

*The MHTS PC Package*

*REFERENCE MANUAL*

©1999, A.I. McLeod & K.W. Hipel

Reference Manual Version: 1.03

*INTRODUCTION*

## *Prologue*

The McLeod-Hipel Time Series (MTS) Package provides a comprehensive decision support system for the analysis and forecasting of many kinds of time series data. In addition to a state-of-the-art approach to Box-Jenkins modelling and forecasting many other topics are covered including: Forecasting, Simulation, Spectral Analysis, Periodic Time Series Modelling, Smoothing and Seasonal Adjustment and much more. The MTS Package does, in fact, furnish an unparalleled coverage of time series modelling and analysis. Many of the methods discussed in the textbook by Hipel and McLeod (1994) are implemented in the MTS Package and this package makes an excellent companion for the textbook.

### Availability

There are three versions of the MTS Package:

1. Student Version
2. Professional Version
3. Network Version

With the Student Version, one can work with about 40 preselected datasets. The Student Version is available at no charge. For the Professional Version, please contact McLeod (e-mail: aim@uwo.ca) or Hipel (519-885-1211). The manual and the Student Version are available on the Internet via anonymous ftp from fisher.stats.uwo.ca in the files pub/mhts/mhts.zip. and pub/mhts/rmts.zip respectively.

You will need to use binary transfer and the -d option to pkunzip.

Assuming you have ftp and an Internet connection. Login as follows:

```
ftp fisher.stats.uwo.ca
```

Give **anonymous** as your login name and for password use your e-mail address. Next enter the following ftp commands:

```
cd pub/mhts
get README
binary
get mhts.zip
get rmts.zip
quit
pkunzip -d mhts.zip
pkunzip rmts.zip
```

The Reference Manual for the MHTS Package is in the file rmts.ps which can be viewed using the unix utility ghostview and can be printed on a postscript printer.

Please check the README file for any additional information.

If you don't have pkunzip.exe on your PC, then you should also download unzip.exe:

```
get unzip.exe
```

As with mhts.zip and rmts.zip binary ftp transfer is needed for downloading unzip.exe.

With the Professional Version any dataset may be used. This version is designed for stand-alone PCs.

The Network Version is designed to be run on a local area network. It may also be run on a stand-alone PC.

The Professional and Network Versions may be obtained by upgrading the Student Version. This upgrade is accomplished by running the executable setup and using a password which will be given to you when you purchase the upgrade.

## Updates to the Package

Please report bugs and/or suggests to A.I. McLeod via e-mail at aim@uwo.ca.

If you have the professional version, you can obtain updates by following the anonymous ftp instructions above. You should unzip the file mhst.zip in a fresh directory and then replace the files mts.exe, mtsa.ovl, ..., mtse.ovl in your old directory with the new ones.

## Version Numbers

In general the version number to the program mhst.exe may be slightly larger than the version number to the MTS Reference Manual since every time bug fixes or corrections are made to the package, the package version number is incremented but the Reference Manual version number will not be changed unless changes to the Reference Manual are also made.

## McLeod-Hipel Datasets

Numerous interesting datasets which are designed to accompany this package and the book by Hipel and McLeod (1994) are also available by anonymous ftp. There is both a unix version, in shar format, and a PC version, in zip format. These datasets are available via anonymous ftp from fisher.stats.uwo.ca in the files pub/mhst/mhsets.sh and pub/mhst/mhsets.zip for the unix and PC versions respectively.

## Getting Started

Simply change to the subdirectory MTS which contains program. Make sure that the mouse driver is loaded. Then enter the command:

```
mts
```

## MTS Directories

The program itself is normally located in a subdirectory mts . On networks, this may be a network drive. The current working directory for mts contains further four

subdirectories DATA, OUTPUT, FITS and STL. The DATA directory contains time series data files which can be input to the package. The OUTPUT directory contains tabulations and analyses produced by mts . The FITS directory contains fitted models and is used directly as input to the forecasting functions in the package. The STL directory contains output from the Seasonal-Trend Loess Adjustment function. You should ensure that there are no more than 80 files in either the DATA or the FITS directory. You may use the Move function in the package to delete and move files to another location. The DATA subdirectory may be changed to any directory located anywhere on the computer system using **Set New DATA Directory** from the **Utility** menu.

## Hardware Requirements

The minimal hardware configuration is a PC computer with an Intel 80386 processor and a 80387 math coprocessor, EGA or VGA display monitor, DOS 3.2 or greater, 640K RAM and a hard disk. A wide variety of dot-matrix and laser printers are supported.

It is desirable to have atleast 2 megabytes of extended memory available.

The package requires about 1.5 megabytes of disk space. When you run the package you should try to have at least 580K of DOS RAM available (use the DOS CHKDSK command to verify this). Extended or expanded memory will automatically be used if it is present.

## Main Menu

The **Main Menu** is comprised of two sections. The first section specifies the current data directory and the **Data Mode**. The data directory may be changed by selecting **Utilities** and then selecting **Set New Data Directory**. The **Data Mode** may be set to PROMPT or REMEMBER. When set to PROMPT, you will be prompted for a new data set every time. If the **Data Mode** is set to REMEMBER, the same data set may be used repeatedly with different menu choices after it has been initially selected. A data set is initially selected the first time you run a time series analysis function such as Trace in the the Exploratory Analysis. To select a new data set, return to **Main Menu** and select **Clear Current File**. This former option, **Clear Current File**, is naturally only available after an initial data set has been selected and the **Data Mode** has been set to REMEMBER.

The second section of the **Main Menu** provides a series of time series analysis topics:

- Exploratory Analysis
- Trend Analysis
- Correlation Function
- Spectral Analysis
- Model Estimation
- Diagnostic Checks
- Forecasting

- Simulation
- Filters
- Utilities
- Quit Package

To quit, you can simply press ESC instead of selecting the menu item **Quit Package**. Also ESC will always pop you back out of a sub-menu.

## MTS Installation

There are two methods of acquiring and then installing the Professional or Network versions of the MHTS PC Package. The preferred method of distributing the Professional and Network versions of this software is via anonymous ftp on the Internet. This method has three big advantages.

1. As a registered user of this software you are entitled to free updates so long as the major version, which is currently set to 1, does not change. This software is under active development. Improvements are constantly being made. However, the major version number, 1, will probably not be changed for at least one or two more years.
2. It is more reliable than shipping out disks which may be damaged in the mail.
3. It reduces overhead.

If you do not have access to the Internet, then you can request three diskettes which contain the software and Reference Manual.

### Upgrading From Internet Version

First, you download the Student Version and the MHTS Reference Manual as was explained above under *Availability*.

Next you upgrade the to the Professional or Network version by supplying the ASCII file, sn.dat, and a password which will be sent to you.

### Installing Using The Three Diskettes Method

If you are unable to access the Internet then you will received three diskettes which are labelled INSTALL, PROGRAM I and PROGRAM II.

To install, insert the INSTALL diskette in drive a. Then enter,

```
a:
```

```
install
```

Notice that you must switch to the diskette drive before you run install.

Then you will be prompted for the necessary information to complete the setup. Instead of drive a, you may also use drive b.

Once the MTS Package has been installed, you can back your copy up. Provided you do not change the computer, the hardware configured on your computer or the logical drive and pathname then you can restore the backed-up copy.

The MHTS PC Reference Manual is available as a zipped postscript file on the third diskette.

### Network Installation Notes

Before using, you should define and set the environment variable `MTSHOME` to the location of `MTS.EXE`. For example, if `MTS.EXE` is in a subdirectory `MTS` on `LAD` or virtual drive `N:` then enter

```
SET MTSHOME=N:\MTS
```

This line can be incorporated in a batch file, to invoke the program.

The MHTS PC package requires that it have disk read/write privileges in the current working directory and the `DATA` directory. The default setting for the `DATA` directory is a subdirectory `DATA` off the current working directory. Alternatively, you may initialize the `DATA` subdirectory used by the MHTS PC Package to any convenient subdirectory by putting its pathname in a file called `MTS.CFG`. The `MTS.CFG` file should be put in the current working directory.

## Running MTS Under Windows

The MTS Package can be conveniently run under Windows 3.1 as a full screen application. It is also possible for the Package to run as a window in Windows 386 enhanced mode however this is much slower. Nevertheless, it is sometimes useful to run the Package in a window in order to copy some of the graphics via the Windows Clipboard to other documents.

An icon is provided for the MTS Package in the file `MTS.ICO` in the home directory. Please see the Microsoft Windows User Guide for information on how to implement this icon.

## Printer Configuration

The MTS package can be configured to work with over 300 dotmatrix and laser printers. The size of the plots produced can also be adjusted by the configuration program. Two different printers *must* be configured: an online printer and a remote printer. An online printer is one which is directly connected to the computer running MTS. A remote printer is usually on a network (which is not connected while running MTS) or on another computer. The configuration program is normally run only once when the package is first installed. To run the configuration program, enter

```
psetup
```

from the directory where MTS is installed. If the printer configuration files `ONLINE.PDT` and `REMOTE.PDT` are not present in the directory where MTS is installed, an error message will be displayed when the package is run.

## Menu Selection

The MTS Package normally works best with a mouse to make menu selections. Alternatively if a mouse is not available, the `SPACE-BAR` can be used to make “mouse clicks”.

The only items in the menu's which don't use a mouse are items involving real numbers, such as Box-Cox transformation parameters  $\lambda$  and  $c$ . Use the keyboard to type in the numerical values for such numbers.

In some items, a large integer ( $> 10$ ) may be represented so that one can select the number by clicking on each digit separately.

To speed up selection of items when faced with a long list, one can more easily use the SPACE-BAR. Holding down the SPACE-BAR will rapidly move through the list.

Pressing the escape key, ESC, exits any menu.

## Data Input to MTS

The data are stored in a regular ASCII files in the MTS subdirectory. By default you are provided with the subdirectory DATA containing various time series. You may also use other subdirectories. The use of other subdirectories for you data files is explained in **Set New DATA Directory** in the **Utility** section.

Data can either be typed in using an editor such as EDLIN or EDIT (or even MTS's own screen editor) or exported from another package and copied to the DATA subdirectory.

The first line in the file is, normally, taken as the title. *Do not omit the title since this will cause the first line of data to be taken as a title!* After the first line giving the title string, comment lines can be inserted before the data to document other information about the data. Comment lines begin with a # in column one. Comment lines are skipped over when the data file is read by a MTS function. Unlike the title string in the first line, comment lines are optional and may be omitted. The data follow the comment lines or, when comment lines are not present, the data follow the first line. The data is stored in free format. For a time series, chronological order is assumed. One exception to the above rules for title and comments occur for the MTS menu items **Screen Editor**. In this case all character and numeric information on the file is available for screen editing. Another exception occurs for the MTS utility **Convert: Column to Row Format**. This utility is provided to help convert data which is stored in columns to the particular row format used in .2 (or .3 etc.) files.

The name of the file can be any valid DOS filename but the extension should be .1 for most MTS menu functions. The extension .1 informs MTS that only one set of data is on the file. For example, LYNX.1 is a valid filename. Some functions in the package allow extensions .2, .3, ..., etc. if there are more than one series on the same file. The format for a .2 file is:

First line: title string

Next lines: any comments beginning with a # in column 1

Next lines: the y-series

Next lines: the x-series Similarly for .3 and higher files.

Next, after the file has been created, you must put this data file in the MTS subdirectory DATA. You can do this with the DOS COPY command. For example,

```
copy lynx.1 data\*.*
```

copies the file lynx.1 from the current directory to the MTS subdirectory DATA.



For best performance, avoid having too many data files on the MTS DATA directory (a maximum limit of 90 is allowed). Use the MTS Move File Function or the DOS delete command to remove unwanted data files.

Initially, when the MTS package is installed the DATA directory is called DATA and it is a subdirectory off the current working directory. However, any directory located anywhere on the computer system may be used as a DATA directory. It does not have to be named DATA. To use a different DATA directory from the default, select **Set New DATA Directory** from the **Utility** menu.

## Data Transformations in MTS

With most of the functions in the MTS package you can make a simple set of transformations of the data sequence  $z_1, z_2, \dots, z_n$ . The general form of these transformations can be expressed by the equation

$$w_t = \nabla^d \nabla_s^{d_s} z_t^{(\lambda)},$$

where,

$w_t$  is the transformed data sequence,

$\lambda$  is the exponent in the Box-Cox transformation,

$\nabla$  is the first differencing operator,

$d$  is the differencing parameter,

$\nabla_s$  is the seasonal differencing operator,

$s$  is the seasonal period,

$d_s$  is the seasonal differencing parameter.

The default settings of these parameters is  $d = d_s = s = 0$  and  $\lambda = 1.0$  which corresponds to no transformation.

The general form of the Box-Cox transformation equation may be written

$$\begin{aligned} z_t^{(\lambda)} &= \frac{(z_t + c)^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0. \end{aligned}$$

In this transformation,  $c$  is usually 0, unless some of the original data values are 0 or negative in which case  $c$  is a small positive constant which is added to all the data values to make them all positive. The default setting of  $c$  is  $c = 0$  in the MTS package.

See Appendix 2 for more about this family of transformations.

## Text Output With MTS

Tabulations of autocorrelations and of fitted models are some examples of text output that may be generated. This output is also shown on the screen at the time it is

generated and is saved in the directory OUTPUT. The filename for the output is shown at the top of the screen when the file is displayed.

You can view the files in the OUTPUT directory at any time by selecting the sub-menu **Editor and Utilities** in the Main Menu. Then select **View File on OUTPUT directory**.

To print a file you are viewing, press the key P. Alternatively, files in the OUTPUT directory, can be printed with the DOS PRINT command.

All files are in ASCII format so they can be edited with a regular editor as well as the screen editor provided in MTS.

## Graphics Hardcopy With MTS

First, in order to obtain graphics hardcopies, it is necessary that the printer be configured correctly. Normally this is done only once when the MTS package is first installed.

Two methods of making hardcopies are supported. If a suitable printer is connected to LPT1: then you can obtain a hardcopy simply by pressing the key H when you see the plot on the screen that you wish to make a copy of.

Another method of getting hardcopies is to save the graph in a file and then print it later. This option is called the remote printer option. The remote printer option is useful for getting hard copies made on network where the network driver takes too much memory to run concurrently with MTS or if your printer is connected to a different PC from the one that MTS is installed on. When you see the graph on the screen, just press the key ‘S’. This causes a dump of the information on the screen to be made to a file. The name of this file is *n.GRF*, where  $n = 000, 001, 002, \dots$ . The first plot you save is in 000.GRF, the next in 001.GRF, etc. *Since these are binary files, be sure to use the DOS copy command rather than the DOS print command for sending your plots to the printer.* For example, to print all your graphics output files on a printer connected to the first printer port, enter

```
copy ??? .grf LPT1:
```

## *Current Restrictions & Limitations*

MTS Object	Maximum Length or Size
length of series	1000
length of series for spectral analysis	1024
seasonal period	52
number of transfer functions	12
number of interventions	12
number of data values in transfer-function modelling	2000
number of parameters estimated (in ARIMA and TFN nonlinear estimation)	15
degree of differencing	5
degree of AR,MA,SAR,SMA operators	24
degree of seasonal plus nonseasonal differencing operators	60
dimension of $\beta$	70
number of lags in autocovariance plots	60
forecasting lead time	200
autoregression order, spectral density	50
title string (number of characters)	80

## *Trouble Shooting*

1. *PROBLEM:* An error message indicating DYNAMIC MEMORY ERROR ALLOCATION is given.

*SOLUTION:* You should increase the amount of DOS memory available. A minimum of about 560K is needed. Remove any unnecessary terminate-and-stay resident utilities.

2. *PROBLEM:* An error message stating the output file cannot be opened is given.

*SOLUTION:* You should increase the amount of DOS memory available. A minimum of about 540K is needed. Remove any unnecessary terminate-and-stay resident utilities.

3. *PROBLEM:* Mouse is not operational.

*SOLUTION:* Make sure mouse driver is loaded prior to running MTS.

4. *PROBLEM:* Printer is not operational.

*SOLUTION:* It may be necessary to load the printer driver. For example, if the printer is connected to the first parallel port (this is the usual situation), then type the DOS command:

```
PRINT /D:LPT1
```

5. *PROBLEM:* Cannot print graphics output or graphics output is too small or too big.

*SOLUTION:* Run the program PSETUP.

6. *PROBLEM:* When output files in the OUTPUT directory are printed, some of the characters are not the same as when viewed on the computer screen.

*SOLUTION:* Make sure that your printer is configured to print the PC character set.

*EXPLORATORY ANALYSIS*

## Trace Plot

The trace plot for  $z_t$ ,  $t = 1, 2, \dots$  is simply a plot of  $z_t$  vs.  $t$ . Although it is very simple, it is very useful in practice. Many interesting features of the series are often visible such as outliers, variance changes or shifts, trends (deterministic or stochastic), interventions, seasonal effects, non-Gaussian or non-linear features.

Five styles of plots are available.

1. Join points with straight line and mark each point. By joining the points with lines, the chronological order is emphasised. If there are not too many values, marking each point is useful. This is the default.
2. Join points with straight line and do not mark points. This is useful when there is so much data that marking each value produces too much clutter.
3. Plot points only. This is only useful if the exact chronological order is not of interest and just the general relationship between  $z_t$  and  $t$  is then the focus.
4. High density line plot. A series of vertical lines from the abscissa to the ordinate is drawn. This provides an alternative to choices 1 and 2 above.
5. Points and lines. This is similar to the previous style number 1, i.e., join points with a straight line and mark each point. However in style number 5, the lines just point in the direction of the next point but do not actually connect the points. This style of time series plot is used in the computing language S.

Significance limits for  $z_t$  can be shown if appropriate. These limits are based on  $\bar{z} \pm 1.96s$ , where  $\bar{z}$  and  $s$  are the sample means and sample standard deviations. These limits assume that the data is normal and independent. This assumption is useful for residuals but perhaps less useful for most other types of data. Still these limits can serve to emphasize the correlation or stochastic trends that exists in the non-independent case. The choices for significance limits are

- (i) at the 5% level (which corresponds to 1.96 sd.)
- (ii) at the 0.27% level (which corresponds to 3 sd.)
- (iii) at 1, 2 and 3 sd.

As another option, you may choose to look at the Tukey 4253H, twice smoother superimposed on the plot.

Another option implements the Nelson pattern checks for nonrandomness. This technique was developed for examining time series plots (Shewart charts) in industrial quality control. The purpose of the Nelson pattern check is to detect an aberrant value in a series assumed to be in a state of statistical control. This assumption means that the series is essentially normally distributed with constant variance and successive observations are independent. The Nelson pattern checks are comprised of eight statistical tests as follows:

1. a value beyond 3 sd.,
2. 2 out of 3 points in a row beyond 2 sd on the same side of mean,
3. 4 out of 5 points in a row beyond 1 sd on the same side of mean,

4. 9 consecutive values on same side of mean,
5. 6 consecutive points all decreasing or increasing,
6. 14 points in a row alternating,
7. 15 points in a row less than 1 sd on either side of mean,
8. 8 points in a row beyond 1 sd on either side of mean.

Tests 1–3 reflect non-normality in the distribution function of the data; tests 4–6 lack of independence; tests 7–8 lack of constant variance (ie., heteroscedasticity). If this option is selected any points failing one of these tests is flagged by marking it red when plot styles 1 or 3 above is selected. In addition, an output file is produced and displayed listing exactly which points failed which test. If all points passed all tests, no output file is created or displayed.

Sometimes it may be desirable to fix the scale on the abscissa in the time series plot. For example, one may wish to compare to closely related series such as before and after an intervention or change. Complete control over the abscissa axis is provided if you select `NO` for the option to `Automatic Scale Selection For Abscissa`.

A final option allows one to conveniently plot a consecutive sub-series of the original series.

## *Bivariate Trace Plot*

Given two time series  $y_t$  and  $x_t$  or a bivariate series  $(y_t, x_t)$ , two trace plots are produced on the same screen. The first plot is a trace plot of  $y_t$  and the second is a trace plot  $x_t$ .

In each case, the ordinate axis is omitted to avoid clutter. It is just the general trends, shape and relationships that we seek to examine with this plot.

An important application of this plot is for identifying non-reversible time series, see for example Lawrance (1991). Such time series are either generated by nonlinear or nonGaussian or nonstationary models. For convenience, MTS allows one the option of reversing the first series.

Sometimes it also is of interest to plot the raw and its transformed values (after differencing and possible Box-Cox transformation) with the bivariate trace plot.

## *Bivariate Slug-Trace Plot*

This is a sort of bivariate time series plot which features an animated slow motion plot of a slug moving along the points  $(x_t, y_t), t = 1, \dots, n$  is shown. The time-ordering is shown by the animation effect of drawing lines between the points at a relatively slow speed. This plot is useful in detecting co-integrated series and other possible interrelationships.

This type of plot was suggested by Ramsey (1988) as preferable to the usual bivariate time series trace plot.

## *Time Series Scatter Plot*

The time series scatter plot shows a plot of  $z_t$  versus  $z_{t-\ell}$  for a selected value of  $\ell$ . This plot graphically shows the autocorrelation at lag  $\ell$ . The plot is useful in detecting outliers or values which unduly influence the correlation value. An instructive example is the series of residuals when an AR(3) is fitted to SERIESJX. If the lag 1 scatter plot is done of the squared-values, then it is clearly seen that the strong correlation effect is due to 3 outliers. In fact if these values are set to 0, no correlation effect remains. Squared values of the residuals are obtained by setting the parameter LAMBDA equal to 2.

## *Periodic Time Series Scatter Plot*

The index  $t$  for a periodic or seasonal time series may be written  $t = r(s - 1) + m$  where  $r = 1, 2, \dots, n$ ,  $m = 1, \dots, s$  and  $s$  is the period or number of seasons. Thus  $z_t$  may be written as  $z_{r,m}$ . For any  $\ell \geq 0$ ,  $t - \ell$ , can also be written in the same form for some values  $r'$  and  $m'$ . For brevity of notation we denote the observation at time  $t - \ell$  by  $z_{r',m'}$  where it is understood that if  $m - \ell \leq 0$  then this observation corresponds to  $z_{r',m'} = z_{t-\ell}$  for suitable  $r'$  and  $m'$ .

The periodic time series scatter plot is then a plot of  $z_{r,m}$  versus  $z_{r',m'}$  for some fixed  $\ell$ . This plot is useful for examining autocorrelations which depend on  $t$  only through the season, i.e. periodic autocorrelations. For more about periodic autocorrelations, see the later section on **Periodic Autocorrelations**.

## *Bivariate Time Series Scatter Plot*

A bivariate time series scatter plot shows a plot of  $y_t$  versus  $x_{t-\ell}$  for some fixed value of  $\ell$ . As in the other scatter plots, this function can be useful in examining the cross correlation structure of a bivariate series.

In addition, a Cleveland Robust Loess smooth can be also be fitted to the data. This function is intended to provide a general nonlinear regression modelling capability for regression with one explanatory variable.

## *Seasonal Subseries Plot*

The seasonal subseries plot shows on one plot the effect of seasonal variation and its evolution over time. For each seasonal period, the subseries corresponding to that period is plotted as vertical lines emanating from a horizontal line. The horizontal line is determined by the midmean of the seasonal period subseries. The midmeans, which are horizontal lines, clearly show the effects of seasonal any seasonal variation. The vertical lines, show its evolution over time. Cleveland and Terpenning (1982) originally suggested this plot for plotting the seasonal component of a time series as estimated by the classical decomposition of a time series into trend, seasonal and remainder. In addition, this plot is useful for examining a seasonal time series, or possibly, the differences or seasonal differences of an original series, to get a more precise idea about the seasonal effects. Another



application of this plot is to intervention analysis. By looking at separate seasonal sub-series plots before and after an intervention, we can get a better idea of what the effect of the intervention might have been. Be sure to use the same scale, for the ordinate axis on both plots.

## *Normal Probability Plots*

Given data  $X_1, X_2, \dots, X_n$  we can assess whether or not this data is normally distributed by using a normal probability plot. Normally probability plots are usefully in assessing not only batches of data but in diagnostic checking of many statistical models such as in ANOVA, regression and time series. The MTS package can be used to produce Normal Probability Plots. The normal probability plot as well as the associated statistical tests of Skewness, Michael and Wilk-Shapiro assume that the data is independent. These tests are not affected very much by small departures from this assumption but are completely vitiated for strong departures. For example, if the data follow a random walk, then it is quite meaningless to look at the normal plot or the associated tests.

### Empirical Quantiles

These are the ordered  $X$ -values:  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ .

### Plotting Positions

Each observed quantile  $X_{(i)}$  corresponds to the percentile:

$$p_i = \frac{i - \frac{1}{2}}{n}.$$

### Normal Quantiles

The normal quantiles  $q_1, q_2, \dots, q_n$  are given by

$$q_i = \Phi^{-1}(p_i), \quad i = 1, \dots, n,$$

where,  $\Phi^{-1}(p)$  denotes the inverse normal cumulative distribution function.

### Normal Probability Plot

This is a plot of the the empirical quantiles  $X_{(i)}$  versus expected normal quantiles  $q_i$ . That is we will put the empirical quantiles on the vertical axes and the normal quantiles on the horizontal axes.

### Interpretation of the Normal Probability Plot

If the data  $X_1, \dots, X_n$  are normally distributed with a mean of  $\mu$  and variance of  $\sigma^2$  then the plot will appear to lie on a straight line with slope  $\sigma$  and horizontal axes intercept  $\mu$ .

The normal probability plot sheds insights into the tail behaviour of the distribution. This tail behaviour can be visualized with the help of a straight line on the plotting passing through the lower and upper quantiles of the empirical and normal quantiles. Then tail, either right or left, of the observed distribution is thicker or thinner than the theoretical normal distribution according as the plot is outside or inside this line in the tail area. The technical terms for thin and thick tails are platykurtosis and leptokurtosis.

In addition to the normal probability plot, MTS produces a boxplot with a jittered-superimposed dotplot. The MTS package also calculates and tests three statistics which are useful in detecting non-normality. The statistics are the skewness coefficient, Michael's Statistic and the Wilk-Shapiro Test Statistic.

### Skewness Coefficient

$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right)^{\frac{3}{2}}}.$$

If  $g_1 > 0$  this means the data is skewed to the right or positively skewed. Similarly, a value  $g_1 < 0$  implies the data are skewed to the left. MTS calculates  $g_1$  and tests the null hypothesis that the data is normally distributed with a two-sided test. The skewness test has high power for alternative distributions possessing non-zero skewness. The significance level is calculated using the method of D'Agnostino (1970).

### Michael's Statistic

Michael's  $D_{SP}$  test statistic (Michael, 1983) enables significance limits to be drawn on the normal probability plot and is much more powerful than the usual standard Kolmogoroff-Smirnov approach. In fact, Michael's statistic is derived by applying a variance-stabilization transformation to the Kolmogoroff-Smirnov method. Let  $p_i = (i - \frac{1}{2})/n$ ,  $i = 1, \dots, n$ . Then  $D_{SP} = \max |g(f_i) - g(p_i)|$ , where  $g(x) = (2/\pi) \sin^{-1}(\sqrt{x})$  and  $f_i = \Phi((X_{(i)} - \bar{X})/\sqrt{v})$ , where  $v = \sum (X_i - \bar{X})^2/n$ . Royston (1993) provides an algorithm for determining the significance level of an observed value of  $D_{SP}$  under the null hypothesis that the data are independent normal with constant variance. Royston also discusses plotting significance limits on the normal probability plot. The plot produced by MTS shows the 0.5% significance limits. The value of  $D_{SP}$  statistic and its two-sided significance level are also displayed.

### Wilk-Shapiro Test

The Wilk-Shapiro statistic  $W$  measures the goodness-of-fit of the straight-line in the normal probability plot. Like  $R^2$ , the coefficient of determination in regression,  $W$  has the following properties:

$$0 \leq W \leq 1$$

$W$  close to 1, implies a good-fit

$W$  not close to 1, implies a poor-fit.

The MTS package, tests if the observed value of  $W$  is significantly smaller than that for normally distributed data. The Wilk-Shapiro test represents the most-powerful all-round test for normality. It is often very good even in small samples.

*TREND ANALYSIS*

## *Linear or Ratio Filter*

Two basic empirical models for nonseasonal time series are the additive trend model,  $z_t = T_t + I_t$ , and the multiplicative trend model,  $z_t = T_t I_t$ , where  $T_t$  represents the trend and  $I_t$  is the irregular or random component. In the multiplicative model,  $z_t$  and  $I_t$  are assumed to be greater than zero. These empirical time series models are still widely used today in seasonal adjustment methods as well as in trend analysis.

In either case trend component may be estimated by a symmetric double-sided moving average filter or data window,

$$\hat{T}_t = \sum_{i=-q}^q w_i z_{t+i},$$

where  $q$  is the half-length of the filter. By symmetry it is assumed that  $w_i = w_{-i}$ . Since it is an average, we have also

$$\sum_{i=-q}^q w_i = 1.$$

The symmetric moving average is applied to the middle part of the series corresponding to  $t = q + 1, \dots, n - q$ . The end-points are smoothed by a re-weighted asymmetric smoother. The starting values in the series  $t = 1, \dots, q$  are smoothed using

$$\hat{T}_t = \frac{\sum_{i=1-t}^q w_i z_{t+i}}{\sum_{i=1-t}^q w_i}.$$

The ending values,  $t = n - q + 1, \dots, n$  are handled in a similar fashion. A limitation of this treatment of end-effects, originally pointed out by Ozaki (1991), is that it sometimes tends to overstate any upward or downward trends in the data near either end of the series.

The *rectangular-data window* corresponds to choosing

$$w_i = \frac{1}{2q + 1}.$$

As shown in Hamming (1977), better results are usually obtained by using the *modified-rectangular-data window* is given by

$$\begin{aligned} w_i &= \frac{1}{2q}, & \text{if } |i| < q, \\ &= \frac{0.5}{2q}, & \text{if } |i| = q. \end{aligned}$$

By iterating the rectangular or the modified-rectangular windows, a variety of bell-shaped symmetry moving average filters can be obtained. For example, Spencer's 15-point formula, which is used in graduation, is equivalent to 3 iterations of rectangular windows with  $q = 4, 4$  and  $5$ . For proof, see Kendall and Stuart (1968, p.372). Note that by the linearity of the filtering operation, the order of application of the filters is immaterial.

In *linear* filtering, the irregular component is simply estimated by

$$\hat{I}_t = z_t - \hat{T}_t$$

whereas in ratio filtering the irregular component is given by

$$\hat{I}_t = \frac{z_t}{\hat{T}_t}.$$

Notice by that taking logarithms of the original data a multiplicative model may be converted into an additive model.

As an option, end-point smoothing can be performed to obtain estimates of  $\hat{T}_t, t = 1, \dots, q$  and  $\hat{T}_t, t = n - q + 1, \dots, n$ . In end-point smoothing,

$$\hat{T}_t = \frac{\sum w_i z_{t+i}}{\sum w_i},$$

where the summations are over those values of  $i$  in  $-q, \dots, 0, \dots, q$  such that  $t + i$  is in  $1, \dots, n$ . Ozaki (1991) has pointed out that end-point smoothing can sometimes exaggerate an apparent trend near the end of a series.

It is always best to show the smooth and the observed time series on the same plot in order to judge of the relative magnitude of the trend.

## *Tukey Smooth*

Tukey has suggested a variety of nonlinear robust smoothers for decomposing time series data  $z_1, \dots, z_n$  into smooth and rough components. Symbolically, we can write

$$\text{data} = \text{smooth} + \text{rough}$$

or

$$z_t = \tilde{z}_t + R_t,$$

where  $\tilde{z}_t$  denotes the smooth and  $R_t$  denotes the rough.

Trace plots of the smooth ( $\tilde{z}_t$  vs.  $t$ ) can reveal trends and changes in level of the series more clearly than plots of the raw data ( $z_t$  vs.  $t$ ).

Of equal importance, trace plots of the rough ( $R_t$  vs.  $t$ ) can reveal outliers, changes in variance or other unusual features.

The MTS package always produces trace plots of both the smooth and the rough for a time series.

It is always best to show the smooth and the observed time series on the same plot in order to judge of the relative magnitude of the trend.

Two different smoothers are available in MTS.

### 3RSSH, Twice

This smoother is recommended by Velleman and Hoalgin (1981) as a good smoother to use when doing the calculations by hand. Here we take medians of 3 and repeat this until convergence. Medians of 3 are used since they facilitate hand computation. Let  $z_t$  denote the original series. Then on the first pass we calculate  $y_t = \text{med}(z_{t-1}, z_t, z_{t+1})$  for  $t = 2, \dots, n-1$ . The endpoints are smoothed by  $y_1 = \text{med}(3y_2 - 2y_3, z_1, y_2)$  and  $y_n = \text{med}(3y_{n-1} - 2y_{n-2}, z_n, y_{n-1})$ . Next replace the  $z$ 's by the  $y$ 's and repeat again and again until there is no change in the resulting series. This yields 3R smoothing. The main problem with medians of 3 is that it tends to generate mesas, which are two consecutive points which form a local maximum or minimum. Mesas are not desirable in the smooth, so at mesas, end-point smoothing is used. This is called splitting (S). Then the 3R smooth is done again and any new mesas are split. This whole process is repeated two times. H refers to Hanning which is defined as taking weighted averages at time  $t-1, t, t+1$  with weights  $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ . This constitutes the first-pass smoother. Twice means that the final smooth is obtained by applying 3RSSH to the residuals obtained from the 3RSSH and adding this smooth to the first-pass smooth.

### 4253H, Twice

This is the preferred smoother when the computations are performed by computer. It is much too laborious to attempt by hand! This is the MTS default choice. As indicated by the name it involves taking medians of 4, then 2, then 5, then 3, then Hanning and then applying 4253H to the residuals of the first pass and adding this to the first pass smoother. In **4** smoothing and in **5** smoothing, lower order smoothing is used near the ends of the series and the end-points themselves are just "copied on". In **2** smoothing the end-points are also just copied on. The endpoints, in the final **3** smoother are obtained by end-point smoothing,  $y_1 = \text{med}(3z_2 - 2z_3, z_1, y_2)$  and  $y_n = \text{med}(3z_{n-1} - 2z_{n-2}, z_n, y_{n-1})$ . In this formula, notice that  $3z_2 - 2z_3$  corresponds to an estimate of  $z_0$  derived by extrapolating backwards the line joining the points  $(2, z_2)$  and  $(3, z_3)$ . Hanning is defined as taking weighted averages at time  $t-1, t, t+1$  with weights  $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ .

As an small illustrative example consider the hypothetical data sequence

$$5, 2, 4, 4, 0, 2, 3, 4$$

Applying **4** smoothing,

$$5, 3.5, 4, 3, 3, 2.5, 2.5, 3.5, 4$$

Applying **2** smoothing,

$$5, 3.75, 3.5, 3, 2.75, 2.5, 3, 4$$

Applying **5** smoothing,

5, 3.75, 3.5, 3, 3, 3, 3, 4

Applying **3** smoothing with end-point smoothing,

4.25, 3.75, 3.5, 3, 3, 3, 3, 3

Hanning,

4.25, 3.8125, 3.4375, 3.125, 3, 3, 3, 3

Taking residuals,

0.75, -1.8125, 0.5625, 0.875, -3, -1, 0, 1

Applying **4** smoothing,

0.75, -0.53125, 0.65625, -0.625, -0.21875, -0.5, -0.5, 0.5, 1

Applying **2** smoothing,

0.75, 0.0625, 0.015625, -0.421875, -0.359375, -0.5, 0, 1

Applying **5** smoothing,

0.75, 0.0625, 0.015625, -0.359375, -0.359375, -0.359375, 0, 1

Applying **3** smoothing with end-point smoothing,

0.15625, 0.0625, 0.015625, -0.359375, -0.359375, -0.359375, 0, 0.71875

Hanning,

0.15625, 0.07421875, -0.06640625, -0.265625, -0.359375, -0.26953125,  
0.08984375, 0.71875

Twicing, produces the final smooth,

4.40625, 3.88671875, 3.37109375, 2.859375, 2.640625, 2.73046875, 3.08984375, 3.71875

Sometimes it may be desirable to fix the scale on the abscissa in the time series plot. For example, one may wish to compare to closely related series such as before and after an intervention or change. Complete control over the abscissa axis is provided if you select **NO** for the option to **Automatic Scale Selection For Abscissa**.

As option, you may choose to look at a plot of the original data superimposed on the smooth.

## *Robust Loess Time Series Smoother*



Cleveland's (1979) robust loess regression smooth is applied to the time series data  $(z_t, t)$ . Smoothers with a smoothing parameter of around  $f = 0.8$  are useful in looking for long-term trends (McLeod et al., 1990). The residuals from the smooth can be saved to a file and examined or modelled.

Since more of the data is taken into account when  $f = 0.8$  than with the Tukey 4253H, twice smooth the plot always much smoother looking.

See Appendix 1 for a general discussion of Cleveland robust loess smoothing.

A limitation of this treatment of end-effects, originally pointed out by Ozaki (1991), is that it sometimes tends to overstate any upward or downward trends in the data near either end of the series.

It is always best to show the smooth and the observed time series on the same plot in order to judge of the relative magnitude of the trend.

## *Runs Test*

The Runs Test is a simple but often effective test of the null hypothesis that a time series is random. Let  $Q$  denote some fixed value such as one of the quartiles of the data series  $z_1, z_2, \dots, z_n$ . If we replace each  $z_t$  by a  $+$  or  $-$  according as  $z_t \leq Q$  or  $z_t > Q$  respectively, then a run is a string of consecutive  $+$  or  $-$ . The total number of runs, say  $R$ , yields a test statistic for randomness. The exact expected number of runs is given by

$$E(R) = 1 + \frac{2n_1n_2}{n_1 + n_2},$$

where  $n_1$  is the total number of  $+$  and  $n_2 = n - n_1$ . If there is persistence in the series, the observed number of runs,  $R$ , will tend to be less than the expected. On the other hand for alternating behaviour the number of runs will exceed  $E(R)$ . The exact variance of  $R$  is given by

$$\text{var}(R) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n - 1)}.$$

Provided that  $n_1$  and  $n_2$  are both greater than 20, the normal approximation can be used to compute the significance level (Swed and Eisenhart, 1943). A two-sided test is used. The Runs Test could be computed about a value other than  $Q$  but it would have less power.

When either  $n_1 \leq 20$  or  $n_2 \leq 20$  exact formulae given by Swed and Eisenhart (1943) for the probability function of  $R$  are used to compute the exact significance level of a two-sided test.

In general tests for lack of randomness based on the autocorrelation function are often the most powerful. In some other cases where a deterministic trend may be present, the Mann-Kendall test is usually more powerful (Hipel, McLeod & Fosu, 1986). If cyclical effects are present the cumulative periodogram test or Fisher test may be best. In certain other cases, the Runs Test may be helpful. The Runs Test makes no distributional assumptions other than independence so in this sense it is slightly more general than the other tests.

Granger (1963) conjectures the Runs Test may have more power than other tests for some forms of nonlinear dependence. An example of this is provided by Freedman's tent map, defined for  $t = 1, \dots, 50$  with initial value  $0 < z_0 < 1$ ,

$$\begin{aligned} z_t &= 2z_{t-1} \quad \text{if } z_{t-1} \leq 0.5, \\ &= 2 - 2z_{t-1}, \quad \text{if } z_{t-1} \geq 0.5. \end{aligned}$$

For the dataset FREEDMAN generated from this map, we found that while the time series trace and the time series scatter plots both showed clear non-random behaviour, the autocorrelation function did not. Taking  $Q$  to the the upper data quartile we obtained  $R = 26$  with  $E(R) = 20.2$  and a two-sided p-value of only 2.7%.

Lehmann (1959, pp.155–156) showed that the Runs Test is equivalent to a likelihood ratio test for a binary time series produced by a Markov chain.

## *Nonseasonal Trend Tests*

At present two tests for monotonic trend with nonseasonal time series are performed. The first is the Mann-Kendall test. This is a test for monotonic trend which was first suggested by Mann (1945). The test statistic is the Kendall rank correlation of  $z_t$  with  $t$ . The null hypothesis of no trend assumes that the  $z_t$  are independently distributed.

For simplicity, we will discuss the case where there are no ties in the values of  $z_t$ . The more general case, which is implemented in MTS is discussed in Ch. 4 of Kendall (1970). In the case of no ties, Kendall's rank correlation for a trend can be written

$$\tau = \frac{S}{\binom{n}{2}},$$

where

$$S = 2P - \binom{n}{2},$$

where  $P$  is the number of times that  $z_{t_2} > z_{t_1}$  for all  $t_1, t_2 = 1, \dots, n$  such that  $t_2 > t_1$ . Thus  $\tau = 2\pi_c - 1$ , where  $\pi_c$  is the relative frequency of positive concordance, i.e., the proportion of time for which  $z_{t_2} > z_{t_1}$  when  $t_2 > t_1$ . Equivalently, the relative frequency of positive concordance is given by  $\pi_c = 0.5(\tau + 1)$ . In the case where there are no ties in either ranking, it is known (Kendall, 1975, p.51) that under the null hypothesis, the distribution of  $S$  may be well approximated by a normal distribution with mean zero and variance,

$$\text{var}(S) = \frac{1}{18}n(n-1)(2n+5),$$

provided that  $n \geq 10$ . Valz and McLeod (1990) have given a simplified derivation of this formula for  $\text{var}(S)$  and Valz, McLeod and Thompson (1994) have examined the adequacy of the normal approximation in the case where there are possible ties.

In the MTS package, the Kendall rank correlation and its significance level are evaluated using the more general formula of Kendall (1970, Ch.4) which allows for ties in the rankings.

In the simple linear trend model  $z_t = \alpha + \beta t + e_t$ , where  $e_t$  is Gaussian white noise, it is known that the Mann-Kendall trend test has 98% efficiency relative to the usual least squares method of testing  $\beta = 0$ . An empirical simulation study of Hipel, McLeod and Fosu (1986) showed that the Mann-Kendall test outperformed the lag one autocorrelation test for detecting a variety of deterministic trends such as a step-intervention or a linear trend.

The second test is due to Abelson and Tukey (1964). The test statistic may be written

$$Z = \frac{\sum \alpha_t z_t}{\sqrt{c_0 \sum \alpha_t^2}},$$

where  $c_0$  is the sample variance of  $z_t$  and

$$\alpha_t = \sqrt{t(1 - \frac{t}{n})} - \sqrt{(t+1)(1 - \frac{t+1}{n})},$$

where  $n$  is the length of the series. Under the null hypothesis of no trend, the statistic  $Z$  is asymptotically normally distributed with mean 0 and variance 1. Large values of  $|Z|$  indicate the null hypothesis is untenable and hence there is the possibility of a trend in the series. The trend is increasing or decreasing according as  $Z$  is  $> 0$  or  $< 0$ . The function  $\alpha_t$  contrasts the values between each end of the series so values near the beginning are given weight close to  $-1$  while those near the other end are given a weight close to  $+1$ .

Both the Abelson-Tukey and the Mann-Kendall trend tests assume that under the null hypothesis, the data  $z_t$  are iid. When this is not the case, the true significance level is overstated. Thus the test may indicate the presence of a monotonic trend when in fact there is no trend but the data is autocorrelated. To overcome this, one can either use the modification to the Abelson-Tukey test suggested by Brillinger (1989) and available as a separate trend test or one could use a suitable systematic sample of the original data. Many autocorrelated time series can be modelling as  $q$ -dependent processes, which means that observations separated by more than  $q$  lags are independent. Hence if we systematically sample observations from the original series separated by  $q + 1$  lags, we obtain a series which is iid under the null hypothesis for no monotonic trend. Note that the `command editor` in the `utility` menu has the capability of constructing a systematic sample for the data. However, for convenience, we have including the systematic sampling option in the menu for this test. The parameters for systematic sampling are the offset parameter, denoted by  $h$ , and the period, denoted by  $k$ . For these parameters, the systematic sample is the  $z_{h+k(r-1)}$ , where  $r = 1, 2, \dots$ .

A nice example is provided by the dataset GREATLP which consists of mean annual precipitation for the Great Lakes, 1900–1986. Both the Mann-Kendall and Abelson-Tukey tests reject the null hypothesis of no trend.

## *Seasonal Trend Test*

This is a test for monotonic trend in a time series with seasonal variation. Hirsch, Slack and Smith (1982) developed such a test by computing the Kendall score separately for each month. The separate monthly scores are then summed to obtain the test statistic. The variance of the test statistic is obtained by summing the variances of the Kendall score statistic for each month. In this test, the null hypothesis is that the time series is of the form  $z_t = \mu_m + e_t$  where  $e_t$  is white noise error and  $\mu_m$  represents the mean for period  $m$ .

## *Brillinger Trend Test*

This test is developed in the article “Consistent detection of a monotonic trend superimposed on a stationary time series” published in *Biometrika* (1989). The important advantage of this test over the Mann-Wald test is its validity in the presence of autocorrelation.

The basic underlying model considered can be written

$$z_t = s_t + \eta_t,$$

where  $z_t$  is the observed time series,  $s_t$  represents a signal or trend component and  $\eta_t$  represents an autocorrelated error component. Under the null hypothesis it is assumed that  $s_t$  is a constant. The alternative hypothesis to be tested assumes that  $s_t$  is either a nondecreasing ( $s_t \leq s_{t+1}$ ) or nonincreasing ( $s_t \geq s_{t+1}$ ) function of time  $t$ .

The test statistic Brillinger develops may be written

$$Z_B = \frac{\sum c_t z_t}{\text{est.sd.}(\sum c_t z_t)},$$

where

$$c_t = \sqrt{t\left(1 - \frac{t}{n}\right)} - \sqrt{(t+1)\left(1 - \frac{t+1}{n}\right)},$$

where  $n$  is the length of the series. Under the null hypothesis of no trend, the statistic  $Z_B$  is asymptotically normally distributed with mean 0 and variance 1. Large values of  $|Z_B|$  indicate the null hypothesis is untenable and hence there is the possibility of a trend in the series. The trend is increasing or decreasing according as  $Z_B$  is  $> 0$  or  $< 0$ . The function  $c_t$  contrasts the values between each end of the series so values near the beginning are given weight close to  $-1$  while those near the other end are given a weight close to  $+1$ . It can be shown that

$$\text{var}\left(\sum c_t z_t\right) = 2\pi f_\eta(0) \sum c_t^2,$$

where  $f_\eta(0)$  denotes the spectral density function of the autocorrelated error component evaluated at 0.

In order to estimate  $f_\eta(0)$  it is first necessary to estimate  $\eta$ . Assuming there are no outliers, an estimate of the trend component is given by the running average of order  $V$ :

$$\hat{s}_t = \sum_{i=-V}^V \frac{1}{2V+1} z_{t+i}.$$

The practitioner should normally choose a value of  $V$  to give a reasonable estimate of the trend component. To assist in verifying the choice of  $V$  the MTS Package produces a plot of the trend component for the user-specified choice of  $V$ . In some cases where there are outliers in the series, a suitable Box-Cox transformation may be used to make the data more normally distributed. The normal probability plot or range-mean analysis may be used to choose the transformation.

After the trend component,  $s_t$ , has been estimated the autocorrelated error component,  $\eta_t$ , can be estimated by  $\hat{\eta}_t = z_t - \hat{s}_t$ . Then an estimate of  $f_\eta(0)$  is given by

$$\hat{f}_\eta(0) = \frac{\sum_{j=1}^L \frac{1}{2\pi n} |\hat{\epsilon}_j|^2}{\sum_{j=1}^L (1 - a_j)^2},$$

where

$$\hat{\epsilon}_j = \sum_{t=V+1}^{n-1-V} \hat{\eta}_t \exp\left\{\frac{-2\pi i t j}{n}\right\},$$

where  $i = \sqrt{-1}$  and

$$a_j = \frac{\sin\left\{\frac{2\pi j(2V+1)}{2n}\right\}}{(2V+1) \sin\left(\frac{2\pi j}{2n}\right)}.$$

The parameter  $L$  determines the degree of smoothing of the periodogram component. To assist in the choice of  $L$  the MTS Package plots the periodogram of the estimated autocorrelated error component  $\hat{\eta}_t$  at the low frequencies and shows the bandwidth corresponding to  $L$ .

Finally

$$\text{est.sd.}(\sum c_t z_t) = \sqrt{\hat{2}\pi \hat{f}_\eta(0) \sum c_t^2}.$$

In practice the Fourier transform  $\hat{\epsilon}_j$  may either be computed using the Discrete Fourier Transform (DFT) or the Fast Fourier Transform (FFT). If the FFT is used, the series is padded with zeros at both ends until it is of length  $n' = 2^p$ , where  $p = [\log_2(n)] + 1$ , where  $[\bullet]$  denotes the integer part. In this case, to avoid leakage it is recommended that data tapering be used. This involves multiplying the series  $\hat{\eta}_t$  by the

cosine tapering function  $u_t$ , where

$$\begin{aligned} u_t &= \frac{1}{2} \left( 1 - \cos \frac{\pi(t - \frac{1}{2})}{\ell} \right), \quad \text{for } t = 1, \dots, \ell, \\ &= 1, \quad \text{for } t = \ell + 1, \dots, n' - \ell - 1, \\ &= \frac{1}{2} \left( 1 - \cos \frac{\pi(n' - t + \frac{1}{2})}{\ell} \right), \quad \text{for } t = n' - \ell, \dots, n' \end{aligned}$$

to form the tapered series  $\hat{\eta}'_t = \hat{\eta}_t u_t$ . The Fourier transform for the tapered series is then evaluated. The percentage of data tapered, say  $\Gamma$ , is then  $\Gamma = \frac{200\ell}{n'}$ . Tukey recommends choosing  $\Gamma = 10$  or  $20$ . Tapering is normally not needed when the DFT method is used. In the MTS Package the user may chose between the DFT or FFT methods.

## *Seasonal-Trend Loess Decomposition*

The seasonal adjustment method, known as seasonal-trend loess or STL given by Cleveland et al. (1991) is implemented. This is a state of the art seasonal adjustment technique which has several advantages over other methods. It is conceptually much simpler than the traditional X-11 method and yet gives results of equal reliability. Moreover STL can be performed for series with any seasonal period not just quarterly or monthly as is the case with X-11. Cleveland et al. (1991) have made the basic Fortran code available by anonymous FTP so it is more accessible than X-11. With suitable extensions such as allowing for trading day and holiday effects and extending the series at both ends using a fitted SARIMA model this technique is expected to perform as well as X-11-ARIMA.

The seasonal decomposition of an observed time series  $z_t$ ,  $t = 1, \dots, N$  can be written

$$z_t^{(\lambda)} = T_t + S_t + R_t,$$

where  $z_t^{(\lambda)}$  denotes the Box-Cox transformation,

$$\begin{aligned} z_t^{(\lambda)} &= \frac{(z_t + c)^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0, \end{aligned}$$

and  $T_t, S_t$  and  $R_t$  are respectively the trend, seasonal and remainder.

Following the notation of Cleveland et al. (1991) we will denote the cycle period by  $n_p$ . For seasonal time series, the cycle period is just the same as the seasonal period which we have previously denoted by  $s$ . The cycle period is a more general concept which applies for example to diurnal data such as that which arises in the study of circadian rhythms.

The primary goal of seasonal adjustment and the basic STL algorithm is to estimate the seasonal component,  $S_t$ , and the seasonally adjusted series. The seasonally adjusted series in the transformed domain is given by

$$z^{(\lambda)} - S_t$$

and in the original data domain by

$$(z^{(\lambda)} - S_t)^{[\lambda]},$$

where  $x^{[\lambda]}$  denotes the inverse Box-Cox transformation,

$$\begin{aligned} x_t^{[\lambda]} &= (\lambda x + 1)^{\frac{1}{\lambda}} - c, \quad \text{if } \lambda \neq 0, \\ &= e^x - c, \quad \text{if } \lambda = 0. \end{aligned}$$

On the other hand if the interest is primarily on estimating a long-term trend, then it is better to first seasonally adjust and then apply loess smoothing to the seasonally adjusted series. See the section below on *Post-Trend Smoothing*.

The general STL algorithm consists of  $n_o$  iterations of an outer loop and  $n_i$  iterations of an inner loop.

## INNER LOOP

Each iteration,  $k = 0, 1, \dots, n_i$  of the inner loop consists of the following steps:

### Step 1. *Detrending*.

Let  $T^{(k)}$  denote the final estimate of trend from Step 6 in each iteration of the inner loop. Initially,  $T_t^{(0)}$  is set equal to the final trend estimate at the  $n_i$ -th iteration of inner loop in the previous iteration of the outer loop. For the first iteration of the outer loop, set  $T_t^{(0)} = 0$ . In Step 1 the detrended series

$$z_t^{(\lambda)} - T_t^{(k)},$$

is calculated.

### Step 2. *Cycle Subseries Smoothing*.

There are  $n_p$  cycle subseries corresponding to each of the separate seasonal periods or cycles. For example, with monthly data, the series of all January values is the first cycle subseries, February values are the next etc. Each cycle subseries is smoothed using a loess smoother with window size,  $n_s$ . The window size,  $n_s$ , indicates the total number data points used in the loess smoother rather than the fraction as was used in the **Cleveland Robust Loess Smoother**. The default setting is to use half the data available. The resulting smoothed cycle subseries are denoted by  $C_t^{(k+1)}$ .

Step 3. *Low Pass Filtering of Smoothed Cycle Subseries.*

The low pass filter consists of: a running average filter of length  $n_p$  followed by another running average filter of length  $n_p$  followed by loess smoothing with a window size of  $n_\ell$ . The output from the low-pass filtering is denoted by  $L_t^{(k+1)}$  and can be calculated for  $t = 1, \dots, N$  by extending smoothed cycle subseries,  $C_t^{(k+1)}$ , obtained in Step 2.

Step 4. *Detrending of Smoothed Cycle Subseries.*

$L_t^{(k+1)}$  is subtracted from  $C_t^{(k+1)}$  to form the estimate of the seasonal component,  $S_t^{(k+1)}$ ,

$$S_t^{(k+1)} = C_t^{(k+1)} - L_t^{(k+1)}.$$

Step 5. *Deseasonalizing.*

An estimate of the deseasonalized series

$$z_t^{(\lambda)} - S_t^{(k)},$$

is calculated.

Step 6. *Trend Smoothing.*

Loess smoothing with a window size,  $n_t$ , is applied to the deseasonalized series from Step 5 to obtain a new estimate of the trend,  $T_t^{(k+1)}$ .

## OUTER LOOP: ROBUSTNESS ITERATIONS

After  $n_i$  iterations of the inner loop have been performed, one iteration of the outer loop is performed. Firstly, in the outer loop, the remainder component

$$R_t = z_t^{(\lambda)} - T_t^{(n_i)} - S_t^{(n_i)},$$

is calculated. Next, robustness weights are calculated as in Cleveland's robust loess method, viz.

$$\delta_t = B\left(\frac{R_t}{6h}\right),$$

where  $B(z)$  is the bisquare function defined by

$$\begin{aligned} B(z) &= (1 - |z|^2)^2, & \text{for } |z| < 1, \\ &= 0, & \text{for } |z| \geq 1, \end{aligned}$$

and  $h$  denotes the median of  $|R_1|, |R_2|, \dots, |R_N|$ . Now  $n_i$  iterations of the inner loop are performed using the robustness weights just calculated in the loess smoothing in Steps 2



and 6. The outer loop is iterated a total of  $n_o$  times. At the conclusion of the outer loop final estimates of  $T_t, S_t$  and  $R_t$  are obtained.

## CHOICE OF SMOOTHING PARAMETERS

Cleveland et al. (1991) have provided a detailed mathematical analysis of the properties of STL which suggests good values for most of the required parameters.

Cleveland et al. (1991) showed that  $n_t$ , the smoothing window size for the trend smoothing in Step 6 should satisfy

$$1.5n_p \leq n_t \leq 2n_p$$

and that in most situations a reasonable choice of  $n_t$  is given by

$$n_t = \left[ \frac{1.5n_p}{1 - \frac{1.5}{n_s}} \right]_{\text{odd}},$$

where  $[\bullet]_{\text{odd}}$  denote the integer part rounded up to the nearest odd integer. The *Trend Diagnostic Plot*, described in the next section, can be used to check or improve upon the default choice.

Cleveland et al. (1991) showed that  $n_\ell$ , the smoothing window for loess in the low pass filter, Step 4, can be chosen as  $n_\ell = [n_p]_{\text{ODD}}$ .

The least automatic and most difficult parameter to select is  $n_s$  the size of the window for the loess filter for cycle subseries smoothing, Step 2. Cleveland et al. (1991) showed that  $n_s$  should be an odd integer greater than or equal to 7. Also it is clear that

$$n_s \leq \frac{N}{n_p}.$$

As a reasonable default choice, we recommend

$$n_s = \left[ \frac{N}{2n_p} \right]_{\text{ODD}}.$$

The *Seasonal Diagnostic Plot*, described in the next section, can be used to check or improve upon the default choice.

The jump parameters,  $n_s^{\text{jump}}, n_\ell^{\text{jump}}, n_t^{\text{jump}}$ , determine the spacing in the evaluation of the weighted regression calculations in loess. The default choices,  $n_s^{\text{jump}} = 1, n_\ell^{\text{jump}} = 1, n_t^{\text{jump}} = 1$ , correspond to evaluating the weighted regression at every data point. To save computer time, the following settings could be used,  $n_s^{\text{jump}} = \frac{n_s}{10}, n_\ell^{\text{jump}} = \frac{n_\ell}{10}, n_t^{\text{jump}} = \frac{n_t}{10}$ .

The default choice for the iteration parameters is  $n_i = 1$  and  $n_o = 2$ . If there are no outliers present then  $n_i = 2$  and  $n_o = 0$  could be used. For more radical outliers  $n_i = 1$  and  $n_o = 5$  or even  $n_i = 1$  and  $n_o = 10$  are recommended.

## DIAGNOSTIC PLOTS

To examine the basic effect of the seasonal adjustment we can examine trace plots of the components  $T_t$ ,  $S_t$  and  $R_t$  as well as trace plots of the data and the seasonally adjusted data in both the transformed domain, if  $\lambda \neq 1$ , and also in the original domain.

An additional very useful plot to show the effect of the seasonal adjustment is the *Cycle Subseries Plot* which was first presented in the article by Cleveland and Terpenning (1982). For each cycle, the cycle subseries of the seasonal component,  $S_t$ , is plotted as vertical lines emanating from a horizontal line. The horizontal line is determined by the midmean of the cycle subseries. The *Cycle Subseries Plot* exhibits the degree of seasonality present, the effectiveness or usefulness of the seasonal adjustment and any trends the seasonal component.

The *Trend Diagnostic Plot* is useful for checking on the adequacy of the choice of  $n_t$  the trend smoothing parameter used in Step 6. The *Trend Diagnostic Plot* is comprised of two plots. The first panel show the points  $T_t + R_t$  with the a solid curve obtained by connecting the points  $T_t$ . If the choice of  $n_t$  is valid there should be no systematic departure from the curve apparent. The second panel is a trace plot of the remainder  $R_t$ . Again there should be no systematic or wandering behaviour present. One should also check for possible outliers and increase the robustness iterations if needed.

The *Seasonal Diagnostic Plot* is useful for checking on the correctness of the choice of  $n_s$ . Actually, the choice of  $n_s$  is somewhat subjective and is determined by considering what part of the variation in the data belongs to the seasonal component. If the data analyst believes that the seasonal component is rapidly evolving then  $n_s$  will be quite small. In this case, some of the variation that might otherwise go into the remainder,  $R_t$ , will be put in the seasonal component,  $S_t$ . On the other hand if the data analyst believes that the seasonal component is quite stable then  $n_s$  will be chosen fairly large. The *Seasonal Diagnostic Plot* shows a separate plot for each cycle or seasonal period. Let  $t = t(r, m)$  where  $r$  denotes the year and  $m$  denotes the seasonal period. So  $m = 1, \dots, n_p$  and  $r = 1, \dots, \frac{N}{n_p}$ . Let  $\bar{s}_m$  denote the mean of the seasonal component,  $S_t$ , for the  $m$ -th period. Then for each period, we examine plots of the points  $S_t + R_t - \bar{s}_m$  with the curve obtained by joining straight lines through the points  $S_t - \bar{s}_m$ . Based on these plots the analyst can choose what part of the variation belongs in the seasonal component.

## POST-TREND SMOOTHING

If the primary object is trend estimation instead of estimation of the seasonal and the seasonally adjusted series then the estimate of the trend component given by STL should be further smoothed. This is called post-trend smoothing. To perform this simply take the seasonally adjusted output from the STL function and input this data to the Cleveland Robust Loess Function in the menu section on Exploratory Techniques.

*View File in STL Directory*

This utility is used to view or print a text file in the STL directory. Time series data files may be viewed with the MTS screen editor.

## *CORRELATION FUNCTIONS*

## *Sample Autocorrelation Function*

The sample autocorrelation at lag  $\ell$  of the observed time series  $z_1, z_2, \dots, z_N$  is defined by

$$r_\ell = \frac{\sum_{i=\ell+1}^N (z_i - \bar{z})(z_{i-\ell} - \bar{z})}{\sum_{i=1}^N (z_i - \bar{z})^2},$$

where

$$\bar{z} = \frac{1}{n} \sum_{t=1}^N z_t.$$

If no mean correction is made (for example, when the series represents residuals or differences) then

$$r_\ell = \frac{\sum_{i=\ell+1}^N z_i z_{i-\ell}}{\sum_{i=1}^N z_i^2}.$$

In practice, we often look at the autocorrelations of such transformed series. The plot of  $r_\ell$  vs.  $\ell$  is referred to as the sample autocorrelation function (SACF). If the data (or transformed data) is independent and identically distributed with constant variance then

$$r_\ell \sim \text{NID}(0, N^{-1}).$$

To test for whiteness then we can examine the plot of  $r_\ell$  versus  $\ell$  with benchmark standard limits at  $\pm 1.96/\sqrt{N}$ . Note that we can expect about 1 in 20 of the  $r$ 's to lie outside these limits even for white noise. Often we pay particular attention to the value of  $r_\ell$  at  $\ell = 1$  and at seasonal lags.

A wider simultaneous 95% limit is also shown. This limit has the property that 95% of the time all values are inside this limit assuming that the time series is independent and identically distributed white noise. See Appendix 3 *Simultaneous Significance Intervals* for the method of determining this interval.

An overall test across a number of lags is given by the Portmanteau Test. In this test we choose a number,  $M$ , typically  $M = 15$  or  $M = 30$ . Two modified Portmanteau Tests are available: first, the Ljung-Box Modified Portmanteau Statistic, given by

$$Q_M^{(1)} = N(N+1) \sum_{i=1}^M \frac{r_i^2}{N-i}$$

and second the Li-McLeod Modified Portmanteau Statistic, given by

$$Q_M^{(2)} = \sum_{i=1}^M r_i^2 + \frac{M(M+1)}{2N}.$$

Under the null hypothesis of independence both of these statistics are  $\chi^2$ -distributed on  $M$  degrees of freedom. These portmanteau statistics and their significance levels are tabulated for  $M = 1, 2, \dots$  on the output file.

If the data (or transformed data) is instead assumed to be such that all theoretical autocorrelations after lag  $Q$  are zero then Bartlett's large-lag formula gives

$$\text{est.sd.}(r_\ell) = \sqrt{\frac{1}{N} \left(1 + \sum_{i=1}^Q r_i^2\right)}.$$

Revised benchmark limits are then drawn for  $\ell = Q + 1, Q + 2, \dots$ . On the input screen,  $Q$ , is referred to as the large-lag in Bartlett's large-lag formula. This capability is useful in identifying moving-average models and determining if high-order autocorrelations are still large when allowance is made for autocorrelation in the data.

When  $Q$  is chosen  $> 0$ , the limits shown at lags  $1, \dots, Q$  are the maximum theoretically attainable values of the theoretical autocorrelations in a MA( $Q$ ) process (see, Davies et al., 1974). This also aids in moving-average identification. If the sample values are outside the limits, the model can not be represented as an MA( $Q$ ).

Two tests for ARCH effects are also available. One can either test using absolute values of the data or squared values.

## *Sample Partial Autocorrelation Function*

The sample partial correlation function (SPACF) at lag  $k$  is the correlation between  $z_t$  and  $z_{t-k}$ , when the intervening  $z_{t-1}, \dots, z_{t-k+1}$  are held constant. The SPACF at lag  $k$  is denoted by  $\phi_{k,k}$ . It can be shown that  $\phi_{k,k}$  can be estimated by fitting an AR( $k$ ),

$$z_t = \mu + \phi_{1,1}(z_{t-1} - \mu) + \dots + \phi_{k,k}(z_{t-k} - \mu) + a_t.$$

Thus by fitting successive AR( $k$ ) models for  $k = 1, 2, \dots$  the sequence of sample partial autocorrelations  $\hat{\phi}_{k,k}$ ,  $k = 1, 2, \dots$  is generated.

The default fitting algorithm used is the Burg algorithm. Alternatively, the Yule-Walker method can also be selected.

It can be shown that if an observed series of length  $N$  can be fitted by an AR( $p$ ) then  $\hat{\phi}_{k,k} \sim NID(0, N^{-1})$ , for  $k > p$ . The plot of  $\hat{\phi}_{k,k}$  vs.  $k$  shows benchmark confidence limits at  $\pm 1.96/\sqrt{N}$ .

A wider simultaneous 95% limit is also shown. This limit has the property that 95% of the time all values are inside this limit assuming that the true model is an AR( $p$ ). See Appendix 3 *Simultaneous Significance Intervals* for the method of determining this interval.

If file output tabulations are requested with this function, the partial autocorrelations  $\hat{\phi}_{k,k}$   $k = 1, 2, \dots$  is also tabulated along with other model selection criteria. The other model selection criteria tabulated include the AIC, BIC, and Residual Variance.

## *Sample Inverse Autocorrelations*

Consider then ARMA( $p,q$ ) model defined by

$$\phi(B)(z_t - \mu) = \theta(B)a_t,$$

where,

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\ a_t &\sim \text{NID}(0, \sigma_a^2).\end{aligned}$$

Then the theoretical inverse autocorrelation function (TIACF), denoted by  $\rho_i(\ell)$ ,  $\ell = 1, 2, \dots$ , is defined as the theoretical autocorrelation function (TACF) for the ARMA( $q, p$ ) defined by

$$\theta(B)(z_t - \mu) = \phi(B)a_t.$$

The first model can be referred to as the primal model and the second model is the corresponding dual. Notice, that the dual of the dual gives back the primal.

The sample inverse autocorrelation function (SIACF) can be estimated by fitting a high-order autoregression and then calculating the autocorrelations in the dual model. More specifically, we select a value of  $P$ , often  $P = 20$  or  $30$  and then fit an AR( $P$ ),

$$\pi(B)(z_t - \mu) = a_t,$$

where  $\pi(B) = 1 - \pi_1 B - \dots - \pi_P B^P$ . Then using the estimated values,  $\hat{\pi}_1, \dots, \hat{\pi}_P$ , the SIACF is given by

$$\begin{aligned}ri(\ell) &= \frac{-\hat{\pi}_\ell + \sum_{i=\ell+1}^P \hat{\pi}_i \hat{\pi}_{i-\ell}}{1 + \hat{\pi}_1^2 + \dots + \hat{\pi}_P^2} \quad \text{if } \ell \leq P \\ &= 0 \quad \text{if } \ell > P.\end{aligned}$$

The user may supply their own value of  $P$  or use the default value specified in the function.

The SIACF is a useful additional tool in ARMA model identification. As an example, consider the problem of identifying a model such as

$$(1 - \phi_1(B) - \phi_5(B))(z_t - \mu) = a_t.$$

In this case the TACF is rather complicated. However for the TIACF, there are nonzero values only at lags 1,4,5,6 and the largest values are at lags 1 and 5. Another example of a situation where the IACF most clearly identifies the model is for the ARMA(2,1). For the ARMA(1,2) we expect that the TACF will damp out as a damped exponential starting at lag 2 but that the values  $\rho_0 = 1$  and  $\rho_1$  will not necessarily follow this pattern. Hence, for the ARMA(2,1) the SIACF may be expected to shown a similar pattern. Similarly, the SIACF could be expected to be useful for identifying ARMA( $p,1$ ) models.

As pointed out by Chang & Dickey (1994), the SIACF is very useful in identifying over-differenced time series. In this situation, there is a moving-average root near

the unit circle, so the SIACF damps out very slowly just like the SACF would for data which needs differencing.

The large-sample distribution of the SIACF has been shown to be identical to that of the SACF from the dual model.

## *Sample Inverse Partial Autocorrelations*

Given the SIACF,  $\rho_i(\ell), \ell = 1, 2, \dots$ , the SIPACF  $\phi_{i\ell, \ell}, \ell = 1, 2, \dots$ , may be calculated using the same algorithm that obtains the SPACF from the SIACF. The SIACF is sometimes helpful in identifying MA(q) models. If the true model is MA(q), the SIACF should cut-off after lag q and this can readily be checked from the confidence limits shown on the plot. The SACF also cuts off after lag q, but to check the significance of higher-order lags it is necessary to look at the revised confidence limits using Bartlett's formula. McLeod (1984) derived the large-sample distribution of the SIACF and discussed its use in testing for MA(q) models.

## *Sample Cross Correlation Function*

Given two time series  $u_t, t = 1, \dots, n$  and  $v_t, t = 1, \dots, n$  the sample cross-correlation function (SCCF) at lag  $\ell$  is given by

$$c_{u,v}(\ell) = \frac{\sum_{t=1}^{n-\ell} (u_t - \bar{u})(v_{t+\ell} - \bar{v})}{\sqrt{\sum_{t=1}^n (u_t - \bar{x})^2 \sum_{t=1}^n (v_t - \bar{y})^2}},$$

where,

$$\bar{u} = \frac{1}{n} \sum_{t=1}^n u_t,$$

and

$$\bar{v} = \frac{1}{n} \sum_{t=1}^n v_t.$$

The SCCF is an estimate of the theoretical cross-correlation function which is defined as

$$\rho_{u,v}(\ell) = \frac{\text{cov}(u_t, v_{t+\ell})}{\sqrt{\text{var}(u_t)\text{var}(v_t)}}$$

Notice that the SCCF unlike the SACF is not symmetric, i.e. in general  $c_{u,v}(\ell) \neq c_{u,v}(-\ell)$ .

Also it should be observed that for  $\ell > 0$ ,  $c_{u,v}(\ell)$  measures the linear relationship between current values of the  $u$ -series and future values of the  $v$ -series. Similarly for  $\ell < 0$ ,  $c_{u,v}(\ell)$  measures the linear relationship between current values of the  $u$ -series and past values of the  $v$ -series.



On the computer output produced by the cross-correlation function, the  $u$ -series is called the first series and the  $v$ -series is called the second series.

The main distributional result for the SCCF states that if the  $u_t$  and  $v_t$  series are uncorrelated white noise sequences then  $c_{u,v}(\ell) \sim NID(0, n^{-1})$ . More precisely, it is required that  $u_t \sim IID(0, \sigma^2)$ ,  $v_t \sim IID(0, \sigma^2)$  and  $\rho_{u,v}(\ell) = 0$  for all  $\ell$ . For this reason the *SCCF* plot shows benchmark confidence limits at  $\pm 1.96\sqrt{n}$ . Often it is of interest to check for dependence at lags  $\ell = -1, 0, 1$  and other lags as well. The following overall portmanteau tests are calculated:

$$Q_M = \sum_{i=-M}^M c_{u,v}^2 + \frac{M(M+1)}{2n}.$$

Under the null hypothesis,  $\mathcal{H}_0 : \rho_{u,v}(\ell) = 0$ , for  $\ell = -M, -M+1, \dots, M-1, M$ , the test statistic  $Q_M$  is approximately  $\chi^2$ -distributed with degrees of freedom  $2M+1$ . This test is appropriate if one is primarily interested in testing for no relationship between the  $u$ -series and the  $v$ -series.

In other situations, more specific tests may be relevant. For example, if  $u_t$  represents mean riverflow per unit time and  $v_t$  represents the precipitation then it may be assumed that  $\rho_{u,v}(\ell) = 0$ , for  $\ell = 1, 2, \dots$  since riverflow does not cause future precipitation. In this situation our interest would focus on the null hypothesis,  $\mathcal{H}_0 : \rho_{u,v}(\ell) = 0$ , for  $\ell = -M, -M+1, \dots, -1$ . The relevant portmanteau test statistic is then

$$Q_M^- = \sum_{i=-M}^{-1} c_{u,v}^2 + \frac{M(M+1)}{2n}.$$

which is approximately  $\chi^2$ -distributed with degrees of freedom  $M$  under our null hypothesis.

To illustrate another application, suppose that  $u_t$  represents the residuals from a transfer-function noise time series model and  $v_t$  is a pre-whitened input series. Then to check that there is no feedback present (that is that the  $u_t$  series is not causally linked to future values of the input series  $v_t$ ), we would be interested in testing the null hypothesis the null hypothesis,  $\mathcal{H}_0 : \rho_{u,v}(\ell) = 0$ , for  $\ell = 1, \dots, M$ . The relevant portmanteau test statistic is then

$$Q_M^+ = \sum_{i=1}^M c_{u,v}^2 + \frac{M(M+1)}{2n}.$$

which is approximately  $\chi^2$ -distributed with degrees of freedom  $M$  under this null hypothesis.

These tests are performed if a tabulation is requested in addition to the SCCF plot. The value of  $M$  used is the same as the number of lags chosen for the plots.

## *Periodic Autocorrelation Function*

Let  $z_t$ ,  $t = 1, \dots, N$  be  $N$  consecutive observations of a seasonal time series with seasonal period  $s$ . Let  $n_m$  denote the number of years of data for period  $m$ .

The time index parameter,  $t$ , may be written  $t = t(r, m) = (r - 1)s + m$ , where  $r = 1, \dots, n_m$  and  $m = 1, \dots, s$ . Thus  $r$  and  $m$  denote the year and month respectively. If

$$\mu_m = E\{z_{t(r,m)}\} \tag{1.1}$$

and

$$\gamma_{\ell,m} = \text{COV}(z_{t(r,m)}, z_{t(r,m)-\ell})$$

exist and depend only on  $\ell$  and  $m$ ,  $z_t$  is said to be periodically correlated (Gladyšev, E.G., 1961). The case where  $\mu_m$  and  $\gamma_{\ell,m}$  do not depend on  $m$  reduces to an ordinary covariance stationary time series.

In general it is understood that  $m$  obeys modular arithmetic, for example,  $\mu_0 = \mu_s$ . The sample periodic autocorrelation function (PeACF) is given by

$$r_{\ell,m} = \frac{c_{\ell,m}}{\sqrt{\{c_{0,m}c_{0,m-\ell}\}}}$$

where

$$c_{\ell,m} = \frac{1}{n_m} \sum_r (z_{t(r,m)} - \hat{\mu}_m)(z_{t(r,m)-\ell} - \hat{\mu}_{m-\ell}),$$

where  $\hat{\mu}_m = \sum_r z_{t(r,m)}/n$ .

A schematic plot which shows the lag values  $\ell$  on the ordinate axis and each period on the abscissa is produced. This plot marks each correlation value with a small horizontal line. The vertical lines show benchmark limits of  $1.96 / \sqrt{n_m}$ . Under the assumption of independence of the  $z_t$  series,  $r_{\ell,m}$  is approximately normally distributed with mean zero and variance  $1/n_m$  and  $r_{\ell,m}$  are statistically independent for different values of  $\ell$  and  $m$ .

The presence of periodic correlation is indicated if the periodic correlations vary markedly between periods.

The user may optionally selected a tabulation of the numerical values of the sample PeACF and sample PePACF if desired.

## *Periodic Partial Autocorrelation Function*

After the presence of periodic correlation has been detected, a suitable Periodic Autoregression (PAR) model can be selected either by examining plots of the sample periodic partial autocorrelation (PePACF). The PePACF can be efficiently implemented using the methods developed by Sakai (1982).

Sakai (1982) extended the celebrated Durbin-Levinson recursion for autoregressive models to PAR models and derived the distribution of the sample PePACF. Let  $\hat{\rho}_{\bullet\ell,m}$  denote the sample PePACF for lag  $\ell$  and period  $m$ . Sakai showed that if the correct order is  $p_m$  for period  $m$ ,  $\text{Est.Sd.}(\hat{\rho}_{\bullet\ell,m}) = 1/\sqrt{n}$ ,  $\ell > p_m$ . The order  $p_m$  can be identified by finding the lowest lag for which the sample PePACF cuts off.

A schematic plot similar to that for the sample PeACF is displayed. A tabulation of the numerical values of the sample PeACF may be obtained from the tabulation that can be produced using the **Periodic Autocorrelation Function** menu item.

## *Theoretical Autocorrelation Function*

The theoretical autocorrelation function (TACF), denoted by  $\rho_\ell$  is defined by

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0},$$

where,  $\gamma_\ell = \text{cov}(z_t, z_{t-\ell})$ ,  $\ell = 1, 2, \dots$  for general multiplicative ARMA models defined by

$$\phi(B)\Phi(B^s)(z_t - \mu) = \theta(B)\Theta(B^s)a_t,$$

where,

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s},$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s},$$

and  $a_t$  is white noise. The algorithm of McLeod (1975) is used.

The TACF is of interest in the identification and interpretation of fitted models.

## *Theoretical Partial Autocorrelation Function*

The theoretical partial autocorrelation function (TPACF) is determined from the TACF sequence  $\rho_\ell, \ell = 1, 2, \dots$ , by the Pagano algorithm (1974).

## *Impulse Response or Inverted Form Coefficients*

The general form of the SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ ) $_s$  model equation for a data series,  $z_t$ ,  $t = 1, \dots, N$ , may be written,

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^{d_s} z_t = \vartheta_0 + \theta(B)\Theta(B^s)a_t,$$

where,  $\vartheta_0$  allows for a possible deterministic trend component. In many applications, there is no deterministic component so  $\vartheta_0 = 0$ . The regular autoregressive operator and regular moving-average components are:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

The seasonal period parameter is  $s$ . For example,  $s = 12$ , for monthly seasonal data. The seasonal autoregressive operator and seasonal moving-average components are:

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s}$$

and

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s}.$$

The differencing and seasonal differencing operations are:

$$\nabla = 1 - B,$$

and

$$\nabla_s = 1 - B^s.$$

In the above  $B$  refers to the backshift operator on  $t$ ,  $B^k z_t = z_{t-k}$ ,  $\forall k$ .

Neglecting the deterministic component, the SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ ) $_s$  model may be written in impulse response form as follows,

$$z_t = \psi(B)a_t,$$

where

$$\psi(B) = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)\nabla^d\nabla_s^{d_s}}.$$

A recursive algorithm to calculate the coefficients,  $\psi_1, \psi_2, \dots$  in

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots,$$

is easily obtained. The  $\psi_k$  are computed recursively using the equation

$$\phi(B)\nabla^d\psi_k = -\theta'_k,$$

where

$$\begin{aligned} \theta'_k &= \theta_k, \text{ for } k = 1, \dots, q \\ &= -1, \text{ for } k = 0, \\ &= 0, \text{ otherwise,} \end{aligned}$$

and  $B$  is the backshift operator on  $k$ . The extension to the SARIMA case is obvious.

Similar impulse response functions arise in transfer-function models.

A plot of  $\psi_k$  vs.  $k$  can be produced.

The theoretical coefficients in the inverted form of the ARIMA( $p, d, q$ ) or SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ ) $_s$  model can also be plotted. In the ARIMA( $p, d, q$ ) case the inverted form of the model is written

$$\pi(B)z_t = \mu(t) + a_t,$$

where  $\mu(t)$  is the deterministic part and

$$\begin{aligned}\pi(B) &= 1 - \sum_{k=1}^{\infty} \pi_k B^k \\ &= \frac{\nabla^d \phi(B)}{\theta(B)}.\end{aligned}$$

The  $\pi_k$  are computed recursively using the equation

$$\theta(B) \nabla^d \pi_k = \phi'_k,$$

where

$$\begin{aligned}\phi'_k &= \phi_k, \text{ for } k = 1, \dots, p \\ &= -1, \text{ for } k = 0, \\ &= 0, \text{ otherwise,} \\ \pi'_k &= \pi_k, \text{ for } k \geq 0, \\ &= -1, \text{ for } k = 0,\end{aligned}$$

and  $B$  is the backshift operator on  $k$ . The extension to the SARIMA case is obvious.

*SPECTRAL ANALYSIS*

## *Cumulative Periodogram Analysis*

Given a data sequence  $z_1, z_2, \dots, z_N$  the periodogram function,  $I(f_j), f_j = j/N, j = 1, \dots, N', N' = \lfloor N/2 \rfloor$ , may be defined by

$$I(f_j) = \frac{1}{n} \left| \sum_{t=1}^n z_t e^{-2\pi f_j(t-1)} \right|^2,$$

where  $|\bullet|$  denotes complex absolute value and  $i = \sqrt{-1}$ .

The units of  $f_j$  are cycles per unit time. The period corresponding to  $f_j$  is  $T_j = 1/f_j$ . Alternatively, some researchers prefer to work with angular frequencies given by  $\omega_j = 2\pi f_j$  which has units radians per unit time. In MTS the user may select which units  $f_j, T_j$  or  $\omega_j$  are used on the abscissa when the spectral function is plotted.

Basically,  $I(f_j)$  measures the strength of the relationship between the data sequence  $z_t$  and a sinusoid with frequency  $f_j$ . Notice that frequencies  $f_j$  satisfy  $0 < f_j \leq 0.5$ .

The cumulative periodogram is defined by

$$C(f_j) = \frac{\sum_{\ell=1}^j (I(f_\ell) + I(f_\ell))}{\sum_{t=1}^N (z_t - \bar{z})^2}$$

Note that,

$$\sum_{j=1}^{N'} (I(f_j) + I(-f_j)) = \sum_{t=1}^N (z_t - \bar{z})^2,$$

where

$$\bar{z} = \frac{1}{N} \sum_{t=1}^N z_t.$$

Thus if the data sequence  $z_t$  is white noise a plot of  $C(f_j)$  vs.  $f_j$  should lie on the straight line joining the origin  $(0, 0)$  with  $(1, 0.5)$ . The plot produced shows significance limits at 25%, 10%, 5% and 1%. The significance limits are determined using a method suggested by Bartlett (1955). The cumulative periodogram test is useful for detecting departures from whiteness due to periodicity. It should be noted that the cumulative periodogram is a consistent estimator for the spectral distribution function even though the periodogram itself is not a consistent estimator for the spectral density function.

### *Fisher Periodicity Test*

Given an observed time series,  $z_1, z_2, \dots, z_n$ , this is a test for the presence of a sinusoidal components of the form  $A \cos(2\pi f_j + B)$ , where  $f_j$  denotes the Fourier frequency

$f_j = \frac{j}{n}$ . The test assumes that the time series is only made up of such deterministic sinusoidal components. In this situation the Fisher-Whittle test detects which ones are statistically significant.

Perhaps, more often, it is more accurate to assume that the spectrum is continuous with some peaks rather than discrete. In this case, Bartlett's test for whiteness would be more suitable. Nevertheless, there are cases in practice where periodicities are more readily detected by the Fisher-Whittle-Priestley test.

For brevity of notation, we will call the Fourier frequency,  $f_j$ , the  $j$ -th harmonic. Then the Fourier frequencies of interest correspond to the harmonics  $1, \dots, m$ , where  $m = \lfloor \frac{(n-1)}{2} \rfloor$ , where  $\lfloor \bullet \rfloor$  denotes the integer part.

First suppose we are interested in testing for the presence of a single sinusoidal term. Then Fisher's test statistic,  $g$ , is the ratio of the maximum periodogram ordinate divided by the sum of all the periodogram ordinates for harmonics  $1, \dots, m$ . That is,

$$g = \frac{\max_j I(f_j)}{\sum_{t=1}^n (z_t - \bar{z})^2},$$

where  $\bar{z}$  denotes the sample mean and

$$I(f_j) = \frac{1}{n} \left| \sum_{t=1}^n z_t e^{-2\pi f_j (t-1)} \right|^2.$$

If  $z_1, \dots, z_n$  is assumed to be Gaussian white noise then Fisher (1929) showed that the exact probability of observing a value of  $g$  as large or larger is given by

$$m(1-g)^{m-1} - \frac{m(m-1)}{2}(1-2g)^{m-1} + \dots + (-1)^a \binom{m}{a} (1-ag)^{m-1},$$

where  $a = \lfloor g^{-1} \rfloor$ .

If a periodicity is detected, then as pointed out by P. Whittle, we can modify Fisher's  $g$ -statistic to test for a second periodicity. The modified  $g$  statistic is obtained by omitting the periodogram ordinate which was significant. In fact as suggested by Priestley (1981), in practice the periodogram ordinates surrounding the one which is significant should also be removed on account of possible leakage when the true periodicity differs from the detected Fourier frequency.

The above modification can be extended to detect multiple periodicities. In the MTS package up to ten distinct harmonics can be tested for in this way.

## *Periodogram Smoothing*

This is the so-called frequency domain approach because in this approach one works with the Fourier transform of the data sequence rather than the original data sequence.



It is the most natural approach to estimating  $p(f)$  although the other approach by Autoregressive Modelling will often produce a more accurate overall estimate in the sense of the integrated mean-square error criterion.

The *Discrete Fourier Transform* (DFT) of the sequence  $z_1, \dots, z_n$ , is defined as

$$Z_j = \sum_{t=1}^n z_t e^{-2\pi i t f_j}, \quad j = -[(n-1)/2], \dots, 0, \dots, [n/2],$$

where  $[\bullet]$  denotes the integer part and  $f_j = j/n$ . The frequencies  $f_j$  are referred to as the Fourier frequencies. The functions  $\cos(2\pi f_j t)$  and  $\sin(2\pi f_j t)$  are orthogonal with respect to the usual inner product when evaluated at the Fourier frequencies.

The periodogram is given by

$$I(f_j) = \frac{1}{n} |Z_j|^2.$$

It follows from the orthogonality mentioned above that

$$\sum_{j=-[(n-1)/2]}^{[n/2]} I(f_j) = \sum_{t=1}^n z_t^2.$$

Thus  $I(f_j)$  can be interpreted as analysis of variance of the data.  $I(f_j)$  shows the amount of variation due to each frequency component. Since the periodogram is symmetric about zero, in practice only positive frequencies need be considered.

The periodogram smoothing approach to the estimation of  $p(f)$  is based on the following two large-sample results:

$$\langle I(f_j) \rangle \approx p(f_j),$$

and

$$\begin{aligned} \text{cov}(I(f_j), I(f_k)) &\approx p^2(f_j), \quad \text{when } j = k, \\ &\approx 0, \quad \text{when } j \neq k. \end{aligned}$$

From the above two equations we see that although  $I(f_j)$  is an unbiased estimator of  $p(f_j)$ , it is not consistent. If  $p(f)$  is assumed to be a smooth function of  $f$ , an estimator with smaller mean-square error can be obtained by averaging values of the periodogram.

The periodogram smoother may be written

$$\hat{p}(f_j) = \sum_{i=-q}^{i=q} w_i I(f_{j+i}),$$

where  $q$  is the half-length of the smoother and the weights  $w_i$  satisfy the following conditions:

- (i)  $w_i \geq 0$ ,

- (ii)  $w_i = w_{-i}$ ,
- (iii)  $\sum_{i=-q}^q w_i = 1$ .

The *Daniell* or *Rectangular* window uses  $w_i = (2q + 1)^{-1}$ ,  $i = -q, \dots, 0, \dots, q$ . As shown by Hamming (1977, §5.8), the *Modified Rectangular*,

$$\begin{aligned} w_i &= 1/(2q), & \text{if } |i| < q, \\ &= 0.5/(2q), & \text{if } |i| = q, \end{aligned}$$

is generally even better. These smoothers can be iterated several times to produce smoothers which yield very smooth and generally more desirable estimates of the spectral density. These iterated smoothers correspond to a single non-iterated smoother whose weights,  $w_i$ , form a symmetric bell-shaped like function. It is sometimes of interest to plot these weights,  $w_i$ . Since the weights are symmetric, it is sufficient to plot the weights for  $i = 0, \dots, q$ . When the weights are plotted on the same graph as the spectral density, it is convenient to rescale the weights since it is just the shape and width of the weight function that we are primarily interested in.

The equivalent degrees of freedom, denoted by edf, is given by

$$\text{edf} = \frac{2}{\sum_{i=-q}^q w_i^2}.$$

For the Rectangular smoother  $\text{edf} = 4q + 2$ . An approximate 95% confidence interval for  $p(f_j)$  is

$$\left( \frac{\text{edf } \hat{p}(f_j)}{\chi_{0.975}^2(\text{edf})}, \frac{\text{edf } \hat{p}(f_j)}{\chi_{0.025}^2(\text{edf})} \right),$$

where  $\chi_{0.025}^2(\text{edf})$  and  $\chi_{0.975}^2(\text{edf})$  denote the 2.5% and the 97.5% points of the  $\chi^2$ -distribution on edf degrees of freedom. The percentage point for fractional degrees of freedom are obtained by linear interpolation. It is important to note that this is *not* a simultaneous confidence interval.

These confidence limits are useful as an additional diagnostic check for fitted ARMA models (McLeod and Hipel, 1995). For this diagnostic check, the spectral density of the fitted ARMA model is plotted with 95% limits determined by the nonparametric smoother. To perform this diagnostic check with the MTS package, simply fit the appropriate ARIMA model (in `Model Estimation` menus) and save it in the FITS directory. Next select `Periodogram Smoothing` in the `Spectral Analysis` menus. After choosing confidence limits, you will be prompted with `Display Fitted ARMA Spectral Density` for which you select `Yes`. Finally, you will then be prompted for the fitted model saved in the FITS directory.

The *Fast Fourier Transform* (FFT) may be used instead of the DFT to calculate the periodogram. In this approach the mean-corrected series  $\tilde{z}_t = z_t - \bar{z}$ , where  $\bar{z} = n^{-1} \sum z_t$  is calculated. Then this series is padded with 0's on both sides until the length of the series can be expressed as a power of 2. Then the FFT is calculated. The

FFT is algebraically equivalent to the DFT but it uses a numerical algorithm which requires much fewer calculations. In addition, the FFT is typically, more accurate numerically. Notice that the padding by 0's does not affect the sample autocorrelations in the time-domain however this padding can lead to distortions in the frequency-domain periodogram. The term given to the effect of these distortions is leakage. To overcome leakage, Tukey (1967) suggested tapering. Tapering has also been shown to be helpful when the DFT is used. With tapering it is possible to resolve individual peaks in the spectrum which would otherwise be masked by leakage. For a rigorous discussion on the effects of tapering, see Brillinger (1981, Chapters 4 and 5).

Tapering the mean-corrected series  $\tilde{z}_t$  involves multiplying  $\tilde{z}_t$  and  $u_t$ , where

$$\begin{aligned} u_t &= \frac{1}{2} \left( 1 - \cos \frac{\pi(t - \frac{1}{2})}{\ell} \right), \quad \text{for } t = 1, \dots, \ell, \\ &= 1, \quad \text{for } t = \ell + 1, \dots, n - \ell - 1, \\ &= \frac{1}{2} \left( 1 - \cos \frac{\pi(n - t + \frac{1}{2})}{\ell} \right), \quad \text{for } t = n - \ell, \dots, n \end{aligned}$$

to form the tapered series  $\tilde{z}'_t = \tilde{z}_t u_t$ . The percentage of data tapered, say  $\Gamma$ , is then  $\Gamma = 200m/n$ . Tukey recommends choosing  $\Gamma = 10$  or  $20$ . Next the FFT of the tapered series is obtained. As shown in Brillinger (1981, p.124) the periodogram should be adjusted by dividing the periodogram of the tapered data series by  $\sum_t u_t^2$ . In the case of the cosine bell taper, Bloomfield (1976, p.194-5) showed that  $\sum_t u_t^2 = 1 - 5\Gamma/8$ . In addition, the equivalent degrees of freedom should be adjusted (Bloomfield, 1975, p.195):

$$\text{edf} = \frac{U_4}{U_2^2} \frac{n'}{n} \frac{2}{\sum_{i=-q}^q w_i^2},$$

where  $U_2 = 1 - 5\Gamma/8$  and  $U_4 = 1 - 93\Gamma/128$ . See Hurvich (1988) for a recent discussion on choosing  $\Gamma$ , the percentage data tapered.

Then periodogram is then smoothed as in the DFT case.

When trends are present in the data, such as is the case with the Beveridge wheat price index, the series may be assumed to be non-stationary and so the spectrum of the series does not exist. In this case, it may be of interest to examine the first-difference of the data series if the first differences look stationary. This corresponds to analyzing the changes rather than the levels. This approach has been advocated by Blackman and Tukey (1959) and Box and Jenkins (1976).

Another traditional approach involves removing the trend with a symmetric two-sided moving average filter. See **Linear or Ratio Filter** for more discussion on the methods available. In this case a trend is estimated and the residuals from the trend are estimated. The spectral analysis is then carried out on the residuals.

Some researchers have noted that estimation by periodogram smoothing is the best approach when one is searching for small peaks. In this case, an additional useful option

is to plot the spectral density function over a smaller range than the standard (0, 0.5) Hertz interval.

An optional output file may be created in the OUTPUT directory which lists the estimates of the spectral density, frequencies and periods.

See Appendix 4 for an overview of spectral density estimation.

## *Autoregressive Spectral Density Estimation*

This is the time-domain approach. Here we estimate  $p(f)$  simply by fitting an autoregression,

$$\phi(B)(z_t - \mu) = a_t,$$

where  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  and  $a_t$  is white noise with variance  $\sigma_a^2$ . Then  $p(f)$  can be determined from the equation

$$p(f) = \frac{\sigma_a^2}{|\phi(e^{2\pi i f})|^2}.$$

Akaike (1969) first suggested this approach and most researchers now agree that this approach often produces the best estimate of the spectral density function. A recent paper by Mackisack and Poskitt (1979) suggests that this is a good method for detecting narrow peak frequencies in data.

Four methods of choosing the autoregression are available in this function. The default method is one proposed by Akaike (1979) and is referred to as the ‘‘AIC-Bayes’’ filter. With this method all autoregressions of orders  $p = 0, 1, \dots, K$ , where  $K$  is some upper limit are determined. The weighted average of all spectral functions forms the final estimate of the spectral function. Each spectral density in the average is weighted according to the quasi-likelihood or Bayes posterior given by

$$e^{-\frac{1}{2} \text{AIC}}.$$

The other methods involve choosing a fixed model order,  $p$ . Two popular criterion are available the AIC or BIC. When one of these criterion is selected, the best model from all autoregressions of orders  $p = 0, 1, \dots, K$ , where  $K$  denotes some maximum value whose default value is 20, is selected.

A final alternative is to select the model order manually oneself. This is done by specifying  $K$  the maximum order of autoregression. In this case, an autoregressive of order  $K$  is fit.

In all of the above autoregressive methods the default maximum order of the autoregression is set to 20 but this value can be increased by the using up to a preset maximum limit which is given in the section on Current Restrictions & Limitations.

An optional output file in the OUTPUT directory may also be created which tabulates the autoregressive parameters and the period and frequency of all peaks or troughs present in the autoregressive spectral density estimate. Troughs at frequencies of 0.0833

often arise with seasonally adjusted monthly data. Peaks correspond to cyclic behaviour and occur in many diverse series.

An optional output file in the FITS directory may also be created for use with other functions in the package such as `Periodogram Smoothing` or `Theoretical Autocorrelation Function of Fitted ARMA Model`.

An optional residual file in the DATA directory may also be created. This is useful for prewhitening a data series or in diagnostic checking the fitted AR model.

See Appendix 4 for an overview of spectral density estimation.

## *Theoretical Spectral Density Function*

Let  $z_t = \mu + \psi(B)a_t$ ,  $a_t \sim \text{NID}(0, \sigma_a^2)$  denote a multiplicative ARMA process with

$$\psi(B) = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)}.$$

Then the spectral density function may be defined as

$$f(\lambda) = \sigma_a^2 |\psi(e^{-2\pi i\lambda})|^2, \quad 0 \leq \lambda \leq 0.5,$$

where  $|\bullet|$  denotes complex absolute value and  $i = \sqrt{-1}$ .

The units of  $\lambda$  are cycles per unit time. The period corresponding to  $\lambda$  is  $t_\lambda = 1/\lambda$ . Alternatively, some researchers prefer to work with angular frequencies given by  $\omega = 2\pi\lambda$  which has units radians per unit time. In MTS the user may select which units  $\lambda$ ,  $t_\lambda$  or  $\omega$  are used on the abscissa when the spectral function is plotted.

This function is evaluated via the FFT and plotted. In general, estimating the a general linear process via an ARMA model and then evaluating the spectral density function often produces a useful estimate of the underlying spectral density function.

## *MODEL ESTIMATION*

## *ARIMA Estimation*

The general form of the SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ ) $_s$  model equation for a data series,  $z_t$ ,  $t = 1, \dots, N$ , may be written,

$$\phi(B)\Phi(B^s)(w_t - \mu_w) = \theta(B)\Theta(B^s)a_t,$$

where,

$$w_t = \nabla^d \nabla_s^{d_s} z_t^{(\lambda)},$$

and

$$\begin{aligned} z_t^{(\lambda)} &= \frac{(z + c)_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0. \end{aligned}$$

The regular autoregressive operator and regular moving-average components are:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

The seasonal period parameter is  $s$ . For example,  $s = 12$ , for monthly seasonal data. The seasonal autoregressive operator and seasonal moving-average components are:

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s}$$

and

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s}.$$

The sequence  $a_t$  is assumed to be informationless white noise with mean zero and variance  $\sigma_a^2$ . In practice, it is often the case that to a good approximation,

$$a_t \sim \text{NID}(0, \sigma_a^2).$$

The parameter  $\mu_w$  represents a deterministic component in the model. If  $d = 0$  and  $d_s = 0$  this deterministic component is simply the mean of  $z_t$ . Otherwise, if  $d \neq 0$  or  $d_s \neq 0$  and if  $\mu_w \neq 0$  then there is a deterministic polynomial trend in  $t$  of degree  $D = d + d_s s$  which is included in the model. The trend coefficient term is

$$\vartheta_0 = \phi(1)\Phi(1)\mu_w.$$

For convenience, we can think of the ARIMA parameters

$$\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_{p_s}, \Theta_1, \dots, \Theta_{q_s}$$

in terms of the  $\beta$ -vector defined by

$$\beta = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \\ \theta_1 \\ \vdots \\ \theta_q \\ \Phi_1 \\ \vdots \\ \Phi_{p_s} \\ \Theta_1 \\ \vdots \\ \Theta_{q_s} \end{pmatrix}$$

Note that the order of  $\beta$ -parameters is  $p, q, p_s, q_s$  which is the same as in the model order specification.

## INPUT SCREEN

The input screen gives the basic ARIMA parameters and their default values as follows:

$$p = 0, d_s = 0, q = 0, p_s = 0, d_s = 0, q_s = 0, s = 0, \lambda = 1.0, c = 0.0.$$

In addition, there are several more prompts for input, viz.,

**Estimate Optimal Box-Cox Transformation:** No

Select **Yes** if the optimal value of  $\lambda$  is to be determined. The method of Box and Cox (1964) which was modified for ARIMA models by McLeod (1974) is used. See also Hipel, McLeod and Lennox (1977).

**Constrain Some  $\beta$ -parameters to zero:** No

If the setting is changed to **Yes** then another input screen will appear later which prompts for the particular  $\beta$ -parameters to constrain to zero.

**Estimation Algorithms**

The following are available:

- **McLeod approximate maximum likelihood.** See McLeod (1977) and McLeod and Salas (1983) for details. This provides a good approximation to exact maximum likelihood and often uses significantly less computer time. This algorithm is the default.
- **Ansley Exact maximum likelihood.** See Ansley (1980) for details. In most cases the only difference is that more computer time is used.



- **Box-Jenkins unconditional sum of squares.** This is given in the textbook by Box and Jenkins (1975). It is provided for comparison but *it is no longer recommended.*
- **Box-Jenkins conditional sum of squares.** Again, this is for interested user's only. *Not recommended for general use.*

#### Filename For Output

The output file is of the form `fname.nnn`, where `fname` is the filename of the input time series data and `nnn` is a counter from 000 to 999. The counter starts at 000 and increases by 1 for each new model which is fitted. This output file will be placed in the MTS OUTPUT directory.

#### Output Residuals As A Time Series Data File: No

If **Yes** is selected, a data file containing  $\hat{a}_t$ ,  $t = 1, \dots, n$  will be created in the MTS DATA directory.

#### Output File For Input To The Forecasting Algorithm: No

If **Yes** is selected, a binary file containing all relevant information will be created in the MTS INPUT directory. This file can be used as the input to the forecasting, simulation and residual-autocorrelation functions.

#### Estimate Trading Day Effects: No

This option is only available for monthly time series, i.e. where  $s = 12$ . In this case, you will be prompted with a special menu. This menu, will determine whether or not you wish to estimate the trading day effects and the exact month and year in which the series started. Only dates on or after January 1900 can be used.

Bell & Hillmer (1984) have discussed the use of trading-day effects in ARIMA modelling.

The ARIMA model with trading day effects may be represented as

$$z_t = \sum_{i=1}^8 \text{td}_i m_t^{(i)} + \eta_t,$$

where  $\eta_t$  is an ARIMA model,  $\text{td}_i$  is the trading-day effect for the  $i$ -th variable and  $m_t^{(i)}$  is the  $i$ -th trading-day variable for the  $t$ -th observation. Define  $m_t^{(8)}$  as the number of days in the month corresponding to the  $t$ -th observation. Now in order to define remaining  $m_t^{(i)}$ , let us first define  $n_t^{(i)}$  for  $i = 1, \dots, 7$  to be the number of Mondays, Tuesdays, Wednesdays, Thursdays, Fridays, Saturdays and Sundays respectively which are in the month corresponding to the  $t$ -th observation. Notice that since

$$m_t^{(8)} = \sum_{i=1}^7 n_t^{(i)},$$

there are really only seven independent parameters. For convenience, we set  $m_t^{(i)} = n_t^{(i)} - n_t^{(7)}$ . So  $\text{td}_t^{(7)} = 0$ . Thus the trading day parameters  $\text{td}_i$ ,  $i = 1, \dots, 7$  show the

trading-day or economic effect of weekdays relative to a zero-effect for Sunday and  $td_8$  shows the economic effect of the number of days in the month. Notice that the variables  $m_t^{(j)}$ ,  $j = 1, \dots, 6$  take on values of zero and one only and that  $m_t^{(8)}$  takes on the values 28, 29, 30 and 31. The trading-day variables are tabulated on the printout.

The standard errors of the trading-day coefficients is not available in the current MTS version but the statistical significance of the effects can be judged by using the AIC, BIC or a likelihood ratio test.

The airline data series (Series G, from Box & Jenkins, 1976) shows pronounced and significant trading day effects. The original identified SARIMA model for the logged data is  $(0, 1, 1)(0, 1, 1)_{12}$ . Next, the model was re-fit to estimate the trading-day effects. The estimated parameters are shown in the table below.

ESTIMATED TRADING-DAY PARAMETERS FOR THE AIRLINE DATA.

Trading-Day Variable	Estimate
Monday	-0.0065
Tuesday	-0.0029
Wednesday	-0.0033
Thursday	-0.0021
Friday	0.0020
Saturday	0.0043
Sunday	0.0000
No. Days	0.0432

From the above table, it is clear that travellers prefer Saturdays most for travelling with Fridays and Sundays as their second and third choices. The number of days in the month as a positive effect.

A comparison with the models estimated with and without trading-day variables is shown in the table below.

COMPARISON OF FITTED MODELS.

method	AIC	BIC	$\sigma_a^2$	Sum of Squares
without trading-day variables	670.11	720.60	$1.3510^{-3}$	11935.456
with trading-day variables	657.12	728.40	$1.1210^{-3}$	9895.405

The AIC selects the trading-day variable model by a large margin. According to the AIC, the relative plausibility of the original model vs. the model with trading-day effects is  $\exp\{-\frac{1}{2}(670.11 - 657.12)\} = 0.0015$ . The AIC selection is confirmed by an hypothesis test. Using the sum-of-squares presented in the table above, one can calculate

the likelihood-ratio test to test the joint significance of the trading-day variables. This yields a  $\chi^2 = 24.5$  on 7 df which corresponds to a p-value of about 0.09%. As is often the case, the BIC criteria favours the most parsimonious model — even though as we have seen the trading-day effects are highly significant statistically. Using the BIC criteria, the relative-plausibility of the model with trading-day effects vs. the original model is  $\exp -\frac{1}{2}(728.40 - 720.60) = 0.0202$ .

## OUTPUT

### Sum of Squares

This is the modified sum of squares,

$$S_m(\hat{\beta}) = F \sum_{t=-\infty}^N \hat{a}_t^2,$$

where  $F$  is a factor which is determined from the estimated covariance determinant and/or the Box-Cox transformation. In principle,  $F$ , is determined so that the log likelihood is given by

$$\log L = -\frac{n}{2} \log \frac{S_m(\hat{\beta})}{n},$$

where  $n = N - D$ . The quantity  $S_m(\hat{\beta})$  is useful in constructing likelihood ratio tests.

### Residual Variance

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=-\infty}^N \hat{a}_t^2.$$

### AIC

The observed value of the Akaike Information Criterion is given by

$$\text{AIC} = -2\delta \log L + 2\kappa,$$

where  $\kappa$  is the number of parameters estimated and  $\delta = \frac{N}{n}$ . The  $\delta$  term is a fudge factor which is introduced to make the AIC's comparable for ARIMA models which contain different orders of differencing. The best model according to the AIC criterion, is the one with the smallest AIC value.

### BIC

The observed value of the Bayes Information Criterion is given by

$$\text{BIC} = -2\delta \log L + 2\kappa \log(n).$$

As with the AIC, the model with the smallest BIC is selected. The best model chosen with the BIC usually has fewer parameters than the AIC. For forecasting purposes, the BIC is recommended.

#### Deterministic Component

The estimated values of  $\mu_w$  and  $\theta_0$  are given along with their standard errors. If **Sample Mean Correction** was set to No, then the values are assumed to be  $\mu_w = \theta_0 = 0$  and no output is given here.

#### Estimated $\beta$ Parameters

$\hat{\beta}$  and  $\text{se}(\hat{\beta})$  are tabulated. The covariance matrix of  $\hat{\beta}$  is also given.

#### Residual Skewness

$$g_1 = \frac{\frac{1}{n} \sum \hat{a}_t^3}{\left(\frac{1}{n} \sum \hat{a}_t^2\right)^{\frac{3}{2}}}$$

So  $g_1$  is calculated and is tested for being significantly different from zero using the method of D'Agnostino, R.B. (1970). Sometimes the use a suitable Box-Cox transformation is useful in reducing the residual skewness.

#### Residual Kurtosis

$$g_2 = \frac{\frac{1}{n} \sum \hat{a}_t^4}{\left(\frac{1}{n} \sum \hat{a}_t^2\right)^2} - 3$$

So  $g_2$  is calculated and is tested for being significantly different from zero. The large-sample approximation  $g_2 \sim N(0, 24n^{-1})$  is used.

#### Tests For Heteroscedasticity

Two tests for changes in variance are performed. In the first test it is assumed that

$$\text{var}(a_t) \propto e^{\tau z_{t-1}(1)},$$

where  $z_{t-1}(1)$  denotes the one-step forecast of  $z_t$  at time  $t - 1$ . This test is useful for detecting changes in variance which depend on the level of the series. In the second test,

$$\text{var}(a_t) \propto e^{\tau t}.$$

So this test will pick up a trend in the variance over time. For each test,  $\hat{\tau}$ , is given along with its estimated standard error. If  $\hat{\tau}$  is significantly different from zero, then a variance change has likely occurred. These tests were developed by McLeod (1974) and are described further in Hipel, Lennox and McLeod (1977).

#### Residual and Squared-Residual Autocorrelations

The following are tabulated for  $l = 1, 2, \dots$ ,

$$\hat{r}_l = \frac{\sum \hat{a}_t \hat{a}_{t-l}}{\sum \hat{a}_t^2}$$

the standard error of  $\hat{r}_l$

$$Q_l = n \sum_{i=1}^l \hat{r}_i^2 + \frac{l(l+1)}{2n}$$

$$\hat{r}_{aa}(l) = \frac{\sum \hat{a}_t^2 \hat{a}_{t-l}^2}{\sum \hat{a}_t^4}$$

$$Q_l^{(aa)} = n \sum_{i=1}^l \hat{r}_{aa}(i)^2 + \frac{l(l+1)}{2n}$$

The standard errors of  $\hat{r}_l$ ,  $l = 1, 2, \dots$  are calculated from equation (15) of McLeod (1978) using estimated values of the  $\beta$ -parameters.

The modified portmanteau statistics,  $Q_l$  and  $Q_l^{(aa)}$  are approximately  $\chi^2$ -distributed on degrees of freedom  $l - k_\beta$  and  $l$  respectively, where  $k_\beta$  is the number of estimated  $\beta$ -parameters (normally,  $k_\beta = p + q + p_s + q_s$ , unless some  $\beta$ -parameters have been constrained to zero). This specific form of the modified portmanteau statistic was introduced by Li and McLeod (1981) and its performance compares favourably with other tests as shown in Kheoh and McLeod (1991).

The use of squared-residual autocorrelations has been discussed by McLeod and Li (1984) and other researchers as well. Basically, these tests are useful in detecting nonlinearities as well as observational outliers.

#### Residual and Squared-Residual Seasonal Autocorrelations

The values tabulated as QAS and QAAS are defined respectively as

$$Q_{l,s} = n \sum_{i=1}^l \hat{r}_{si}^2$$

and

$$Q_{l,s}^{(aa)} = n \sum_{i=1}^l \hat{r}_{aa}^2(si).$$

Assuming all  $\beta$ -parameters have been estimated and there is no seasonal residual autocorrelation then

$$Q_{l,s} \sim \chi_{(l-p_s-q_s)}^2$$

and

$$Q_{l,s}^{(aa)} \sim \chi_{(l)}^2.$$

These provide more sensitive tests of model inadequacy due to seasonality.

#### Test For Periodic Correlation At Lag 1

Assuming  $s > 1$  then we can write  $t = (r-1)s+m$ , where  $r = 1, 2, \dots, n'$  corresponds to year and  $m = 1, 2, \dots, s$  corresponds to the season (for example, month). Then

$$Q_1(\text{periodic}) = n' \sum_{m=1}^s r_m^2(1),$$

where,

$$r_m(1) = \frac{\sum \hat{a}_{r,m} a_{r,m-1}}{\sqrt{\sum \hat{a}_{r,m}^2 \sum \hat{a}_{r,m-1}^2}}.$$

Assuming model adequacy,

$$Q_1(\text{periodic}) \sim \chi_{(s)}^2.$$

## Transfer-Function Noise Estimation & Intervention Analysis

The transfer-function noise and intervention analysis model may be written

$$y_t^{(\lambda)} = \mu + \sum_{i=1}^k \frac{\omega_i(B)}{\delta_i(B)} B^{b_i} x'_{i,t} + \sum_{i=k+1}^{k+h} \frac{\omega_i(B)}{\delta_i(B)} B^{b_i} \xi_{i',t} + N_t$$

where

$$y_t^{(\lambda)} = \frac{(y+c)_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0,$$

$$= \log(y_t + c), \quad \text{if } \lambda = 0.$$

$k$  is the number of transfer functions

$h$  is the number of interventions

$i' = i - k$  denotes the  $i'$ -th intervention

$$\omega_i(B) = \omega_0^{(i)} - \omega_1^{(i)} B - \dots - \omega_{v_i}^{(i)} (B^{v_i})$$

$v_i$  is the degree of the  $\omega_i(B)$  operator

$$\delta_i(B) = 1 - \delta_1^{(i)} B - \dots - \delta_{u_i}^{(i)} (B^{u_i})$$

$u_i$  is the degree of the  $\delta_i(B)$  operator

$b_i \geq 0$  is the delay parameter

and  $N_t$  is the dynamic-disturbance noise term which is generated by the ARIMA model

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^{d_s}N_t = \theta(B)\Theta(B^s)a_t,$$

The  $i$ -th input series,  $x_{i,t}$ , may be used directly or else the transformed input series  $x'_{i,t}$ ,

$$\begin{aligned} x'_{i,t} &= x_{i,t}^\lambda - \bar{x} \quad \text{if } \lambda \neq 0, \\ &= \log(x_{i,t}) - \bar{x}, \quad \text{if } \lambda = 0. \end{aligned}$$

where  $\bar{x}$  denotes a possible mean correction. Usually a mean correction is applied to each input series, but this term can be omitted if necessary.

Several types of intervention terms, denoted by  $\xi_{i',t}$ , may be specified. Each intervention series is specified by a parameter  $T$  which indicates the observation number at which the intervention occurred. An additional parameter  $\pi$  may be used to specify a seasonal or periodic intervention effect. If  $\pi$  is 0 or 1, no periodic effect is assumed. The simplest intervention is a pulse intervention,  $\xi_t$  which is specified by

$$\begin{aligned} \xi_t &= 0, \quad t \neq T, \\ &= 1, \quad t = T. \end{aligned}$$

Pulse interventions are useful for dealing with missing values or outliers. Perhaps, the most common intervention type is a step intervention which is defined by

$$\begin{aligned} \xi_t &= 0, \quad \{t < T\} \vee \{t \not\equiv 0 \pmod{\pi}\}, \\ &= 1, \quad \text{otherwise.} \end{aligned}$$

The third type of intervention is a ramp intervention,

$$\begin{aligned} \xi_t &= 0, \quad \{t < T\} \vee \{t \not\equiv 0 \pmod{\pi}\}, \\ &= (t - T + 1), \quad \text{otherwise.} \end{aligned}$$

In all cases the exact intervention series used is tabulated on the output file.

Current limitations include  $h + k \leq 15$ ,  $N \leq 1000$  and  $Nk \leq 2000$ .

The regular autoregressive operator and regular moving-average components are:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

The seasonal period parameter is  $s$ . For example,  $s = 12$ , for monthly seasonal data. The seasonal autoregressive operator and seasonal moving-average components are:

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s}$$

and

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s}.$$

The sequence  $a_t$  is assumed to be informationless white noise with mean zero and variance  $\sigma_a^2$ . In practice, it is often the case that to a good approximation,

$$a_t \sim \text{NID}(0, \sigma_a^2).$$

The parameter  $\mu$  represents a deterministic component in the model. If  $d = 0$  and  $d_s = 0$  this deterministic component is simply the mean of  $z_t$ . Otherwise, if  $d \neq 0$  or  $d_s \neq 0$  and if  $\mu \neq 0$  then there is a deterministic polynomial trend in  $t$  of degree  $D = d + d_s s$  which is included in the model. The trend coefficient term is

$$\theta_0 = \phi(1)\Phi(1)\mu.$$

For convenience, we can think of the ARIMA parameters

$$\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_{p_s}, \Theta_1, \dots, \Theta_{q_s}$$

in terms of the  $\beta$ -vector defined by

$$\beta = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \\ \theta_1 \\ \vdots \\ \theta_q \\ \Phi_1 \\ \vdots \\ \Phi_{p_s} \\ \Theta_1 \\ \vdots \\ \Theta_{q_s} \end{pmatrix}$$

Note that the order of  $\beta$ -parameters is  $p, q, p_s, q_s$  which is the same as in the model order specification.

Similarly with each transfer-function component, the associated  $\beta$ -vector is defined by

$$\beta_i = \begin{pmatrix} \delta_1^{(i)} \\ \vdots \\ \delta_{v_i}^{(i)} \\ \omega_1^{(i)} \\ \vdots \\ \omega_{u_i}^{(i)} \end{pmatrix}$$



## INPUT SCREEN

The input screen gives the basic ARIMA parameters and their default values as follows:

$$p = 0, d_s = 0, q = 0, p_s = 0, d_s = 0, q_s = 0, s = 0, \lambda = 1.0, c = 0.0.$$

In addition, there are several more prompts for input, viz.,

**Constrain Some  $\beta$ -parameters to zero:** No

If the setting is changed to **Yes** then another input screen will appear later which prompts for the particular  $\beta$ -parameters to constrain to zero.

**Estimation Algorithms**

The following are available:

- **McLeod approximate maximum likelihood.** See McLeod (1977) and McLeod and Salas (1983) for details. This provides a good approximation to exact maximum likelihood and often uses significantly less computer time. This algorithm is the default.
- **Ansley Exact maximum likelihood.** See Ansley (1980) for details. In most cases the only difference is that more computer time is used.
- **Box-Jenkins unconditional sum of squares.** This is given in the textbook by Box and Jenkins (1975). It is provided for comparison but *it is no longer recommended*.
- **Box-Jenkins conditional sum of squares.** Again, this is for interested user's only. *Not recommended for general use.*

**Filename For Output**

The output file is of the form `fname.nnn`, where `fname` is the filename of the input time series data and `nnn` is a counter from 000 to 999. The counter starts at 000 and increases by 1 for each new model which is fitted. This output file will be placed in the MTS OUTPUT directory.

**Output Residuals As A Time Series Data File:** No

If **Yes** is selected, a data file containing  $\hat{a}_t, \quad t = 1, \dots, n$  will be created in the MTS DATA directory.

**Output File For Input To The Forecasting Algorithm:** No

If **Yes** is selected, a binary file containing all relevant information will be created in the MTS INPUT directory. This file can be used as the input to the forecasting, simulation and residual-autocorrelation functions.

## OUTPUT

**Sum of Squares**

This is the modified sum of squares,

$$S_m(\hat{\beta}) = F \sum_{t=-\infty}^N \hat{a}_t^2,$$

where  $F$  is a factor which is determined from the estimated covariance determinant and/or the Box-Cox transformation. In principle,  $F$ , is determined so that the log likelihood is given by

$$\log L = -\frac{n}{2} \log \frac{S_m(\hat{\beta})}{n},$$

where  $n = N - D$ . The quantity  $S_m(\hat{\beta})$  is useful in constructing likelihood ratio tests.

#### Residual Variance

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=-\infty}^N \hat{a}_t^2.$$

#### AIC

The observed value of the Akaike Information Criterion is given by

$$\text{AIC} = -2 \log L + 2\kappa,$$

where  $\kappa$  is the number of parameters estimated. The best model according to this criterion, is the one with the smallest AIC value.

#### BIC

The observed value of the Bayes Information Criterion is given by

$$\text{BIC} = -2 \log L + 2\kappa \log(n).$$

The best model chosen with the BIC usually has fewer parameters than the AIC. For forecasting purposes, the BIC is recommended.

#### Deterministic Component

The estimated values of  $\mu$  and  $\theta_0$  are given along with their standard errors. If **Sample Mean Correction** was set to **No**, then the values are assumed to be  $\mu = \theta_0 = 0$  and no output is given here.

#### Estimated $\beta$ Parameters

$\hat{\beta}$  and  $\text{se}(\hat{\beta})$  are tabulated. The covariance matrix of  $\hat{\beta}$  is also given.

#### Transfer-Function Parameters

The estimates of the transfer-function parameters and their standard deviations are given.

### Intervention Analysis Parameters

The estimates of the intervention analysis parameters and their standard deviations are given.

### Correlation Matrix of Transfer Function & Intervention Analysis Parameters

The order of the parameters in the correlation matrix is:  $\delta_1^{(i)}, \dots, \delta_{u_i}^{(i)}, \omega_0^{(i)}, \omega_1^{(i)}, \dots, \omega_{v_i}^{(i)}$  where  $i = 1, \dots, k + h$ . Note that the estimated parameters of the Transfer Function and Intervention Analysis,  $\hat{\tau}$ , is uncorrelated with the estimated parameters in the noise component,  $\hat{\beta}$ .

### Lagrange Multiplier Test

Let  $\tau = (\delta_1^{(1)}, \dots, \delta_{u_i}^{(1)}, \omega_0^{(1)}, \omega_1^{(1)}, \dots, \omega_{v_1}^{(1)}, \dots, \delta_1^{(k+h)}, \dots, \delta_{u_i}^{(k+h)}, \omega_0^{(k+h)}, \omega_1^{(k+h)}, \dots, \omega_{v_{k+h}}^{(k+h)})$ . We test the null hypothesis:  $\mathcal{H}_0 : \tau = 0$  vs the alternative hypothesis  $\tau \neq 0$ . The test statistic we use is  $X^2 = \hat{\tau}' \text{est. var}(\hat{\tau})^{-1} \hat{\tau}$  which under  $\mathcal{H}_0$  is  $\chi^2$  distributed on  $k + h$  df. This type of test is called a Lagrange Multiplier Test and the general properties of such tests are reviewed in the article by Hosking (1983). The Lagrange multiplier test is asymptotically equivalent to the likelihood ratio test.

### Residual Skewness

$$g_1 = \frac{\frac{1}{n} \sum \hat{a}_t^3}{\left(\frac{1}{n} \sum \hat{a}_t^2\right)^{\frac{3}{2}}}$$

So  $g_1$  is calculated and is tested for being significantly different from zero using the method of D'Agostino, R.B. (1970). Sometimes the use a suitable Box-Cox transformation is useful in reducing the residual skewness.

### Residual Kurtosis

$$g_2 = \frac{\frac{1}{n} \sum \hat{a}_t^4}{\left(\frac{1}{n} \sum \hat{a}_t^2\right)^2} - 3$$

So  $g_2$  is calculated and is tested for being significantly different from zero. The large-sample approximation  $g_2 \sim N(0, 24n^{-1})$  is used.

### Tests For Heteroscedasticity

Two tests for changes in variance are performed. In the first test it is assumed that

$$\text{var}(a_t) \propto e^{\tau z_t - 1(1)},$$

where  $z_{t-1}(1)$  denotes the one-step forecast of  $z_t$  at time  $t - 1$ . This test is useful for detecting changes in variance which depend on the level of the series. In the second test,

$$\text{var}(a_t) \propto e^{\tau t}.$$

So this test will pick up a trend in the variance over time. For each test,  $\hat{\tau}$ , is given along with its estimated standard error. If  $\hat{\tau}$  is significantly different from zero, then a variance change has likely occurred. These tests were developed by McLeod (1974) and are described further in Hipel, Lennox and McLeod (1977).

### Residual and Squared-Residual Autocorrelations

The following are tabulated for  $l = 1, 2, \dots$ ,

$$\hat{r}_l = \frac{\sum \hat{a}_t \hat{a}_{t-l}}{\sum \hat{a}_t^2}$$

the standard error of  $\hat{r}_l$

$$Q_l = n \sum_{i=1}^l \hat{r}_i^2 + \frac{l(l+1)}{2n}$$

$$\hat{r}_{aa}(l) = \frac{\sum \hat{a}_t^2 \hat{a}_{t-l}^2}{\sum \hat{a}_t^4}$$

$$Q_l^{(aa)} = n \sum_{i=1}^l \hat{r}_{aa}(i)^2 + \frac{l(l+1)}{2n}$$

The standard errors of  $\hat{r}_l$ ,  $l = 1, 2, \dots$  are calculated from equation (15) of McLeod (1978) using estimated values of the  $\beta$ -parameters.

The modified portmanteau statistics,  $Q_l$  and  $Q_l^{(aa)}$  are approximately  $\chi^2$ -distributed on degrees of freedom  $l - k_\beta$  and  $l$  respectively, where  $k_\beta$  is the number of estimated  $\beta$ -parameters (normally,  $k_\beta = p + q + p_s + q_s$ , unless some  $\beta$ -parameters have been constrained to zero). This specific form of the modified portmanteau statistic was introduced by Li and McLeod (1981) and its performance compares favourably with other tests as shown in Kheoh and McLeod (1991).

The use of squared-residual autocorrelations has been discussed by McLeod and Li (1984) and other researchers as well. Basically, these tests are useful in detecting nonlinearities as well as observational outliers.

### Residual and Squared-Residual Seasonal Autocorrelations

The values tabulated as QAS and QAAS are defined respectively as

$$Q_{l,s} = n \sum_{i=1}^l \hat{r}_{si}^2$$

and

$$Q_{l,s}^{(aa)} = n \sum_{i=1}^l \hat{r}_{aa}^2(si).$$

Assuming all  $\beta$ -parameters have been estimated and there is no seasonal residual autocorrelation then

$$Q_{l,s} \sim \chi_{(l-p_s-q_s)}^2$$

and

$$Q_{l,s}^{(aa)} \sim \chi_{(l)}^2.$$

These provide more sensitive tests of model inadequacy due to seasonality.

## *Periodic Autoregression Estimation*

Let  $z_t, t = 1, \dots, N$  denote a periodic time series with period  $s$ . Then the time index can be written  $t = (s-1)r + m$ , where  $m = 1, 2, \dots, s$  and  $r = 1, 2, \dots, n$ . For notational convenience, it is assumed in the description below that  $N = ns$ , although in the program itself this assumption is not required, i.e. the length of the time series does not have to be a multiple of the seasonal period. For seasonal monthly periodically correlated time series,  $s = 12$ ,  $m$  represents the month and  $r$  represents the year. Thus the  $t$ -th observation may be denoted by  $z_t$  or  $z_{r,m}$ . In general we define  $z_{r,m-\ell}$  to be  $z_{t-\ell}$ .

The periodic autoregression (PAR) of order  $(p_1, p_2, \dots, p_s)$  may be written

$$\phi_m(B)(z_t - \mu_m) = a_t,$$

where

$$\phi_m(B) = 1 - \phi_{m,1}B - \dots - \phi_{m,p_m}B^{p_m}.$$

The backshift operator  $B$  on  $t$  shifts back one time unit. For example,

$$\begin{aligned} Bz_t &= z_{t-1} = z_{r,m-1}, & \text{if } m > 2, \\ &= z_{r-1,s}, & \text{if } m = 1. \end{aligned}$$

More generally, models involving a Box-Cox transformation can be fitted. So the general form of the PAR model may be written

$$\phi_m(B)(z_t^\lambda - \mu_m) = a_t.$$

Parameter Estimates

The mean parameters,  $\mu_m$ ,  $m = 1, \dots, s$ , are estimated by the means

$$\hat{\mu}_m = \bar{z}_m = \frac{1}{n} \sum_{\ell=1}^n z_{r,m}.$$

The parameters,  $(\phi_{m,1}, \dots, \phi_{m,p_m})$ , are estimated from the equations,

$$\hat{\phi}_{m,1}r_{m-i,0} + \hat{\phi}_{m,2}r_{m-i,1} + \dots + \hat{\phi}_{m,p_m}r_{m-i,p_m} = r_{m-i,1}, \quad i = 1, \dots, p_m,$$

where  $r_{m,\ell}$  denotes the periodic autocorrelation (PRACF) for period  $m$  and lag  $\ell$  defined by,

$$r_{m,\ell} = \frac{c_{m,\ell}}{\sqrt{c_{m,0}c_{m-\ell,0}}},$$

where

$$c_{m,\ell} = \frac{1}{n} \sum_r (z_{r,m} - \hat{\mu}_m)(z_{r,m-\ell} - \hat{\mu}_{m-\ell}).$$

The residuals,  $\hat{a}_t$ ,  $t = 1, \dots, N$ , are then obtained from

$$\hat{a}_t = \hat{\phi}_m(B)(z_t - \hat{\mu}_m).$$

For diagnostic checking, the periodic residual autocorrelations,  $r_{m,\ell}^{(a)}$ , defined by

$$r_{m,\ell}^{(a)} = \frac{c_{m,\ell}^{(a)}}{\sqrt{c_{m,0}^{(a)}c_{m-\ell,0}^{(a)}}},$$

where

$$c_{m,\ell}^{(a)} = \frac{1}{n} \sum_r \hat{a}_{r,m} \hat{a}_{r,m-\ell}.$$

Note that the residual variances for each period are given by,

$$\sigma_a^2(m) = c_{m,0}^{(a)}$$

### Log Likelihood

The maximized log-likelihood function is may now be written as

$$\log \mathcal{L} = -\frac{n}{2} \sum_{i=1}^s \log(\sigma_a^2(m)).$$

### AIC and BIC

The model order  $(p_1, p_2, \dots, p_s)$  may be input by the user or chosen automatically using either the AIC or BIC criterion. The AIC and BIC values are given by

$$\text{AIC} = -2 \log \mathcal{L} + 2k,$$

and

$$\text{BIC} = -2 \log \mathcal{L} + 2k \log N,$$

where  $k$  is the number of parameters estimated. Thus

$$k = 2s + \sum_{i=0}^s p_i + I_{\lambda \neq 1},$$

where

$$\begin{aligned} I_{\lambda \neq 1} &= 1, \text{ if } \lambda \neq 1, \\ &= 0, \text{ if } \lambda = 1. \end{aligned}$$

The model order  $(p_1, p_2, \dots, p_s)$  which minimizes the value of the AIC or BIC criterion is determined.

### Model Order Selection

In practice, we have found that using the AIC criterion results in too many parameters. So if an automatic criterion is used, we recommend the use of the BIC criterion. In fact, Noakes, McLeod & Hipel (1985) found that the best approach was to choose the model order by examining the periodic partial autocorrelation (PRPACF) and determining the most parsimonious adequate model. In practice this is done by choosing the model order for the  $m$ -th period,  $p_m$ , by finding the value of the PRPACF after which the next value is not significantly different from 0. Then after the model order has been chosen, the portmanteau statistics and residual autocorrelations for the model are examined. In most cases,  $p_m = 0$  or 1.

### Residual Autocorrelation

The residual autocorrelations,  $ra$ , and their standard deviations,  $sd(ra)$ , are tabulated for lags  $1, \dots, p_m$ . Possible inadequacy is suggested if any of  $ra$  are much more than twice  $sd(ra)$  in magnitude.

### Squared Residual Autocorrelation

The squared-residual autocorrelations,  $raa$ , and their standard deviations,  $sd(raa)$ , are tabulated for lags  $1, \dots, p_m$ . Possible inadequacy is suggested if any of  $raa$  are much more than twice  $sd(raa)$  in magnitude.

### Portmanteau Tests

A portmanteau test for using the residual autocorrelations over lags  $1, \dots, M$  is calculated separately for each period.

*DIAGNOSTIC CHECKS*



## Unit Root Test

Unit root tests provide a test of the null hypothesis that the time series may require differencing. It should be noted that the null hypothesis being tested is that the time series is homogenous nonstationary. If this null hypothesis is not rejected, it does not mean that it is true or even that the best model is necessarily one involving differencing. For example, the stationary AR(1) model,  $z_t = 0.8z_{t-1} + a_t$  with a series length of  $n = 50$  was simulated 10 times and it was found that the null hypothesis of homogenous nonstationarity was only rejected at the 5% level 2 times. The Dickey-Fuller type tests have also been criticized for their lack of power against fractional ARMA models (Robinson, 1994). Whether to use differencing or not should also be based on nature of the time series data as well as on the trace plot and the sample autocorrelation function.

Two types of unit root tests, the Dickey-Fuller and the Said-Dickey tests are available.

In the Dickey-Fuller test (Dickey and Fuller, 1979) the underlying model of the time series  $z_t, t = 1, \dots, n$  is assumed to be a stochastic difference equation of the form  $z_t = \mu + \rho z_{t-1} + a_t$  where  $a_t$  is iid( $0, \sigma_a^2$ ),  $|\rho| \leq 1$  and  $z_0 = 0$ . This model includes the stationary AR(1) as well as the ARIMA(0,1,0) as special cases. This model may be written in a form like a simple linear regression as

$$y_t = \mu + \rho x_t + a_t,$$

where  $t = 1, \dots, n - 1$ .  $x_t = z_t$  and  $y_t = z_{t+1}$ . A consistent estimator for  $\rho$  in the stochastic difference equation model is given by the least squares regression estimator

$$\hat{\rho} = \frac{\sum_{t=2}^n (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=2}^n (x_t - \bar{x})^2},$$

where

$$\bar{x} = \frac{1}{n-1} \sum_{t=2}^n x_t$$

and

$$\bar{y} = \frac{1}{n-1} \sum_{t=2}^n y_t.$$

A  $t$ -like statistic for testing  $\mathcal{H}_0 : \rho = 1$  is given

$$\hat{\tau}_\mu = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\tau_\mu}},$$

where

$$\hat{\sigma}_{\tau_\mu} = \frac{\hat{\sigma}_a^2}{\sum_{t=2}^n (x_t - \bar{x})^2},$$

where

$$\hat{\sigma}_a^2 = \frac{1}{n-3} \sum_{t=2}^n \hat{a}_t^2,$$

where

$$\hat{a}_t^2 = y_t - \hat{\mu} - \hat{\rho}x_t,$$

where  $\hat{\mu} = \bar{y} - \hat{\rho}\bar{x}$ . The distribution of  $\hat{\tau}_\mu$  is quite complex and percentage points which were derived by simulation are given by Fuller (1976, p.373). In practice the alternative hypothesis is  $\mathcal{H}_a : \rho < 1$  so a one-sided test is used.

If  $\mathcal{H}_0$  is accepted we can test for an ARIMA(0,2,0) model by differencing the original series and applying the Dickey-Fuller test to the differenced series.

The Said-Dickey test (Said and Dickey, 1984) is more general than the Dickey-Fuller test. In the Said-Dickey test the underlying stochastic difference equation model for  $z_t, t = 1, \dots, n$  is

$$z_t = \mu + \rho z_{t-1} + N_t,$$

where  $N_t$  is assumed to follow a ARMA( $p, q$ ) model. Then the ARMA( $p, q$ ) model is approximated by a high order autoregression, AR( $p'$ ). As a default  $p' = 6$  is suggested but the user can choose a larger or smaller value if it is warranted. The parameters in the resulting model are then estimated by a multiple linear regression technique. Said and Dickey showed that the resulting  $t$ -like statistic for testing  $\mathcal{H}_0 : \rho = 1$ ,

$$\hat{\tau}_\mu = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\tau_\mu}},$$

has the same asymptotic distribution as for the Dickey-Fuller statistic. As before a one-sided significance test is used. As before the test may be used on the differenced series to test if  $d = 2$  is needed.

## *Residual Autocorrelation Analysis of Fitted ARIMA Model*

A plot of the residual autocorrelations and their 95% significance intervals is given using the technique derived by McLeod (1978). This plot is sometimes useful in checking model adequacy and choosing an improved model based on the lags or pattern of lags of the correlations.

## *Theoretical Autocorrelation Function Of Fitted ARMA Model*

Often it is of interest to compare the sample autocorrelation with that of a fitted ARMA model. If there is a large discrepancy between these, it likely indicates model inadequacy. For example, comparing a fitted AR(2) model of the annual sunspot series (file SUNSPT.1) with the SACF we see their apparent sinusoidal amplitudes are quite different. This inadequacy is backed up with careful use of other diagnostics.

## *Theoretical Partial Autocorrelation Function Of Fitted ARMA Model*

Just as a comparison between the fitted ACF and sample ACF is of interest in assessing the goodness-of-fit of a model, so is a comparison between the fitted PACF and sample PACF.

## *Theoretical Spectral Density Function Of Fitted ARMA Model*

This provides an alternative method for spectral estimation.

## *Theoretical Impulse-Response Function Of Fitted ARIMA Model*

This plot show  $\psi_k$  vs.  $k$ , where the  $\psi_k$  are defined by

$$\begin{aligned}\hat{\psi}(B) &= \frac{\hat{\theta}(B)\hat{\Theta}(B^s)}{\hat{\phi}(B)\hat{\Phi}(B^s)\nabla^d\nabla_s^{d_s}} \\ &= 1 + \hat{\psi}_1 B + \hat{\psi}_2 B^2 + \hat{\psi}_3 B^3 + \dots\end{aligned}$$

## *Theoretical Autocorrelation Function Of Fitted ARMA Model*

The theoretical autocorrelation function of the stationary ARMA( $p, q$ ) component of a fitted model is plotted.

The stationary component of a fitted ARIMA( $p, d, q$ ) model is defined as follows. For notational simplicity we shall only consider the ARIMA model since the extension to the SARIMA model is obvious. The general form of the ARIMA( $p, d, q$ ) model equation for a data series,  $z_t$ ,  $t = 1, \dots, N$ , may be written,

$$\phi(B)(w_t - \mu_w) = \theta(B)a_t, \quad (1)$$

where,

$$w_t = \nabla^d \nabla_s^{d_s} z_t^{(\lambda)},$$

where,  $z_t^{(\lambda)}$  indicates a Box-Cox transformation. Then eq. (1) defines the stationary component.

## *Theoretical Partial Autocorrelation Function Of Fitted ARMA Model*

The theoretical partial autocorrelation function of the stationary ARMA( $p, q$ ) component of a fitted model is plotted.

## *Theoretical Spectral Density Function Of Fitted ARMA Model*

The theoretical spectral density function of the stationary ARMA( $p, q$ ) component of a fitted model is plotted.

## *Theoretical Impulse Response Function Of Fitted ARMA Model*

The theoretical impulse response function of the ARIMA( $p, d, q$ ) or SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ )<sub>s</sub> model is plotted. In the ARIMA( $p, d, q$ ) case the impulse response form the model is written

$$z_t = \mu(t) + \psi(B)a_t,$$

where  $\mu(t)$  is the deterministic part and

$$\begin{aligned} \psi(B) &= \sum_{k=0}^{\infty} \psi_k B^k \\ &= \frac{\theta(B)}{\nabla^d \phi(B)}. \end{aligned}$$

The  $\psi_k$  are computed recursively using the equation

$$\phi(B)\nabla^d \psi_k = -\theta'_k,$$

where

$$\begin{aligned} \theta'_k &= \theta_k, \text{ for } k = 1, \dots, q \\ &= -1, \text{ for } k = 0, \\ &= 0, \text{ otherwise,} \end{aligned}$$

and  $B$  is the backshift operator on  $k$ . The extension to the SARIMA case is obvious.

## *Theoretical Inverted Form Coefficients Of Fitted ARMA Model*

The theoretical coefficients in the inverted form of the ARIMA( $p, d, q$ ) or SARIMA( $p, d, q$ )( $p_s, d_s, q_s$ )<sub>s</sub> model are plotted. In the ARIMA( $p, d, q$ ) case the inverted form of the model is written

$$\pi(B)z_t = \mu(t) + a_t,$$

where  $\mu(t)$  is the deterministic part and

$$\begin{aligned} \pi(B) &= 1 - \sum_{k=1}^{\infty} \pi_k B^k \\ &= \frac{\nabla^d \phi(B)}{\theta(B)}. \end{aligned}$$

The  $\pi_k$  are computed recursively using the equation

$$\theta(B)\nabla^d\pi_k = \phi'_k,$$

where

$$\begin{aligned}\phi'_k &= \phi_k, \text{ for } k = 1, \dots, p \\ &= -1, \text{ for } k = 0, \\ &= 0, \text{ otherwise,} \\ \pi'_k &= \pi_k, \text{ for } k \geq 0, \\ &= -1, \text{ for } k = 0,\end{aligned}$$

and  $B$  is the backshift operator on  $k$ . The extension to the SARIMA case is obvious.

## *Spread-Level Diagnostic Check*

Spread-level analysis enables one to examine the assumption of constant variance in a data series  $z_t, t = 1, 2, \dots, n$  and get a rough estimate of the best Box-Cox transformation,

$$\begin{aligned}z_t^{(\lambda)} &= \frac{z_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t), \quad \text{if } \lambda = 0,\end{aligned}$$

to stabilize the variance. Spread-level analysis is based on the theory of variance stabilizing transformations developed by Bartlett (1947).

Let  $\mu_t$  and  $\sigma_{z_t}$  denote the local mean and standard deviation of  $z_t$  at time  $t$ . If the standard deviation depends on the mean so that

$$\sigma_{z_t} \propto \mu_t^p,$$

then using a Taylor series approximation it can be shown that

$$\sigma_{z_t^{(\lambda)}} \propto \mu_t^{\lambda+p-1}.$$

In order to achieve constant variance, we should choose  $\lambda = 1 - p$ .

In the diagnostic plot, we plot the log of the absolute residual vs log of the absolute fitted value. A line is fitted using the robust bisquare method. Then one minus the slope of this line estimates  $\lambda$ . A 90% confidence interval is obtained by bootstrapping. using the  $BC_a$  method of Efron and Tibshirani (1993).

## *Likelihood Ratio Test*

In diagnostic checking, a model is sometimes overfit by introducing one or several extra parameters. If the original model is denoted by  $\omega$  and the model containing some

additional parameters is denoted by  $\Omega$ . Sometimes  $\Omega$  is called the **full model** and  $\omega$  is called the **reduced model**.

To test the adequacy of the reduced model  $\omega$  we can use a likelihood ratio test. The likelihood-ratio test statistic is given by

$$X = -2 \log \frac{L(\omega)}{L(\Omega)},$$

where  $L(\omega)$  and  $L(\Omega)$  denote the maximized value of the likelihoods under each model. It may be shown that

$$X = \left( \frac{S(\Omega)}{S(\omega)} \right)^{\frac{n}{2}},$$

where  $S(\Omega)$  and  $S(\omega)$  denote the residual sum-of-squares functions.  $S(\Omega)$  and  $S(\omega)$  are tabulated on the output of the **ARIMA Estimation and Transfer-Function-Noise and Intervention Analysis** functions.

Under the assumption of model adequacy of the  $\omega$  model,  $X \sim \chi_k^2$ , where  $k$  is the number of additional parameters in the enlarged model  $\Omega$ . Notice that

$$k = \dim(\Omega) - \dim(\omega),$$

where  $\dim(\Omega)$  and  $\dim(\omega)$  denote the number of parameters in each model.

For example, if  $k = 1$ , so one additional parameter is included then we can reject at the 1% level the original model is  $X > 6.635$ .

The input screen prompts for  $S(\Omega)$  and  $S(\omega)$  and the number of parameters in each model. The output shows  $X$  and its observed significance level.

## *View File in OUTPUT Directory*

This utility is used to view or print a text file in the OUTPUT directory. Time series data files may be viewed with the MTS screen editor.

## *FORECASTING*

## Forecast A Fitted ARIMA Model

The purpose of this function is to produce optimal forecasts for any ARIMA models which have been fitted by the function `ARIMA Estimation`. Consult the section of the reference manual entitled “ARIMA Estimation” for the definition of the ARIMA model and the basic notation. A plot of the forecasts, prediction interval and observed values is produced. In addition, the minimum-mean-square-error forecasts and their prediction intervals may also be tabulated. The basic notions of minimum-mean-square-error forecast and the prediction intervals are summarized in the remainder of this section.

The optimal forecast of  $z_{n+\ell}$  given data  $z_1, \dots, z_n$  and a fitted ARIMA model is denoted by  $z_n(\ell)$ , where  $n$  is said to be the forecast origin and  $\ell$  is the forecast lead time. The optimal forecast  $z_n(\ell)$  in the sense of minimum-mean-square-error is the conditional expectation and *when there is no Box-Cox transformation* then it may be obtained recursively by using

$$z_n(\ell) = \theta_0 + \varphi_1 z_n(\ell - 1) + \dots + \varphi_{p^*} z_n(\ell - p^*) + \theta_1^* \langle a_{t-1} \rangle_t + \dots + \theta_{q^*}^* \langle a_{t-q^*} \rangle_t,$$

where,

$$\varphi(B) = \nabla^d \nabla_s^{d_s} \phi(B) \Phi(B^s),$$

$$\theta^*(B) = \theta(B) \Theta(B),$$

$$\langle a_s \rangle_t = a_s, \quad \text{when } s \leq t,$$

$$= 0, \quad \text{when } s > t,$$

and

$$z_t(\ell - i) = z_{t+\ell-i} \quad \text{when } i \geq \ell.$$

The  $\varphi$ 's and  $\theta^*$ 's are referred to as the generalized autoregressive and moving-average coefficients.

Prediction intervals with coverage probability  $\alpha$  for the forecasts are given by

$$z_n(\ell) \mp \mathcal{P}\left(1 - \frac{\alpha}{2}\right) \sqrt{V(\ell)},$$

where  $V(\ell)$  is the variance of the forecast error and  $\mathcal{P}(\alpha)$  is the inverse normal cumulative distribution function. Some typical values are  $\alpha = 0.95, 0.9, 0.75, 0.5$  which correspond respectively to the values of  $\mathcal{P}\left(1 - \frac{\alpha}{2}\right) = 1.96, 1.64, 1.15, 0.674$ .

The error in the lead  $\ell$  forecast, denoted by,  $a_n(\ell)$  is given by

$$a_n(\ell) = z_{n+\ell} - z_n(\ell)$$

$$= \sum_{i=0}^{\ell-1} \psi_i a_{n+\ell-i}.$$



Thus the variance  $V(\ell) = \text{var}(a_n(\ell))$  is given by

$$V(\ell) = \sigma_a^2 \sum_{i=0}^{\ell-1} \psi_i^2,$$

where,

$$\begin{aligned} \psi(B) &= \sum_{i=0}^{\infty} \psi_i B^i, \\ &= \frac{\theta^*(B)}{\varphi(B)}. \end{aligned}$$

The  $\psi$ 's are referred to as the coefficients in the random shock form of the model.

In the case where  $\lambda \neq 1$ , i.e. the data have been transformed with a Box-Cox transformations, the minimum-mean-square-error forecasts needs a slight adjustment. Data which have been transformed (for example by taking logs) are said to be in the transformed domain. The original data is said to be in the untransformed domain. For most forecasting situations we are most interested in the forecasts in the untransformed domain.

Forecasts and their prediction intervals are first produced for the transformed domain as described above. Next the inverse Box-Cox transformation,

$$\begin{aligned} z_t &= (z^{(\lambda)} \lambda + 1)^{\frac{1}{\lambda}} - c, \\ &= e^{z^{(\lambda)}} - c, \end{aligned}$$

is applied to the forecasts and their prediction intervals in the transformed domain to obtain estimates in the untransformed domain. The forecasts thus produced in the untransformed domain are said to be naive forecasts since they are not the exact minimum-mean-square-error forecast but rather an approximation to it. Notice that the inverse Box-Cox transformation is valid for all  $z_t$  when  $\lambda = 0$  but is only valid for  $z_t > -1/\lambda$  when  $\lambda \neq 0$ . It is possible, although extremely rare in practice, that  $z_t > -1/\lambda$  when  $\lambda \neq 0$  so the inverse transformation can not be made. In this case a warning message is given and the inverse values are replaced by the smallest feasible value, i.e.  $-c$ .

The exact minimum-mean-square-error forecast in the untransformed domain is determined from the fact that its transformed value follows a Normal distribution with expected value  $z_n^{(\lambda)}(\ell)$  and variance  $V(\ell)$ . The expected value of the inverse Box-Cox transformed value is the desired minimum-mean-square-error forecast. Thus the minimum-mean-square-error forecast,  $z_t(\ell)$ , is given by

$$\begin{aligned} z_t(\ell) &= \frac{1}{\sqrt{2\pi V(\ell)}} \int_{-\infty}^{\infty} (\lambda y + 1)^{\frac{1}{\lambda}} e^{-\frac{1}{2} \frac{(y - z_t^{(\lambda)}(\ell))^2}{V(\ell)}} dy, \quad \lambda \neq 0, \\ &= e^{z_t^{(\lambda)}(\ell) + \frac{1}{2} V(\ell)}, \quad \lambda = 0. \end{aligned}$$

The required integral may be determined numerically by Hermite polynomial integration.

In practice, it is found that the minimum-mean-square-error forecasts are slightly smaller than the naive forecasts. Also, studies with real data have shown that these minimum-mean-square-error forecasts do perform better than the naive forecasts.

## *Comparison Of Forecast And Actuality*

It sometimes happens that we have knowledge that a possible change or intervention occurred at some point in a time series. For example, a new pollution abatement regulation which reduces the amount of phosphates in detergents could be expected to reduce the level of phosphorous concentration in river water quality time series. Similarly seat-belt legislation, may be expected to reduce the number of fatal car accidents. In fact there are numerous situations with environmental and socio-economic series where it may be suspected that a known change may have caused a change in the level of the data. A very simple and intuitive method for examining the statistical significance of the possible change is to compare the actual data after the intervention with the forecasts for the data based on a model fit prior to the intervention.

More exactly, suppose that the series  $z_1, \dots, z_T$  represents the data before the intervention has occurred and that we have data  $z_{T+1}, \dots, z_{T+L}$  after the intervention has occurred. Then we fit a suitable ARIMA model to the  $z_1, \dots, z_T$  using the **ARIMA Estimation** function and save the output file in the FITS directory. Next we select the function **Comparison Of Forecast And Actuality** and input the fitted model from the FITS directory and also the second time series  $z_{T+1}, \dots, z_{T+L}$  as a **separate** data file in the DATA directory.

To test the statistical significance of the discrepancy between these forecasts and actual values, Box & Tiao (1976) proposed the test statistic

$$Q = \zeta' \Upsilon^{-1} \zeta,$$

where  $\zeta' = (z_{T+1} - z_T(1), \dots, z_{T+L} - z_T(L))'$ , and  $\Upsilon$  is the covariance matrix of the forecast errors. Under the null hypothesis of no intervention effect,  $Q \sim \chi_L^2$ . In fact, it can be shown that  $Q$  can more easily be obtained using the following result,

$$Q = \frac{1}{\sigma_a^2} \sum_{i=1}^L a_{T+i-1}(1)^2,$$

where  $a_{T+i-1}(1) = z_{T+i} - z_{T-i+1}(1)$ . The output tabulation includes these one-step forecasts and forecast errors and the value of  $Q$  and its significance level.

Note that the Box-Tiao statistic given above is essentially equivalent to the Janus quotient proposed by Gadd and Wold (1964). The Janus,  $J$ , coefficient is defined as the ratio of the mean square error of the one-step ahead forecast errors in the prediction period divided by the residual variance of the fitted model. Hence,

$$J = \frac{Q}{L}.$$

The plot shows the forecasts at origin time  $T$  for lead times  $i = 1, \dots, L$ , i.e.,  $z_T(1), \dots, z_T(L)$ , and the actual values  $z_{T+1}, \dots, z_{T+L}$ . Notice that this is different from the optional tabulation which just shows the one-step forecast and errors.

### *Forecast A Fitted Periodic Autoregression*

Not available in current version.

*SIMULATION*

## *Simulate An ARIMA Model*

This function can simulate a time series which can be transformed via a Box-Cox transformation to normality or a time series with innovations from the Johnson curve family of distributions (see Appendix 5). The user can select from the menu whether to use the normal distribution or a Johnson curve distribution. If the Johnson curve is selected, a further menu prompts for the values of the skewness coefficient,  $g_1$  and kurtosis coefficient,  $g_2$ .

Let  $z_1, \dots, z_N$  denote the time series to be simulated. Then  $z_t$  satisfies the equations:

$$\phi(B)\Phi(B^s)(w_t - \mu_w) = \theta(B)\Theta(B^s)a_t,$$

where,

$$\begin{aligned} w_t &= \nabla^d \nabla_s^{d_s} z_t^{(\lambda)}, \\ z_t^{(\lambda)} &= \frac{(z+c)_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0, \\ \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\ \Phi(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s} \end{aligned}$$

and

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s}.$$

It is assumed that

$$a_t \sim \text{NID}(0, \sigma_a^2).$$

The first step in the simulation is to generate the  $a$ 's. The random number generator, Superduper, developed by Marsaglia (1975) is used to generate uniform (0,1) random variables. The structure of Superduper is described in Appendix 6. Note that Superduper requires two integer seed values. In the menu prompt the first seed ISEED refers to the seed for the congruential part of Superduper and JSEED refers to shift-register part (see Appendix 6 for details). The current values of these seeds are stored in an ASCII file SEEDS.RNG. Next the generated uniform random variables are transformed to NID(0,1) via the Marsaglia polar transformation. Then for  $t = D + q + 1, \dots, N$  the  $a$ 's are obtained by multiplying by  $\sigma_a$ .

The next step is to generate the  $w_t$ 's. This is done by first generating random initial values of  $(w_{1+D}, \dots, w_{p+D}, a_{1+D}, \dots, a_{q+D})$ , where  $D = d + sd_s$ . These initial values are generated using the multivariate normal distribution with the correct means and covariance matrix. Then the remaining  $w_t$ 's are obtained recursively.

After this the transformed sequence  $z_t^{(\lambda)}$  is calculated by integrating the  $w_t$ 's using user-supplied starting values  $z_1, \dots, z_D$ . These starting values are first transformed via

the Box-Cox transformation. The starting values may be supplied either using values at the beginning or end of another time series data file or may be input directly.

Finally, an inverse Box-Cox transformation is applied to generate the final simulated series. Sometimes, data can be generated for which the inverse Box-Cox transformation is not valid. In this case a warning message is produced and the invalid data values are replaced by the smallest valid values, i.e. `-CONS`, where `CONS` is the parameter in the original Box-Cox transformation.

The series is output for  $t = N_\omega + 1, \dots, N$ , where  $N_\omega$  is the number of values allowed for warm-up. Since an exact procedure is used for stationary series, one should choose  $N_\omega = 0$  when  $D = 0$ .

Up to 1,000 time series replications can be simulated in one run. The simulations are put in files `SIM000.1`, `SIM001.1`, etc. in the `DATA` directory.

Another option allows an optional time series plot of the series to be done immediately after it is generated.

## *Simulate A Fitted ARIMA Model*

This works similar to the function `Simulate An ARIMA Model`. The main difference is that the parameters from a fitted ARIMA model can be automatically fed into the function. Either a normal or Johnson curve distribution may be selected for generating the white noise.

There is one major difference however. In the case of a fitted model, the user can select one of three different options for the starting values:

1. **Random Starting Values.** In this case the starting values are chosen using the correct joint distribution of the initial time series values and their innovations.
2. **Starting Values From Start Of Series.** In this case the values of the time series and estimated innovations at the start of the series are used.
3. **Starting Values From End of Series.** The most recent time series values and their innovations are used to start the simulation. This technique is useful if the purpose of the simulation is to show possible *future* realizations of a particular time series.

If a Box-Cox transformation is used in the modelling of the original data then an inverse Box-Cox transformation must be applied if it is desired to simulate the data in the untransformed domain. Sometimes, data can be generated for which the inverse Box-Cox transformation is not valid. In this case a warning message is produced and the invalid data values are replaced by the smallest valid values, i.e. `-CONS`, where `CONS` is the parameter in the original Box-Cox transformation.

## *Filters*

## *MTS Command Editor*

The Command Editor can be used to perform many editing and other manipulations of time series. The basic capabilities are now discussed.

- (1) **Extracting a Subset.** Often it is desired to extract a portion of the such as all values corresponding to say the month of May in a time series of monthly values. This can be done by choosing a **Systematic Sample** with  $k = 1$  and dropping the first 4 values from the beginning of the series.

Another common situation, is to look a just a portion of the series. Say from  $t = n_1$  to  $t = n_2$ .

- (2) **Reversing the Chronological Order of a Time Series.** For example, sometimes one wishes to provide estimates of the past values. This could be done by modelling and forecasting the reversed time series.
- (3) **Converting the Units of Measurement.** Sometimes the series is in one unit of measurement (example, English system) and we wish to convert to metric. This can be effected by a suitable location-scale transformation,

$$z'_t = \alpha + \delta z_t.$$

- (4) **Recode a Data Value.** Values such as missing value codes or detection limits can be altered.
- (5) **Numeric Formatting.** The series may be re-formatted in either exponential or fixed form. This is sometimes convenient when the data file is to be listed in a table in a report.
- (6) **Concatenation.** Time series data files can be concatenated to produce longer series or bivariate series data files. By repeatedly using the command editor, trivariate series could be also be formed.

## *Aggregate Time Series*

This utility is used to convert seasonal data to non-seasonal data by converting, say, the monthly series to an annual series either by taking the total or mean of the monthly values.

## *ARIMA Filter*

Given a data sequence  $x_t, t = 1, \dots, N$ , an ARIMA filter transformation is applied to produce the filtered output sequence  $\beta_t, t = 1, \dots, N$ . Thus,

$$x_t \xrightarrow{\text{ARIMA filter}} \beta_t.$$



The general ARIMA( $p, d, q$ )( $p_s, d_s, q_s$ ) $_s$  filter for a data series,  $x_t$ ,  $t = 1, \dots, N$ , may be written,

$$\phi(B)\Phi(B^s)(w_t - \mu_w) = \theta(B)\Theta(B^s)\beta_t,$$

where,

$$w_t = \nabla^d \nabla_s^{d_s} x_t^{(\lambda)},$$

and

$$\begin{aligned} x_t^{(\lambda)} &= \frac{(x+c)_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(x_t + c), \quad \text{if } \lambda = 0. \end{aligned}$$

The regular autoregressive operator and regular moving-average components are:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

The seasonal period parameter is  $s$ . For example,  $s = 12$ , for monthly seasonal data. The seasonal autoregressive operator and seasonal moving-average components are:

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp_s}$$

and

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq_s}.$$

To start the recursive calculation of  $\alpha_t$ , initial values of  $\alpha_t$  are set to 0.0.

The principal use of this calculation is in the input prewhitening method of model identification of single input transfer-functions (see Chapter 11 of Box & Jenkins, 1975). Some other applications include: calculation of residuals, prefiltering before spectral analysis, simulation of explosive or non-invertible time series.

## *Deseasonalize a Trend-Free Time Series*

Let  $z_t = z_{r,m}$  denote a seasonal time series with observation numbers  $t = 1, \dots, n$  and seasonal period  $s$ . The notation  $z_{r,m}$  means we think of the  $t$ -th observation as occurring in the  $r$ -th year and  $m$ -th season. Thus  $t = s(r-1) + m$ . If the seasonal means  $\mu_m$  and seasonal standard deviations  $\sigma_m$  of the series were known, the deseasonalized series is given by,

$$\tilde{z}_{r,m} = \frac{z_{r,m} - \mu_m}{\sigma_m}.$$

When the seasonal variance is the same for each month, then it is simpler to use the deseasonalized series given by,

$$\tilde{z}_{r,m} = z_{r,m} - \mu_m.$$

In practice, estimates of  $\mu_m$  and  $\sigma_m$  are used. The simplest estimates are just the sample estimates given by

$$\hat{\mu}_m = \frac{1}{n_m} \sum_{r=1}^{n_m} z_{r,m},$$

and

$$\hat{\sigma}_m = \sqrt{\frac{1}{n_m} \sum_{r=1}^{n_m} (z_{r,m} - \hat{\mu}_m)^2},$$

where  $n_m$  is the number of time series observations occurring in the  $m$ -th month.

One way of reducing the number of parameters,  $2s$  in total — counting seasonal means and standard deviations, is to fit a harmonic regression. This is especially useful for large values of  $s$ . In this approach the seasonal estimates  $\hat{\mu}_m$  and  $\hat{\sigma}_m$  are regressed on  $\cos(\omega jm)$  and  $\sin(\omega jm)$ , where  $j = 1, \dots, J$  is the harmonic component and

$$\omega = \frac{2\pi}{s}$$

denotes the fundamental frequency.

For brevity we will only discuss the harmonic regression for the seasonal means  $\hat{\mu}_m$  since the details for the seasonal standard deviations are virtually the same. When  $J$  is chosen equal  $F_s$  where

$$\begin{aligned} F_s &= \frac{1}{2}s, & \text{if } s \text{ is even,} \\ &= \frac{1}{2}(s-1), & \text{if } s \text{ is odd,} \end{aligned}$$

then the full harmonic regression for the seasonal means is given by

$$\hat{\mu}_m = A_0 + \sum_{j=1}^{F_s} \{A_j \cos(\omega j) + B_j \sin(\omega j)\}$$

where

$$A_0 = \frac{1}{s} \sum_{m=1}^s \hat{\mu}_m$$

and for  $j = 1, \dots, F_s$ ,

$$A_j = \frac{2}{s} \sum_{m=1}^s \hat{\mu}_m \cos(\omega jk)$$

and

$$B_j = \frac{2}{s} \sum_{m=1}^s \hat{\mu}_m \sin(\omega jk).$$

Notice that if  $s$  is even,  $B_{F_s} = 0$  and

$$A_{F_s} = \sum_{m=1}^s (-1)^j \hat{\mu}_m.$$

In the case of the full harmonic regression, there is a perfect fit so the coefficient of determination,  $R^2$ , which indicates the fraction of variation accounted for the model is 100%. The number of parameters required can be reduced by choosing the number of harmonic terms,  $J$ , to be less than  $F_s$ . In this case the estimated seasonal means are given by,

$$\tilde{\mu}_m = A_0 + \sum_{j=1}^J \{A_j \cos(\omega j) + B_j \sin(\omega j)\}.$$

Note that since the independent variables are orthogonal the regression coefficients  $A_0$ ,  $A_m$  and  $B_m$  for  $m = 1, \dots, J$  are identical to the coefficients in the full harmonic regression. It can be shown that the coefficient of determination for this harmonic regression based on the first  $J$  harmonics is given by

$$R_J^2 = \frac{s}{2} \sum_{j=1}^J (A_j^2 + B_j^2).$$

Our program tabulates the value of  $R_J^2$  for  $J = 1, \dots, F_s$  so the user can decide the number of terms needed in the harmonic regression. Generally an adequate number of terms will produce a coefficient of determination in excess of 90%. The same procedure can be used to obtain the fitted standard deviations  $\hat{\sigma}_m$ .

The precise seasonal adjustment is determined by the parameters the user selects for the Number of Harmonic Components for the mean and standard deviation, say  $J_\mu$  and  $J_\sigma$ . The deseasonalized series is given by,

$$\tilde{z}_{r,m} = \frac{(z_{r,m} - \tilde{\mu}_m)}{\tilde{\sigma}_m},$$

or when  $J_\sigma = 0$  by

$$\tilde{z}_{r,m} = z_{r,m} - \tilde{\mu}_m.$$

## *Algebraic Combination Of Two Series*

Given two series  $x_t$  and  $y_t$ , we may be interested in various combinations such as  $z_t = \frac{1}{2}(x_t + y_t)$  or  $z_t = x_t - y_t$ . For example, in water quality time series we may be interested in the total load,  $z_t$ , given the concentration,  $x_t$ , and the instantaneous discharge,  $y_t$ . This would be given by  $z_t = x_t y_t$ . As another example, given the original data,  $x_t$ , and the residuals,  $y_t$ , from some model we may wish to calculate the fitted values  $z_t = x_t - y_t$ . As a final example, given the total unemployment series,  $x_t$ , and the

total labor force,  $y_t$ , we may wish to calculate the percentage unemployment given by  $z_t = 100.0x_t \div y_t$ .

In general, let

$$u_t = \nabla^{d^{(x)}} \nabla_{s_x^{d^{(x)}}} x_t^{(\lambda_x)},$$

where,

$u_t$  is the transformed data sequence,

$\lambda_x$  is the exponent in the Box-Cox transformation,

$\nabla_x$  is the first differencing operator,

$d^{(x)}$  is the differencing parameter,

$\nabla_{s_x}$  is the seasonal differencing operator,

$s_x$  is the seasonal period,

$d^{(x)}_s$  is the seasonal differencing parameter,.

and let

$$v_t = \nabla^{d^{(y)}} \nabla_{s_y^{d^{(y)}}} y_t^{(\lambda_y)},$$

where,

$u_t$  is the transformed data sequence,

$\lambda_y$  is the exponent in the Box-Cox transformation,

$\nabla_y$  is the first differencing operator,

$d^{(y)}$  is the differencing parameter,

$\nabla_{s_y}$  is the seasonal differencing operator,

$s_y$  is the seasonal period,

$d^{(y)}_s$  is the seasonal differencing parameter.

Then the general form of the algebraic combination of  $x_t$  and  $y_t$  is given by

$$z_t = (\alpha_x u_{t+\ell}) \text{ f } (\alpha_y v_{t+k}),$$

where  $\alpha_x$  and  $\alpha_y$  are user-specified constants,  $\ell$  and  $k$  are user-specified lags, and f denotes one of the user-specified binary operations: addition (+), subtraction (−), multiplication (×) or division (÷).

## *Difference Or Integrate A Time Series*

Differencing and integrating a time series are inverse operations. They are not precisely inverse operations however since in differencing a time series we lose the initial values in the series and in integrating a time series we must start with certain initial values.

Given an observed time series  $z_t$ ,  $t = 1, \dots, n$  the differenced series  $w_t$ ,  $t = d + 1, \dots, n$  satisfies the equation

$$w_t = \nabla^d z_t$$

where  $\nabla = 1 - B$  is the backward difference operator which is equal to the identity minus the backshift operator  $B$  and  $d$  is a nonnegative integer which indicates the degree of differencing. The simplest and best algorithm for differencing a time series involves applying first differences  $d$  times. The length of the resulting series is  $n - d$ .

Given an observed time series  $z_t$ ,  $t = 1, \dots, n$  the integrated series  $u_t$ ,  $t = 1 - d, \dots, n$  satisfies the equation

$$u_t = \mathcal{S}^d z_t$$

where  $\mathcal{S} = \nabla^{-1} = 1 + B + B^2 + B^3 + \dots$  is the backward summation operator which is equal to the inverse of the backward differencing operator. Notice that when differencing the series we lose  $d$  initial values whereas when integrating with  $d \geq 1$  we must start with  $d$  initial values  $u_0, \dots, u_{1-d}$ . The following algorithm produces an integrated time series of length  $n + d$ :

*Step 0:* Set  $i \leftarrow 1$ .

*Step 1:* For  $t = 1, \dots, n$ , set  $u'_t \leftarrow u_t$

*Step 2:* For  $t = 0, \dots, 1 - d - i$ , set  $u'_t \leftarrow \nabla^{d-i} u_t$ .

*Step 3:* For  $t = 1, \dots, n$ , set  $u'_t \leftarrow u'_t + u'_{t-1}$

*Step 4:* Increment  $i \leftarrow i + 1$

*Step 5:* If  $i \leq d$ , return to Step 2. Otherwise continue.

*Step 6:* For  $t = 1, \dots, n$ , set  $u_t \leftarrow u'_t$ .

Both the differencing and integration algorithms extend directly to the more general case where seasonal differencing and integration are used. That is in the case of seasonal differencing we have

$$w_t = \nabla_s^{d_s} \nabla^d z_t,$$

where  $\nabla_s = 1 - B^s$  and  $s$  denotes the seasonal period and  $d_s$  denotes the degree of seasonal differencing. Similarly in the case of seasonal integration,

$$u_t = \mathcal{S}_s^{d_s} \mathcal{S}^d z_t,$$

where  $\mathcal{S}_s = 1 + B^s + B^{2s} + \dots$

In the case of integration the  $D = d + sd_s$  initial values are taken from the beginning of the input time series data file. If necessary you can concatenate the initial values,  $u_0, \dots, u_D$  with the data series  $z_1, \dots, z_n$  using the Command Editor.

## Box-Cox Or Inverse Box Cox Transformation

The Box-Cox transformation of  $z_t$ ,  $t = 1, \dots, n$  is given by

$$\begin{aligned} z_t^{(\lambda)} &= \frac{(z_t + c)^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0, \end{aligned}$$

where  $c$  is some constant. The Box-Cox transformation is only defined when  $z_t + c > 0$  for all  $t = 1, \dots, n$ .

Its inverse is given by

$$\begin{aligned} z_t^{[\lambda]} &= (\lambda z_t + 1)^{\frac{1}{\lambda}} - c, \quad \text{if } \lambda \neq 0, \\ &= e^{z_t} + c, \quad \text{if } \lambda = 0, \end{aligned}$$

where  $c$  is some constant. Notice that the inverse Box-Cox transformation is valid for all  $z_t$  when  $\lambda = 0$  but is only valid for  $z_t > -1/\lambda$  when  $\lambda \neq 0$ .

An error message is produced for an invalid transformation.

*UTILITIES*

## *Data Screen Editor and Viewer*

The screen editor may be used to modify an existing file create a new file or simply to view a data file. When you are prompted for **Filename For Output**, set this to the name of the file you wish to save after the editing session is complete. It may be set to the same name as an existing data file, in which case the file will be overwritten *after the editing session has been complete*. You are queried **Edit An Existing Dataset**. Reply **Yes** if you wish to modify or edit an existing data file and **No** if the file you wish to enter a new data file using the editor.

The default name for the **Filename For Output** is left blank. If this is not changed, then no output will be generated from the editor. This setting should be used if you just want to look at the data file.

Key Code	Function
→	move right one character
←	move left one character
↑	move up one line
↓	move down one line
HOME	move to beginning of current line
END	move to end of current line
PAGEUP	move up one screen
PAGEDOWN	move down one screen
ALT—S	search for text string
ALT—R	search for and replace text string
ALT—M	turns marking mode on if its off; turns marking mode off if its on;
GREYPLUS	copies the marked block to the cut buffer
GREYMINUS	cuts the marked block to the cut buffer
INSERT	pastes the cut buffer and the current cursor location
DELETE	delete the marked block if there is one else delete single char

In addition to these commands, the contents in the window can be scrolled using the scroll bar on the right of the text window. At present the Screen Editor can handle files up 9000 characters (or about 100 lines).

## *Data Documentation Editor & Viewer*

DOC is a subdirectory of the current DATA directory which contains documentation files for each dataset in the current DATA directory. Initially, the documentation



file for a particular dataset may contain only the title string from the original data file. As the analysis of a particular dataset proceeds, the analysis may wish to note some of the results in this file. Also, special features of the data or work done by other researchers can be noted in this file. The documentation file has the same name as the dataset file. This function allows one to examine and up-date the documentation file.

## *Describe DATA Directory*

For the current DATA directory, a list showing all the file names and the first line, which is normally the title string, is shown. This list of files and titles is saved in the current DATA directory in a file named DESCRIBE.

## *View or Edit DATA Directory SUMMARY*

It is often convenient to give a brief description of the contents of a DATA directory. For example, the DATA directory may contain a collection of time series all from a certain project or textbook. The DATA directories supplied with the MTSpackage all have such summaries. For user created DATA directories, the DATA directory SUMMARY will initially be empty. The contents of the DATA directory SUMMARY are stored in a file named SUMMARY in the DATA directory itself.

## *Set New DATA Directory*

The standard default DATA directory when the MTSpackage is installed is called DATA and it is always a subdirectory of the current working directory. The user may change his DATA directory to any directory located anywhere on in their computer system, although the use of floppy drives or optical drives, is not recommended as these drives would be too slow. If an invalid drive is entered in the function, the DATA directory reverts to the original default DATA directory located off the current working directory. Any change in the DATA directory will be saved and remain effective from the current session to the next. The DATA directory is saved in the file MTS.CFG which is located off the current working directory.

Changing the current DATA directory does not in any way affect or change the subdirectories OUTPUT, INPUT and FITS which remain the same. These directories are always subdirectories off of the current working directory.

A similar facility has not been provided by the OUTPUT, INPUT and FITS subdirectories since you can conveniently manipulate the contents of these subdirectories by using the Delete, Copy or Move Files utility.

## *View File in OUTPUT Directory*

This utility is used to view or print a text file in the OUTPUT directory. Time series data files may be viewed with the MTS screen editor. For convenience this utility is also included submenu for `model estimation`.

## *Delete, Copy or Move Files*

This is a general utility which can be used for cleaning up directories and also backing important files to another directory or drive. Files may be deleted or copied or moved. Moving a file is the same as copying it to a new directory and deleting it in the old one. A convenient interface is provided so that the operations can be performed simultaneously on selected files in any given directory.

## *Conversion From Column Format To Row Format*

This function is useful for importing data from Spreadsheets to MTS . The input file is assumed to contain the data by column. So the first column corresponds to the first time series, the second column to the second time series, etc. *No title string on the input is assumed* since the menu prompts for the title for the output file. *The extension of the input file must be .DAT*. An output file with extension `.k` is produced where `k` is the number of columns on the `.DAT` file.

For example, an input `.DAT` file which looks like this:

```
200.10  10.01
199.50  10.07
199.40  10.32
198.90  9.75
.....
.....
.....
.....
.....
```

could be converted to an MTS data file with extension `.2` which looks like this:

```
SERIES M, SALES DATA WITH LEADING INDICATOR
200.10  199.50  199.40  198.90  ...
.....
10.01  10.07  10.32  9.75  ...
.....
```

## *Conversion From Row Format To Column Format*

Spreadsheets and other statistical packages usually store variables by columns. However, since the most frequent data type in MTS are univariate time series the most

natural representation in MTS is by rows. This utility function is provided to make it convenient to export MTS data files to other Spreadsheets and other packages.

This conversion utility can convert an MTS data file or files to a file with the variables arranged by column. For example the bivariate data file SERIESM.2 which looks like this:

SERIES M, SALES DATA WITH LEADING INDICATOR

200.10 199.50 199.40 198.90 ...

.....

10.01 10.07 10.32 9.75 ...

.....

Could be converted to a file which looks like this:

200.10 10.01

199.50 10.07

199.40 10.32

198.90 9.75

.....

.....

.....

.....

.....

The data files to be converted can be a series of MTS data files with extensions .1, .2, etc. The number of columns on the output file is then the sum of the extensions of the input files. For example, a data file with extension .1 could be combined with another file containing residuals (which also has extension .1) to produce an output file with two columns. The output file normally has extension .DAT and is placed in the DATA subdirectory. There is no title string included on the .DAT output file.

*APPENDICIES*

## Appendix 1: Robust Locally Weighted Regression

### Introduction and Summary

This methodology was introduced by Cleveland (1979) for modelling and smoothing bivariate data  $(x_t, y_t)$ ,  $t = 1, \dots, n$ . This technique provides a general and flexible nonlinear family of models to fit bivariate data. Basically, the model may be written as

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t.$$

The parameters  $\alpha_t$  and  $\beta_t$  are estimated using a locally weighted regression (loess) technique which puts higher weight on data values close to  $t$  than those data values which are further away. The specific weighting used is determined by the smoothing parameter  $f$  which the user selects. The parameter  $f$  indicates the fraction of data used in the procedure. If  $f = 0.5$ , only half the data is used in the estimation and in effect, a moderate amount of smoothing is done. On the other hand, if  $f = 0.8$ , then eighty percent of the data is used and much more smoothing is done. In all cases however, locally weighing is used which weights the data close to  $t$  more than that further away.

The estimation procedure is not least-squares but rather a robust method which guards against outliers. A plot of the fitted values

$$\hat{y}_t = \hat{\alpha}_t + \hat{\beta}_t x_t$$

versus  $x_t$  is useful for summarizing the relationship between  $y_t$  and  $x_t$ . The **Bivariate Scatter Plot** function has an option for performing robust loess smoothing. In checking the goodness of fit of a robust loess regression it is useful to look at the usual regression residual plots, viz. (i) residual vs. observation number, (ii) residual vs. fitted value, (iii) residual vs. explanatory variable and (iv) residual normal probability plot. As pointed out by Cleveland, it may be preferable to use the absolute residual vs. fitted value instead of plot (ii) with a robust loess smooth to check for trends or other systematic features.

Another useful application of robust loess is to smoothing a time series. Here  $y_t = z_t$  and  $x_t = t$ . Robust loess smooths with a smoothing parameter of around  $f = 0.8$  are useful in looking for trends in environmental series (McLeod, Hipel & Bodo, 1990). The **Robust Loess Smooth** function in MTS calculates and plots the robust loess smooth of a time series.

Another application is to the diagnostic checking of fitted models (regression or ARIMA). In diagnostic checking it is often recommended to look at a plot of the fitted values versus the residual. As pointed out by Cleveland, this can be misleading when the data is concentrated around the high values. A better technique in this situation is to plot the absolute residuals versus fitted values and look for a trend using the robust loess smooth. This can also be done in MTS by importing the fitted values and residuals as two separate files. Then use the **MTS Command Editor** to take the absolute values of the residuals. Next select the **Bivariate Scatter Plot**. Input first the residuals, then the fitted values data series and then select the robust loess option.

When  $n > 100$ , the computations can be lengthy. In this case one can take a short-cut in the calculations by computing the estimates  $\hat{\alpha}$  and  $\hat{\beta}$  at values separated by at least  $\delta$  units. The parameter  $\delta$  can be selected by the user although it is recommended to use the default values which are shown on the menu. These default values are determined as follows:

$$\begin{aligned} &\text{if } n \leq 100, \quad \delta = 0, \\ &\text{if } n > 100, \quad \delta = 0.03 \text{ IQ, where IQ denotes the inter-quartile range.} \end{aligned}$$

With these values the full computations are carried out at about 100 data points.

## Technical Details

The basic underlying model supposes that

$$y_t = g(x_t) + \epsilon_t, \quad t = 1, \dots, n,$$

where  $g(x)$  is a smooth function and  $\epsilon_t$  is an error which has mean zero and constant scale. Then by the smoothness of  $g$  we can approximate by a locally linear function to obtain,

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t.$$

In order to define the algorithm, we introduce the concept of the local weight function,  $w(x_t)$  and the robustness weights  $\delta(x_t)$ .

### Local Weight Function.

Let

$$\begin{aligned} W(z) &= (1 - |z|^3)^3, \quad \text{for } |z| < 1, \\ &= 0, \quad \text{for } |z| \geq 1, \end{aligned}$$

For the smoothing parameter  $0 < f \leq 1$ , let  $r$  denote the quantity  $fn$  rounded to the nearest integer. Let  $h_t$  denote the distance to the  $r$ -th nearest neighbour of  $x_t$ . Then the local weight for  $x_t$  for any value of  $x$  is

$$w_t(x) = W\left(\frac{(x - x_t)}{h_t}\right).$$

### Robustness Weight Function.

Let  $\tilde{\epsilon}_t$  denote an initial estimate of the error term  $\epsilon_t$  which is obtained after *Step 2* below. Let  $s$  denote the median of  $|\tilde{\epsilon}_1|, |\tilde{\epsilon}_2|, \dots, |\tilde{\epsilon}_n|$ . Then the robustness weight for  $(x_t, y_t)$  is denoted by  $\delta_t$  and given by

$$\delta_t = B\left(\frac{\tilde{\epsilon}_t}{6s}\right),$$

where  $B(z)$  is the bisquare function defined by

$$\begin{aligned} B(z) &= (1 - |z|^2)^2, \quad \text{for } |z| < 1, \\ &= 0, \quad \text{for } |z| \geq 1, \end{aligned}$$

### Robust Loess Method.

- Step 1.* For each  $t = 1, \dots, n$  a weighted simple linear regression of the  $y$ 's on the  $x$ 's is fitted using weights  $w_t$ . This produces estimates  $\tilde{\alpha}_t$  and  $\tilde{\beta}_t$ . Also a fitted value  $\tilde{y}_t = \tilde{\alpha}_t + \tilde{\beta}_t$  is produced.
- Step 2.* The initial residuals are obtained as  $\tilde{\epsilon}_t = y_t - \tilde{y}_t$ . Then the robustness weights  $\delta_t$  are calculated.
- Step 3.* For each  $t = 1, \dots, n$  a weighted simple linear regression of the  $y$ 's on the  $x$ 's is fitted using weights  $\delta_t w_t$ . This produces new estimates  $\tilde{\alpha}_t$  and  $\tilde{\beta}_t$ . Also a fitted value  $\tilde{y}_t = \tilde{\alpha}_t + \tilde{\beta}_t$  is produced.
- Step 4.* The next set of residuals are obtained as  $\tilde{\epsilon}_t = y_t - \tilde{y}_t$ . Then the next set of robustness weights  $\delta_t$  are calculated.
- Step 5.* For each  $t = 1, \dots, n$  a weighted simple linear regression of the  $y$ 's on the  $x$ 's is fitted using weights  $\delta_t w_t$ . This produces the final estimates of  $\hat{\alpha}_t, \hat{\beta}_t, \hat{y}_t = \hat{\alpha}_t + \hat{\beta}_t$  and  $\hat{\epsilon}_t = y_t - \hat{y}_t$ .

Cleveland has provided an updating algorithm which speeds the fitting of the  $n$  regressions in Steps (1), (3) and (5). However, if  $n > 100$  the computations can still be quite time-consuming.

## Appendix 2: Box-Cox Transformations

In many cases a batch of data may be made more symmetric by use of a power transformation such as square-root, log or negative reciprocal. All of these transformations preserve the ordering of the data but change the shape of the distribution. The shape of the distribution can be seen graphically by examining stem-and-leaf and/or box plots. Thus by examining these plots for various power transformations we can choose the most effective transformation.

An alternative method of choosing the transformation is to estimate from the data the optimal Box-Cox transformation. This method is based on the principle of maximum likelihood which provides the best estimators in many situations. In addition, this method can be extended to improve many statistical methods which assume the normal distribution. These methods include regression, ANOVA, two-sample tests, experimental design and time series analysis.

The family of Box-Cox transformations (Box & Cox, 1964) for a batch of data  $y_1, y_2, \dots, y_n$  is defined by

$$y_t^{(\lambda)} = \frac{y_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0,$$

$$= \log(y_t), \quad \text{if } \lambda = 0.$$

Note that subtracting 1 and dividing by  $\lambda$  do not change the shape of the data, so  $\lambda = 0.5, 0, -1$  correspond to the square-root, log and negative reciprocal transformations. It is also important, that the Box-Cox transformation family may be regarded as a continuous function of  $\lambda$ .

### Application to a Batch of Data

If it is assumed that for some value of  $\lambda$  that  $y_1^{(\lambda)}, y_2^{(\lambda)}, \dots, y_n^{(\lambda)}$  are independent normal with mean,  $\mu$ , and variance,  $\sigma^2$ , i.e.

$$y_t^{(\lambda)} \sim \text{NID}(\mu, \sigma^2).$$

Then it may be shown using the methods of mathematical statistics that the log likelihood function for the parameter  $\lambda$  may be written

$$\log(L(\lambda)) = -\frac{n}{2} \log(\hat{\sigma}^2) + J$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (y_t^{(\lambda)} - \bar{y}^{(\lambda)})^2,$$

$$\bar{y}^{(\lambda)} = \frac{1}{n} \sum_{t=1}^n y_t^{(\lambda)},$$

and

$$J = (\lambda - 1) \sum_{t=1}^n \log(y_t).$$

The function  $\log(L(\lambda))$  can be tabulated or graphed to determine the value of  $\lambda$  which maximizes the likelihood. This gives the maximum likelihood estimate of  $\lambda$  which is denoted by  $\hat{\lambda}$ . The optimal  $\hat{\lambda}$  may also be determined by numerical optimization as in the MTS package.

### Application to the Random Walk Model

In this case we have a *data sequence*  $z_1, z_2, \dots, z_n$  which are consecutive values at times  $t = 1, 2, \dots, n$ . The model equation may be written,

$$z_t - z_{t-1} = a_t$$

where

$$a_t \sim \text{NID}(0, \sigma^2).$$



Then to apply the Box-Cox method, we assume that the model holds for some transformation of  $z_t$ ,  $z_t^{(\lambda)}$ . Thus the new model may be written

$$z_t^{(\lambda)} - z_{t-1}^{(\lambda)} = a_t$$

where

$$a_t \sim \text{NID}(0, \sigma^2).$$

The log-likelihood function for this new model is

$$\log(L(\lambda)) = -\frac{(n-1)}{2} \log(\hat{\sigma}^2) + J$$

where

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=2}^n \hat{a}_t^2,$$

$$\hat{a}_t = z_t^{(\lambda)} - z_{t-1}^{(\lambda)},$$

and

$$J = (\lambda - 1) \sum_{t=2}^n \log(z_t).$$

As before  $\log(L(\lambda))$  can be tabulated or graphed to determine the value of  $\lambda$  which maximizes the likelihood or an exact value may be found with the MTS package.

### Application to the ARIMA Model

The general form of the ARIMA( $p, d, q$ ) model equation for a data series,  $z_t$ ,  $t = 1, \dots, N$ , may be written,

$$\phi(B)(w_t - \mu_w) = \theta(B)a_t,$$

where,

$$w_t = \nabla^d z_t^{(\lambda)},$$

and

$$\begin{aligned} z_t^{(\lambda)} &= \frac{(z+c)_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0, \\ &= \log(z_t + c), \quad \text{if } \lambda = 0. \end{aligned}$$

The regular autoregressive operator and regular moving-average components are:

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

The log-likelihood function of  $\lambda$  maximized over the other parameters is

$$\log(L(\lambda)) = -\frac{n}{2} \log(\hat{\sigma}^2) + J$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=-\infty}^N \hat{a}_t^2,$$

$\hat{a}_t$  denotes the residual at time  $t$ ,  $n = N - d$  and

$$J = (\lambda - 1) \sum_{t=d+1}^N \log(z_t).$$

The derivation of the above equation and the application of Box-Cox transformations to time series data was discussed by Hipel & McLeod (1977).

### *Appendix 3: Simultaneous Significance Intervals*

In the case of the sample autocorrelation function of iid noise, it is known that the autocorrelations are normally distributed with mean zero and variance  $1/n$ , where  $n$  is the length of the series. Thus a 5% significance limit for the sample autocorrelation at lag  $k$ ,  $r_k$ , is  $\pm 1.96/\sqrt{n}$ . For any significance limit  $\alpha$  let  $c$  denote the value such that

$$P\{|r_k| < c, k = 1, \dots, m\} = 1 - \alpha.$$

Since the  $r_k$  for  $k = 1, \dots, M$  are iid, it follows that  $c = \Phi^{-1}\left((1 - \alpha/2)^{1/m}\right)/\sqrt{n}$ .

TABLE I

$$\Phi^{-1}\left((1 - \alpha/2)^{1/m}\right)$$

$m = 1$	$m = 2$	$m = 10$	$m = 20$	$m = 40$	$m = 60$
1.96	2.24	2.81	3.02	3.23	3.34

As shown by Hosking and Ravishanker (1993) Bonferroni's inequality may be used to determine simultaneous significance intervals which can be used to test if all sample residual autocorrelations are zero.

To understand how this works, let's first consider two random events denoted by  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . Then

$$P\{\mathcal{E}_1 \cup \mathcal{E}_2\} = P\{\mathcal{E}_1\} + P\{\mathcal{E}_2\} - P\{\mathcal{E}_1 \cap \mathcal{E}_2\}$$

Utilizing the fact that

$$P\{\mathcal{E}_i\} = 1 - P\{\bar{\mathcal{E}}_i\} \quad \text{for } i = 1, 2,$$

and that

$$P\{\mathcal{E}_1 \cup \mathcal{E}_2\} \leq 1,$$

we obtain Bonferroni's inequality for the case of two events  $\mathcal{E}_1$  and  $\mathcal{E}_2$ ,

$$P\{\mathcal{E}_1 \cap \mathcal{E}_2\} \geq 1 - P\{\bar{\mathcal{E}}_1\} + P\{\bar{\mathcal{E}}_2\}$$

Bonferroni approximation for the joint probability of  $m$  events,  $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_m$ , is given by

$$P\left\{\bigcap_{i=1}^m \mathcal{E}_i\right\} \approx 1 - \sum_{i=1}^m P\{\bar{\mathcal{E}}_i\}.$$

If the events are statistically independent as in the case of the sample autocorrelations of white noise, the approximation is exact.

Now consider the problem of testing the residual autocorrelations  $r_1, r_2, \dots, r_m$  from an observed time series of length  $n$ . Then if the model is correct,  $r_1, r_2, \dots, r_m$ , have mean zero and some variance,  $\text{var}(r_k)$ , which can be calculated using the formula given by McLeod (1977). Let

$$\mathcal{E}_i = \{|r_i| < \frac{z_\alpha}{\sqrt{n}} \text{est.sd.}(r_i)\},$$

where  $z_\alpha$  denotes upper  $1 - \alpha/2$ -point of a standard normal distribution. So under  $\mathcal{H}_0$ , we have for each  $i = 1, \dots, m$ :  $P\{\bar{\mathcal{E}}_i\} \doteq \alpha$ . Hence the simultaneous or joint probability of the event  $\bigcap_{i=1}^m \mathcal{E}_i$  is approximated with the Bonferroni limit  $m\alpha$ . It follows that in order to get a simultaneous bound such that  $|r_i| < c$ , for all  $i = 1, \dots, m$  we should take

$$c = \frac{z_\alpha}{\sqrt{n}}.$$

The Bonferroni bounds are equivalent to the bounds already derived in Table I above. However, in this case the events  $\mathcal{E}_i$  are not statistically independent so the Bonferroni approximation is needed. Hosking and Ravishanker (1993) demonstrated that the Bonferroni approximation works well for residual autocorrelations.

## *Appendix 4: Spectral Analysis Primer*

Spectral analysis can be regarded as the development of a Fourier analysis for stationary time series. Just as in classical Fourier analysis a real function  $z(t)$  is represented by a Fourier series, in spectral analysis the autocovariance function of a stationary time series has a frequency representation in terms of a Fourier transform. This representation was first given by Herglotz (1911) who showed that any positive-definite function, such as the autocovariance function,  $\gamma_k$ , of a stationary time series can be represented as

$$\gamma_k = \int_{(-\pi, \pi]} e^{i\omega k} dP(\omega), \quad (2)$$

where  $P(\omega)$  is the spectral distribution function. In the case where a spectral density function exists, we have  $dP(\omega) = p(\omega)d\omega$  and Herglotz's equation, eq. (2), can be written

$$\gamma_k = 2 \int_0^{\pi} p(\omega) \cos(\omega k) d\omega. \quad (3)$$

The function  $p(\omega)$ ,  $-\pi \leq \omega \leq \pi$  is called the spectral density function and shares many properties of the probability density function. In addition, note that  $p(\omega)$  is symmetric,  $p(\omega) = p(-\omega)$ . For mathematical convenience the units of  $\omega$  are in radians per unit time, which is known as angular or circular frequency. In practice, however, it is more convenient to work in units of cycles per unit time or Hertz, which is related to  $\omega$  by the equation  $\omega = 2\pi f$ , where  $f$  is now in cycles per unit time.

Notice that there is an upper limit to the highest frequency that can be observed in the time series. This upper limit, which is 0.5 cycles per unit time or  $\pi$  radians per unit time, is called the Nquist frequency. This upper limit arises because of the discrete time nature of our time series (there is no such upper limit in the continuous time case). To see why this occurs, let  $\tilde{f}$  denote any frequency in the interval  $[0, 0.5]$  and let  $\check{f} = \tilde{f} + 0.5$ . Then it is easily shown for all integer  $t$  that  $\cos(\check{f}) = \cos(\tilde{f})$  and  $\sin(\check{f}) = -\sin(\tilde{f})$ . The frequencies  $\tilde{f}$  and  $\check{f}$  are said to be *aliases*. Aliased frequencies, such as  $\check{f}$ , are observationally indistinguishable from frequencies in the range  $[0, 0.5]$ .

A natural estimator of the spectral density function given  $n$  observations  $z_1, \dots, z_n$  from a covariance stationary time series, is given by the periodogram,

$$I(f_j) = \frac{1}{n} \left| \sum_{t=1}^n z_t e^{-2\pi f_j(t-1)} \right|^2, \quad (4)$$

where  $f_j = \frac{j}{n}$ ,  $j = [-(n-1)/2], \dots, 0, \dots, [n/2]$ , where  $[\bullet]$  denotes the integer part function. Since  $I(f_j) = I(-f_j)$  the periodogram is symmetric about 0, and so when the periodogram or spectral density is plotted we only plot the part where  $f_j > 0$ . When  $f_j = 0$ ,  $I(0) = n\bar{z}^2$ , where  $\bar{z} = \sum_{t=1}^n z_t/n$ . This component,  $I(0)$ , is usually very large due to a non-zero mean and so usually ignored in the periodogram and spectral plots.

Schuster (1898) developed the periodogram for searching for periodicities in time series. Sometimes  $I(f_j)$  is plotted against its period  $1/f_j$ .

Taking  $k = 0$  in eq. (3) we obtain for  $\text{var}(z_t) = \gamma_0$ ,

$$\text{var}(z_t) = 2 \int_0^{2\pi} p(\omega) d\omega. \quad (5)$$

It can be shown that spectral analysis provides an anova like decomposition of a time series into its frequency components and this fact is illustrated in eq. (5). The sample analogue of eq. (5) is

$$\sum_{t=1}^n (z_t - \bar{z})^2 = 2 \sum_{j=1}^{[(n-1)/2]} I(f_j) + I(f_{[n/2]}),$$

where the last term  $I(f_{[n/2]})$  is omitted when  $n$  is an odd.

Another important property for the interpretation is that it can be shown that  $I(f_j)$  is proportional to the square of the multiple correlation between the observed data sequence  $z_1, \dots, z_n$  and a sinusoid having frequency  $f_j$ . Specifically, consider the regression,

$$z_t = A_0 + A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t) + e_t, \quad (6)$$

where  $e_t$  is the error term. Then the least-squares estimates of  $A_0$ ,  $A_j$  and  $B_j$  are given by

$$A_0 = \frac{1}{n} \sum_{t=1}^n z_t$$

$$A_j = \frac{2}{n} \sum_{t=1}^n z_t \cos(2\pi f_j t)$$

$$B_j = \frac{2}{n} \sum_{t=1}^n z_t \sin(2\pi f_j t).$$

The multiple correlation coefficient,  $R_j^2$ , can be shown to be given by

$$R_j^2 = A_j^2 + B_j^2.$$

In terms of the periodogram we have then

$$I(f_j) = \frac{n}{2} R_j^2.$$

Thus  $p(f)$  can be interpreted as measuring the strength of a random sinusoidal component having a period  $1/f$  in the data sequence. Time series exhibiting cycles or oscillatory behaviour will have a peak in the spectral density function at the frequency which corresponds to the cycle period. For example, if there is a ten-year cyclical component, then there will be a peak in the spectral density function at  $f = 0.1$ . The sharpness of the peak depends on how closely the period  $1/f$  appears in the data sequence, and the relative size of the peak depends on the relative amplitude of the cycle in the time series. The units of  $f_j$  are cycles per unit time. The period corresponding to  $f_j$  is  $T_j = 1/f_j$ . Alternatively, some researchers prefer to work with angular frequencies given by  $\omega_j = 2\pi f_j$  which has units radians per unit time in which case the period is  $2\pi/\omega_j$ .

In the case where the spectral density function exists, the expected value of  $I(f_j)$  can be shown to be approximately equal to  $p(f_j)$ , where  $p(f)$  is the spectral density function. In fact, in large samples,  $I(f_j)$  for  $j = 1, \dots, [(n-1)/2]$  are statistically independent and exponentially distributed with mean  $p(f_j)$ .

The spectral density function can be derived by taking the inverse Fourier transformation to eq. (3) which yields,

$$p(f) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-2\pi f k}, \quad |f| \leq 0.5,$$

The sample analogue of this formula also holds, viz.

$$I(f) = \sum_{k=-(n-1)}^{n-1} c_k e^{-2\pi f k},$$

where  $c_k$  denotes the sample autocovariance function given by

$$c_k = \frac{1}{n} \sum_{\ell=k+1}^n (z_\ell - \bar{z})(z_{\ell-k} - \bar{z}) \quad \text{for } k \geq 0,$$

and for  $k < 0$ ,  $c_k = c_{-k}$ .

It should be mentioned that some writers prefer to define the spectral density function by the equation

$$p(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-\omega k}, \quad |\omega| \leq \pi.$$

Accordingly, the periodogram is then given by

$$I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n z_t e^{-\omega_j(t-1)} \right|^2,$$

where  $\omega_j = 2\pi j/n$ ,  $j = [-(n-1)/2], \dots, 0, \dots, [n/2]$ . Notice, that since the spectral function is a density function, like a probability density function, the rules for transformation of a density function apply here. In other words, the difference in notation arises according to whether one uses frequency in Hertz or in circular measure.

The periodogram,  $I(f_j)$ , and the estimate of the spectral density function,  $p(f)$ , are useful for providing a nonparametric description and characterization of a time series — particularly a long one. As its name implies it is also useful in determining and characterizing periodicities in the data. Some common types of spectra are listed in the table below.

#### COMMON TYPES OF SPECTRA

Spectral Type	Illustrative AR Model	Example Dataset
high-frequency spectrum	AR(1) $\phi_1 = -0.42$	Series F Batch Data
low-frequency spectrum	AR(3) $\phi_1 = 0.62, \phi_2 = 0, \phi_3 = 0.18$	Mean Annual St. Lawrence Riverflow
peak spectrum	AR(9) $\phi_1 = 1.22, \phi_2 = -0.52$ $\phi_3 = \dots \phi_8 = 0$ $\phi_9 = 0.20$	Annual Sunspot Numbers
trough spectrum	AR(2) $\phi_1 = 0.12, \phi_2 = 0.20$	Annual temperature central England

The spectral density,  $p(f)$ , can be estimated by periodogram smoothing or by fitting autoregressive or ARMA models.

## Appendix 5: Johnson Curve Distributions

The family of Johnson curve distributions, denoted by  $S$  can be used to match the first four moments or equivalently the mean ( $\mu$ ), variance ( $\sigma^2$ ), skewness ( $g_1$ ) and kurtosis ( $g_2$ ) to a fitted statistical distribution. The family of Johnson curve distribution is defined in terms of a transformation on a standard normal variable  $\mathcal{Z}$  (with mean 0 and variance 1). Let  $\mathcal{X}$  denote the Johnson curve random variable. Then for some function  $f$ ,

$$\mathcal{X} = \xi + \lambda\mathcal{Y},$$

where

$$\mathcal{Z} = \gamma + \delta f(\mathcal{Y}).$$

There are three types of functions  $f$  which are used:

1. The unbounded family,  $S_U$ ,

$$\begin{aligned} f(\mathcal{Y}) &= \log\{\mathcal{Y} + (1 + \mathcal{Y}^2)^{\frac{1}{2}}\} \\ &= \sinh^{-1} \mathcal{Y}, \quad -\infty < \mathcal{Y} < \infty. \end{aligned}$$

2. The lognormal family,  $S_L$ ,

$$f(\mathcal{Y}) = \log \mathcal{Y}, \quad 0 < \mathcal{Y} < \infty.$$

3. The bounded range family,  $S_B$ ,

$$f(\mathcal{Y}) = \log \frac{\mathcal{Y}}{1 - \mathcal{Y}}, \quad 0 < \mathcal{Y} < 1.$$

The method of moments can be used to fit a Johnson curve distribution to data (Hill et al., 1976). Then random variables from the Johnson curve distribution can be simulated (Hill, 1976).

## *Appendix 6: Random Number Generator Superduper*

Marsaglia (1976) has suggested a generator which overcomes the lattice structure defect of the usual mixed congruential generators. On account of its desirable attributes, this generator has become known as Superduper. Superduper is comprised of a congruential generator

$$X_n = 69069 X_{n-1} \text{ mod } 2^{32}$$

and another type of generator which is called a shift-register generator. The shift-register generator produces a pseudo-random sequence of bits (0s and 1s) using the recurrence

$$b_i = b_{i-17} + b_{i-32} \text{ mod } 2$$

Each consecutive 32 bits  $(b_1, \dots, b_{32}), (b_{33}, \dots, b_{64}), \dots$  forms a random integer

$$Y_n = \sum_{j=1}^{32} b_{32n+j} 2^{j-n}$$

Then  $X_n$  and  $Y_n$  are combined by adding the individual bits modulo 2 in the base 2 (or bit) representations of each one. The resulting integer is divided by  $2^{32}$  to produce a (0,1) variable.

The maximum period is about  $10^{18}$ . To get the maximum possible period simply choose the seed of the congruential generator,  $X_0 > 0$  and the seed of the shift-register generator,  $0 < Y_0 < 2048$ .

## *Appendix 7: Parsimony, Model Adequacy & Periodic Correlation*

### **7.1 Introduction**

The main purpose of this appendix is to discuss some general statistical principles which are elucidated by our recent work in Hipel & McLeod (1994) and McLeod (1993).

Also, based on case studies, a new diagnostic check for periodic correlation in the residuals of fitted ARMA models is developed. This diagnostic check is suitable for routine use when fitting seasonal ARMA models.



Briefly our experience with river flow time series suggests that the best forecasting results are obtained by following the general model building philosophy implicit in Box & Jenkins (1976) with suitable modifications and improvements. In general terms, this approach is iterative and advocates choosing the most parsimonious adequate statistical model. Two basic principles of special relevance are:

**PRINCIPLE 1: MODEL ADEQUACY.**

The model is considered adequate if it incorporates all relevant information and if when calibrated to the data, no important significant departures from the statistical assumptions made can be found.

**PRINCIPLE 2: MODEL PARSIMONY.**

The principle of parsimony means that the simplest possible model should be chosen.

One can view the problem of statistical modelling as choosing an adequate statistical model which is the most parsimonious. In mathematical programming terminology we could say that the problem of statistical modelling has an objective function which is to minimize the model complexity (Model Parsimony) subject to the constraint of Model Adequacy.

In §7.2 the results of a case study of forecasting monthly river flow time series is summarized. Here the importance of incorporating periodic correlation in the forecasting model is demonstrated. For a seasonal time series denoted by  $z_{r,p}$  where  $r$  denotes the year and  $p$  denotes the seasonal period, the periodic correlation coefficient is defined by

$$\rho_m(\ell) = \frac{\gamma_m(\ell)}{\sqrt{(\gamma_m(0)\gamma_{m-\ell}(0))}},$$

where

$$\gamma_m(\ell) = \text{cov}(z_{r,m}, z_{r,m-\ell}).$$

In some situations, as in the case study in §7.3, a comprehensive modelling approach which satisfies both adequacy and parsimony principles may not be practical either for reasons of expediency or because a suitable model cannot be found with available methodology. In this case, we have found combined forecasts to be useful. On the other hand, if a good model can be found, our experience suggests that the forecast cannot be significantly improved by combining it with forecasts from models which are less parsimonious or less adequate. This latter result is at variance with the results reported by Winkler & Makridakis (1983) and Newbold & Granger (1974). Perhaps this is due to the fact that the river flow time series used in our studies are generally longer and more homogeneous than the economic series used by the aforementioned authors. The skill of the modeller in developing an adequate model could also be a factor.

## 7.2 Monthly River Flow Case Study

The data in this case study consisted of thirty mean monthly river flows for periods of from 37 to 64 years. Various models and model selection procedures were used to calibrate a model to each data set omitting the last three years of data. The one-step ahead forecasts were then compared for the last three years (36 values). The best forecasts as judged by the root mean-square error and other criteria were obtained with the family of periodic autoregressive models.

The periodic autoregression model equation may be written

$$\phi_m(B)(Z_{r,m} - \mu_m) = a_{r,m} \quad (1)$$

where  $Z_{r,m}$  denotes the logarithmic flow for the  $r^{\text{th}}$  year and  $m^{\text{th}}$  month,  $\mu_m$  denotes the corresponding monthly mean,  $a_{r,m}, r = 1, 2, \dots, m = 1, 2, \dots, 12$  are a sequence independent normal random numbers with mean zero and variance,  $\sigma_m^2$ , and

$$\phi_m(B) = 1 - \phi_{m,1}B - \dots - \phi_{m,p_m}B^{p_m} \quad (2)$$

where  $B$  is the backshift operator on  $t$ , where  $t = 12(r - 1) + m$ . Several model selection techniques were used to select  $p_m$  ( $m = 1, 2, \dots, 12$ ). It was found that a periodic autoregression which was determined by choosing  $p_m$  as small as possible to achieve an adequate fit gave the best forecasts. This was accomplished by initially determining  $p_m$  based on a plot of the periodic partial autocorrelation function and then checking the adequacy of the fitted model. Our approach is thus a natural extension of that of Box & Jenkins (1970).

On the other hand, a subset periodic autoregression approach was found to produce comparatively very poor forecasts. In this approach, for each period all possible autoregressions with some parameters constrained to zero and with  $p_m = 12$  were examined ( $2^{12}$  possibilities) and the best model was selected with the Akaike Information Criterion (Akaike, 1974) as well as the Bayes Information Criterion (Schwarz, 1978). It was also noticed that the resulting models were always less parsimonious than that selected by the first approach.

The seasonal ARMA model developed by Box & Jenkins (1970, Ch. 9) did not perform very well either. In this case, the diagnostic check, developed in the next §7.4, indicates that this is due to model inadequacy.

The periodic autoregression and seasonal ARMA represent quite different families of time series models. Not only do the models differ in the correlation structure but the specification of seasonality is purely stochastic in the seasonal ARMA model and purely deterministic in the case of the periodic autoregression. Moreover neither specification is likely to be absolutely correct. Thus although the periodic autoregression model forecasted best and was considered to represent a more valid statistical model, it might be thought from the experience reported by Newbold & Granger (1974) and Winkler & Makridakis (1983) that combining the periodic autoregression and seasonal ARMA forecasts would be helpful. As shown in Hipel & McLeod (1994) this is not the case. In particular with method 1 of Winkler & Makridakis (1983, p. 152) the periodic autoregression forecast had a smaller mean square error at least 17 times out of 30. Thus combined forecasts cannot be recommended in this situation.

### 7.3 Quarter-Monthly River Flow Case Study

The object of this study was to obtain one-step-ahead forecasts of the quarter-monthly, i.e. almost weekly, inflows to the Lac St. Jean reservoir system operated by Alcan Limited. Complete time series on past quarter-month inflows, precipitation and snowmelt in the river basin were available for 30 years. A Box-Jenkins multiple transfer-function noise model with precipitation and snowmelt as inputs was found to provide an adequate fit to the deseasonalized data in many respects except that it did not account for the periodic correlation effect. A periodic autoregression model was also fit but this model did not take into account the covariates precipitation and snowmelt. It could be suggested that at this stage a periodic-transfer-function noise model should be developed to take into account both factors. However such a model could easily involve too many parameters and, in any case, it was not possible to calibrate it with our existing computer software. Perhaps future work will result in a suitable model. Finally, a third model which was a semi-theoretical hydrological model which incorporates various hydrological and meteorological information in a conceptual model of river flow. The conceptual modelling approach has been strongly advocated by certain hydrologists who feel that time series methods are too empirical.

All three models were calibrated on data for 27 years and then used to produce one-step-ahead forecasts over the next three years (144 periods). The root mean square error for transfer-function noise, periodic autoregression and conceptual model for forecasting logarithmic flows were respectively 0.2790, 0.3009 and 0.3894. When the forecasts were combined by simple averaging the root mean square error dropped to 0.1355. More sophisticated combination techniques were found to lead to even further improvements.

It is interesting to note that the empirical time series approach outperformed the more theoretical conceptual approach which has been strongly advocated by some hydrologists. A similar phenomenon with macro-economic time series forecasting as previously been found (Naylor et al., 1972).

### 7.4 A New Diagnostic Check For Periodic Autocorrelation

The seasonal ARMA model of order  $(p, d, q)(p_s, d_s, q_s)_s$  may be written

$$\Phi(B^s)\phi(B)\nabla_s^{d_s}\nabla^d Z_t = \Theta(B^s)\theta(B)a_t, \quad (3)$$

where  $Z_t$  is the observation at time  $t$  and  $a_t$  is a sequence of independent normal random variables with mean zero and variance  $\sigma^2$ . For monthly time series  $s = 12$  and  $t = 12(r - 1) + m$ , where  $r$  and  $m$  represent the year and month respectively. The polynomials  $\Phi(B^s)$ ,  $\phi(B)$ ,  $\Theta(B^s)$  and  $\theta(B)$  of degrees  $p_s$ ,  $p$ ,  $q_s$  and  $q$  specify the autoregressive and moving average components of the model. The terms  $\nabla_s = 1 - B^s$  and  $\nabla = 1 - B$  represent the seasonal and non-seasonal differencing operators. Using standard model selection techniques (Hipel & McLeod, 1994) it was found that most monthly river flow time series could be tentatively modelled as a seasonal ARMA model of order  $(p, 0, 1)(0, 1, 1)_{12}$ , where  $p = 0, 1$  or  $2$ . The diagnostic check described below can be used to check for model inadequacy due to periodic correlation in the residuals of such fitted models.

The residual periodic autocorrelation at lag  $k \geq 1$  may be written

$$\hat{r}_m(k) = \frac{\sum_r \hat{a}_{r,m} \hat{a}_{r,m-k}}{\sqrt{\sum_r \hat{a}_{r,m}^2 \sum_r \hat{a}_{r,m-k}^2}}, \tag{4}$$

where  $\hat{a}_{r,m}$  denotes the seasonal ARMA model residual for period  $t = 12(r - 1) + m$  ( $r = 1, \dots, N; m = 1, \dots, 12$ ). If the seasonal ARMA model is adequate then using the methodology in McLeod (1978) it can be shown for any fixed  $M \geq 1$ ,  $\sqrt{N}\hat{\mathbf{r}}^{(m)} = \sqrt{N}(\hat{r}_m(1), \dots, \hat{r}_m(M))$  is asymptotically normal with mean zero and covariance matrix  $(1_M - Q/12)/N$ , where  $1_M$  is the identity matrix of order  $M$  and  $Q = XI^{-1}X^T$ , where  $X$  and  $I$  are given in eq. (44) of McLeod (1978). Moreover,  $\sqrt{N}\mathbf{r}^{(m)}$  and  $\sqrt{N}\mathbf{r}^{(m')}$  are asymptotically independent when  $m \neq m'$ . Since the diagonal elements of  $Q$  are all less than one, it follows that to a good approximation,  $\hat{r}_m(1), m = 1, \dots, 12$  are jointly normally distributed with mean vector zero, diagonal covariance matrix and  $\text{var}(\hat{r}_m(1)) = N^{-1}$ . A diagnostic check for detecting periodic autocorrelation in seasonal ARMA model residuals is given by

$$S = N \sum_{m=1}^{12} \hat{r}_m^2(1) \tag{5}$$

which should be approximately  $\chi^2$ -distributed on 12 df.

As a check on the asymptotic approximation involved, a brief simulation experiment was performed. A  $(1, 0, 0)(0, 0, 0)_{12}$  model with  $\phi_1 = -0.9, -0.6, -0.3, 0.3, 0.6$  and  $0.9$  was simulated. Table 1 summarizes the results on  $S$  for one thousand simulations with  $N = 17$ . The empirical significance level of a nominal 5% test was estimated by counting the number of times that  $S$  exceeded 21.0261. From Table 1, the approximation is seen to be adequate for practical purposes. In further experiments with  $N = 34$  and  $68$ , the approximation was seen to improve although the empirical significance level was still slightly less than 0.05 in all cases. This suggests that in general the significance will be slightly overestimated. For example, if the observed value of  $S$  indicates significant periodic correlation at the 5% level, the true significance level will be slightly less than 5%. The data on the Saugeen River (1915–1976) is illustrative of the usefulness of this new diagnostic check. A  $(1, 0, 1)(0, 1, 1)_{12}$  model was fit to the logarithmic flows and passed all diagnostic checks given in Box & Jenkins (1976). However, it was found that  $S = 59.6$  indicating very significant residual periodic correlation. As indicated in the next section, it appears that many seasonal economic time series also exhibit such periodic residual correlations.

**Table 1**

*Empirical mean, variance and significance level of  $S$  with  $N = 17$  in 1000 simulations using a nominal 5% test.*

$\phi_1$	Mean	Variance	Significance level
-0.9	11.9	19.3	0.032
-0.6	11.5	18.6	0.025
-0.3	11.3	19.0	0.023
0.0	10.9	16.6	0.016
0.3	11.1	19.7	0.027
0.6	11.5	18.6	0.026
0.9	11.7	19.7	0.030

### 7.5 Application to Forecasting Economic Time Series

Many seasonal economic time series may exhibit periodic correlation which most of the standard approaches do not take into account. The diagnostic check of §4 may be applied routinely when fitting seasonal ARMA models. Table 2 shows the results of testing the seasonal ARMA models fitted by Miller & Wichern (1977, p.432) to four Wisconsin series. It is seen that in two out of the four series there is very significant periodic correlation. In these cases, models which take this correlation into account may be expected to produce improved forecasts.

**Table 2**

*Diagnostic Test For Residual Periodic Correlation  
For Four Wisconsin Series From Miller & Wichern*

Category	$S$	$d.f.$	Significance level
Food Products	25.36	12	0.013
Fabricated Metals	36.8	12	0.0002
Transportation Equipment	11.6	12	0.478
Trade	6.98	12	0.859

*REFERENCES*

- Ableson, R.P. & Tukey, J.W. (1963), "Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order", *The Annals of Mathematical Statistics*, Vol. 34, pp.1347–1369.
- Akaike, H. (1979), "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting", *Biometrika*, Vol.66, pp.237–242.
- Ansley, C.F. (1979), "An algorithm for the exact likelihood of a mixed autoregressive moving-average process", *Biometrika*, Vol.66, pp.59–65.
- Bartlett, M.S. (1947), "The use of transformations", *Biometrics*, Vol. 3, No.1, pp.37–52.
- Bartlett, M.S. (1955), *An Introduction to Stochastic Processes*, Cambridge University Press.
- Blackman, R.B. & Tukey, J.W. (1959). *The measurement of power spectra*, New York: Dover.
- Bell, W.R. & Hillmer, S.C. (1984), "Modeling time series with calendar variation", *Journal of the American Statistical Association*, Vol.78, pp.526–534.
- Bloomfield, P. (1976), *Fourier Analysis of Time Series: An Introduction*, New York: Wiley.
- Box, G.E.P. & Cox, D.R. (1964), "An analysis of transformations", *Journal of the Royal Statistical Society Series B (Methodological)*, Vol.26, pp.211–252.
- Box, G.E.P. & Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*, Expanded Edition, San Francisco: Holden-Day.
- Brillinger, D.R. (1989), "Consistent detection of a monotonic trend superimposed on a stationary time series", *Biometrika*, Vol.76, pp.23–30.
- D'Agnostino, R.B. (1970), "Transformations to normality of the null distribution of  $g_1$ ", *Biometrika*, Vol.57, pp.679–680.
- Chatfield, C. & Prothero, D.L. (1973), "Box-Jenkins seasonal forecasting: Problems in a case-study", *Journal of the Royal Statistical Society, Series A*, Vol. 136, pp. 295–308.
- Cleveland, W.S. (1979), "Robust locally weighted regression and smoothing scatter-plots", *Journal of the American Statistical Association*, V. 74, pp. 829–836.
- Cleveland, W.S. & Terpenning, I.J. (1984), "Graphical methods for seasonal adjustment", *Journal of the American Statistical Association*, V. 77, pp.52–62.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E. & Terpenning, I. (1991), "STL: A seasonal-trend decomposition procedure based on loess", *Journal of Official Statistics*, Vol. 6, pp.3–73.
- Davies, N, Pate, M.B, & Frost, M.G. (1974), "Maximum autocorrelations for moving-average processes", *Biometrika*, Vol.66, pp.59–65.

- Dickey, D.A. & Fuller, W.A. (1979), "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, Vol. 74, pp.427–431.
- Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- Fisher, R.A. (1929), "Tests of significance in harmonic analysis", *Proceedings of the Royal Society, Series A*, V.125, pp.54–59.
- Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York: Wiley.
- Gadd, A. & Wold, H. (1964), "The Janus quotient: a measure for the accuracy of prediction". In H. Wold (ed.) *Econometric Model Building*. Amsterdam: North Holland.
- Gladyšev, E.G. (1961), "Periodically correlated random sequences", *Soviet Math. Dokl.*, Vol. 2, 385–388.
- Granger, C.W.J. (1963), "A quick test for serial correlation suitable for use with non-stationary time series", *Journal of the American Statistical Association*, V.58, pp.728–736.
- Granger, C.W.J. and Hughes, A.O. (1971), A new look at some old data. *Journal of the Royal Statistical Society, A*, 134, 413–428
- Hamming, R.W. (1977), *Digital Filters*, New Jersey: Prentice-Hall.
- Herglotz, G. (1911), "Über potenzreihen mit positivem reellem teil im einheitskreis", *Sitzgaber. Sachs Akad. Wiss.*, Vol.63, pp.501–511.
- Hipel, K.W. & McLeod, A.I. (1977), "Advances in Box-Jenkins modelling", *Water Resources Research*, V.13, pp.567–586.
- Hipel, K.W. & McLeod, A.I. (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam: Elsevier. ISBN: 0-444-89270-2
- Hipel, K.W., McLeod, A.I. & Fosu, P. (1983), "Empirical power comparisons of some tests for trend", in *Statistical Aspects of Water Quality Monitoring, Developments in Water Science*, V.27, pp.347–362, Edited by A.H. El-Shaarawi and R.E. Kwiatkowski.
- Hill, I.D., Hill, R. & Holder, R.L. (1976), "Algorithm AS 99. Fitting Johnson curves by moments", *Applied Statistics*, V.25, pp.180–189.
- Hill, I.D. (1976), "Algorithm AS 100. Normal-Johnson and Johnson-Normal Transformations", *Applied Statistics*, V.25, pp.190–192.
- Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982), "Techniques for trend assessment for monthly water quality data", *Water Resources Research*, Vol. 18, pp.107–121.
- Hosking, J.R.M. (1983), "Lagrange multiplier test", *Encyclopedia of Statistical Sciences*, Vol. 4, pp. 456–459.



- Hosking, J.R.M. & Ravishanker, N. (1993), "Approximate simultaneous significance intervals for residual autocorrelations of autoregressive moving-average time series models", *Journal of Time Series Analysis*, V.14, pp.19–26.
- Hurvich, C. M. (1988), "A mean squared error criterion for time series data windows", *Biometrika* 75, 485–490.
- Kendall, M.G. & Stuart, A. (1968), *The Advanced Theory of Statistics*, Volume 3, 2nd Ed.
- Kheoh, T.S. & McLeod, A.I. (1992), "Comparison of Modified Portmanteau Tests", *Computational Statistics and Data Analysis*, V.14, pp.99–106.
- Kendall, M.G. (1970), *Rank Correlation Methods*, 2nd Ed., New York: Hafner.
- Lawrance, A.J. (1991), "Directionality and reversibility in time series", *International Statistical Review*, Vol.59, pp.67–79.
- Lawrance, A.J. & Lewis, P.A.W. (1985), "Modelling and residual analysis of nonlinear autoregressive processes", *Journal of the Royal Statistical Society B* Vol. 47 No. 2, pp.165–202.
- Lawrance, A.J. & Lewis, P.A.W. (1992), "Reversed residuals in autoregressive time series analysis", *Journal of Time Series Analysis*, Vol. 13, No. 3, pp.253–266.
- Lehmann, E.L. (1959). *Testing Statistical Hypotheses*, New York: Wiley.
- Li, W.K. & McLeod, A.I. (1981), "Distribution of residual autocorrelations in multivariate ARMA time series models", *Journal of the Royal Statistical Society, Series B*, V.43, pp.231–239.
- Mann, H.B. (1945), "Nonparametric tests against trend", *Econometrica*, Vol. 13, 245–259.
- Marsaglia, G. (1976), "Random Number Generation". In *Encyclopedia of Computer Science*, ed. A. Ralson, pp. 1192–1197. New York: Petrocelli and Charter.
- Mackisack, M.S. and Poskitt, D.S. (1979), "Autoregressive frequency estimation", *Biometrika*, Vol.76, pp.565–575.
- McLeod, A.I. (1977), "Improved Box-Jenkins estimators", *Biometrika*, V.64, pp.531–534.
- McLeod, A.I. (1975), "Derivation of the theoretical autocorrelation function of autoregressive moving-average time series", *Applied Statistics*, V.24, pp.255–256.
- McLeod, A.I. (1978), "On the distribution of residual autocorrelations in Box-Jenkins", *Journal of the Royal Statistical Society, Series B*, V.40, pp.296–302.
- McLeod, A.I. (1985), "AS R58. Remark on AS 183", *Applied Statistics*, V.34, No. 2, pp.198–200.
- McLeod, A.I., (1993), "Parsimony, Model Adequacy and Periodic Correlation in Forecasting Time Series", *International Statistical Review*, Vol. 61, No. 3, pp.387–393.

- McLeod, A.I. (1994) "Diagnostic Checking Periodic Autoregression Models with Application", *The Journal of Time Series Analysis*, Vol. 15, No. 2, pp.221-233.
- McLeod, A.I. & Li, W.K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis*, V.4, pp.269-273.
- McLeod, A.I., Hipel, K.W. & Bodo, B.A. (1990), "Trend analysis methodology for water quality time series", *Environmetrics*, Vol. 2, pp.169-200.
- McLeod, A.I. & Hipel, K.W. (1995), "Exploratory spectral analysis of hydrological time series", *Stochastic Hydrology and Hydraulics*, No.2.
- McLeod, A.I. & Sales, P.R.H. (1983), "An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models", *Applied Statistics*, V.32, pp.211-223.
- Michael, J.R. (1983), "The stabilized probability plot", *Biometrika*, Vol. 70, 11-17.
- Miller, R. B. & Wichern, D. W. (1977). *Intermediate Business Statistics*. New York: Holt, Reinhart and Winston.
- Naylor, T. II, Seaks, T.G. & Wichern, D.W. (1972). Box-Jenkins methods: An alternative to econometric models. *Int. Statist Rev.* **40**, 123-137.
- Newbold, P. & Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *J. R. Statist. Soc. A* **137**, 131-165.
- Percival, D.B. & Walden, A.T. (1993). *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge University Press.
- Ozaki, T. (1991), Contribution to the discussion of the paper "STL: A seasonal-trend decomposition procedure based on loess", by R. B. Cleveland, W. S. Cleveland, J. E. McRae & I. Terpenning, *Journal of Official Statistics*, Vol. 6, pp.3-73.
- Pagano, M. (1978). On periodic and multiple autoregressions. *Ann. Statist.* **6**, 1310-1317.
- Pagano, M. (1972), "An algorithm for fitting autoregressive schemes", *Applied Statistics*, V.21, pp.274-281.
- Priestley, M.B. (1981), "Spectral Analysis and Time Series", New York: Academic Press.
- Sakai, H. (1982). Circular lattice filtering using Pagano's method. *IEEE Trans. Acoust. Speech Signal Process.* **30**, 279-287.
- Schuster, A. (1898), "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena", *Terr. Magn.*, Vol.3, pp.13-41.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.
- Ramsey, F.L. (1988), "The slug trace", *The American Statistician*, Vol. 42, No. 4, p.290.

- Robinson, P.M. (1994), "Efficient tests of the nonstationary hypotheses", *Journal of the American Statistical Association* 89, 1420–1437.
- Royston (1982), "The  $W$  test for normality. Algorithm AS 181" *Applied Statistics*, V.31, pp.176–224.
- Royston (1993), "Graphical detection of non-normality by using Michael's statistic", *Applied Statistics*, V.42, No.1, pp.153–158.
- Said, E.S. & Dickey, D.A. (1984), "Testing for unit roots in autoregressive-moving average models of unknown order", *Biometrika*, 71, pp.599–607.
- Sakai, H. (1982) "Circular lattice filtering using Pagano's method", *IEEE, Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-30, No.2, pp.279–287
- Swed, F.S. & Eisenhart, C. (1943), "Tables for testing randomness of grouping in a sequence of alternatives", *Annals of Mathematical Statistics*, Vol.14, pp.66–87.
- Tukey, J. W. (1967), "An introduction to the calculations of numerical spectrum analysis" in *Advanced Seminar on Spectral Analysis of Time Series*, edited by B. Harris, pp.25–46, Wiley, New York.
- Tukey, J.W. (1977), "Exploratory Data Analysis", Addison-Wesley.
- Valz, P.D. & McLeod, A.I. (1990), "A simplified derivation of the variance of Kendall's tau.", *The American Statistician*, V.44, pp.39–40.
- Valz, P., McLeod, A. I. and Thompson, M. E., (1994a, to appear) "Cumulant generating function and tail probability approximations for Kendall's score with tied rankings", *Annals of Statistics*.
- Velleman, P.F. & Hoaglin, D.C. (1981), "ABC of Exploratory Data Analysis", Duxbury Press.
- Wichmann, B.A. & Hill, I.D. (1982), "Algorithm AS 183. An efficient and portable pseudo-random number generator", *Applied Statistics*, V.21, pp.188–190.
- Winkler, R. L. & Makridakis, S. (1983). The combination of forecasts. *J. R. Statist. Soc. A* **146**, 150–157.

# TABLE OF CONTENTS

Introduction	1
Prologue	2
Getting Started	2
MTS Directories	2
Hardware Requirements	2
Installation	5
Running MTS Under Windows	6
Printer Configuration	6
Menu Selection	6
Data Input	7
Graphics Hardcopy	8
Hardcopy Text Output	8
Data Transformations	9
Current Limitations	10
Trouble Shooting	10
Exploratory Analysis	12
Trace Plot	13
Bivariate Trace Plot	14
Bivariate Slug-Trace Plot	14
Time Series Scatter Plot	14
Periodic Time Series Scatter Plot	15
Bivariate Time Series Scatter Plot	15
Seasonal Subseries Plot	15
Normal Probability Plot	16
Trend Analysis	19
Linear or Ratio Filter	20
Tukey Smooth	21
Cleveland Robust Loess Smoother	23
Runs Test	24
Nonseasonal Trend Tests	25
Seasonal Trend Test	26
Brillinger Trend Test	27
Seasonal-Trend Loess Decomposition	29
View File In STL Directory	33
Correlation Functions	35
Sample Autocorrelation Function	36
Sample Partial Autocorrelation Function	37
Sample Inverse Autocorrelation Function	37
Sample Inverse Partial Autocorrelation Function	38
Sample Cross Correlation Function	39
Sample Periodic Autocorrelation Function	40
Sample Periodic Partial Autocorrelation Function	41

Theoretical Autocorrelation Function .....	41
Theoretical Partial Autocorrelation Function .....	42
Impulse Response or Inverted Form Coefficients .....	42
Spectral Analysis .....	45
Cumulative Periodogram Analysis .....	46
Fisher Periodicity Test .....	46
Periodogram Smoothing .....	47
Autoregressive Spectral Density Estimation .....	51
Theoretical Spectral Density of ARMA Process .....	52
Model Estimation .....	53
ARIMA Estimation .....	54
Transfer-Function Noise Estimation & Intervention Analysis .....	61
Periodic AR Estimation .....	68
Diagnostic Checks .....	74
Unit Root Test .....	72
Residual Autocorrelation Analysis of Fitted ARIMA Model .....	73
Theoretical Autocorrelation Function Of Fitted ARMA Model .....	74
Theoretical Partial Autocorrelation Function Of Fitted ARMA Model .....	74
Theoretical Spectral Density Function Of Fitted ARMA Model .....	74
Theoretical Impulse-Response Function Of Fitted ARIMA Model .....	75
Theoretical Inverted Form Coefficients Of Fitted ARMA Model .....	75
Spread-Level Diagnostic Check .....	76
Likelihood Ratio Test .....	76
View File In OUTPUT Directory .....	77
Forecasting .....	78
Forecast Fitted ARIMA Model .....	79
Comparison Of Forecast and Actuality .....	81
Forecast A Fitted Periodic Autoregression .....	82
Simulation .....	83
Simulate Fitted ARIMA Model .....	85
Simulate ARIMA Model (parameters input) .....	84
Filters .....	86
Command Editor .....	87
Aggregate Time Series .....	87
ARIMA Filter .....	87
Deseasonalize Trend-Free Time Series .....	88
Algebraic Combination Of Two Time Series .....	90
Difference Or Integrate A Time Series .....	91
Box-Cox Or Inverse Box Cox Transformation .....	92
Utilities .....	94
Screen Editor .....	95
Data Documentation Editor & Viewer .....	95
Describe DATA Directory .....	96
View or Edit DATA Directory SUMMARY .....	96

Set New DATA Directory .....	96
View File in OUTPUT Directory .....	96
Delete, Copy or Move Files .....	97
Convert From Column Format to Row Format .....	97
Convert From Row Format to Column Format .....	97
Appendices .....	99
Appendix 1: Robust Loess Regression .....	100
Appendix 2: Box-Cox Transformations .....	102
Appendix 3: Simultaneous Significance Intervals .....	111
Appendix 4: Spectral Analysis Primer .....	106
Appendix 5: Johnson Curve Distributions .....	110
Appendix 6: Random Number Generator Superduper .....	111
Appendix 7: Parsimony, Model Adequacy & Periodic Correlation .....	
References .....	117