REVUE FRANÇAISE D'AUTOMATIQUE, D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

O. D. ANDERSON

The Box-Jenkins approach to time series analysis

Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle, tome 11, n° 1 (1977), p. 3-29.

<http://www.numdam.org/item?id=RO_1977__11_1_3_0>

R.A.I.R.O. Recherche opérationnelle/Operations research
(vol. 11, n° 1, février 1977, p. 3 à 29)

THE BOX-JENKINS APPROACH TO TIME SERIES ANALYSIS (*) (1)

O. D. ANDERSON (2)

Summary. — *The Box-Jenkins approach to time series analysis and forecasting is currently a subject of major interest. At present, its successful application requires considerable skill from the practitioner. However, the potential gains of the method over other established, but less sophisticated extrapolation procedures, make it imperative that all workers concerned with time-series should have an appreciation of the approach.*

This paper is designed for the numerate reader, with some knowledge of statistics; and it will be of greatest value to the analyst already wrestling with time series by other means. To the generally interested reader, this account will be self-contained; but, for the specialist, it will serve as a relatively simple introduction to this increasingly important methodology, and the references cited will indicate how he can pursue the matter further.

1. INTRODUCTION TO TIME SERIES

A *time series* is a set of observations ordered in some dimension, usually time. We will only consider *discrete* series with observations y_t taken at various (relatively precise) instants. These instants are chosen at equispaced intervals; so, if we make n observations, we can consider them taken at times $t = 1, 2, \ldots, n$—just by suitably choosing the unit of time and the starting point. For instance, according to an H.M.S.O. [13] publication, the numbers of women unemployed in the United Kingdom on the first of each month, from January 1967 to July 1972, are as given in the appendix and shown in figure 1.1. (The data points are joined up by straight lines which help the eye to follow the development through time, especially for more volatile series histories.) There are 67 observations, so we say that the series has *length* $n = 67$.

The Women Unemployed data is an example of a *sampled* series. The H.M.S.O. publication has chosen to record the numbers on the first of each month, but of course there are unemployed women at other times. Alternatively, discrete time series can be obtained by accumulating a quantity for a period of time. For instance, production figures are examples of *accumulated* series. We do not speak of the production of say paraffin on the last day of the month, but for the whole of that month. In economics, these two types of series are usually referred to as "stock" and "flow",

(*) Reçu mai 1975.
(1) This paper was first presented at the University of Surrey Seminar in Economics, 25th February 1975.
(2) Statistician, Division of Statistics and Operational Research, Civil Service College, London.

respectively. The various Index Numbers provide good examples of them. Thus, the *Wholesale Price Index*, constructed by the Department of Industry, gives a sampled time series; whilst the *Index of Industrial Production*, prepared by the Central Statistical Office, provides an accumulated series.

Some readers will have noticed that the Women Unemployed series does not have a strictly constant sampling interval – months are not all of the

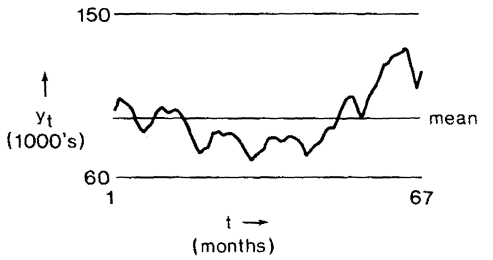

Figure 1.1

Women unemployed (1000's) in UK on 1st of each month, January 1967-July 1972.

same duration. For a sampled series, this does not usually matter very much; but, for an accumulated one, it is more serious. Especially so when it is not the calendar month that needs to be considered; but the effective working month, which depends on how many week-ends and public holidays it contains. Adjustments are consequently often made to flow series, though it is frequently difficult to draw the line – for instance, should strikes be included, and what about a work-to-rule?

The basic time series property

Most statistical methodology is concerned with independent sets of observations. A lack of independence is usually considered highly undesirable, and one of the objects of good experimentation is to eliminate dependence. However, with time series analysis, we are concerned with data which develops through time; and where, in general, each observation depends on earlier observations. It is, in fact, this dependence which is of interest. Contrast this with regression analysis, where a fundamental assumption is that the observations are independent of each other.

Thus a time series behaves as if it possesses a "memory". Again, the series to date to some extent determines its future values; so it also contains a certain degree of "foresight". Unlike many previous approaches to time series analysis, the Box-Jenkins method uses this fundamental time series property virtually to the full.

The purpose of time series analysis

Before we look at the mechanics of the subject, let us discuss why we wish to analyse time series. There are three practical interrelated reasons for doing so.

First, we might want to make inferences about the statistical structure which gave rise to a particular series history. Should the series look like figure 1.2, there are no prizes.

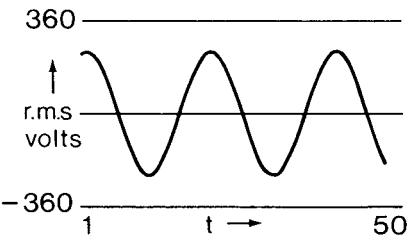

Figure 1.2

A.C. Single Phase Voltage across Household Mains, at intervals of .001 s.

Such a series is fully deterministic (apart from negligible disturbances) and requires no further attention from the statistician. However, in economics, series are for ever fickle; and any patterns, or *trends*, will be buried in irregularity. It is an object of time series analysis to obtain as full an explanation as possible for the series, by building as satisfactory a model as is possible. (Provided the situation merits the effort.)

Such a *structural* model may give an indication of the "physical" mechanism which generated the series, and so increase our theoretical understanding in the particular area.

However, even if this does not happen, the inferred structure, as represented by the model, can be used to forecast future values of the series—with, hopefully, realistic stated degrees of uncertainty. For, other things being equal, the same sort of dependence, observed in the past, can be expected to continue into the future. This is why the interdependence of the elements of a time series is of prime interest. Standing at the present, at the time n = now, we study the past in order to get a glimpse, inevitably distorted, of the future. This is the second object of time series analysis.

Having obtained a forecast, in a given situation, the possibility of altering some of the conditions presents itself. This, the third object of analysis, is *control*—and, in Government, it is usually the prime purpose. For instance, consider a balance of payments series. This is analysed and a model built. There is little interest in the model *per se*, but forecasts made from it may

be alarming. Then Government will try to avoid these future values by appropriate action.

It is an unfortunate fact of life, that though it is relatively easy to provide a plausible explanation of a series to date, it is much more difficult to forecast effectively. It is conjectured that, with the type of time series Government encounters, control is even more difficult. "Conjectured", because who can tell whether the control ever had the desired effect? If no action is taken, one can observe later whether the forecast was close or not; but, if control is attempted and a poor result occurs, is this due to poor control or poor forecasting? However as with all statistics, just because you can never succeed, does not mean you should not try – usually one can do better than merely guess.

At present there is a considerable gap between what government and business analysts appear to be doing in time series, and some of the possibilities now available. This, of course, is the familiar tale of how practice trails behind theory, and no doubt will be remedied in the next few years. Indeed there are indications that modern methods are being tried, but in some areas the old may prove more practicable for a long time yet. However, in this paper, we are going to discuss the recent and powerful, but not simply manipulated, tool now available to the time series analyst – the Box-Jenkins approach.

2. SIMPLE BOX-JENKINS MODELS

The theory of Professors Box and Jenkins' approach to discrete time series analysis, incorporating an iterative cycle of Identification, Estimation and Verification, has been fully discussed in their book [9]. Anderson [5] gives a more concise account, including recent research. The approach deals with what is termed the *time domain*, where an observation y_t is related to previous y_{t-j}, $j > 0$. This is in contrast to complementary methods, based on harmonic analysis, which treat the frequency domain. In this section, we will give a brief description of the simplest Box-Jenkins models.

Time processes

A *time process* is a sequence of random variables $\{ Y_t \}$, which are not generally independent, but serially correlated. We shall interest ourselves in *linear* processes of the form

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + A_t + \theta_1 A_{t-1} + \ldots + \theta_q A_{t-q} \quad (2.1)$$

satisfying the Gaussian assumption, which is that $\{ A_t \}$ is a structureless process of independent zero-mean normal random variables with constant variance σ_A^2, called a *white noise* process. Without loss of generality, we can choose p and q sufficiently small so that φ_p, $\theta_q \neq 0$.

Introducing the backshift operator B, with the property that

$$BZ_i = Z_{i-1}$$

for any process $\{ Z_t \}$ and any time i, (2.1) can be written

$$\left(1 - \sum_{j=1}^{p} \varphi_j B^j\right) Y_t = \left(1 + \sum_{j=1}^{q} \theta_j B^j\right) A_t$$

or, using an obvious notation,

$$\varphi_p(B)\, Y_t = \theta_q(B)\, A_t. \tag{2.2}$$

When $q = 0$, (2.2) is termed a pth order *Auto-Regressive* process, or AR (p); while if $p = 0$, it is a *Moving Average* process of order q, or MA (q). For general p and q, (2.2) is an ARMA (p, q) process, which is called *proper* when it does not degenerate to either AR (p) or MA (q). Note, that when the process is of form (2.2), say, we will use the notation

$$\{ Y_t \} \sim \text{ARMA}(p, q).$$

Simple examples are the AR (1), MA (1) and ARMA (1,1) processes of respective form

$$Y_t = \varphi\, Y_{t-1} + A_t, \tag{2.3}$$

$$Y_t = A_t + \theta\, A_{t-1}, \tag{2.4}$$

$$Y_t = \varphi\, Y_{t-1} + A_t + \theta\, A_{t-1} \tag{2.5}$$

which can be alternatively written as, respectively,

$$(1 - \varphi B)\, Y_t = A_t,$$

$$Y_t = (1 + \theta B)\, A_t,$$

$$(1 - \varphi B)\, Y_t = (1 + \theta B)\, A_t.$$

The process (2.2) is *stationary* if $\varphi_p(\zeta)$, a polynomial in the complex variable ζ, has all its zeros outside the unit circle. It is *invertible* if a similar condition holds for $\theta_q(\zeta)$. We will also allow zeros of $\theta_q(\zeta)$ to lie on the unit circle, giving marginal non-invertibility.

For a stationary process, by definition $\varphi_p(1) \neq 0$; so, taking expectations in (2.1), gives

$$E[Y_t] = 0$$

and, for all integers k, defining the *autocovariance* γ_k at *lag* k by Cov $[Y_t, Y_{t-k}]$, we have

$$\gamma_k = E[Y_t\, Y_{t-k}] = \gamma_{-k}.$$

In particular $\gamma_0 = \sigma_Y^2$. (Should $E[Y_t] = \mu_Y \neq 0$ then, instead of working with $\{ Y_t \}$, we work with the mean-corrected process $\{ \tilde{Y}_t = Y_t - \mu_Y \}$.)

A general process is fully described by

$$\underset{\sim}{\Psi} = (\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q)$$

and σ_A^2. Alternatively it can be represented by

$$(\gamma_0, \gamma_1, \ldots, \gamma_{p+q}).$$

Defining the autocorrelation at lag k by

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

precisely the same information as in $\underset{\sim}{\Psi}$ is contained in

$$(\rho_1, \ldots, \rho_{p+q}).$$

The complete set of autocorrelations $\rho_1, \rho_2, \ldots$ is termed the autocorrelation function, or a. c. f.

Associated with the a. c. f. is the partial autocorrelation function, the p. a. c. f. This is a set $\pi_1, \pi_2, \ldots$ defined by

$$\pi_k = |\underset{\sim}{P}_k^*| / |\underset{\sim}{P}_k| \tag{2.6}$$

where $\underset{\sim}{P}_k$ is the $k \times k$ autocorrelation matrix, with general r, sth element $= \rho_{|r-s|}$; and $\underset{\sim}{P}_k^*$ is $\underset{\sim}{P}_k$, with every r, kth element replaced by ρ_r. In simple language, each π_k gives the conditional correlation between Y_t and Y_{t-k}, given the intervening $Y_{t-1}, \ldots, Y_{t-k+1}$; and each π_k can be interpreted as the φ_k of that AR (k) model which comes closest to representing the process.

The forms for the a. c. f. and p. a. c. f., associated with AR, MA and ARMA models, are summarised in table 2.1; while the actual functions, for particular low-order processes, are shown in figure 2.1.

TABLE 2.1

Characteristics of a. c. f. and p. a. c. f. for linear processes

Process	a. c. f.	p. a. c. f.
AR (p).....	Damps out	Cuts off after lag p
MA (q).....	Cuts off after lag q (*)	Damps out
ARMA.....	Damps out	Damps out

(*) Stronger results have been given by Anderson [2, 3].

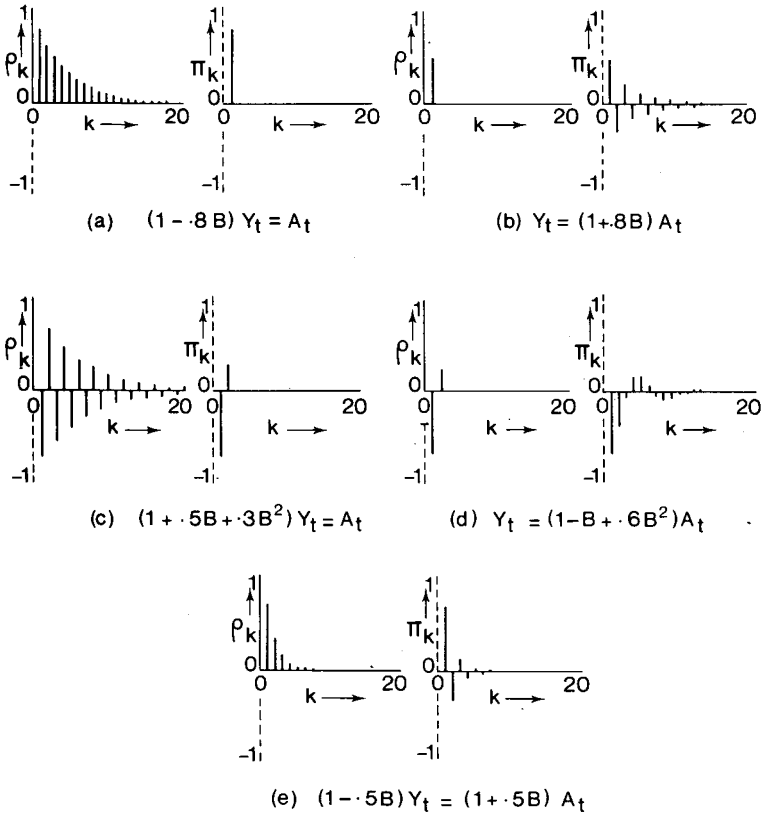

Figure 2.1

Theoretical a. c. f. and p. a. c. f. for some low-order models.

Time Series

Time processes are important because their realisations occur as sets of ordered observations such as $y_1, y_2, \ldots, y_n$, a *time series* of *length n*. Provided the process is *ergodic*, when the probabilistic structures of all its possible realisations are the same, the properties of such a series are expected to mimic, to some extent, those of its parent process, though "sampling error" will distort them. Thus one would expect the sample mean $\bar{y}$ and variance s_y^2 not to be significantly different from zero and γ_0 respectively, given the model (2.1).

The estimated a. c. f. and p. a. c. f. are sets $\{ r_k \}$ and $\{ p_k \}$, where

$$r_k = \frac{c_k}{c_0}$$

with

$$c_k = \frac{1}{n} \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}), \qquad k = 0, 1, \ldots$$

and p_k is obtained from $\{ r_k \}$ as π_k was from $\{ \rho_k \}$.

Though it will usually be quite easy to recognise the similarity between the sampled and theoretical functions when the model is known, it will be more difficult to deduce the process from the estimated functions due to the sampling errors and the fact that the estimated sets are themselves auto-correlated. Thus, for instance, the rules that, for an ARMA (p, q), the a. c. f. mimics that of an AR (p) process after $q-p$ lags, while the p. a. c. f. resembles that of an MA (q) after $p-q$ lags, are of little practical value in process modelling.

For an MA (q) process, from Bartlett [8],

$$\mathrm{Var}[r_k] \simeq \frac{1}{n}\left(1 + 2\sum_{1}^{q} r_i^2\right)$$

while for an AR (p) process, Quenouille [21] gave

$$\mathrm{Var}[p_k] \simeq \frac{1}{n}.$$

When n is fairly large, the distributions of r_k, $k > q$, for an MA (q), and p_k, $k > p$, for an AR (p), are roughly normal, with zero mean. These results *are* useful for identifying processes.

However, significant r_k and p_k values have to be considered with care. The "significance" refers to an individual estimate, whereas one wishes to interpret the set of non-independent estimates. Thus, amongst say $r_1, \ldots, r_{20}$, one does not "expect" to have just one value significant at the 5 % level. However in 20 such sets, one would expect to have many more than one such value. In fact, due to the serial dependence of the estimates, when one chance significant value occurs, there is a tendency towards having several significant values. Plotting the functions for known simulations seems to be the best way of gaining experience of how the sample a. c. f. and p. a. c. f. should be interpreted.

Finally we remark, that with our interest in forecasting in mind, the general model (2.2) can be written in *random shock* form

$$Y_t = \psi(B) A_t. \tag{2.7}$$

Here $\psi(\zeta)$ is a polynomial in ζ, in general of infinite degree, defined by

$$\psi(B) = \varphi_p^{-1}(B)\theta_q(B).$$

(2.7) is then of course an MA (∞) representation. Wold [25] has shown that *every* stationary process, from which any deterministic part has been removed, can be represented in this way.

3. THE BOX-JENKINS CYCLE

Given a time series history $y_1, y_2, \ldots, y_n$, the problem is to make an inference about a process $\{ Y_t \}$, which may be considered to have given rise to the realisation. Note that usually it is the particular series at hand which is of interest, and so $\{ Y_t \}$ does not necessarily represent the general ensemble of possible series, but just that subensemble which *is* ergodic with the given series.

The first step, as with any applied statistical problem, is to get the feel of the data. Ideally the series is plotted against time, and visual inspection will indicate whether it is plausible to assume that the process is stationary. The writer also likes to construct a histogram of the y_t, to see whether a gaussian assumption is reasonable, and to further test this assumption by obtaining the realisation skewness and kurtosis. (*Cf.* Webb [23] and Lomnicki [19].) In this section, we will assume that the hypothesis of a stationary gaussian process is acceptable; and, that the plot does not indicate that $\{ Y_t \}$ is seasonal.

In such a situation, one might expect that an adequate representation will be

$$\{ Y_t \} \sim \text{ARMA}(p, q),$$

where, from experience, $p+q$ is small. This last point is reasonable since, in practice, series are rarely sufficiently long to make any high order process a substantially superior fit to a carefully selected low order alternative. And, apart from the importance of avoiding a spuriously good fit by "data-mining", there are notably diminishing returns for effort in fitting more complex models.

Identification

In order to identify tentative initial choices for p and q, the a. c. f. and p. a. c. f. are calculated, and preferably plotted, for the first K lags, where K is, say, min (20, $n/4$). A suitable program has been given by Anderson [4]. Two questions are then asked:–

(*a*) Is $p_k \underset{\cdot}{\sim} N(0, 1/n)$ for $k > p$? If so, an AR (p) is indicated.

(*b*) Is $r_k \underset{\cdot}{\sim} N\left(0, 1/n\left(1+2\sum_1^q r_i^2\right)\right)$ for $k > q$? If so, an MA (q) is suggested.

If neither (*a*) nor (*b*) occurs, then neither the a. c. f. nor the p. a. c. f. "cut off", and an ARMA model is inferred.

Box and Jenkins [9] claim that, as a rule, one can take $p+q \leqq 2$. Then, generally, questions (*a*) and (*b*) should indicate whether the process should be tentatively identified as AR (1), AR (2), MA (1) or MA (2); while if none of these are indicated, by default an ARMA (1, 1) would be tried.

To give the reader some idea of how well this strategy might do, four simulated ARMA (p, q) series of length 200, with $p+q \leqq 2$, are analysed. (At the time of such analyses the generating processes should be unknown, but can be checked afterwards.) Even so the situation is better than can be expected with "real" series, since the processes are chosen of exactly the right form, whereas in practice they will only more or less approximate to this. Also a history length of 200 is much longer than is normally available. The results from the identification program are shown in table 3.1.

TABLE 3.1

Results from Identification program for 4 simulated series

	A		B		C		D	
k	r_k	p_k	r_k	p_k	r_k	p_k	r_k	p_k
1	-.800	-.800	.449	.449	-.719	-.719	.427	.427
2	.670	.085	-.056	-.324	.337	-.375	.475	.358
3	-.518	.112	-.023	.218	-.083	-.048	.169	-.160
4	.390	-.046	.028	-.118	.075	.239	.253	.106
5	-.310	-.061	.013	.077	-.088	.173	.126	.035
$\bar{y}$	.03		-.34		-.05		.09	
s_y^2	3.34		1.34		2.32		1.15	

For all four series $2/\sqrt{n} = .1414$.

Series A : $\{r_k\}$ disqualifies the possibility of MA, $\{p_k\}$ very strongly suggests AR (1).

Series B : $\{r_k\}$ suggests MA (1), $\{p_k\}$ seems compatible with this.

Series C : $\{r_k\}$ suggests MA (2), $\{p_k\}$ compatible with this.

Series D : $\{r_k\}$ disqualifies MA, $\{p_k\}$ suggests AR (2).

In fact all four identifications are correct, but we have not done quite so well as this suggests. For series D, p_3 is just significant. So, if we had not known that $p+q \leqq 2$, an AR (3) would have been identified. This model will be tried as an "overfit" later in the section.

For shorter realisations, the identification becomes much more difficult. Figure 3.1 shows the sampled functions for a simulated series of length 80 from the ARMA (1, 1) model

$$Y_t = .5\, Y_{t-1} + A_t + .5\, A_{t-1}.$$

This series is still relatively long, compared with lengths which usually occur in practice; but, comparing figures 3.1 and 2.1 (*e*), we see that there is a considerable difference between the estimated and theoretical functions.

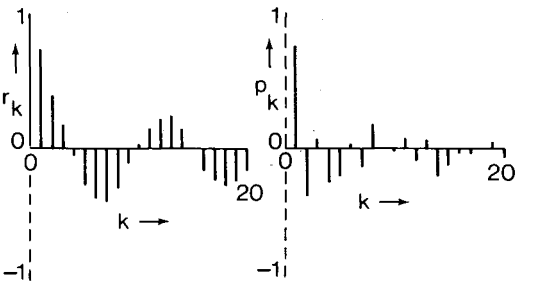

Figure 3.1

Estimated a. c. f. and p. a. c. f. for simulation of length 80 from $(1 - .5\,B)\, Y_t = (1 + .5\,B)\, A_t$.

As a general rule, for a Box-Jenkins analysis, the series length should be at least 40 (rather longer for the seasonal models of the next section). Otherwise, due to sampling error, the estimated functions will not contain sufficient information for a meaningful identification. Put another way, if the series is too short, any subtle patterns observed in it are as likely as not to be purely fortuitous.

Estimation

Once a model has been tentatively identified, its parameters have to be efficiently estimated, and the resulting fit assessed, mainly by an analysis of residuals, to see whether it can be accepted as a plausible explanation of the series. If the model is found to be inadequate, then this assessment should indicate promising modifications to the identification, and the cycle is repeated; and so on until the analyst is satisfied.

The efficient fitting can be left to a suitable computer package – for instance I.C.L. [14]. First, however, one should test that

$$E[Y_t] = 0 \tag{3.1}$$

is plausible. For, otherwise, one should work with the series

$$z_t = y_t - y.$$

The test is to compare $\bar{y}$, in the usual way, with its standard error, assuming (3.1) is true. For $p+q \leqq 2$, the standard errors can be obtained from table 3.2.

TABLE 3.2

Approximate Var $[\bar{y}]$ *for* ARMA (p, q) *processes,* $p+q \leqq 2$

p \ q	0	1	2
0	1	$1+2r_1$	$1+2r_1+2r_2$
1	$\frac{1+r_1}{1-r_1}$	$1+\frac{2r_1^2}{r_1-r_2}$	
2	$\frac{1+r_1}{1-r_1}\cdot\frac{1-2r_1^2+r_2}{1-r_2}$	All multiplied by c_0/n.	

Of course the cases with $p+q < 2$ are easily deduced from those with $p+q = 2$. Thus for MA (1), $\rho_2 = 0$, and the MA (2) and ARMA (1, 1) results both reduce to $1+2r_1$, on putting $r_2 = 0$. Similarly, putting $r_2 = r_1^2$ in the AR (2) and ARMA (1, 1) results gives that for AR (1). Putting $r_1 = 0$ reduces those of MA (1) and AR (1) to that of ARMA (0, 0), white noise.

Depending on the computer program, either rough parameter estimates ([1]) or null values are needed as initial values, from which the efficient estimates are obtained by iteration. A non-linear least squares procedure is used to obtain the vector of parameter estimates

$$(\underset{\sim}{\hat{\varphi}}, \underset{\sim}{\hat{\theta}}) \equiv (\hat{\varphi}_1, \ldots, \hat{\varphi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$$

which minimises the shock sum of squares

$$S(\underset{\sim}{\varphi}, \underset{\sim}{\theta}) = \sum_1^n \alpha_t^2,$$

where the

$$\alpha_t = \theta^{-1}(B)\varphi(B)\, y_t$$

are the estimated shocks given the model and the series.

([1]) For instance, for an AR (1) process, it is easy to show that $\rho_1 = \varphi$. So a sensible initial estimate for φ is given by $\hat{\varphi}_0 = r_1$.

Denoting the $\{ \alpha_t \}$ which minimise S by $\{ \hat{a}_t \}$, the efficient shock estimates, we have

$$y_t = \hat{a}_t + \hat{\omega}_1 y_{t-1} + \ldots,$$

where $(1 - \hat{\omega}_1 B - \ldots) = \hat{\theta}^{-1}(B)\,\hat{\varphi}(B)$, and

$$\hat{y}_t = \hat{\omega}_1 y_{t-1} + \ldots.$$

Subtracting these two equations gives

$$\hat{a}_t = y_t - \hat{y}_t$$

so the $\{ \hat{a}_t \}$ are in fact the residuals after fitting.

If it is possible, S contours should be displayed, since these often provide useful visual information as to how perhaps the fit should be modified. In fact, a preliminary plot of the S surface can indicate any peculiarities in the estimation situation, and for instance prevent the estimation program converging on a minor dip in the surface. For models with $p+q \leqq 2$, the plotting is straightforward, but for higher order processes the technique is much less convenient.

For series A to D, table 3.3 gives the estimates.

TABLE 3.3

Estimates for series A *to* D

Series	Parameter Estimates	S.E's of estimates	$\hat{\sigma}_A^2$
A	$\hat{\varphi} = -.805$	.042	1.091
B	$\hat{\theta} = .845$	.038	.833
C	$\hat{\theta}_1 = 1.010$	.053	.864
	$\hat{\theta}_2 = .635$	.054	
D	$\hat{\varphi}_1 = .285$	.067	.805
	$\hat{\varphi}_2 = .360$	.067	

The standard errors of the estimated parameters are needed to test the significance of these estimates. For low order models, the S.E.'s can be obtained from table 3.4, substituting estimates for the parameters in the expressions.

TABLE 3.4

Approximate Variances for parameter estimators of low order models

AR(1)	$\mathrm{Var}[\hat{\varphi}] \simeq \frac{1-\varphi^2}{n}$,
AR(2)	$\mathrm{Var}[\hat{\varphi}_1], \mathrm{Var}[\hat{\varphi}_2] \simeq \frac{1-\varphi_2^2}{n}$,
MA(1)	$\mathrm{Var}[\hat{\theta}] \simeq \frac{1-\theta^2}{n}$,
MA(2)	$\mathrm{Var}[\hat{\theta}_1], \mathrm{Var}[\hat{\theta}_2] \simeq \frac{1-\theta_2^2}{n}$,
ARMA(1, 1)	$\mathrm{Var}[\hat{\varphi}] \simeq \frac{(1-\varphi^2)(1+\varphi\theta)^2}{n(\varphi+\theta)^2}$,
	$\mathrm{Var}[\hat{\theta}] \simeq \frac{(1-\theta^2)(1+\varphi\theta)^2}{n(\varphi+\theta)^2}$.

Verification

The final portion of the Box-Jenkins cycle is to subject the identified and estimated model to "diagnostic checks" of its adequacy. Such checks should be designed to test for any suspected departures from the fit, and also to show up any other serious discrepancies.

If we suspect that a more elaborate model might be necessary, one with extra parameters can be "overfitted", and tested to see if it is indeed superior.

As indicated earlier in the section, the p. a. c. f. for series D suggests an AR (3) overfit should be tried. The estimates for this are

$$\hat{\varphi}_1 = .348, \quad \text{S.E.} = .072,$$
$$\hat{\varphi}_2 = .405, \quad \text{S.E.} = .070,$$
$$\hat{\varphi}_3 = -.168, \quad \text{S.E.} = .071, \quad \hat{\sigma}_A^2 = .803.$$

Comparing these with table 3.3, $\hat{\sigma}_A^2$ is not much smaller, nor are the parameters $\hat{\varphi}_1$ and $\hat{\varphi}_2$ significantly different from the corresponding estimates for the AR (2) fit. However, $\hat{\varphi}_3$ is significantly different from zero, as was suspected at the identification stage, so the AR (3) overfit is justified. An incorrect conclusion, as we know, but supported by the evidence.

Note that one should never increase the orders of both the AR and MA operators simultaneously, since this can easily lead to parameter redundancy. Thus, for instance, if we tried to fit an ARMA (1, 1), as an overfit to a truly

white noise model, we would run into extreme parameter instability, and the estimation program might well fail to converge. This is because a white noise process

$$Y_t = A_t$$

can be written as

$$(1-\varphi B)\, Y_t = (1+\theta B)\, A_t$$

for *any* choice of φ and $\theta = -\varphi$.

If we have to rely on the results themselves pointing to any model inadequacy, an analysis of the residuals is, as usual, helpful. Suppose an ARMA (p, q) model was identified when an ARMA (p^*, q^*) should have been. Then the residuals ought to follow a realisation of

$$\varphi_{p^*}(B)\,\theta_q(B)\, A_t = \varphi_p(B)\,\theta_{q^*}(B)\, C_t, \tag{3.1}$$

where $\{ A_t \}$ is not a white noise process, but $\{ C_t \}$ is. Thus we would expect the a. c. f. and p. a. c. f. of the residual series to mimic those for an appropriate ARMA $(p^*+q, p+q^*)$ model. However the fitted ARMA (p, q) is trying to approximate to the true ARMA (p^*, q^*), and so there will in general be some rough cancellation in (3.1) yielding a more parsimonious fit to the actual residuals, say

$$\varphi_{\tilde{p}}(B)\, a_t = \theta_{\tilde{q}}(B)\, c_t.$$

So, if a residual series shows evidence of not being a white noise realisation, an ARMA $(\tilde{p}, \tilde{q})$ model can be identified and fitted to it, and then combined with the original ARMA (p, q), to give ARMA $(p+\tilde{p}, q+\tilde{q})$. A rather better approach is to first identify $\tilde{p}$ and $\tilde{q}$ and then overfit an ARMA $(p+\tilde{p}, q+\tilde{q})$. It is reasonable to suppose that this single estimation will be more efficient than the product of two sequentially staged estimations. Of course, theoretically $\tilde{p}$ and $\tilde{q}$ must be replaced by p^*+q and $p+q^*$, but then the resulting ARMA $(p+p^*+q, q+p+q^*)$ has a common factor $\varphi_p(B)\,\theta_q(B)$ in both the AR and MA parts, and so on cancellation the ARMA (p^*, q^*) is retrieved.

For moderately long series and k not too small, the S.E.'s for the autocorrelations and partial autocorrelations of the residual series, $r_k(\hat{a})$ and $p_k(\hat{a})$, are obtained in the usual way. But for small k, Box and Pierce [10] show that this can no longer be done. They demonstrate that, under the assumption that a model of the correct form has been fitted, the early $r_k(\hat{a})$ can have variances much lower than the white noise value $1/n$. This extra sensitivity can show up otherwise unnoticed $r_k(\hat{a})$ as significant, and thus indicate a sensible overfit.

For the AR (1) fit to series A,

$$r_1(\hat{a}) = .072, \qquad r_2(\hat{a}) = .111,$$

$$r_3(\hat{a}) = -.009, \qquad r_4(\hat{a}) \text{ to } r_{20}(\hat{a}) < .135$$

in magnitude. The Box-Pierce theory gives

$$\text{S.E.}[r_1(\hat{a})] \simeq .057 < \text{S.E.}[r_2(\hat{a})] < \text{S.E.}[r_3(\hat{a})],$$
$$\text{S.E.}[r_4(\hat{a})] \simeq .068 < \text{S.E.}[r_k(\hat{a})], \qquad k > 4.$$

Thus none of the observed $r_k(\hat{a})$ are more than twice their standard errors away from zero, and the AR (1) fit seems satisfactory, no overfit being called for.

An easy, but low powered check is the *Portmanteau* lack of fit test, again due to Box and Pierce [10]. This can give little support to the model, should it prove not significant, but is simple to include in the estimation program. The statistic

$$R = n\sum_{1}^{K} r_k^2(\hat{a})$$

is computed, where K is sufficiently large. R evidently contains information on the first K of the $\rho_k(\hat{a})$, taken as a whole, and when the fitted model is appropriate

$$R \dot{\sim} \chi^2_{K-p-q}.$$

A significant R indicates model inadequacy. For convenience, K can be taken to satisfy

$$K = 20+p+q$$

when the model identified is ARMA (p, q), and then only χ^2_{20} points will be needed.

For instance, consider incorrectly identifying series C as AR (2). When fitted, this gives the very highly significant R value of 47.12, and so would be rejected. The fit for the correct MA (2) model has $R = 20.98$, which is not significant, even at the 40 % level.

Finally we consider the *Cumulative Periodogram* check on the residuals. Define

$$I(r) = \frac{2}{n}\left\{\left(\sum_{i=1}^{n} \hat{a}_i \operatorname{Cos} 2\pi \frac{r}{n} i\right)^2 + \left(\sum_{i=1}^{n} \hat{a}_i \operatorname{Sin} 2\pi \frac{r}{n} i\right)^2\right\}$$

for $r = 1, \ldots, [(n-2)/2]$, where [ ] denotes the "integer part of". Then we plot $C(j)$ against j/n, where

$$C(j) = \frac{\sum_{r=1}^{j} I(r)}{n\hat{\sigma}_A^2}, \qquad j = 1, \ldots, \left[\frac{n-2}{2}\right],$$
$$C\left(\left[\frac{n}{2}\right]\right) = 1.$$

If the residual series comes from a white noise process, the plot will be scattered randomly about the join of (0,0) to (.5,1). Inadequacies in the fit show up as systematic deviations from this line, and the significance of such deviations is assessed by the Kolmogorov-Smirnov test, (for instance, Siegel [22]).

Usually the results are presented by superimposing parallel dotted lines above and below the white noise join. These are so positioned that, if the plot crosses either, the results are significant at the appropriate level. Causes of significant values will be both model inadequacy and fitting error, (and of course chance). However, the test is very insensitive and rarely spots inadequacies which the other tests have missed if the identification and

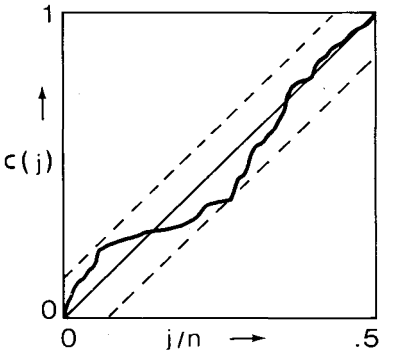

Figure 3.2

Cumulative Periodogram for residuals, after fitting AR (2) to series C; with 95 % confidence lines, for white noise process of same length, dotted in.

estimation have been sensibly carried out. In figure 3.2, the incorrect, and already rejected, AR (2) fit to series C shows up as only just significant. Apart from when the seasonality of a series has been blatantly overlooked, the test usually fails to spot the most evidently incorrect models, and consequently it seems doubtful whether it is really worth applying. However, visual comparison of plots for various competing fits can be useful.

Should a fit stand up to diagnostic checks it is not, of course, proved correct, but just shown to be plausible, and is adequate only in this sense. However, goodness of fit is not the only criterion when analysing a time series. Often the fitted model will be required to generate forecasts, and best fits do not necessarily give the best predictions. The closeness of fit depends very much on the peculiarities of the series at hand, whilst future values will depend more on the actual generating process. A perfect fit can always be obtained by choosing a high enough order model, but this is a pointless exercise for a statistical series.

4. INTEGRATED AND SEASONAL MODELS

Many observed non-stationary time series exhibit a certain homogeneity and can be accounted for by a simple modification of the ARMA model, the Autoregressive *Integrated* Moving Average model. This ARIMA (p, d, q) is written

$$\Phi(B)\, Y_t = \theta_q(B)\, A_t,$$

where $\Phi(B)$, the *generalised* autoregressive operator, is a polynomial of degree $p+d$, with exactly d zeros equal to unity and all the others outside the unit circle. So

$$\Phi(B) = \varphi_p(B)(1-B)^d = \varphi_p(B)\nabla^d,$$

where $\varphi_p(B)$ is a stationary autoregressive operator of order p, and the operator ∇ effects a differencing.

If we replace $\nabla^d Y_t$ by W_t, the ARIMA (p, d, q) process $\{ Y_t \}$ is reduced to an ARMA (p, q) process $\{ W_t \}$. Should a realisation show evidence that $E[W_t] \neq 0$, then the series $\{ w_t \}$ is replaced by $\{ z_t = w_t - \bar{w} \}$. Thus such a non-stationary series can, after the appropriate degree of differencing d, be treated by the methods of the previous section. Having obtained the $\{ w_t \}$ fit, the corresponding $\{ y_t \}$ fit can be obtained by the operation inverse to the differencing ∇^d, that is by summing or *integrating* the stationary $\{ w_t \}$ fit d times. Since this is a linear transformation, the optimal properties of the fit are retained.

As an example, consider a series which appears to be stationary, except that there are superimposed randomly occurring shifts in its level. Evidently we require a model whose behaviour is not influenced by the local level of the process, such that given any constant C

$$\Phi(B)(y_t + C) = \Phi(B)\, y_t.$$

This implies $\Phi(B)\, C = 0$ or equivalently that $\Phi(1) = 0$, and so $\Phi(B)$ has a factor $(1-B)$. If it has only one such factor, differencing once will result in a stationary series.

Should there also occur random shifts of *slope*, the model needs to have $\Phi(B)$ with a factor $(1-B)^2$, and differencing twice is necessary to produce stationarity. And so on, should stochastic *trends* of higher order be present — though, in practice, it is seldom found necessary to difference more than twice. In general, then, we have $d \leqq 2$; and, for non-seasonal series, $p+q \leqq 2$. The procedure for such homogeneous non-stationarity is to first recognise it, by visual inspection of the plotted series, and then remove it by the necessary degree of differencing. Alternatively, the fact that for a non-stationary process the a. c. f. follows a gentle linear decline, whilst for a stationary one the decline is rapid, allows d to be decided by inspection of the sampled a. c. f.'s for $\{ y_t \}$, $\{ \nabla y_t \}$ and $\{ \nabla^2 y_t \}$.

Differencing can also be used to remove deterministic trends, though it rapidly builds up the noise variance, and so should not be overdone.

Consider the I.C.I. closing stock price series, shown in figure 4.1, which exhibits a varying level. Its first two differenced series are shown in figure 4.2; (*a*) looks stationary, whilst (*b*) appears overdifferenced, the variance evidently being greater. The corresponding a. c. f.'s and p. a. c. f.'s are shown in figure 4.3 and support this conclusion.

Figure 4.1

I.C.I. closing stock prices (new pence), 25 August 1972-19 January 1973 (Financial Times).

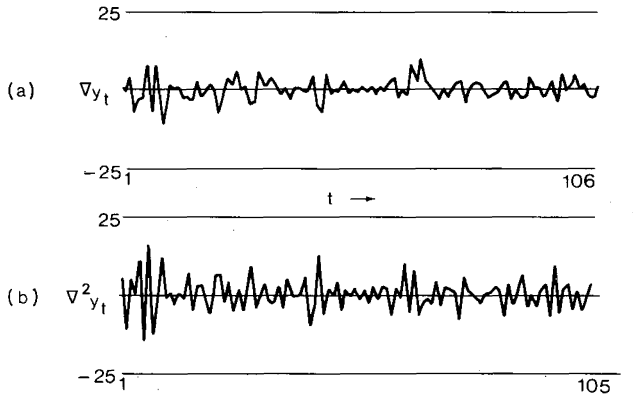

Figure 4.2

First two differenced series for I.C.I. closing stock price data.

Figure 4.3 (*a*) suggests the tentative identification of

$$\nabla Y_t \sim \text{white noise.} \tag{4.1}$$

This is on the basis of the low order r_k and p_k, but we note that inserting the resulting 2 S.E. lines gives r_7 approaching significance and p_7, p_{14} just significant, (which in fact, as we will see, suggests a seasonal component of period 7). Of course, for model (4.1), there is nothing to estimate, but the residuals can still be used to check the model's adequacy, and simple overfits can be tried. However doing this does not suggest anything preferable, so the white noise model will be retained for the present.

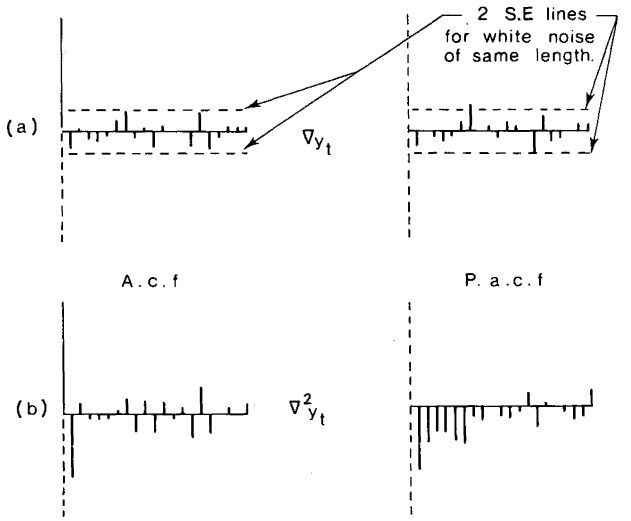

Figure 4.3
a. c. f.'s and p. a. c. f.'s for series of figure 4.2.

Figure 4.4 gives the daily Ben Nevis temperatures for a period in 1884; it gives the first differenced series and also the corresponding a. c. f.'s and p. a. c. f.'s. Looking at the temperatures plot, there appears to be a diagonal trend, which suggests differencing. The differenced series appears stationary. (Of course the trend is only the first half of the annual cycle, so for purposes of extrapolation a differenced model will be much more sensible than a straight line regression plus noise.) Again, for the temperatures, $\{ r_k \}$ and $\{ p_k \}$ suggest differencing; whilst the plotted functions for the differenced series suggest perhaps an ARIMA (0, 1, 2) model.

The differenced series was efficiently fitted and this gave

$$\nabla y_t = (1 - .238\,B - .305\,B^2)\,a_t, \qquad \hat{\sigma}_A^2 = 17.91.$$
$$[.068] \quad [.069]$$

where the numbers in brackets are the S.E.'s associated with the parameters directly above them. There is some trouble with the residuals initially, five of the first eleven being significant – but this can be attributed to the problem of "starting up". Apart from these, only four of the next 188 are significant. Only one of the residual autocorrelations is significant, and this only just, whilst the portmanteau χ^2 is very low. However the residual autocorrelations do present a wavelike pattern, though this is quite a common

"chance" phenomenon (and the cumulative periodogram for the residuals, which is not significant, does not suggest any missed periodicities). The fit would thus appear not too unreasonable.

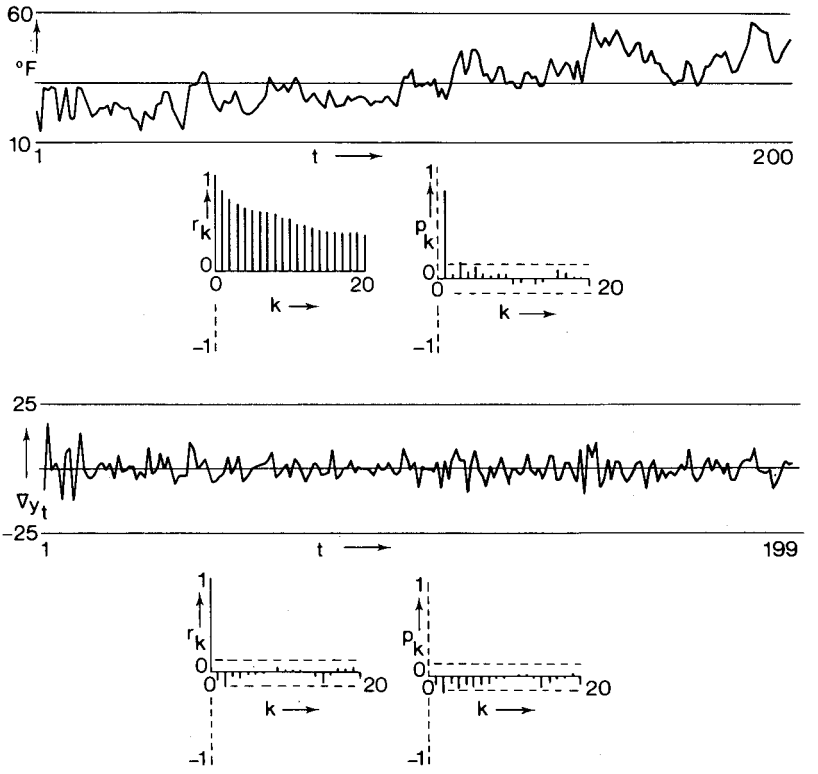

Figure 4.4

Ben Nevis noon temperatures, February 1st-August 18th 1884, with first differenced series and corresponding a. c. f.'s and p. a. c. f.'s.

Seasonal Models

We now extend our ideas to include the analyses of series with seasonal components. Figure 1.1 showed the Women Unemployed data, which display a pronounced periodic pattern as well as a changing trend. Here the seasonal period T is 12, the basic time interval being one month.

For modelling such a series, we introduce the (stationary) seasonal autoregressive operator of order P,

$$\Phi_P(B^T) \equiv 1 - \Phi_1 B^T - \ldots - \Phi_P B^{TP}$$

and the (invertible) seasonal moving average operator of order Q

$$\Theta_Q(B^T) \equiv 1+\Theta_1 B^T+\ldots+\Theta_Q B^{TQ}$$

and the seasonal difference operator

$$\nabla_T \equiv 1-B^T.$$

A model of form

$$\Phi_P(B^T)\nabla_T^D Y_t = \Theta_Q(B^T) A_t$$

will be called a SARIMA model of order $(P, D, Q)_T$. It has its theoretical a. c. f. and p. a. c. f. identical to those of the corresponding ARIMA model, obtained by replacing T by 1, except that the values will occur at intervals of T instead of consecutively. However note that, for the SARIMA, the *unit* interval between observations is not the same as the structural interval T, though of course every observation will help to fit the model. Because of the function sprawl, longer estimated functions are required; subject to not going beyond about $N/4$, where $N = n-d-TD$.

A general Box-Jenkins model can be written as

$$\varphi_p(B)\Phi_P(B^T)\nabla^d\nabla_T^D Y_t = \theta_q(B)\Theta_Q(B^T) A_t$$

which is the *multiplicative* $(p, d, q)\times(P, D, Q)_T$ model.

Such a model is useful in explaining many series with a marked periodicity. Thus, for the Women Unemployed data, one would expect autocorrelation between neighbouring months in the same year, and between the same months in adjacent years. Both the monthly and annual intervals are important. The idea can be extended to combine several distinct periodicities, and modified to give non-multiplicative models. Further generalisations are possible. (Certain other non-stationary series can be reduced to stationarity by an appropriate non-linear transformation. Thus, if, as with many economic series, the variance appears to increase or decrease in step with the local level; then a logarithmic transformation, to stabilise the variance, might be tried. But beware of the basic fact that, when transformed back, the model will no longer be optimal.)

The cycle of Identification, Estimation and Verification is unaltered in principle, though now the seasonal period must first be identified and then the degree of seasonal differencing, as well as that of unit differencing, decided on. The problem of identifying p, P, q and Q is usually extremely difficult ([2]) for the non-expert, and even the skilled analyst generally has to repeat the cycle several times, before a satisfactory fit obtains ([3]).

([2]) The functions for, say, the multiplicative model will reflect the characteristics of both the ARIMA and SARIMA components, but these will also interact to give extra terms. The sampling errors considerably confuse the already complicated theoretical results.

([3]) If available, on line facilities greatly spead up the modelling.

The autocorrelations and partial autocorrelations for a "suitable" transformation of the Women Unemployed series are shown in figure 4.5, and the final Box-Jenkins fit is

$$(1 - .349\,B)\nabla\nabla_{12}\,y_t = a_t,$$

the R value having then a probability of .93 of being exceeded.

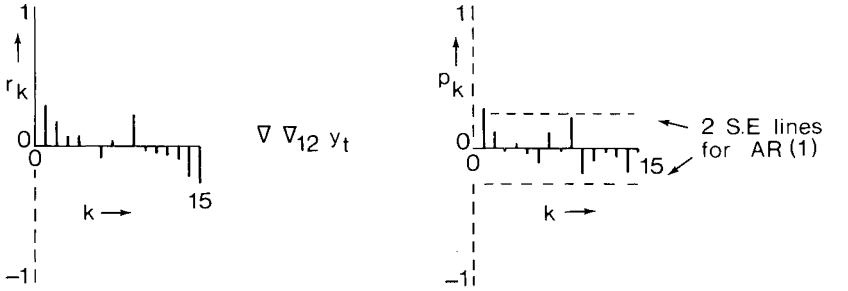

Figure 4.5

a. c. f. and p. a. c. f. for transformation of Women unemployed series.

For the I.C.I. data we get the model

$$(1 + .186\,B)\nabla\,y_t = (1 + .253\,B^7)\,a_t$$

with a chance .74 of R being exceeded. The seasonal period of 7 is hard to explain, as the stock exchange has a *five* day week.

5. MODEL INTERPRETATION

In practice, it is often insufficient to just model a series, but further necessary to relate the model to theory. For many workers will only accept a fit when they can see some theoretical explanation of how the model comes about. Whereas pure AR or MA models can have simple interpretations, proper ARMA models are more difficult to explain. However, in a fascinating paper, Granger [17] has shown how such more complex mixed models can arise, in a variety of ways, from basically pure AR or MA situations. This *interpretation* completes the Box-Jenkins analysis.

First of all, the sum of a number of independent simple processes, including at least one which is not pure MA, gives a proper mixed process. This depends on a result, discussed in Anderson [6], which we will call Granger's Lemma. This states that the sum of two independent MA processes, of orders q_1 and q_2, is also MA and of order max $[q_1, q_2]$.

For instance, suppose we have an AR (1) process

$$(1 - \varphi B)X_t = C_t \qquad (5.1)$$

buried in independent white noise

$$Y_t = D_t. \tag{5.2}$$

(The process $\{ Y_t \}$ might represent observation error.) Then what is the observed process $\{ Z_t = X_t + Y_t \}$?

Evidently

$$(1-\varphi B) Z_t = C_t + (1-\varphi B) D_t$$

and the right of this is MA (0) + MA (1), which, by Granger's lemma can be written in the form $(1+\theta B) A_t$. Thus

$$Z_t \sim \text{ARMA}(1, 1).$$

Conversely, it can be shown that a process

$$(1-\varphi B) Z_t = (1+\theta B) A_t$$

can be explained as arising from the sum of (5.1) and (5.2), provided certain *realisability* conditions hold, namely

$$\theta\varphi < 0$$

and

$$|\theta| < |\varphi|.$$

A second situation, where a pure process can give rise to an observed mixed process, arises when an incorrect choice of sampling interval is made. If a series is observed too frequently, evidently redundant data amasses, whereas if it is recorded too rarely, high-frequency detail is lost. But, look more closely at, say, the AR (2) model

$$(1-\alpha B)(1-\beta B) Z_t = A_t \tag{5.3}$$

when it is sampled at twice its structural interval. Then we would observe the process

$$(1-\alpha \Lambda^{1/2})(1-\beta \Lambda^{1/2}) Z_t = A_t, \tag{5.4}$$

where $\Lambda = B^2$ is the backshift operator for the *sampled* process. Then, multiplying (5.4) through by $(1+\alpha \Lambda^{1/2})(1+\beta \Lambda^{1/2})$ gives

$$(1-\alpha^2 \Lambda)(1-\beta^2 \Lambda) Z_t = A_t + (\alpha+\beta) A_{t-1} + \alpha\beta A_{t-2}, \tag{5.5}$$

where the R.H.S. is easily shown to have an MA (1) representation, in terms of a sampling interval of two. Thus (5.5) is ARMA (2,1), and alternate terms of an AR (2) process give rise to such a mixed process. (*See* also Anderson [7].)

As a further example, we give one of a number of results from some interesting work by Amemiya and Wu [1]. If an AR (p) process is observed as an *accumulated* series, observations being made every m structural intervals, then for $m > p$, the accumulated series follows an ARMA (p, p). (Also *see* Brewer [11].)

Finally, suppose that a pair of processes $\{ X_t \}$, $\{ Y_t \}$ are generated by the bivariate autoregressive scheme

$$\left.\begin{aligned} \alpha(B)X_t &= \beta(B)\, Y_t + A_t, \\ \gamma(B)\, Y_t &= \delta(B)X_t + C_t, \end{aligned}\right\} \tag{5.6}$$

where $\alpha(B)$, $\beta(B)$, $\gamma(B)$, $\delta(B)$ are finite polynomials in B of order a, b, c, d respectively, with $\alpha_0 = \gamma_0 = 1$, $\beta_0 = \delta_0 = 0$; and $\{ A_t \}$, $\{ C_t \}$ are independent white noise processes. Then, using Granger's lemma, it is straightforward to show that

$$\{X_t\} \sim \text{ARMA}(\max[ac, bd], \max[c, b]),$$
$$\{ Y_t\} \sim \text{ARMA}(\max[ac, bd], \max[a, d])$$

and so the simple "feed back" situation of (5.6) again gives rise to ARMA models.

6. FORECASTING

Now consider the problem of predicting a future value y_{h+n}, (h = hence from n = now), of a stationary zero-mean series, given the realisation to date $\{ y_1, \ldots, y_n \}$ but no other data. Any forecast of y_{h+n} will evidently be some function of $y_1, \ldots, y_n$, and we will restrict ourselves to just linear functions, that is to the class of *linear forecasts*. We will also assume that the best forecast is the one which has least e. m. s. e., (expected mean square error).

The process has a unique invertible, or marginally non-invertible, MA representation

$$Y_t = \sum_{j=0}^{\infty} \psi_j A_{t-j} = \psi(B) A_t, \qquad A_t \sim IN(0, \sigma_A^2), \tag{6.1}$$

where $\psi_0 \equiv 1$, and the zeros of $\sum_{j=0}^{\infty} \psi_j \zeta^j$ all lie on or outside the unit circle. The optimal forecast is then, as shown in Whittle [24],

$${}_hf_n = \sum_{j=0}^{n-1} \psi_{h+j} \hat{a}_{n-j}, \tag{6.2}$$

where the $\{ \hat{a}_t \}$ are the residuals after fitting the model, assumed to be correctly identified and exactly estimated. Then ${}_hf_n$ has variance given by

$${}_hV = \sigma_A^2 \sum_{j=0}^{h-1} \psi_j^2 \tag{6.3}$$

which is independent of n, and increases monotonically with forecast lead. So the further ahead one forecasts, the worse one expects to do, on average – which is intuitively reasonable.

In practice, the model might well be misidentified or misestimated, and so the forecasts are likely to have an error variance larger than (6.3), and to be biased. (Granger [18] even suggests inspection of the forecast errors, when the future values eventually come available, as a useful diagnostic check, since systematic departures from the theoretical error pattern can indicate a misspecification and point to a suitable model modification.) More serious is the fact that, even if the fitted model does closely explain the series history, it is very likely to change in the future. So the expectation of realistic forecasts, for anything but the short-run, is rather optimistic.

Of course, the optimal forecast function has been derived on the assumption that the aim of minimising forecast e. m. s. e. is valid. But this will only be so when the cost function associated with making forecast errors is quadratic. In practice, this is seldom realistic, as frequently the function will evidently not even be symmetric about zero error, under and over-forecasting by the same amounts not being equally "expensive". However, Granger [16] has shown that even when the assumption is not valid, a fairly efficient procedure is to forecast as if it were; but, in the case of an unsymmetrical cost function, to appropriately bias the resulting predictions, so that the errors on the more expensive side will be reduced.

REFERENCES

1. T. AMEMIYA and R. Y. WU, *The Effect of Aggregation on Prediction in the Autoregressive Model*, J.A.S.A., vol. 67, 1972, p. 628-632.
2. O. D. ANDERSON, *An Inequality with a Time Series Application*, J. Econom., vol. 2, 1974, p. 189-193.
3. O. D. ANDERSON, *Topics in Time Series*, Paper presented to the Mathematics, Computing and Statistics Applications Conference of the British Sociological Association, Guildford, 1974 (to be published).
4. O. D. ANDERSON, *Identifying Box-Jenkins Processes*, Computer Applications, vol. 2, 1974, p. 275-283.
5. O. D. ANDERSON, *Time Series Analysis and Forecasting, the Box-Jenkins Approach*, Butterworths, London, 1975.
6. O. D. ANDERSON, *On a Lemma Associated with Box, Jenkins and Granger*, J. Econom., vol. 3, 1975, p. 151-156.
7. O. D. ANDERSON, *On the Collection of Time Series data*, O.R. Quart., vol. 26, 1975, p. 331-335.
8. M. S. BARTLETT, *On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series*, J.R.S.S., B 8, 1946, p. 27-41.
9. G. E. P. BOX and G. M. JENKINS, *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco, 1970.

10. G. E. P. Box and D. A. Pierce, *Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models*, J.A.S.A., vol. 65, 1970, p. 1509-1526.
11. K. R. W. Brewer, *Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models*, J. Econom., vol. 1, 1973, p. 133-154.
12. A. Buchan, *Meteorology of Ben Nevis*, Trans. Roy. Soc. of Edinburgh, vol. 34, 1890.
13. H.M.S.O., *Unemployment Flow Statistics*, Department of Employment Gazette, 1973, p. 793-795.
14. I.C.L., 1900 *Series Technical Publication* 4284, 1972.
15. Financial Times, August 26th 1972-January 20th 1973.
16. C. W. J. Granger, *Prediction with a Generalised Cost of Error Function*, O.R. Quart., vol. 20, 1969, p. 199-207.
17. C. W. J. Granger, *Time Series Modelling and Interpretation*, Paper presented to the European Econometric Congress, Budapest, 1972.
18. C. W. J. Granger, *Multi-Step Forecast Errors and Model Mis-specification*, Paper presented to the Econometric Society, Oslo, 1973.
19. Z. A. Lomnicki, *Tests for Departure from Normality in the Case of Linear Stochastic Processes*, Metrika, vol. 4, 1961, p. 37-62.
20. *Nottingham City Engineer and Surveyor*, 1920. Meteorology of Nottingham.
21. M. H. Quenouille, *Approximate Tests of Correlation in Time Series*, J.R.S.S., B 11, 1949, p. 68-84.
22. S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, 1956.
23. R. A. J. Webb, *A Simulation Study of Lomnicki's Test for Departure from Normality in the Case of Linear Stochastic Processes*, MSc thesis, Nottingham University, 1972.
24. P. Whittle, *Prediction and Regulation by Linear Least-Squares Methods*, E.U.P., London, 1963.
25. H. Wold, *A Study in the Analysis of Stationary Time Series*, Almqvist and Wiksell, Stockholm, 2nd Edn, 1954.

APPENDIX

Women unemployed (1000's) in UK on 1st of each month, January 1967-July 1972.

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1967	96.4	104.1	102.0	100.3	96.4	87.3	85.1	87.9	89.8	96.3	99.5	95.6
1968	97.3	97.8	94.5	90.2	85.6	77.5	73.4	76.6	76.8	85.3	85.9	82.9
1969	84.5	84.5	82.3	78.3	74.6	68.7	71.9	74.1	74.9	82.4	83.3	80.1
1970	80.5	83.0	82.7	81.5	78.3	71.6	75.4	78.6	79.8	84.8	87.5	86.5
1971	92.0	98.8	104.9	104.8	99.2	91.7	100.6	106.1	110.3	117.0	123.1	122.9
1972	127.5	128.4	129.5	131.9	120.1	109.1	118.3					