
2006	12	6	8	10
2007	14	7	8	13
2008	16	9	10	13

Example: British Airways passenger figures

$$y_3^* = \frac{12 + 2(6 + 8 + 10) + 14}{8}$$

$$= \frac{12 + 48 + 14}{8}$$

$$= 9.25$$

$$y_4^* = \frac{6 + 2(8 + 10 + 14) + 7}{8}$$

$$= 9.625$$

Example: British Airways passenger figures

Similarly,

$$y_5^* = \frac{8 + 2(10 + 14 + 7) + 8}{8} = 9.75$$

$$y_6^* = \frac{10 + 2(14 + 7 + 8) + 13}{8} = 10.125$$

$$y_7^* = \frac{14 + 2(7 + 8 + 13) + 16}{8} = 10.75$$

Example: British Airways passenger figures

$$y_8^* = \frac{7 + 2(8 + 13 + 16) + 9}{8} = 11.25$$

$$y_9^* = \frac{8 + 2(13 + 16 + 9) + 10}{8} = 11.75$$

$$y_{10}^* = \frac{13 + 2(16 + 9 + 10) + 13}{8} = 12$$

Example: British Airways passenger figures

These should be entered into Table 2.3 in the notes:

	Q1 (Jan–Mar)	Q2 (Apr–Jun)	Q3 (Jul–Sep)	Q4 (Oct–Dec)
2006	*	*	9.25	9.625
2007	9.75	10.125	10.75	11.25
2008	11.75	12	*	*

Minitab

Monthly data

By similar reasoning, i.e. to

- ensure our moving averages are centred around an integer time value;
- avoid undue weight being given to a particular “season”,

we use the following formula to obtain moving averages for monthly data:

$$y_t^* = \frac{y_{t-6} + 2(y_{t-5} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+5}) + y_{t+6}}{24}.$$

Example: British visitors to Menorca

	J	F	M	A	M	J	J	A	S	O	N	D
2003	5	3	4	8	10	12	14	20	19	14	6	3
2004	7	4	8	10	15	16	17	21	20	16	8	4
2005	8	5	8	10	16	18	20	22	21	17	9	5

$$\begin{aligned}
 y_7^* &= \frac{5 + 2(3 + 4 + 8 + 10 + 12 + 14 + 20 + 19 + 14 + 6 + 3) + 7}{24} \\
 &= \frac{238}{24} \\
 &= 9.917.
 \end{aligned}$$

Example: British visitors to Menorca

Space has been left on page 55 for you to try some of these calculations yourself; the answers are shown in Table 2.5:

	J	F	M	A	M	J
2003	*	*	*	*	*	*
2004	11.46	11.63	11.71	11.83	12.00	12.13
2005	12.71	12.88	12.96	13.04	13.13	13.21

	J	A	S	O	N	D
2003	9.92	10.04	10.25	10.50	10.79	11.17
2004	12.21	12.29	12.33	12.33	12.38	12.50
2005	*	*	*	*	*	*

Minitab

Using simple linear regression

Look at the plots in Figures 2.3, 2.4 and 2.5:

- Once we've smoothed out the data, the moving averages seem to follow a straight line
- From a forecasting point-of-view, this is great...
- ... we can use some of the ideas from the last chapter to *model* this straight line relationship!

Example: BA passengers data

Let Y^* be our moving averages and let T represent the time index.
Then we have:

t	y^*	t^2	ty^*
3	9.25	9	27.75
4	9.625	16	38.5
5	9.75	25	48.75
6	10.125	36	60.75
7	10.75	49	75.25
8	11.25	64	90
9	11.75	81	105.75
10	12	100	120
52	84.5	380	566.75

Example: BA passengers data

We now reformulate the simple linear regression model from Chapter 1:

$$Y^* = \beta_0 + \beta_1 T + \epsilon,$$

Thus, we now have

$$\hat{\beta}_1 = \frac{S_{TY^*}}{S_{TT}} \quad \text{and}$$

$$\hat{\beta}_0 = \bar{y}^* - \hat{\beta}_1 \bar{t};$$

Example: BA passengers data

The quantities S_{TY^*} and S_{TT} are defined analogously to S_{XY} and S_{XX} in Chapter 1:

$$S_{TY^*} = \sum_{i=3}^{10} t_i y_i^* - n \bar{t} \bar{y}^* \quad \text{and}$$

$$S_{TT} = \sum_{i=3}^{10} t_i^2 - n \bar{t}^2.$$

Example: BA passengers data

Using the sums from the previous table gives:

$$S_{TY^*} = 566.75 - 8 \times \frac{52}{8} \times \frac{84.5}{8}$$

$$= 17.5,$$

$$S_{TT} = 380 - 8 \times \left(\frac{52}{8}\right)^2$$

$$= 42.$$

Example: BA passengers data

Thus, we have

$$\begin{aligned}\hat{\beta}_1 &= \frac{17.5}{42} \\ &= 0.417 \quad \text{and} \\ \hat{\beta}_0 &= \frac{84.5}{8} - 0.417 \times \frac{52}{8} \\ &= 7.852.\end{aligned}$$

So the regression equation is given by

$$Y^* = 7.852 + 0.417T + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

Minitab

Questions

Use the estimated regression equation to forecast total BA passenger numbers in Jan–March 2009.

Why might the global economic situation in 2009–2010 invalidate this forecast?

What else have we not accounted for here?

Announcements

- Computer practical today **cancelled** – I am trusting you to work through practical 3 in your own time
- Lecture next week: **Friday 3–4pm**, BEDB.1.75 (this room)
- Week after – last week of term: Lecture on Tuesday, 12–2; computer practical Tuesday, 4–5pm.
- Up-and-coming deadline: CBA3 – 23:59, Friday 14th March

Isolating the seasonal effects

In the last section we examined how to isolate trend in our time series data.

We did this by

- “smoothing out” the data by finding moving averages;
- fitting a regression line to the series of moving averages

Isolating the seasonal effects

However, as we noted in the last example, any forecasts we make based on the regression line alone do not take into account the *seasonal cycles* around that line.

We will now review the methods used in MAS1403 to identify *seasonal effects*, but will also see this in action in Minitab.

MAS1403 review

In MAS1403 we used several steps to obtain our seasonal effects:

1. Find the *seasonal deviations* ($y_t - y_t^*$, $t = 1, \dots, n$);
2. Calculate the *seasonal means*;
3. Calculate the *seasonal effects*;
4. Obtain the *adjusted seasonal effects*.

Example: BA passenger data

Recall from table 2.2 and 2.3 the quarterly British Airways passenger figures (in millions for 2006–2008), and the corresponding moving averages, respectively:

	Q1 (Jan–Mar)	Q2 (Apr–Jun)	Q3 (Jul–Sep)	Q4 (Oct–Dec)
2006	12	6	8	10
2007	14	7	8	13
2008	16	9	10	13
	Q1 (Jan–Mar)	Q2 (Apr–Jun)	Q3 (Jul–Sep)	Q4 (Oct–Dec)
2006	*	*	9.25	9.625
2007	9.75	10.125	10.75	11.25
2008	11.75	12	*	*

Step 1: Seasonal deviations

The first seasonal deviation is for time-point 3, as there's no moving average for time-points 1 and 2. This is found as

$$8 - 9.25 = -1.25.$$

You can find the others in a similar way; these are shown in the table below, and should be entered into table 2.6:

	Q1	Q2	Q3	Q4
2006	*	*	-1.25	0.375
2007	4.25	-3.125	-2.75	1.75
2008	4.25	-3	*	*
Seasonal means				

Step 2: Seasonal means

The first seasonal mean is for Q1 (Jan–Mar), and this is given by

$$\frac{4.25 + 4.25}{2} = 4.25$$

The others can be found in a similar way:

$$\frac{-3.125 - 3}{2} = -3.0625$$

$$\frac{-1.25 - 2.75}{2} = -2$$

$$\frac{0.375 + 1.75}{2} = 1.0625$$

Step 2: Seasonal means

These should be entered into table 2.6:

	Q1	Q2	Q3	Q4
2006	*	*	-1.25	0.375
2007	4.25	-3.125	-2.75	1.75
2008	4.25	-3	*	*
Seasonal means	4.25	-3.0625	-2	1.0625

Step 3: Seasonal effects

The seasonal effects are found by subtracting the *overall mean* from each of the seasonal means in the last table.

The overall mean is given by

$$\frac{-1.25 + 0.375 + 4.25 - 3.125 - 2.75 + 1.75 + 4.25 - 3}{8} = 0.0625.$$

So the first seasonal effect is

$$\begin{aligned}\hat{s}_{Q1} &= 4.25 - 0.0625 \\ &= 4.1875\end{aligned}$$

The others are given on the next slide.

Step 3: Seasonal effects

$$\begin{aligned}\hat{s}_{Q2} &= -3.0625 - 0.0625 \\ &= -3.125\end{aligned}$$

$$\begin{aligned}\hat{s}_{Q3} &= -2 - 0.0625 \\ &= -2.0625 \quad \text{and}\end{aligned}$$

$$\begin{aligned}\hat{s}_{Q4} &= 1.0625 - 0.0625 \\ &= 1\end{aligned}$$

Step 4: Adjusted seasonal effects

Recall from MAS1403 that the seasonal effects must sum to zero.
Ours give:

$$\begin{aligned}\hat{s}_{Q1} + \hat{s}_{S2} + \hat{s}_{Q3} + \hat{s}_{Q4} &= 4.1875 - 3.125 - 2.0625 + 1 \\ &= 0\end{aligned}$$

and so no adjustments are necessary!

If they didn't add to give zero, you would

- take the mean of your seasonal effects;
- subtract this from each seasonal effect.

Minitab

Forecasting

Recall the question at the top of page 60 in your lecture notes notes:

Use the estimated regression equation to forecast total BA passenger numbers in Jan–March 2009.

We can now do this more realistically by adjusting our original forecast obtained via the regression equation for the seasonal effect for Jan–March.

Forecasting

Recall that the regression equation for the moving averages was found to be:

$$Y^* = 7.852 + 0.417T + \epsilon.$$

January–March 2009 would be time–point 13, and so using this regression equation gave us a forecast of

$$\begin{aligned} Y^* &= 7.852 + 0.417 \times 13 \\ &= 13.273, \end{aligned}$$

or just over 13 million passengers.

Forecasting

However, you'll notice from figure 2.7 that the first quarter of each year always seems to record higher than average passenger figures...

... we now adjust this initial forecast by the seasonal effect for January–March, which was found to be +4.1875, giving a *full forecast* of

$$13.273 + 4.1875 = 17.4605,$$

or just under 17.5 million passengers.

Note that this has still not taken into account the global financial situation of late!

Obtaining the residual series

In the next section, we will consider some special *probability models* for time series data.

These models assume that our data are *stationary*, i.e. have no trend or seasonality.

Most of the time, our time series data exhibit either trend, or seasonality, or *both* – in fact, this is what makes time series data so interesting – and so these probability models are not immediately useable.

Obtaining the residual series

However, if we can estimate the trend and seasonal components of our data – and we have shown how to do this in the previous two sections – we can attempt to *make* our time series stationary by *de-trending* and *de-seasonalising*.

Minitab

The $AR(1)$ time series model

Even when we removed the trend and seasonality from the long haul dataset, there still remained significant correlation between observations one time–point apart.

There was *significant lag 1 autocorrelation* present in the series.

This is often the case with time series data, and is one of the things that makes such datasets interesting!

The $AR(1)$ time series model

Not only can we model the trend and the seasonal variation, but there are also techniques available for modelling the autocorrelation.

One method is via a class of time series models known as *Autoregressive models*, or *AR* models for short.

The $AR(1)$ time series model

The simplest type of AR model is the $AR(1)$ model, whereby each observation is dependent on the observation at lag 1 – i.e. the observation immediately preceding it.

The model is defined as:

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

- μ is the mean
- α is known as the “autoregressive parameter”
- ϵ_t represents random error.

All autoregressive time series models assume that our series is *stationary* and so our Y_t would represent the residual series.

The $AR(1)$ time series model

In Minitab, we

- click on Stat–Time Series–ARIMA;
- enter our time series for analysis in Series;
- enter **1** in the Autoregressive (Nonseasonal) box;
- click OK.

Minitab

The $AR(1)$ time series model

From this output, we see that:

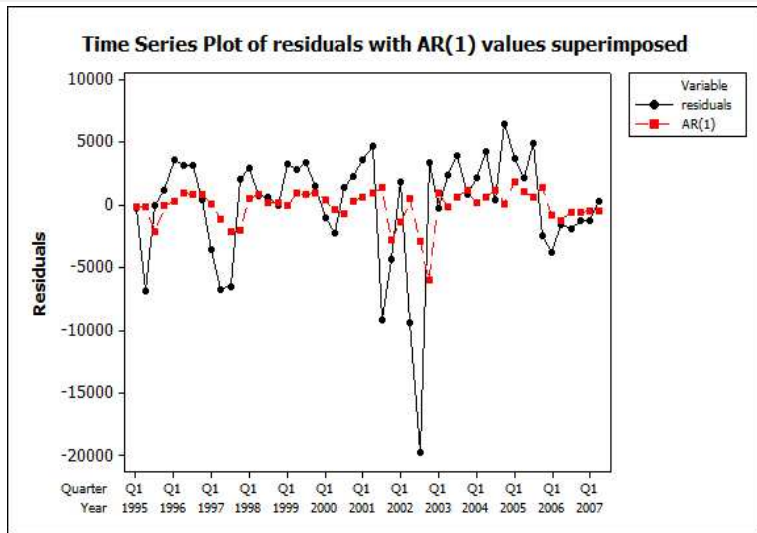
- The estimate of μ is -110.1
- The estimate of α is 0.3005 , and this is significant since it has a p -value less than 0.05

Thus, we obtain the following $AR(1)$ model for our residuals:

$$Y_t = -110.1 + 0.3005(Y_{t-1} - (-110.1)) + \epsilon_t, \quad \text{i.e.}$$

$$Y_t = -110.1 + 0.3005(Y_{t-1} + 110.1) + \epsilon_t.$$

The $AR(1)$ time series model



The $AR(1)$ time series model

Comments

- Notice how the AR model “follows” the data
- In order to forecast Y_t , we need to know the previous value in the series
- Such a forecast would not include trend nor seasonal effects – to obtain a “proper” forecast, we would have to add these effects back in

Forecasting using the AR(1) model

We can now obtain forecasts using the fitted $AR(1)$ model.

For example, suppose we want to forecast the number of long-haul holidays in the 3rd quarter of 2007.

The last observed value was for the 2nd quarter of 2007, and this was time-point 50.

Thus, the 3rd quarter of 2007 would be time-point 51. This gives an estimated value *from the residual series* of

$$\begin{aligned}Y_{51} &= -110.1 + 0.3005(\mathbf{309.5} + 110.1) \\ &= 15.9898\end{aligned}$$

309.5??

Forecasting using the AR(1) model

Recall that the simple linear regression equation for the moving averages was

$$Y_t^* = 30779.0 + 4943.52T + \epsilon.$$

Thus, an estimate of the trend at time–point 51 would be

$$\begin{aligned} Y_{51}^* &= 30779.0 + 4943.52 \times 51 \\ &= 282898.52. \end{aligned}$$

Forecasting using the AR(1) model

Recall also that the seasonal effect for the 3rd quarter was found to be

$$\hat{s}_{Q3} = +20284.6.$$

Adding the trend and seasonal effects back in gives a *full forecast* of

$$15.9898 + 282898.52 + 20284.6 = 303199.1098,$$

or about 303,199 long-haul holidays.

Minitab

Confidence interval for forecasted values

$$\text{Lower} : -8643.86 + 282898.52 + 20284.6 = 294539.26$$

$$\text{Upper} : 8675.85 + 282898.52 + 20284.6 = 311858.97,$$

i.e. about (294,539, 311,859) long-haul holidays.

The AR(p) time series model

A time series is said to be an *autoregressive process of order p* , abbreviated $AR(p)$, if

$$\begin{aligned} Y_t &= \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \dots + \alpha_p(Y_{t-p} - \mu) + \epsilon_t, \\ &= \mu + \sum_{i=1}^p \alpha_i(Y_{t-i} - \mu) + \epsilon_t. \end{aligned}$$

The name “autoregressive” comes from the similarity to multiple regression, but the predictor variables $X_1, X_2, \dots, X_p$ have been replaced by *past values* of Y .

The $AR(p)$ time series model

We have already met an $AR(1)$ model, where the current value in the process depends only upon the previous value.

Likewise, the current value only influences the very next value.

You could think of this as

“what happens today is only influenced by what happened yesterday – not the day before”

But it is often the case that what happens today *will be* influenced by what happened two days ago, or even three days ago, or even the day before that!

The AR(p) time series model

This is where AR(p) processes come in handy, where $p > 1$.

For example, if you think what happens today is influenced by what happened yesterday *and* the day before, you could use an AR(2) time series model, i.e.

$$Y_t = \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \epsilon_t.$$

Or maybe you also think that what happened *three* days ago could also play a part; hence, an AR(3) process, i.e.

$$Y_t = \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \alpha_3(Y_{t-3} - \mu) + \epsilon_t.$$

The AR(p) time series model

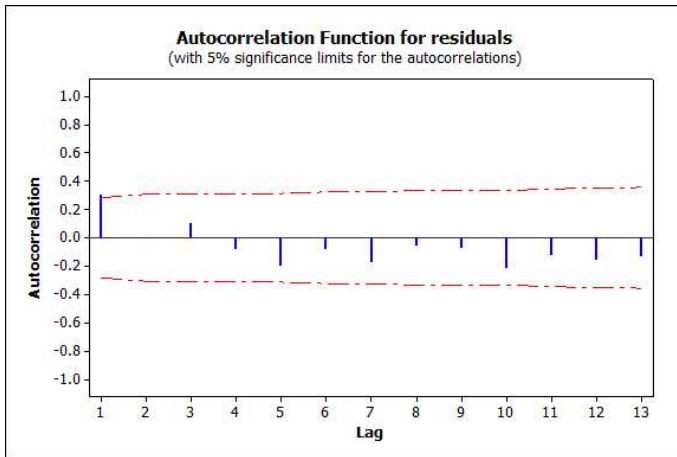
We can fit such models in Minitab.

But how do we know which of these models would be the most appropriate?

In other words, how can we find out the *order* of our *AR* model?
How do we know how many past values will influence what we observe today?

The AR(p) time series model

Recall the *correlogram* shown in figure 2.13:



The $AR(p)$ time series model

The only significant autocorrelation was at lag 1, suggesting that the only significant serial dependence is between values one time–point apart.

However, choosing the most appropriate AR process from a correlogram *can* be quite hard – sometimes a very strong autocorrelation at lag 1 can induce a significant autocorrelation at lag 2 or beyond, even if there is no genuine lag 2 (or lag 3, or lag 4 etc.) autocorrelation.

The AR(p) time series model

A plot of the *partial autocorrelations* provides a way around this.

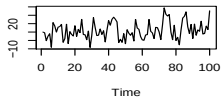
The *lag k partial autocorrelation* more genuinely represents the dependence between observations k time-points apart, after taking into account any intermediate dependencies.

Thus, a plot of these partial autocorrelations can provide a diagnostic tool for the selection of an appropriate autoregressive time series model.

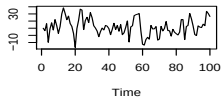
Minitab

The AR(p) time series model

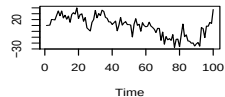
Series 1



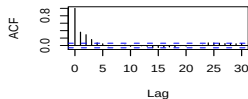
Series 2



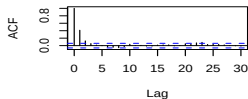
Series 3



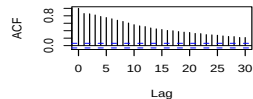
Autocorrelations



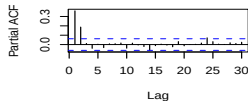
Autocorrelations



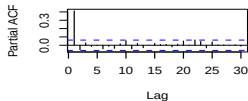
Autocorrelations



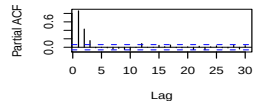
Partial autocorrelations



Partial autocorrelations

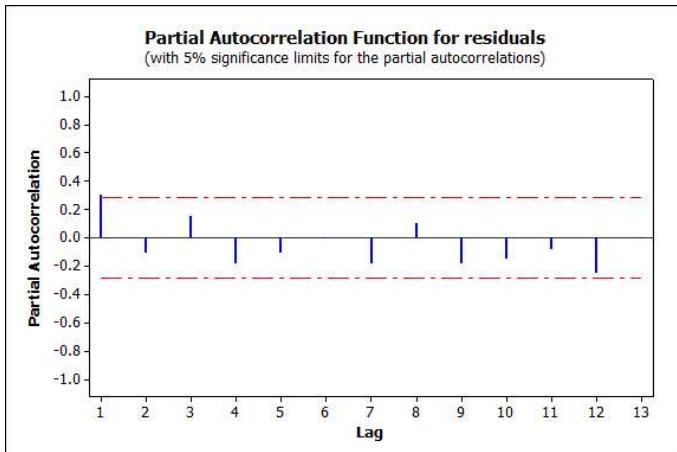


Partial autocorrelations



The AR(p) time series model

Figure 2.16 shows a plot of the partial autocorrelations for the residual series for the long haul dataset:



The $AR(p)$ time series model

But what would happen if we *were* to fit, say, an $AR(2)$ model in Minitab? We would

- click on Stat–Time Series–ARIMA;
- enter our time series for analysis in Series;
- enter **2** in the Autoregressive (Nonseasonal) box;
- click OK.

Minitab

The AR(p) time series model

From this output, you should see that the fitted model is

$$Y_t = -87.5 + 0.3355(Y_{t-1} - (-87.5)) \\ - 0.1165(Y_{t-2} - (-87.5)) + \epsilon_t, \quad \text{i.e.}$$

$$Y_t = -87.5 + 0.3355(Y_{t-1} + 87.5) - 0.1165(Y_{t-2} + 87.5) + \epsilon_t,$$

where Y_{t-1} is the previous value in the series and Y_{t-2} is the value before that.

The AR(p) time series model

You should also notice that

- The estimate of the first AR parameter is $\hat{\alpha}_1 = 0.3355...$
- ... this *is* significant in the model – it has a p -value less than 0.05
- The estimate of the second AR parameter is $\hat{\alpha}_2 = -0.1165 ...$
- ... this is *not* significant in the model since its p -value is 0.426, which is *greater* than 0.05.

Thus, the second AR term is unnecessary in our model and so the simpler $AR(1)$ model is more appropriate.

Recap of probability models

Once we have a stationary series (the “residual series”) we can try to fit **Autoregressive models**.

We use a plot of the partial autocorrelations to suggest the **order** of the model.

We then use Minitab to **fit** this model, e.g.:

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

Recap of probability models

We should check the **significance** of our autoregressive parameters – as with linear regression, the simpler the model, the better!

We can then use our model to **forecast**:

- Use $y_{t-1}, \dots$ from the residual series
- Add back in the trend
- Add back in the seasonal effects
- This will give a **full forecast** accounting for time dependence, trend and seasonality

We can also use Minitab to help us obtain **confidence intervals** for our forecasts.

Full worked example

Consider the Monsters dataset, available from the ACE2013 website.

In 1999 a friend of mine set up an online party pack business. Her idea, initially, was to sell party packs online for Hallowe'en parties, and so she set up the *Monster Parties* website. She was a poor student and needed some extra cash for the pub at weekends! Her business went from strength to strength; she finished her Marketing degree in 2002 and went into the business full-time, with help from a local Business Enterprise centre.

The figures shown in the Monsters worksheet are monthly sales between 2000 and 2009 (inclusive), giving 120 observation in total (12 observations per year over ten years). A time series plot for these data is shown in figure 2.17.