

# An Introductory Course on Time Series Analysis



[timeseries@abs.gov.au](mailto:timeseries@abs.gov.au)

May 2001

# Contents

1. An Introductory Look at Time Series
  - 1.1 What is a time series?
  - 1.2 Stock and flow series
  - 1.3 What is seasonal adjustment?
  - 1.4 How does Time Series Analysis fit into the ABS?
  
2. Decomposition Models
  - 2.1 What is a trend?
  - 2.2 What is seasonality?
    - Identifying seasonality
  - 2.3 What is a residual/irregular?
  - 2.4 Models for decomposing a series
    - Additive decomposition model
    - Multiplicative decomposition model
    - Pseudo-additive decomposition model
    - Summary: Model comparisons
  - 2.5 Example: Decomposition of a series
  
3. Seasonal adjustment philosophies
  - 3.1 Model based methods
  - 3.2 Filter based methods
  - 3.3 Comparison of the two approaches
  
4. Frequency Domain Analysis
  - 4.1 Autocovariance
  - 4.2 Fourier analysis
  - 4.3 Spectral analysis
  - 4.4 Gain functions
  - 4.5 Phase shift
  
5. Filtering - the empirical approach
  - 5.1 General framework
  - 5.2 Example: Application of a linear filter
  - 5.3 Trend moving averages (trend filters)
    - Henderson trend moving averages
    - Symmetric Henderson moving averages
    - Asymmetric Henderson moving averages
  - 5.4 Seasonal moving averages (seasonal filters)
    - Symmetric seasonal moving averages
    - Asymmetric seasonal moving averages

6. Issues with seasonal adjustment
  - 6.1 Moving holidays
  - 6.2 Trading day
    - Determining trading day
    - Trading day for quarterly series
    - Length of month effect
  - 6.3 Extremes (outliers)
    - Detecting extremes
    - Querying an extreme
    - Calculation of a correction factor
  - 6.4 Designing a survey for trend or seasonally adjusted estimates
    - What do users really want?
    - Background: Rotation patterns
    - Designing a survey for sampling variance
    - Designing a survey for revision of trend estimates
    - Summary: The key points
  - 6.5 Indirect (aggregate) versus Direct (disaggregate)
    - When to use Aggregative/Disaggregative
    - Methods of aggregation
  - 6.6 Forward factors versus concurrent adjustment
    - Using forward factors
    - Using concurrent adjustment
    - Emergency seasonal adjustment procedures for concurrent adjustment
7. The mysterious X11
  - 7.1 A simple seasonal adjustment approach
  - 7.2 Outline of the steps involved within X11
  - 7.3 Output from X11
  - 7.4 Beyond X11?
  - 7.5 Using X11
8. SEASABS
  - 8.1 Introduction
  - 8.2 SEASABS
  - 8.3 Time Series Toolkit
  - 8.4 Time Series Update
  - 8.5 Upholding ABS quality standards
  - 8.6 The basic steps in running SEASABS
  - 8.7 SEASABS demo
9. Exercises
10. Solutions
11. Further Reading

## Preface

The mission of the Australian Bureau of Statistics is to encourage informed decision making, by providing a high quality statistical service. Time series analysis is a fundamental component of this vision as seasonal adjustment techniques can be used to highlight important characteristics that exist in data which can aid in the decision making process.

The course notes describe some of the issues that are relevant to time series analysis in the Australian Bureau of Statistics. Included is an introduction to some of the theory underlying time series analysis, a discussion of different seasonal adjustment philosophies and issues relating to seasonal adjustment. The course notes concentrate on the filter based approach to seasonal adjustment as this is the method most commonly applied in the ABS.

The course notes are intended for methodologists newly recruited to the Australian Bureau of Statistics. Exercises and solutions are included at the end of the course notes. The purpose of these course notes and exercises is to encourage understanding of time series analysis issues.

## Acknowledgements

The course notes, exercises and solutions were put together with the help of the following people (alphabetically):

Kate Bradbury, Jeff Cannon, Lujuan Chen, Katherine Fewell,  
Gemma Van Halderen, Joseph Hudson, Shuyan Huo,  
Craig McLaren, Fiona Mackie, Tom Outteridge,  
Nick von Sanden, Andrew Sutcliffe, Frank Yu, Mark Zhang

Benoit Quenneville (Statistics Canada) and Godfrey Lubulwa (Australian Bureau of Statistics) provided many helpful and insightful comments on earlier versions of the course notes. Of course, any errors or omissions which remain are entirely the responsibility of the Time Series Analysis team and we would appreciate them being brought to our attention.

Time Series Analysis  
timeseries@abs.gov.au

First edition: July 2000  
Second edition: May 2001

## Notation

ABS	Australian Bureau of Statistics
SSB	Statistical Services Branch
TSA	Time Series Analysis
SEASABS	SEASonal analysis ABS standard software package
MA	Moving average
HMA	Henderson moving average
SMA	Seasonal moving average
TMA	Trend moving average
ARIMA	AutoRegressive Integrated Moving Average method of modelling
X11	Filter based seasonal adjustment program which SEASABS is based on
ABSDB	ABS Data Base
FAME	Forecasting, Analysis and Modelling Environment
PAS	Publishing Assembly System
TST	Time Series Toolkit
TSU	Time Series Update
STAMP	Along with SABL and SEATS/TRAMO are other seasonal adjustment software packages
ewp	End weight parameter
$t$	Time index
$y_k$	Data point at time $t = k$
$Y_t$	Time series
$\varepsilon_k$	Error term at time $t = k$
$O_t$	Original data
$S_t$	Seasonal component
$T_t$	Trend component
$f_s(y_t)$	Symmetrically filtered value
$I_t$	Irregular/residual component
$SA_t$	Seasonally adjusted data
$f_a(y_t)$	Asymmetrically filtered value
$OM_t$	Outlier modified original data
SI	Seasonal $\times$ Irregular (S-I)
I/C	I/C ratio: noise to trend ratio
P	Prior factors
OP	Original data prior corrected
OT	Original data modified for trading day
SAM	Modified seasonally adjusted series
IM	Modified irregular/residual series
L	Length of month factors (trading day)
TD	Trading day factors
SIM	Modified SI values
TDL	Trading day factors $\times$ length of month factors
D	Daily factor weight
DI	Daily factor weight $\times$ irregular/residual (D-I)
$r_t^{(k,i)}$	Revision between two smoothed values for time $t$ using data up to times $t + i$ and $t + k$
$\omega$	Frequency

$p$	Period
$B$	Backshift operator, $B^k y_t = y_{t-k}$
$\Delta$	Difference operator, $\Delta y_t = y_t - y_{t-1}$
$\Delta^k$	$k^{\text{th}}$ difference operator, $\Delta^k y_t = y_t - y_{t-k} = (1 - B^k)y_t$
$\gamma_k$	Autocovariance at time $t = k$
$\gamma(B)$	Autocovariance generating function
$G(\omega)$	Gain function
$\Phi(\omega)$	Phase function
$S_y(\omega)$	Spectral function
$w_i$	Filter weight at time $i$ for a symmetric filter
$u_i$	Filter weight at time $i$ for an asymmetric filter
$m_1$	Number of terms to the left of centre in a filter
$m_2$	Number of terms to the right of centre in a filter
$S_{x \times y}$	Seasonal $x \times y$ filter

# 1 An introductory look at time series

## 1.1 What is a time series?

A time series is a collection of observations of well-defined data items observed through time. For example, a measure of the number of tonnes of wool produced each month of the year would comprise a time series. This is because wool is well defined, it is easy to determine what wool is, and the amount produced is measured at equally spaced intervals. Data collected irregularly or only once are not time series. For example, the survey of household usage of electrical appliances was only run once and so is not a time series.

Time series are of interest primarily for economic analysis because they help identify cycles, such as business cycles, and turning points in those business cycles. A time series can be decomposed into various components: regular or systematic variations, irregular influences such as those associated with one-off events, and underlying movements. For example, analysts are sometimes interested in long term cycles, which are indicators of growth measures, and so it is helpful if shorter length cycles are removed.

Some of the economic time series collected by the Australian Bureau of Statistics (ABS) that are of interest include:

- Current Account Deficit
- Balance of Payments
- Retail Turnover
- Employment/Unemployment
- Motor Vehicle Registrations

Overall, the ABS analyses around 2600 different types of time series.

## 1.2 Stock and flow series

Time series can be classified into two different types: stock and flow. The ABS produces both stock series and flow series.

Stock series are measures of activity at a point in time and can be thought of as “stocktakes”. For example, the Monthly Labour Force Survey is a stock measure because it takes stock of whether a person was employed in the reference week. Other examples of stock series include Inventories and Average Weekly Earnings.

Flow series are series which are a measure of activity to a date, for example, the number of bicycles produced this month. Manufacturing is also a flow measure because a certain amount is produced each day, and then these amounts are summed to give a total value for production for a given reporting period. Examples of flow series include Current Account Deficit, Balance of Payments, Housing Finance, Capital Expenditure, and Manufacturing.

The main difference between a stock and a flow series is that flow series can contain effects related to the calendar (trading day effects). Both types of series can still be adjusted using the same seasonal adjustment process.

### 1.3 What is seasonal adjustment?

Seasonal adjustment is an analysis technique that estimates and then removes from a series influences that are systematic and calendar related. A seasonally adjusted series can be formed by removing the systematic calendar related influences from the original series. A trend series is then formed by removing the remaining irregular influences from the seasonally adjusted series. Alternatively the trend can be estimated directly by removing both the calendar related influences and the irregular influences from the original series.

Bell and Hillmer (1984) suggested that seasonal adjustment has three main purposes:

1. to aid in short term forecasting
2. to aid in relating time series to other series or extreme events
3. to allow series to be compared from month to month

Seasonal adjustment helps ABS clients to assess underlying patterns in a series without significant loss of information.

Seasonal adjustment is a relatively simple analysis technique. However, there are a few issues that can complicate the process and lead to a seasonally adjusted and/or a trend series that is unusual. These issues will be covered in greater depth in Chapter 6. They are

1. Outliers (unusual estimates):  
The focus is on unusual estimates, not unusual observations as in the sampling sense. Outliers can cause blips in an original series, seasonally adjusted series and trend series unless they are modified or corrected during the seasonal adjustment process.
2. Revisions:  
The seasonal adjustment process leads to revisions to the seasonally adjusted and trend series. Revisions are not desirable, either for the ABS or the users of the series. The analysis technique chosen aims to strike a balance between revisions and quality of the seasonally adjusted and trend series. This issue is commonly referred to as the End Weight problem.
3. Aggregation and Disaggregation:  
Regular and irregular influences are often estimated and removed from series at fine levels of disaggregation, such as at the State by Industry level. Higher level seasonally adjusted series, such as at the Australia level, can be constructed by adding up component series to a higher level (to form an indirectly adjusted series) or by directly seasonally adjusting the higher level series (to form a directly adjusted series). The resulting series will not be identical. A common issue faced by time series analysts is explaining why the two approaches do not result in the same series.

Seasonal adjustment techniques can be categorised into two main approaches or philosophies. These are the model based approaches and filter based approaches to obtaining estimates of the seasonal, irregular and trend components of a series.

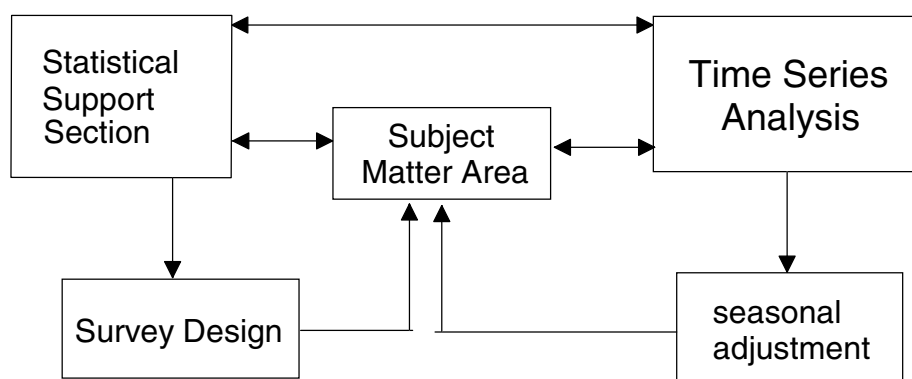


## 1.4 How does Time Series Analysis fit into the ABS?

The seasonal adjustment process is performed at the end of the survey cycle. The survey has been designed and tested; data has been collected, processed and edited; and estimates are produced. These estimates are now referred to as original estimates.

The seasonal adjustment process starts once the original estimates are available. The original estimates are then collected as a time series which forms an original series. Seasonally adjusted series are derived by removing systematic calendar related effects from the original series. A trend series is derived by removing irregular influences from the seasonally adjusted series.

Figure 1.1 illustrates how this process works within the ABS.



**Figure 1.1: The Statistical Support Branch (SSB) support cycle within the Australian Bureau of Statistics**

The Time Series Analysis (TSA) section is responsible for ensuring the methods used to derive seasonally adjusted and trend estimates are appropriate and correctly implemented. This has been achieved by using the ABS seasonal adjustment package, SEASABS. SEASABS is used to reanalyse each time series to update estimates of the systematic calendar related effects and irregular influences that are removed during the seasonal adjustment process.

Many issues can affect a seasonally adjusted and trend series. These include outliers, methodological changes, and real world events such as tax reform. A major responsibility of a time series analyst is to help subject matter areas understand the behaviour of their seasonally adjusted and trend series, and to inform ABS clients about the underlying patterns in the series.

## 2 Decomposition models

Decomposing a time series highlights important features of the data. This can help with the monitoring of a series over time, especially with respect to the making of policy decisions.

A simple approach to decomposition is to assume that any time series can be decomposed into three basic components:

- the trend, or the underlying level of the original series
- the seasonal, or the portion of the variation about the trend attributable to factors that reoccur systematically one or more times a year
- the irregular, or the residual variation, remaining after the trend and seasonal components have been removed from the original series

Decomposition of a time series into a model is not unique. The problem is to define and estimate components that fit a reasonable theoretical framework as well as being useful for interpretation and policy formulation. When choosing a decomposition model, the aim is to choose a model that yields the most stable seasonal component, especially at the end of the series. Decomposition models are typically multiplicative or additive, but on occasions can take other forms such as pseudo-additive.

This chapter introduces the three key components in the decomposition of a time series.

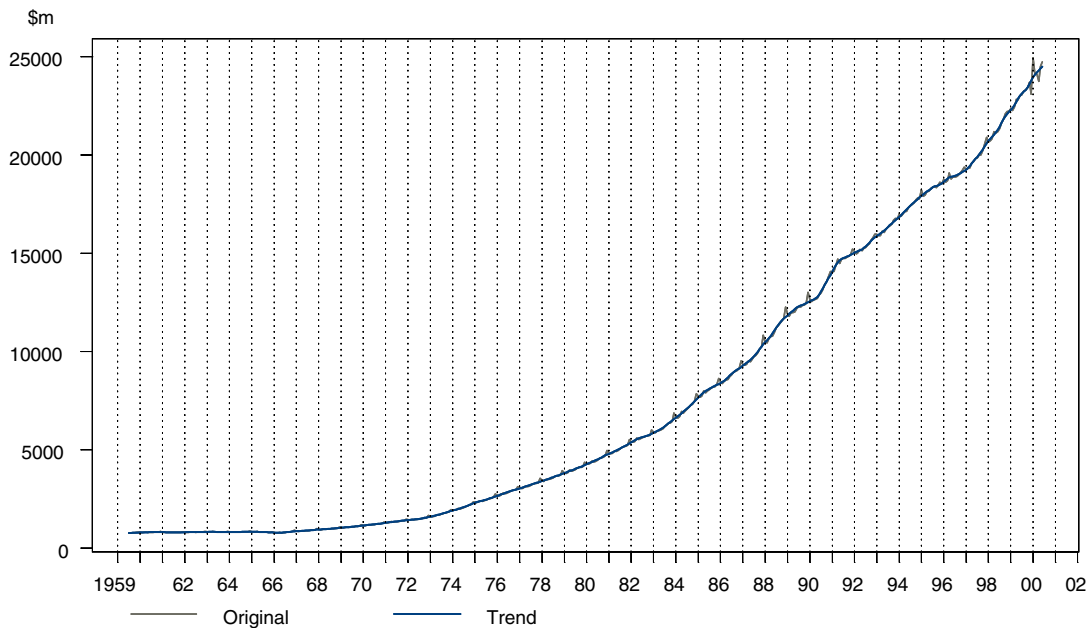
### 2.1 What is a trend?

The trend component, as used by the ABS, is defined as the long term movement in a series. The trend is a reflection of the underlying level of the series. This is typically due to influences such as population growth, price inflation and general economic development. For the purposes of seasonal adjustment, both the long term trend and the medium term cycles are treated as 'the trend' component of a time series. For this reason, the trend component is sometimes referred to as the trend cycle.

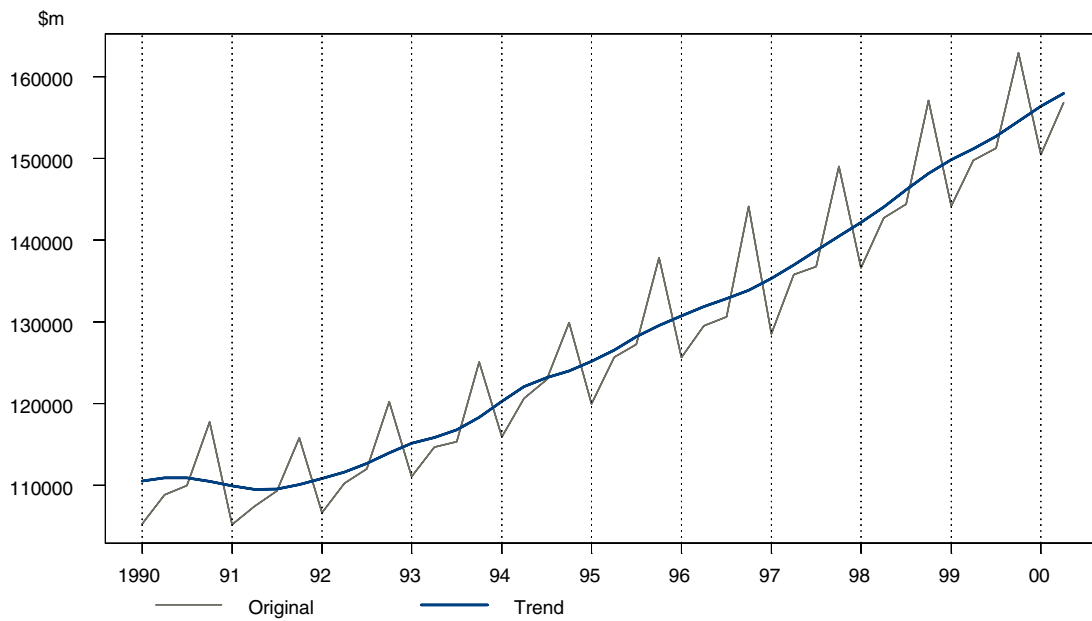
Problems arise when trying to estimate the trend at the end of a time series, as the problem then becomes a forecasting, as well as smoothing, exercise. See Chapter 5 for more details.

Included below are some examples of time series which exhibit a strong, easily identifiable trend.

Figure 2.1 shows the amount of currency in circulation. Over the last 40 years the level of currency in circulation has been gradually and steadily increasing. The original data and the trend series are both plotted. These lines are almost identical as there is little seasonality or irregularity in the data. Overall the long term movement of the series has been to trend upwards.



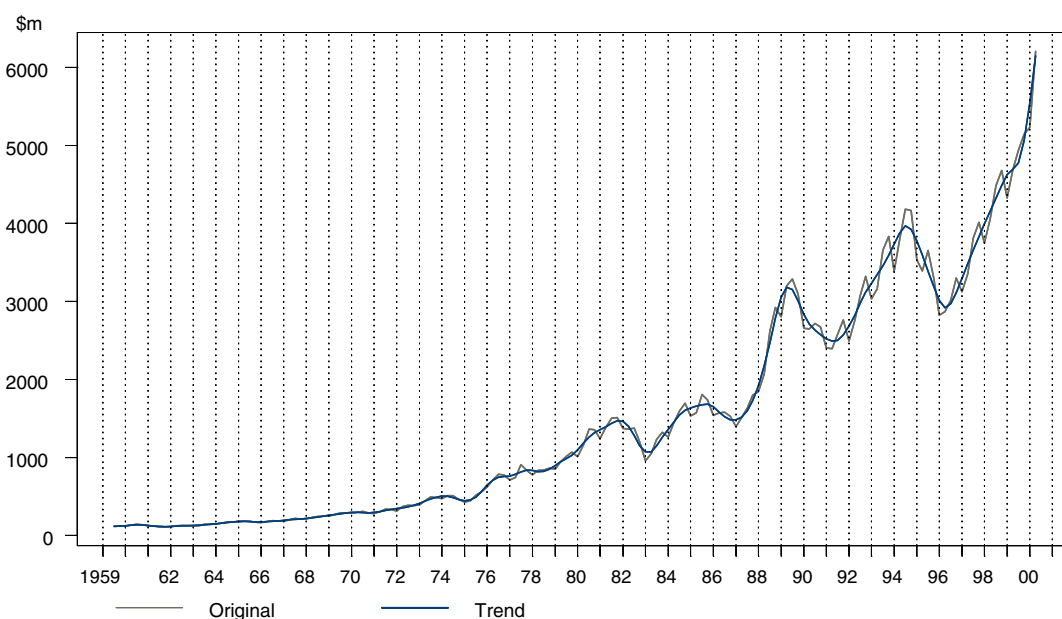
**Figure 2.1: Monthly Currency in Public Circulation - Original and Trend**



**Figure 2.2: Quarterly Gross Domestic Product - Original and Trend**

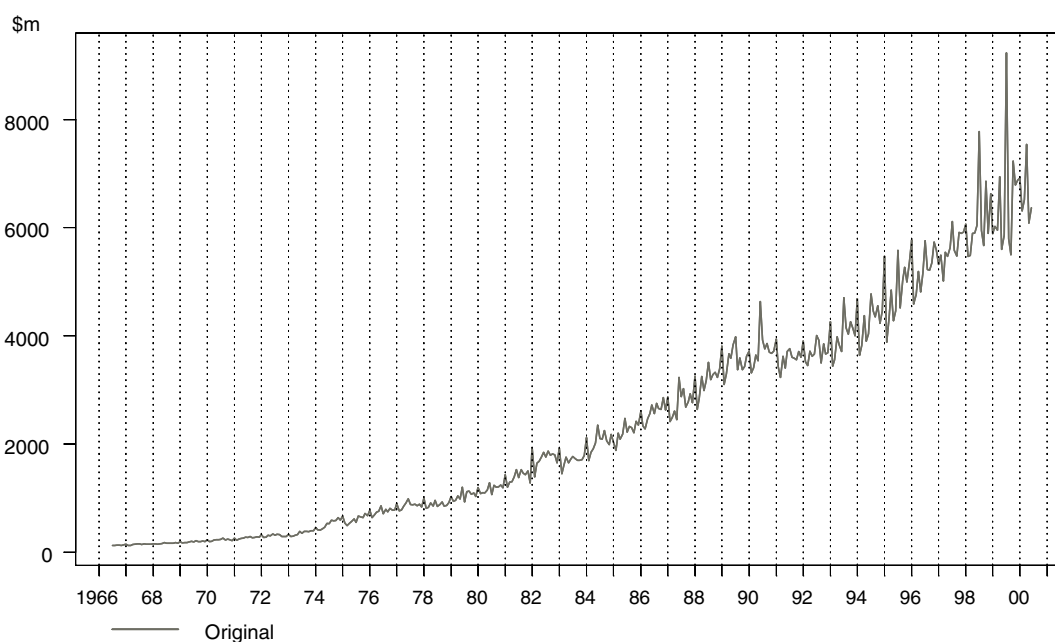
Figure 2.2 shows a very seasonal series (seasonality is discussed further in Section 2.2) which has also been trending steadily upwards over time. Both the original and the trend series have been included. The trend shows the general long term movement of the series in the absence of seasonal and irregular effects. In this case the trend of Australia's GDP has generally been increasing except for a slight dip during the recession of the early 1990's.

Figure 2.3 shows the household gross fixed capital expenditure in original and trend terms. In this case the trend experienced several turning points as it tracked the medium-term cycles in the data. In comparison, visual inspection suggests that over the last 40 years the long term trend of this series has been gradually upwards.



**Figure 2.3: Quarterly Household Gross Fixed Capital Expenditure - Original and Trend**

Figure 2.4 shows monthly Pay As You Earn (PAYE) tax installments. Over the last 30 years the level of monthly PAYE tax installments has been gradually and steadily increasing. In this case the long term movement of the series has been to trend upwards. Notice also that the irregularity in the series, which can be viewed as the level of random movement around the trend, is also increasing over time. This suggests that the amount of irregular movement in the series is related to the level of the trend and that it can best be modelled using a multiplicative decomposition (see Section 2.4).



**Figure 2.4: Gross Pay As You Earn (PAYE) Monthly Tax Installments - Original**

## 2.2 What is seasonality?

Seasonality can be thought of as factors that recur one or more times per year. A seasonal effect is reasonably stable with respect to timing, direction and magnitude. The seasonal component of a time series comprises three main types of systematic calendar related influences: seasonal influences, trading day influences and moving holiday influences.

Seasonal influences represent intra-year fluctuations in the series level, that are repeated more or less regularly year after year. For example, seasonal influences may:

- be representative of the seasons such as warmth in Summer and cold in Winter. Weather conditions that are out of character for a particular season, such as snow in a summer month, would appear in irregular, not seasonal influences
- reflect traditional behaviour associated with the calendar and the various social (Chinese New Year), business (quarterly provisional tax payments), administrative procedures (tax returns) and effects of Christmas and the holiday season

Trading day influences refer to the impact on the series, of the number and type of days in a particular month. A calendar month typically comprises four weeks (28 days) plus an extra one, two or three days. The activity for the month overall will be influenced by those extra days whenever the level of activity on the days of the week are different.

Moving holiday influences refer to the impact on the series level of holidays that occur once a year but whose exact timing shifts systematically. Examples of moving holidays include Easter and Chinese New Year where the exact date is determined by the cycles of the moon.

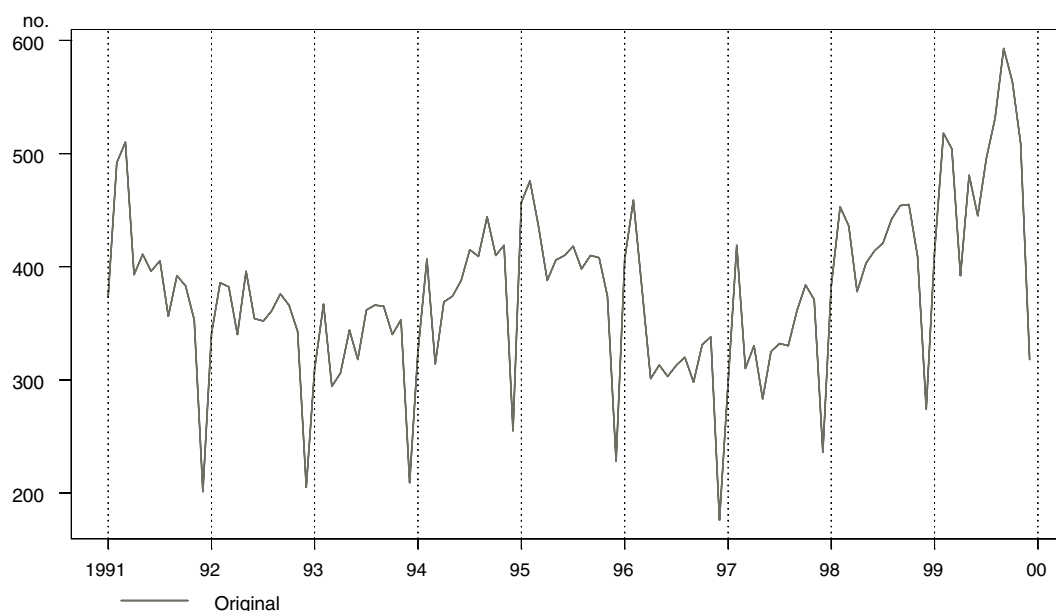
Trading day and moving holidays are considered in more detail in Chapter 6.

## Identifying seasonality

Seasonality can be identified as regularly spaced peaks and troughs (every 12 observations for monthly data and every four observations for quarterly series) which have a consistent direction and have approximately the same magnitude every year. The following examples present some ‘typical’ time series which exhibit a strong seasonal component.

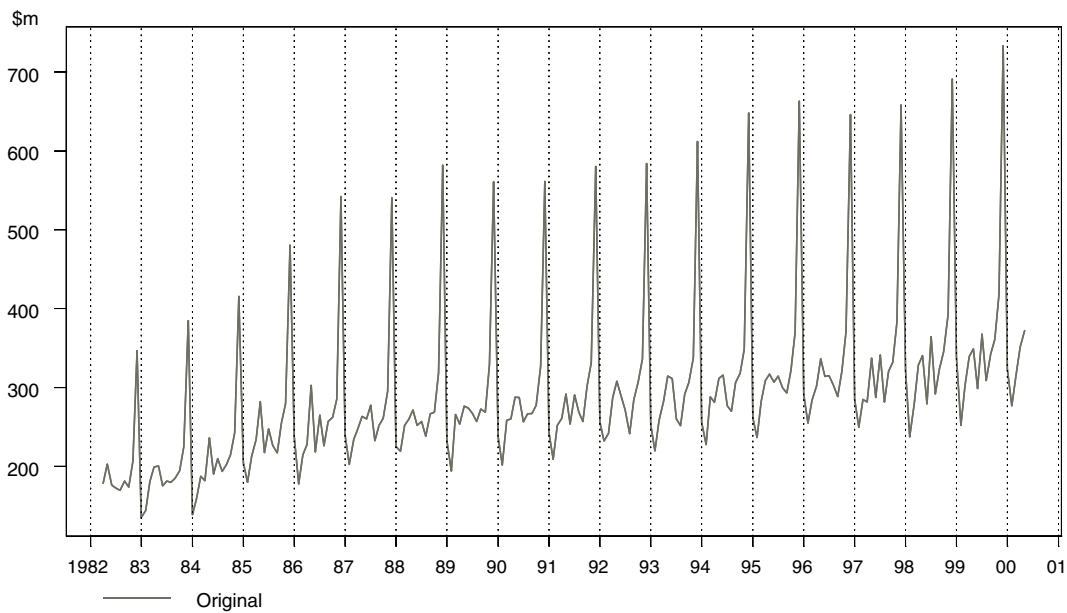
Estimates of seasonal and trading day influences in ABS time series can either remain fixed or evolve over time. By allowing them to move, the combined seasonal and trading day estimates, known as factors, reflect changes in the intra-year activity over the life of the time series.

Figure 2.5 shows the number of job advertisements in the Australian Capital Territory (ACT). In December of every year there is a seasonal low in the number of job advertisements due to the holiday season. This seasonality can be seen as sharp regular dips in the series.



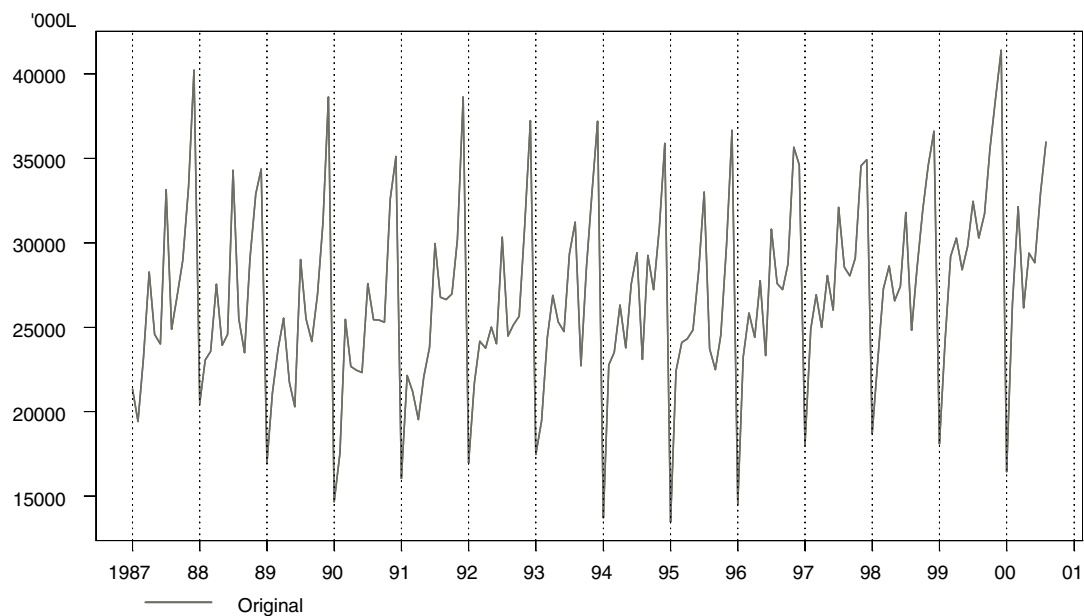
**Figure 2.5: Monthly ANZ Job Advertisements, Australian Capital Territory (ACT) - Original**

Figure 2.6 shows the monthly retail sales for New South Wales (NSW) retail department stores. There is an obvious large seasonal increase in retail buying in December due to Christmas purchasing and a much smaller seasonal increase during the middle of the year. The seasonal highs can be seen as the regular sharp upward spikes in the series.



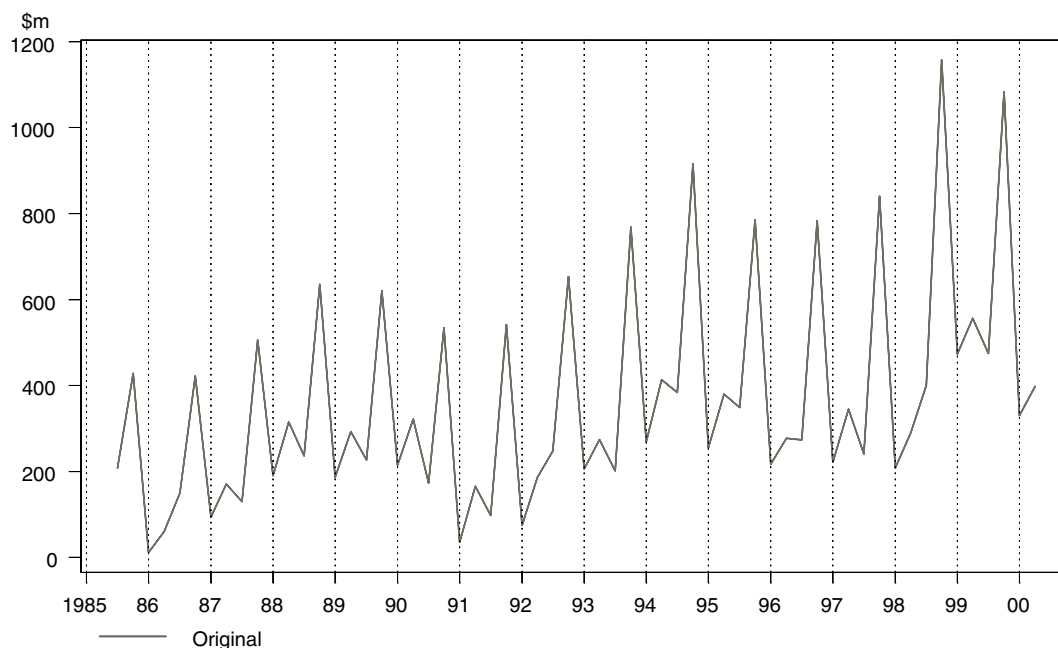
**Figure 2.6: Monthly New South Wales (NSW) Retail Department Stores - Original**

Wine production (Figure 2.7) follows a seasonal pattern reflecting the grape harvesting cycle. Despite some irregularity in the series seasonal highs can still be seen every December and seasonal lows each January.



**Figure 2.7: Monthly Wine Production - Original**

Figure 2.8 shows a quarterly series of retail profits. In quarterly series the annual pattern is often easier to see as it repeats every fourth observation. In this case there is a seasonal high in profits in the December quarter every year.



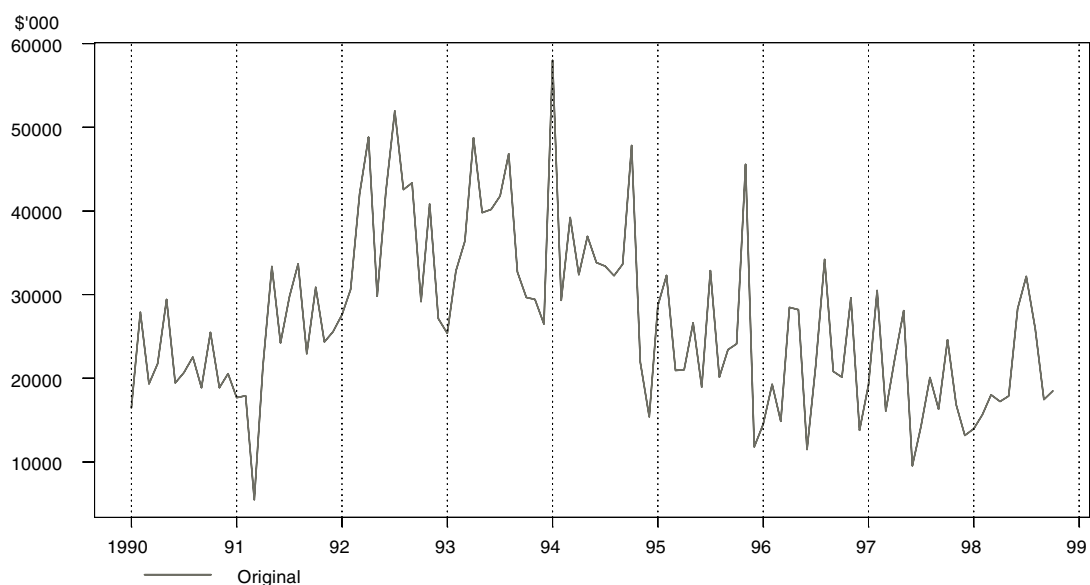
**Figure 2.8: Quarterly Retail Profits Before Tax - Original**

### 2.3 What is a residual/irregular?

The residual/irregular component is the remaining component of the series after the seasonal and trend components have been removed from the original data. It attempts to capture the remaining short term fluctuations in the series which are neither systematic nor predictable. The residual/irregular component of a time series may or may not be random. It can contain both random effects (white noise) or artifacts of non-sampling error, which are not necessarily random. Most time series contain some degree of volatility, causing original and seasonally adjusted values to oscillate around the general trend level. However, on occasions when the degree of irregularity is unusually large, the values can deviate from the trend by a large margin, resulting in an extreme value. Some examples of residual/irregular variation are natural events and industrial disputes.

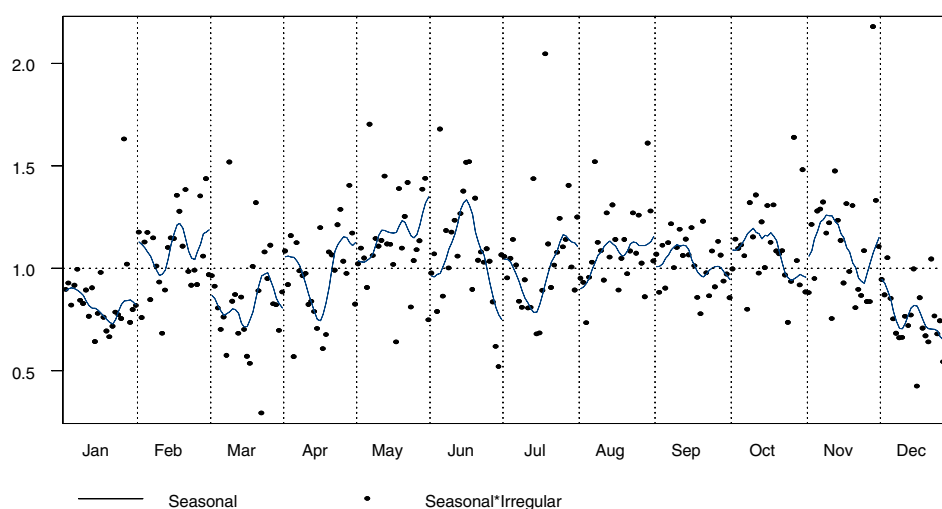
Irregularity in a series appears as short-term movements around the trend that do not follow any systematic or repeated pattern. In a highly irregular series the residuals can dominate movements causing the seasonality and trend to be difficult to identify. An example of a very irregular series is shown in Figure 2.9.





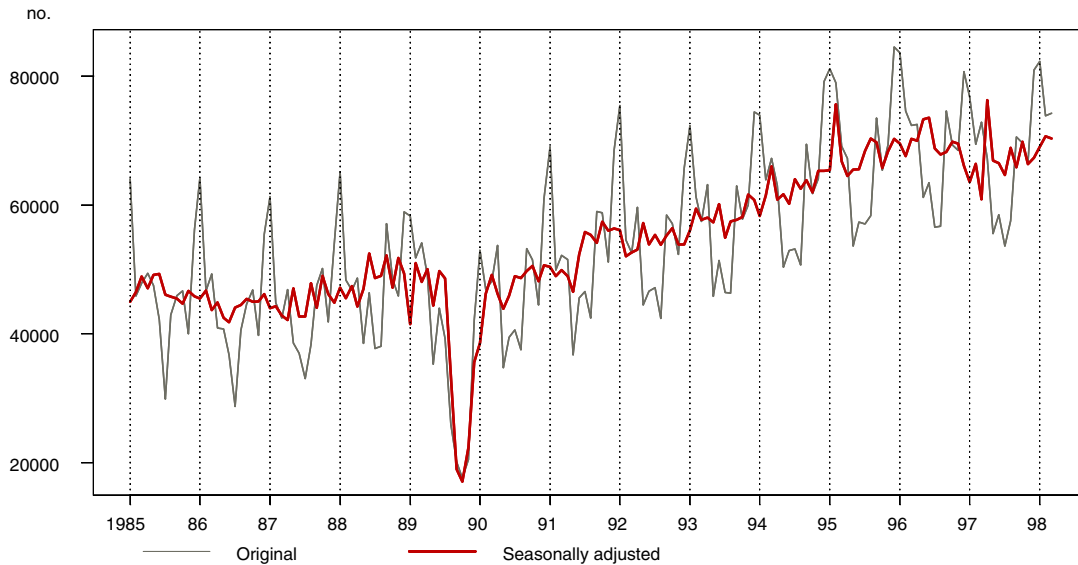
**Figure 2.9: Monthly Value of Building Approvals, ACT - Original**

In the building approvals series for the ACT it is difficult to determine whether the short-term movements are caused by seasonal or irregular influences. This distinction can be aided by the use of a seasonal/irregular (S-I) chart. S-I charts plot the seasonal and the seasonal-irregular factors of a series for specific months. The S-I chart for the ACT building approvals series is shown in Figure 2.10. The points in the chart are the seasonal-irregular factors while the solid line shows the seasonal factors calculated for the series. Large irregular influences in the series will cause the seasonal and irregular factors to differ markedly from the seasonal factors. From Figure 2.10 the seasonal-irregular factors can be observed to fluctuate widely around the seasonal factors indicating a dominating irregular component to the series.

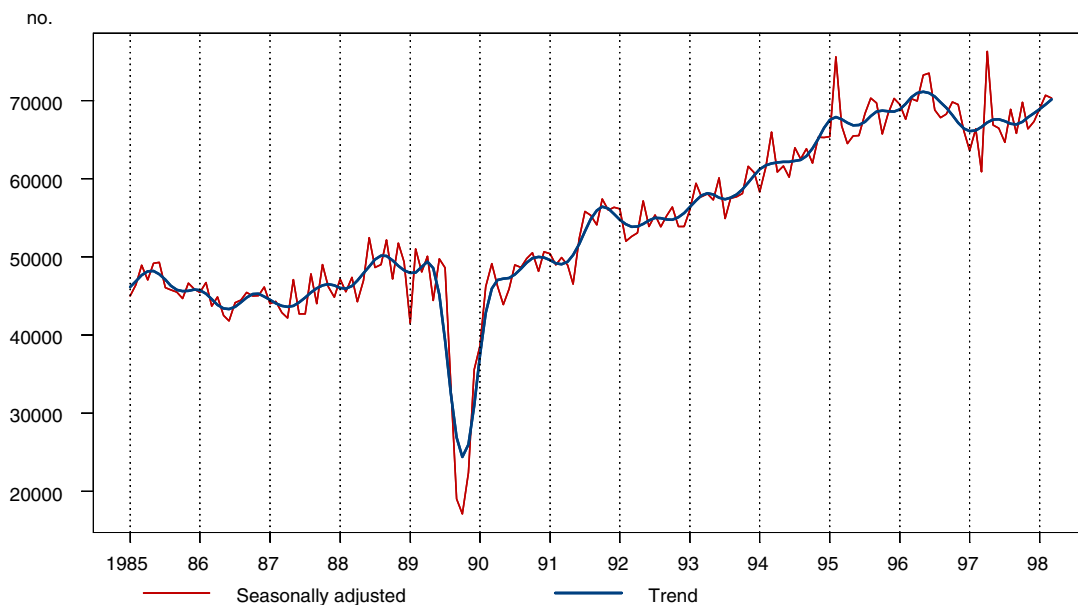


**Figure 2.10: Seasonal/Irregular (S-I) Chart - Value of Building Approvals, ACT**

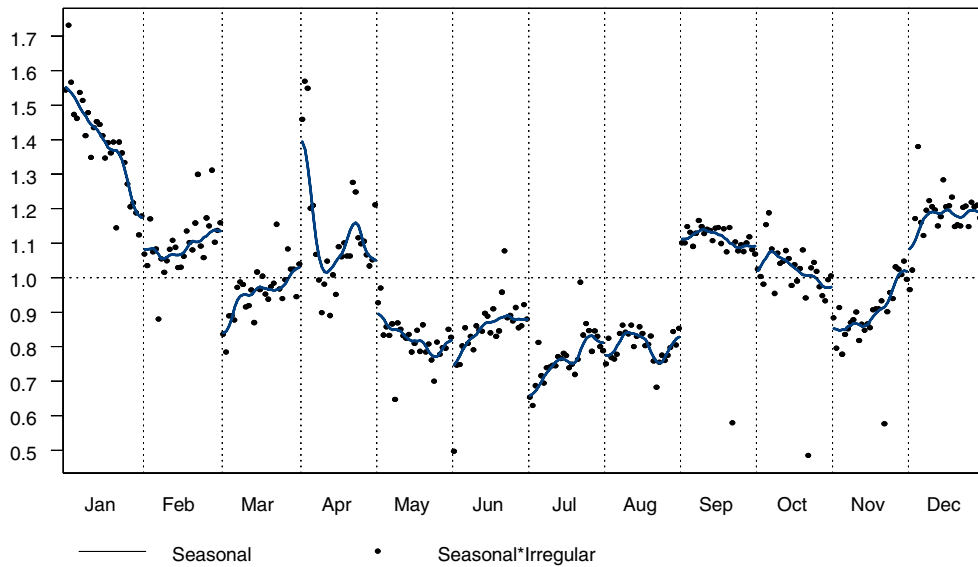
Figures 2.11-2.13 are of the original, seasonally adjusted, trend and seasonal-irregular series for Arrivals in Tasmania by Air. In September, October and November 1989 there is a large low irregular which was caused by a pilots' dispute. In particular, Figure 2.13 shows the S-I (seasonal-irregular) chart for this series and the extremes for September, October and November are particularly noticeable.



**Figure 2.11: Monthly Arrivals by Air in Tasmania : Original and Seasonally Adjusted**

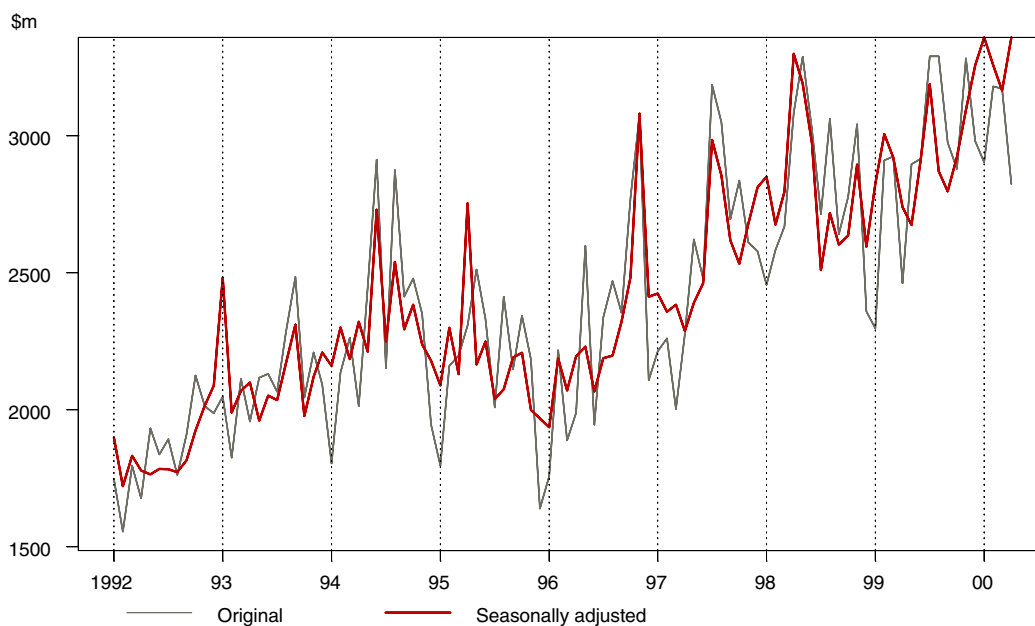


**Figure 2.12: Monthly Arrivals by Air in Tasmania - Seasonally Adjusted and Trend**

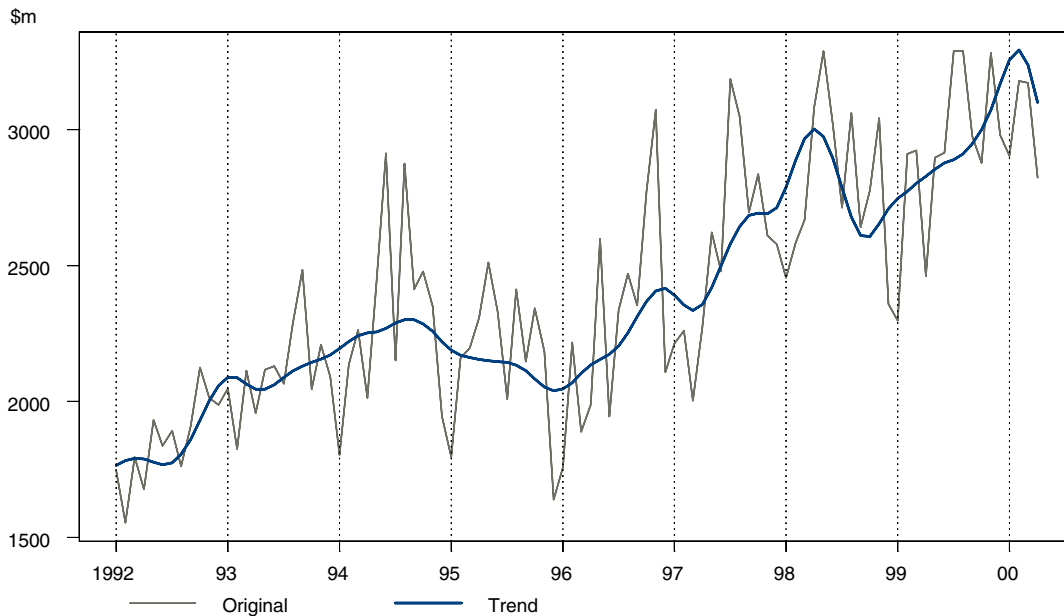


**Figure 2.13: S-I Chart - Monthly Arrivals by Air in Tasmania**

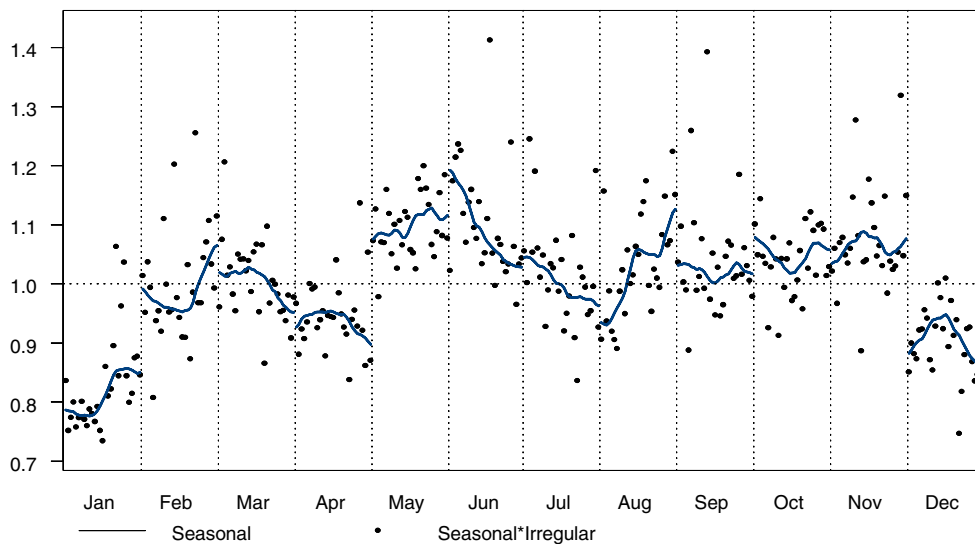
Graphs 2.14-2.16 show the original, seasonally adjusted, trend and seasonal-irregular series for the Buildings Approval series. In June 1994, April 1995 and November 1996 an effect on the series was caused by the approval to construct the Melbourne casino (1994), Sydney casino (1995), and Olympic stadium (1996) respectively. These once off approvals caused unusually high extremes in the series. In particular, Figure 2.16 shows the S-I (seasonal-irregular) chart for this series and highlights the extremes in June, April and November.



**Figure 2.14: Monthly Value of Building and Construction Approval - Original and Seasonally Adjusted**



**Figure 2.15: Monthly Value of Building and Construction Approval - Original and Trend**



**Figure 2.16: S-I Chart - Value of Building and Construction Approvals**

In the first example (Arrivals to Tasmania by Air, Graphs 2.11-2.13), the extremes are clearly visible in the original and adjusted, adjusted and trend, and SI charts. This is because the series is comprised of a strong seasonal and trend component but not much volatility. Hence, the adjusted series is fairly smooth, and as the irregular component is small the SI's are

relatively stable. This means that irregular values will stand out (ie. the SI points stay close to the seasonal factor line).

In the second example, (Buildings Approvals, Graphs 2.14-2.16), the extremes are almost impossible to detect in the original and adjusted, and SI charts. They are more visible in the adjusted and trend figure. This is because the series is not particularly seasonal, so the original and adjusted series are quite similar and quite irregular. Hence the adjusted series, which contains both the irregular and trend components, is substantially different from the trend. Being a volatile series also makes extremes difficult to detect in the SI charts because there is a lot of variability in the seasonal and irregulars.

## 2.4 Models for decomposing a series

To choose an appropriate decomposition model, the time series analyst will

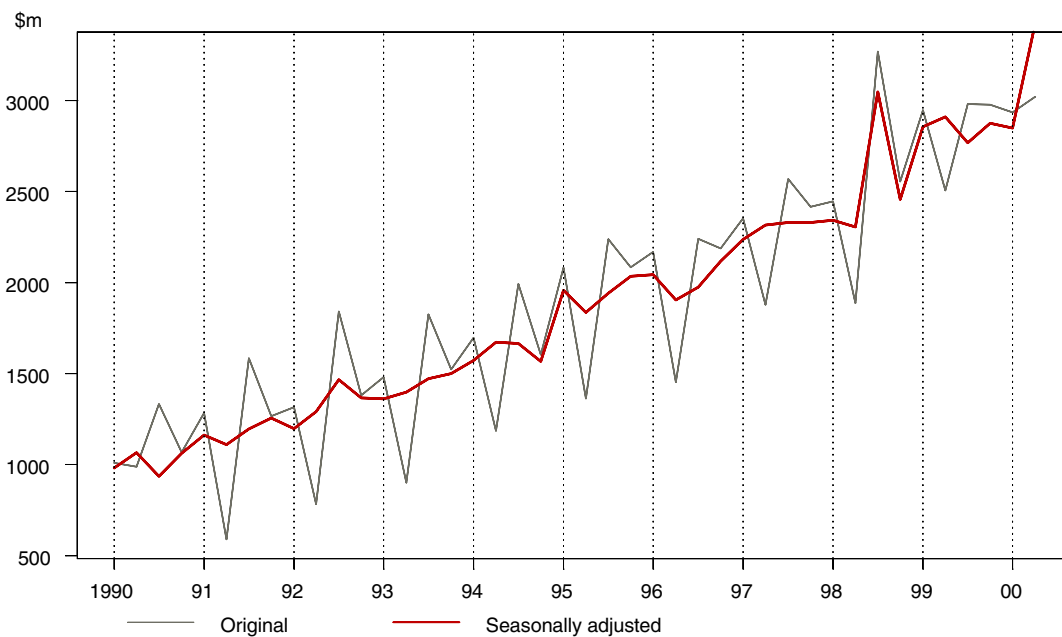
- a) examine the figure of the original series
- b) try a range of decomposition models and choose the one that yields the more stable seasonal component, especially at the end of the series

In practice, the original series is run through a seasonal adjustment program and then the seasonal component is checked for stable seasonality. The decomposition model which gives the more stable seasonal component is the more appropriate one to fit to the series.

There are two decomposition models which are commonly used: additive and multiplicative. The pseudo-additive model is occasionally used within the ABS.

### Additive Decomposition Model

The additive decomposition model assumes that the components of the series behave independently of each other. For example, an increase in the trend cycle will not cause an increase in the seasonal component. This model is used if the irregular/residual and the seasonal effects are independent of the trend behaviour, ie. if the seasonal effects are the same from year to year. Figure 2.17 shows a typically additive series. The trend of the series fluctuates yet the amplitude of the adjusted series (magnitude of the seasonal spikes) remain approximately the same, implying an additive model.



**Figure 2.17: General Government and Other Current Transfers to Other Sectors - Original and Seasonally Adjusted**

Mathematically, the original time series  $O_t$ , at time  $t$ , is considered to be a sum of the three components: the seasonal  $S_t$ , the trend  $T_t$ , and the residual/irregular  $I_t$ :

$$O_t = S_t + T_t + I_t$$

Therefore the seasonally adjusted data can be expressed by

$$\begin{aligned} SA_t &= O_t - S_t \\ &= T_t + I_t \end{aligned}$$

The constraints for an additive model are:

- the seasonal and irregular components are assumed to be independent of the trend component
- the seasonal component remains stable from year to year
- the seasonal fluctuation averages out to zero over the year:

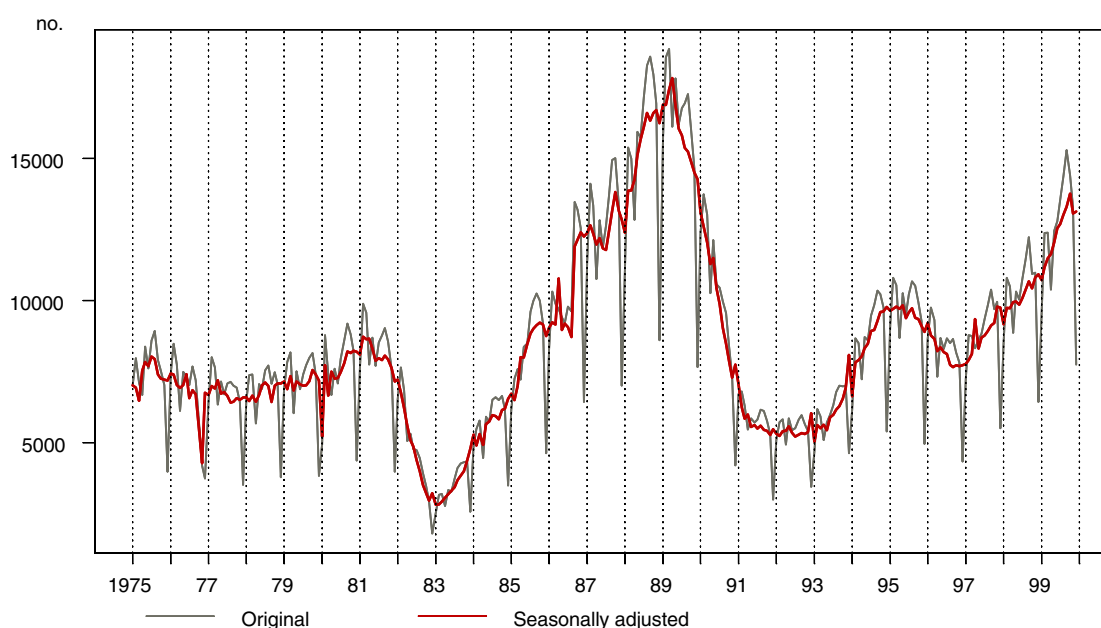
$$\sum S_{t+j} = 0$$

summing from  $j = 1$  to 12 (for monthly series) or 4 (for quarterly series).

## Multiplicative Decomposition Model

The multiplicative decomposition model assumes that the amplitude of the seasonal component is directly proportional to the level of the trend. As the trend increases, the amplitude of the seasonal component increases. Similarly if the trend decreases, the amplitude of the seasonal component decreases. It is assumed that the variance of the irregular component is directly proportional to the combined seasonal and trend cycle.

Most of the series analysed by the ABS display the characteristics of the multiplicative time series model. An example of a multiplicative time series is the NSW ANZ job advertisements series and is shown in Figure 2.18. As the trend of the series increases, the magnitude of the seasonal dips also increases, implying a multiplicative model.



**Figure 2.18: Monthly NSW ANZ Job Advertisements - Original and Seasonally Adjusted**

In this model, the original time series  $O_t$ , at time,  $t$ , is considered to be a multiple of the seasonal  $S_t$ , the trend  $T_t$ , and the irregular  $I_t$

$$O_t = S_t \times T_t \times I_t$$

The seasonally adjusted data can be expressed as

$$SA_t = O_t / S_t = T_t \times I_t$$

The multiplicative model requires the constraint

$$\sum S_{t+j} = n$$

where summing from  $j = 1, \dots, n$  with  $n = 12$  for monthly series or  $n = 4$  for quarterly series, requires that the seasonal component is fixed in relation to the trend from one year to the next.

The multiplicative model can also be written in an additive form by taking logarithms.

$$\log O_t = \log S_t + \log T_t + \log I_t$$

Clearly the multiplicative or log additive models cannot be used when the original series contains zero values. In this case a pseudo-additive decomposition model can be used.

### Pseudo-additive Decomposition Model

This pseudo-additive decomposition model combines elements of both the additive and the multiplicative relationships. It assumes that the seasonal pattern and the irregular element are both dependent on the level of the trend  $T_t$ , but independent of each other.

Removing the multiplicative link between the seasonal and irregular components is particularly useful if the series contains one or more zero values. This is because the additive relationship between the seasonal and irregular components allows either one of them to absorb the zero values without disturbing the trend behaviour of the series. In this way the method can accommodate one off zero values as well as regular zero values occurring as the result of a consistent seasonal pattern.

Algebraically this means the original data can be expressed as:

$$\begin{aligned} O_t &= T_t + T_t \times (S_t - 1) + T_t \times (I_t - 1) \\ &= T_t(S_t + I_t - 1) \end{aligned}$$

and the seasonally adjusted data is defined to be

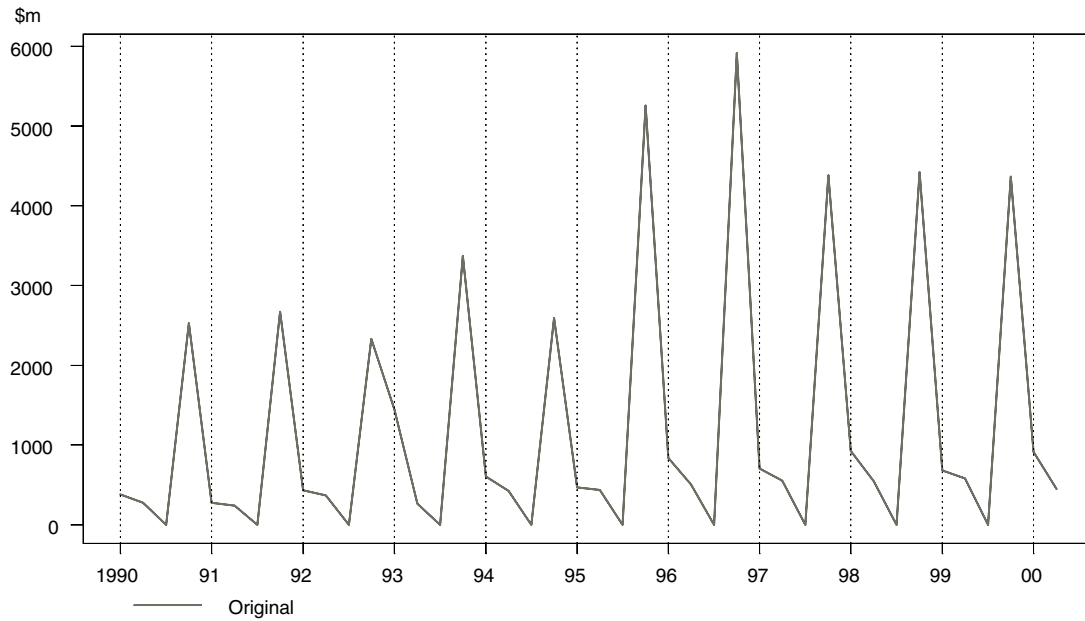
$$\begin{aligned} SA_t &= O_t - T_t \times (S_t - 1) \\ &= T_t \times I_t \end{aligned}$$

One is subtracted from the multiplicative seasonal pattern and the irregular element because each component is centered around one in the multiplicative case so that the level of the original series is not modified when the components are multiplied together. In the pseudo-additive relationship, the seasonal and irregular are both multiplicatively related to the trend and added to the trend to form the original data. If the seasonal pattern and irregular element were left centered around one, this would cause the level of the series to be grossly inflated. Instead, each of these components is modified to be centered around zero before being multiplied by the trend.

An example of a series that requires pseudo-additive adjustment is shown in Figure 2.19. This decomposition is used as cereal crops are only produced during certain months with the cereal



production being virtually zero for one quarter each year (See ABS (1999a) for a detailed discussion of the Australian Crop series).



**Figure 2.19: Quarterly gross value for the production of cereal crops - Original**

## Summary: Model comparisons

All three decomposition models are compared in Table 2.1. Column 2 illustrates the decompositions assuming that there are only three components: trend, seasonal and irregular. Column 3 assumes an additional component of trading day. Trading day is discussed in detail in Section 6.2.

	Three component model	Four component model
Components	$O_t$ = original $T_t$ = trend $S_t$ = seasonal $I_t$ = irregular	$O_t$ = original $T_t$ = trend $S_t$ = seasonal $I_t$ = irregular $TD_t$ = trading day
Multiplicative	$O_t = T_t \times S_t \times I_t$ $SA_t = O_t / S_t = T_t \times I_t$ $SI_t = O_t / T_t = S_t \times I_t$	$O_t = T_t \times S_t \times TD_t \times I_t$ $SA_t = O_t / (S_t \times TD_t) = T_t \times I_t$ $SI_t = O_t / (T_t \times TD_t) = S_t \times I_t$ $DI_t = TD_t \times I_t$
	$S_t, I_t, TD_t$ are distributed around 1	
Additive	$O_t = T_t + S_t + I_t$ $SA_t = O_t - S_t$ $SI_t = S_t + I_t$	$O_t = T_t + S_t + TD_t + I_t$ $SA_t = O_t - S_t - TD_t$ $SI_t = S_t + I_t$ $DI_t = TD_t + I_t$
	$S_t, I_t, TD_t$ are distributed around 0	
Pseudo-additive	$O_t = T_t \times (S_t + I_t - 1)$ $SA_t = O_t - T_t \times (S_t - 1)$ $SI_t = S_t + I_t - 1$	$O_t = T_t \times TD_t \times (S_t + I_t - 1)$ $SA_t = (O_t / TD_t) - T_t \times (S_t - 1)$ $SI_t = S_t + I_t - 1$
	$S_t, I_t, TD_t$ are distributed around 1	

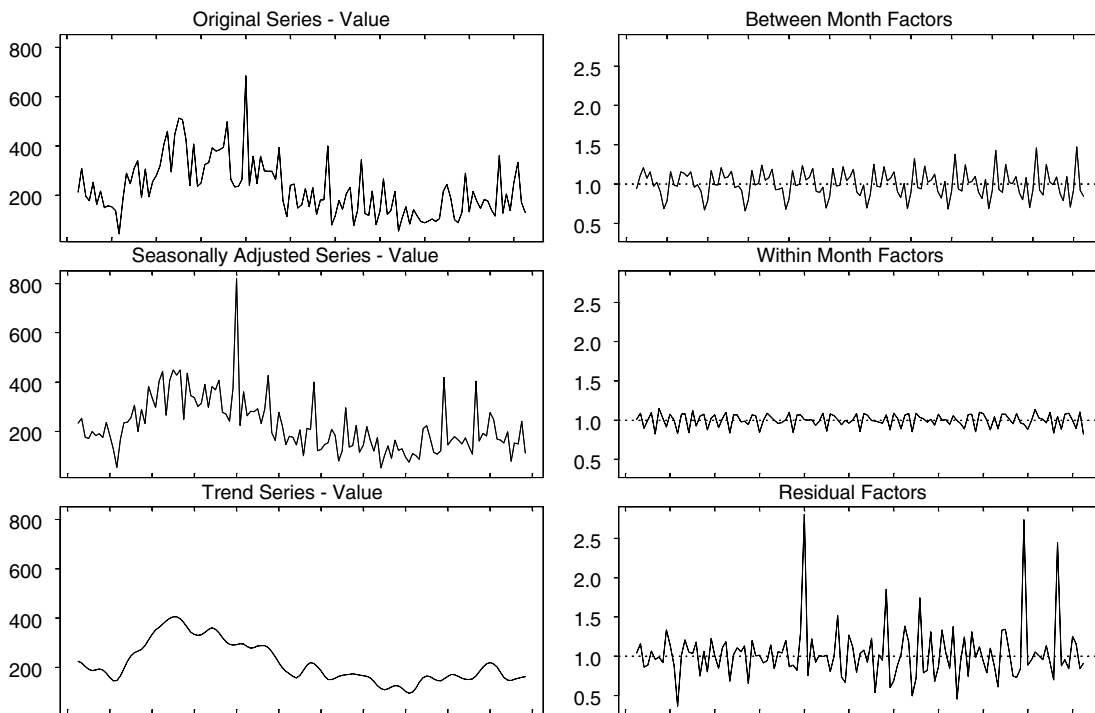
Table 2.1: Comparison of the multiplicative, additive and pseudo-additive decomposition models

## 2.5 Example: Decomposition of a series

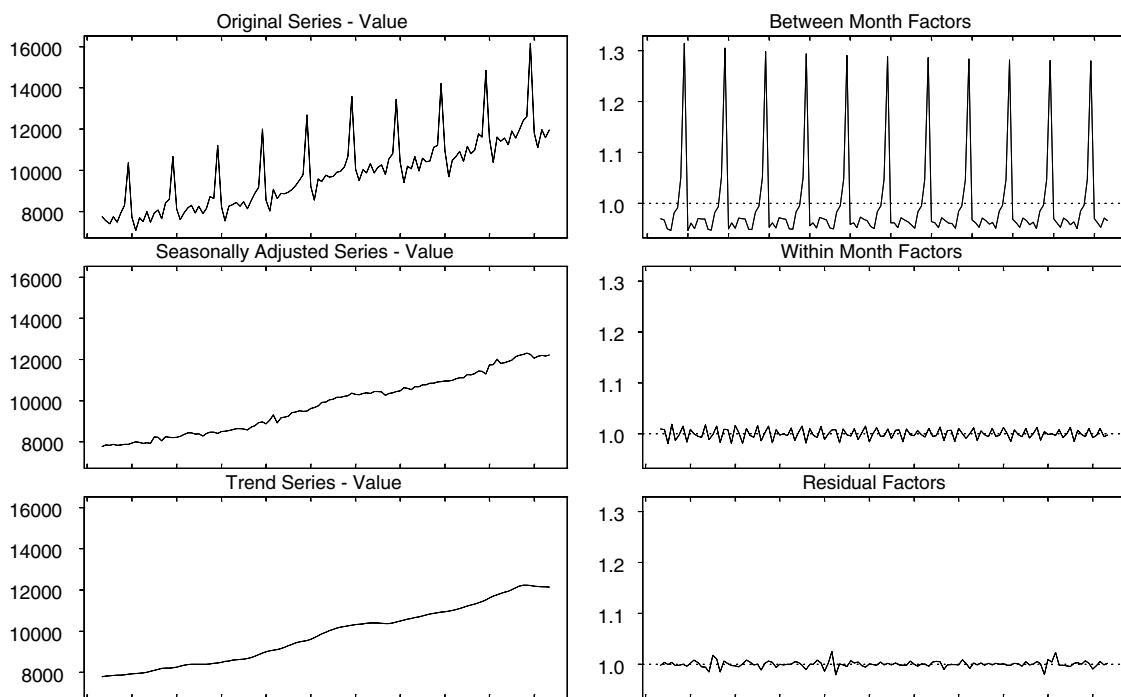
The Shiskin decomposition gives graphs of the original series, seasonally adjusted series, trend series, residual (irregular) factors and the between month (seasonal) and within month (trading day) factors that are combined to form the combined adjustment factors. The residual (irregular) factors are found by dividing the seasonally adjusted series by the trend series.

Figure 2.20 shows a Shiskin decomposition for the number of ACT Building Approvals. This series was adjusted using a multiplicative model.

Figure 2.21 shows a Shiskin decomposition for the Australian Retail series. Note that in contrast with the Building Approvals series, the variation of the trading day factors, is greater than for the Building Approvals series. This indicates that, for the retail series, activity is more dependent on the number of different types of days of the week in any given month.



**Figure 2.20: Shiskin decomposition for the Number of ACT Dwelling Approvals, April 1990 to April 2000**



**Figure 2.21: Shiskin decomposition for Australian Total Retail Turnover, May 1990 to May 2000**

## 3 Seasonal adjustment philosophies

The two main philosophies for seasonal adjustment are the model based method and the filter based method.

In this section, an outline of the basis of the two philosophies is given. For more details see Sutcliffe (1999). Chapters 4 and 5 describe in more depth the technical basis of the filter based method commonly applied in the ABS. In general these notes concentrate on the filter based method.

### 3.1 Model based methods

The model based approach requires the components of an original time series, such as the trend, seasonal and irregular to be modelled separately. Alternatively, the original series could be modelled and from that model, the trend, seasonal and irregular component models can be derived.

Models for each of the components are usually estimated simultaneously using the Kalman filter or related techniques (see Harvey (1990) for further discussion of these methods). Each component is assigned its own model which determines the broad spectral characteristics (i.e. the cycles which are present in the original series) of that component. Each component is also assigned its own innovation variance which determines the relative strength of each component. The contributions of the components to the overall model are usually estimated using maximum likelihood methods.

Model based methods assume the irregular component is “white noise” i.e. the irregular has no structure, zero mean and a constant variance.

Major computational differences between the various methods in the model based family are usually due to model specification. In some cases, the components are modelled directly. In other cases, the original time series is modelled and the component models are decomposed from that model. The decomposition of the component series is not generally unique. STAMP and TRAMO/SEATS are two software packages that use model based methods.

### 3.2 Filter based methods

This method applies filters (moving averages) to decompose the time series into a trend, seasonal and irregular component. Typically, symmetric linear filters are applied to the middle of the series, and asymmetric linear filters are applied to the ends of the series.

The general philosophy is to remove a range of cycles (i.e. band of spectral power) from the original data. To produce a seasonally adjusted series, cycles around 12, 6, 4, 3, 2.4 and 2 cycles per year in a monthly time series, need to be removed. These cycles are referred to as the seasonal harmonics. They correspond to 1 (monthly), 2, 3, 4 (quarterly), 5 and 6 (half-yearly) occurrences per year in a monthly time series. The trend is always required to be

‘smooth’. This can be done by removing the shorter term cycles in the seasonally adjusted series.

Filter based methods, because of the criteria for the seasonal harmonics and ‘smooth’ trend, result in an irregular component that does not necessarily display white noise characteristics. No explicit model of the components is required for filter based methods. Features of the adjustment process derive directly from the properties of the filters used.

Filter based methods can be thought of as X11 style methods. Methods include X11, X11ARIMA88, X12-ARIMA, STL, SABL and SEASABS.

X11 style methods are generally iterative. The trend, seasonal and irregular components are estimated separately in a computation loop. For example, in X11 a preliminary estimate of the trend is computed. The original data is then detrended to give an estimate of the combined seasonal and irregular component. This is then smoothed month by month to yield a separate estimate of the seasonal component. The computations continue through three iterations.

Major computational differences between the various methods in the X11 family are usually due to different techniques used at the ends of the time series. For example, some methods use asymmetric filters at the ends while others extend the series using ARIMA models and apply symmetric filters to the extended series.

### 3.3 Comparison of the two approaches

Model based approaches allow for the stochastic properties of the series under analysis. In other words, model based approaches tailor the filter weights based on the stochastic nature of the series. Checks for model identification are performed for the models capability for describing the behaviour of the series. Statistical inferences are available for the estimates if the assumption that the irregular component is white noise is true.

On the other hand, filter based approaches are less dependent on the stochastic properties of the series under analysis. It is up to the analysts to select the most adequate filter given from a limited collection for a particular series. No rigorous checks may be performed, and exact measures of precision or statistical inference are not available. Therefore, the building of a confidence interval around the estimate is not possible.

The following table contrasts the various elements of the two approaches to seasonal adjustment.

<b>Model based</b>	<b>Filter (X11) based</b>
<p>Irregular component is defined as white noise.</p> <p>Model fitting and identification is done first before decomposition. The correlation structure of the components are well described by the model. The noise is re-partitioned to component noises (innovations).</p>	<p>Irregular component is defined as residual or what is left in the original once the trend and seasonal components have been removed. It is specified last after trend and seasonal.</p>
<p>Trend component has a random walk look (plus cycle can also be modeled with trend).</p>	<p>Trend component defined as cycles longer than a certain length.</p>
<p>Seasonal component is a stochastic process with its own noise element, and allows evolution.</p>	<p>Seasonal component defined as the band around the seasonal harmonics. The 'smooth' seasonal factors are a particular characteristic of this method.</p>
<p>Theoretically, leaves some power at the seasonal frequencies in the irregular component and also in the trend. That is, as white noise is left over, some of the cycles of the seasonal and trend are still in the irregular component.</p>	<p>Aims to remove all spectral power at the seasonal frequencies</p>
<p>Structure of trend, seasonal and irregular for a given model is defined by the variance of the innovations for each component model.</p>	<p>Size of the bands around the seasonal harmonics are determined by the seasonal filter length, and the frequency cut off for the trend is determined by the length of the trend filter.</p>
<p>Simultaneous estimates are obtained for the trend, seasonal and irregular components.</p>	<p>Computations are iterative to provide estimates for the trend, seasonal and irregular.</p>

**Table 3.1: Comparison of two different model philosophies**

## 4 Frequency domain analysis

Two common filter based approaches to analysing time series are time domain analysis and frequency domain analysis. Time domain analysis uses parametric models to describe the time series. Frequency domain analysis describes a time series using the cyclical behavior at various frequencies.

The ABS uses a filter-based approach to analysing time series. This chapter concentrates on frequency domain analysis which allows filters to be represented in terms of cycles. In particular, the concepts of spectral analysis, gain functions and phase shifts are introduced. Spectral analysis lets us investigate the cycles which are present in a series. Properties of linear filters can then be investigated using gain and phase functions. The gain and phase summarise what happens to particular cycles, such as business cycles, seasonal cycles and noise, when linear filters are applied.

Some preliminary concepts are needed. The following sections briefly discuss the autocovariance, fourier analysis and spectral analysis of a series. For more information on these topics see, for example, Box and Jenkins (1970) and Wei (1993). The gain and phase functions are then discussed.

### 4.1 Autocovariance

Consider a stationary process  $\{Y_t\}$ . This has a constant mean  $E(Y_t) = \mu$ , constant variance  $Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$ , and covariances  $Cov(Y_t, Y_s)$  which are functions of the difference  $|t - s|$ . The autocovariance function,  $\gamma_k$ , is defined as the covariance between  $Y_t$  and  $Y_{t+k}$

$$\begin{aligned}\gamma_k &= Cov(Y_t, Y_{t+k}) \\ &= E(Y_t - \mu)(Y_{t+k} - \mu)\end{aligned}\quad (4.1)$$

For a given sequence of autocovariances  $\gamma_k, k = 0, \pm 1, \pm 2, \dots$ , the autocovariance generating function is defined as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \quad (4.2)$$

Note the use here of the backshift operator  $B$ , which is defined as

$$B^k x_t = x_{t-k}$$

Consider  $Z_t$ , which has a zero mean,  $Z_t = Y_t - \mu$ . Then  $Z_t$  can be expressed as a linear combination of random variables

$$\begin{aligned}Z_t &= \theta(B)a_t \\ &= \sum_{j=0}^{\infty} \theta_j a_{t-j}\end{aligned}\quad (4.3)$$

where  $\theta_j$  are coefficients and  $a_t \sim N(0, \sigma^2)$ . Then

$$\text{Var}(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \theta_j^2$$

and

$$\begin{aligned} \gamma_k &= E(Z_t - 0)(Z_{t+k} - 0) \\ &= E(Z_t Z_{t+k}) \\ &= E\left(\sum_{j=0}^{\infty} \theta_j a_{t-j} \sum_{i=0}^{\infty} \theta_i a_{t+k-i}\right) \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+k} \end{aligned}$$

So

$$\begin{aligned} \gamma(B) &= \sum_{k=-\infty}^{\infty} \gamma_k B^k \\ &= \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \theta_i \theta_{i+k} B^k \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_i \theta_j B^{j-i} \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \theta_j B^j \sum_{i=0}^{\infty} \theta_i B^{-i} \\ &= \sigma_a^2 \theta(B) \theta(B^{-1}) \end{aligned} \tag{4.4}$$

where  $j = i + k$  and  $\theta_j = 0$  for  $j < 0$ . This can be applied to any linear combination of weights of the form given in (4.3).

## 4.2 Fourier analysis

Fourier analysis is a form of frequency domain analysis which can be used to represent any periodic function as a series of harmonically related sinusoids.

A fourier transform of a discrete time function  $Y_t$ , such that  $Y_t = 0$  for  $|t| > M$  is defined as

$$f(\omega) = \sum_{k=-\infty}^{\infty} Y_k e^{-i\omega k} \quad \text{where} \quad \pi \leq \omega \leq \pi \tag{4.5}$$

The spectrum of a stationary process is the fourier transform of the absolutely summable autocovariance function of the process.



### 4.3 Spectral analysis

If  $Y_t$  is a stationary process with an absolutely summable autocovariance sequence  $\gamma_k$ , this means that the sum of the process  $\gamma_k$  exists and is of finite duration. This means that the fourier transform exists.

Substituting  $\gamma_k$  into (4.5) gives the spectrum of a series  $Y_t$  with autocovariance  $\gamma_k$

$$\begin{aligned} f(\omega) &= \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k} \\ &= \sum_{k=-\infty}^{\infty} \gamma_k (\cos(-\omega k) + i \sin(-\omega k)) \\ &= \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\omega k) \end{aligned} \quad (4.6)$$

where  $\gamma_k = \gamma_{-k}$  and  $0 \leq \omega \leq \pi$ . Also  $\sin(0) = 0$ ,  $\sin(\omega(-k)) = -\sin(\omega k)$ ,  $\cos(\omega(-k)) = \cos(\omega k)$  and  $e^{-i\omega t} = \cos(-\omega k) + i \sin(-\omega k)$ .

Recall that the autocovariance generating function is defined by (4.2) as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$$

By equating (4.6) and (4.2) the spectra and the autocovariance generating function can be shown to be related by

$$f(\omega) = \gamma(e^{-i\omega}) \quad (4.7)$$

This means that once the autocovariance generating function is known for a process the above relationship can then be used to derive the spectrum for that process.

As an aside, the sequence  $\gamma_k$  can be recovered through the inverse Fourier transform defined as

$$\gamma_k = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega \quad (4.8)$$

By setting  $k = 0$  this gives the variance for the process  $Y_t$  as

$$\begin{aligned} \text{Var}(Y_t) &= \gamma_0 \\ &= \int_{-\pi}^{\pi} f(\omega) d\omega \end{aligned}$$

This shows that the spectrum  $f(\omega)$  can be interpreted as the decomposition of the variance process. A peak in the spectrum indicates an important contribution to the variance from the components at frequencies in the corresponding interval.

## 4.4 Gain functions

Generally, economic data is made up of a wide range of cycles such as business cycles, seasonal cycles and noise. Gain functions can be used to look at the effect of a linear filter at a given frequency on the amplitude of a cycle in a given series. In other words, it shows what happens to the amplitude of a given input cycle on the application of a moving average.

Theoretically, the properties of a linear filter can be studied by analysing the fourier transform defined by (4.5) as

$$f(\omega) = \sum_{j=-m_1}^{m_2} w_j e^{i\omega j}$$

where  $w_{-m_1}, \dots, w_{m_2}$  are the weights of the filter and  $\omega$  is the frequency with  $-\pi \leq \omega \leq \pi$ . The frequency response function can be written in terms of the contribution of a sine and cosine functions as follows

$$f(\omega) = C(\omega) + iS(\omega)$$

where

$$C(\omega) = \sum_{j=-m_1}^{m_2} w_j \cos(\omega j)$$

$$S(\omega) = \sum_{j=-m_1}^{m_2} w_j \sin(\omega j)$$

The gain function of a filter with weights  $w_{-m_1}, \dots, w_{m_2}$  is then defined by

$$G(\omega) = (C(\omega)^2 + S(\omega)^2)^{\frac{1}{2}} \quad (4.9)$$

For a symmetric moving average of length  $2m + 1$  where  $m_1 = m_2 = m$  and  $w_{-k} = w_k$  the gain function simplifies to

$$G(\omega) = |w_0 + 2 \sum_{k=1}^m w_k \cos(\omega k)| \quad (4.10)$$

The gain can be defined to be both positive and negative. If this is the case then a gain of -1 means that peaks have become troughs and troughs have become peaks. The size of the magnitude from a trough to a peak is still the same. The gain is symmetric over the frequencies  $-\pi$  to  $\pi$ . For this reason, the gain is typically plotted from 0 to  $\pi$ .

If a process  $Y_t$  has spectrum  $S_Y(\omega)$  then the spectrum of the filtered data,  $S_T(\omega)$ , is given by

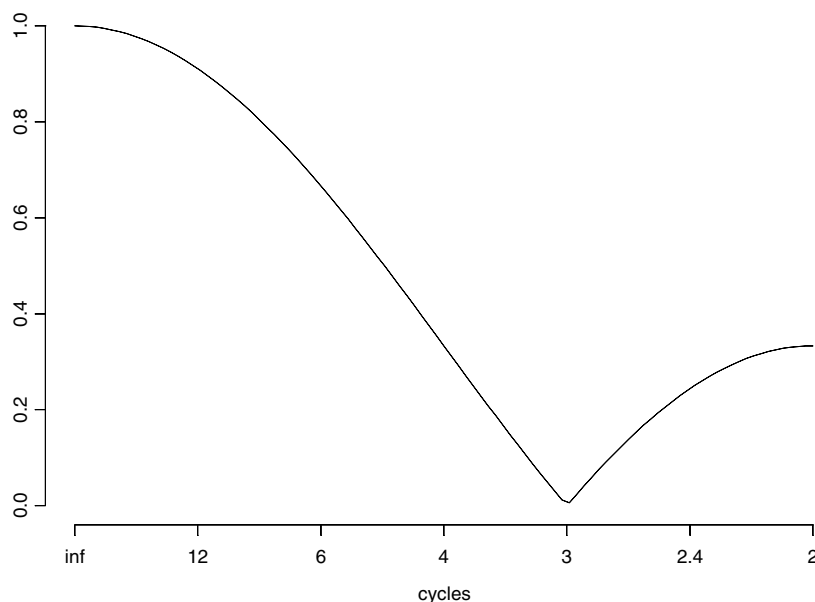
$$S_T(\omega) = G^2(\omega) \times S_Y(\omega) \quad (4.11)$$

Note: Spectra are useful for analysing stationary time series only. Spectrum for non-stationary time series may be better analysed by first transforming the data to obtain a stationary series and then obtaining the spectrum.

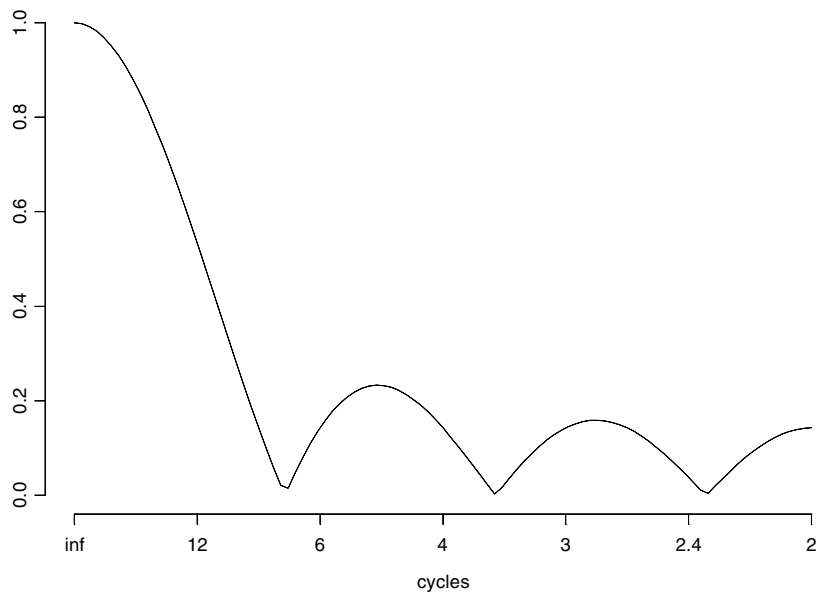
Examples of gain functions which have been defined to be all positive are shown in Figures 4.1-4.3. All of the gain functions have been plotted showing the periods along the x-axis. To convert frequencies ( $\omega$ ) to periods ( $p$ ) the following relationship can be used:

$$\omega = (2 \times \pi)/p$$

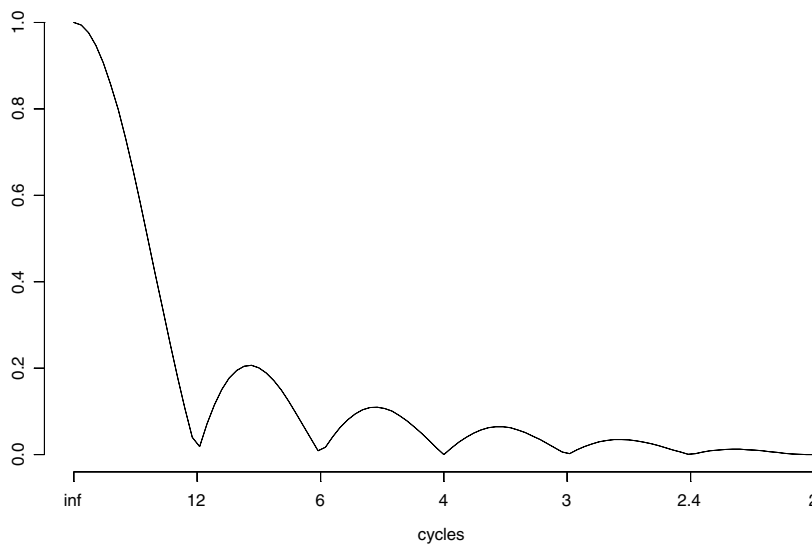
For example, the seasonal frequencies occur at 1, 2, 3, 4, 5 and 6 cycles per year, these correspond to 12, 6, 4, 3, 2.4 and 2 period cycles. In frequencies these are given by  $(2 \times \pi)/12 = \pi/6$  (12 period cycle),  $(2 \times \pi)/6 = \pi/3$  (6 period cycle), through to  $(2 \times \pi)/2 = \pi$  (2 period cycle). All spectral plots are labeled with periods along the x-axis. Figure 4.1 shows a gain function for a simple three term moving average. This illustrates that the application of this filter to an input series removes cycles of approximately length 3 periods from the input series. Figure 4.2 shows a gain function for a simple seven term moving average. This filter would remove cycles of approximate length 7, 3.5 and 2.3 periods.



**Figure 4.1: Moving average with weights (1/3,1/3,1/3)**



**Figure 4.2** Moving average with weights  $(1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7)$



**Figure 4.3:** Moving average weights  $(1/24, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/24)$

Figure 4.3 illustrates an example of a filter which would remove constant seasonal cycles from an input series. The seasonal frequencies occur at 12, 6, 4, 3, 2.4 and 2 months. This filter would therefore be good to use for removing seasonal effects in an input series. Cycles which are indicative of an irregular component (cycles less than 3 periods) would be significantly reduced by the application of this filter.

## 4.5 Phase shift

Phase shifting occurs when the timing of turning points is distorted. The turning points will then occur either earlier or later than in the original series. Symmetric moving averages, where the result is centrally placed, do not cause time phase shifting. An important property of the trend moving averages is that the moving average retains the time phase and hence the timing of any turning points.

If the result of the moving average were placed off-centre, such as on the last or first term of the moving average, time phase shifting would occur. Indeed the off-centred simple 12 term moving average is said to cause a shift in time phase of about 6 months. The ABS uses only odd length trend moving averages in the middle of the series so that the result will be centred on a value, rather than between values.

A phase shift is the time shift between the filtered cycle and the unfiltered cycle. In the case of the definitions below a positive phase shift means that the filtered cycle is shifted backwards and a negative phase shift it is shifted forwards. If the gain is defined to be strictly positive then the phase shift,  $\Phi(\omega)$  can be calculated by the following set of equations:

$$\begin{aligned}
 C > 0 \text{ then } \Phi(\omega) &= \arctan(S/C) \\
 C < 0 \text{ and } S \geq 0 \text{ then } \Phi(\omega) &= \arctan(S/C) + \pi \\
 C < 0 \text{ and } S < 0 \text{ then } \Phi(\omega) &= \arctan(S/C) - \pi \\
 C = 0 \text{ and } S > 0 \text{ then } \Phi(\omega) &= \pi/2 \\
 C = 0 \text{ and } S < 0 \text{ then } \Phi(\omega) &= -\pi/2 \\
 C = 0 \text{ and } S = 0 \text{ then } \Phi(\omega) &\text{ is undefined}
 \end{aligned} \tag{4.12}$$

If the moving average is symmetric the phase is 0 or  $\pm\pi$ . Generally, for filters designed for smoothing (trend filter), a phase shift of  $\pm\pi$  will be at frequencies where the amplitude has been almost completely eliminated and will therefore not have a great impact on the series.

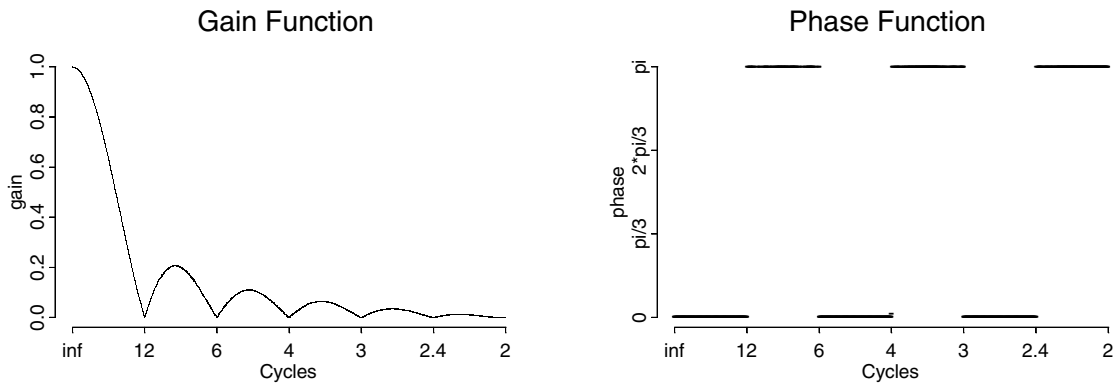
The phase is also given by the alternate set of equations:

$$\begin{aligned}
 C \neq 0 \text{ then } \Phi(\omega) &= \arctan(S/C) \\
 C = 0 \text{ and } S > 0 \text{ then } \Phi(\omega) &= \pi/2 \\
 C = 0 \text{ and } S < 0 \text{ then } \Phi(\omega) &= -\pi/2 \\
 C = 0 \text{ and } S = 0 \text{ then } \Phi(\omega) &\text{ is undefined.}
 \end{aligned} \tag{4.13}$$

While most authors prefer the first definition of the phase function (4.12) there is support for the second definition (4.13). For example Chatfield (1996) states:

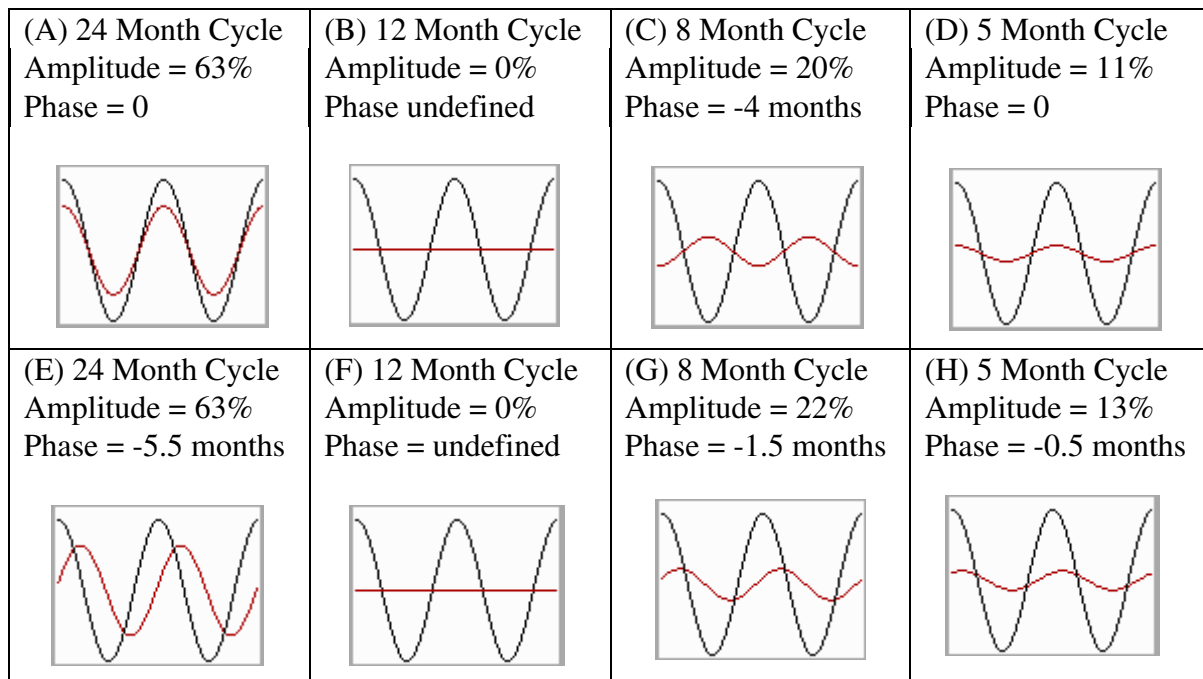
*"In fact there are physical reasons why engineers prefer to plot the phase as a continuous unconstrained function allowing  $G(\omega)$  to be positive or negative..."*

These two different definitions are possible since a half cycle phase shift is equivalent to changing peaks into troughs and vice versa, this can also be achieved by a gain of minus one. The phase can be computed in periods by using  $\Phi(\omega)/\omega$ .



**Figure 4.4: Gain for 2x12 symmetric moving average and a phase for 2x12 symmetric moving average**

Some examples of the effects of cycles of different lengths for the 2x12 symmetric moving average are shown in the graphs below labeled A,B,C and D. Remember that gain is the effect on the amplitude and phase the effect on the timing. Graphs E-H show the same cycles with a off-center 12-term asymmetric moving average.



**Figure 4.5: Gain and phase for different length cycles after applying a 2x12 symmetric moving average**

## 5 Filtering - the empirical approach

In Chapter 4, the theory behind the filtering approach to seasonal adjustment was discussed. The concept of linear filters, gain functions, phase functions and spectra was introduced. This chapter, describes how this theory is put into practice within the ABS.

### 5.1 General framework

Moving averages or linear filters successively average a shifting time span of data and have long been used to estimate the components of decomposition models. Depending on their weighting structure they can also be used to eliminate cycles of certain length. Moving averages work on the assumption that the trend and seasonal components are stochastic (the future is only partly determined by past values) and not deterministic.

Moving averages are often applied to smooth out the irregularities in a series. This smoothed series can be considered to have been derived by running an input series through a process which filters out irregularities. Consequently, a moving average is often referred to as a filter.

As an example, a moving average of the original data can be used to smooth the data and obtain an initial estimate of the trend. This estimate of the trend can then be removed from the original data to get the combined seasonal and irregular components. These can then be smoothed, a month or quarter at a time, with another moving average to obtain an estimate of the seasonal pattern. This seasonal pattern can in turn be removed from the original data to leave an estimate of the seasonally adjusted data which can be smoothed to get an improved estimate of trend.

This is the basic process undertaken by Time Series Analysis to produce seasonally adjusted and trend estimates. The moving averages used in this estimation process have been selected for their properties, such as highlighting the appropriate frequencies and not changing the level of the series or the timing of turning points. This chapter discusses both the trend and the seasonal moving averages used within the seasonal adjustment process at the ABS.

A variety of other techniques can be applied to analysing time series. These include simple averages, transformations, differencing, regression, spectral analysis, curve-fitting and looking at year-apart and year-to-date movements. These techniques should be applied with care as some of these techniques are deficient or have results that can be easily misinterpreted.

#### Definition:

Define a set of weights of length  $m_1 + m_2 + 1$  as:

$$w_{-m_1}, w_{-(m_1-1)}, \dots, w_{-1}, w_0, w_1, \dots, w_{m_2-1}, w_{m_2} \quad (5.1)$$

Note: a symmetric set of weights has  $m_1 = m_2$  and  $w_j = w_{-j}$ .

**Definition:**

A filtered value at time  $t$  can be calculated by

$$f(y_t) = \sum_{j=-m_1}^{m_2} w_j Y_{t+j} \quad (5.2)$$

**5.2 Example: Application of a linear filter**

An example of a first difference linear filter is given by

$$f(y_t) = Y_t - Y_{t-1} \quad (5.3)$$

Then, in this case (5.1) and (5.2) would be

$$m_1 = 1, m_2 = 0, w_0 = 1 \text{ and } w_{-1} = -1$$

To investigate the gain function properties of this filter consider the two components given in (4.9)

$$C(\omega) = 1 - \cos(-\omega) = 1 - \cos(\omega)$$

$$S(\omega) = -\sin(-\omega) = \sin(\omega)$$

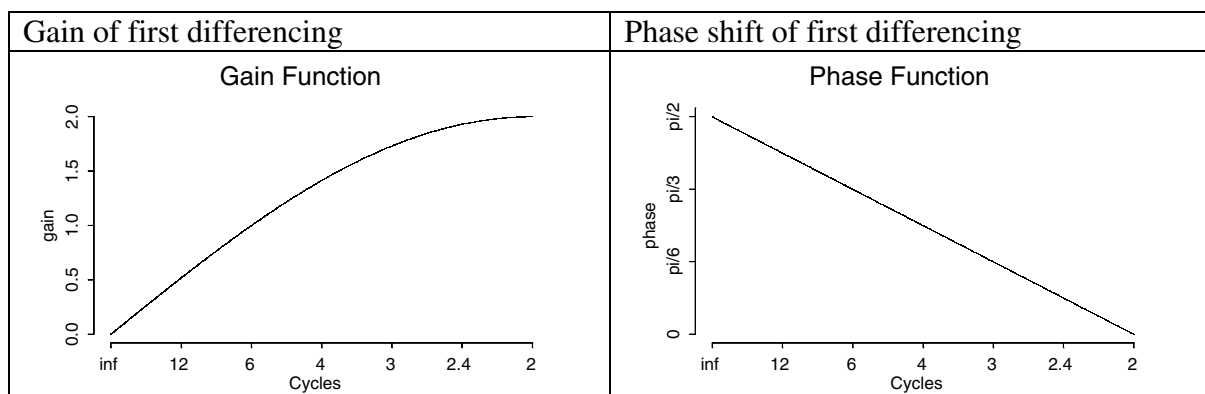
Since  $C(\omega) > 0$  for all  $\omega$  the gain function is defined to be

$$G(\omega) = \sqrt{(1 - \cos(\omega))^2 + (\sin(\omega))^2}$$

and the phase function is

$$\Phi(\omega) = \arctan\left(\frac{\sin(\omega)}{1 - \cos(\omega)}\right) = (\pi - \omega)/2$$

Figure 5.1 below shows the gain and phase for the first difference operator.



**Figure 5.1: Gain and phase for the first difference operator**



For example, for a 10 period cycle

$$\omega = (2\pi)/p = \pi/5, \quad \text{and}$$

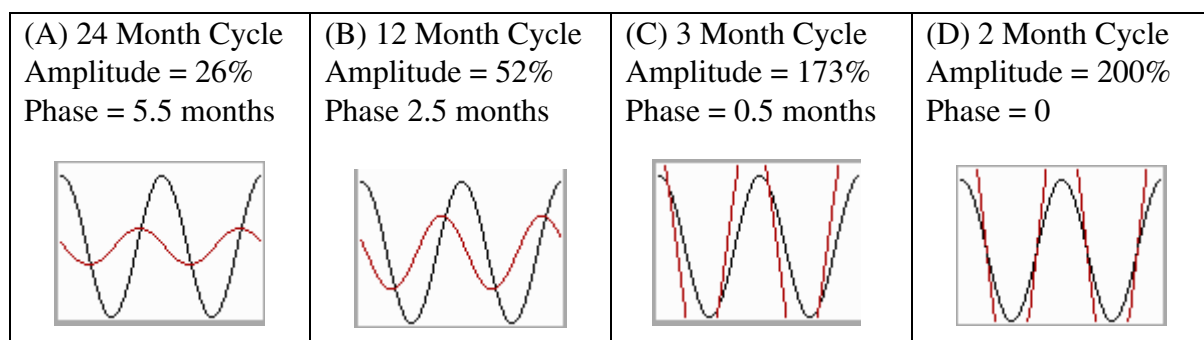
$$G(\pi/5) = \sqrt{(1 - \cos(\pi/5))^2 + (\sin(\pi/5))^2} = 0.62, \quad \text{then}$$

$$\Phi(\pi/5) = (\pi - \pi/5)/2$$

which is given in periods by

$$\Phi(\pi/5)/(\pi/5) = 2 \text{ periods}$$

Results for a range of input cycles are shown in Figure 5.2.



**Figure 5.2** The effect of applying a first difference on various cycles of an input series

It can be seen above that first differencing completely distorts the spectral characteristics of a time series. For example the short term movements (i.e. those that follow a 2 month cycle) get amplified by about 200%. Its only redeeming feature is that it removes “long term growth”. Publishing such movements has been fundamental for about the last 100 years and, in addition, sampling theory for time series is based on minimising the standard errors of the movements in a time series. Of course even a “bad” measure can be improved by developing a “memory map” of historical behaviour. For example, experience suggests that a percentage movement in seasonally adjusted Australian Total Retail turnover of 1% is a moderate movement while 4% would be considered to be a very large movement.

The relationship between the gain and phase function can be readily obtained and understood by analysing the effect of the linear filter given in (5.2) on the process

$$Y_t = \cos(\omega t), -\pi < \omega < \pi$$

The filtered data is given by

$$\begin{aligned} f(\cos y_t) &= \sum_{k=-m_1}^{m_2} w_k \cos(\omega(t+k)) \\ &= \sum_{k=-m_1}^{m_2} w_k (\cos(\omega t) \cos(\omega k) - \sin(\omega t) \sin(\omega k)) \\ &= \left( \sum_{k=-m_1}^{m_2} w_k \cos(\omega k) \right) \cos(\omega t) - \left( \sum_{k=-m_1}^{m_2} w_k \sin(\omega k) \right) \sin(\omega t) \\ &= G(\omega) \cos(\omega t + \Phi(\omega)) \end{aligned}$$

where  $G(\omega)$  and  $\Phi(\omega)$  are given in (4.9) and (4.12) respectively and the following result was used

$$x \cos(\omega) - y \sin(\omega) = \text{sgn}(x) \times \sqrt{x^2 + y^2} \cos(\omega + \arctan(\frac{y}{x})) \quad (5.4)$$

An example of a negative gain follows. If  $C < 0$  then peaks become troughs and troughs become peaks. If the gain is defined to be positive then a correction is required to the phase shift which is given in (4.12). These effects are shown in the example below for a symmetric moving average.

Consider the linear filter given by

$$f(y_t) = 0.5 \times Y_{t+1} + 0.5 \times Y_{t-1} \quad (5.5)$$

then

$$\begin{aligned} G(\omega) &= \sqrt{\cos(\omega)^2} = |\cos(\omega)| \\ \Phi(\omega) &= 0 \text{ for } 0 < \omega \leq \frac{\pi}{2} \\ \Phi(\omega) &= \pi \text{ for } \frac{\pi}{2} < \omega \leq \pi \text{ (} \cos(\omega) < 0 \text{)} \end{aligned}$$

For a 10 month cycle the original cycle and filtered cycle are shown in Figure 5.3 Graph A. It can be seen that there is no phase shift. Figure 5.3 Graph B shows a 3 month cycle which has a maximum phase shift of 1.5 months. It can also be noted here that for a 3 month cycle 1.5 months back and 1.5 months forward give the same cycle.

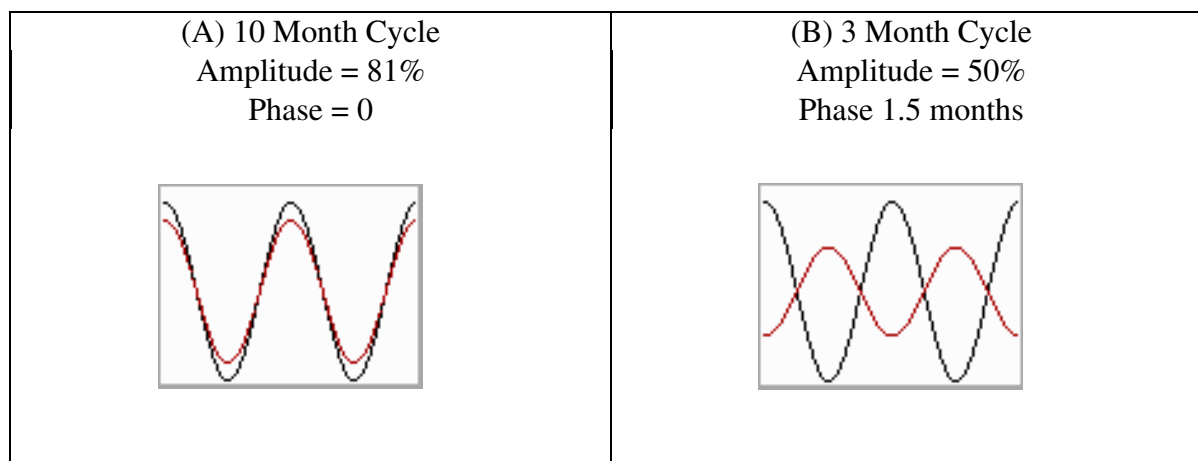


Figure 5.3: Effect of the linear filter presented above on various cycles

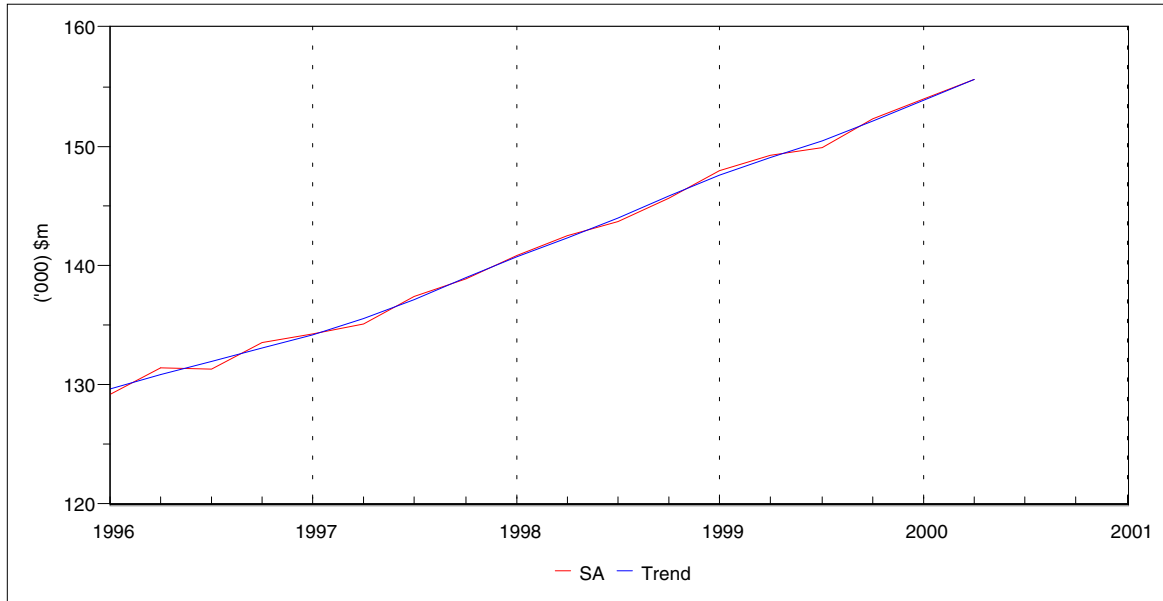
### 5.3 Trend moving averages (trend filters)

#### Henderson trend moving averages

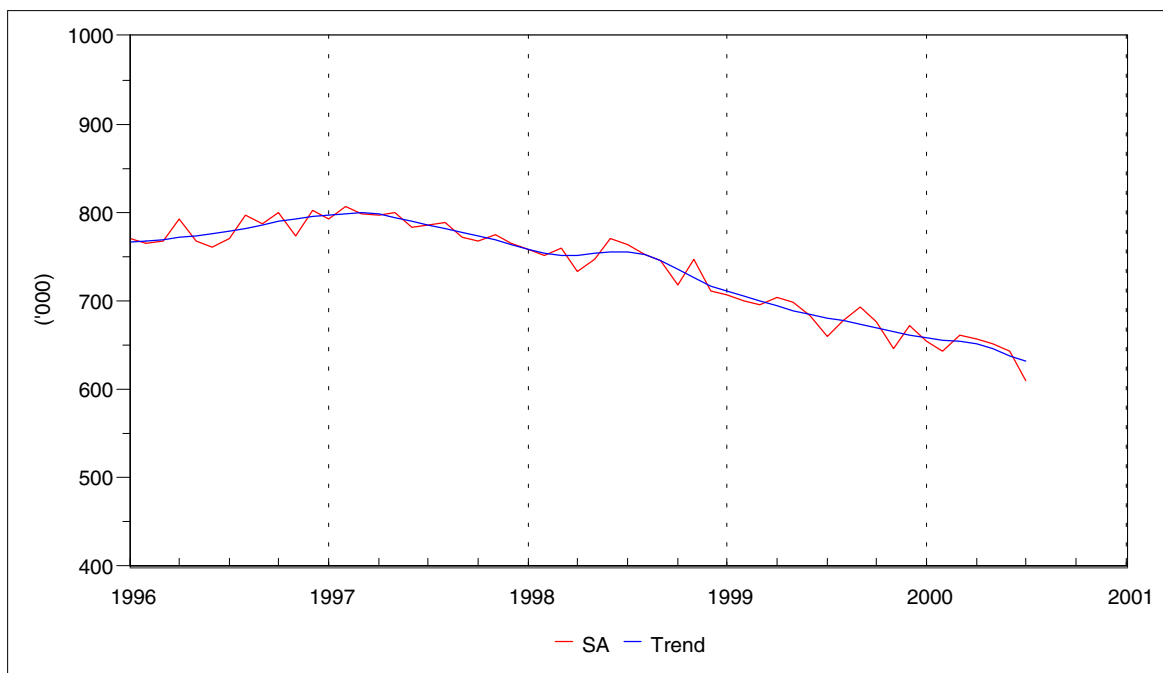
Henderson moving averages (derived by Henderson (1916) for use in actuarial applications) are a filter commonly used to smooth data, and have the property that they are able to track a cubic polynomial without distortion. To obtain the weights, a compromise was struck between two of the characteristics that are generally expected of a trend series; i.e. that the trend should be able to reproduce or represent a wide range of curvatures, and that it should be as smooth as possible. They can be used, for example, to obtain estimates of trend by smoothing a series containing trend and irregular components. Henderson moving averages are used in preference to simpler moving averages as they can reproduce polynomials up to degree three. This means that they are able to adequately capture trend turning points and points of inflection.

The ABS uses Henderson trend moving averages to produce trend estimates from the seasonally adjusted estimates. The ABS publishes trends derived using 13 term Henderson filters for monthly series while 7 term Henderson filters are used for quarterly series. There are two reasons for this: these filters have many desirable properties, and using a restricted set of filters means that users can, theoretically, reproduce the ABS trend estimates. It is important to note that the ABS uses other length Henderson filters, primarily 5, 9 and 23 term during the seasonal adjustment analysis process (see Chapter 7 for more details).

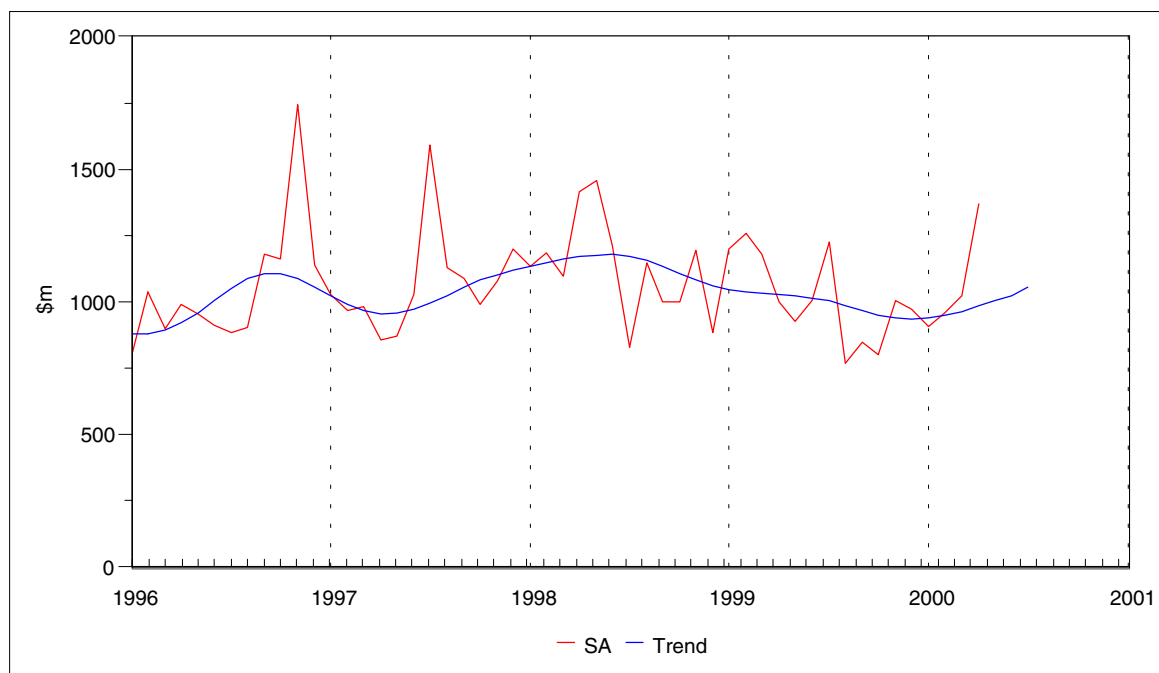
Figures 5.4 to 5.6 shows ABS time series trended with the 7, 13 and 23 term Henderson moving averages, respectively. In Figure 5.6, while the trend is derived using a 23 term Henderson filter, the published ABS trend is obtained from applying a 13 term Henderson to the published seasonally adjusted estimate.



**Figure 5.4: 7-Term Henderson filter : GDP (Chain Volume Measure)**



**Figure 5.5: 13-Term Henderson filter - Number of Unemployed Persons**



**Figure 5.6: 23-Term Henderson Filter - Value of Non-residential Building Approvals**

The Henderson trend moving averages are non-simple, meaning not all the observations necessarily have equal weight. The weights of a Henderson moving average are computed by

- (i) Minimising the sum of squares of the third differences (smoothness criteria)
- (ii) Allowing polynomials up to degree three to be exactly reproduced

Henderson moving averages can either be symmetric or asymmetric. Symmetric Henderson moving averages can be used in regions sufficiently far away from the start or end of a series. In this case a smoothed value for a given time series is calculated from equal numbers of values from previous and subsequent times.

Using symmetric trend moving averages results in what is commonly known as the 'end point problem'. This occurs because there will not be enough data available at the start or end of the series to apply the symmetric moving average. To solve this, asymmetric weights can be developed to carry out smoothing at the ends of the series. Asymmetric weights used for trend analysis can be tailored to suit the characteristics of the series. The ABS uses asymmetric weights which have similar filter properties to the symmetric Henderson trend moving averages. These weights are used within the seasonal adjustment package, X11, and have been derived from a formula developed by John Musgrave for the US Bureau of the Census. This formula, which is independent of the series to which it is applied, results in asymmetric trend moving averages which have similar dampening properties to the Henderson and weights that sum to one (for more information see subsection on asymmetric Henderson).

The following sections show how the symmetric and asymmetric Henderson filters are derived (see Kenny and Durbin (1982), Laniel (1985), Doherty (1992), Gray and Thomson (1996a), Findley *et. al* (1998) or McLaren (1999) for more details).

## Symmetric Henderson moving averages

Consider a linear filter of length  $n = 2r + 1$  and denote the coefficients of this filter by  $w_k$  where  $k = -r, \dots, r$ . Applying this filter to a time point sufficiently far away from the start of a non-seasonal input series  $\{y_t : t = 1, \dots, n\}$  gives

$$\begin{aligned}\hat{A}_{t|\infty} &= \sum_{i=-r}^r w_i y_{t+i} \\ &= w_r' y_t\end{aligned}\quad (5.6)$$

where  $\hat{A}_{t|\infty}$  represents a smoothed series at time  $t$  given that there is information up to a time sufficiently far away from time  $t$  in the filter window,  $w_r = (w_{-r}, \dots, w_r)'$  is the vector of unknown coefficients and  $y_t = (y_{t-r}, \dots, y_{t+r})'$  is a non-seasonal input series. The coefficients in  $w_r$  are subject to

$$\sum_{i=-r}^r w_i = 1 \quad \text{and} \quad \sum_{i=-r}^r i^j w_i = 0 \quad (5.7)$$

where  $j = 0, 1, 2$  and  $3$  to ensure that the local cubic criteria is satisfied, and the first condition ensures the coefficients  $w_i$  are required to sum to one. The second condition is automatically satisfied for symmetric weights  $w_i = w_{-i}$ .

Henderson assumed that  $y_t$  consists of a trend can be represented by a locally cubic polynomial, over the length of the filter window at time  $t = 1, \dots, n$  and a residual term  $\varepsilon_t$ . That is

$$\begin{aligned}y_t &= T_t + \varepsilon_t \\ &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t \\ &= P_t^3 + \varepsilon_t\end{aligned}\quad (5.8)$$

where the  $\beta_j$ 's are arbitrary coefficients and  $\varepsilon_t$  has the following properties

$$\begin{aligned}E(\varepsilon_t) &= 0, \\ \text{Var}(\varepsilon_t) &= \text{Var}(\varepsilon_{t-k}) \forall k, \\ \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) &= 0\end{aligned}\quad (5.9)$$

Rewriting (5.8) in matrix form gives

$$\begin{aligned}y_t &= \begin{pmatrix} \beta_0 + \beta_1(t-r) + \beta_2(t-r)^2 + \beta_3(t-r)^3 \\ \vdots \\ \beta_0 + \beta_1(t+r) + \beta_2(t+r)^2 + \beta_3(t+r)^3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{t-r} \\ \vdots \\ \varepsilon_{t+r} \end{pmatrix} \\ &= P_t^3 + \varepsilon_t\end{aligned}\quad (5.10)$$

where  $P_t^3$  and  $\varepsilon_t$  are vectors of length  $n$  centered at time  $t$ . Substituting (5.10) into (5.6) results in

$$\hat{A}_{t|\infty} = w_r'(P_t^3 + \varepsilon_t)$$

Then applying the difference operator gives

$$\begin{aligned}\Delta \hat{A}_{t|t\infty} &= \hat{A}_{(t+1)|t\infty} - \hat{A}_{t|t\infty} \\ &= w_r' P_{t+1}^3 + w_r' \varepsilon_{t+1} - (w_r' P_t^3 + w_r' \varepsilon_t) \\ &= \Delta(w_r' P_{t+1}^3) + \Delta(w_r' \varepsilon_t)\end{aligned}$$

The range of the first differential series now extends to include  $t+r+1$ . The difference operator is applied again to obtain third differences as per the Henderson criteria

$$\begin{aligned}\Delta^3 \hat{A}_{t|t\infty} &= \Delta^3(w_r' P_{t+1}^3) + \Delta^3(w_r' \varepsilon_t) \\ &= a + \Delta^3(w_r' \varepsilon_t)\end{aligned}$$

where  $a$  is a vector of constants since the third differences of a polynomial of order three is a constant. The range of the second differenced series extends to  $t+r+3$ . To obtain values for the coefficients  $w_r$  the variance of the third differences of the smoothed series has to be minimised

$$\begin{aligned}\text{Var}(\Delta^3 \hat{A}_{t|t\infty}) &= \text{Var}(a) + \text{Var}(\Delta^3 w_r' \varepsilon_t) \\ &= (\Delta^3 w_r)' \text{Var}(\varepsilon_t) (\Delta^3 w_r)\end{aligned}\quad (5.11)$$

where  $\Delta^3 w_r = (\Delta^3 w_{-r}, \dots, \Delta^3 w_r)$  is the vector of the third differences of length  $n$ .

By using the properties in (5.9),  $\text{Var}(\varepsilon_t)$  can be written as the identity matrix and (5.11) then becomes equivalent to minimising

$$\sum_{k=-r}^r (\Delta^3 w_r)^2$$

subject to the constraint (5.7).

Using Lagrange multipliers the symmetric  $n = 2r + 1$  term Henderson filter can be written in the form

$$w_i = \{(r+1)^2 - i^2\} \{(r+2)^2 - i^2\} \{(r+3)^2 - i^2\} \{\beta_0 + \beta_1 i^2\} \quad (5.12)$$

where

$$\begin{aligned}\beta_0 &= \left(\frac{315}{8}\right) (3r^2 + 12r - 4) / (64r^9 + 1152r^8 + 8592r^7 + 34272r^6 + 78204r^5 \\ &\quad + 99288r^4 + 57203r^3 - 4302r^2 - 19323r - 5670) \\ \beta_1 &= \left(\frac{3465}{8}\right) 1 / (64r^9 + 1152r^8 + 8592r^7 + 34272r^6 + 78204r^5 + 99288r^4 \\ &\quad + 57203r^3 - 4302r^2 - 19323r - 5670)\end{aligned}$$

The weighting pattern for a range of symmetric Henderson moving averages is given in Table 5.1.

### Gain and Phase Functions of the Symmetric Henderson Moving Averages

It is useful to summarise the properties of a moving average through the use of the gain function. Figures 5.7 to 5.11 consist of graphs of the gain and phase functions of the 5, 7, 9, 13 and 23 term symmetric Henderson moving averages. The gain functions show the strength remaining at certain cycles. The cycle strength is shown on the y-axis, with 1 meaning cycle strength is unchanged, greater than 1 meaning the strength of the cycle is increased and less than 1 indicating the strength of the cycle is decreased. From the figures, the trend moving averages dampen the short term cycles which can be considered to represent the irregular component.

Weighting pattern for symmetric Henderson filters	
5 Term Henderson	(-0.073, 0.294, 0.558, 0.294, -0.073)
7 Term Henderson	(-0.059, 0.059, 0.294, 0.412, 0.294, 0.059, -0.059)
9 Term Henderson	(-0.041, -0.010, 0.119, 0.267, 0.330, 0.267, 0.119, -0.010, -0.041)
13 Term Henderson	(-0.019, -0.028, 0.0, 0.066, 0.147, 0.214, 0.240, 0.214, 0.147, 0.066, 0.0, -0.028, -0.019)
23 Term Henderson	(-0.004, -0.011, -0.016, -0.015, -0.005, 0.013, 0.039, 0.068, 0.097, 0.122, 0.138, 0.148, 0.138, 0.122, 0.097, 0.068, 0.039, 0.013, -0.005, -0.015, -0.016, -0.011, -0.004)

Table 5.1: Weighting pattern for a range of symmetric Henderson filters

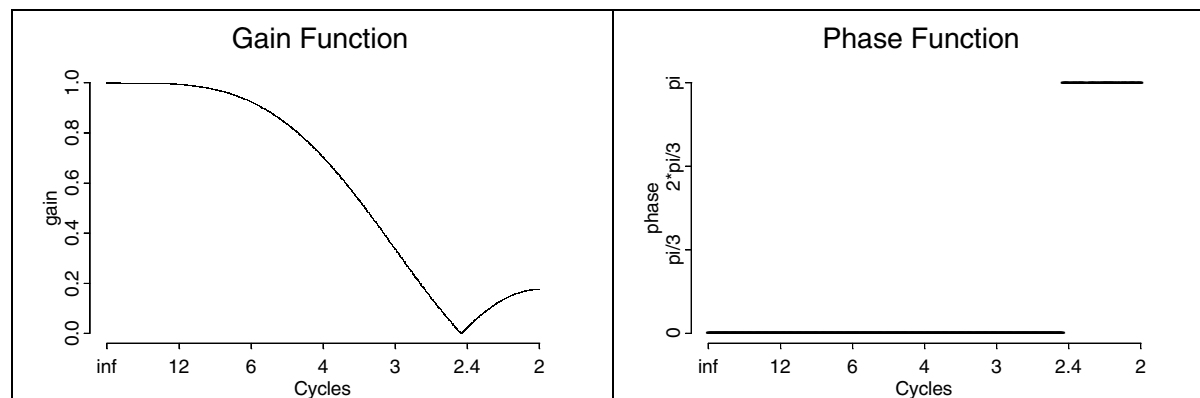


Figure 5.7: Gain and Phase Functions using Symmetric 5 Term Henderson filters



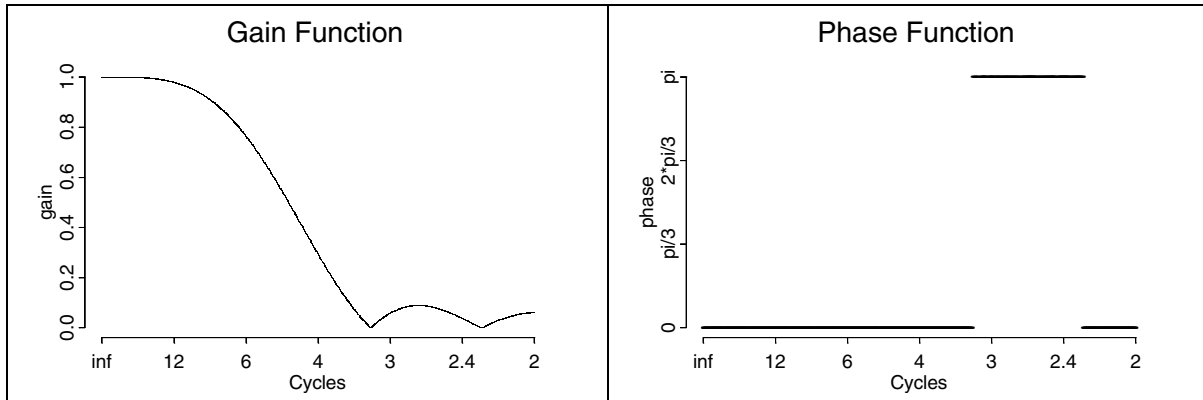


Figure 5.8: Gain and Phase Functions using Symmetric 7 Term Henderson filters

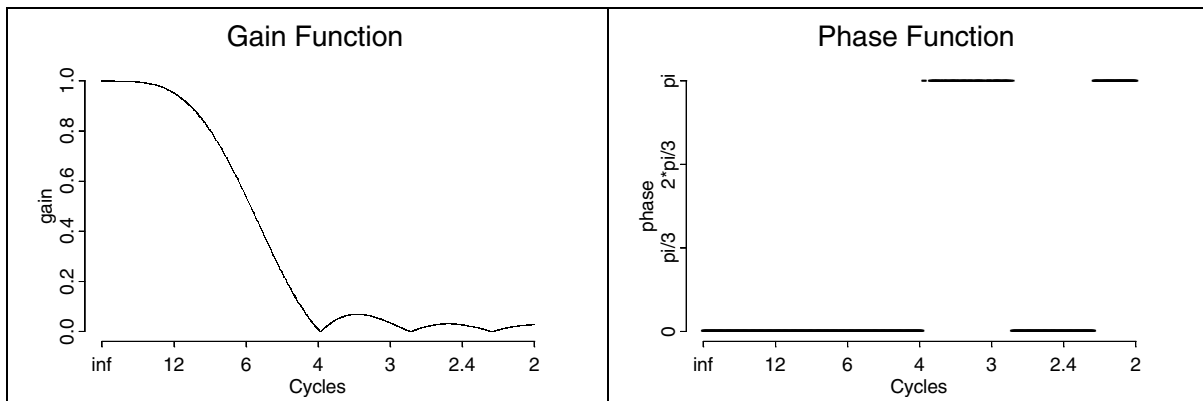


Figure 5.9: Gain and Phase Functions using Symmetric 9 Term Henderson filters

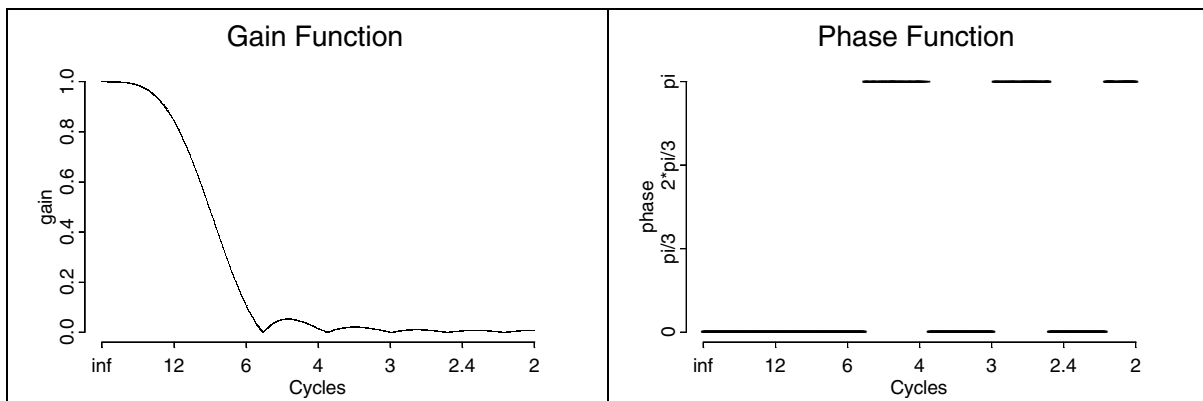


Figure 5.10: Gain and Phase Functions using Symmetric 13 Term Henderson filters

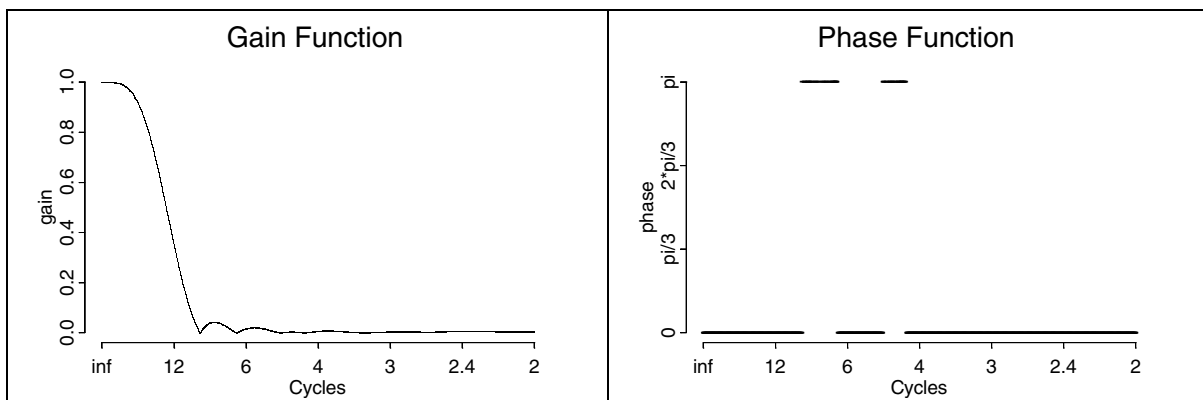


Figure 5.11: Gain and Phase Functions using Symmetric 23 Term Henderson filters

Figure 5.10 shows that for monthly data the 13 term Henderson almost exactly reproduces cycles in the range 18 months to infinity while in the range 8 months to 18 month the cycles are reduced to various degrees.

In general the longer the trend moving average, the smoother the resulting trend, as evident from Figure 5.11. When using a 5 or 7 term Henderson, cycles of about 8 periods or longer are left with their original amplitude while the 23 term Henderson only leaves cycles of about 36 periods or longer with their original amplitude. The 5 term Henderson reduces cycles of about 2.4 periods or less by at least 80% whereas the 23 term Henderson reduces cycles of about 8 periods or less by at least 90%. When using a 23 term Henderson, cycles less than about 4 periods are completely removed. In the range 2 to 8 months the amplitude of the cycles is almost eliminated.

Note that the Henderson moving averages also dampen the seasonal cycles to varying degrees, although not significantly enough, especially with respect to cycles of 12 periods, to warrant applying them to an original series to obtain a measure of trend. This is why they should only be applied to seasonally adjusted series where the calendar related effects have been removed with specifically designed seasonal filters.

### **Asymmetric Henderson Moving Averages**

The symmetric Henderson moving average can only be directly applied to the historical part of the data. For example the standard 13 term Henderson can only be applied to monthly data that is at least 6 observations from the start or end of the data. To provide trend estimates of the most recent data points a modified or asymmetric moving average must be used.

Trend estimates at the end of the series allow the current direction of the series to be assessed. Estimates of the current direction of the series are often very important to users of our data and it is therefore important to provide 'good' asymmetric filters. Calculation of the asymmetric Henderson filters can be generated by a number of different methods which generate similar results. The four main methods are the Musgrave method, the Minimisation of the Mean Square Revision method, the Best Linear Unbiased Estimates method, and the Kenny and Durbin method. A brief description of the Musgrave method is given below. For a more detailed discussion of asymmetric weights see Doherty (1992) and Findley *et. al* (1998).

Shiskin *et. al* (1967) derived the original asymmetric weights for the Henderson moving average which are used within the X11 package. The weights have been chosen to minimise the mean square of the revision between the initial and final trend estimate. No documentation was kept of the original derivation of the weights and Wallis (1982 p.30) stated

*'The technical manual for the X11 program gives a set of asymmetric filters, ...but the source of these is somewhat mysterious. Certainly, no reference is given by the authors of the technical manual.'*

Laniel (1985) rediscovered the original criteria for the asymmetric filters used within X11. The following is based on a mean square revision (MSR) criteria and reproduces the Henderson Moving Averages as used within the X11 package.

Similar to (5.6) the asymmetric weights can be calculated using

$$\hat{A}_{t+t+k} = \sum_{i=-r}^k u_i y_{t+i}$$

where  $r$  is an indicator determining the length of the asymmetric weights with  $0 \leq r \leq (k-1)$ ,  $u_i$  are a set of asymmetric weights, and the data of the series  $y_t$  is up to time point  $t = t+k$ .

The revision between two smoothed values for time  $t$  using data up to times  $t+i$  and  $t+k$ ,  $0 \leq i < k \leq m$  is then defined as

$$r_t^{(k,i)} = \hat{A}_{t+t+i} - \hat{A}_{t+t+k}$$

It is assumed that the trend cycle of the original series follows a straight line and the random noise component  $\varepsilon_t$  follows an independent distribution with variance  $\sigma_a^2$ . That is  $y_t = a_1 t + \varepsilon_t$ . The revision between the final trend estimate calculated using the symmetric filter and an initial trend estimate can then be expressed as

$$\begin{aligned} r_t^{(k,r)} &= \hat{A}_{t+t+r} - \hat{A}_{t+t+k} \\ &= \sum_{j=-r}^r w_j y_{t-j} - \sum_{j=-r}^k u_j y_{t-j} \\ &= \sum_{j=-r}^r w_j (A_{t-j} + \varepsilon_{t-j}) - \sum_{j=-r}^k u_j (A_{t-j} + \varepsilon_{t-j}) \end{aligned}$$

where  $\{w_i : i = -r, \dots, r\}$  are the symmetric Henderson weights and  $\{u_i : i = -r, \dots, k\}$  are the asymmetric Henderson weights.

The mean square revision criteria is defined as

$$E\{(r_t^{(k,r)})^2\} = E^2(r_t^{(k,r)}) + V(r_t^{(k,r)})$$

Laniel (1985) showed that the mean square revision can be written

$$E\{(r_t^{(k,r)})^2\} = a_1^2 \left(t - \sum_{j=-r}^k u_j (t-j)\right)^2 + \sigma_a^2 \sum_{j=-r}^r (w_j - u_j)^2 \quad (5.13)$$

where  $a_1$  and  $\sigma_a^2$  are unknown.

To solve (5.13) the unknown parameter  $a_1^2$  can be written in terms of  $\sigma_a^2$  using a noise to trend ratio. This is typically denoted as  $I/C$ . The  $I/C$  ratio is defined as the absolute value of the averages of the first difference of the irregulars divided by the absolute value of the averages of the first difference of the trend-cycle. That is, under normality of the  $\varepsilon_t$ 's:

$$I/C = \left(\frac{4\sigma_a^2}{\pi}\right)^2 \left(\frac{1}{|a_1|}\right)$$

Different values for the  $I/C$  ratio result in different smoothness properties. For example, for monthly series the recommended approach is to set  $I/C=3.5$  which gives the standard 13 term

Henderson Moving Average (HMA) trend filter and  $I/C=4.5$  gives the standard 23 term HMA trend filter (Dagum, 1988).

Substituting  $I/C$  into (5.13) gives

$$\left( \frac{E\{(r_i^{(i,k)})^2\}}{\sigma_a^2} \right) = \left( \frac{4}{\pi(I/C)^2} \right) \left( \sum_{j=-r}^k w_j i \right)^2 + \sum_{j=-r}^r (w_j - u_j)^2 \quad (5.14)$$

Setting  $I/C$  and then minimising (5.14) gives the HMA asymmetric weights. The asymmetric filters derived using this method are the standard Henderson moving averages.

An  $I/C$  parameter allows for a wide range of data to be smoothed with the same set of end weights. However, data not sufficiently close to the global parameter such as very smooth data, may have higher revision than choosing the optimal  $I/C$  parameter. In practice, the  $I/C$  ratio can be estimated directly from the data. The ABS currently calculates an "optimal" value using the last few years of data. An alternative method could allow the  $I/C$  ratio to change slowly over time.

The properties of the asymmetric moving averages at the end of the data are only approximations to the final symmetric moving average. An example of asymmetric weights for the 13 term Henderson filter are given in Table 5.2.

The asymmetric Henderson trend moving averages do not have all the properties of symmetric Henderson filters, although their weights do sum to one. For example, while the Henderson moving averages reproduce polynomials of degree three or less, the asymmetric filters do not even reproduce straight lines, unless an end weight parameter  $I/C = 0$  is used.

The different properties, in terms of the gain function, of the asymmetric 13 term Henderson trend moving averages and the symmetric 13 term Henderson moving average can be seen in Figure 5.12.

	N-12	N-11	N-10	N-9	N-8	N-7	N-6	N-5	N-4	N-3	N-2	N-1	N
N	0	0	0	0	0	0	-0.092	-0.058	0.012	0.120	0.244	0.353	0.421
N-1	0	0	0	0	0	-0.043	-0.039	0.002	0.080	0.174	0.254	0.292	0.279
N-2	0	0	0	0	-0.016	-0.025	0.003	0.068	0.149	0.216	0.241	0.215	0.148
N-3	0	0	0	-0.008	-0.020	0.004	0.066	0.144	0.208	0.230	0.201	0.130	0.045
N-4	0	0	-0.011	-0.022	0.003	0.066	0.146	0.210	0.233	0.205	0.135	0.051	-0.017
N-5	0	-0.016	-0.026	0.001	0.066	0.147	0.213	0.238	0.211	0.144	0.061	-0.005	-0.034
N-6	-0.019	-0.028	0.000	0.066	0.147	0.214	0.240	0.214	0.147	0.066	0.000	-0.028	-0.019

Table 5.2: Asymmetric weighting patterns ( $I/C=3.50$ ) for a 13 term symmetric Henderson filter

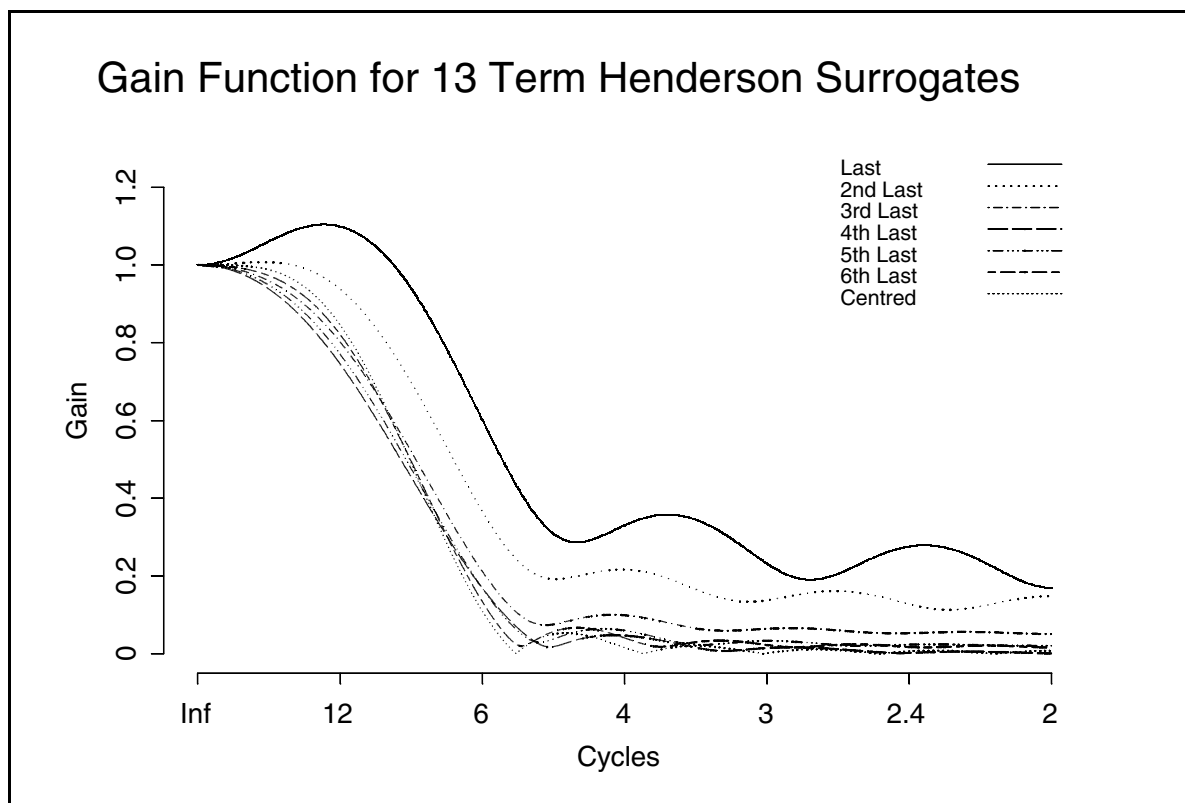


Figure 5.12: Gain functions for the asymmetric Henderson trend moving averages for the 13 term Henderson

In Figure 5.12, it is clear that the surrogates for estimating the trend for the third last to the last observation of the input series do not dampen the same cycles as the 13 term Henderson. Indeed the surrogate for the last observation even amplifies the strength of the 12 period cycle. It can also be seen that the surrogates for the fourth to the sixth last observations are very similar in their filter properties to the 13 term Henderson. The Henderson surrogates also produce some time phase shifting, unlike the symmetric Henderson moving averages.

One solution to the end point problem, that may seem appropriate, is using shorter Henderson trend moving averages towards the end of the series. This approach leads to two problems. Firstly, smoothed estimates cannot be calculated for the second last or the last observation because the weighting pattern of a 3 term Henderson is  $(0.0, 1.0, 0.0)$  (for the second last observation) and there is no symmetric Henderson moving average to smooth one

observation. Secondly, the gain functions above show that different length Henderson moving averages have different filtering properties and so this approach would not be consistent in smoothing the input series, particularly with respect to the current movements which are often scrutinised the most. This could also lead to greater revisions in the trend estimate over time.

A summary of some of the advantages and disadvantages of fixed end weights (standard I/C ratio) against end weights designed to be optimal in terms of the I/C ratio (optimally generated I/C ratio) are given in Table 5.3.

<b>Fixed weights used on all time series</b>	<b>Advantages</b>	<b>Disadvantages</b>
	Simple to understand and apply	May exhibit noticeable bias in smooth growing times
	Comparable over different time series	May not be optimal for a wide range of criteria
	Additive over components	
	Gives reasonable results for a range of time series	
<b>Weights designed for individual time series</b>	Optimal for a given criteria	Much more complicated to apply
	Bias can be minimised	Potential to be affected by large outliers
		Not additive over components

**Table 5.3: Advantages and disadvantages of fixed end weights**

## 5.4 Seasonal moving averages (seasonal filters)

Almost all of the data investigated by the ABS have seasonal characteristics. Since the Henderson moving averages used to trend the data do not eliminate seasonality, the data must be seasonally adjusted first. The X11 approach to seasonal adjustment attempts to remove all of the cyclical power near the seasonal harmonics. This can be achieved by using filters designed to remove seasonal harmonics.

Typically ABS data is seasonally reanalysed every 12 months. Under this approach seasonal adjustment is performed once a year and seasonal factors are calculated for the next year and are kept fixed over a 12 month period. As with the trend at the ends of the data the seasonally adjusted data changes from what is initially published to what is finally published. In most cases it takes longer for the seasonally adjusted data to become final. An alternative to seasonally adjusting the data every 12 months is seasonally adjusting the data every month. This issue is further discussed in Section 6.6.

### Symmetric seasonal moving averages

A seasonal filter has weights which are applied to the same month over time. An example of a seasonal filter would be

$$(1/3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1/3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1/3)$$

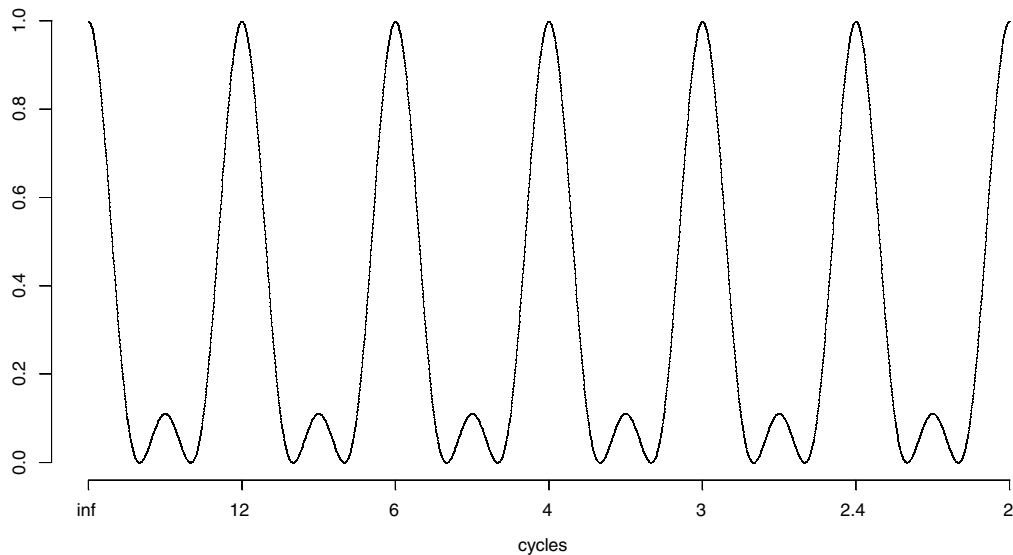
where, for example, the weight 1/3 is applied to consecutive Januarys.

Within X11 a range of seasonal filters are available to choose from. These are a weighted 3-term moving average (ma)  $S_{3 \times 1}$ , weighted 5-term ma  $S_{3 \times 3}$ , weighted 7-term ma  $S_{3 \times 5}$  and a weighted 11-term ma  $S_{3 \times 9}$ . The gain functions for the  $S_{3 \times 3}$ ,  $S_{3 \times 5}$ , and  $S_{3 \times 9}$  term symmetric seasonal filters are produced in Figures 5.13, 5.14, and 5.15 respectively. The choice of the seasonal filter will depend on how the seasonal terms change over time. For example, the greater the movement in the seasonal component, the shorter the filter choice i.e. a  $S_{3 \times 1}$  or a  $S_{3 \times 3}$ .

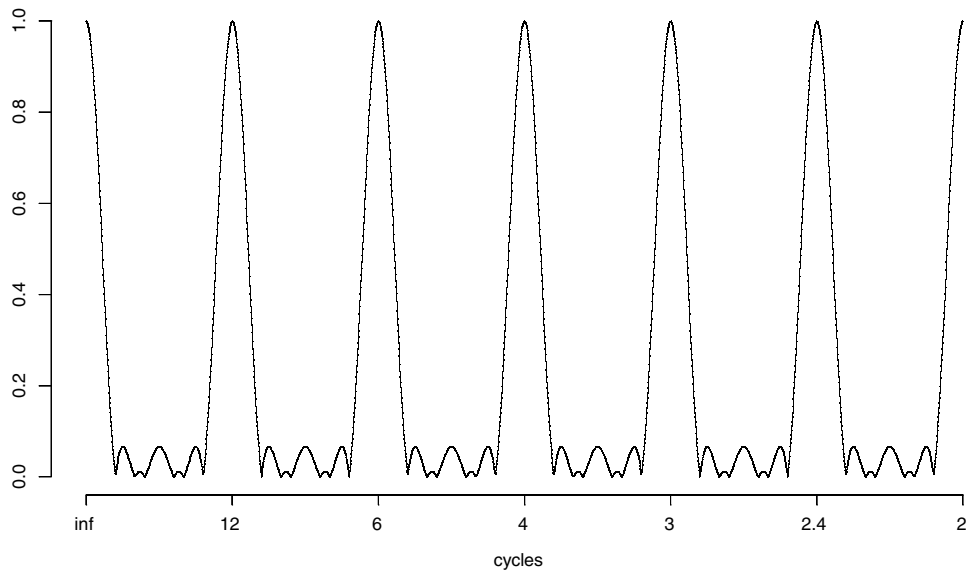
Dagum *et. al* (1996) gives a general form for the symmetric seasonal filters as

$$S_{3 \times k}(B) = (3k)^{-1} B^{-12[1+(k-1)/2]} (1 + B^{12} + B^{24}) (1 + \dots + B^{(k-1)12})$$

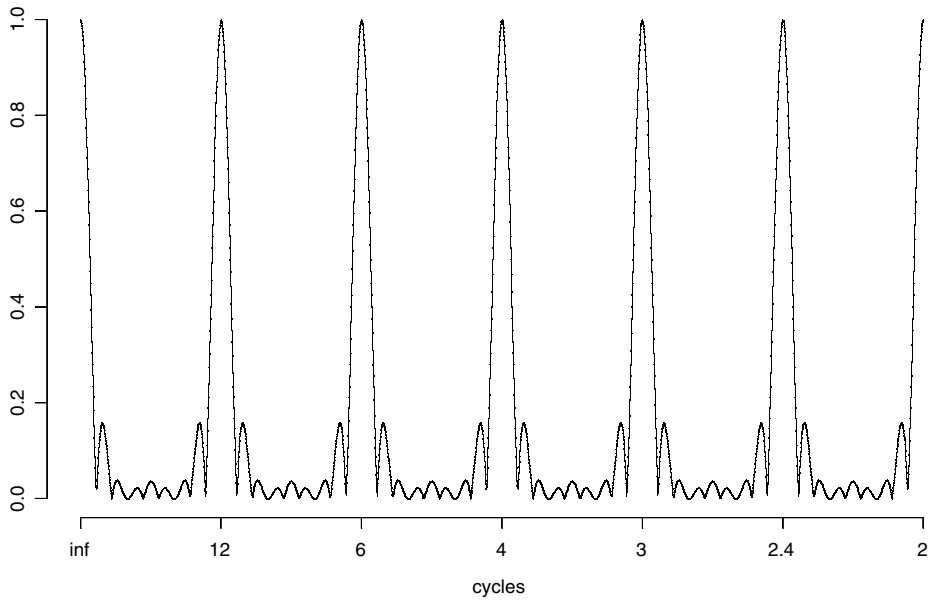
for  $k = 1, 3, 5, 9$ .



**Figure 5.13: Gain Function for  $S_{3 \times 3}$  Symmetric Seasonal Filter**



**Figure 5.14: Gain Function for  $S_{3 \times 5}$  Symmetric Seasonal Filter**



**Figure 5.15: Gain Function for  $S_{3 \times 9}$  Symmetric Seasonal Filter**



## Asymmetric seasonal moving averages

Asymmetric weights used for the seasonal factors obtained from the seasonal-irregular (SI's) factors have not been extensively documented. As with the asymmetric Henderson weights the asymmetric seasonal weights depend on implicit forecasts of the Seasonal-Irregular component.

### 3 × 1 Term Asymmetric Weights

The symmetric weights for a 3 × 1 term symmetric seasonal filter are

$$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

The asymmetric weights used within X11 are given in Shiskin *et. al* (1967), and are included below in Table 5.4.

Factor for year	Weight given S-I ratios in year		
	$N-2$	$N-1$	$N$
$N$	0	0.39	0.61
$N-1$	0.333	0.333	0.333

**Table 5.4: Published X11 Asymmetric Weights for  $S_{3 \times 1}$**

These seem to have been generated using the same system (Musgrave) as used for the Henderson weights with an I/C ratio of 1.5. Using this approach gives the weights in Table 5.5.

Factor for year	Weight given S-I ratios in year		
	$N-2$	$N-1$	$N$
$N$	0	0.389	0.61
$N-1$	0.333	0.333	0.333

**Table 5.5: Henderson Asymmetric Weights with I/C Ratio of 1.5**

### 3 × 3 Term Asymmetric Weights

The symmetric weights for a 3 × 3 term symmetric seasonal filter are

$$\left(\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}\right)$$

The published asymmetric weights are shown below in Table 5.6.

Factor for year	Weight given S-I ratios in year				
	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0.185	0.407	0.407
$N-1$	0	0.111	0.259	0.37	0.259
$N-2$	0.111	0.222	0.333	0.222	0.111

**Table 5.6: Published Asymmetric Weights for  $S_{3 \times 3}$**

The weights shown in Table 5.6 can be derived in the following way. Let the data be given by

$$X_{N-3}, X_{N-2}, X_{N-1}, X_N, X_{N+1}, X_{N+2}$$

Then let

$$X_{N+1} = \frac{1}{3}(X_{N-2} + X_{N-1} + X_N)$$

$$X_{N+2} = X_{N-1}$$

Then the  $N-1$  weights are

$$\begin{aligned} & \frac{1}{9}X_{N-3} + \frac{2}{9}X_{N-2} + \frac{3}{9}X_{N-1} + \frac{2}{9}X_N + \frac{1}{9}\left(\frac{1}{3}(X_{N-2} + X_{N-1} + X_N)\right) \\ &= \frac{3}{27}X_{N-3} + \frac{7}{27}X_{N-2} + \frac{10}{27}X_{N-1} + \frac{7}{27}X_N \end{aligned}$$

The  $N$  weights are

$$\begin{aligned} & \frac{1}{9}X_{N-2} + \frac{2}{9}X_{N-1} + \frac{3}{9}X_N + \frac{2}{9}\left(\frac{1}{3}(X_{N-2} + X_{N-1} + X_N)\right) + \frac{1}{9}X_{N-1} \\ &= \frac{5}{27}X_{N-2} + \frac{11}{27}X_{N-1} + \frac{11}{27}X_N \end{aligned}$$

This gives the weights which are shown in Table 5.7.

Factor for year	Weight given S-I ratios in year				
	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0.185	0.407	0.407
$N-1$	0	0.111	0.259	0.370	0.259
$N-2$	0.111	0.222	0.333	0.222	0.111

**Table 5.7: Derived Asymmetric Weights for  $S_{3 \times 3}$**

The gain function for the  $S_{3 \times 3}$  asymmetric seasonal filter (row  $N$  in Table 5.7) is shown in Figure 5.16.

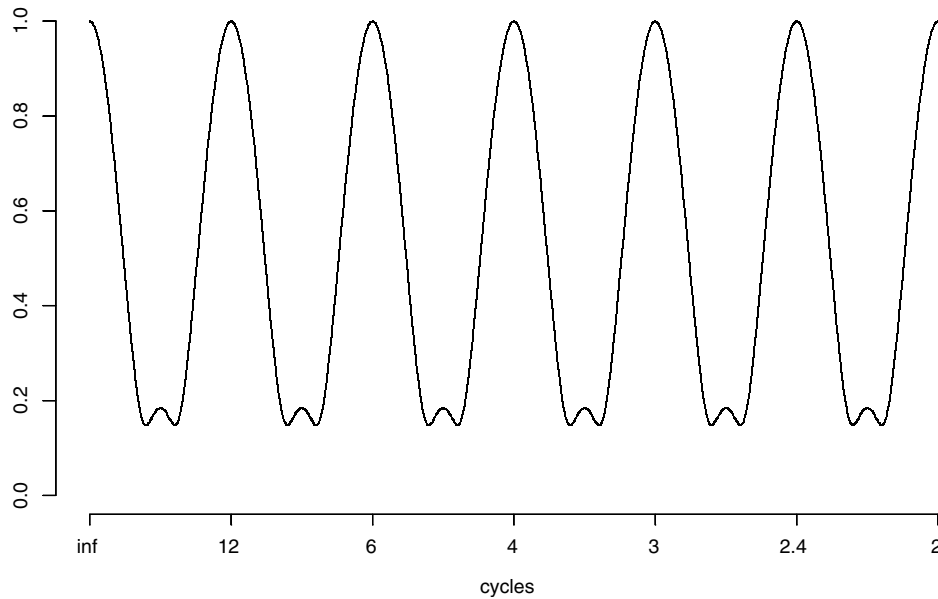


Figure 5.16: Gain Function for  $S_{3 \times 3}$  Asymmetric Seasonal Filter (factor year  $N$  - row one in Table 5.7)

### $3 \times 5$ Term Asymmetric Weights

The symmetric weights for a  $3 \times 5$  term symmetric seasonal filter are

$$\left(\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{3}{15}, \frac{2}{15}, \frac{1}{15}\right)$$

The  $3 \times 5$  seasonal factors are the most commonly used. They have forecasts of one and two steps ahead of the average of the last four SI's and three steps ahead of the third last SI.

Factor for year	Weight given S-I ratios in year						
	$N-6$	$N-5$	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0	0.150	0.283	0.283	0.283
$N-1$	0	0	0.067	0.183	0.250	0.250	0.250
$N-2$	0	0.067	0.133	0.217	0.217	0.217	0.150
$N-3$	0.067	0.133	0.200	0.200	0.200	0.133	0.067

Table 5.8: Published Asymmetric Weights for  $S_{3 \times 5}$

The weights in Table 5.8 can be derived as follows. Let the data be

$$\begin{aligned}
 &X_{N-5}, X_{N-4}, X_{N-3}, X_{N-2}, X_{N-1}, X_N, X_{N+1}, X_{N+2}, X_{N+3} \\
 X_{N+1} &= \frac{1}{4}(X_{N-3} + X_{N-2} + X_{N-1} + X_N) \\
 X_{N+2} &= \frac{1}{4}(X_{N-3} + X_{N-2} + X_{N-1} + X_N) \\
 X_{N+2} &= \frac{1}{4}(X_{N-3} + X_{N-2} + X_{N-1} + X_N) \\
 X_{N+3} &= X_{N-2}
 \end{aligned}$$

Using similar algebra to the  $3 \times 3$  case gives the following weights

$$N-2 \quad \left( \frac{4}{60}, \frac{8}{60}, \frac{13}{60}, \frac{13}{60}, \frac{13}{60}, \frac{9}{60} \right)$$

$$N-1 \quad \left( \frac{4}{60}, \frac{11}{60}, \frac{15}{60}, \frac{15}{60}, \frac{15}{60} \right)$$

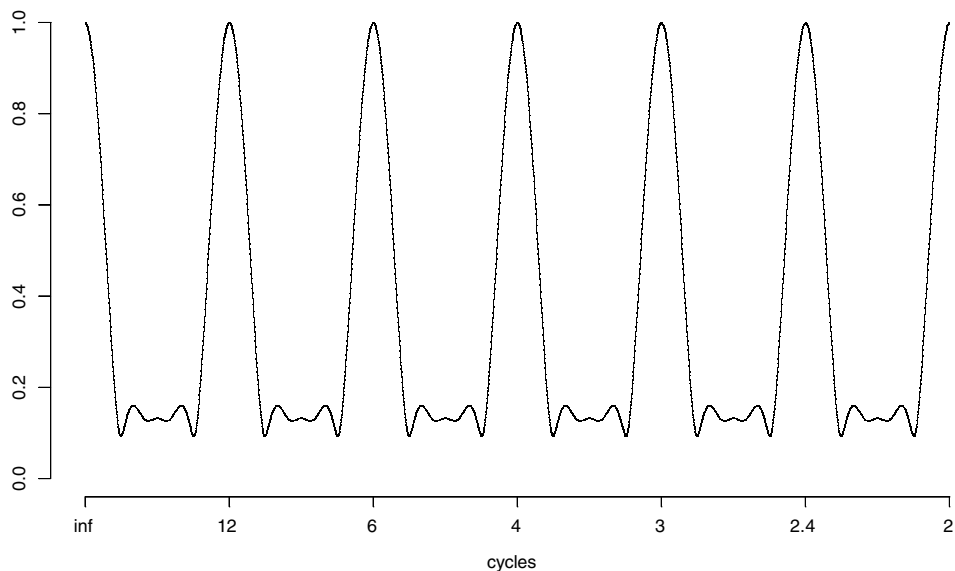
$$N \quad \left( \frac{9}{60}, \frac{17}{60}, \frac{17}{60}, \frac{17}{60} \right)$$

The final weights are shown in Table 5.9.

Factor for year	Weight given S-I ratios in year						
	$N-6$	$N-5$	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0	0.150	0.283	0.283	0.283
$N-1$	0	0	0.067	0.183	0.250	0.250	0.250
$N-2$	0	0.067	0.133	0.217	0.217	0.217	0.150
$N-3$	0.067	0.133	0.200	0.200	0.200	0.133	0.067

**Table 5.9: Derived Asymmetric Weights for  $S_{3 \times 5}$**

The gain function for the  $S_{3 \times 5}$  asymmetric seasonal filter (row  $N-1$  in Table 5.9) is shown in Figure 5.17.



**Figure 5.17: Gain Function for  $S_{3 \times 5}$  Asymmetric Seasonal Filter (factor year  $N-1$  row two in Table 5.7)**

### 3 × 9 Term Asymmetric Weights

The symmetric weights for a 3 × 9 term seasonal filter are

$$\left(\frac{1}{27}, \frac{2}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{2}{27}, \frac{1}{27}\right)$$

The published weights appear in Table 5.10.

Factor for year	Weight given S-I ratios in year										
	$N-10$	$N-9$	$N-8$	$N-7$	$N-6$	$N-5$	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0	0	0	0.051	0.112	0.173	0.197	0.221	0.246
$N-1$	0	0	0	0	0.028	0.092	0.144	0.160	0.176	0.192	0.208
$N-2$	0	0	0	0.032	0.079	0.123	0.133	0.143	0.154	0.163	0.173
$N-3$	0	0	0.034	0.075	0.113	0.117	0.123	0.128	0.132	0.137	0.141
$N-4$	0	0.034	0.073	0.111	0.113	0.114	0.116	0.117	0.118	0.120	0.084
$N-5$	0.037	0.074	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.074	0.037

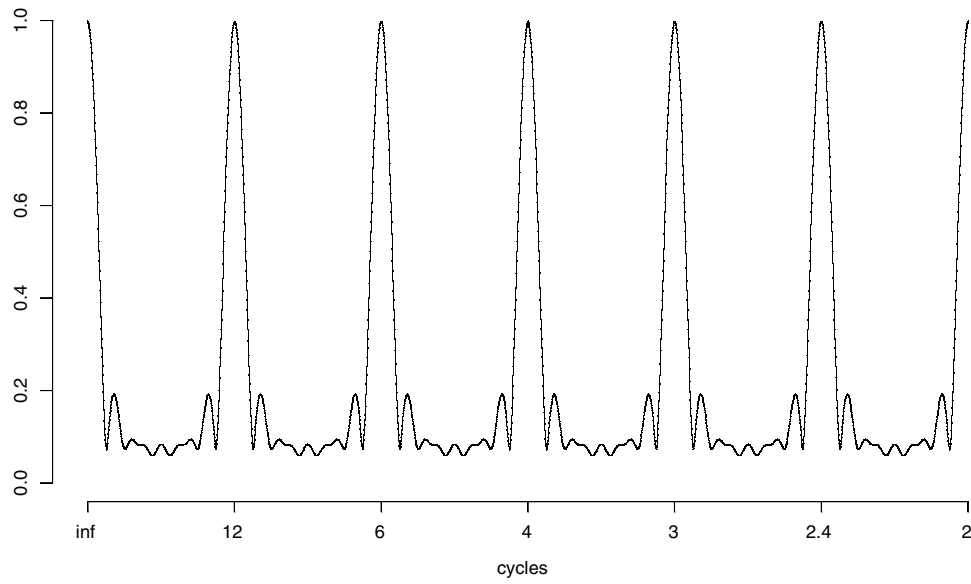
**Table 5.10** Published Asymmetric Weights for  $S_{3 \times 9}$

The computation of the asymmetric weights for the 3 × 9 is a mystery. They do not appear to allow for simple forecasting weights such as those used for the 3 × 3 and 3 × 5. A close approximation can be made by using the formula for the Henderson weights with an  $I/C$  ratio of about 9.8. This gives the weights shown below in Table 5.11. They have been rounded to 3 decimal places but not scaled to sum to 1. The "fit" is good except at the smallest weights.

Factor for year	Weight given S-I ratios in year										
	$N-10$	$N-9$	$N-8$	$N-7$	$N-6$	$N-5$	$N-4$	$N-3$	$N-2$	$N-1$	$N$
$N$	0	0	0	0	0	0.050	0.112	0.173	0.197	0.222	0.246
$N-1$	0	0	0	0	0.035	0.089	0.142	0.159	0.175	0.192	0.208
$N-2$	0	0	0	0.031	0.077	0.124	0.134	0.144	0.154	0.163	0.173
$N-3$	0	0	0.032	0.073	0.114	0.119	0.123	0.128	0.132	0.137	0.141
$N-4$	0	0.035	0.073	0.112	0.113	0.114	0.115	0.117	0.118	0.119	0.084
$N-5$	0.037	0.074	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.074	0.037

**Table 5.11:** Henderson approximation to the Published Asymmetric Weights for  $S_{3 \times 9}$

The gain function for the  $S_{3 \times 9}$  asymmetric seasonal filter (row  $N-3$  in Table 5.11) is shown in Figure 5.18.



**Figure 5.18: Gain Function for  $S_{3 \times 9}$  Term Asymmetric Seasonal Filter (for factor year  $N - 3$ )**

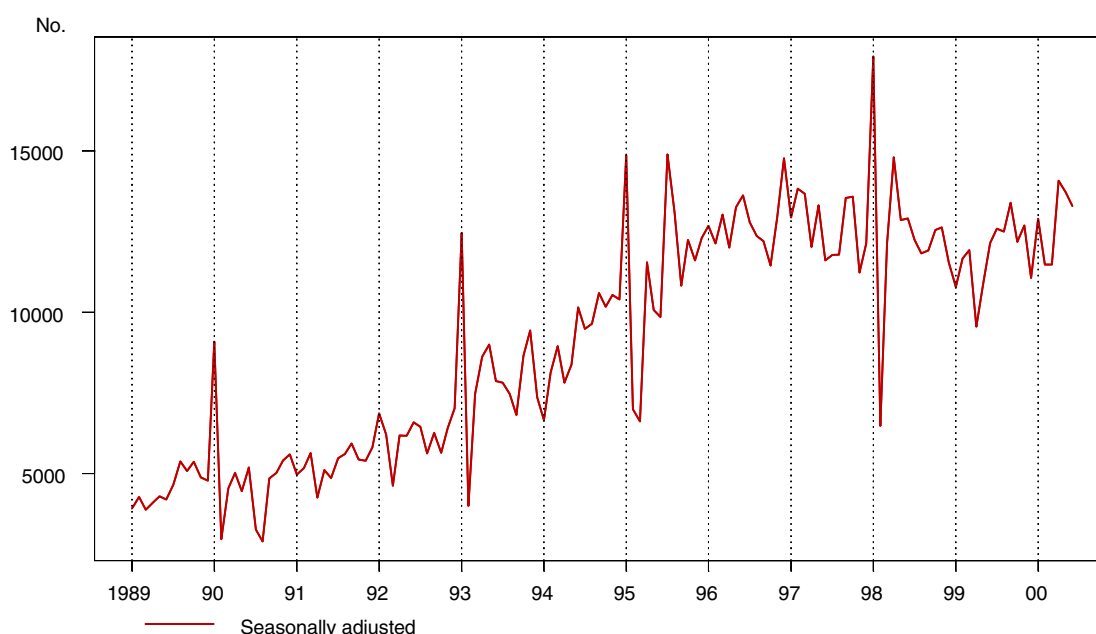
## 6 Issues with seasonal adjustment

The seasonal adjustment packages used by government statistical agencies around the world vary. The Office of National Statistics (ONS), Statistics Canada, the US Bureau of the Census (USBC) and the US Bureau of Labour Statistics (BLS) all use the X11ARIMA or X12ARIMA software package, even though some of the methodological options they choose are different. In the ABS, the standard X11 software has been enhanced with an expert system to produce the SEASABS (SEASonal analysis ABS standard) program.

This chapter examines some of the issues faced by the time series analyst and how they are generally approached within the ABS.

### 6.1 Moving holidays

Moving holidays are holidays that occur at seasonal frequencies but their exact timing shifts systematically each year. Examples of moving holidays include Easter and Chinese New Year. Both holidays follow the cycles of the moon and always occur exactly once a year. Easter generally falls in April but can also fall in late March and affects series such as Tourism because people frequently holiday in this period. Chinese New Year mostly occurs in February but can also occur in January and affects the Overseas Arrivals and Departures series from some Asian countries as people in these countries plan their travel to fit in with holidays in their own country versus holidays in their destination country. Both of these effects have a consistent magnitude and direction (see Figure 6.1).

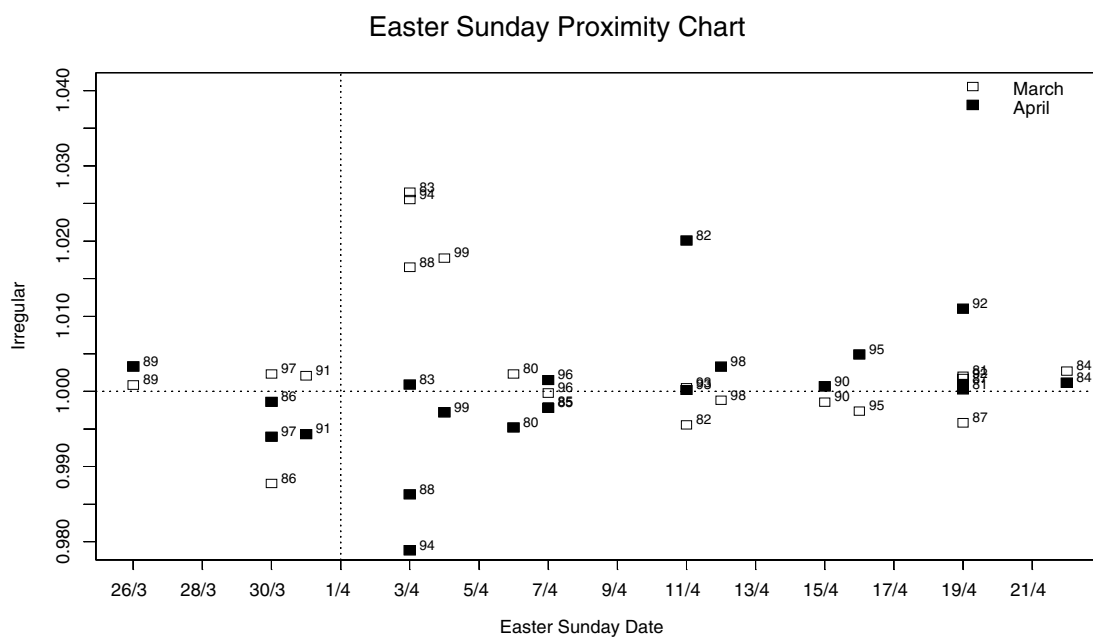


**Figure 6.1: Hong Kong Short Term Arrivals - Chinese New Year Effect**

Unlike Chinese New Year, Easter may effect quarterly ABS series. The influence of these moving holidays can include the loss of working days and changes in economic and social behavior. Such effects are usually country specific, making it difficult to build them into standard routines about the holiday. Each of these effects is estimated by regression techniques and can be corrected for in X11 by using the associated prior correction factors.

A moving holiday effect may also affect a month or quarter other than the one in which the holiday falls if the holiday occurs at either the very beginning or very end of the month or quarter. This is called a proximity effect. Proximity charts can be used to check for the existence, if any, of this effect.

An investigation into the likely magnitude of an Easter proximity effect in Australian Total Retail Turnover was conducted by the Australian Bureau of Statistics in April 1999 (see Leung, McLaren, Zhang (1999)). It was found that the magnitude of the Easter proximity effect for March and April seasonally adjusted estimates was no greater than 1.5%. To illustrate this a concurrent seasonal adjustment was performed using SEASABS on the original Australia Total Retail Turnover series. The proximity chart in Figure 6.2 shows how the irregulars from this concurrent seasonal adjustment for both March and April, from 1980 onwards, are distributed about one as the date of Easter Sunday changes. An Easter proximity effect is observed as the March and April irregular values for 1983, 1988, 1994 and 1999 are not distributed around one. If there was no Easter proximity effect the values would be evenly distributed about one. The labels indicate which year the irregular was observed.



**Figure 6.2: Easter Proximity chart for original Australia Total Retail Turnover - truncated span: January 1980 to April 1999 inclusive.**



## 6.2 Trading day

Trading day is a calendar related effect often found in economic time series. This effect is related to the number each day of the week occurs in any given month. As different days of the week often have different levels of activity in economic populations, the recorded figure for any month may include the effect of having certain extra days.

One way to think of the trading day effect is to consider each month of the year as a block of four sets of each type of day plus one, two or three extra days. If the level of activity in each type of day is constant through the years, the only difference between the same months in different years will be due to the activity on the extra days. For example, the number of working days in successive May's is shown below.

May	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Working days	23	21	21	22	23	23	22	21	21	23	23

This effect is best seen in an example.

May 1993 had an extra Saturday, Sunday and Monday whilst May 1994 had an extra Sunday, Monday and Tuesday. Suppose the level of activity in this particular population was: 10 units produced on Mondays and Fridays, 20 units produced on Tuesday through Thursday and no activity on weekends.

This means that in a week of seven days, activity would be  $10 + 20 + 20 + 20 + 10$  units = 80 units. In May 1993, the extra Saturday, Sunday and Monday would yield an extra  $0+0+10$  units = 10 units of activity, giving a total for May 1993 of  $(80 \text{ units/week} \times 4 \text{ weeks}) + 10 \text{ units} = 330 \text{ units}$ . In May 1994, the extra Sunday, Monday and Tuesday would yield an extra  $0 + 10 + 20$  units = 30 units of activity, giving a total for May 1994 of  $(80 \text{ units/week} \times 4 \text{ weeks}) + 30 \text{ units} = 350 \text{ units}$ . The difference between the two years for the same month is  $350 - 330 \text{ units} = 20 \text{ units}$ , which can be solely attributed to the type of extra days in that month in those years.

There are 7 types of 31 day months, 7 types of 30 day months, 7 types of 29 day months and 1 type of 28 day months. There is a 28 year cycle for the types of months because a leap year comes every four years and the extra day in February in each leap year is different. This means that there are four years times seven different types of days which gives a cycle of 28 years. The existence and isolation of a trading day affect is difficult because the analyst needs around seven years of data to have enough examples of the different types of months. Trading day is usually not analysed in series with less than five years of data.

Trading day effects need to be accounted for because they lead to apparent changes in level of activity when the underlying level is unchanged. In the example above, the actual level of activity in May for 1993 and 1994 was the same, yet the recorded figures indicate increased activity in 1994. The way to allow for this affect is to assign to each day of the week a weight, relative to the level of activity of that day. For a population with equal levels of activity on the five weekdays and no activity on weekends, assign weights of 1.4 to the weekdays and 0.0 to the weekend days. If a multiplicative model is used, the daily weights sum to 7 and the default weights are 1.0 for each day. If an additive model is used, the daily weights sum to 0.

The trading day weights sum to 7 (multiplicative) and 0 (additive) because trading day is, in theory, a conceptually neutral process.

When daily weights have been assigned, each month in a time series is given a weight in proportion to the value obtained by summing the daily weights for that month. These monthly weights are divided by either 30.4735 (365.25/12), the average number of days in a month (if a length of month correction is applied - see below) or by the number of days in each month, i.e., 31, 30 or 29.

In practice, the appropriate daily weights are not obvious. In some cases the weights are contrary to what anecdotal evidence would suggest. For example, in banking industry series, one would expect weights for Saturday and Sunday to be 0.0 or low, as most banks are shut on weekends. However, daily weights derived from regression may be significantly high on Saturday because of the nature of the bank's reporting practices. It is always important to remember that weights derived from a regression analysis use only the information in the data itself, regardless of the population the data is supposed to represent. Often, the nature of this information is heavily dependent on reporting and surveying practices exhibited by both the ABS and the population being measured. It is therefore more accurate to use the statistically derived weights than any derived from general knowledge.

### **Determining trading day**

To determine if a trading day effect is present in a time series, a regression analysis can be performed. This regression is done using the monthly irregular values calculated by X11 and regressing these values against a calendar that contains the number of times each day of the week occurs in each particular month.

The results of the regression analysis give trading day weights for each day of the week and reflect the level of activity for that day.

See Section 7.2 for further details on the calculation of trading day within X11. Ladiray and Quenneville (2001) also describe the calculation of trading day in greater detail.

### **Trading day for quarterly series**

There are three types of quarters in a year. ABS naming conventions have the type of division of months into quarters identified by the name of the last month in that quarter. Hence, March Quarter refers to a year where the first quarter is January, February and March, February Quarter refers to a year where the first quarter is December, January and February and January Quarter refers to a year where the first quarter is November, December, January.

Quarters have different numbers of days. For example, years which have March Quarters and February Quarters only have quarters with 90, 91 or 92 days. Years with January Quarters have quarters with 89, 90 or 92 days. Hence, the maximum difference in the number of days between quarters is 3 in a year with a January Quarter. If the activity on these three extra days is significant, a trading day effect may be identified.

Typically, March Quarter, June Quarter, September Quarter and December Quarter are used most often in the ABS. These quarters have a maximum difference of 2 days. Because of this, a trading day effect is rarely seen in quarterly ABS time series.

### **Length of month effect**

The different months of the year have different lengths: 28, 29, 30 or 31 days. Hence, if June and July have the same levels of activity on the respective days of the week, the total level of activity for July may still be greater than that for June purely because July has an extra day. This effect is called a length of month effect. If a series does not have a trading day correction, then the length of month effect will be accounted for automatically in the seasonal factors. If the series does have a trading day correction, the length of month can still be accounted for in the seasonal factors or alternatively in the trading day factors.

If a length of month correction is included in the prior factors, then it is assumed that this correction will be appropriate and hence the monthly weights obtained in the trading day regression routine will be divided by the average number of days in a month (30.4735). Alternatively, to allow for the length of month in the trading day correction, the monthly weights are divided by the number of days in the respective month. The structure of the length of month correction that appears in the prior factors is a weight for each month depending on the number of days in that month (and hence the factor is one of four different choices).

There can be other systematic calendar related variations included in the seasonal component. For example, Public Service paydays occur as a function of the calendar with most months having exactly two paydays, but occasionally a month may have three paydays. These systematic variations are not strictly seasonal events because they do not occur at the seasonal frequencies but we do attempt to estimate them as part of this component. The occurrence of payday variations can affect series such as finance data.

## **6.3 Extremes (outliers)**

In the ABS, detection of extremes is conducted by both X11 during an adjustment and by the time series analyst.

Extremes need to be detected and corrected because this process improves the estimation of the three main time series components: seasonal, irregular and trend. In particular, extremes are corrected such that they do not distort the trend path. This component is intended as a measure of longer term growth, so it is not desirable for it to respond to a one-off irregular movement.

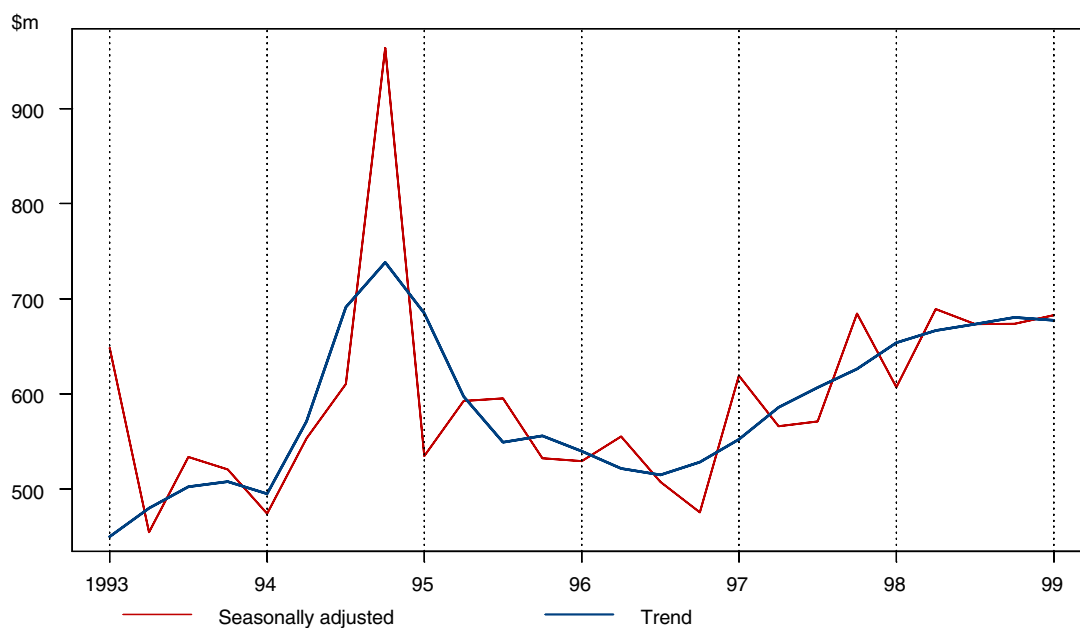
During an adjustment, X11 runs algorithms to look for and correct extreme values in the irregular component. These algorithms run independently of any prior extreme modifications that the user may supply. The detection works by calculating the standard deviations of moving five year periods of the irregular component. The irregular value at the centre of each five year window is compared with a confidence band between 0.0 and 9.9 for these standard deviations. The standard confidence band is 0.0 - 1.5 (indicating not an extreme), 1.5 - 2.5 (indicating near-extreme value) and 2.5 + (indicating extreme value). The confidence band

can be modified by the user. Irregular values found in these confidence bands are given weights of 0.0, between 0.0 and 100.0 and 100.0 respectively. X11 calculates a replacement value if the weight is less than 100.0. Irregular values are automatically replaced by their modified values.

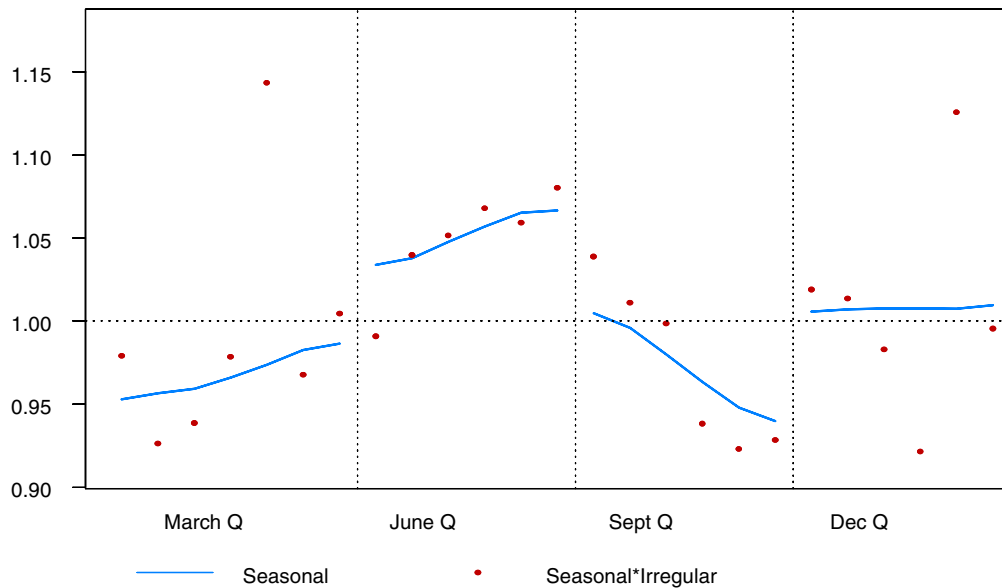
## Detecting Extremes

X11 is not always successful at detecting unusual values and because of this, the analyst should always check for extremes after the adjustment is run. Doing this involves using charts and numerical diagnostics, as well as any anecdotal information that may be available.

The SI chart for the extreme value in Figure 6.3 is shown in Figure 6.4. The extreme value is at 4th quarter 1994. The extreme SI is clearly visible, as it is approximately 70% above the trend seasonality as shown by the factor path. Note that the shoulder values for the extreme, in the 1st and 3rd quarters, also appear extreme. This is because the trend has been distorted upwards towards the 4th quarter value and hence the shoulder values appear to be unusually low compared to the trend.



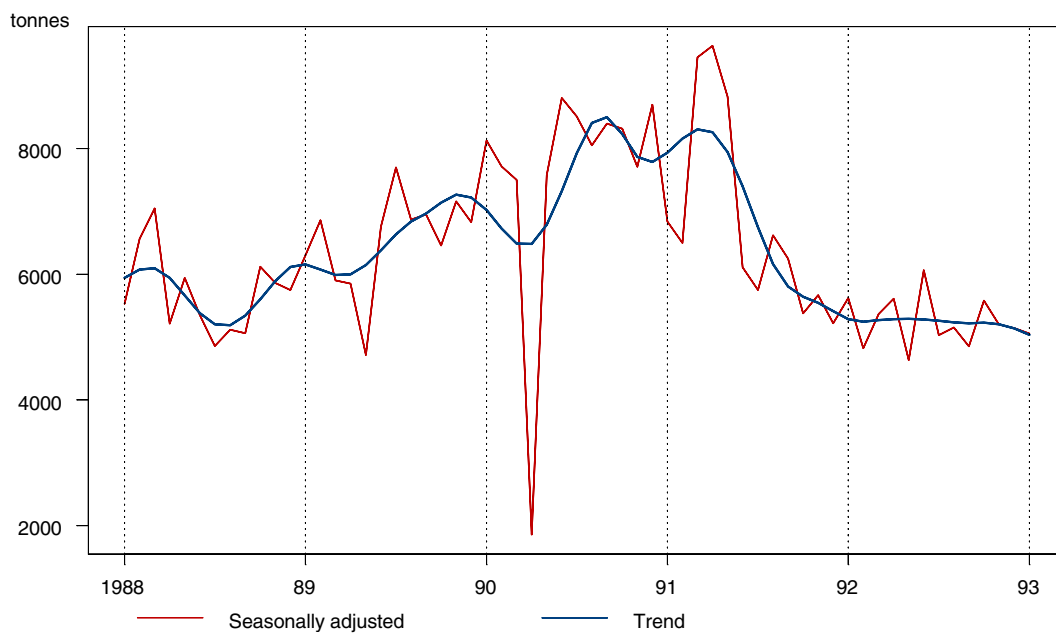
**Figure 6.3: Quarterly Imports of Industrial Transport Equipment - Seasonally Adjusted and Trend**



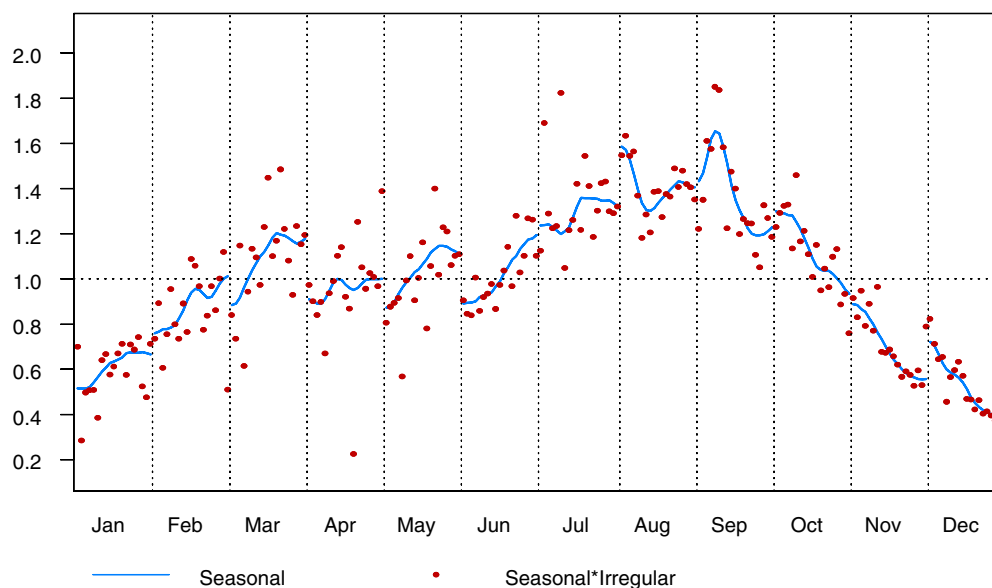
**Figure 6.4: SI chart for Figure 7.3 - 1st quarter 1986 to 1st quarter 1996**

Note, however, that the seasonal factor path in Figure 6.4 does not seem to have been particularly distorted by the 4th quarter extreme.

Figure 6.6 shows an SI chart where the factor path is distorted by a extreme value in April 1992 (the corresponding adjusted and trend series are shown in Figure 6.5).



**Figure 6.5: Receipt of Taxable Wool by Brokers, Queensland - Seasonally Adjusted and Trend**



**Figure 6.6: SI chart for Figure 7.5 - July 1978 to June 1996**

What is apparent in adjusted and trend charts for both these examples above is that the trend is being greatly distorted by the extreme. This is the main reason to correct an extreme. There are several issues to consider before an extreme value can be corrected:

- is the extreme historical?
- is the extreme at the current end?
- do we have a reason for the extreme behaviour in terms of the population it is measuring?
- how large is the extreme?

If the extreme is historical, i.e. occurring at the earlier part of the series, then the analyst may not bother correcting it, especially if the data for that period and surrounding periods is not being published. The reason why the surrounding data span must be considered is that as moving averages are used to estimate the seasonal and trend components, the modification of one value impacts on the estimates of the values around it. How long the impact lasts depends on the length of moving average being used. For example, if a 7 term Henderson is used to estimate the trend component, any trend value (not at either end of the series) is a function of itself and the three values surrounding it on either side. For more information see Chapter 5.

If the extreme is at the current end, and in particular, if it occurs in the span of data that is published, then special consideration needs to be applied to the decision to modify it. Also, as current end estimates of the components of a time series use moving averages with surrogate weights, these estimates are subject to revision, as are the original values themselves. Correcting an extreme can have a large impact in this instance. Finally, a current end extreme may in fact represent the start of a trend break. For this reason the analyst will usually wait for at least two additional data points before correcting for an extreme. Usually, an analyst

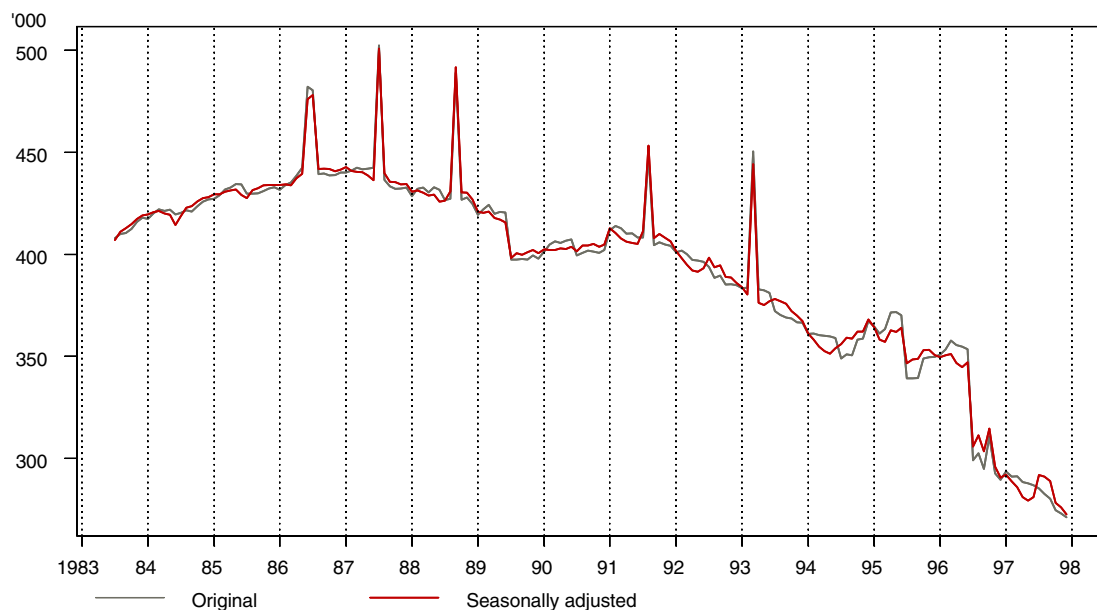
will not modify an extreme at the current end unless they have a substantial reason for the behaviour.

### Querying an extreme

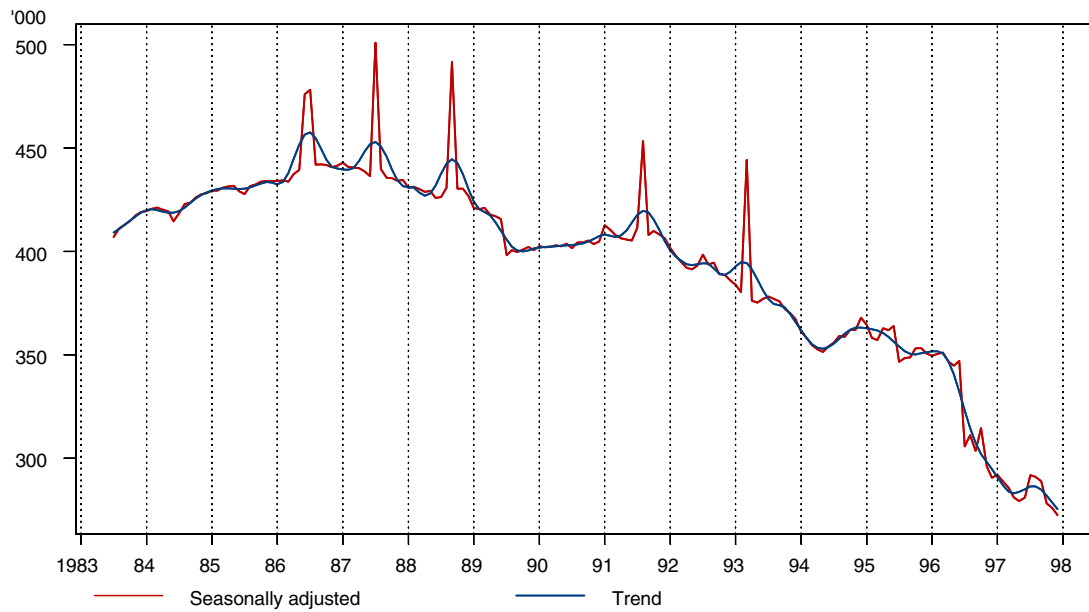
The extremes in Figures 6.7 - 6.9 occurred for the following reasons:

- June 1986 - Census
- July 1986 - Census
- July 1987 - Federal Election
- September 1988 - Referendum
- August 1991 - Census
- March 1993 - Federal Election

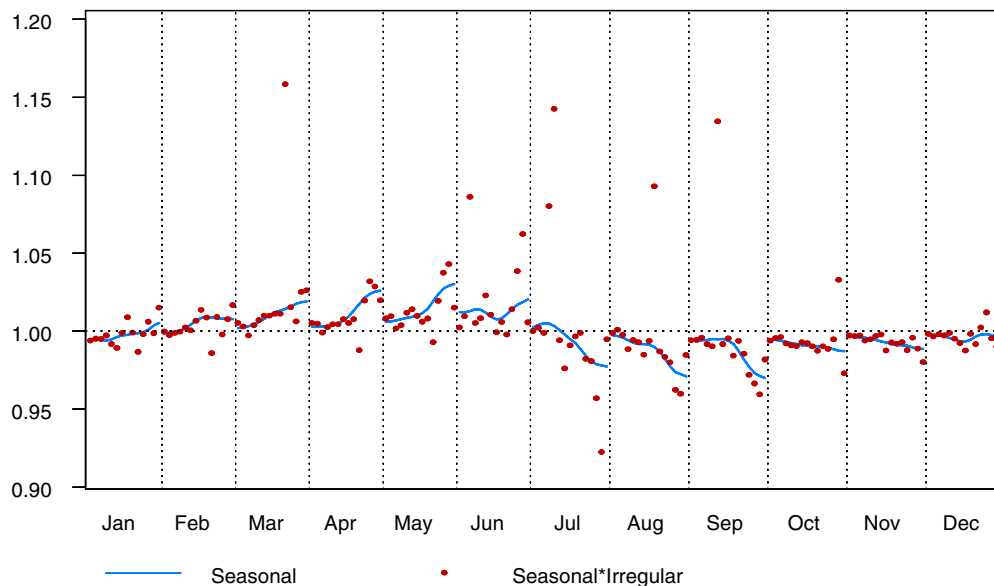
They are suitable reasons because they detail the timing, nature and the behaviour that these events would bring about, namely an increase in employment in the public sector around the country. This matches the behaviour that is observed in the time series. Extremes are usually only corrected if anecdotal evidence such as this is supplied (commonly by the subject matter area).



**Figure 6.7: Monthly Number of wage and salary earners in Commonwealth Government sector - Original and Seasonally Adjusted**



**Figure 6.8: Monthly Number of wage and salary earners in Commonwealth Government sector - Seasonally Adjusted and Trend**



**Figure 6.9: Seasonal-Irregular (SI) Chart for Figure 6.8**

There are instances where extremes are corrected without a reason being given. For example, if a consistent moving holiday affect is observed, such as Easter. However, the analyst still

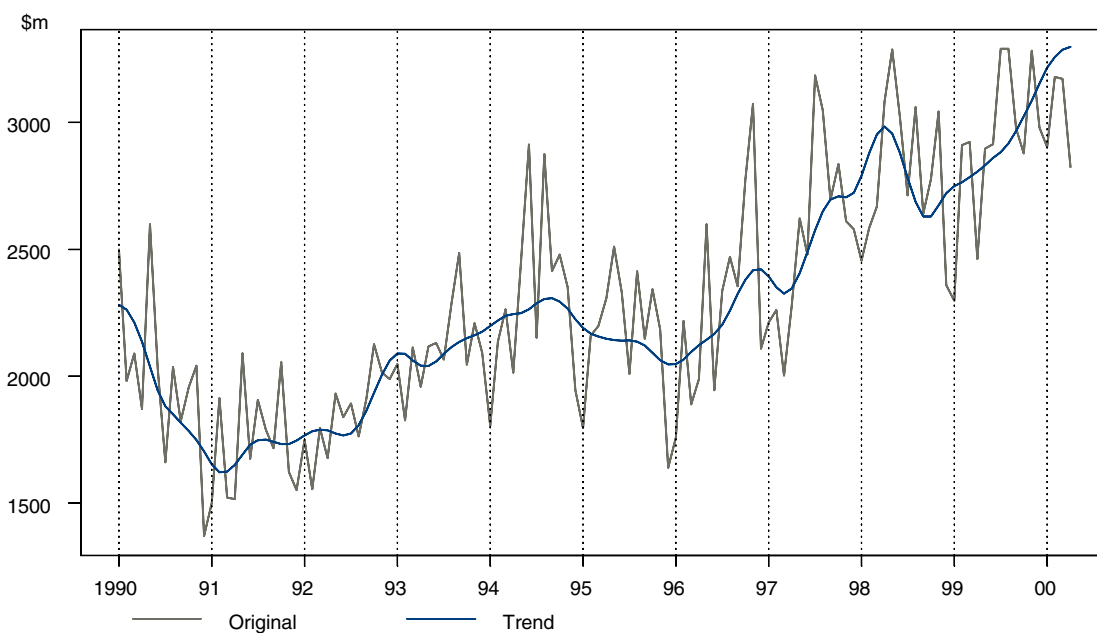


needs to ensure that such a holiday having an affect makes sense. Chinese New Year effects would be expected in overseas travel series because of increased movement due to celebrations, but it would not make much sense in a series about milk production.

Extremes are sometimes also modified without a reason if they are historical or so large that the quality of the adjustment would be greatly reduced if they were left uncorrected. However, the analyst should always make every effort to determine if there is a reason behind the observed behaviour, and think carefully before modifying an extreme for which no evidence is available.

### Calculation of a correction factor

If the extreme is modified, the analyst needs to work out a correction factor. Sometimes an indication of this is supplied by the subject matter area. For example, in the case of Australian Building Approvals (Figure 6.10), one of the extremes is caused by approval of the Olympic Stadium and this value is available from the subject matter area. If this happens, the value supplied is taken away from the current original value to have an idea of what the raw data would have looked like without the irregular.



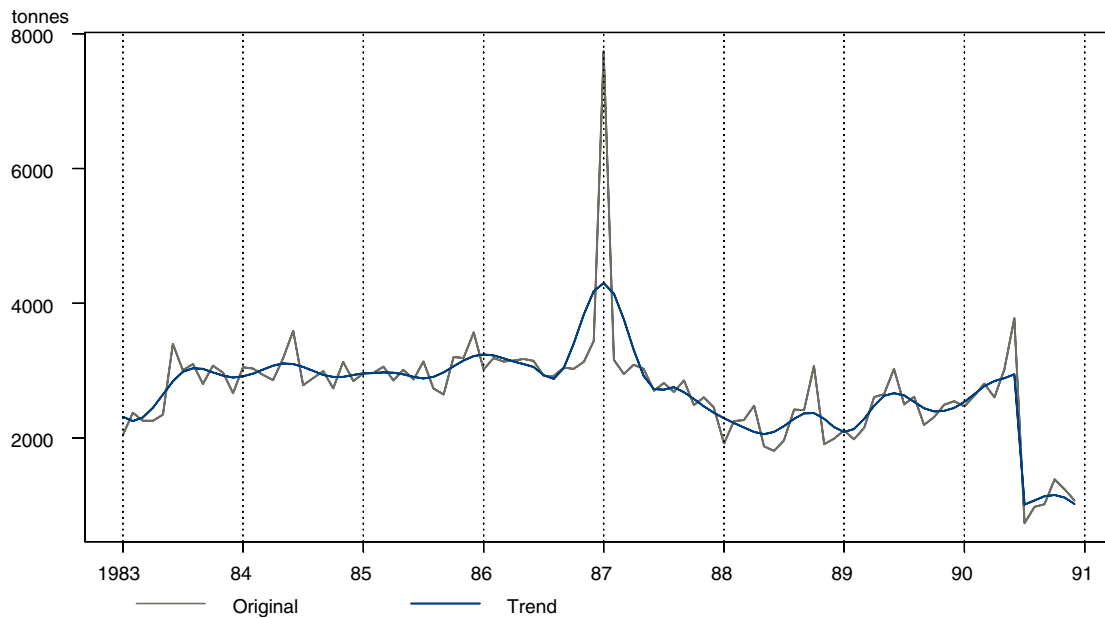
**Figure 6.10: Monthly Value of Building and Construction Approvals - Original and Trend**

Usually no such figure is available and the analyst has to use personal judgement to calculate a correction. The way this is usually done is as follows:

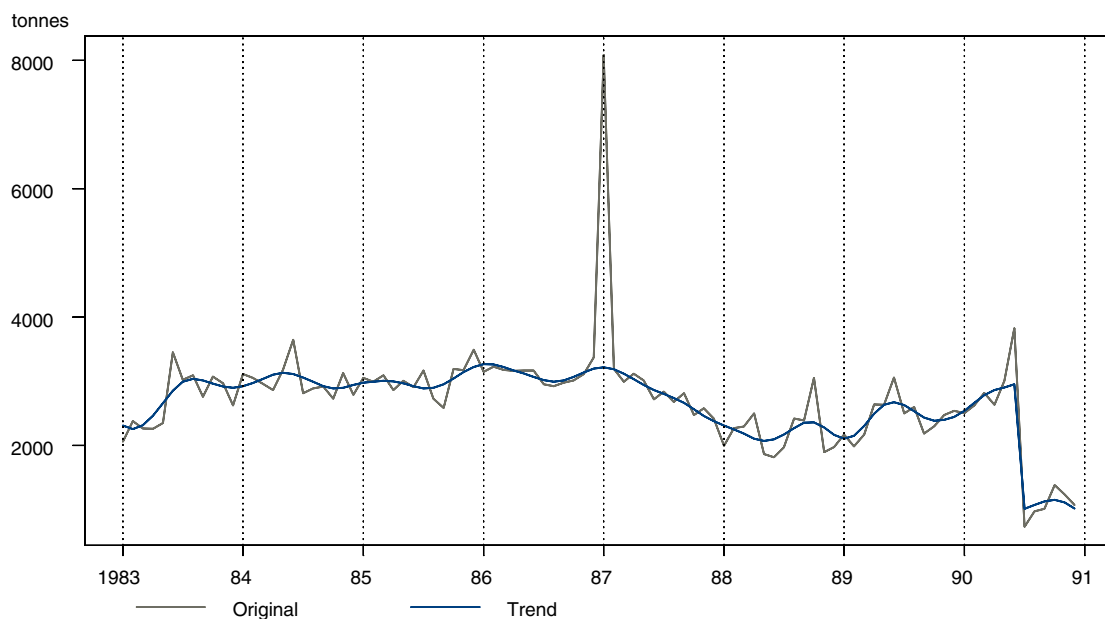
1. Make an estimate of the seasonally adjusted value at the time point of the extreme in question (call this  $sa$ )
2. Make an estimate of the where the trend value should be at the time point of the extreme in question (call this  $t$ )
3. Derive the factor  $k$ , where  $k = sa/t$ .

This method works because, in the multiplicative case, dividing the seasonally adjusted component by the trend leaves the irregular component.

Applying the above method for the extreme shown in Figure 6.11, estimates for the seasonally adjusted and trend values are 52,500 and 13,750 respectively. This gives  $k = 3.82$ . This factor was used to produce the modified extreme in Figure 6.12.



**Figure 6.11: Receipt of Taxable Wool by Brokers, Victoria - Without an Extreme Correction**



**Figure 6.12: Receipt of Taxable Wool by Brokers, Victoria - With an Extreme Correction**

Another method is to estimate the seasonal (using SI charts) and desired trend components and then remove these from the original series to obtain an estimate of the correction factor.

For an additive series, the method of calculating a factor is the same as described above, except the trend value is subtracted from the adjusted value as opposed to divided from.

## 6.4 Designing a survey for trend or seasonally adjusted estimates

### What do users really want?

Users of official statistics are often interested in the current direction of a time series. Most of the repeated sample surveys currently in use have been designed to give good estimates of the level and/or movements in the original and seasonally adjusted series. The month-to-month movement of the original and seasonally adjusted series essentially gives a rough indication of the current direction, or the trend, of the series. However, the current direction of the series can be more appropriately assessed by specifically designing the repeated survey to detect the trend. The ABS prefers users to focus on trend estimates rather than original or seasonally adjusted estimates.

Repeated sample surveys are usually based on samples that have a degree of overlap from period. This is done to reduce costs and the standard errors associated with the estimates of change between two consecutive time periods. For example, Binder and Hidioglou (1988) p. 187 state:

*"The level of overlap between occasions will depend on the objectives of the survey, which include measures of level and change, as well as operational constraints ..."*

The issue of then designing a repeated sample survey for trend estimates is a natural question.

The sample design, in particular the rotation pattern, can be chosen by (at least) two different criteria:

- minimising the sampling variance of an estimate (this could be the original/seasonally adjusted/trend)
- reducing the revisions of an estimate (this could be the original/seasonally adjusted/trend)

The issue of designing repeated surveys for seasonally adjusted and trend estimates has been considered by Sutcliffe and Lee (1995), McLaren (1999), Bell (1999). A brief summary of some of the main issues and results from McLaren (1999) are given below.

## Background: Rotation patterns

Rotation patterns vary in terms of the number of times a unit, for example a dwelling in the case of a household survey, is included in the survey and the time interval between inclusions.

The sample overlap at different lags for a range of rotation patterns are as follows:

*in-for-m*: This pattern leads to an overlap of  $1 - s/m$  between samples  $s$  months apart, for  $s = 1, \dots, m - 1$  and no overlap for months  $m$  or more apart. Unless  $m$  exceeds 12 there will be no overlap for months a year apart. Australia uses  $m = 8$  and Canada uses  $m = 6$  for their monthly population labour force survey

1-2-1( $m$ ): This pattern is repeated until dwellings are included for a total of  $m$  times. This leads to no sample overlap between months one or two months apart, but an overlap of  $1 - s/(3m)$  for  $s = 3, 6, \dots, 3m$ . The overlap between months a year apart is  $1 - 4/m$  provided  $m$  is 5 or more.

4-8-4(8): This leads to an overlap of  $1 - s/4$  for months  $s$  months apart,  $s = 1, 2, 3$ . For  $s = 12$  the overlap factor is  $4/8$ . In fact the overlap is  $4/8 - |s - 12|/8$  for  $s = 9, \dots, 15$ . There is no overlap for  $s = 4, \dots, 8$ .

2-10-2(4): This leads to an overlap of  $1/2$  for adjacent months. For  $s = 12$  the overlap factor is also  $1/2$ .

These rotation patterns are special cases of a general class of rotation patterns in which selected dwellings are included for  $a$  consecutive months, removed from the survey for  $b$  months and then included again for a further  $a$  months. The pattern is repeated so that dwellings are included for a total of  $m$  occasions. These rotation patterns can be denoted  $a - b - a(m)$ . Setting  $b = 0$  gives an *in-for-m* rotation pattern.

## Designing a survey for sampling variance

Consider  $y_t = (y_{t-r}, \dots, y_{t+r})$  as a vector of length  $n = 2r + 1$  centred at time  $t$ . Let  $y_t$  be a time series of survey estimates and  $Y_t$  be the true unknown values.

Using sampling to obtain a time series means that the value obtained at each time period has a component of variability attributed to the sampling error. That is,

$$y_t = Y_t + e_t$$

For many series the sampling error will be a major source of variability. The use of rotation patterns means that the sampling error component will have a complex structure and be correlated over time.

The sampling variance of the original series is denoted by

$$V_p(y_t) = V(y_t|Y_t)$$

It is the sampling variance that is affected by an alteration the sample design. This is because the rotation pattern used in the survey will affect the structure of the sampling error, which will then affect the sampling variance.

We can consider how the rotation patterns impact on filtered values, such as the seasonally adjusted and trend estimates. Consider a linear filter which is used to obtain filtered values from  $y_t$  by applying a vector of weights  $w_t$  to give the filtered value at time  $t$

$$f(y_t) = w_t' y_t$$

The sampling variance of the filtered value at time  $t$  is

$$V_p(f(y_t)) = w_t' V(y_t|Y_t) w_t$$

which is conditional on the values of the true unknown series,  $Y_t$ , and the filter  $w_t$  depends on the time period for which the value  $f(y_t)$  refers. The filter  $w_t$  could represent an original estimate where  $w_t = (0, \dots, 0, 1, 0, \dots, 0)$ , a seasonally adjusted estimate where  $w_t$  is an appropriate choice, or a trend estimate where  $w_t$  is an appropriate choice. The filter  $w_t$  could represent an initial estimate or a final estimate of either the seasonally adjusted or trend.

To assess the impact of different rotation patterns and filters, consider filters designed for use at the end of the series. A summary of the effect of different rotation patterns on the level of seasonally adjusted and trend estimates and also the one month difference between seasonally adjusted and trend estimates is shown in Figure 6.13 (reproduced from McLaren, 1999). The different rotation patterns considered are: 4-8-4(8) (used in the United States for the Labour Force Survey), 2-10-2(4) (used in Japan for the Labour Force Survey), 2-2-2(8), 1-2-1(5), 1-2-1(8), 1-1-1(6) and a range of *in-for-m* ( $m=1$  to 30) rotation patterns.

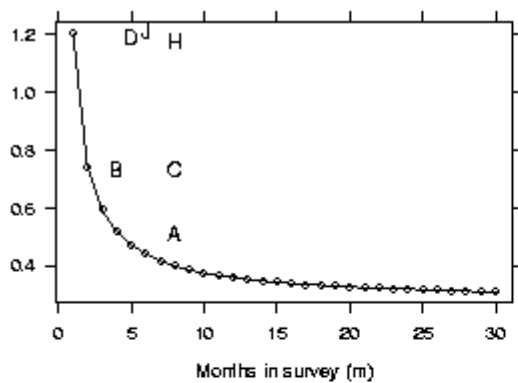


Figure 1a) One month difference in seasonally adjusted

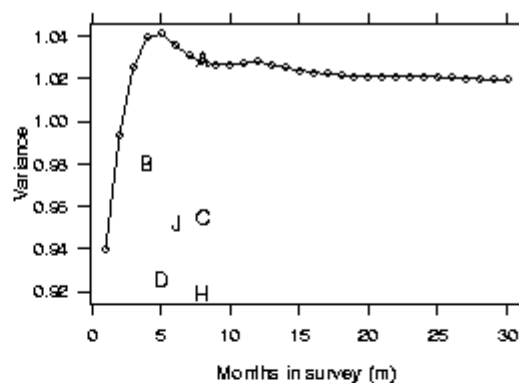


Figure 1b) Seasonally adjusted level

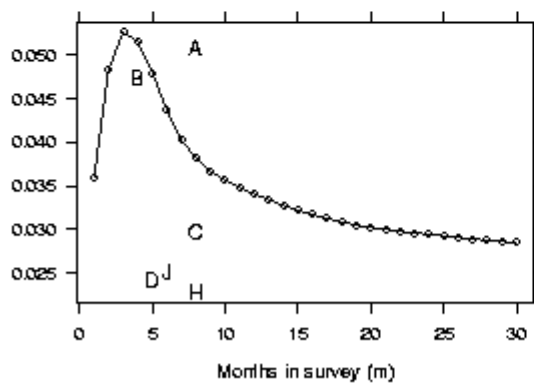


Figure 1c) One month difference in trend

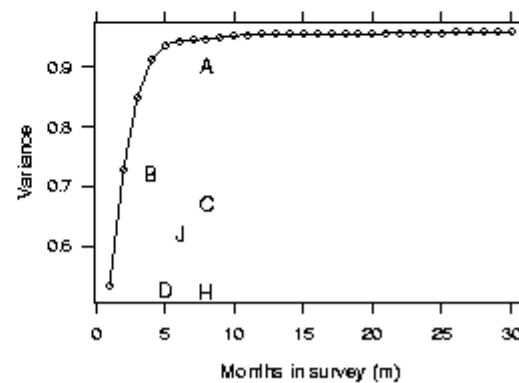


Figure 1d) Trend level

**Figure 6.13: Ratio of the sampling variance to the variance of the original series at the end for chosen rotation patterns for a standard X11 for the Australian Labor Force Series employment where A=4-8-4(8), B=2-10-2(4), C=2-2-2(8), D=1-2-1(5), H=1-2-1(8), J=1-1-1(6). (reproduced from McLaren, 1999).**

The key point from Figure 6.13 is that a low variance is good. Figure 6.13 shows that the *in-for-m* rotation patterns (given by the solid line) are sensible if the one month change in seasonally adjusted estimates are the key statistic to be analysed. If the trend is the key statistic to be analysed, then quite different rotation patterns (1-2-1(*m*)) type should then be used. Australia uses an *in-for-8* rotation pattern for the Australian Monthly Labour Force survey.

### Designing a survey for revision of trend estimates

Given that users are interested in trend estimates, an alternative criteria which can be assessed is minimising the revisions of trend estimates.

Revision of an estimate occurs when an initial estimate is made based on the current data available. When additional information becomes available this leads to a new estimate for each previous time period which is now based on this additional information. The difference between the preliminary estimate and the new estimate for a given time period is the revision.

Users of trend estimates observe and are concerned about the revision of trend estimates, particularly as data up to the next three time points are added to the series. One approach to reducing the revision of trend estimates at the end of a series is to use trend filters which minimise the difference between trend estimates at two different time points. One example of this approach is to use ARIMA models to predict values at the end of a series. An alternative approach is to change the survey design through changing the rotation pattern and see what effect this has on the revision of estimates of the trend.

The revision between a trend estimate made at time  $t$  using information up to time  $t$ ,  $T_{t|t}$ , and a trend estimate made at time  $t+k$  using information up to time  $t$ ,  $T_{t|t+k}$  can be calculated as

$$\begin{aligned}\widehat{T}_{t|t} - \widehat{T}_{t|t+k} &= a'_{t|t} y_t - a'_{t|t+k} y_{t+k} \\ &= (a^*_{t|t} - a_{t|t+k})' y_{t+k} \\ &= b'_{t|t+k} y_{t+k}\end{aligned}\tag{6.1}$$

where  $y_{t+k}$  is a vector of observed values ending  $k$  terms on from  $t$ , the filter  $a_{t|t}$  is a trend filter,  $a^*_{t|t} = (a'_{t|t}, 0, \dots, 0)'$  contains  $k$  zeros and  $b_{t|t+k} = a^*_{t|t} - a_{t|t+k}$  is the trend revision filter. In practice, there will become a point where the degree of revision will be negligible and a minimal difference will be observed. This usually occurs after the addition of three time points ie. between a trend estimate  $\widehat{T}_{t|t+3}$  and trend estimates up to the historical trend,  $\widehat{T}_{t|\infty}$ .

### Conditional on the Population: Sampling Variance

Consider the mean square error of the revision between two trend estimates at time  $t$  with true unknown values  $Y_t$  as fixed. Conditioning on the vector  $Y_{t+k} = (Y_1, \dots, Y_{t+k})'$  gives the revision mean square error

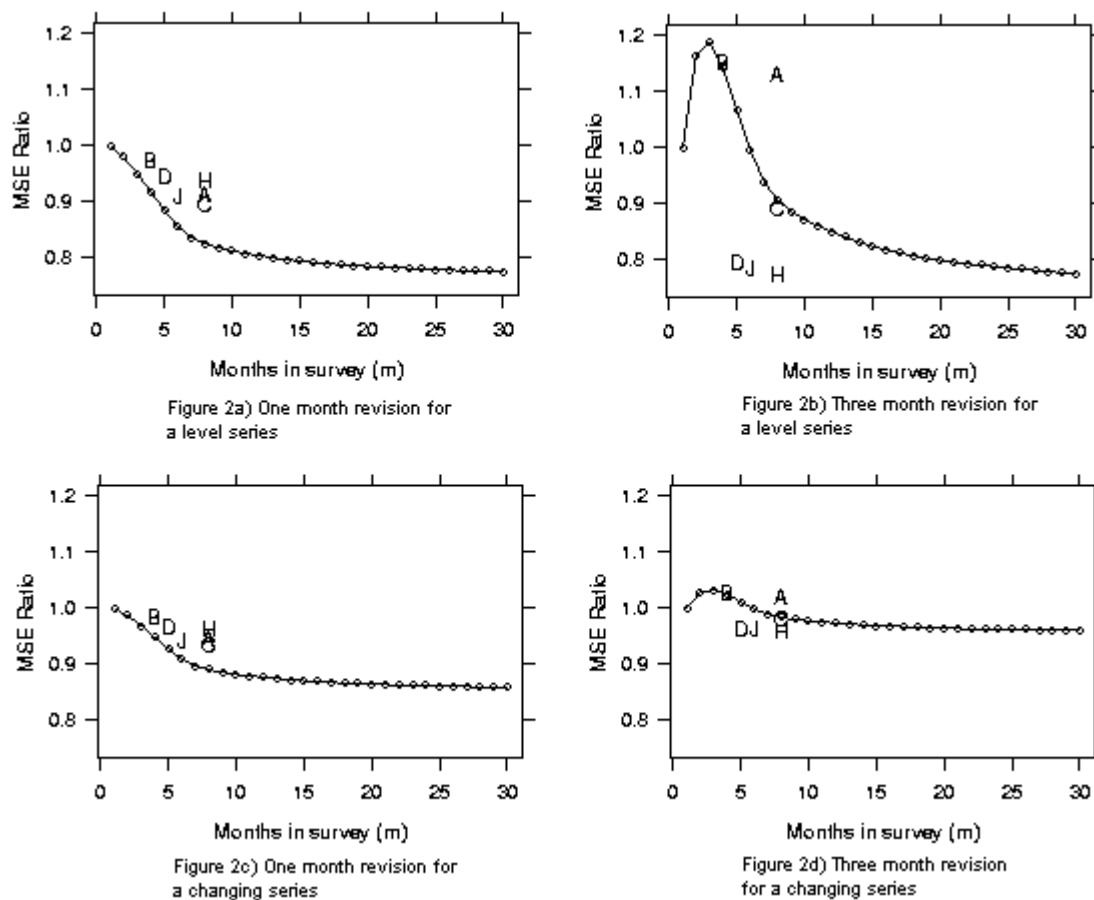
$$\begin{aligned}R^c_{t|t+k} &= E_p \{ (\widehat{T}_{t|t} - \widehat{T}_{t|t+k})^2 \} \\ &= V_p(\widehat{T}_{t|t} - \widehat{T}_{t|t+k}) + \{ E_p(\widehat{T}_{t|t} - \widehat{T}_{t|t+k}) \}^2 \\ &= b'_{t|t+k} V_p(y_{t+k}) b_{t|t+k} + \{ E_p(b'_{t|t+k} y_{t+k}) \}^2 \\ &= b'_{t|t+k} V_p(y_{t+k}) b_{t|t+k} + (b'_{t|t+k} Y_{t+k})^2\end{aligned}\tag{6.2}$$

where  $b_{t|t+k} = a^*_{t|t} - a_{t|t+k}$  with  $a^*_{t|t} = (a'_{t|t}, 0, \dots, 0)'$  having  $k$  zeros and assuming  $E_p(y_t) = Y_t$ .

The first term in (6.2) can be thought of as a design component as the sampling variance  $V_p(y_{t+k}) = V(y_{t+k}|Y_{t+k})$ . The sampling variance will depend on the correlation structure of the series of sample estimates  $y_t$  and hence can be affected by the rotation pattern used. The second term in (6.2) can be thought of as a signal component and is dependent on the filters chosen and the true underlying series. The second term does not depend on the sampling variance and hence is independent of the rotation pattern and other features of the sample design, such as sample size. The second term acts as a level shift and will reduce the impact of using different rotation patterns.

The magnitude of the signal component relative to the design component is important when looking at the mean square error of the revision. The relative contribution of each term will depend on the sample size. As sample size decreases the variance increases, so the contribution of the variance component becomes more important.

Figure 6.14 (reproduced from McLaren, 1999) shows how the revision of trend estimates at different lags, for different points in the series, is affected by the rotation pattern. The results have been calculated for periods in the original series which are approximately level and also periods in the original series which are changing rapidly, ie. A turning point. This is now referred to as a “changing series”.



**Figure 6.14. Ratio of the mean square error of the revision for chosen rotation patterns divided by the mean square error of the revision for an independent design under standard X11 for the variable employment were A=4-8-4(8), B=2-10-2(4), C=2-2-2(8), D=1-1-1(5), H=1-2-1(8), J=1-2-1(5). (reproduced from McLaren, 1999).**

Figures 6.14-2a) and 6.14-2c) show that for the one month revision in either a level or a changing series, the currently used rotation patterns perform well in reducing the variance of the one month revision. In contrast, Figures 6.14-2b) and 6.14-2d) show that for a level series the 1-2-1( $m$ ) type rotation patterns actually outperform the currently used rotation patterns for reducing the variance of the three month revision. This is also true for a changing series (Figure 6.14-2d)) but not to the same extent. Remember, Australia uses an *in-for-8* rotation pattern for the Australian Monthly Labour Force survey.



See McLaren (1999) for results for the case where the variance is unconditional on the population values.

### **Summary: The key points**

Users needs should be considered when designing a repeated sample survey. The choice of rotation pattern used in the repeated sample survey should depend on the estimates that the users consider the most important.

There is a trade off between designing a repeated sample survey for seasonally adjusted and trend estimates. A good rotation pattern for minimising the variance of trend estimates may not be good for the variance of the seasonally adjusted.

## **6.5 Indirect (aggregate) versus Direct (disaggregate)**

There are two ways to seasonally adjust time series that are related in an aggregative way. The first method is called the indirect or aggregative method while the second method is the direct or disaggregate method. That is

1. seasonally adjust all of the lower component series and sum up all of the seasonally adjusted values and consider this the seasonally adjusted series for the total (indirect (aggregate))
2. sum all of the original data from all the lower series to form a total and seasonally adjust this total directly (direct (disaggregate))

For a detailed discussion on the issues of aggregation see Cannon (2000).

For example, consider a set of series for which there is original data for each state in Australia and that all of the original data for a set of series sums to the Australian total. After each state is adjusted directly (disaggregatively) in the normal way then the Australian total can be seasonally adjusted either indirectly or directly. In this case a number of different variables could be used to sum to an Australian total:

- by state
- Male/female
- age groups
- education levels
- type of good manufactured
- working full time/part time
- type of dwelling approved

It is also possible for the total to be defined as being the sum of only some of the components. For example, the ANZ Job Advertisements are derived by counting the number of job vacancy advertisements for a month in the major newspapers of each capital city in Australia except Darwin. This yields an average number of job advertisements per week. An Australian total for this series is published, by summing the values for the 7 capital cities in scope. While

this is labelled Australian Job Advertisements, it does not cover all of Australia, as it omits Darwin, non-capital cities, rural areas and smaller newspapers. In this instance, the Australia total is adjusted aggregatively. A similar example is the Number of Chickens Slaughtered series. The Australia level series is the sum of the series for New South Wales, Victoria, Queensland, South Australia and Western Australia. The other states are not included because their population is too small and their values are confidentialised. The Australian level series is adjusted aggregatively as the sum of the adjusted series for the above states.

Intuitively, it may appear that seasonally adjusting series at a component level will lead to improvements in the adjustment as more information about the data at hand is available. For example, one or more components may be non-seasonal, while other components could have reasonably significant seasonal patterns. In instances such as these, various problems may be encountered in performing a direct analysis of the total level series and treatment of these problems may be better handled at a component level. Empirical studies suggest this is not necessarily so. In some cases, disaggregation of an aggregate series into component series yields component series which are subject to substantially higher levels of irregularity than the aggregate series. This could lead to a greater margin of error in the estimation of the seasonal components.

If the behaviour of the various component series is significantly different a direct seasonal adjustment of the total is not recommended. If the components have very similar behaviour, then using the same method of analysis on the components and the total is appropriate. In these instances, differences between indirect and direct adjustments are often negligible. By chance, the aggregate series may exhibit similar abnormal movements in the same period of consecutive years, suggesting a possible seasonal break. However, disaggregation may reveal that these are irregular movements which affect different components. Similarly, component series may exhibit extremes which cancel each other out in the aggregate series. Ideally, corrections or specific options that are applied to one or more component series are applied to their corresponding aggregate series.

Often it is desirable by users that the total seasonally adjusted series is equal to the sum of the seasonally adjusted components. In this instance, the analyst has two possible approaches: seasonally adjust the series aggregatively or adjust the series disaggregatively and apportion any differences between the total and the sum of the components across those components. This latter method is used for Australian Total Retail Turnover for which Australia level is split by state and industry. The existence of differences between the aggregate series and the components when using indirect adjustment highlights that seasonal adjustment is an estimation process. One reason the sum of component adjusted series may not equal a direct adjustment of the aggregate series is related to span. Finer level components often have shorter spans than total series as particular levels of disaggregation may be requested as time goes on. The longer the span of a series, the more information it contains, especially with regard to features such as seasonal patterns and trading day. Hence we often get more accurate estimates of behaviour for total series than components, leading to discrepancies in aggregated totals.

## When to Use Aggregative/Disaggregative

There are several factors that need to be considered when deciding between an aggregative and a disaggregative adjustment.

The first thing to check before using an indirect adjustment is that the original total series is derived as the sum of the original series components. For example, if there are component series for age groups 15-19 and 20-39 only, but the total series is for all age groups, then the total must be adjusted directly. If data for a population was available for both males and females, then since these components cover the population, the total series could be adjusted indirectly. However, care should be taken in circumstances where the original components are obtained in a different manner to the original totals.

Component series may not be available for many reasons. In some cases they are too volatile or their population too small. For example, many series for the smaller states of Australia: Tasmania, Northern Territory and ACT, contain very low levels of activity or even zero values. Series with zero values would require additive or pseudo-additive adjustment and this will not be compatible with other states which are adjusted multiplicatively. It may also be the case that finer level data is not available at the same time as the aggregate series. The analyst needs to determine whether the effort involved in producing disaggregate series is outweighed by any gains in the estimation process.

## Methods of aggregation

Aggregating series can be done using seasonally adjusted or trend series but not combined seasonal factors. That is, it is not wise to sum the combined seasonal factors for each industry series to produce a combined seasonal factor for total across all industries. The reason behind this can be shown in a simplified algebraic manner. Suppose we have three time series:  $O$ ,  $O'$ ,  $O''$  which fit under the simple multiplicative model

$$\begin{aligned} O &= S \times T \times I \Rightarrow O/S = T \times I \\ O' &= S' \times T' \times I' \Rightarrow O'/S' = T' \times I' \\ O'' &= O + O' \end{aligned}$$

Then the method of calculating the seasonally adjusted series for the original series  $O''$  is

$$\left(\frac{O}{S}\right) + \left(\frac{O'}{S'}\right) = \left(\frac{OS' + O'S}{SS'}\right)$$

This is not necessarily equal to

$$\left(\frac{O''}{S+S'}\right) = \left(\frac{O+O'}{S+S'}\right)$$

which is using the sum of the seasonal factors to obtain the seasonally adjusted series for the original series  $O''$ . Hence, using this latter method is not robust, especially not in a world where time series do not usually fit a simple model.

## 6.6 Forward Factors versus Concurrent Adjustment

There are two approaches to deriving seasonal and trading day factors.

1. Forward factors rely on an annual analysis of the latest available data to determine seasonal and trading day factors that will be applied in the forthcoming 4 or 12 months (depending if the series is quarterly or monthly). For example for a monthly series under the forward factor method, data up to the June 1999 reference month is used to calculate seasonal and trading day factors which will be used to obtain seasonally adjusted estimates for the July 1999 to June 2000 period. Data up to June 2000 is used to calculate seasonal and trading day factors for application to the July 2000 to June 2001 period, and so on. This method is known as a forward factor adjustment.
2. Concurrent adjustment uses the data available at each reference period to re-estimate seasonal and trading day factors. Under this method data for the current month are used in estimating seasonal and trading day factors for the current and previous months. This method continually fine tunes the estimates whenever new data becomes available. For example, data up to the July 2000 reference month are used in calculating seasonal and trading day factors which will be used to obtain seasonally adjusted estimates for the period up to July 2000, data up to and including the August reference month will be used to obtain seasonally adjusted estimates up to August 2000, and so on. This method is known as a concurrent adjustment.

The concurrent adjustment method is a relatively computationally intensive seasonal adjustment method compared with the forward factor method. However technological advances have now made it possible to apply the concurrent adjustment method with little operational burden or cost and at considerable speed (ABS, 1999b). Under concurrent adjustment, the seasonal factors will also be more responsive to underlying dynamic changes than forward factors.

### Using forward factors

For a monthly series, it is recommended by TSA in the ABS that a preliminary analysis is performed when 11 months of data exists from the last analysis. This allows any problems with the series which have arisen in the last 11 months to be corrected to aid the timeliness of the results from the annual analysis. An annual analysis should then be performed when an entire extra year of data is available. The factors calculated at this time are consistent with a forward factor adjustment.

When analysing quarterly series using forward factor adjustment, a preliminary analysis is not required since three months exists between publication dates for the analysis to take place.

### Using concurrent adjustment

The reanalysis cycle for concurrent adjustment consists of an annual analysis and then monthly adjustments where the seasonality and moving trading day are re-estimated by including the most recent month's data. Following from this, a new set of forward factors is calculated and the appropriate month's factor is applied.

For example, consider a monthly series whose annual analysis is performed in June each year. The first step is to perform the annual analysis, calculate a set of forward factors and apply the June factor to the June data. When the July data becomes available, a new set of factors is calculated on the data up to and including July. In this instance only the seasonal component and the moving trading day are re-estimated, the remainder of the parameters remain as they were calculated during the June analysis. This means that structural breaks and the length of the moving averages are not examined between annual analyses. The July factor from the new set of forward factors is applied. In August the process is repeated with a new set of adjustment factors calculated and the appropriate one applied, only this time on the data up to and including August. Therefore, this process causes the adjustment to be more closely related to the available data.

For a quarterly series the cycle remains almost identical to that for a monthly series with an annual analysis and then concurrent adjustments each quarter as the new data becomes available.

### **Emergency seasonal adjustment procedures for concurrent adjustment**

On the rare occasion when a set of concurrent adjustment factors are deemed to be inappropriate for use because they have been distorted by some recent or current behavior in the original series, it is sometimes necessary to perform a concurrent analysis which automatically reassesses the parameters. A concurrent adjustment is then used using the settings derived at the concurrent analysis.

Another option is to revert to a set of seasonal adjustment factors from an earlier period, specified by the analyst. In SEASABS only the set of forward factors calculated during the annual analysis are saved to the series knowledge for concurrent adjustment, the factors applied at each TS Update are overwritten each month/quarter. Therefore, the analyst cannot simply return to a saved version of the forward factors from an earlier period. Instead it is necessary to alter the end date of the series so that the last month/quarter is that from which the analyst wishes to use the factors. Next a concurrent adjustment must be performed and forced to an annual analysis. Finally the analyst must change the Adjustment Type within the TS Update Facility Build menu so that the series is updated using forward factor adjustment rather than concurrent for the next TS Update. This set of previous forward factors can be applied until it is considered suitable to revert to concurrent seasonal adjustment, or until an annual analysis is performed, whichever occurs sooner.

## 7 The mysterious X11

The ratio-to-moving average technique of seasonal adjustment was originally devised in the 1920s by Frederick R. Macaulay, at the National Bureau of Economic Research USA. Refinements of this method, and completely new methods, have since been developed. However, it was not until the advent of the electronic computer that relatively sophisticated methods could be applied to large numbers of series. Since then there has been a steady stream of seasonal adjustment techniques developed, including regression techniques, spectral analysis, the Box-Jenkins (1970) approach and X11 (Shiskin *et. al* 1967). X11 was developed by the US Bureau of the Census and began operation in the United States in 1965. It was soon adopted by the ABS and has been used with improvements specific to the ABS, ever since.

The X11 method is based on the standard “ratio to moving average” seasonal adjustment procedure of Macaulay (1931). This consists of the following steps:

1. Estimate the trend by a moving average
2. Remove the trend leaving a seasonal and irregular component
3. Estimate the seasonal component using moving averages.

Seasonality cannot be identified until the trend is known and an estimate of the trend cannot be found until the series has been seasonally adjusted. Hence an iterative approach is needed.

The first version of X11 (Census I) was developed in 1954 on the first commercial computer ever available. One of the first electronic seasonal adjustments was shown to the President of the United States, the story told was:

“I have an appointment with President Eisenhower at 3 o’clock, and would like to show him the seasonally adjusted series for female unemployment. You have that wonderful machine, and can do it in a minute. How about delivering it in time for my appointment?”

I said, “Okay, okay.”

But you know what’s involved. First, I had to gather the data, have them punched, and, finally, we all had to pray that the machine would work. You know, believe it or not, we were able to go through all those steps that morning. It may be the only time in history that everything worked right. Murphy’s law in reverse: Everything that could go wrong went right. The data was delivered about 2 o’clock that afternoon. From there, President Eisenhower was completely sold on the method and strongly supported it.

### 7.1 A simple seasonal adjustment approach

An example of a simple seasonally adjusted process is as follows. The series is first smoothed by a moving average. If there are  $p$  seasons in a year ( $p = 12$  monthly data,  $p = 4$  for quarterly data), and  $p$  is even, then the initial trend estimate is given by

$$\hat{T}_t = \sum_{j=-p/2}^{p/2} w_j O_{t-j} \quad t = (p/2) + 1, \dots, n - (p/2) - 1$$

where  $w_j = 1/p$  for  $0, \pm 1, \dots, \pm(1/(2p) - 1)$  and  $w_{-j} = w_j = 1/(2p)$  for  $j = \pm p/2$ . This is equivalent to a  $2 \times p$  moving average. For example, setting  $p = 12$  gives the  $2 \times 12$  moving average

$$\hat{T}_t = \sum_{j=-6}^6 w_j O_{t-j} \quad t = 7, \dots, n-6$$

where the weights  $w_j$  are given by  $1/24, 1/12, \dots, 1/12, 1/24$ .

The residuals  $O_t - \hat{T}_t$  are then averaged over each season for a number of years to give seasonal factors  $S_h$  for  $h = 1, \dots, s$ . Subtracting these from the original series gives the seasonally adjusted series  $SA_t = O_t - S_t$ . More elaborate procedures such as X11 are extensions of this basic procedure.

## 7.2 Outline of the steps involved within X11

A brief description of the X11 procedure follows. The standard model option used within X11 is the multiplicative model. A multiplicative model is assumed, although options for an additive model are given by addition/subtraction rather than division/multiplication. (For a detailed description of all the steps within X11 and X11-ARIMA see Ladiray and Quenneville, 2001).

While X11 is an adhoc procedure using local smoothing methods, it could be thought of as a form of non-parametric smoothing. The advantage over the model based methods is that the use of local filters ensure that data points sufficiently far from the end of the series are only slightly adjusted, if at all, when new data points are added. One disadvantage is the high variability of the most recent data points. Some people argue that the lack of a specific model for the series is also disadvantageous, although some people may argue the opposite!

The basic steps are outlined below, these correspond to the “D” tables in Section 7.3.

### Step One: Initial estimate of trend

Step one involves a first estimate of the trend component. This is done by computing ratios between the original series  $O_t$  and a centered 12-term moving average ( $2 \times 12$  ma) of the original series to give an estimate of the trend  $\hat{T}_t$ . This gives first estimates of the seasonal and irregular components ( $S_t \cdot I_t$ ). Six values at the end of the series are lost due to using symmetric filters.

$$\left( \frac{O_t}{\hat{T}_t} \right) = \left( \frac{T_t \cdot S_t \cdot I_t}{\hat{T}_t} \right) \approx S_t \cdot I_t$$

A centered 12-term moving average is used for a preliminary estimate of the trend cycle. This will remove the seasonal pattern (see Figure 4.4 and 4.5). The main limitation of this filter is that it misses peaks and troughs of short cycles, i.e., two to three years. Irregular components

are comprised of short cycles and some of these short cycles will still remain in the smoothed data. Cycles with periods of five years or more will pass through this filter with a small reduction in amplitude. It has been shown that most economic time series have long trend-cycle variations of approximately 40 months or more. So this is a good filter for a preliminary estimation of the trend cycle as it includes the longer term cycles.

#### Step Two: Preliminary estimate of the seasonal component

A preliminary estimate of the seasonal component is then found by applying a weighted seven term average ( $3 \times 5\text{ma}$ ) to each month separately of the  $S_t \cdot I_t$ 's from step 1. The application of the  $3 \times 5\text{ma}$  is specific to SEASABS (see Chapter 8 for more information on SEASABS). The default moving average used within X11 at this step is the  $3 \times 3\text{ma}$ . Figures 5.13 and 5.14 show the gain function for the symmetric  $3 \times 3\text{ma}$  and  $3 \times 5\text{ma}$ . These type of moving averages enable estimation of linearly moving seasonality and can approximate gradual seasonal changes that follow non-linear patterns over the whole range of the series. The seasonal components are adjusted to add to 12 approximately for a multiplicative model (0 for an additive model) over a 12 month period by dividing them by a centered moving average (extended at the ends with the last known value). The missing values from the seasonal component at each end are computed by repeating the value in the previous year.

#### Step Three: Preliminary estimate of the adjusted data

A preliminary seasonally adjusted series is found by dividing the second estimate of the seasonal from the previous step into the original series

$$\left( \frac{O_t}{\hat{S}_t} \right) = \left( \frac{T_t \cdot S_t \cdot I_t}{\hat{S}_t} \right) \approx T_t \cdot I_t$$

#### Step Four: A better estimate of the trend

A 9,13 or 23 term Henderson moving average is applied to the seasonally adjusted values, depending on the volatility of the series, to give an improved estimate of the trend cycle (see Section 5.3). The resulting trend cycle series is divided into the original series to give a second estimate of the seasonal and irregular components. At the end of the series asymmetric moving averages are used, hence there are no missing values at the end as in step 1.

$$\left( \frac{O_t}{\hat{T}_t} \right) = \left( \frac{T_t \cdot S_t \cdot I_t}{\hat{T}_t} \right) \approx S_t \cdot I_t$$

#### Step Five: Final estimate of the seasonal component

A final estimate of the seasonal component is then found by applying a weighted seven term average ( $3 \times 5\text{ma}$ ) to each month separately of the  $S_t \cdot I_t$ 's from step 4. The seasonal components are adjusted to add to 12 approximately for a multiplicative model (0 for an additive model) over a 12 month period by dividing them by a centered moving average (extended at the ends with the last known value).



### Step Six: Final estimate of the adjusted data

A final seasonally adjusted series is found by dividing the second estimate of the seasonal from the previous step into the original series

$$\left( \frac{O_t}{\hat{S}_t} \right) = \left( \frac{S_t \cdot T_t \cdot I_t}{\hat{S}_t} \right) \approx T_t \cdot I_t$$

### Step Seven: Final estimate of the trend

A 9, 13 or 23 term Henderson moving average is applied to the seasonally adjusted series, which has been modified for extreme values. This gives an improved estimate of the trend cycle.

### Step Eight: Final estimate of the irregular component

The irregulars can then be estimated by dividing the trend estimates into the seasonally adjusted data.

$$\left( \frac{T_t \cdot I_t}{\hat{T}_t} \right) \approx I_t$$

The “B” tables and “C” tables in X11 follow a slightly different iteration to the “D” tables and are designed to estimate outliers and other components such as easter and trading day.

Different computations are used for different models within X11. These are the multiplicative, additive and pseudo-additive models. These differences are indicated in Section 7.3. There are small differences in the steps in X11 between various versions. Different versions include SAS proc x11, SEASABS, X11, X11ARIMA, X11ARIMA-88, X12ARIMA, and various other commercial and academic packages.

There are three major outlier steps in X11.

### Outlier step One: Weights used for outlier modification

Calculate a moving five year standard deviation,  $\sigma$ , of the estimates of the irregular components ( $I_t$ ). The standard deviation used for each of the 12 monthly values of  $I_t$  for a given year is found from a block of five years centered on that year. It is assumed that the mean of the irregulars is 1 (0 for the additive model).

A test is then constructed to determine extreme values. A weight between 0 and 1 is assigned to each irregular value using

$$\begin{array}{ll} |I_t - \mu| \leq 1.5\sigma & w = 1 \\ 1.5\sigma \leq |I_t - \mu| \leq 2.5\sigma & w = 2.5 - \left| \left( \frac{I_t - \mu}{\sigma} \right) \right| \\ |I_t - \mu| > 2.5\sigma & w = 0 \end{array}$$

Any values beyond  $2.5\sigma$  are classed as extreme and are given a zero weight.

#### Outlier step Two: Modified $SI$ values

Extreme values are then replaced as follows. For the first two years  $\sigma$  is set to  $\sigma$  for the third year, and  $\sigma$  for the last two years is set to  $\sigma$  for the third last year. To replace an extreme  $S_t \cdot I_t$  values, the weighted mean of five numbers is taken: the value multiplied by its weight and the two nearest  $S_t \cdot I_t$  values each side with corresponding months having  $w = 1$ . If there are not enough  $S_t \cdot I_t$  values each side and at the ends of the data special rules are used.

#### Outlier step Three: Modified original values

The original data is modified for outliers using

$$OM_t = O_t(1 + w_t(I_t - 1))/I_t$$

X11 has fairly robust outlier techniques. However, because X11 is an iterative procedure, if an outlier is too outrageous X11 will break down.

#### Computation of trading day factors

Let

$$\begin{array}{l} X_{t,1} \text{ be the number of Mondays in month } t \\ X_{t,2} \text{ be the number of Tuesdays in month } t \\ \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ X_{t,7} \text{ be the number of Sundays in month } t \end{array}$$

Then a trading day factor (without a length of month correction) in month  $t$  with daily activity weights  $d_1, \dots, d_7$  is given by

$$td_t = (\sum_{j=1}^7 d_j \cdot X_{t,j})/lom_t$$

where  $lom_t$  is the length of the month. A daily weight greater than 1 signifies high activity, less than 1 low activity and equal to 1 neutral activity.

Computation of the trading day factors is done by regression. The estimation of the regression coefficients excludes  $D \cdot I$  values identified as potential extremes.

Let  $D \cdot I$  be the trading day times the irregular modified for outliers. These are transformed so that the daily activity weights sum to zero. The transform is

$$D \cdot I'_t = lom_t \cdot D \cdot I_t - lom_t$$

Set

$$\hat{X}_{t,j} = X_{t,j} - X_{t,7} \text{ for } t = 1, \dots, n \text{ and } j = 1, \dots, 6$$

where  $n$  is the number of observations in the series. Let  $b_i = d_i - 1$  and then

$$\begin{bmatrix} D \cdot I'_1 \\ D \cdot I'_2 \\ \cdot \\ \cdot \\ D \cdot I'_n \end{bmatrix} = \begin{bmatrix} \hat{X}_{1,1} & \hat{X}_{1,2} & \dots & \hat{X}_{1,6} \\ \hat{X}_{2,1} & \hat{X}_{2,2} & \dots & \hat{X}_{2,6} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \hat{X}_{n,1} & \hat{X}_{n,2} & \dots & \hat{X}_{n,6} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_6 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

where  $e_i$  is  $D \cdot I'_i$ . Solving gives

$$b = [b_1 \ b_2 \ \dots \ b_6]' = (\hat{X}'\hat{X})^{-1}\hat{X}'e$$

Then

$$b_7 = -(b_1 + b_2 + \dots + b_6)$$

To estimate the daily activity weights the rank of the design matrix  $X$  only needs to be greater than or equal to six. If this occurs, the daily activity weights are computed as

$$d_j = 1 + b_j \text{ for } j = 1, \dots, 7$$

The trading day factors are then computed as

$$td_t = \left( \sum_{j=1}^7 d_j \cdot X_{t,j} \right) / lom_t$$

### 7.3 Output from X11

Legend:

- $O$  = Original series
- $P$  = Prior factors
- $OP$  = Original series prior corrected
- $OT$  = Original series modified for trading day
- $SA$  = Adjusted
- $SAM$  = Modified adjusted
- $T$  = Trend
- $I$  = Irregular (residual)
- $IM$  = Modified irregular
- $L$  = Length of month factors (trading day)
- $TD$  = Trading day factors
- $TDL$  =  $TD \cdot L$
- $SI$  = Seasonal  $\cdot$  Irregular factors
- $SIM$  = Modified seasonal  $\cdot$  irregular factors
- $S$  = Seasonal factors

The X11 tables shown below are for a monthly X11 with moving trading day,  $3 \times 5$  seasonal filter and a 13 term Henderson filter. The standard monthly X11 refers to the use of a  $3 \times 3$  seasonal filter.

The “A” tables include prior modifications to the original data such as very large outliers, abrupt changes in the trend and/or seasonal pattern, and other special corrections.

Table	Multiplicative	Additive	Pseudo-additive
A1	$O$	$O$	$O$
A2	$P$	$P$	$P$
A3	$O/P$	$O-P$	
A4	not used	not used	not used
A4B	$L$	$NA$	$L$

The “B” tables are designed to estimate outliers and trading day.

Table	Multiplicative	Additive	Pseudo-additive
B1	$OP = O/(P \cdot L)$	$OP = O - P$	$OP = O/(P \cdot L)$
B2	$T = (2 \times 12)(OP)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OP)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OP)$ Note: 6 observations lost at each end
B3	$SI = OP/T$	$SI = OP - T$	$SI = OP/T$
B4	$SI$ replacement values (*)	$SI$ replacement values (*)	$SI$ replacement values (*)
B5	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends
B6	$SA = OP/S$	$SA = OP - S$	$SA = OP - T \cdot (S - 1)$
B7	$T = (13H)(SA)$	$T = (13H)(SA)$	$T = (13H)(SA)$
B8	$SI = OP/T$	$SI = OP - T$	$SI = OP/T$
B9	$SI$ replacement values (*)	$SI$ replacement values (*)	$SI$ replacement values (*)
B10	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.
B11	$SA = OP/S$	$SA = OP - S$	$SA = OP - T \cdot (S - 1)$
B12	not used	not used	not used
B13	$I = SA/T$	$I = SA - T$	$I = SA/T$
B14	$I$ excluded from tday	$I$ excluded from tday	$I$ excluded from tday
B15	Daily trading day weights	Daily trading day weights	Daily trading day weights
B16	$TD$	$TD$	$TD$
B17	$W(**)$	$W(**)$	$W(**)$
B18	$TDL = TD \cdot L$	$TDL = TD$	$TDL = TD \cdot L (?)$
B19	$OT = OP/TDL$	$OT = OP - TDL$	$OT = OP/TDL$

Notes: (\*) is the weighted average of the nearest two full weight  $SI$  values on each side plus the actual  $SI$  by its weight. Special rules are used if the program cannot find two full weight values on one side.

(\*\*) refers to a double iteration of a 5 year moving sigma limit on the irregular.

The “C” tables are designed to estimate outliers and trading day.

Table	Multiplicative	Additive	Pseudo-additive
C1	$OM = OT \cdot (W \cdot (I - 1) + 1) / I$	$OM = OT + I \cdot (W - 1)$	$OM = OT + (I - 1) \cdot (W - 1) \cdot T$
C2	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end
C3	not used	not used	not used
C4	$SI = OM / T$  $SI$ replacement values (*)	$SI = OM - T$  $SI$ replacement values (*)	$S = OM / T$  $SI$ replacement values (*)
C5	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected. $S$ is also extended at the ends
C6	$SA = OM / S$	$SA = OM - S$	$SA = OM - T \cdot (S - 1)$
C7	$T = (13H)(SA)$	$T = (13H)(SA)$	$T = (13H)(SA)$
C8	not used	not used	not used
C9	$SI = OM / T$  $SI$ replacement values (*)	$SI = OM - T$  $SI$ replacement values (*)	$S = OM / T$  $SI$ replacement values (*)
C10	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$  Note: $S$ is also level corrected.
C11	$SAM = OM / S$	$SAM = OM - S$	$SAM = OM - T \cdot (S - 1)$
C12	not used	not used	not used
C13	$IM = SAM / T$	$IM = SAM - T$	$IM = SAM / T$
C14	$I$ excluded from tday	$I$ excluded from tday	$I$ excluded from tday
C15	Daily trading day weights	Daily trading day weights	Daily trading day weights
C16B	$TD$	$TD$	$TD$
C17	$W (**)$	$W (**)$	$W (**)$
C18	$TDL = TD \cdot L$	$TDL = TD$	$TDL = TD \cdot L (?)$
C19	$OT = OP / TDL$	$OT = OP - TDL$	$OT = OP / TDL$

The “D” tables provide final estimates of all components.

Table	Multiplicative	Additive	Pseudo-additive
D1	$OM = OT \cdot (W \cdot (I - 1) + 1) / I$	$OM = OT + I \cdot (W - 1)$	$OM = OT + (I - 1) \cdot (W - 1) \cdot T$
D2	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end	$T = (2 \times 12)(OM)$ Note: 6 observations lost at each end
D3	not used	not used	not used
D4	$SI = OM / T$  SI replacement values	$SI = OM - T$  SI replacement values	$S = OM / T$  SI replacement values
D5	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected. $S$ is also extended at the ends
D6	$SA = O / S$	$SA = O - S$	$SA = O - T \cdot (S - 1)$
D7	$T = (13H)(SA)$	$T = (13H)(SA)$	$T = (13H)(SA)$
D8	$SI = OT / T$	$SI = OT - T$	$SI = OT / T$
D9	$SI = OM / T$  SI replacement values	$SI = OM - T$  SI replacement values	$S = OM / T$  SI replacement values
D10	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.	$S = (3 \times 5)(SIM)$ Note: $S$ is also level corrected.
D11	$SA = OP / S$	$SA = OP - S$	$SA = OP - T \cdot (S - 1)$
D12	$T = (13H)(OM / S)$	$T = (13H)(OM - S)$	$T = (13H)(OM - T \cdot (S - 1))$
D13	$I = SA / T$	$I = SA - T$	$I = SA / T$

## 7.4 Beyond X11?

Another method of time series analysis employed by some statistical agencies, such as the US Bureau of the Census, is to use ARIMA modelling of the data (see Findley *et. al* (1998) for more details). The use of ARIMA modelling on the original series helps reduce revisions in the seasonally adjusted series by extending the original series so that the effect of the end point problem is reduced. Kenny and Durbin (1982) discuss this issue in more detail.

Experience at the ABS has shown that generally ARIMA models do not seem to fit Australian data very well. This may be related to the increased variability in Australian data that comes with smaller sample sizes. Indeed Statistics Canada state that if a series does not fit one of the models, it may be "so much contaminated by the irregulars that its systematic movement is unidentifiable"! (see Dagum (1980), p8).

ARIMA models can also be used to prior adjusted an original series for effects such as trading day, easter, additive outliers and level shifts. This approach is used within X12-ARIMA as part of the reg-ARIMA framework.

Research into the use of ARIMA models with Australian data is an ongoing research topic at the ABS.

## 7.5 Using X11

A version of X11 exists in SAS which can be used to try X11. Simple SAS coding to do this might be (it is assumed that an input text file exists of the form month year observation):

```

data f2;
infile "c:\test\last.dis";
input month year orig;
run;

data f1;
set f2;
date=mdy(month,1,year);
tme=year+(2*month-1)/24;
output;
run;

proc x11 data=f1;
var orig;
monthly      date=date printout=standard tdregr=adjust
              trendma=13;
macurves JAN='3X5' FEB='3X5' MAR='3X5' APR='3X5' MAY='3X5' JUN='3X5'
          JUL='3X5' AUG='3X5' SEP='3X5' OCT='3X5' NOV='3X5' DEC='3X5';
output out=factors b1=orig d8=si d11=adj d12=trend d10=seas
          c17=weight d13=irreg;
title 'dummy series';
run;

proc print data=factors;
run;

```

Other versions such as X12ARIMA by the United States Bureau of the Census could be used at home by downloading it from their web site. The program is freeware, in fact, they encourage people to use it.

The ABS' version of X11 is SEASABS. To use this you would need to obtain a stand alone version that could be installed on your hard drive.

## 8 SEASABS

### 8.1 Introduction

**SEASABS** (**SEAS**onal Analysis, **ABS** standards) is a "knowledge-based" seasonal analysis and adjustment system used by the ABS to seasonally adjusted time series data. SEASABS is one part of the ABS time series processing system. Other components include the ABSDB (ABS information warehouse), FAME (Forecasting, Analysis and Modelling Environment - which is used to store and manipulate time series data) and PAS (Publishing Assembly System).

### 8.2 SEASABS

SEASABS performs four major functions:

- Data review
- Seasonal reanalysis of time series
- Investigation of time series
- Upholding ABS quality standards

SEASABS allows expert and client use of the X11 method (which has been enhanced significantly by the ABS) of seasonal adjustment, i.e. a user does not need detailed knowledge of the X11 package to arrive at an acceptable seasonal adjustment of a time series. An intelligent interface guides the user through the seasonal analysis process, making appropriate choices of parameters and adjustment methods. SEASABS keeps records of the previous analysis of a series so it can compare X11 diagnostics over time and "knows" what parameters and prior factors led to the acceptable adjustment at the last analysis.

SEASABS is a knowledge based system for analysis of time series. It is used mainly by Time Series Analysis Section to undertake reanalysis of time series datasets compiled by the subject matter areas of the ABS. SEASABS creates seasonal adjustment factors for the series and gives the user an indication of the suitability of these factors. It identifies and corrects trend and seasonal breaks as well as extreme values, inserts trading day factors if necessary, chooses appropriate moving averages for the computation of trends and seasonal factors, and allows for moving holiday corrections.

The history of changes to the parameters and prior factors can be viewed. They are contained in the 'series knowledge' component of SEASABS and are stored in the ABSDB. Graphs of original, seasonally adjusted, trend, seasonal/irregular, X11 outputs and toolset facilities are available.

SEASABS is available in both an expert and client version. The main difference between the two versions is that the client version does not allow you to change any parameters. This means that apart from the start and end dates of the series and whether or not dialogue is displayed, no feature of the series can be changed. Any corrections for structural breaks or other problems must be calculated by SEASABS and can not be manually inputted. As a result, in the client version, prior correction factors can be viewed, but not changed.



The following systems are also used in conjunction with SEASABS.

### **8.3 Time Series Toolkit**

The Time Series Toolkit (TST) provides a time series view of data held in ABSDB macrodatasets. It allows time series functionality on the ABSDB, allowing users to tabulate, graph and manipulate time series in their native format as opposed to a multidimensional tabular format. The TST cannot alter data stored on the ABSDB. The results can be printed or saved for use with other products.

### **8.4 Time Series Update**

This system creates (updates) seasonally adjusted and trend data. It is used mainly by subject matter areas in applying seasonal factors (estimated by Time Series Analysis section) to the original data to derive seasonally adjusted and trend data. The updated data results in the seasonally adjusted and trend data being available to the same end date as the original series. This is done for both the forward factor adjustment and concurrent adjustment. This functionality is performed within the TS Update facility which is separate from SEASABS. The subject matter areas typically do not have user access to SEASABS.

If the series is adjusted using the forward factor adjustment approach, a set of factors for the following year will have been calculated and stored. In this instance, a time series update is performed when the next month/quarter of data becomes available which automatically selects the stored factor corresponding to the new data and applies it (as well as the factors for all preceding months/quarters) to calculate the new seasonally adjusted, trend series and sensitivity analysis.

For example, consider a series which is adjusted using the forward factor method and has been analysed to February 1999. When the March 1999 data becomes available, running Time Series Update will cause the March factor derived at the February analysis to be applied to the March original data. The factors from February 1999 back to the beginning of the series will also be applied to their corresponding original data points to create a seasonally adjusted series for the entire data span. This series is then trended.

If the series is adjusted using the concurrent adjustment approach only the first of the factors calculated at the annual analysis is ever applied. Then, when the next month/quarter of data becomes available and the Time Series Update is performed, the seasonal adjustment procedure determined at the last analysis is reapplied. This will update the factors for previous months/quarters to incorporate the information of the latest data point. Together with these new factors, an updated seasonally adjusted series, trend and sensitivity analysis are then calculated and saved to the ABSDB.

## 8.5 Upholding ABS quality standards

Seasonally adjusted or trend data must be cleared by the Time Series Analysis area prior to release. Clearance procedures ensure that data quality standards are observed as closely as possible. Data may be considered as "cleared" when combined or seasonal adjustment factors are provided to the client area and all necessary prior corrections have been created.

SEASABS forms a key part of the clearance process in that it will only allow combined adjustment factors to be used in those circumstances where ABS quality standards have been met.

In using SEASABS, the **client area** is responsible for:

- performing preliminary or concurrent reanalysis
- performing initial investigations into data not currently seasonally adjusted or trended
- reviewing data for release or client servicing purposes
- running the monthly or quarterly time series update
- releasing the trend and seasonally adjusted data
- handling routine client queries about the trend and seasonally adjusted data

In using SEASABS, the **Time Series Analysis Section** is responsible for:

- finalising all reanalyses
- finalising investigations into data not previously seasonally adjusted or trended
- quality assurance
- data clearance
- troubleshooting problems with the data
- handling more complex client queries about the trend and seasonally adjusted data or methods of adjustment

## 8.6 The basic steps in running SEASABS

1. Open a group. A group is a set of series which may have some common characteristics and which, for operational convenience, are handled as a block.
2. Open or insert a series, via the File menu.
3. Set the X11 options (only start, end date, file name and format are mandatory).
4. Select a Run option from the Seasonal Analysis menu.

SEASABS will execute X11 and diagnose the results, iterating until it cannot improve the adjustment further.

The Seasonally adjusted series, the ABS trend series and the combined adjustment factors can be saved via the File Menu.

Results can be viewed either graphically or in tabular form using Graphs or Analysis Output respectively. Further information on the adjustment is available using Tools. If an analysis

has been performed, information on the various hypotheses tested can be examined using the Diagnosis menu.

There are various factors which need to be considered to help select the most appropriate method of adjustment. It is necessary to decide if forward factors are required or if it is only necessary to re-adjust the series every month/quarter. Further considerations are whether the series can be best adjusted by a multiplicative, additive or pseudo-additive model and how the series has been previously adjusted (if this is not the first adjustment).

SEASABS offers four analysis choices: Preliminary Analysis, Annual Analysis, Research Analysis and Concurrent Adjustment.

- (i) A preliminary analysis is recommended for examining a series prior to the entire new year of data being available so that, any problems with the series can begin to be identified before the annual analysis. This type of analysis does not result in a new set of forward factors.
- (ii) An annual analysis is recommended if a benchmark run for the series has already been created and an extra year's data from this benchmark run exists. It is possible to perform an annual analysis on a new series without a benchmark run in which case SEASABS will perform two runs, one on a shorter span of data to create a benchmark and one on the entire span of data for the annual analysis.
- (iii) A research analysis is recommended if the series is new or if the user wishes to investigate any problems encountered with the analysis or experiment with methods of handling these problems. The results of a research analysis are not recorded unless the run is forced to annual or the series knowledge is updated.
- (iv) A concurrent adjustment is also useful for investigating the behavior of a series. An adjustment of this kind does not test the various hypotheses inherent in SEASABS and so will not modify extremes or detect seasonal and trend breaks. Since these hypotheses are not tested this run is relatively fast so that basic results can be viewed almost immediately. Currently, this option must be employed if the series is to be additively or pseudo-additively adjusted.

As well as choosing the analysis option most appropriate for the current needs of the user, it is necessary for a processing procedure to be selected. SEASABS is equipped with two adjustment options: Forward Factor adjustment and Concurrent adjustment. Both methods begin with an annual analysis which is performed once per year. This annual analysis updates the series knowledge on the ABSDB and these updates are then set for the following year. In both cases a set of forward factors is calculated and stored on the ABSDB.

## 8.7 SEASABS Demo

Ask a friendly Time Series Analyst (email: [timeseries@abs.gov.au](mailto:timeseries@abs.gov.au)) for an example of how SEASABS works in practice.

## 9 Exercises

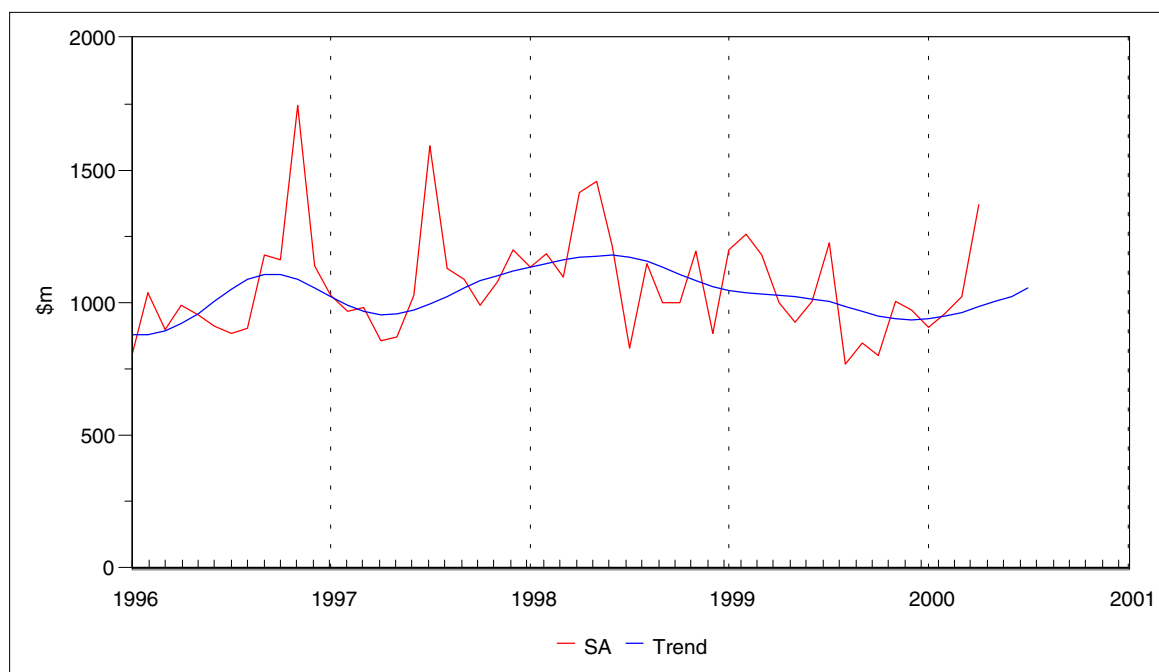
Q1.

Is it better to apply a short term moving average or a longer term moving average if we want to smooth an adjusted series? Why?

Q2.

For each of the three types of time series models (additive, multiplicative and pseudo-additive) can you derive an expression for the vertical distance between the trend and the seasonally adjusted series? Can you interpret this quantity in words?

The vertical distance between the trend and seasonally adjusted series can easily be seen in a graph such as the one below



Q3. Below are some hypothetical time series. Decide which series are either:

- A) additive
- B) multiplicative
- C) pseudo-additive
- D) all of the above
- E) no idea

Give reasons for your choice.

Figure 1

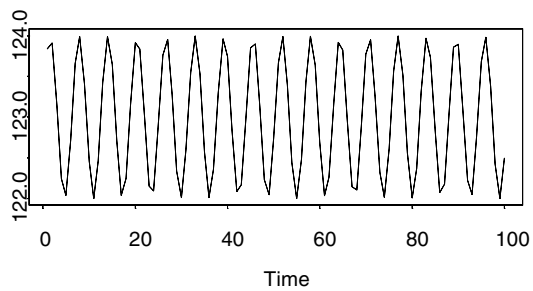


Figure 2

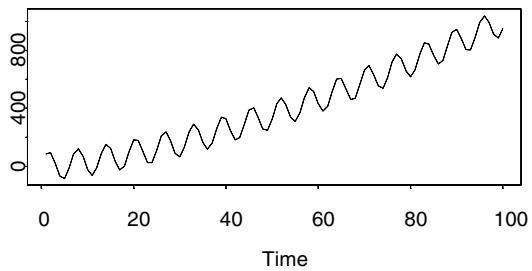


Figure 3

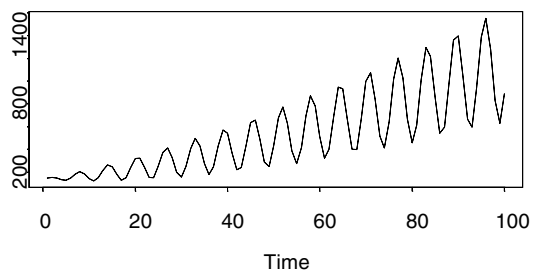


Figure 4

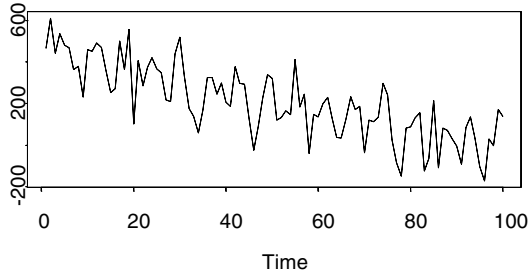


Figure 5

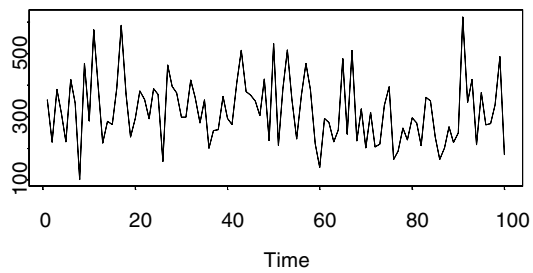


Figure 6

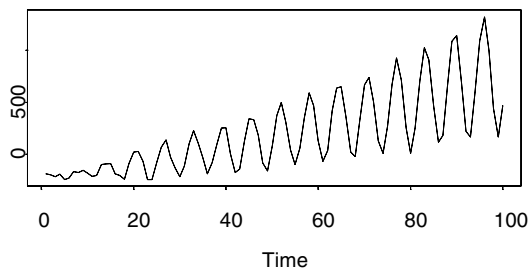


Figure 7

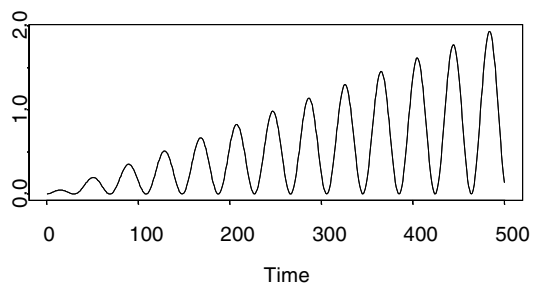
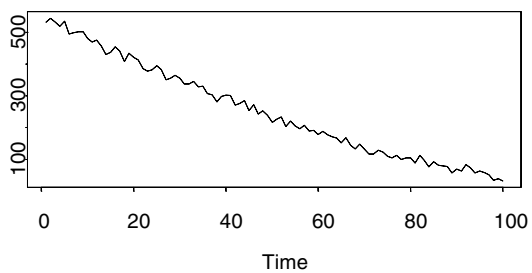


Figure 8



Q4. Englebert worked at the Australian Bureau of Statistics and did not live very far away from work. One day Englebert's car broke down. It broke down very badly and had to be totally rebuilt over a period of one month. He now had to walk to work. Being a keen and enthusiastic time series person, he kept a record of the time it took him to walk to and from work over a period of one month.

The two datasets are:

Walking to work:

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	31.34	23.19	28.86	19.04	20.05
Week 2	30.49	25.33	26.20	23.10	25.59
Week 3	29.10	24.50	29.01	22.50	24.52
Week 4	Sick Day	32.07	29.43	30.20	21.22

Walking home from work:

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	25.00	24.30	24.00	31.00	23.50
Week 2	23.20	22.50	22.10	33.40	22.30
Week 3	21.00	21.20	20.50	29.10	20.20
Week 4	20.00	19.10	19.40	45.50	19.20

From visual inspection of the data (unless you want to plot them!) describe what is going on.

In particular, you should consider aspects like:

- are there any problems with either of the datasets?
- what factors would you consider if you were analysing the datasets?
- describe what the components of the series are doing.

Q5. The spectra of a stationary process  $X_t$  with an absolutely summable autocovariance sequence  $\gamma_{Xk}$  is given by

$$f_X(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{Xk}(\omega) e^{-i\omega k}$$

Given a process  $Z_t$  which is the sum of two independent stationary process  $X_t$  and  $Y_t$ ,

$$Z_t = X_t + Y_t$$

Derive the spectra for  $Z_t$  in terms of the spectra of  $X_t$  and  $Y_t$ .

Q6.

Compute and plot the gain for first differencing twice  $(1 - B)^2$ ?  
This has weights  $w_0 = 1, w_{-1} = -2, w_{-2} = 1$ .

Simplify your answer using the identities

$$\begin{aligned}\cos(2w) \cos(w) + \sin(2w) \sin(w) &= \cos(w) \\ \cos(2w) &= 2 \cos^2(w) - 1 \\ \cos^2(w) + \sin^2(w) &= 1\end{aligned}$$

For a monthly time series what does second differencing do to the amplitude of a 2 month cycle?

Q7.

Show analytically (difficult) or numerically that the phase shift, in periods, for a 12 term uncentered moving average with equal weights is given by

$$\Phi(\omega)/\omega = -5.5, \text{ for } \omega \in (0, \pi/6)$$

Given business cycles are usually defined in the range 18 months to 5 years, what does the above result mean for detecting turning points with this moving average?

The following formulas may be of use

$$\begin{aligned}\sum_{j=0}^{n-1} \cos(-jx) &= 0.5 * (1 - \cos(x) - \cos(nx) + \cos((n-1)x)) / (1 - \cos(x)) \\ \sum_{j=0}^{n-1} \sin(-jx) &= -0.5 * (\sin(x) - \sin(nx) + \sin((n-1)x)) / (1 - \cos(x))\end{aligned}$$

If you are using this uncentred 12 term moving average what does the above result imply for detecting turning points in the data?

Q8.

X11 is a commonly used seasonal adjustment procedure. Referring to the attached edited output and assuming the default model for X11 answer the following:

- (i) What is the default model used for X11?
- (ii) Illustrate that the decomposition chosen in part (i) holds, by choosing a row from the output below.
- (iii) Are there any “unusual values” in the original series?
- (iv) What should the sum of the seasonal components add up to over what period? Show this using the output below.

Date	Orig	Weight	SI	Seas	Adj	Trend	Irreg
JAN99	441.7	1.000	1.09842	1.08431	409.277	403.668	1.01389
FEB99	448.5	0.950	1.06786	1.10775	391.195	405.267	0.96528
MAR99	415.6	0.968	1.01975	1.04351	399.464	408.150	0.97872
APR99	408.0	1.000	0.99322	0.99717	410.421	411.528	0.99731
MAY99	416.6	1.000	1.00751	0.99062	421.625	414.250	1.01780
JUN99	409.3	1.000	0.97970	0.96899	419.992	415.303	1.01129
JUL99	387.6	1.000	0.93964	0.94636	411.501	414.491	0.99279
AUG99	394.5	1.000	0.95967	0.95952	412.143	412.183	0.99990
SEP99	407.6	1.000	0.99087	0.99224	408.305	409.018	0.99826
OCT99	378.5	1.000	0.92896	0.93328	403.856	406.082	0.99452
NOV99	359.6	0.000	0.90155	0.90716	400.585	403.702	0.99228
DEC99	435.7	1.000	1.08383	1.06958	405.678	401.095	1.01142
JAN00	433.8	1.000	1.09570	1.08198	401.962	397.471	1.01130

## Q9.

- (i) Read some monthly data into SAS. See your friendly time series analyst to be emailed a dataset!
- (ii) Create a date and time variable
- (iii) Use proc X11 to run X11 using the original series as the input series
- (iv) What do the options *tdregr*, *trendma* and *additive* do within the *monthly* command within X11.
- (v) To set up the data in printable form add the commands:

```
output out=factors b1=orig d8=si d11=adj d12=trend d10=seas
c17=weight d13=irreg;
title 'x11 output';
run;
```

```
proc print data=factors;
format
date MONYY. ;
run;
```

- (vi) Given that we have used the additive option within X11 choose a 12 month period (and write down which period you used!) and show that the sum of the seasonals satisfy the appropriate condition for an additive model.
- (vii) Choose three observations and show that by using the appropriate model for an additive decomposition, that the values given for trend, seasonal and irregular for that observation approximately equal the original value.

## Q10.

- (i) Seasonal adjustment is a complex procedure. Write half a page on some of the issues surrounding seasonal adjustment, including, why should seasonal adjustment be done? Also include any advantages and disadvantages of doing seasonal adjustment.
- (ii) Should seasonally adjusted estimates be smoothed. If so, why? If not, why?



- (iii) Trend estimates are commonly calculated at the end of a series. Assuming a filter based approach to trend estimation, briefly describe some of the issues involved in calculating trend estimates at the end of a series.

Q11.

- (i) The choice of a surrogate based filter will be determined by what factor?
- (ii) How many surrogate filters are needed to estimate the trend at the end of the series, for a Henderson filter of length  $2m + 1$ .

Q12.

The following table contains a quarterly time series of gas production. Using the original data, complete the corresponding columns, 2-5. Complete the seasonally adjusted series (column 6) using the values obtained in the previous columns.

Quarterly Gas Production							
		[1]	[2]	[3]	[4]	[5]	[6]
Year	Qtr	Original	Fourth Qtr moving average	2x4 moving average	S-I Ratio	Seasonal Factors	Seasonally Adjusted Series
1981	1	52.8	-	-	-	?	?
	2	71.3	-	-	-	?	?
	3	78.9	?	?	?	?	?
	4	62.1	?	?	?	?	?
1982	1	52.3	?	?	?	?	?
	2	70.0	?	?	?	?	?
	3	78.8	?	?	?	?	?
	4	59.4	?	?	?	?	?
1983	1	53.9	?	?	?	?	?
	2	72.0	?	?	?	?	?
	3	80.5	?	?	?	?	?
	4	63.0	?	?	?	?	?
1984	1	56.0	68.6	69.0	0.811	?	69.6
	2	74.7	69.5	69.6	1.073	?	69.1
	3	84.4	69.7	69.8	1.209	?	70.4
	4	63.7	69.9	70.2	0.908	?	69.4
1985	1	56.7	70.5	70.5	0.804	?	70.4
	2	77.1	70.6	70.8	1.088	?	71.3
	3	84.9	71.1	71.2	1.192	?	70.9
	4	65.6	71.4	71.6	0.916	?	71.4
1986	1	57.8	71.9	72.2	0.801	?	71.8
	2	79.4	72.4	72.4	1.097	?	73.5
	3	86.7	72.4	-	-	?	72.4
	4	65.7	-	-	-	?	71.5

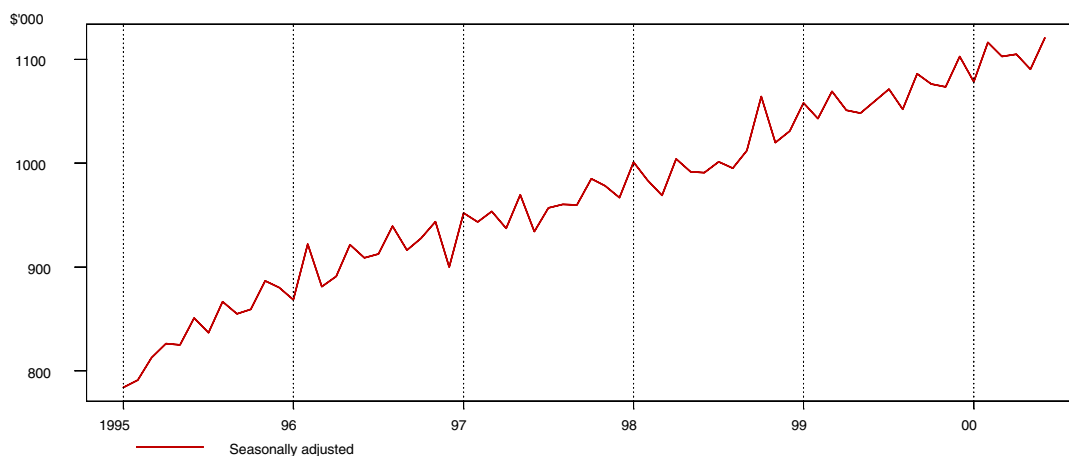
Q13.

Why does the ABS use only odd length moving averages? Justify your answer.

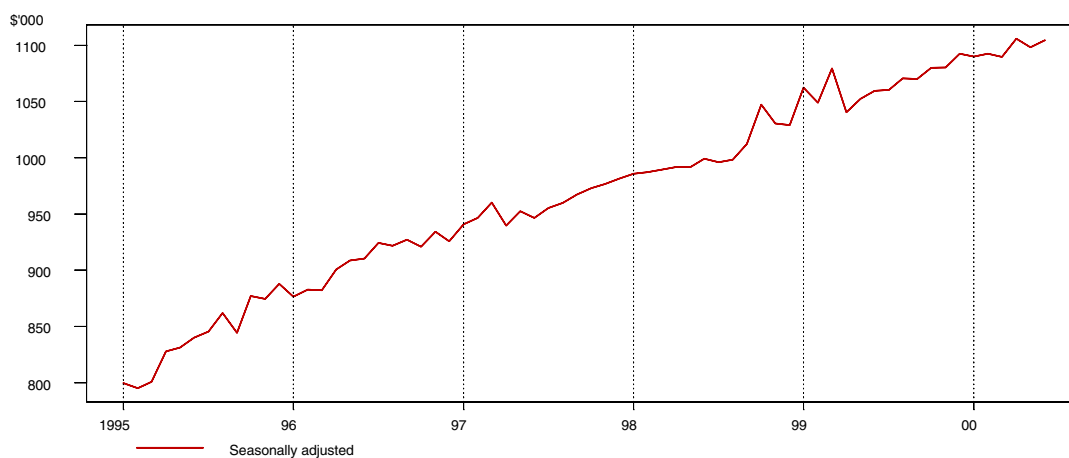
Q14.

The following two graphs are seasonally adjusted series of NSW supermarkets and grocery stores turnover. They are identical except that one has been adjusted for trading day variation and other has not. Which graph has been adjusted for trading day variation? Justify your answer.

**Graph 1**



**Graph 2.**



Q15.

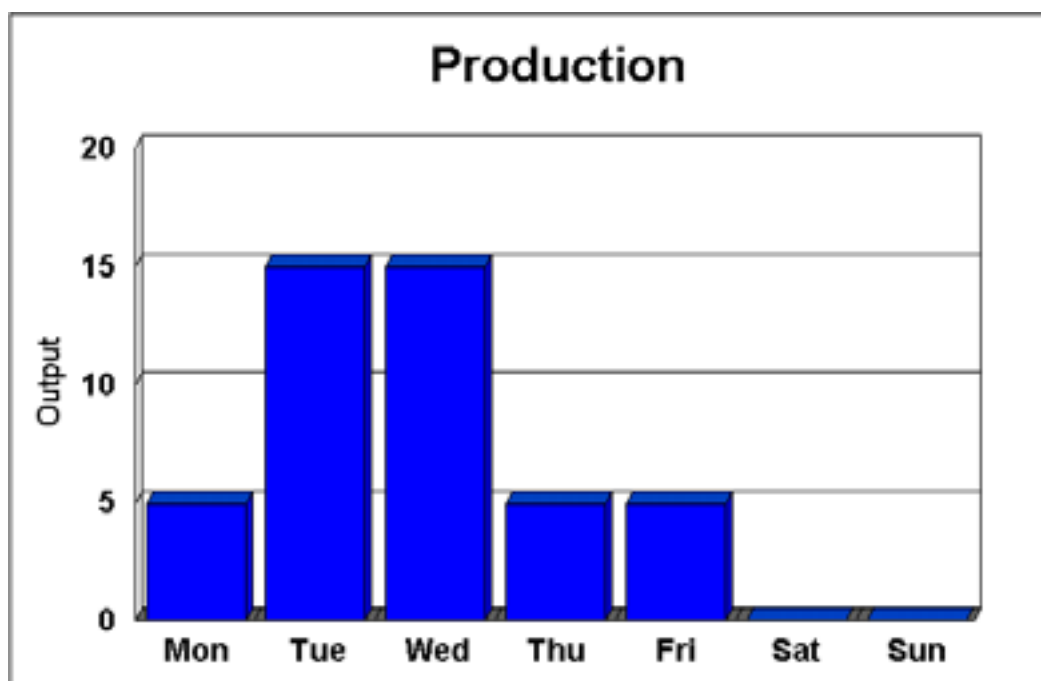
- (i) What is the difference between a 'stock' and 'flow' series?
- (ii) Which type of series is more likely to display calendar variation? Why?

Q16.

- (i) What is the purpose of trading-day adjustment?
- (ii) What is meant by the term 'within-month variation'? Why is it a problem?
- (iii) Name the component of the time series model where trading day variation would most likely occur? Justify your answer.

Q17.

The following graph shows a producer's different levels of output during the week. Using the calendar below, show how the monthly output levels can vary despite the underlying level of output staying the same. For this example show the output levels for the months January through to May.



January						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

March						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
31					8	2
3	4	5	6	7		9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

May						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

February						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29		

April						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

June						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
30						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

Q18.

Show using Taylor Series expansion that the multiplicative model requires the constraint

$$\sum_{j=1}^n S_{t+j} = n$$

Hint: Use the fact that

$$\prod_{j=1}^n S_{t+j} = 1$$

## 10 Solutions

Q1.

It depends on how volatile the adjusted series is. A more volatile series should be smoothed using a longer term moving average.

Q2.

In the additive case:

$$\begin{aligned}
 SA_t - T_t &= (O_t - S_t) - T_t \\
 &= ((T_t + S_t + I_t) - S_t) - T_t \\
 &= (T_t + I_t) - T_t \\
 &= I_t
 \end{aligned}$$

In the multiplicative case:

$$\begin{aligned}
 SA_t - T_t &= \frac{O_t}{S_t} - T_t \\
 &= \frac{T_t S_t I_t}{S_t} - T_t \\
 &= T_t I_t - T_t \\
 &= T_t (I_t - 1)
 \end{aligned}$$

The same result is achieved in the pseudo-additive case:

$$\begin{aligned}
 SA_t - T_t &= O_t - T_t(S_t - 1) - T_t \\
 &= T_t + T_t(I_t - 1) - T_t \\
 &= T_t(I_t - 1)
 \end{aligned}$$

Therefore we can interpret the vertical distance between the trend and the seasonally adjusted series seen in the graph above as follows:

In the additive case, the distance between the two lines is the irregular. In the multiplicative and pseudo-additive case this distance is not proportional to the irregular, though in periods of stable (flat) trend this may be a reasonable approximation.

Q3.

Figure 1: Additive. The seasonal pattern remains constant over time.

Figure 2: Additive. The seasonal pattern remains constant over time. There is an increasing trend and the series dips below the zero value.

Figure 3: Multiplicative. The seasonal pattern increases as time increases. The series does not dip below the zero value.

Figure 4: Additive. Series has a decreasing trend and a seasonal pattern. Pretty hard to see with all the noise!

Figure 5: Random noise.

Figure 6: Not a multiplicative series. Note that it has zero and possibly some negative values. A pseudo-additive model may be appropriate here.

Figure 7: True pseudo-additive model. Has the characteristics of a multiplicative model but includes zero values.

Figure 8: Use an additive model, although a multiplicative model may also be satisfactory here. Series appears to be just a decreasing trend with an irregular component.

Q4.

The answer would need to consider such issues as:

- Trading day?
- Missing values: what to do?
- Decreasing trend
- Rounding on the way home from work
- What are the units in?
- Appears to be a weekly effect on Thursday. Maybe this is the 4.51 social club?!

Q5.

Let  $\gamma_z(B)$ ,  $\gamma_x(B)$  and  $\gamma_y(B)$  be autocovariance generating functions for  $Z_t$ ,  $X_t$  and  $Y_t$  respectively.

Then

$$\begin{aligned}
 \gamma_z(B) &= \sum_{k=-\infty}^{\infty} \gamma_z(k)B^k \\
 &= \sum_{k=-\infty}^{\infty} \text{cov}(z_t, z_{t+k})B^k \\
 &= \sum_{k=-\infty}^{\infty} \text{cov}(x_t + y_t, x_{t+k} + y_{t+k})B^k \\
 &= \sum_{k=-\infty}^{\infty} \{\text{cov}(x_t + x_{t+k}) + \text{cov}(y_t, y_{t+k})\}B^k
 \end{aligned}$$

$$\gamma_z(B) = \gamma_x(B) + \gamma_y(B)$$

Therefore

$$\begin{aligned}
 f_z(\omega) &= \sum_{k=-\infty}^{\infty} \gamma_z^{(k)} e^{-i\omega k} \\
 &= \sum_{k=-\infty}^{\infty} (\gamma_x^{(k)} + \gamma_y^{(k)}) e^{-i\omega k} \\
 &= \sum_{k=-\infty}^{\infty} \gamma_x^{(k)} e^{-i\omega k} + \sum_{k=-\infty}^{\infty} \gamma_y^{(k)} e^{-i\omega k} \\
 &= f_x(\omega) + f_y(\omega)
 \end{aligned}$$

Q6.

In this case the weights have been included so we don't have to use the backshift operator  $B$  to calculate the weights, however an example of this can be seen below.

$$\begin{aligned}
 f(y_t) &= (1 - B)^2 \times y_t \\
 &= (1 - 2B + B^2) \times y_t \\
 &= y_t - 2y_{t-1} + y_{t-2}
 \end{aligned}$$

which implies the given weights of  $w_0 = 1$ ,  $w_{-1} = 2$ , and  $w_{-2} = 1$ .

In this case we have

$$m_1 = 2 \text{ and } m_2 = 0$$

and we know that

$$G(\omega) = \sqrt{C(\omega)^2 + S(\omega)^2}$$

and

$$\begin{aligned}
 C(\omega) &= \sum_{j=-m_1}^{m_2} w_j \cos(\omega j) \\
 &= \sum_{-2}^0 w_j \cos(\omega j) \\
 &= \cos(-2\omega) - 2 \cos(-\omega) + \cos(0) \\
 &= \cos(2\omega) - 2 \cos(\omega) + 1 \\
 S(\omega) &= \sum_{j=-m_1}^{m_2} w_j \sin(\omega j) \\
 &= \sum_{-2}^0 w_j \sin(\omega j) \\
 &= \sin(-2\omega) - 2 \sin(-\omega) + \sin(0) \\
 &= 2 \sin(\omega) - \sin(2\omega)
 \end{aligned}$$

Then we can express the gain function as

$$\begin{aligned}
 G(\omega) &= \sqrt{(1 - 2 * \cos(\omega) + \cos(2\omega))^2 + (-2 * \sin(\omega) + \sin(2\omega))^2} \\
 &= \sqrt{1 - 4 \cos(\omega) + 2 \cos(2\omega) + 4 \cos^2(\omega) - 4 \cos(\omega) \cos(2\omega) + \cos^2(2\omega) + 4 \sin^2(\omega) - 4 \sin(\omega) \sin(2\omega) + \sin^2(2\omega)} \\
 &= \sqrt{6 - 8 \cos(\omega) + 2 \cos(2\omega)} \\
 &= \sqrt{6 - 8 \cos(\omega) + 4 \cos^2(\omega) - 2} \\
 &= \sqrt{4 - 8 \cos(\omega) + 4 \cos^2(\omega)} \\
 &= \sqrt{4 * (1 - \cos(\omega))^2} \\
 &= 2 * (1 - \cos(\omega))
 \end{aligned}$$

Or alternatively,

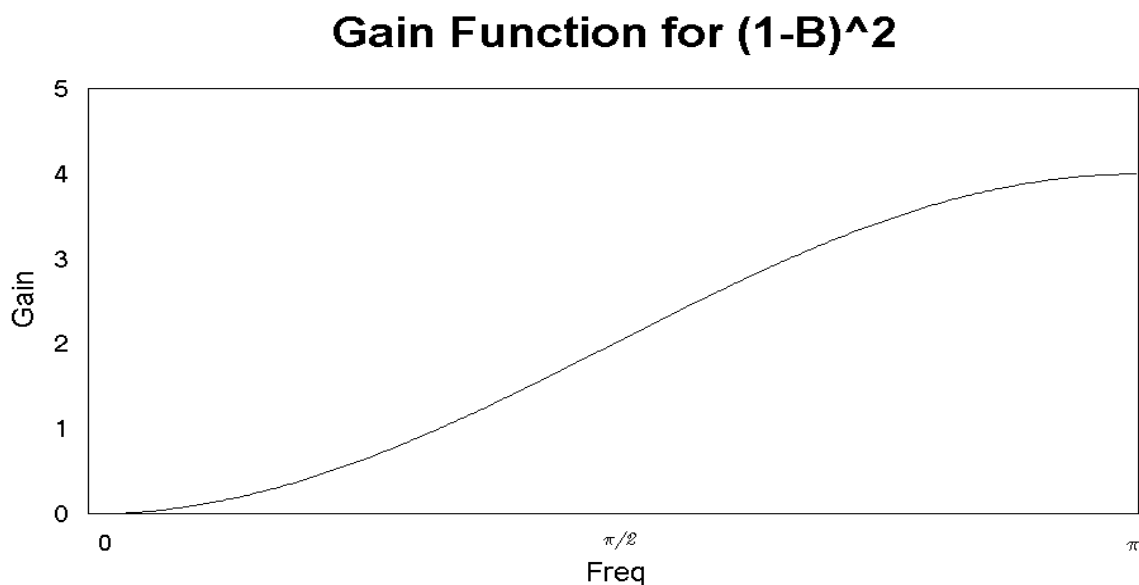
The gain of the differencing operator is given by

$$G(\omega) = \sqrt{2 - 2 \cos(\omega)}$$

Hence the gain of the second difference operator is given by the square of the Differencing operator, that is

$$G(\omega) = (\sqrt{2 - 2 \cos(\omega)})^2 = 2 - 2 \cos(\omega)$$

A plot of the gain is  $2 - 2 \cos(\omega)$ ,  $\omega = 0, \pi$  over a suitable grid. This can be seen below,



A 2 month cycle is at frequency  $(2\pi)/2 = \pi$ , hence the gain at this frequency is  $2 - 2 \cos(\pi) = 4$ . Hence a 2 month cycle is amplified 4 times.



Q7.

In this case the weights of the filter are as follows

$$w_0 = \frac{1}{12} \text{ to } w_{11} = \frac{1}{12}$$

As we are dealing with a 12 term uncentered moving average with equal weights.

Then

$$\begin{aligned} C(\omega) &= \sum_{j=-m_1}^{m_2} w_j \cos(\omega j) \\ &= \sum_0^{11} w_j \cos(\omega j) \\ &= \frac{1}{12} \{ \cos(0) + \dots + \cos(12\omega) \} \end{aligned}$$

$$\begin{aligned} S(\omega) &= \sum_{j=-m_1}^{m_2} w_j \sin(\omega j) \\ &= \sum_0^{11} w_j \sin(\omega j) \\ &= \frac{1}{12} \{ \sin(0) + \dots + \sin(12\omega) \} \end{aligned}$$

Note that this expression can be simplified using the formula given in the question. ie.

$$\begin{aligned} C(\omega) &= \frac{1}{12} \sum_{j=0}^{11} \cos(-j\omega) \\ &= 6 \{ 1 - \cos(\omega) - \cos(12\omega) + \cos(11\omega) \} / \{ 1 - \cos(\omega) \} \end{aligned}$$

$$\begin{aligned} S(\omega) &= \frac{1}{12} \sum_{j=0}^{11} \sin(-j\omega) \\ &= -6 \{ \sin(\omega) - \sin(12\omega) + \sin(11\omega) \} / \{ 1 - \cos(\omega) \} \end{aligned}$$

Now for any point in  $\omega \in (0, \pi/6)$ ,  $C(\omega) > 0$  and  $S(\omega) < 0$ . We can calculate the phase function as

$$\begin{aligned} \Phi(\omega) &= \arctan\left(\frac{S(\omega)}{C(\omega)}\right) \\ &= \arctan\left(\frac{\sin(0) + \dots + \sin(12\omega)}{\cos(0) + \dots + \cos(12\omega)}\right) \\ &\text{or} \\ &= \arctan\left(\frac{-(\sin(\omega) - \sin(12\omega) + \sin(11\omega))}{1 - \cos(\omega) - \cos(12\omega) + \cos(11\omega)}\right) \end{aligned}$$

and if we substitute in some values for  $\omega$  we find that  $\Phi(\omega)/\omega$  is a constant that can

be calculated by substituting in any value for  $\omega$  in the range  $\omega \in (0, \pi/6)$ . Therefore

$$\Phi(\omega)/\omega = -5.5$$

So for the uncentered 12 term moving average with equal weights the phase shift is -5.5 periods.

Some code to produce the phase is shown below. Note: if we don't allow for the cosine terms going negative they won't get the right answer. It is possible to show analytically, but very messy.

```

pi = 3.1415926#
FOR deg = 1 TO 30
  freq = deg * pi / 180
  c = 1 - COS(freq) - COS(12 * freq) + COS(11 * freq)
  s = -(SIN(freq) - SIN(12 * freq) + SIN(11 * freq))
  IF c > 0 THEN
    PRINT deg; ATN(s / c) / freq
  ELSE
    PRINT deg; (ATN(s / c) - pie) / freq
  END IF
NEXT deg

```

This means that a turning point for any cycles in the business cycle range will be detected 5.5 months after the actual turning point, which effectively means 6 months late.

Q8.

- (i) X11 uses the multiplicative model as its default model.

$$Y_t = T_t \times S_t \times I_t$$

- (ii)

$$\begin{aligned}
 \text{Use January 1980} &= 1.08431 \times 403.668 \times 1.01389 \\
 &= 443.78 \\
 &\simeq 441.7
 \end{aligned}$$

$$\begin{aligned}
 \text{Use January 1981} &= 1.08198 \times 497.471 \times 1.01130 \\
 &= 434.91 \\
 &\simeq 443.8
 \end{aligned}$$

- (iii) The unusual values are those with weights of zero

- (iv) For the multiplicative model the sum of the seasonal components should be

twelve. The sum from January 1980 to December 1980 is 12.00049.

Q9.

In SAS construct the following

```
data dataset;
infile "dataset.sas";
input month $ year orig ;
run;
```

These four commands input the *dataset.sas* dataset and label it as *dataset* within SAS. The three variables *month*, *year* and *orig* are the month represented as numbers (Jan=1, Feb=2,...), and the year, and the estimates respectively.

The following commands setup the date in terms of months, days and years.

```
data dataset2;
set dataset;
date=mdy(month,1,year);
tme=year+(2*month-1)/24;
if tme ge 1978 and tme lt 1996.8 then output;
run;
```

The above commands create a new dataset within SAS called *dataset2* which has the new variables *date* and *tme*. To use X11 within SAS the following procedure needs to be used

```
proc x11 data=dataset2;
```

This tells the X11 procedure to use the dataset *dataset2* which includes the *date* and *tme* variables.

The next set of commands relate to options within the X11 procedure.

For more examples of the different types of options check out *help* within the *modelling and analysis tools* and then to *econometrics and time series* and then to *x11*.

Next add the following to your file

```
var orig;
```

This specifies the variables to be analysed by X11. We are only interested in the original series from the dataset *dataset2*.

The following commands define the dataset as containing monthly data

```
Monthly date=date charts=none printout=none
```

*tdregr=adjust trendma=13 additive;*

All charts and normal X11 output have been turned off. This usually comes out to approximately 40 pages!

Q10.

- (i)
  - It is a complex procedure to use, ie need X11
  - Seasonal adjustment enables users to compare data from year to year more easily
  - Makes the data easier for users to interpret
  - Various filter lengths can be used, and using X11 is bad at the ends of the data
  - Enables us to measure and/or eliminate seasonal effects
- (ii)
  - Eliminates the irregular component, making the data easier to interpret
  - Yes, smoothed
- (iii)
  - Asymmetric filters vs. symmetric filters
  - Revision to the estimates

Q11.

- (i) The amount of revision that occurs between the surrogate based estimates of trend and the main Henderson based estimates of trend. The preferred surrogate filter will be the one that minimise the amount of revision.
- (ii)  $m$  surrogate filters. So, although only one 13 term Henderson filter is required for smoothing a monthly series, a set of six surrogate filters is required for the end-point problem.

Q12.

The completed table appears below:

Quarterly Gas Production							
		[1]	[2]	[3]	[4]	[5]	[6]
Year	Qtr	Original	Fourth Qtr moving average	2x4 moving average	S-I Ratio  =col[1]/col[3]	Seasonal Factors	Seasonally Adjusted Series =col[1]/col[5]
1981	1	52.8				<b>0.805</b>	<b>65.6</b>
	2	71.3				<b>1.081</b>	<b>66.0</b>
	3	78.9	<b>66.3</b>	<b>66.2</b>	<b>1.192</b>	<b>1.198</b>	<b>65.9</b>
	4	62.1	<b>66.2</b>	<b>66.0</b>	<b>0.941</b>	<b>0.918</b>	<b>67.6</b>
1982	1	52.3	<b>65.8</b>	<b>65.8</b>	<b>0.795</b>	<b>0.805</b>	<b>65.0</b>
	2	70.0	<b>65.8</b>	<b>65.5</b>	<b>1.069</b>	<b>1.081</b>	<b>64.8</b>
	3	78.8	<b>65.1</b>	<b>65.3</b>	<b>1.206</b>	<b>1.198</b>	<b>65.8</b>
	4	59.4	<b>65.5</b>	<b>65.8</b>	<b>0.903</b>	<b>0.918</b>	<b>64.7</b>
1983	1	53.9	<b>66.0</b>	<b>66.2</b>	<b>0.814</b>	<b>0.805</b>	<b>67.0</b>
	2	72.0	<b>66.5</b>	<b>66.9</b>	<b>1.076</b>	<b>1.081</b>	<b>66.6</b>
	3	80.5	<b>67.4</b>	<b>67.6</b>	<b>1.191</b>	<b>1.198</b>	<b>67.2</b>
	4	63.0	<b>67.9</b>	<b>68.2</b>	<b>0.924</b>	<b>0.918</b>	<b>68.6</b>
1984	1	56.0	68.6	69.0	0.811	0.805	69.6
	2	74.7	69.5	69.6	1.073	1.081	69.1
	3	84.4	69.7	69.8	1.209	1.198	70.4
	4	63.7	69.9	70.2	0.908	0.918	69.4
1985	1	56.7	70.5	70.5	0.804	0.805	70.4
	2	77.1	70.6	70.8	1.088	1.081	71.3
	3	84.9	71.1	71.2	1.192	1.198	70.9
	4	65.6	71.4	71.6	0.916	0.918	71.4
1986	1	57.8	71.9	72.2	0.801	0.805	71.8
	2	79.4	72.4	72.4	1.097	1.081	73.5
	3	86.7	72.4			1.198	72.4
	4	65.7				0.918	71.5

Note: Columns 2 and 3 involve estimating the trend by a process of moving averages. Column 2 is a 4-qtr moving average obtained by average the values, first of periods 1 to 4, then of periods 2 to 5, etc.. The averages are taken to represent the ‘smoothed-out’ values relating to the period in the middle of the span of the average. Thus the 4-qtr average of periods 1 to 4 is centred between periods 2 and 3. A second averaging process must be employed to centre the moving averages **at** periods rather than **between** periods. This is the reason for Column 3. For example, in column 2 (1981, qtr 3),  $66.3=(52.8+71.3+78.9+62.1)/4$ , and column 3,  $66.2=(66.2+66.3)/2$

Columns 4 and 5 involve estimating the seasonal component of the time series. Column 4 divides the trend values into the original values to obtain the ‘de-trended’ originals (this assumes a multiplicative model, ie  $O_t=T_t \times S_t \times I_t$ ). For example, in column 4 (1981, qtr 3),  $1.192=78.9/66.2$ . Column 5 is a moving average of the seasonal-irregulars for each quarter to give estimates of the seasonal factors. For example, in column 5 (1981, qtr 3),  $1.198=(1.192+1.206+1.191+1.209+1.192)/5$ . Column 6 divides the original by the seasonal factors to obtain the seasonally adjusted estimate. For example, in column 6 (1981, qtr 3),  $65.9=78.9/1.198$ .

Q13.

To retain the time phase. *Odd* length seasonal and trend moving averages have the averaged value on an observable and publishable observation point. However, an *even* length averaging span centres the average value between the periods, introducing a half month or half quarter time phase shift.

Q14.

Graph 2, as it contains less volatility.

Q15.

(i) A flow series can be regarded as the sum of daily (or weekly, etc) figures to form a monthly (or quarterly) aggregate. A stock series relates to a particular point in time, such as the last working day of the month or the last Wednesday in the month.

(ii) A flow series. Since flow series figures are more or less related to the daily activity during a given period and the composition of a period may vary both from period to period and year to year then there is a possibility that there will be a non-seasonal variation in the series linked to the calendar.

Q16.

(i) It reduces the month-to-month variation in the seasonally adjusted data so that the trend cycle is revealed more clearly. In other words, the trading day effect leads to apparent changes in the level of activity when the underlying level is unchanged.

(ii) The variation between months of equal length in the number of times a particular day or days of the week occur in a calendar month, and the length-of-month variation in the calendar. Trading day variation, like seasonality, can complicate comparisons of values in a series from one month to the next, or those for a given month from one year to the next.

(iii) The irregular component. Trading day variation frequently reverses direction, over a period of months and years, while the trend is a smooth curve. Over a period of years, within a month, trading day variation also frequently reverses direction, while seasonal variation is constant or gradually increasing or decreasing curve.

Q17.

		Sun	Mon	Tue	Wed	Thu	Fri	Sat	Monthly Total
Jan	Days	4	5	5	5	4	4	4	
	Output	0	5	15	15	5	5	0	
	Total	0	25	75	75	20	20	0	<b>215</b>
Feb	Days	4	4	4	4	5	4	4	
	Output	0	5	15	15	5	5	0	
	Total	0	20	60	60	25	20	0	<b>185</b>
Mar	Days	5	4	4	4	4	5	5	
	Output	0	5	15	15	5	5	0	
	Total	0	20	60	60	20	25	0	<b>185</b>
Apr	Days	4	5	5	4	4	4	4	
	Output	0	5	15	15	5	5	0	
	Total	0	25	75	60	20	20	0	<b>200</b>
May	Days	4	4	4	5	5	5	4	
	Output	0	5	15	15	5	5	0	
	Total	0	20	60	75	25	25	0	<b>205</b>

Q18.

$$S_{t+1} \cdot S_{t+2} \cdot \dots \cdot S_{t+n} = 1$$

Using Taylor Series expansion our function is

$$\begin{aligned} \theta &= S_{t+1} \cdot S_{t+2} \cdot \dots \cdot S_{t+n} \\ &= f(S_{t+1}, \dots, S_{t+n}) \end{aligned}$$

Which has an estimate of

$$\begin{aligned} \hat{\theta} &= f(s_{t+1}, \dots, s_{t+n}) \\ &= f(S_{t+1} = \Delta s_{t+1}, \dots, S_{t+n} = \Delta s_{t+n}) \\ &= f(S_{t+1}, \dots, S_{t+n}) + \Delta s_{t+1} \frac{\partial f}{\partial S_{t+1}}(S_{t+1}, \dots, S_{t+n}) + \dots + \\ &\quad \Delta s_{t+n} \frac{\partial f}{\partial S_{t+n}}(S_{t+1}, \dots, S_{t+n}) \end{aligned}$$

Evaluating this expression gives

$$\hat{\theta} = 1 + n - \sum_{j=1}^n S_{t+j}$$

So

$$\begin{aligned} 1 + n - \sum_{j=1}^n S_{t+j} &= 1 \\ \sum_{j=1}^n S_{t+j} &\approx n \end{aligned}$$

## 11 Further reading

Australian Bureau of Statistics (1987). A Guide to Smoothing Time Series - Estimates of "Trend". *Australian Bureau of Statistics*, 1316.0, Canberra, Australia.

Australian Bureau of Statistics (1993). A Guide to Interpreting Time Series Monitoring "Trend" - an Overview. *Australian Bureau of Statistics*, 1348.0, Canberra, Australia.

Australian Bureau of Statistics (1999a). The new method for seasonally adjusting crop production data. *Australian Economic Indicators*, 1350.0 June 1999.

Australian Bureau of Statistics (1999b). Introduction of Concurrent Seasonal Adjustment into the Retail Trade Series. *Australian Bureau of Statistics*, 8514.0, Canberra, Australia.

Bell, P. (1999). The Impact of Sample Rotation Patterns and Composite Estimation on Survey Outcomes, Working Paper No. 99/1, *Australian Bureau of Statistics*, Canberra, Australia.

Bell, W.R. and Hillmer, S.C. (1984). Issues Involved with the Seasonal Adjustment of Economic Time Series, *Journal of Business and Economic Statistics*, 2, 291-320.

Binder, D.A. and Hidiroglou M. (1988). Sampling in Time, in Handbook of Statistics, Vol. 6, ed. By P.R. Krishnaiah and C.R. Rao, Elsevier Science Publishers, B.V., 187-211.

Box, G.E.P; Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control, published by Holden-Day.

Cannon, J. (2000). Diagnostic Measures for Comparing Direct and Aggregative Seasonal Adjustments, Working Paper No. 2000/1, *Australian Bureau of Statistics*, Canberra, Australia.

Cantwell, P.J., and Caldwell, C.V. (1998). Examining the Revisions in Monthly Retail and Wholesale Trade Surveys Under a Rotating Panel Design. *Journal of Official Statistics*, Vol. 14, No. 1, 47-59.

Chatfield, C. (1996). The analysis of time series : an introduction. (5th edition). London, Chapman and Hall.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess, *Journal of Official Statistics*, Vol. 6, No. 1., 3-73.

Dagum, E.B. (1980). The X11ARIMA Seasonal Adjustment Method. Ottawa: Statistics Canada, cat. 12-564E.

Dagum, E.B., Chhab, N. and Chiu, K. (1996). Derivation and Properties of the X11ARIMA and Census X11 Linear Filters. *Journal of Official Statistics*, 12, No. 4, 329-347.



Doherty, M. (1992). The Surrogate Henderson Filters in X11, *Statistics New Zealand*, Working Paper.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen B. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business and Economic Statistics*, Vol. 16, No. 2., 127-177.

Gray, A. and Thomson, P. (1996a). Design of moving-average trend filters using fidelity, smoothness and minimum revisions criteria, Bureau of the Census, RR96/01

Gray, A. and Thomson, P. (1996b). Design of moving-average trend filters using fidelity and smoothness criteria in Vol 2: Time Series Analysis in Memory of E.J. Hannan. ed. P. Robinson and M. Rosenblatt. Springer Lecture Notes in Statistics 115, 205-219.

Gray, A. and Thomson, P. (1996c). On a family of moving-average trend filters for the ends of series. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1996.

Henderson, R. (1916). Note on Graduation by Adjusted Average. Transactions of the American Society of Actuaries, 17, 43-48.

Harvey, A. (1990) Forecasting, structural time series models, and the Kalman filter. Published by Cambridge University Press.

Kenny, P.B., and Durbin, J. (1982). Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series. *Journal of the Royal Statistical Society, Series A*, 145, 1-41.

Ladiray, D. and Quenneville, B. (2001). Seasonal Adjustment with the X-11 method, New York: Springer Verlan, Lecture notes in statistics, 158.

Laniel, N. (1985). Design criteria for 13 term Henderson end-weights. Technical Report Working paper TSRA-86-011, *Statistics Canada*, Ottawa K1A 0T6.

Leung C., McLaren C.H., Zhang X,. (1999) Adjusting for an Easter Proximity Effect, Working Paper No. 99/3, *Australian Bureau of Statistics*, Canberra, Australia.

Macaulay, F. (1931). The smoothing of time series. *National Bureau of Economic Research*, New York.

McLaren, C.H. (1997) A short course on Seasonal Adjustment Using X11, *University of Wollongong*, NSW, Australia.

McLaren, C.H. (1999) Designing Rotation Patterns and Filters for Trend Estimation in Repeated Surveys *Unpublished PhD. Thesis, University of Wollongong*, NSW, Australia.

Pierce, D.A. (1980). Data Revisions with Moving Average Seasonal Adjustment Procedures. *Journal of Econometrics*, 14, 95-114.

Shiskin, J., Young, A. H. And Musgrave, J.C. (1967). The X11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Sutcliffe, A. and Lee, G. (1995). Seasonal Analysis and Sample Design. Paper presented at the Conference of Survey Measurement and Process Quality, Bristol 1995.

Sutcliffe, A. (1993). X11 Time Series Decomposition and Sampling Errors. Working Papers in Econometrics and Applied Statistics, No. 93/2. *Australian Bureau of Statistics*, Canberra, Australia.

Sutcliffe, A. (1999). Seasonal adjustment: Comparison of Philosophies. Working Papers in Econometrics and Applied Statistics, No. 99/2. *Australian Bureau of Statistics*, Canberra, Australia.

Wallis, K.F. (1982). Seasonal Adjustment and Revision of Current Data: Linear filters for the X11-method. *Journal of the Royal Statistical Society, Series A*, 145, 74-85.

Wei, W.W.S. (1993). Time Series Analysis: Univariate and Multivariate Methods. Published by Addison-Wesley.