EXERCISES IN FORECASTING II

Differencing to Achieve Stationarity

The first step, which is taken before attempting to identify an autoregressive moving-average (ARMA) model which might be fitted to a data series, is to examine the time plot of the data and to determine whether or not the series could be have been generated by a stationary process. If a trend is evident in the data, then it must be removed. A variety of techniques of trend removal, which include the fitting of parametric curves and of spline functions, have been discussed in previous lectures. When such a function is fitted, it is to the sequence of residuals that the ARMA model is applied.

However, many empirical series can be modelled adequately by supposing that some suitable difference of the process is stationary. Thus the process generating the observed series $y(t)$ might be modelled by the ARIMA(p, d, q) equation

$$\alpha(L)\nabla^d y(t) = \mu(L)\varepsilon(t), \tag{1}$$

wherein $\nabla^d = (I - L)^d$ is the dth power of the difference operator. In that case, the differenced series $z(t) = \nabla^d y(t)$ will be described by a stationary ARMA(p, q) model. The inverse operator ∇^{-1} is the summing or integrating operator, which accounts for the fact that the model depicted by equation (1) is described an autoregressive integrated moving-average (ARIMA) model.

To determine whether stationarity has been achieved, either by trend removal or by differencing, one can examine the autocorrelation sequence of the residual or processed series. The sequence corresponding to a stationary process should converge quite rapidly to zero as the value of the lag increases. An empirical autocorrelation function which exhibits a smooth pattern of significant values at high lags indicates a nonstationary series.

An example is provided by Figure 1 where a comparison is made between the autocorrelation function of the original series and that of its differences. Although the original series does not appear to embody a systematic trend, it does drift in a haphazard manner which suggests a random walk; and it is appropriate to apply the difference operator.

Once the degree of differencing has been determined, the autoregressive and moving-average orders are selected by examining the sample autocorrelations and sample partial autocorrelations. The characteristics of pure autoregressive and pure moving-average process are easily spotted. Those of a mixed autoregressive moving-average model are not so easily unravelled.

As an exercise, you are invited to wander through the data sets which accompany *MESOSAUR* in pursuit of series which can be reduced to stationarity by the application of one or more differencing operations.

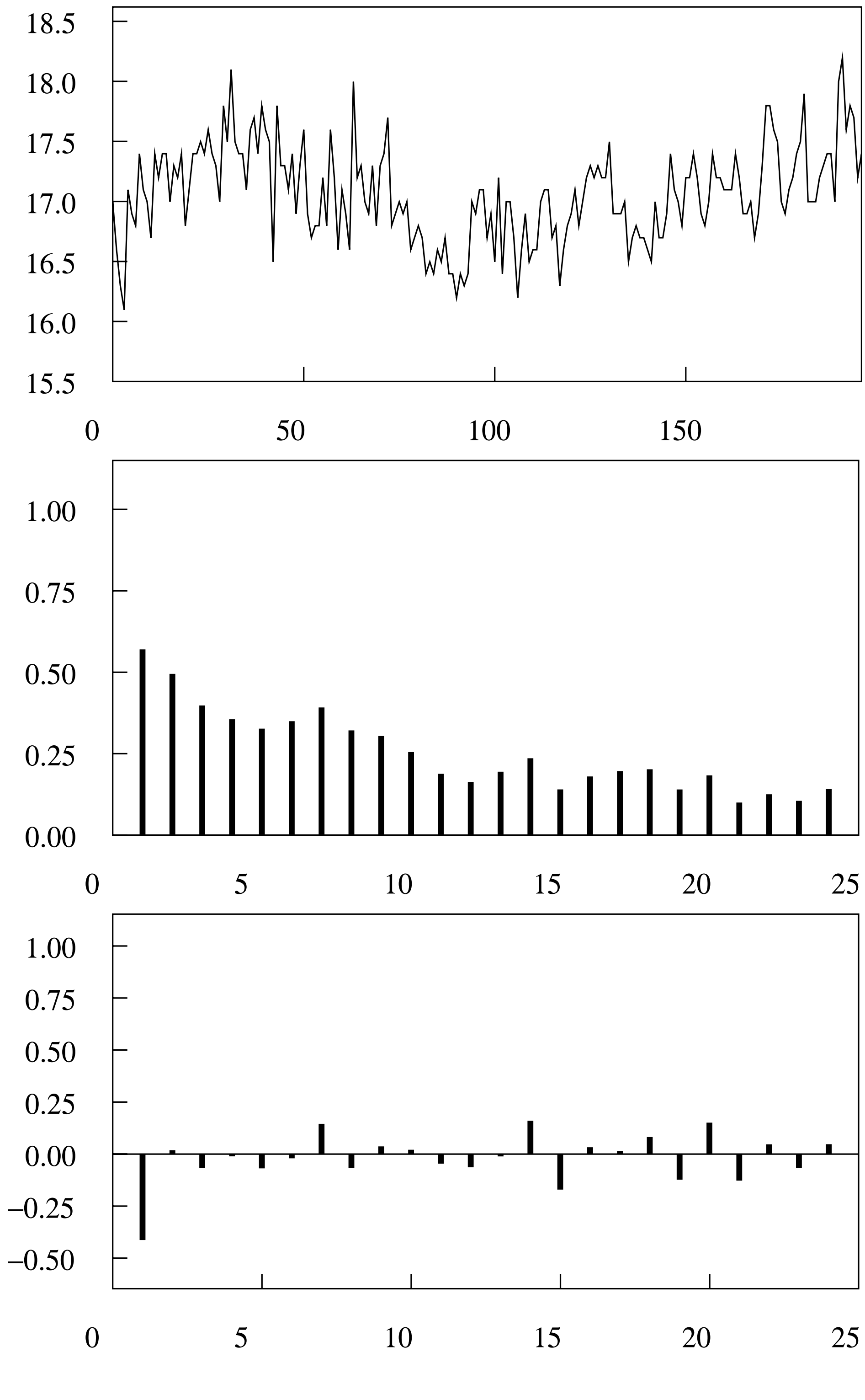

Figure 1. The concentration readings from a chemical process with the autocorrelation function and the autocorrelation function of the differences.

It should be of some interest to discover whether the financial series which will be found amongst the data can be reduced to stationarity by simple differencing. An crude version of the efficient-markets hypothesis asserts that such series follow random walks; in which case their first differences should be white noise.

The theoretical autocovariance function $\{\gamma_\tau\}$ of a white-noise process has $\gamma_\tau = 0$ for all $\tau > 0$. The series to be examined include `IBMSTOCK` and `ICISTOCK`. There are also numerous exchange-rate series to be found amongst the `DEMO` data under `J:\APPS\MESO` which might be examined.

The Removal of Seasonality by Annual Differencing

Economic time series often manifest seasonal patterns which are of such regularity that it is difficult to accept the the idea that they can be modelled as ARIMA processes. The `AIRPASS` data is typical of such seasonal series. Further examples are provided by `APPAREL`, `B&JSALES` etc. You should search the data for others.

A seasonal pattern is liable to strike the eye quite forcibly. However, seasonal fluctuations also manifest themselves strongly in the sequence of autocovariances and in the periodogram. You should gain a clear idea of what these effects consist of.

The typical ARIMA model for a seasonal series, which was proposed by Boy and Jenkins, takes the form of

$$\alpha(L)\phi(L^s)\nabla_s y(t) = \mu(L)\theta(L^s)\varepsilon(t) \tag{2}$$

wherein ∇_s is the seasonal differencing operator. This becomes $\nabla_4 = 1 - L^4$ for quarterly data and $\nabla_{12} = 1 - L^{12}$ for monthly data. The model suggests that the seasonality in a series which is otherwise stationary can be eliminated by a process of seasonal differencing. You are invited to investigate the efficacy of this method of de-seasonalisation. You should examine the autocovariances and the periodogram of the series in question both before and after taking differences.

A de-seaonalising filter may be at fault for removing too little of the seasonal component or for removing too much besides the component. The seasonal differencing operators have the second of these faults. At a later stage in the course, we shall be proposing some de-seasonalising filters of a more sophisticated nature.

In some cases, it is advisable to remove a linear trend before attempting the de-seasonalisation. In the case of `AIRPASS`, you should begin by taking logarithms of the series in order to overcome the tendency for its dispersion about the trend to increase with time. Then you should remove the trend by whatever means strikes you as appropriate.

The Removal of Seasonality via Trigonometrical Regressions

A model of the sort depicted in (2) is capable of generating quite regular cycles. However, in the long run, such cycles are unbounded in amplitude and their phase shows no central tendency. This suggests that it might be more appropriate to model the seasonality in an economic series by using sinusoidal functions which are guaranteed to remain in phase with the annual calender.

You will find the appropriate trigonometrical series, which are generated by sine and cosine function of various frequencies, within the files `QUATRO.TXT` and `DUODEC.TXT`. These series may be used as explanatory variables in a multiple regression in which the seasonal data series is the dependent variable. You should consult the chapter on “Seasons and Cycles” in the yellow book for a fuller explanation of this procedure.

You might also investigate the effects of applying the seasonal differencing operators to the trigonometrical series.

The Multiple Regression Facility within MESOSAUR

You may prefer to perform a regression analysis within *EXCEL* or *MICROFIT*. However, there is a perfectly serviceable facility for multiple regression within *MESOSAUR* itself. The facility if to be found in the **Multivariate Models** submenu within the **Models** menu. The **Regression** command allows you to compute a regression model comprising up to 15 variables. You must first mark the variables to be included in the model.

To mark one of the variables which is displayed in a legend across the screen, you highlight it with the cursor keys and then you press `<Enter>`. To unmark it, you select it and press enter again. You may select all of the variables displayed in legends by typing `<Alt-M>`. To unmark all of the marked variables, you type `<Alt-U>`.

To perform a regression analysis, you first mark the independent variables. Then you select the dependent variable. The dependent variable is shown at the top of the screen. To change the dependent variable, you must press `<Esc>` and repeat the process. When the appropriate variables have been selected, press `<Insert>` to obtain the local menu and type `<G>` for **Go** to run the regression.