



GLOBAL JOURNAL OF SCIENCE FRONTIER RESEARCH: F
MATHEMATICS AND DECISION SCIENCES
Volume 15 Issue 9 Version 1.0 Year 2015
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Time Series Decomposition and Seasonal Adjustment

By Maskurul Alam, Matiur Rahman, Sharmin Akter Sumy & Yasin Ali Parh

Islamic University, Bangladesh

Abstract- Many forecasting method are based on some notion that when an underlying pattern exists in a data series. That data can be distinguished from randomness by smoothing (averaging) past values. The effect of this smoothing is to eliminate randomness so the pattern can be projected into the future. It goes without saying that when a data is good enough and have nice pattern then forecast could be done more precisely. One of the main objectives for decomposition is to estimate seasonal effects that can be used to create and present seasonally adjusted values. A seasonally adjusted value removes the seasonal effect from a value so that trends can be seen more clearly. My main aim is to choose a best decomposition method and forecast the data more precisely.

Keywords: *time series decomposition, decomposition models, seasonal adjustment, moving average smoother.*

GJSFR-F Classification : *MSC 2010: 49M27*



Strictly as per the compliance and regulations of :



© 2015. Maskurul Alam, Matiur Rahman, Sharmin Akter Sumy & Yasin Ali Parh. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License <http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



Time Series Decomposition and Seasonal Adjustment

Maskurul Alam ^α, Matiur Rahman ^σ, Sharmin Akter Sumy ^ρ & Yasin Ali Parh ^ω

Abstract- Many forecasting method are based on some notion that when an underlying pattern exists in a data series. That data can be distinguished from randomness by smoothing (averaging) past values. The effect of this smoothing is to eliminate randomness so the pattern can be projected into the future. It goes without saying that when a data is good enough and have nice pattern then forecast could be done more precisely. One of the main objectives for decomposition is to estimate seasonal effects that can be used to create and present seasonally adjusted values. A seasonally adjusted value removes the seasonal effect from a value so that trends can be seen more clearly. My main aim is to choose a best decomposition method and forecast the data more precisely.

Keywords: *time series decomposition, decomposition models, seasonal adjustment, moving average smoother.*

I. INTRODUCTION

In many instances the pattern of the data can be broken down (decomposed) into sub pattern that identify each component of the time series separately. Such break down of the data can give the better ideas about the understanding the behavior of the series which facilitates improves accuracy in forecasting. Decomposition method usually try to identify two separate useful components of the basic underline pattern that tend to characterize economic and business series. These are the trend cycle and seasonal factors. The seasonal factors relates to periodic fluctuations of constant length that are usually caused by such things as temperature, rainfall, month of the year, timing of holidays and corporate policies. The trend cycle represents the longer term changes in the level of the series. The trend cycle sometime could be separated into two components. These are trend and cycle components. But the distinction is somewhat artificial and most procedures leave the trend and cycle as a single component known as the trend-cycle.

II. TIME SERIES DECOMPOSITION

Decomposition assume that the data are made as follows

$$Data = pattern + error$$

Now it is necessary to mention that the pattern of any data may form trend cycle and seasonality. This means a trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend “changing direction” when it might go from an increasing trend to a decreasing trend.

Author ^{α σ ρ ω}: Islamic University, Kushtia, Bangladesh. e-mail: maskurulra@gmail.com

On the other hand a seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Seasonality is always of a fixed and known period.

So we could give a standard form of the decomposition time series on the basis of the pattern of the data.

$$Data = pattern + error$$

$$=f(trend - cycle, seasonality, error)$$

An element of the error or randomness is also assumed to be present in the data. It is actually the combined effect of the two sub patterns of the series. This means the combined effect of the trend-cycle, seasonality and the actual data. This is often called the “irregular” or the “reminder” component.

It goes without saying that there are several alternative approaches to decomposing a time series all of which aim to isolate each component of the series with great accuracy and precisely. Actually the main substance is to remove the trend cycle and then isolating the seasonal component.

III. DECOMPOSITION MODELS

As we discussed earlier about the decomposition we could give the following form

$$Y_t = f(S_t, T_t, E_t)$$

Where Y_t it is the time series value (actual data) at period t ,

S_t Is the seasonal component (or index) at period t ,

T_t Is the trend cycle component at period t ,

And E_t is the irregular (or reminder) component at period t

There are two common method of decomposition these are

1. Additive decomposition and
2. Multiplicative decomposition

Additive decomposition:

A common approach is to assume the addition of seasonal component, trend cycle component and the irregular component.

$$Y_t = S_t + T_t + E_t$$

An additive model is usually appropriate if the magnitude or the span of the seasonal fluctuation doesn't vary with the level of the series. It actually means when the magnitude of the seasonal fluctuation remain same then additive decomposition is used.

Multiplicative decomposition

A common approach is to multiply the seasonal, trend cycle and irregular components together to give the observed series.

$$Y_t = S_t * T_t * E_t$$

A multiplicative decomposition is usually apply when the seasonal fluctuations increase and decrease proportionally with increase and decreases in the level of the series. Multiplicative decomposition is more apposite for the economic series because most seasonal economic series have seasonal variation and even it may vary for day, week, and month as well as for year.

Now either choosing an additive or multiplicative decomposition we could use a transformation. When the original data are not additive then logarithm transformation turns a multiplicative relationship into an additive relationship.

$$Y_t = S_t * T_t * E_t$$

Then $\log Y_t = \log S_t + \log T_t + \log E_t$

So we can fit a multiplicative relationship by fitting an additive relationship to the logarithm of the data.

IV. DECOMPOSING SEASONAL DATA WITH THE HELP OF ADDITIVE DECOMPOSITION MODEL

To estimate the trend component and seasonal component of a seasonal time series that can be described using an additive model, we can use the “decompose ()” function in R. This function estimates the trend, seasonal, and irregular components of a time series that can be described using an additive model.

We explain it through an example. We are considering a birth time series data of New York.

| Time | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1946 | 26.663 | 23.598 | 26.931 | 24.740 | 25.806 | 24.364 | 24.477 | 23.901 | 23.175 | 23.227 | 21.672 | 21.870 |
| 1947 | 21.439 | 21.089 | 23.709 | 21.669 | 21.752 | 20.761 | 23.479 | 23.824 | 23.105 | 23.110 | 21.759 | 22.073 |
| 1948 | 21.937 | 20.035 | 23.590 | 21.672 | 22.222 | 22.123 | 23.950 | 23.504 | 22.238 | 23.142 | 21.059 | 21.573 |
| 1949 | 21.548 | 20.000 | 22.424 | 20.615 | 21.761 | 22.874 | 24.104 | 23.748 | 23.262 | 22.907 | 21.519 | 22.025 |
| 1950 | 22.604 | 20.894 | 24.677 | 23.673 | 25.320 | 23.583 | 24.671 | 24.454 | 24.122 | 24.252 | 22.084 | 22.991 |
| 1951 | 23.287 | 23.049 | 25.076 | 24.037 | 24.430 | 24.667 | 26.451 | 25.618 | 25.014 | 25.110 | 22.964 | 23.981 |
| 1952 | 23.798 | 22.270 | 24.775 | 22.646 | 23.988 | 24.737 | 26.276 | 25.816 | 25.210 | 25.199 | 23.162 | 24.707 |
| 1953 | 24.364 | 22.644 | 25.565 | 24.062 | 25.431 | 24.635 | 27.009 | 26.606 | 26.268 | 26.462 | 25.246 | 25.180 |
| 1954 | 24.657 | 23.304 | 26.982 | 26.199 | 27.210 | 26.122 | 26.706 | 26.878 | 26.152 | 26.379 | 24.712 | 25.688 |
| 1955 | 24.990 | 24.239 | 26.721 | 23.475 | 24.767 | 26.219 | 28.361 | 28.599 | 27.914 | 27.784 | 25.693 | 26.881 |
| 1956 | 26.217 | 24.218 | 27.914 | 26.975 | 28.527 | 27.139 | 28.982 | 28.169 | 28.056 | 29.136 | 26.291 | 26.987 |
| 1957 | 26.589 | 24.848 | 27.543 | 26.896 | 28.878 | 27.390 | 28.065 | 28.141 | 29.048 | 28.484 | 26.634 | 27.735 |
| 1958 | 27.132 | 24.924 | 28.963 | 26.589 | 27.931 | 28.009 | 29.229 | 28.759 | 28.405 | 27.945 | 25.912 | 26.619 |
| 1959 | 26.076 | 25.286 | 27.660 | 25.951 | 26.398 | 25.565 | 28.865 | 30.000 | 29.261 | 29.012 | 26.992 | 27.897 |

It is roughly could be mentioned from the above time series data the number of births per month in New York city is seasonal with a peak every summer and trough every winter, and can probably be described using an additive model since the seasonal and random fluctuations seem to be constant in size over time. This means the number of birth in New York per month roughly constant.

To see it visually we could use R program
`birthstimeseries= ts (births, frequency=12, start=c(1946,1))`
`birthstimeseries`

plot (birthstimeseries)

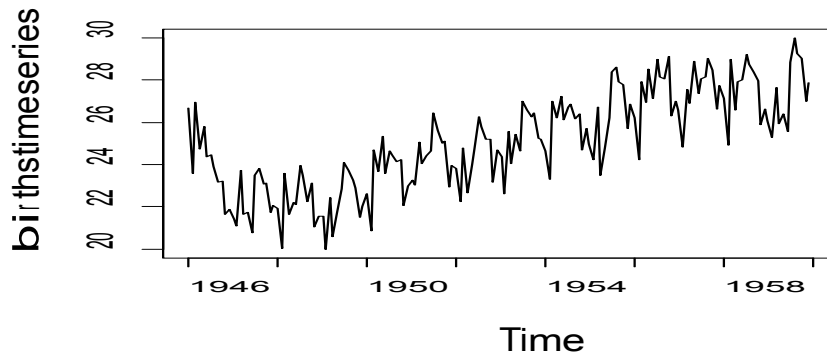


Figure : the plot of births time series from 1946 to 1959.

Now it is manifested to us from above figure that the number of birth per month in New York roughly constant but the data is seasonal in peak summer and trough every winter.

Now we estimate the trend-cycle, seasonal and irregular components in order to use additive model .

These could be estimated through R programming.

```
Births time series components=decompose(births time series)
Births time series components
```

Now estimated all the components (trend-cycle, seasonal and irregular) are stored into the birth time series components variable.

We can plot the estimated trend, seasonal and irregular components by using R command

Plot(

```
Plot(Births time series components)
```

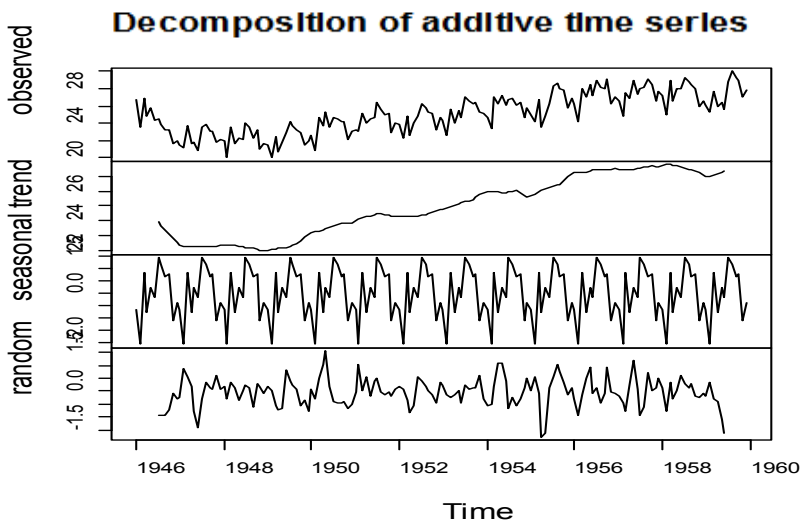


Figure : The plot births time series of Time series components

It could be mentioned from the above figure that the estimated seasonal component doesn't change much over time. The seasonal pattern at the start of the

series is almost same as the pattern at the end of the series. On the other hand trend component shows a small decrease from about 24 in 1947 to about 22 in 1948, followed by a steady increase from then on to about 27 in 1959.

V. SEASONAL ADJUSTMENT

It is necessary to say that when a seasonal data is subtracted from the main data then the resulting values are referred to as “seasonal adjustment” data. The additive model is given by

$$Y_t = S_t + T_t + E_t$$

So the seasonal adjusted holds the following form

$$Y_t - S_t = T_t + E_t$$

This means leaving only trend cycle and irregular component.

And for multiplicative data seasonal data can be found by dividing the main data to the seasonal data. Mathematically it is given by

$$Y_t/S_t = T_t * E_t$$

Most published economic data series are seasonally adjusted because seasonal variation is typically not primary interest. The seasonally adjusted data series shows the data after any seasonal variation has been removed.

For an example we consider the previous example of birth time series data.

The seasonal component can be found with the help of R programming. We type simply “decompose ()” in R and then subtract the seasonal component from the main time series data. It is usually done in order to remove the seasonal variation.

```
Birthstimeseriescomponents=decompose (birthstimeseries)
Birthstimeseriescomponents
Birthstimeseriesseasonallyadjusted=birthstimeseries-
birthstimeseriescomponents$seasonal
Birthstimeseriesseasonallyadjusted
```

We can plot the seasonally adjusted data in order to see the other components. we can type the following command in R

```
“plot(Birthstimeseriesseasonallyadjusted)”
```

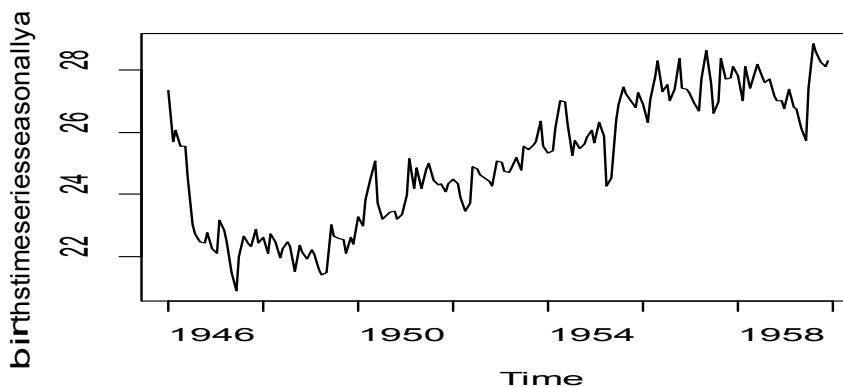


Figure : The plot of seasonal adjusted time series

It is manifested to us that the seasonal variation has been removed from the seasonally adjusted data. It is certainly say from here that the seasonally adjusted data just trend and irregular variation contain. Clearly saying first of all we had observed data, seasonal component, trend component and irregular component. After removing the seasonal variation from the adjusted data we have just trend and irregular component. It can be shown after observing the two figures.

VI. MOVING AVERAGE SMOOTHER

It is known to all that the trend cycle can be estimated by smoothing the series to reduce the random variation. There are number smoothers are available for reducing the random variation. The oldest and simplest method is moving average that could be used in order to reduce the random variation.

A moving average for m order can be written as

$$Tt = \frac{1}{k} \sum_{j=-m}^{km} yt + j$$

Where k is moving average of order k (or MA), k is an odd integer and it is defined as the average consisting of an observation and the m= (k-1)/2 points of either side. Observations that are nearby in time are also likely to be close in value, and the average eliminates some of the randomness in the data, leaving a smooth trend-cycle component. We call this an k-MA meaning a moving average of order k

We explain through an example

Suppose we have sales of detergent (in liters) over three year's period.

| year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1989 | 266.0 | 145.9 | 183.1 | 119.3 | 180.3 | 168.5 | 231.8 | 224.5 | 192.8 | 122.9 | 336.5 | 185.9 |
| 1990 | 194.3 | 149.5 | 210.1 | 273.3 | 191.4 | 287.0 | 226.0 | 303.6 | 289.9 | 421.6 | 264.5 | 342.3 |
| 1991 | 339.7 | 440.4 | 315.9 | 439.3 | 401.3 | 437.4 | 575.5 | 407.6 | 682.0 | 475.3 | 581.3 | 646.9 |

Now our main purpose is to reduce the trend variation as much as possible. Taking average of the points near an observation will provide the reasonable estimate of the trend cycle at that observation. The average eliminates some of the randomness in the data. It is necessary to know how many data points to include in each average. Suppose we use the average of the three points, namely the observation at which we are calculating trend cycle and the points on either side. Clearly saying if we want to estimate the trend cycle of February month at 145.9 in 1989 then we just consider this value previous values(266.0) and and the next value(183.1) of it.This is called the moving average of order Three.

Mathematically it is given by

$$T2 = \frac{1}{3} (Y1 + Y2 + Y3) = \frac{1}{3} (266.0 + 145.9 + 183.1) = 198.3$$

Generally a moving average of order 3 centered at time t is $= \frac{1}{3} (Yt - 1 + Yt + Yt + 1)$

The 3MA can be determined with the help R programming. But it is bear in mind that first of all data must be read in R. We type the following command for reading as well as for determining 3 MA.

```
timeseries<- ts(mas, frequency=12, start=c(1989,1))
timeseries
plot(timeseries)
z=mav (timeseries,3)
z.
```

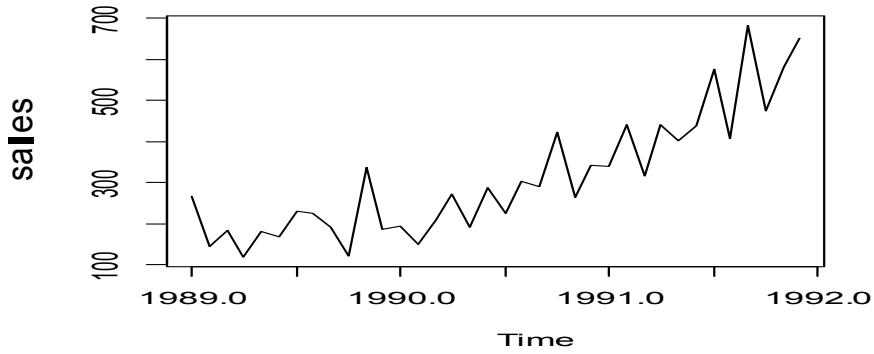


Figure : The sales of detergents over three years

The output of 3 MA in R

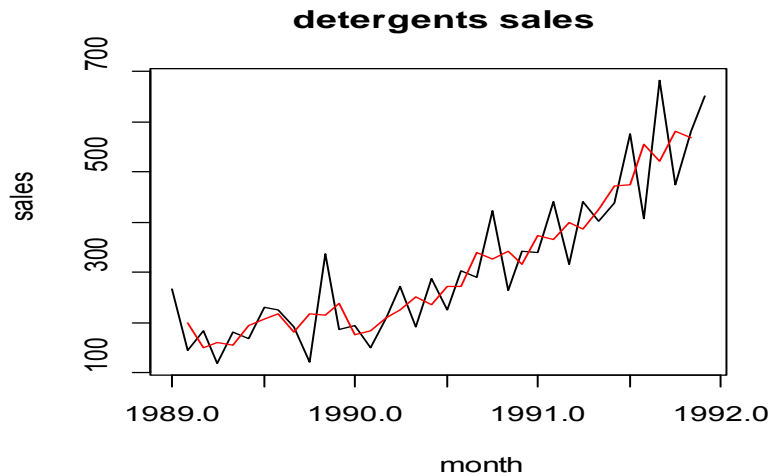


Figure : 3 MA smoother

```
[1,] NA
[2,] 198.3333
[3,] 149.4333to see what the trend estimate look like we plot it wth the original data.
[4,] 160.9000 we could make the above 3 order MA plot with the help of R command.
[5,] 156.0333plot(timeseries, main="detergents sales", ylab="sales",xlab="month")
[6,] 193.5333 lines(mav(timeseries,3),col="red") .....
[34,] 579.5333
[35,] 568.8333
[36,] NA
```

In the second column of this table a moving average of order 3 providing an estimate of trend cycle.

Notice how the trend (in red) is smoother than the original data and captures the main movement of the time series without all the minor fluctuations. It goes without saying that the order of the moving average determines the smoothness of the trend cycle estimate. It is certainly could be told that the higher order means a smoother curve. So we could make 5 order moving average in order make smoother. For

the 5 order moving average we could determine the trend cycle with the help of R that saves our time compare to other methodology.

We could use the following command in order to see the how the trend look like and the

```
k=mav(timeseries,5)
k
plot(k)
plot(mas,main="detergents sales",ylab="sales", xlab="month")
lines(mav(mas,5),col="green")
output of the 5 order Moving average.
```

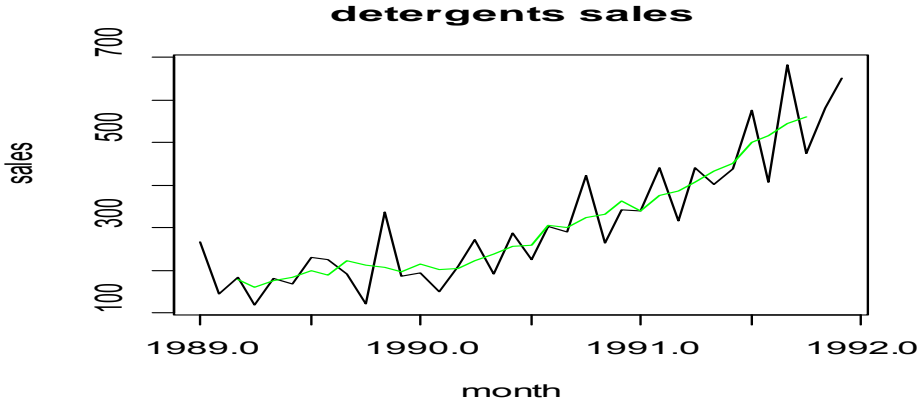


Figure : 5 MA smoother

- [1,] NA
- [2,] NA
- [3,] 178.92
- [4,] 159.42 so it could be mentioned from the above two 3 MA and 5 MA that
- [5,] 176.603 MA is smoother leaves too much randomness in the trend cycle
Estimate. It could be mentioned that the 5 MA smoother is better.
- [34,] 559.22 but the trend cycle is probably smoother for other orders.
- [35,] NA
- [36,] NA

It should be bear in mind that determining the appropriate length of a moving average is an important task in decomposition method. As a rule a larger number of terms in the moving average increase the likelihood that randomness will be eliminated.

VII. CONCLUSION

Time series data can exhibit a huge variety of patterns and it is helpful to categorize some of the patterns and behaviors that can be seen in time series. It is also sometimes useful to try to split a time series into several components, each representing one of the underlying categories of pattern. Decomposition often plays the vital role to make time series better as well as improve the forecast.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. New York: Oxford University press Kendall, M. G. (1976). *Time-Series (Second ed.)*. Charles Griffin

- ANDERSON, o. and U. NoCHMALS (1914) The elimination of spurious correlation due to position in time or space, *Biometrika*, 10, 269-276
2. BAXTER, M.A. (1994) "A guide to seasonal adjustment of monthly data with X-11," 3rd ed., Central Statistical Office, United Kingdom.
 3. BELL, W.R. and S.C. HILLMER (1984) Issues involved with the seasonal adjustment of economic time series, *Journal of Business and Economic Statistics*, 2, 291-320
 4. BROWN, R.G. (1963) Smoothing, forecasting and prediction of discrete time series, Englewood Cliffs, N.J.: Prentice Hall BURMAN, J.P. (1979) Seasonal Adjustment: A survey, *TIMS Studies in management Sciences*, 12, 45-57
 5. CLEVELAND, W.S. I.J. TERPENNING (1992) Graphical methods for seasonal adjustment, *Journal of the American statistical Association*, 77, 52-62
 6. DAGUM, E.B. (1982) Revisions of time varying seasonal filters, *Journal of Forecasting*, 1, 20-28.
 7. DAGUM, E.B. (1988) X-11 ARIMA/88 seasonal adjustment method: foundations and users manual, Statistics Canada.
 8. DEN BUTTER, F.A.G and M.M.G FASE (1991) *Seasonal adjustment as a practical problem*, Amsterdam: north- Holland.
 9. FINDLEY, D.F. and B.C. MONSELL (1989) Reg- ARIMA based pre-processing for seasonal adjustment. In *Analysis of data in time*, ed, A.c. Singe and P whitridge, 117-123 Ottawa, Canada.

This page is intentionally left blank