F E A T U R E   A R T I C L E – Interpreting Time Series Data

INTRODUCTION

Users of statistics regularly analyse time series data in an attempt to understand real world dynamics. For example, changes in the characteristics of the population, cyclical movements in economic markets, etc. This type of analysis is difficult when examining the original data as there can be seasonal or other influences masking the true direction of the series. For this reason, the ABS publishes seasonally adjusted and trend data for many of its series. While publication of these additional series can be extremely useful to the experienced analyst, it may result in some confusion for the general user in terms of understanding what each series is indicating. This article aims to explain the basic concepts of time series analysis, discuss issues users should be aware of, and provide an indication of the most appropriate series to use in different circumstances.

WHAT IS A TIME SERIES?

A *time series* is a collection of well-defined data points that have been measured at regular intervals of time. For example, the number of litres of milk produced each quarter would be a time series because a litre of milk is a well-defined concept and each measurement is taken over a period of three months.

WHOLE MILK INTAKE BY FACTORIES, Western Australia
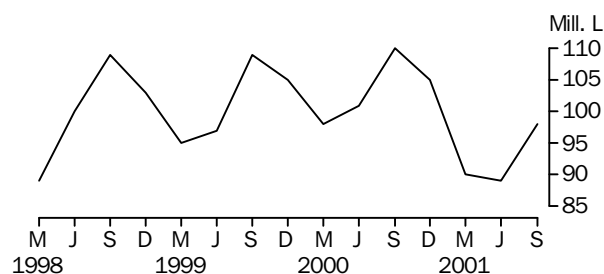
Source: Livestock Products, Australia (Cat. no. 7215.0)

Data which are collected irregularly or only once cannot be defined as a time series. For example, a one-off count of the total number of persons who received the government's $14,000 First Home Owner Grant is not a time series.

TYPES OF TIME SERIES

Time series can be classified as being either a *stock* or a *flow* series, depending on the type of measurements being taken.

Stock series are measures, or counts, taken at a point in time. For example, the number of bicycles in a store on a particular day. This figure will change from day to day depending on the amount of stock received that day and the number of bicycles sold. Similarly, the Labour Force Survey takes stock of the number of people employed in a particular reference week and is therefore considered to be a stock series.

Flow series are measures of activity over a given period of time. For example, the number of bicycles sold by a store in a particular month. This figure will change day by day, depending on the number of bicycles sold each day. At the end of the month, the total number of sales can be calculated. Similarly, the number of new motor vehicle sales each month is the sum of all new motor vehicles sold during each day of the month.

The main difference between a stock and a flow series is that a flow series can be affected by trading day effects (see Trading Day Effect section on page 16 for further information). Apart from this, both stock and flow series are treated in much the same way in the time series analysis process.

· ·

14   A B S • W E S T E R N   A U S T R A L I A N   S T A T I S T I C A L   I N D I C A T O R S • 1 3 6 7 . 5 • M A R C H   2 0 0 2

COMPONENTS OF A TIME SERIES

A time series can be thought of as comprising three separate components:

- the trend,

- any calendar related effects, and

- the residual effects.

The trend component is a measure of the underlying behaviour of the series over time. That is, whether the series is generally increasing, decreasing or remaining stable over time. This underlying behaviour could be due to influences such as population growth, price inflation or general economic development, and can often be hidden in the original time series data by the calendar related and/or residual effects.

For example, consider the original data in the figure below. A superficial examination of the data at the current end of the series would suggest that the number of employed persons in WA has taken a downward turn in January 2002. However, upon further examination, it can be seen that there is also a downward turn for the previous three Januarys, which would indicate that there may be a seasonal factor influencing the original data. The fact that January appears to be a low seasonal month could be caused, for example, by a high number of employees ending their contracts in January after working over the Christmas period. An examination of the underlying behaviour of the series shows that the number of employed persons in WA has actually remained relatively stable over most of 2001 and, if anything, the series seems to be slowly increasing, not decreasing.

EMPLOYED PERSONS, Western Australia
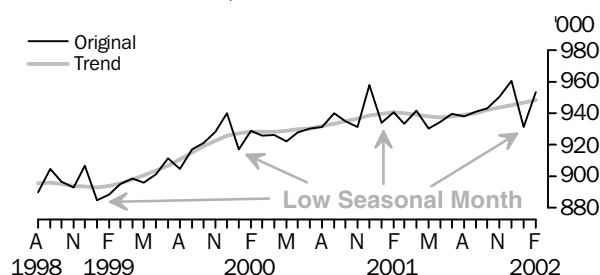
Source: Labour Force, Australia, Preliminary (Cat. no. 6202.0)

CALENDAR RELATED EFFECTS

Calendar related effects are systematic influences on the source data. They are predictable and persistent, and are sometimes referred to as 'seasonal effects' even though they encompass more than just seasonality. The four main types of calendar related effects are:

- seasonal effects;

- trading day effects;

- moving holiday effects; and

- other systematic effects.

Seasonal Effects

Seasonal effects are factors which recur one or more times per year. They are reasonably stable with respect to annual timing, consistent direction and predictable magnitude. They can be due to natural factors (eg. seasons, harvests), administrative or legal matters (eg. tax payments) or social traditions (eg. Christmas).

For example, the following figure shows large increases in the December retail turnover figures over the last five years. These increases are most likely due to increased Christmas spending in December.

Seasonal Effects *continued*

MONTHLY RETAIL TURNOVER, Western Australia
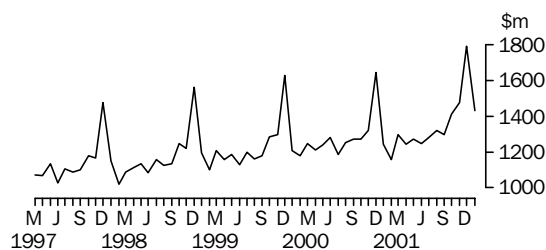
Source: Retail Trade, Australia (Cat. no. 8501.0)

The presence of seasonal effects can also be seen in the following gas production graph. There are distinct increases each winter, when gas heating is in high demand, and marked decreases over the summer months.

QUARTERLY GAS PRODUCTION, Australia
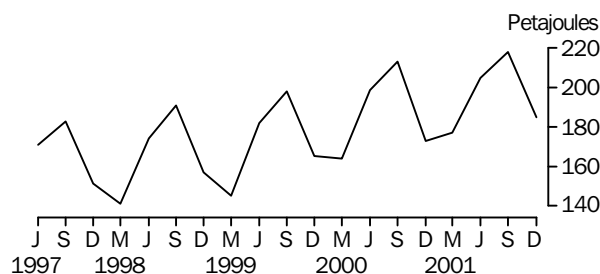
Source: Manufacturing Production, Australia (Cat. no. 8301.0)

Trading Day Effect

A trading day effect is caused by the number of high and low activity days in a given month. That is, since each month in the year has 28 days, plus one, two or three extra days, time series data can be affected by whether these extra days are high or low activity days. For example, in a 31 day month, if the three extra days were Sunday, Monday and Tuesday, then it would be expected that less retail sales would be recorded than if the three extra days were Thursday, Friday and Saturday, since there is generally a higher level of retail activity towards the end of the week.

Series are also affected by the varying number of extra days in the month. For example, suppose that a factory's average production of jelly beans has the following distribution.

AVERAGE JELLY BEAN PRODUCTION

Day of Week	Number
Sunday	0
Monday	4 000
Tuesday	6 000
Wednesday	6 000
Thursday	5 000
Friday	3 000
Saturday	0
Weekly Total	**24 000**

If the above distribution remains consistent from year to year, then the only difference between the production of jelly beans in the same month across different years will be due to the activity on the extra days. As shown below, the number of working days in March 1999, March 2000, March 2001 and March 2002 were 23, 23, 22 and 21 respectively, and the extra working days were Monday, Tuesday & Wednesday in 1999, Wednesday, Thursday & Friday in 2000, Thursday & Friday in 2001, and Friday in 2002.

MARCH 1999

Su	Mo	Tu	We	Th	Fr	Sa
	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 2000

Su	Mo	Tu	We	Th	Fr	Sa
			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 2001

Su	Mo	Tu	We	Th	Fr	Sa
				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 2002

Su	Mo	Tu	We	Th	Fr	Sa
31					*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

Knowing the distribution of the factory's jelly bean production, the average number of jelly beans produced in each full four week period would be 24,000 x 4 = 96,000 and, hence, the total production of jelly beans in each March would have been:

TOTAL JELLY BEAN PRODUCTION

. . . .

March	Total number of jelly beans produced
1999	96 000 + 4 000 (Mo) + 6 000 (Tu) + 6 000 (We) = **112 000**
2000	96 000 + 6 000 (We) + 5 000 (Th) + 3 000 (Fr) = **110 000**
2001	96 000 + 5 000 (Th) + 3 000 (Fr) = **104 000**
2002	96 000 + 3 000 (Fr) = **99 000**

If no consideration was given to the trading day effect, it would appear as though jelly bean production had declined over the last four years, from 112,000 jelly beans in March 1999 to 99,000 jelly beans in March 2002. In reality, the production has remained constant and it is only the working days in March that have changed.
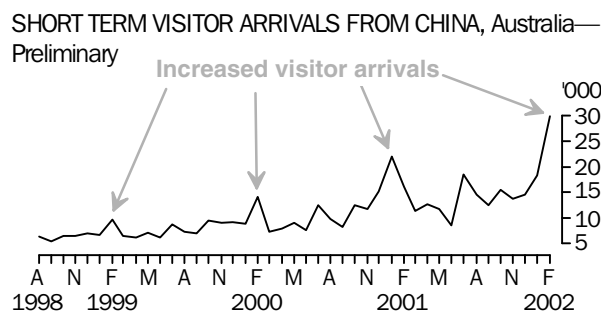
Moving Holiday Effect

Moving holiday effects are caused by regular holidays which do not occur at the same time each year. For example, both Easter and Chinese New Year occur once a year but, since they follow the cycles of the moon, the exact month in which they occur can vary.

In most years, Easter falls in April, but can occur in March or March/April. The effects of Easter can be expected to be seen in confectionery production figures and tourism series as many people travel over the Easter holidays.

Moving Holiday Effect *continued*

Similarly, Chinese New Year normally occurs in February, but will sometimes fall in late January. Effects from this holiday are evident in Overseas Arrivals and Departure series from some Asian countries as many people travel over this holiday period. For example, the following graph shows short term visitor arrivals from China. Chinese New Year started on 16 February in 1999, 5 February in 2000, 24 January in 2001 and 12 February in 2002. Correspondingly, sharp increases in visitor arrivals can be seen in February 1999, February 2000, *January 2001* and February 2002.

SHORT TERM VISITOR ARRIVALS FROM CHINA, Australia—
Preliminary

Source: Overseas Arrivals and Departures, Australia (Cat. no. 3401.0)

Moving holidays can also affect data for months or quarters adjacent to the one where the holiday falls. This is called a *proximity effect* and will occur if the holiday falls close to the beginning or end of the month or quarter of interest. For example, the Retail Trade series is sometimes adjusted for an Easter proximity effect, depending on whether Easter falls in late March or early April.

Other Systematic Effects

Other systematic effects can have an impact on time series. For example, government social security payments are typically paid fortnightly. In some months, this will result in two payments and in other months there will be three. A series measuring the total monthly government outlays on, say, the Age Pension, would be affected by this systematic effect.

RESIDUAL EFFECT

Residual effects (sometimes referred to as 'irregulars') are short term fluctuations in the data which are generally not systematic or predictable with regards to timing, duration and degree of impact. These random fluctuations are typically caused by sampling and non-sampling errors in the data. Sampling errors are found in data collected through sample surveys and exist as a result of not enumerating the entire population. Non-sampling errors are all other errors in the data (such as reporting errors, processing errors, coverage errors, etc) and can affect collections regardless of whether or not they are sample surveys.

Aside from these random fluctuations, large impacts can sometimes be observed in the residual effect. For example, the effect of a flood on agricultural production data, or the effect of The New Tax System (TNTS) on retail turnover figures.

As it is not possible to identify the cause, timing or magnitude of most irregular effects, they cannot normally be individually removed from the series (except for some large irregular effects). Instead, the ABS uses a generalised statistical procedure known as filtering, or smoothing, to remove the short term residuals from the series, as described further in the section Calculating the Trend (see page 20).

SEASONAL ADJUSTMENT

Seasonal adjustment is the process by which calendar related effects are removed from the original series. A seasonally adjusted series, then, will be the combination of the underlying trend of the series and the irregular factors. Whether the seasonally adjusted series is a good estimate of the trend will depend on the strength of the irregulars in the series.

For example, as discussed above, the Monthly Retail Turnover series has strong seasonal factors (there are large spikes each December due to Christmas trading). When the series is seasonally adjusted, these factors are removed, as shown below.

MONTHLY RETAIL TURNOVER, Western Australia
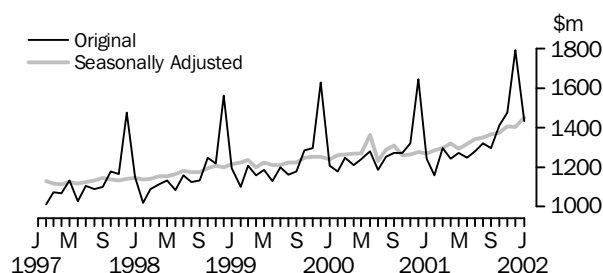
Source: Retail Trade, Australia (Cat. No. 8501.0)

The seasonally adjusted series can be seen to be quite similar to the underlying trend of the series. This is because the strength of the irregulars is generally small relative to that of the trend component (except in mid-2000 where a strong GST-related irregular can be observed).
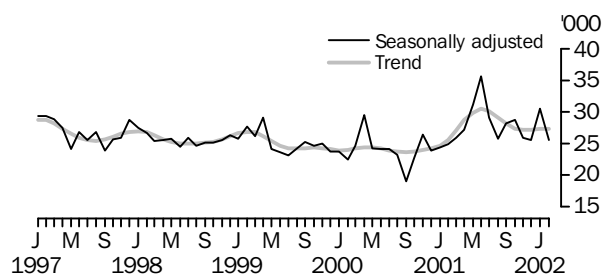
MONTHLY RETAIL TURNOVER, Western Australia
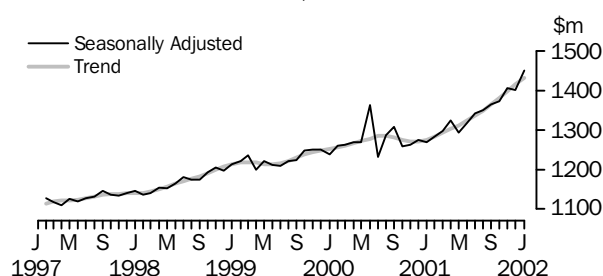
Source: Retail Trade, Australia (Cat. No. 8501.0)

In comparison, the seasonally adjusted Unemployed Females series shown below is relatively more volatile than retail sales and is therefore not as clear an indicator of the underlying direction of the series.

UNEMPLOYED FEMALES, Western Australia

Source: Labour Force, Australia, Preliminary (Cat. no. 6202.0)

The actual process for removing the calendar related effects is complex and will not be discussed in this article. Users who are interested in a technical explanation are referred to *Information Paper: An Introductory Course on Time Series Analysis* (Cat. no. 1346.0). In general, there are two approaches to the seasonal adjustment process:

- forward factors – where seasonal factors are estimated once a year and then kept fixed for a 12 month period; or

- concurrent adjustment – where the seasonal factors are re-estimated each time there is new data available.

Most ABS series use the forward factor approach.

The ABS recommends that at least seven years of data be used to ensure that the results of the seasonal adjustment process are reliable, as it can take some time for seasonal patterns to evolve. Experimental estimates are possible with fewer observations, although a minimum of five years of data is preferable.

CALCULATING THE TREND

Once the original data has been seasonally adjusted, the underlying trend of that series can be estimated by removing the irregular effects. This can be done by applying a moving average to the seasonally adjusted series. The ABS uses a *Henderson moving average* because it is able to dampen the irregular component without distorting the timing of turning points, it is relatively reliable and is easy to produce.

A 7-term Henderson moving average is generally used to smooth quarterly series while a 13-term is used for monthly series. This means that there are seven and thirteen data points respectively used to calculate the smoothed figure. The Henderson moving average is described as being 'centred' because the resulting values are placed in the centre of the series. For example, in the case of the 7-term moving average, the smoothed figure at time t is calculated using three past data points (up to time t–3), the data point at time t, and three future data points (up to time t+3), and the resulting moving average value is placed at time t.

The mathematical formula for the 7-term Henderson moving average is:

$$A_t = \sum_{i=t-3}^{t+3} w_i x_i$$

where A_t is the smoothed data at time t (the trend),

w_i are the weights, and

x_i are the seasonally adjusted data points.

The weights assign an importance to each data point in the calculation. There are specific techniques for deriving weights for different moving averages. For the 7-term symmetric Henderson moving average, the weighting pattern is:

(–0.059, 0.059, 0.294, 0.412, 0.294, 0.059, –0.059)

That is, the trend figure at time t is calculated as:

$$A_t = -0.059x_{t-3} + 0.059x_{t-2} + 0.294x_{t-1} + 0.412x_t + 0.294x_{t+1} + 0.059x_{t+2} - 0.059x_{t+3}$$

For example, suppose the following hypothetical data corresponds to seasonally adjusted
quarterly jelly bean production data from the factory discussed in an earlier example.

HYPOTHETICAL JELLY BEAN PRODUCTION ('000)

|	1998		1999				2000				2001				2002
	Sep	Dec	Mar	Jun	Sep	Dec	Mar	Jun	Sep	Dec	Mar	Jun	Sep	Dec	Mar
Seasonally Adjusted	291.2	300.3	313.7	318.4	320.0	309.7	298.2	298.1	283.9	285.5	286.2	283.0	277.1	285.2	295.6
Henderson Moving Average (Trend)				318.7	317.3	310.5	301.6	293.3	288.1	285.6	283.7	281.7			

The trend figure for June 1999 would be calculated as:

$$A_t = -(0.059 \times 291.2) + (0.059 \times 300.3) + (0.294 \times 313.7) + (0.412 \times 318.4) +$$
$$(0.294 \times 320.0) + (0.059 \times 309.7) - (0.059 \times 298.2)$$

$$= 318.7$$

The trend series can only be calculated using this formula for the middle time periods
because there are insufficient data points available at the ends of the series. That is, the
above table shows that the latest time period for which trend data are available is June 2001
(281.7). To calculate a trend figure for September 2001 would require data for June 2002,
which is yet to be collected. This is known as the *end point problem* and can be overcome
by using asymmetric Henderson moving averages. That is, instead of using the symmetric
weights provided above, asymmetric weighting patterns (which do not require the three
future data points) are used. The asymmetric weighting patterns vary for each time period
and across data series, hence have not been included here.

The appropriate asymmetric weighting patterns have been used to calculate a trend figure
for September 2001, December 2001 and March 2002 in the above table and the following
graph shows the full jelly bean production series. It can be seen that the seasonally
adjusted jelly bean production data is relatively stable with respect to the trend series, and
that the factory's production of jelly beans is slowly starting to increase after declining since
September 1999.

HYPOTHETICAL PRODUCTION OF JELLY BEANS
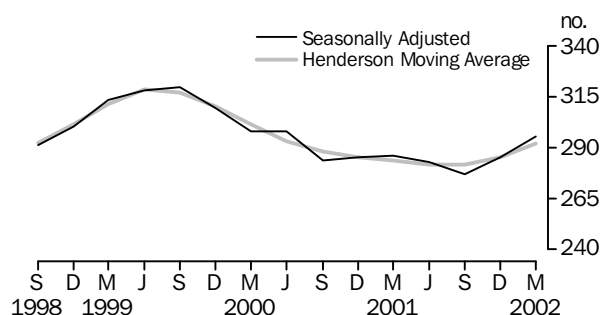
ISSUES TO BE AWARE OF

When analysing seasonally adjusted or trend data, there are a number of important issues that users need to be aware of. These are described below.

Revisions

Revisions to the seasonally adjusted and trend data are common and can occur for a number of reasons.

One of the major reasons for trend data revision is the 'end point problem' discussed earlier. That is, since there are insufficient data points available toward the ends of the series to use the standard smoothing technique, asymmetric Henderson moving averages are used. When the next data point becomes available, the type of moving average used (i.e. symmetric or asymmetric) is shifted across to the next time period, which results in changes to the trend estimates.

For example, the following table shows that when data for the March 2002 reference period is released, the September 2001, December 2001 and March 2002 trend estimates are calculated using asymmetric Henderson moving averages. When data for the June 2002 reference period become available, the September 2001 trend estimates are re-calculated using the standard symmetric moving average. Furthermore, the availability of a new data point affects the values calculated in December 2001 and March 2002, which are also revised.

END POINT PROBLEM: Timing of Symmetric and Asymmetric Moving Averages

TREND ESTIMATES ....

Reference Period	Jun 2000	Sep 2000	Dec 2000	Mar 2001	Jun 2001	Sep 2001	Dec 2001	Mar 2002	Jun 2002	Sep 2002
Mar 2002	Sym	Sym	Sym	Sym	Sym	**A-Sym**	**A-Sym**	**A-Sym**		
Jun 2002	Sym	Sym	Sym	Sym	Sym	Sym	**A-Sym**	**A-Sym**	**A-Sym**	
Sep 2002	Sym	Sym	Sym	Sym	Sym	Sym	Sym	**A-Sym**	**A-Sym**	**A-Sym**

As a result of the end point problem, the most current trend estimate can be revised up to three times in a quarterly series and up to six times in a monthly series. Typically, the largest trend revisions occur the first time new data are available and are generally negligible after the first revision for quarterly series and after the third revision for monthly series.

Revisions can also be made to the seasonally adjusted series as a result of evolving seasonal patterns and/or trading day effects. Unlike trend revisions, which typically affect the last few data points, the method used to revise seasonal factors results in a minimum of five years worth of seasonally adjusted data being affected.

Any revisions which are made to the seasonally adjusted data will flow through to trend series revisions (although they have a small impact on the trend data). Similarly, any amendments made to the original data will flow through to both the seasonally adjusted and trend series. Generally, the degree of revision of the seasonally adjusted and trend data depends on the irregularity of the original series.

Structural Breaks

Long spans of time series data are rarely consistent. They are prone to the effects of structural changes, such as changes in data item definitions, changes in the coverage of the collection, changes in administrative practices, technological innovation and social changes. Such changes can result in an abrupt discontinuity in the underlying level of the original series. This effect is generally referred to as a 'trend break'. For example, consider the new motor vehicle sales series shown below. There is a clear and abrupt increase in the underlying level of the series between June and July 2000 due to the introduction of TNTS.

NEW MOTOR VEHICLE SALES, Western Australia
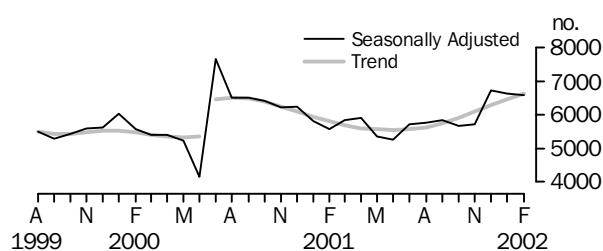
Source: Sales of New Motor Vehicles, Electronic Delivery (Cat. no. 9314.0.55.001)

A 'seasonal break' can occur when the seasonal behaviour of the series abruptly changes from one year to the next. For example, consider the Commonwealth Government benefit payments series below. This series includes education and training payments such as Austudy. The mild seasonal pattern which can be observed from 1990 to 1995 changes abruptly in 1996 when the timing of Austudy payments changed. The seasonal pattern changed again in 1998 when the timing of fortnightly government payments were changed to be made on any day of the week. The series also shows a trend break in 2000 due to the Sole Parents Pension being taken over by Centrelink (and the corresponding data being included in another series).

COMMONWEALTH GOVERNMENT BENEFIT PAYMENTS:
Other Than Health and Social Security
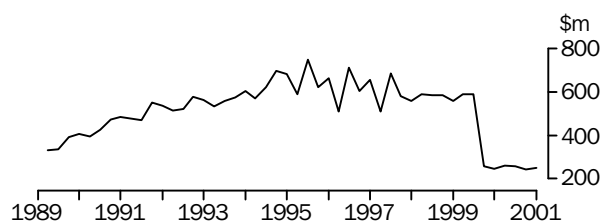
Source: Australian National Accounts: National Income, Expenditure and Product (Cat. no. 5206.0)

Outliers

Time series data can be subject to large, one-off effects. These effects will remain in the seasonally adjusted series and can distort the trend path if they are not corrected during the trending process.

For example, the following graph shows extremes in the number of Commonwealth wage and salary earners during the conduct of the 1986 Census, the 1987 Federal election, the 1988 Referendum, the 1991 Census and the 1993 Federal election, due to the employment of additional temporary staff. More recent elections and censuses have possibly used different employment arrangements which do not appear as large extremes.
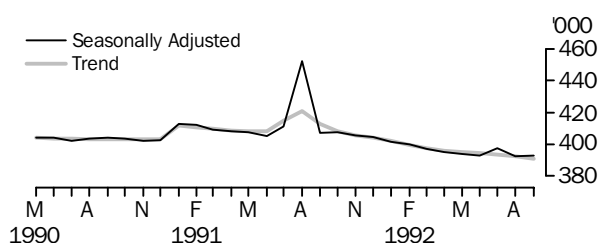
Outliers *continued*

COMMONWEALTH WAGE AND SALARY EARNERS,
Australia
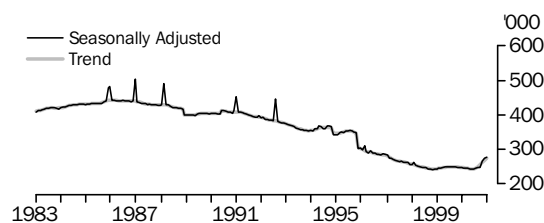
Source: Wage and Salary Earners, Australia (Cat. no. 6248.0)

If these extremes were not taken into consideration during the trending process, the trend line would be distorted, as shown below using the 1991 Census as an example.

COMMONWEALTH WAGE AND SALARY EARNERS,
Australia

Source: Wage and Salary Earners, Australia (Cat. no. 6248.0)

WHICH SERIES TO USE?

The original, seasonally adjusted and trend series are all useful measures for time series analysts. They do, however, serve different purposes and it is important to be able to distinguish which is the most appropriate series to use under different circumstances.

Often, users are interested in analysing the underlying direction of the series, unobscured by any seasonal or irregular effects, and in detecting possible turning points in the series. In such circumstances, the trend series would be the most appropriate to use as all seasonal and irregular effects have been removed.

While the trend series provides useful information about the underlying direction of the data, it does not provide any information about the seasonal patterns in the data. Some users may be interested in, for example, the relative magnitudes of the seasonal peaks and troughs from year to year, or how the seasonal effects have evolved over the years. In this case, the original data, which has not had the seasonal effects removed, would be the most appropriate. Users who are interested in comparing one month to the next may find the seasonally adjusted data more useful than the original as it is not obscured by seasonal patterns.

Some users may be interested in which months are the most or least irregular, or how much the irregularity is changing over time. Since the irregularity is removed from the trend series, the user would be interested in analysing the seasonally adjusted data. Other users may be interested in measuring the magnitude of the irregular so as to line it up with economic events or a change in government policy. For example, users may be interested in the magnitude of the impact of the Goods and Services Tax on retail turnover figures. Again, seasonally adjusted data would be the most appropriate for such purposes.

NON-ABS TIME SERIES

Time series data are collected by a wide range of government and non-government organisations and the concepts described above, regarding the analysis of such data, are not solely applicable to ABS series.

This article describes basic time series analysis concepts. It does not explain the complex statistical techniques actually used. ABS statistical consultants are available to assist external organisations with analysis of non-ABS time series. For further information and advice, contact the manager of Statistical Consultancy on (08) 9360 5144.

RELATED ABS PUBLICATIONS

Information Paper: A Guide to Smoothing Time Series - Estimates of "Trend" (Cat. no. 1316.0)

Information Paper: Time Series Decomposition - An Overview (Cat. no. 1317.0)

Information Paper: An Introductory Course on Time Series Analysis (Cat. no. 1346.0)

Information Paper: A Guide to Interpreting Time Series - Monitoring "Trends" An Overview (Cat. no. 1348.0)

Australian Economic Indicators, April 1991 (Cat. no. 1350.0) — Article titled "Picking Turning Points in the Economy"

Australian Economic Indicators, March 1992 (Cat. no. 1350.0) — Article titled "Smarter Data Use"

Australian Economic Indicators, January 1995 (Cat. no. 1350.0) — Article titled "A Guide to Interpreting Time Series"