(6) if there is no evidence of non-randomness in the one-step-ahead forecast errors, either initially or after model adaption, the forecasts are printed out by the computer. However, as new data comes to hand the forecast errors are monitored and abnormal changes recorded – for example, a large forecast error, or a change in the mean of the forecast errors, has occurred at a particular point in time.

In this way, after the preliminary analysis has been undertaken, large numbers of time series can be handled on an automatic basis, with abnormal situations signalled on an *exception* basis, so that an investigation can be started as to the cause of the anomalies. Applications of this approach have been made not only in manufacturing industry but also in forecasting deposits and withdrawals on an area basis for banks.

3.2. Three applications of transfer function models

Three applications of transfer function modelling will be described:

(1) The relationship between Market Share, Relative Price and Relative Advertising for a *non-seasonal* consumer product,

(2) the relationship between Electricity Consumption and Temperature (both *seasonal* series) for a national economy,

(3) the relationship between Manufacturing Employment and Manufacturing Output for the Federal Republic of Germany using quarterly and annual data.

3.2.1. Relationship between market share, price and advertising

This analysis formed part of a wider study to investigate the effect of various advertising and pricing policies on a range of consumer products. The following analysis relates to one product only.

Figure 18 shows the structure of the model, which relates the output variable, Market Share (Y_t) to two input variables: Relative Advertising (X_{1t}) and Relative Price (X_{2t}). Relative Advertising is defined as the ratio of the company's advertising on a consumer product to the total industry advertising on this particular type of product. Relative Price is the ratio of the company's price to a volume-weighted price for the industry.

Figure 19 shows time series data relating to the three variables in the model. The data series were recorded at intervals of four weeks for a period of 38 time intervals (nearly three years). Table 5 shows univariate models fitted to each time series separately.

The univariate model of the Market Share series is needed for two reasons:

(i) so that the reduction in residual variance, as a result of introducing the input variables X_{1t} and X_{2t}, can be measured,

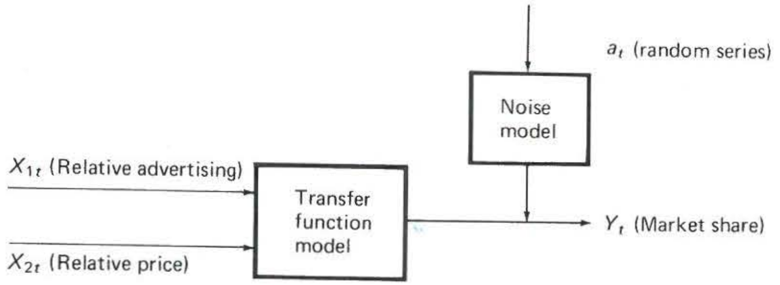

Fig 18. Structure of market share model relating market share (Y_t) to relative advertising (X_{1t}) and relative price (X_{2t}).

(ii) to provide a first approximation to the structure of the noise model to be used in the transfer function model.

The univariate models of the two input variables are needed to calculate the prewhitened cross correlation function between Market Share and each of the input variables, as is now illustrated in the case of Relative Advertising. The univariate model for relative advertising:

$$(1 - 0.80B + 0.53B^2)\nabla X_{1t} = c + (1 - 0.77B)(1 - 0.76B^{13})\alpha_{1t}$$

converts the highly correlated relative advertising series X_{1t} into an (approximately) random series α_{1t}. The identical operation

$$(1 - 0.80B + 0.53B^2)\nabla Y_t = c + (1 - 0.77B)(1 - 0.76B^{13})\beta_{1t}$$

is then applied to the Market Share Series Y_t, converting it into another time series β_{1t} which will not be random in general. The prewhitened cross correlation function between Market Share and Relative Advertising (Figure 20b) is then defined as the cross correlation function between β_{1t} and α_{1t}. This function has a large positive value at lag zero and is small thereafter, implying that an increase in Relative Advertising during the present four week period increases market share in the same period but has no influence on market share in the following periods. This observation suggests that a simple relationship

$$Y_t = \omega_0 X_{1t} + N_t \tag{1}$$

can be postulated between Market Share and Relative Advertising, where N_t is a

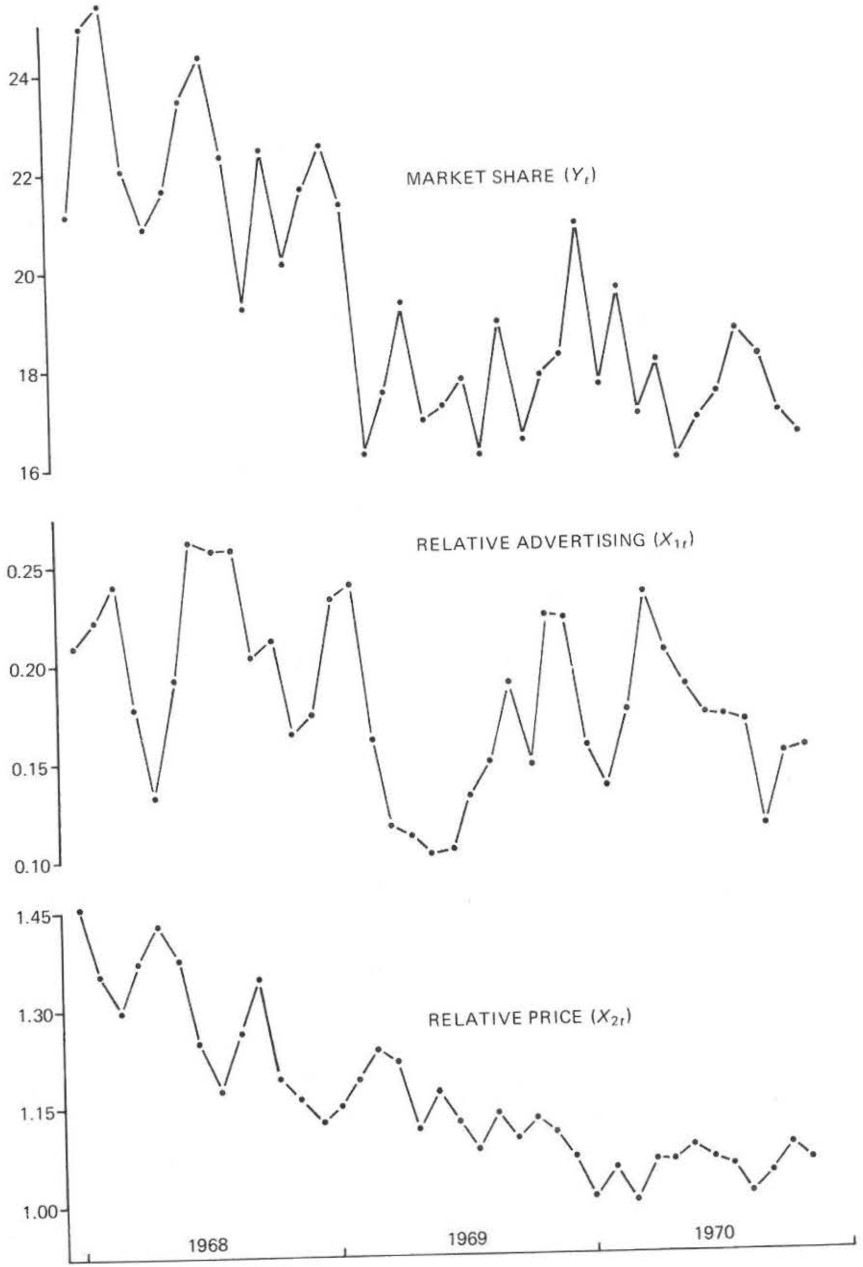

Fig 19. Market share, relative advertising and relative price of a consumer product: 4 week data from 1968, period 1 to 1970, period 12.

Table 5
Summary of univariate models fitted to each variable separately

Variable	Fitted model	Residual variance
Market share	$\nabla Y_t = (1 - \underset{\pm 0.13}{0.63}B)a_t$	3.30
Relative advertising	$(1-\underset{\pm 0.15}{0.80}B + \underset{\pm 0.14}{0.53}B^2)\nabla X_{1t} = \underset{\pm 0.0014}{-0.0017} + (1-\underset{\pm 0.14}{0.77}B)(1-\underset{\pm 0.08}{0.76}B^{13})\alpha_{1t}$	0.001103
Relative price	$(1-\underset{\pm 0.16}{0.41}B)\nabla X_{2t} = \underset{\pm 0.0008}{-0.0053} + (1-\underset{\pm 0.15}{0.95}B)\alpha_{2t}$	0.000819

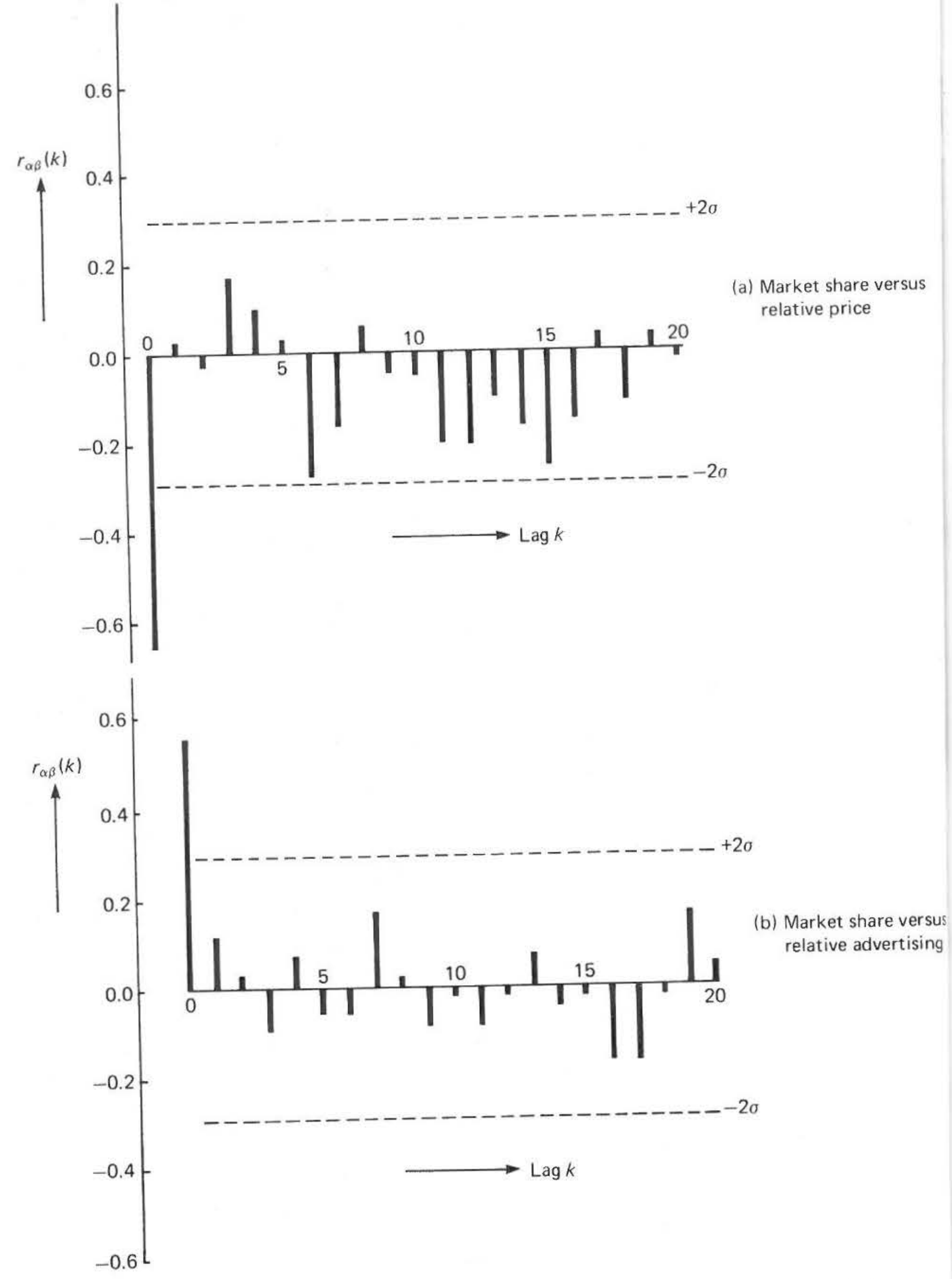

Fig 20. Prewhitened cross correlation functions between market share and (a) relative price, (b) relative advertising.

noise or error term. As a first approximation to the noise model N_t, we can take the univariate model for Y_t given in Table 5, that is

$$\nabla N_t = (1 - \theta B)a_t \qquad (2)$$

On combining (1) and (2), we obtain the transfer function-noise model

$$\nabla Y_t = \omega_0 \nabla X_{1t} + (1 - \theta B)a_t \qquad (3)$$

in which ω_0 and θ must now be estimated simultaneously. The justification for the use of the univariate model for Y_t as a first approximation to the model for the noise N_t is that if $\omega_0 = 0$ in (1), then $Y_t = N_t$. If $\omega_0 \neq 0$, as will be the case in this example, we are essentially fitting the transfer function model by perturbing about the univariate model. The fitted model is shown in the second row of Table 6.

Figure 20(a) shows the prewhitened cross correlation function between Market Share and Relative Price, using the univariate model for Relative Price given in Table 5. The cross correlation function has a large negative spike at lag zero and small correlations thereafter, suggesting a model with the same structure as the Relative Advertising model. The fitted model is shown in the third row of Table 6, from which it is seen that the residual variance is much smaller than for the Relative Advertising model, indicating that Relative Price has a much greater influence on Market Share than Relative Advertising.

A final model was fitted with both Relative Advertising and Relative Price as input variables. Because there is little correlation between X_{1t} and X_{2t}, we can use as an initial guess of the structure of the two input model, the structures of the two single input models. The fitted two input model is shown in the last row of Table 6, all the parameters in the model being estimated simultaneously. It can be seen that there is a further reduction in the residual variance as compared with the single input models. The dominance of Relative Price is again apparent from the magnitude of the estimate of the price parameter, compared with its standard error, in relation to the magnitude of the Advertising parameter, compared with its standard error. The two-input model implies that a unit increase in Advertising increases Market Share by 7.9% ± 3.8% and that a unit increase in price decreases Market Share by 40.5% ± 6.2%.

Manipulation of the two-input model in Table 6 enables it to be written in the alternative form

$$(Y_t - \bar{Y}_{t-1}) = 7.9(X_{1t} - \bar{X}_{1,t-1}) - 40.5(X_{2t} - \bar{X}_{2,t-1}) + a_t$$

Table 6
Summary of univariate and transfer function models relating market share (Y_t) to relative advertising (X_{1t}) and relative price (X_{2t})

Model type	Fitted Model	Residual Variance
Univariate (market share)	$\nabla Y_t = (1 - \underset{\pm 0.13}{0.63}B)a_t$	3.30
Single input Transfer function (versus relative advertising)	$\nabla Y_t = \underset{\pm 5.3}{16.1}\nabla X_{1t} + (1 - \underset{\pm 0.15}{0.65}B)a_t$	2.62
Single input transfer function (versus relative price)	$\nabla Y_t = \underset{\pm 6.2}{-44.5}\nabla X_{2t} + (1 - \underset{\pm 0.15}{0.43}B)a_t$	1.29
Two input transfer function (versus relative advertising and relative price)	$\nabla Y_t = \underset{\pm 3.8}{7.9}\nabla X_{1t} - \underset{\pm 6.2}{40.5}\nabla X_{2t} + (1 - \underset{\pm 0.15}{0.45}B)a_t$	1.08

where $\bar{Y}_{t-1}$, $\bar{X}_{1,t-1}$ and $\bar{X}_{2,t-1}$ are exponentially weighted moving averages with smoothing constant 0.45 and starting at time $t-1$. Thus, although the models of Table 6 look deceptively simple, containing few parameters, in effect they are quite sophisticated since they involve applying exponential smoothing to the three time series and then relating the smoothed series by what is effectively classical regression analysis with random residuals. Such smoothing is not arbitrarily chosen, of course, but is determined by the structure of the model as developed during the model building process.

In this application, the main objective was not to forecast as such but to gain an understanding of the mechanisms relating sales of various products to price and advertising expenditure. Based on models such as those in Table 6, an investigation can then be made of the effect of changes in price and advertising on market share, sales volume, sales revenue and profitability.

3.2.2. Relationship between electricity consumption and temperature

This section describes an application of transfer function modelling to the relationship between two seasonal time series: an output series Y_t consisting of the monthly electricity consumption for a national economy and an input series X_t, consisting of the corresponding mean monthly temperatures. The latter were

defined as the weighted average of the temperatures in five regions, the weights being proportional to the mean electricity consumption in a particular region. In this application greater emphasis will be placed on the technical aspects of the model building.

The upper part of Figure 21 shows a plot of the Y_t and X_t series and the lower part of Figure 21 their *range-mean plots.* This data provides an opportunity to emphasize another aspect of the identification stage (see Table 2 and Figures A.1, A.3) of the model building process, namely the transformation of the data before building a model. This is necessary because, in some situations, if no transformation is applied the variability of the residuals may increase with time, thus violating one of the assumptions made in the model. A rough indication of the nature of the transformation can be obtained by dividing the time series into sub-series and plotting the range against the mean for each sub-series (see Appendix A.1). The linear relationship between the range and mean of the electricity consumption series Y_t suggests the need for a logarithmic transformation of this series whilst the random scatter present in the range-mean plot of the temperature series X_t suggests that no transformation of this series is needed. Figure 22(a) shows the logarithms of the earlier part of the electricity consumption series together with the autocorrelation function of the whole series. The latter is characterised by a period of 12 and a failure to damp out, suggesting that the series is non-stationary and that nonseasonal differencing $\nabla \ln Y_t$ is needed. Figure 22(b) shows the series $\nabla \ln Y_t$ and its autocorrelation function. Whereas the differencing has removed the trend in the original series, seasonal non-stationarity is indicated by a failure to damp out at lags 12, 24, 36 etc. Indeed, the seasonal behaviour is more clearly visible than in the autocorrelation function of the original series. Thus, the need for further seasonal differencing ∇_{12} is demonstrated and Figure 22(c) shows the series $\nabla\nabla_{12}\ln Y_t$ and its autocorrelation function. The latter indicates that no further differencing is required and that the series $\nabla\nabla_{12}\ln Y_t$ is approximately stationary.

Figure 22(c) shows that the largest autocorrelations of $\nabla\nabla_{12}\ln Y_t$ occur at lags 1 and 12, suggesting that we may take as our initial guess of the structure of the model:

$$\nabla\nabla_{12}\ln Y_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

Based on the first-lag autocorrelation $r_1 = -0.27$ of $\nabla\nabla_{12}\ln Y_t$, we may take $\hat{\theta} = 0.30$ as an initial estimate of θ. Based on the twelfth-lag autocorrelation $r_{12} = -0.33$ of $\nabla\nabla_{12} \ln Y_t$, an initial estimate of $\hat{\Theta} = 0.35$ is suggested for Θ (see Box and Jenkins, 1970, pages 517, 518).

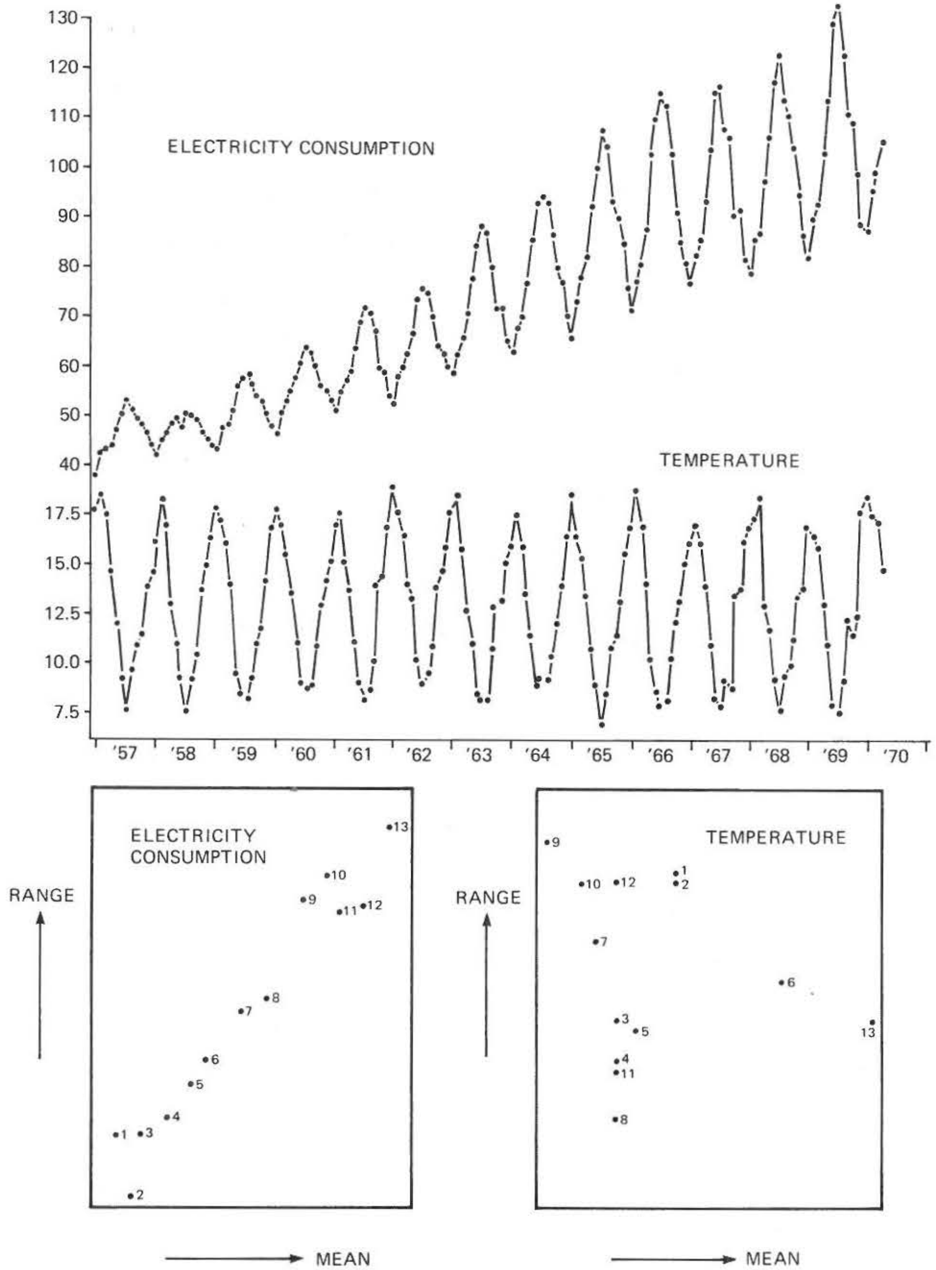

Fig 21. Plot of national electricity consumption and temperature, together with their range-mean plots: monthly data from January 1957 to April 1970. (Sub-series of 12 were used for calculation of range-mean plots; figure shows the sequence number of each sub-series).

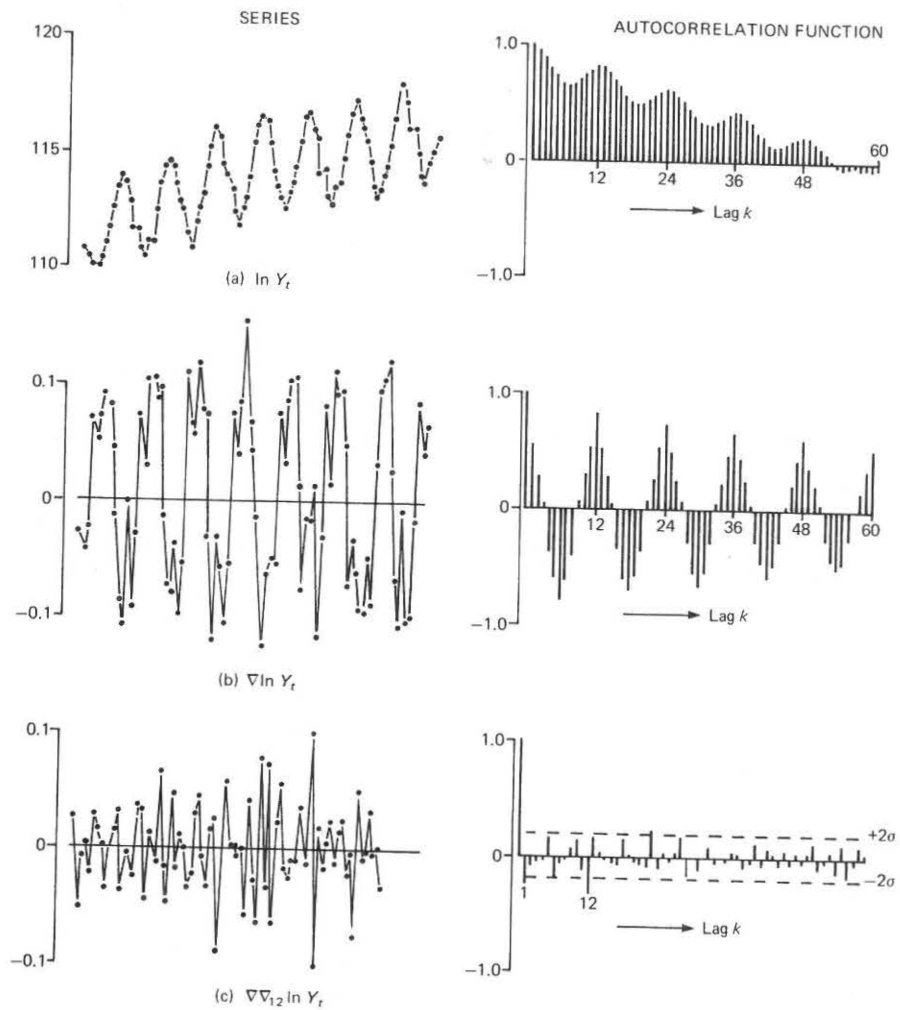

Fig 22. Various differences of electricity consumption series, together with their corresponding autocorrelation functions.

Table 7
Summary of univariate models fitted to national electricity consumption

Data set	Estimated model	Residual variance (per cent s.d.)
Whole series ($N = 160$)	$\nabla\nabla_{12}\ln Y_t = (1-0.30B)(1-0.55B^{12})a_t$ $\pm 0.08 \quad \pm 0.07$	0.0008010 (2.8%)
First half ($N = 64$)	$\nabla\nabla_{12}\ln Y_t = (1+0.14B)(1-0.83B^{12})a_t$ $\pm 0.12 \quad \pm 0.04$	0.0005616 (2.4%)
Second half ($N = 96$)	$\nabla\nabla_{12}\ln Y_t = (1-0.73B)(1-0.83B^{12})a_t$ $\pm 0.08 \quad \pm 0.05$	0.0006481 (2.55%)

The fitted model, based on $N = 160$ monthly observations, is shown in the first row of Table 7. The estimate of the residual standard deviation $\hat{\sigma} = 0.0283$ has a simple interpretation when a logarithmic transformation has been applied, as in this case. On multiplying by 100 we obtain a value of 2.8% which implies that the standard deviation of the residuals is approximately 2.8% of the 'level' of the series. Stated another way, it implies that approximately 2 out of 3 of the forecast errors one-step-ahead resulting from the use of the model can be expected to be less than 2.8% (in absolute value) of the level of the series at any point.

Figure 23(a) shows the residuals a_t and Figure 23(b) the residual autocorrelation function for the 'whole series' model. It can be seen that there are residual autocorrelations larger than two standard deviations at lags 6, 13, 23, 25, 29 and 31 and the corresponding chi-squared statistic (56.6 on approximately 30 degrees of freedom) corresponds to a probability level of approximately 0.002 under the assumption that the residual series is random, indicating model inadequacy. Further elaboration of the model by adding in turn non-seasonal and seasonal parameters did not result in a reduction in the size of these large residual autocorrelations. This state of affairs sometimes occurs with long time series. Whereas problems occur with short series because it may not be possible to obtain a satisfactory diagnosis of the structure of the model and accurate estimates of the parameters, the problem with longer series is different. With long series it may happen that changing circumstances cause the structure of the series to change slowly with time. To test this hypothesis, the series was split into approximately equal halves, the second half being somewhat longer than the first half in order to preserve a reasonable length of recent data to model for forecasting purposes. Table 7 shows the models fitted to the two halves

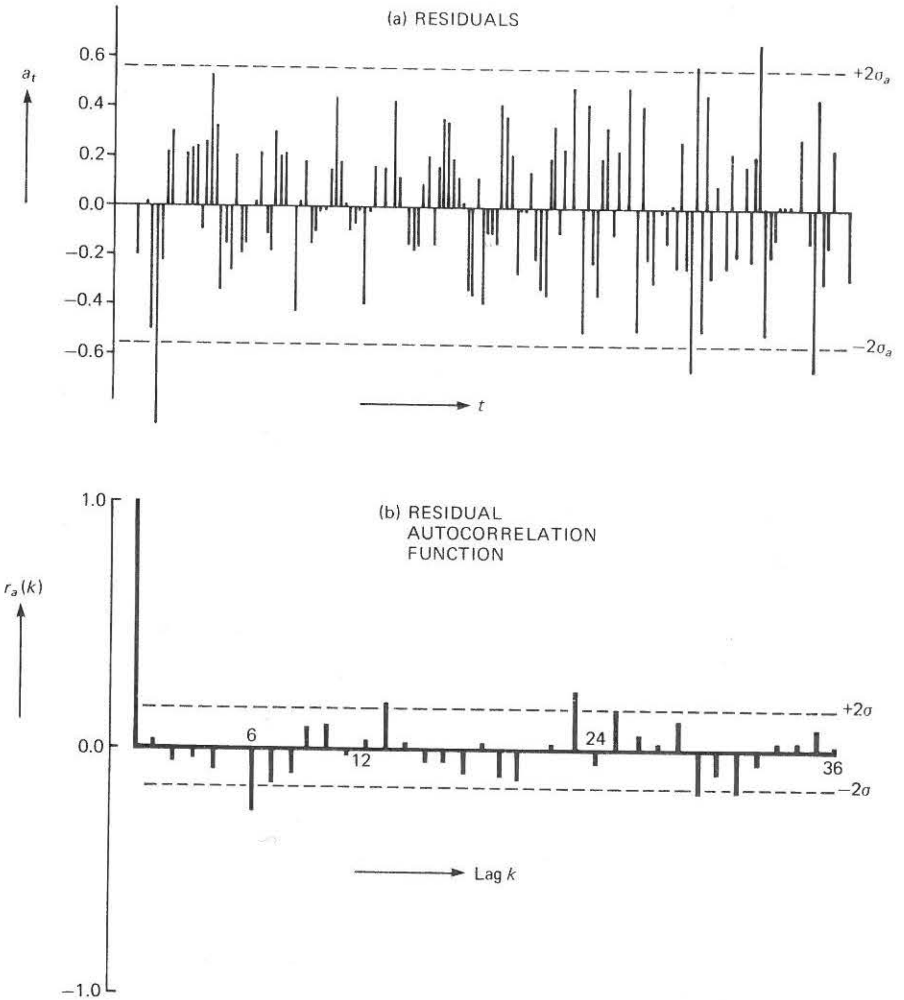

Fig 23. Residuals a_t and residual autocorrelation function $r_a(k)$ for univariate model

$$\nabla\nabla_{12}\ln Y_t = (1-0.30B)(1-0.55B^{12})a_t$$

fitted to the whole series of electricity consumption ($N = 160$ observations).

separately. Whereas there is reasonable agreement between the estimates of the seasonal parameter for the two halves, there is clear evidence that the estimates of the non-seasonal parameter differ significantly from one half to the next. However, the residual autocorrelation functions for the two models fitted to each half separately do not display abnormal values, confirming that these models are representationally adequate. A possible explanation for this behaviour is that the X-series behaves differently as between the first and second half of the series. However, the behaviour of the temperature series is fairly regular and so this possibility can be ruled out. Accordingly, it was decided to work with the second half of the series for transfer function modelling.

Prewhitening. The modelling of the temperature series was relatively straightforward and is not described in detail here. Since the temperature series does not contain a trend, no non-seasonal differencing was needed. However, since the temperature series is highly seasonal, seasonal differencing ∇_{12} was required. The final model was

$$\underset{\pm 0.08}{(1 - 0.27B)}\nabla_{12}X_t = \underset{\pm 0.03}{(1 - 0.88B^{12})}\alpha_t. \tag{4}$$

Proceeding as in the example of Section 3.2.1, the prewhitened cross-correlation function is calculated by applying the temperature model (4) to the logarithm of the consumption series, that is

$$(1 - 0.27B)\nabla_{12}\ln Y_t = (1 - 0.88B^{12})\beta_t$$

and then calculating the cross-correlation function between β_t and α_{t-k} at different lags k. Figure 24 shows the prewhitened series and the cross-correlation function, together with its approximate two standard error limits under the assumption that the two series are unrelated (see Box and Jenkins, 1970, p.376). The outstanding feature of the prewhitened cross correlation function is a large negative value at lag zero, implying that an increase in temperature this month decreases electricity consumption this month and vice versa. This behaviour suggests a relationship of the form

$$\ln Y_t = \omega_0 X_t + N_t \tag{5}$$

between electricity consumption and temperature. As a first guess of the structure of the noise model we can use the univariate model for $\ln Y_t$ given in Table 7, namely

$$\nabla_{12}N_t = (1 - \theta B)(1 - \Theta B^{12})a_t. \tag{6}$$

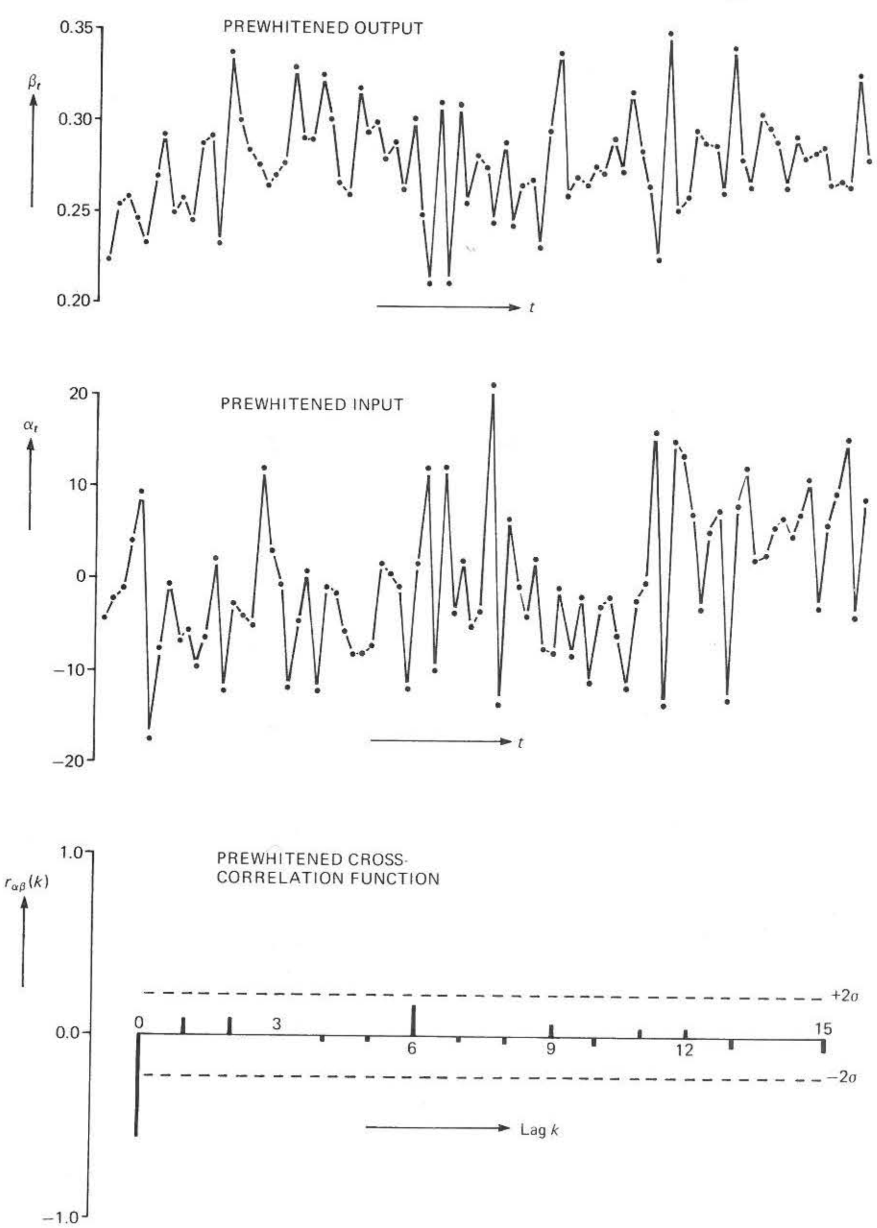

Fig 24. Prewhitened output (electricity consumption) series β_t, prewhitened input (temperature) series α_t and prewhitened cross correlation function $r_{\alpha\beta}(k)$.

Combining (5) and (6) we obtain for the overall transfer function-noise model.

$$\nabla\nabla_{12}\ln Y_t = \omega_0\, \nabla\nabla_{12}X_t + (1-\theta B)(1-\Theta B^{12})a_t.$$

In addition to estimates of the prewhitened cross correlation function, the computer program MTID (see Appendix A.3) also calculates preliminary estimates of the *impulse response weights* v_j in the impulse response representation*

$$\ln Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots + N_t \qquad (7)$$

relating $\ln Y_t$ and X_t. Comparing this representation with (5) above we see that an initial estimate of ω_0 is provided by an estimate of v_0. In this example this was $v_0 = -0.0015$. Similarly, initial estimates of θ and Θ are obtained from the univariate model (based on $N = 96$) given in Table 7, namely $\hat{\theta} = 0.73$ and $\hat{\Theta} = 0.83$.

Fitted transfer function model. The final model, with ω_0, θ and Θ estimated simultaneously by fully efficient likelihood methods is shown in Table 8, together with the best univariate model for comparison. The estimate of ω_0 is approximately 10 times its standard error and there is a very large reduction in residual variance as compared with the univariate model. Written in the form

$$\ln Y_t = \underset{\pm 0.00020}{-0.00222}\, X_t + N_t$$

where

$$\nabla\nabla_{12} N_t = (1 - \underset{\pm 0.09}{0.64}B)(1 - \underset{\pm 0.09}{0.75}B^{12})a_t$$

the model is capable of a simple interpretation. It implies that a 1° Centigrade rise in temperature in a given month decreases electricity consumption by 0.22% ± 0.02% in that same month.

*As a starting point in the identification of the lag structure, it is convenient to assume that there is a parameter v_j, called the impulse response weight, associated with each lag. The prewhitened cross correlation when multiplied by a constant provides estimates of the impulse response weights v_j, and hence can be used to suggest a more parsimonious representation of the lag structure based on a few parameters – which can then be estimated efficiently.

Table 8
Comparison of univariate and transfer function models fitted to second half (N = 96) of electricity consumption (Y_t) and temperature (X_t) series.

Model type	Estimated model	Residual Variance (% s.d.)
Univariate	$\nabla\nabla_{12}\ln Y_t = (1 - \underset{\pm 0.08}{0.74}B)(1 - \underset{\pm 0.05}{0.83}B^{12})a_t$	0.0006481 (2.55%)
Transfer function	$\nabla\nabla_{12}\ln Y_t = \underset{\pm 0.00020}{-0.00222}\nabla\nabla_{12}X_t + (1 - \underset{\pm 0.09}{0.64}B)(1 - \underset{\pm 0.09}{0.75}B^{12})a_t$	0.0003032 (1.74%)

Correlations between parameter estimates. Another important aspect of an estimation situation is the *correlation matrix* of the parameter estimates. High correlations between parameter estimates indicate that the likelihood surface is elongated in certain directions, implying that there are many combinations of the parameters in the neighbourhood of the maximum likelihood estimates with very nearly equal likelihood. Such ambiguity in the estimation situation is undesirable and sometimes indicates that the model is too elaborate or that it is mis-specified. By the latter is meant that, for example, several autoregressive parameters may have been introduced into the model whereas fewer moving average parameters would have sufficed. When all attempts to remove high correlations have failed, it may be necessary to live with them, recognising that they are a property of the particular data set being analysed and a consequence of one's inability to *design experiments* to collect time series data with the objective of achieving uncorrelated (orthogonal) parameter estimates.

The correlation matrix of the three parameters in the transfer function model of Table 8 is shown below.

	$\hat{\omega}_0$	$\hat{\theta}$	$\hat{\Theta}$
$\hat{\omega}_0$	1.00		
$\hat{\theta}$	0.14	1.00	
$\hat{\Theta}$	0.00	−0.07	1.00

It is rather remarkable that what, at first sight, seems a rather complex relationship between two series can be described by a model containing only three parameters and that these parameters are virtually orthognal.

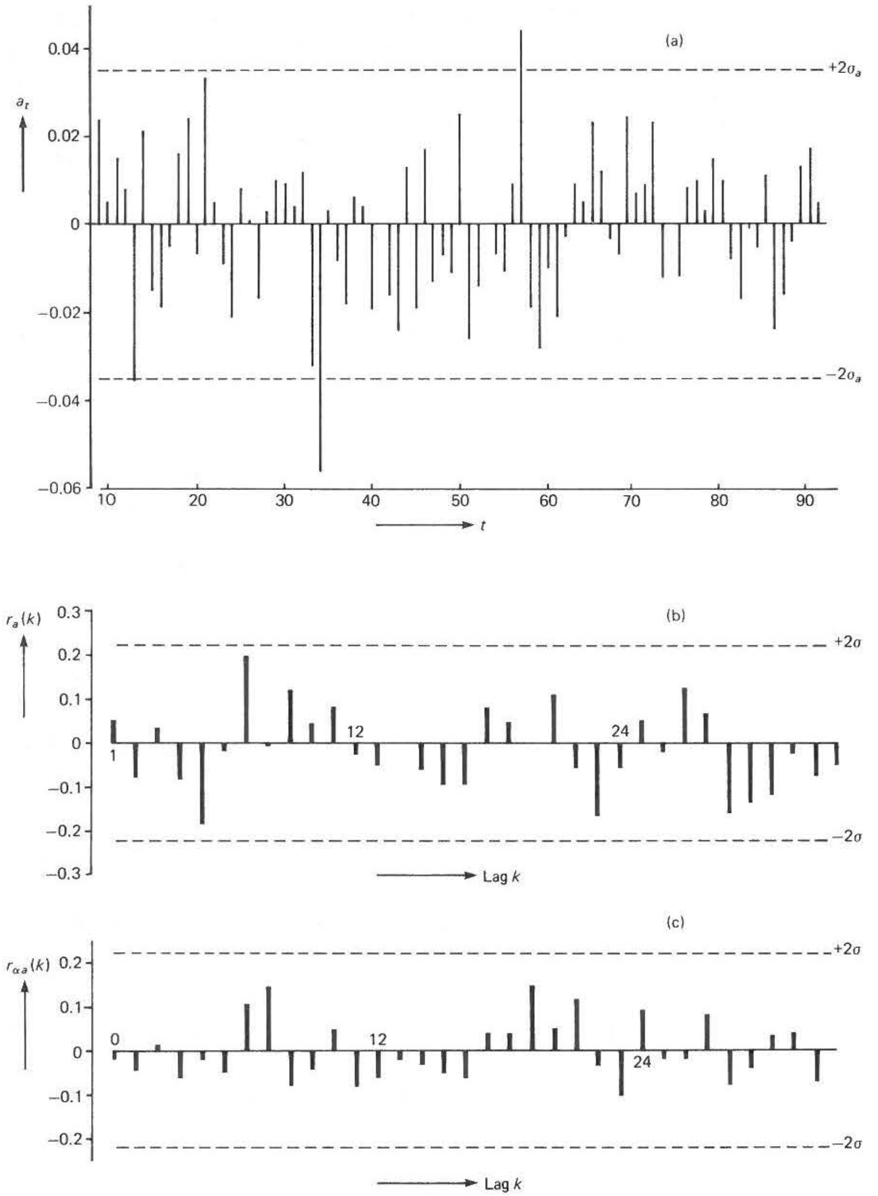

Fig 25. Diagnostic checks for transfer function model relating electricity consumption to temperature: (a) residuals a_t, (b) residual autocorrelation function $r_a(k)$, (c) cross-correlation function $r_{\alpha a}(k)$ between residuals and prewhitened temperature series.

Figure 25 shows the residuals a_t from the transfer function model given in Table 8, the autocorrelation function $r_a(k)$ of the residuals and the cross-correlation function $r_{\alpha a}(k)$ between the residuals and the prewhitened temperature. There is no obvious evidence of model inadequacy. However, there are two large consecutive residuals at $t = 38$ $(-1.9\sigma_a)$ and $t = 39$ $(-3.3\sigma_a)$ but no explanation could be found for these.

The nature of the checks shown in Figure 25 should be noted:

(1) A plot of the residuals a_t (Figure 25(a)), together with 'control limits' $\pm 2\sigma_a$, can indicate points where the residuals seem to be unrepresentative, when compared with the overall distribution of residuals. Such unrepresentative, or abnormal, residuals are indicative of large external shocks, such as a strike or, quite often, an anomaly in the data due to misrecording or wrong transcription.

(2) The residual autocorrelation function (Figure 25(b)) looks for evidence of non-randomness in the residuals. Such non-randomness could be indicative of inadequacies in the transfer function model and the noise model. (See Box and Jenkins, 1970, pages 392, 393).

(3) The cross correlation function between the residuals and the temperature series (Figure 25(c)) provides evidence of inadequacy in the transfer function model. (see Box and Jenkins, 1970, pages 393, 394).

Most econometric models quote as evidence of the 'adequacy' the Durbin–Watson (DW) statistic (see Durbin and Watson, 1950, 1951). Whereas this statistic has served a useful purpose in the past, it should be remembered that it is equivalent to the first lag of the residual autocorrelation function. Thus, a model could be inadequate despite an 'acceptable' Durbin–Watson statistic, due to the presence of large autocorrelations at lags greater than 1. More importantly, even if the residual autocorrelations are satisfactory, the model could still be unsatisfactory since significant cross-correlations might exist between the residuals and the 'independent variables'. Such cross correlation checks are very important since they provide clues as to the inadequacy of the transfer function. The absence of such checks from econometric models represents a serious deficiency in the building of such models.

Forecasting. The transfer function model forecasts of electricity consumption can be calculated only if there exist forecasts of temperature for various lead times. Table 9a and Figure 26 show forecasts of the electricity consumption series for lead times 1, 2, . . . , 12 months from origin $t = 96$, based on:

(a) the univariate model of Table 8,

(b) the transfer function model of Table 8, with temperature forecast using

Initially, trade promotions were classified into 'strong', 'medium' and 'weak' categories according to an assessment made by the organisation's Marketing Department. It was found that 'strong' and 'medium' promotions had very similar effects whereas 'weak' promotions had no discernible effect on sales. Accordingly, in the subsequent analysis, 'strong' and 'medium' trade promotions were classified together, and called 'trade promotions', while 'weak' promotions were ignored.

The first 5 rows in Table 11 show respectively,

(a) the univariate model fitted to the normalised sales series, ignoring the effect of promotions,

(b) an intervention model relating sales to trade promotions, whether 'strong' or 'medium',

(c) an intervention model relating sales to 'strong' trade and 'medium' trade promotions,

(d) an intervention model relating sales to consumer promotions,

(e) an intervention model relating sales to consumer and trade promotions.

It is not possible to use prewhitening to identify intervention models (see Appendix A.4). Thus, to illustrate how these models were fitted, suppose Y_t is the normalised sales series and ξ_t is a variable representing the occurrence of a particular type of promotion. It was believed by some in the organisation that the effect of a promotion this month is to increase sales in that same month. Thus, the model may be written

$$Y_t = \omega_0 \xi_t + N_t \tag{17}$$

where N_t is a noise or error, representing that part of the sales series that can not be explained by trade promotions. However, others believed that a promotion this month affected not only sales this month but also sales the following month. In this case the model

$$Y_t = (\omega_0 - \omega_1 B)\xi_t + N_t \tag{18}$$

could be entertained. As in previous examples, the univariate model for Y_t given in the first row of Table 11 can be used as an initial guess of the structure of the noise, that is

$$\nabla N_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^{13} - \Theta_2 B^{26})a_t. \tag{19}$$

Combining (19) with, for example (18), models of the form

$$\nabla Y_t = (\omega_0 - \omega_1 B)\nabla \xi_t + (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^{13} - \Theta_2 B^{26})a_t \tag{20}$$

Table 11
Summary of models fitted to sales data (orders received) normalized by dividing by the number of delivery days per accounting period (N = 38 data points).

Model	Intervention Variables	Estimated model	Residual standard deviation	Largest residuals
(a)	None: univariate model only	$\nabla Y_t = (1 - \underset{\pm 0.15}{0.45}B - \underset{\pm 0.15}{0.48}B^2)(1 + \underset{\pm 0.15}{0.40}B^{13} + \underset{\pm 0.10}{0.52}B^{26})a_{1t}$	1439	$a_{24} = 2.4\sigma$
(b)	$\xi_{1t} = \begin{cases} 1, & \text{Trade Promotion} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 382}{2387}\nabla\xi_{1t} + (1 - \underset{\pm 0.16}{0.53}B - \underset{\pm 0.16}{0.38}B^2)(1 + \underset{\pm 0.20}{0.55}B^{13})a_{2t}$	1090	$a_{10} = -2.0\sigma$ $a_{11} = 2.4\sigma$
(c)	$\xi_{2t} = \begin{cases} 1, & \text{strong trade promotion} \\ 0, & \text{otherwise} \end{cases}$ $\xi_{3t} = \begin{cases} 1, & \text{medium trade promotion} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 462}{2213}\nabla\xi_{2t} + \underset{\pm 594}{2619}\nabla\xi_{3t} + (1 - \underset{\pm 0.17}{0.54}B - \underset{\pm 0.17}{0.37}B^2)(1 + \underset{\pm 0.21}{0.57}B^{13})a_{3t}$	1101	$a_{10} = -1.9\sigma$ $a_{11} = 2.3\sigma$
(d)	$\xi_{4t} = \begin{cases} 1, & \text{consumer promotion} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 528}{2184}\nabla\xi_{4t} + (1 - \underset{\pm 0.15}{0.38}B - \underset{\pm 0.15}{0.51}B^2)(1 + \underset{\pm 0.21}{0.35}B^{13})a_{4t}$	1353	$a_5 = 2.5\sigma$ $a_{10} = -1.7\sigma$ $a_{11} = 2.0\sigma$
(e)	$\xi_{1t} = \begin{cases} 1, & \text{trade promotion} \\ 0, & \text{otherwise} \end{cases}$ $\xi_{4t} = \begin{cases} 1, & \text{consumer promotion} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 615}{2778}\nabla\xi_{1t} - \underset{\pm 638}{572}\nabla\xi_{4t} + (1 - \underset{\pm 0.17}{0.56}B - \underset{\pm 0.17}{0.35}B^2)(1 + \underset{\pm 0.20}{0.61}B^{13})a_{5t}$	1095	$a_{10} = -2.0\sigma$ $a_{11} = 2.3\sigma$
(f)	$\xi_{5t} = \begin{cases} 1, & \text{trade \textit{and} consumer promotion} \\ 0, & \text{otherwise} \end{cases}$ $\xi_{6t} = \begin{cases} 1, & \text{trade promotion only} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 403}{2206}\nabla\xi_{5t} + \underset{\pm 615}{2778}\nabla\xi_{6t} + (1 - \underset{\pm 0.17}{0.56}B - \underset{\pm 0.17}{0.35}B^2)(1 + \underset{\pm 0.20}{0.61}B^{13})a_{6t}$	1095	$a_{10} = -1.9\sigma$ $a_{11} = 2.3\sigma$
(g)	$\xi_{1t} = \begin{cases} 1, & \text{trade promotion} \\ 0, & \text{otherwise} \end{cases}$ $\xi_{7t} = \begin{cases} 1, & \text{year 3, period 4 (2 sales forces)} \\ 0, & \text{otherwise} \end{cases}$	$\nabla Y_t = \underset{\pm 353}{2012}\nabla\xi_{1t} - \underset{\pm 949}{1607}\nabla\xi_{7,t-1} + (1 - \underset{\pm 0.17}{0.51}B - \underset{\pm 0.17}{0.46}B^2)(1 + \underset{\pm 0.21}{0.57}B^{13})a_{7t}$	1045	$a_{10} = -1.9\sigma$ $a_{11} = 2.3\sigma$

were entertained and then fitted for different categories of promotional variable ξ_t. Then parameters were omitted if they turned out to be small compared with their standard errors and the simpler models refitted. In this way conflicting views as to the effect of certain promotions could be resolved.

The following comments may be made concerning the final models which are shown in Table 11:

(i) The univariate model (a) displays strong seasonality, most of it due to the strongly seasonal nature of the promotional activity.

(ii) Model (b) shows that trade promotions have a strong effect on sales, the estimate $\hat{\omega}_0 = 2387$ being over six times its standard error. Note that the seasonal part of the model has simplified considerably as a result of introducing the promotional effect. Note also that the large residual at 1974, period 11 ($t = 24$) in the univariate model (a) has now disappeared since it has been explained by the promotion variable at that point.

(iii) Model (c) shows that if promotions are split into 'strong' promotions and 'medium' promotions, these two types of promotion have effects on sales that are similar in magnitude. Moreover, since model (c) does not result in a better fit than model (b), there is no advantage in distinguishing between 'strong' and 'medium' trade promotions, as mentioned above.

(iv) Model (d) suggests that consumer promotions also affect sales but that the effect is weaker since the residual variance is much larger for model (d) than for model (b).

(v) Model (e) shows the effect of introducing trade and consumer promotions simultaneously. It indicates that consumer promotions have no effect, seeming to negate the result implied by Model (d). However, closer examination of the promotion series reveals that whereas trade promotions are sometimes held in the absence of consumer promotions, consumer promotions are always held at the same time as trade promotions – that is, there are no times when consumer promotions are held on their own. Hence, it is difficult to disentangle the two effects. Thus, the strong effect due to consumer promotions in Model (d) could be due to the fact that trade promotions were held simultaneously.

(vi) To clarify the position, Model (f) was fitted using two other intervention variables defined as follows:

$$\xi_{5t} = \begin{cases} 1, & \text{trade and consumer promotions occur simultaneously,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\xi_{6t} = \begin{cases} 1, & \text{trade promotions occur on their own,} \\ 0, & \text{otherwise.} \end{cases}$$

Model (f) shows that it makes no difference to the magnitude of the trade promotion effect if it occurs with or without consumer promotions – again suggesting that consumer promotions have no effect and confirming the results implied by Model (e).

This example raises interesting questions as to how promotional activity should be carried out most effectively. It suggests that organisations should not only undertake promotional activity but also, at the same time, should take more care to generate the right kind of information in order to determine the effectiveness of their promotional activity. In this example, this could have been achieved if some periods had trade promotions on their own, some periods had consumer promotions on their own and some periods had neither trade nor consumer promotions. If this had been done it would have been possible to separate out the effects of trade and consumer promotions with greater precision.

(vii) Models (b) to (f) have negative residuals from 1976, Period 5 to 1976, Period 9. Since this was unusual behaviour, the organisation were asked what happened at this point in time. It transpired that 1976, Period 4 corresponded to the splitting of the sales force into two separate sales forces. The reorganisation resulting from this change resulted in a temporary loss in sales for a few months. In Model (g) the introduction of the two sales forces is represented by adding a further intervention variable into Model (b), the best model obtained previously. It is seen that the introduction of the two sales forces has a significant negative effect on sales in the period after its introduction. Model (g) shows that the inclusion of just two variables, namely the occurrence of trade promotions and the introduction of the two sales forces, has resulted in a very big reduction in residual standard deviation, as compared with the univariate model (a).

Figure 31(b) shows forecasts of the renormalised sales (that is the forecasts of the normalised sales multiplied by the number of delivery days per accounting period) based on model (g), using data up to a time origin 1976, period 12 and future knowledge that promotions would be held in certain periods. Also shown are the actual sales during the first seven of these months, when they came to hand.

3.4. *An application of multivariate stochastic models to the sales of competitive products*

This analysis was part of a wider study of competition between several products in the European market. Figure 32 shows a plot of the logarithms of quarterly sales data for two such products, which are designated Product 1 and

Product 2. The data had been previously seasonally adjusted using a standard package, against the advice of the writer. Some consequences of this act are described later.

Figure 32 shows that both products are characterised by sustained growth which is approximately linear in the logarithms and hence exponential for the original series. The range-mean plots (not shown) for the two products revealed that there was a well defined linear dependence between the range and mean for both series, suggesting that a logarithmic transformation would be appropriate in each case. (See Appendix A.1 for a further discussion of data transformations). In this example a logarithmic transformation not only ensures that the scatter of the transformed variables is independent of the level, but also converts the approximately exponential growth into approximately linear growth.

There was prior evidence that a transfer function model, (implying a unidirectional relationship between the two series) would not be appropriate in this case since it was known that sales of Product 1 affected sales of Product 2 and vice versa. Hence, it was assumed initially that a multivariate stochastic model would be more appropriate than a transfer function model to describe this situation. If such an assumption turned out to be incorrect, then the model building process should reveal that a transfer function is more appropriate, in which case a transfer function model could then be fitted.

The first step in a multivariate stochastic analysis is to build univariate models for each time series separately (another illustration of 'learning to grovel before attempting to walk' as emphasised in Section 2.6). Denoting the time series for sales of Product 1 by z_{1t} and the sales of Product 2 by z_{2t}, the usual identification, estimation and checking stages for univariate models (see Box and Jenkins, 1970) led to the following univariate models fitted separately to each series:

$$\underset{}{\nabla \ln z_{1t}} = \underset{\pm 0.010}{0.056} + \alpha_{1t} \tag{21}$$

where α_{1t} is a random series with variance $\sigma_1^2 = 0.002991$ (per cent standard deviation = 5.5%) and

$$\nabla \ln z_{2t} = \underset{\pm 0.002}{0.042} + (1 - \underset{\pm 0.11}{0.70}B)\alpha_{2t} \tag{22}$$

where α_{2t} is a random series with variance $\sigma_2^2 = 0.002044$ (per cent standard deviation = 4.5%).

Figure 33 shows the prewhitened cross correlation function between the two series, that is the cross correlation function between α_{1t} and α_{2t} at positive and

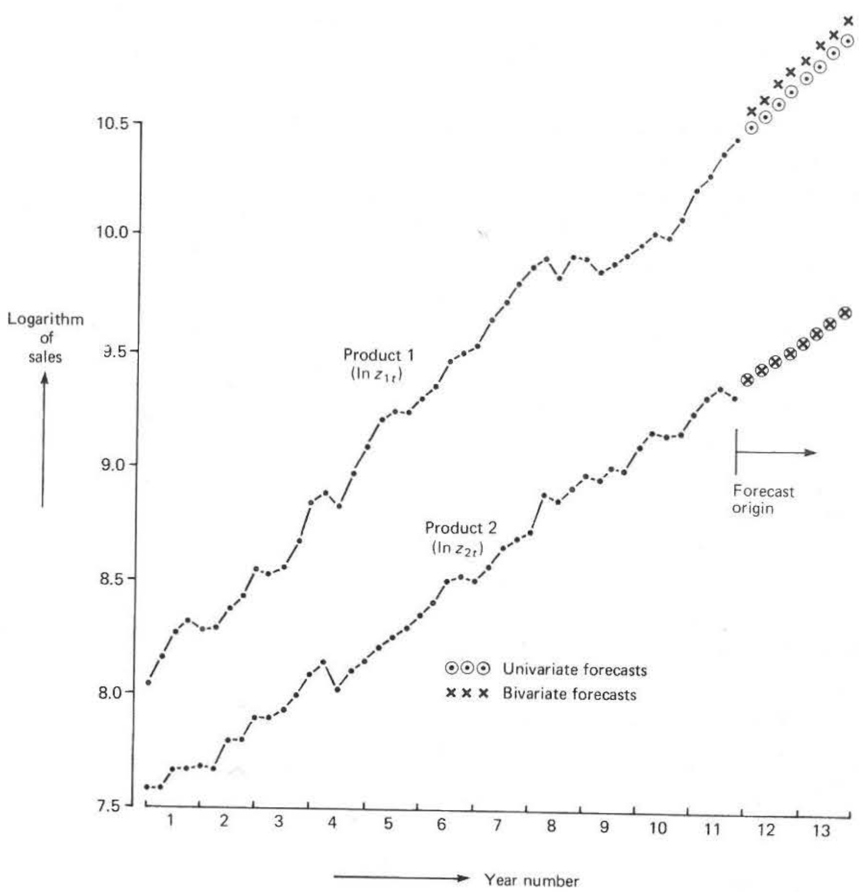

Fig 32. **Sales of two competitive products, together with forecasts for lead times 1, 2, . . . , 8 from origin $t = 44$; quarterly observations for 11 years.**

negative lags. The two correlations which stand out above the two standard error limits are at lags $k = -1$ and $+1$. The positive correlation at lag $k = +1$ implies that there is a tendency for an increase in the sales of Product 1 this quarter to be accompanied by an increase in sales of Product 2 in the next quarter. The negative correlation at lag $k = -1$ implies that there is a tendency for an increase in sales of Product 2 this quarter to be accompanied by a decrease in sales of Product 1 in the next quarter. This result stems from the fact that Product 2 can genuinely be substituted for Product 1 but not vice versa – the end use for Product 1 requires that Product 2 be also used.

where

$$y_t = \nabla^{dY} Y_t^{(\lambda_Y)}, \qquad dY = d + dN,$$

$$x_t = \nabla^{dX} X_t^{(\lambda_X)}, \qquad dX = d_1 + dN.$$

Proceeding as in ordinary regression analysis, the estimation of the constant c can be made approximately orthogonal to the other model parameters if the x-series is mean corrected, that is

$$y_t = c + \frac{\omega(B)}{\delta(B)} (x_{t-b} - \bar{x}) + \frac{\theta(B)}{\phi(B)} a_t. \tag{A.9}$$

Multiple input models. If several input variables $X_{1t}, X_{2t}, \ldots, X_{lt}$ are to be related to the output Y_t, then (A.9) may be generalised to

$$y_t = c + \sum_{j=1}^{l} \frac{\omega_j(B)}{\delta_j(B)} (x_{j,t-b_j} - \bar{x}_j) + \frac{\theta(B)}{\phi(B)} a_t \tag{A.10}$$

each X-variable having a transfer function with its own moving average operator $\omega_j(B)$, autoregressive operator $\delta_j(B)$ and pure delay b_j.

Seasonal models. Finally, if Y_t and the X_{it} are seasonal series with seasonal period s, we obtain a seasonal transfer function model

$$\nabla^d Y_t^{(\lambda_Y)} = \sum_{j=1}^{l} \frac{\omega_j(B)}{\delta_j(B)} \nabla^{d_j} X_{jt}^{(\lambda_{X_{j,t}})} + N_t$$

where

$$\nabla^{dN} \nabla_s^{DN} N_t = \frac{\theta(B)\Theta(B^S)}{\phi(B)\Phi(B^S)} a_t. \tag{A.11}$$

The model (A.11) can also be written in the form (A.10) with suitable definition of y_t, x_t and the inclusion of seasonal operators in the noise.

Figure A.5 shows a filter representation of the seasonal multiple input transfer function model (A.11) and Figure A.6 a flow diagram for building transfer function models. For further details of transfer function model building, see Box and Jenkins (1970).

A.4. *Intervention models*

During the course of building the four main types of model described in these Appendices, it may be necessary to introduce further modifications to the model

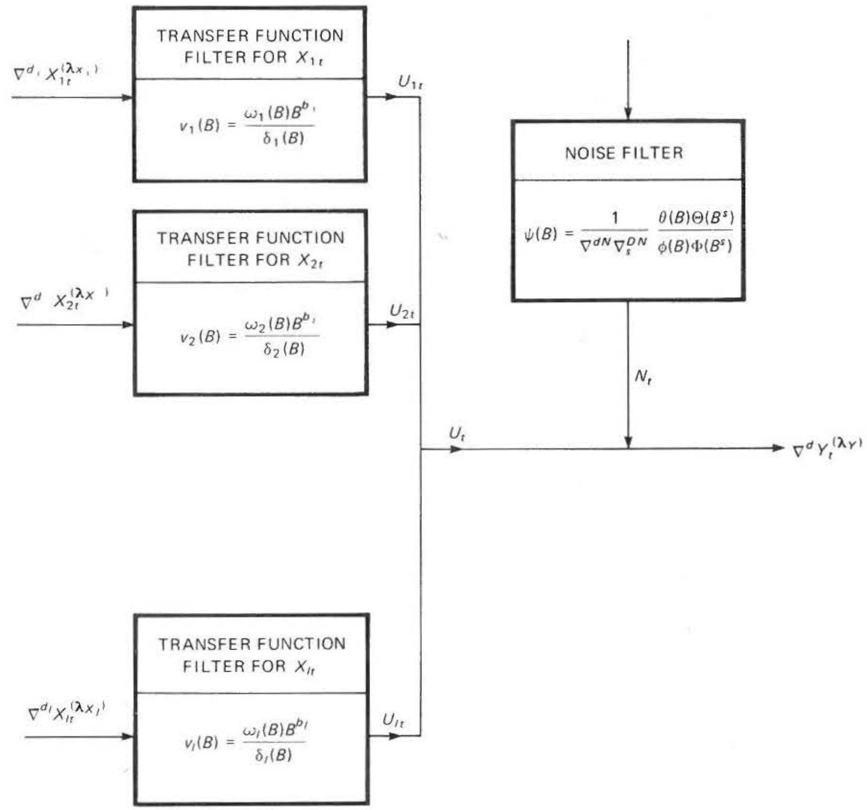

Fig A.5. Filter representation of seasonal multiple input transfer function model.

to deal with abnormal events or other forms of effects which are not easy to quantify. The following *dummy variables* are useful for representing such behaviour:

(i) *pulse* variables, which take on the value '1' when an anomalous event (such as a holiday or a strike) occurs and are '0' everywhere else,

(ii) *step* variables, which take on the value '0' *before* a change (such as a policy change, or a new law, or a change in definition in an economic variable) and the value '1' *after* such a change.

To investigate the effect of such an intervention variable ξ_t on the variable being modelled, we may postulate a lag structure

$$Y_t^{(\lambda)} = \frac{\omega(B)}{\delta(B)} \xi_{t-b} \tag{A.12}$$

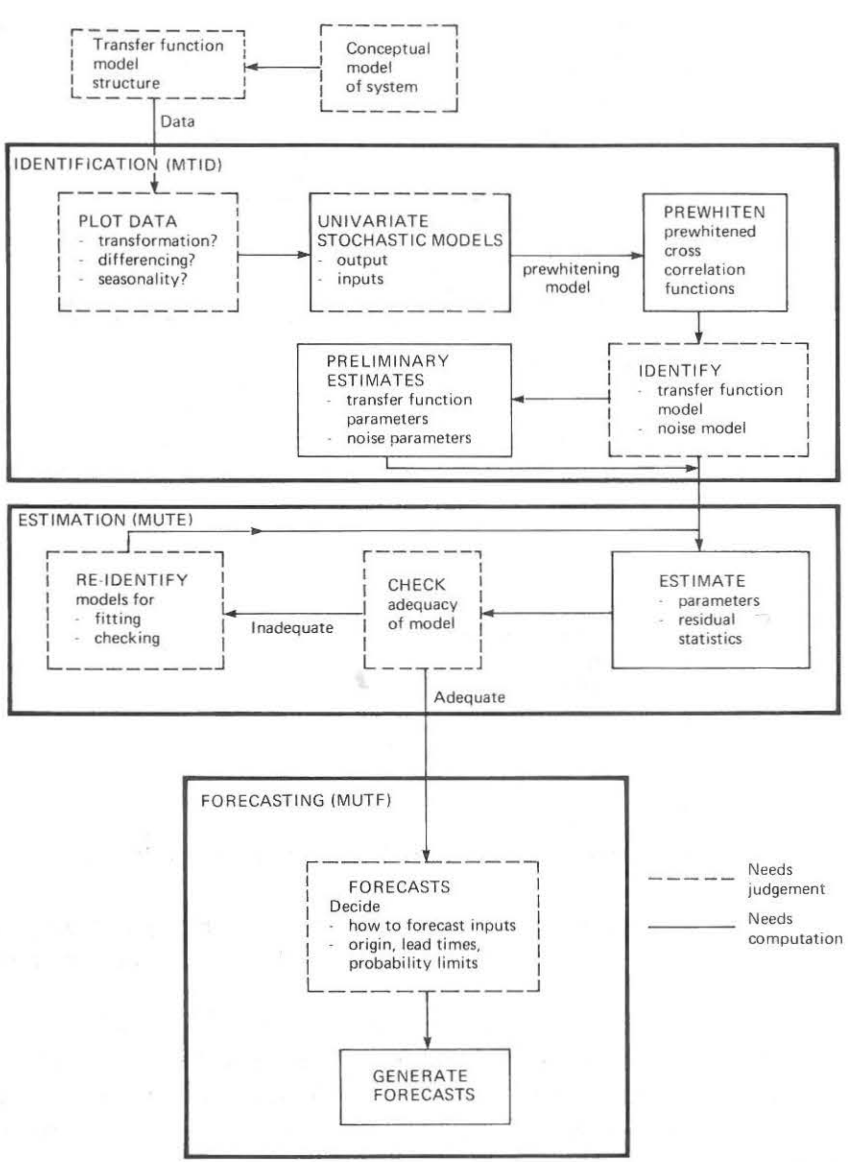

Fig A.6. Flow diagram for transfer function model building and forecasting, based on three computer programs: MTID (Multiple Input Transfer Function Identification), MUTE (Multiple Input Transfer Function Estimation), MUTF (Multiple Input Transfer Function Forecasting).

whose parameters can be estimated as in a transfer function model.

It is not possible to use prewhitening to identify the structure of an intervention model (A.12) as in the case of transfer function models. Instead the following guidelines may be used in complement with each other to identify intervention models:

(1) As a result of a known external event, inspection of the data itself may suggest ways in which that event has changed the course of the series. For example, inspection of many consumer price indices will indicate that their rate of change has increased as a result of the dramatic oil price increase in the last quarter of 1973. This effect can be modelled by introducing an intervention term into an existing model as follows:

$$\nabla Y_t^{(\lambda)} = \omega_0 \xi_t \tag{A.13}$$

where ξ_t is a step function jumping from '0' to '1' at the point of the so-called 'oil crisis'.

(2) Supplementary evidence may also be obtained by examining the residuals from the model fitted before an intervention variable is introduced. For example, a large negative residual followed by a large positive residual in a univariate model may be due to a loss of sales during the period of a 'strike' and a catching up in deliveries in the period following the strike. Such an effect may be described by the model:

$$Y_t^{(\lambda)} = (\omega_0 - \omega_1 B)\xi_t \tag{A.14}$$

where ξ_t is a pulse of unit height at the point where the strike occurred.

(3) Whereas (1) and (2) are useful in providing visual clues that an abnormal change has taken place in a series, a better way to postulate an intervention model is to discuss the *mechanisms* that might be causing the change with those who have some understanding of the situation being considered. If the postulated model is too elaborate, those parameters which are small compared with their standard errors in the fitted model can be omitted and the simpler model fitted. Alternatively, if it is suspected that the model is inadequate, further terms can be added. Figure A.7 shows further examples of the effects on the Y_t variables in (A.12) which can be modelled by simple transfer function models when the intervention variable ξ_t is a step or pulse.

Introducing the noise into an intervention model. To complete the specification of an intervention model, it is necessary to assume a structure for the behaviour of the series that would have occurred if no abnormal event had taken place.

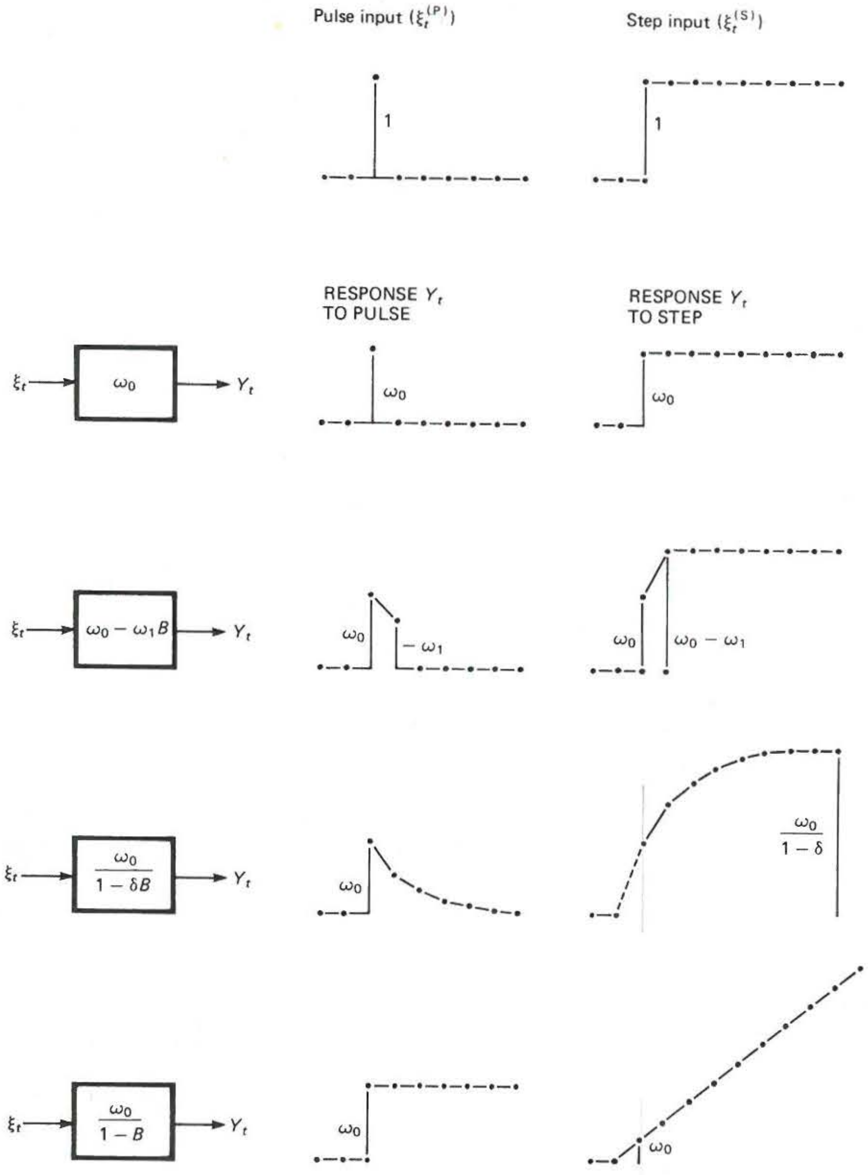

Fig A.7. Examples of dynamic effects which can be simulated in intervention analysis using a 'pulse' input and step input.

To illustrate the approach, suppose that a univariate model:

$$\nabla\nabla_{12} Y_t^{(\lambda)} = (1 - \theta B)(1 - \Theta B^{12}) a_t \tag{A.15}$$

had been fitted to a series, preferably to data not including the period when the abnormal event occurred. Then, if the intervention mechanism was (A.14), we could postulate a model

$$Y_t^{(\lambda)} = (\omega_0 - \omega_1 B)\xi_t + N_t \tag{A.16}$$

where N_t is a noise term describing the behaviour of the series in the absence of the abnormal event. If ω_0 and ω_1 were both zero in (A.16) then $Y_t^{(\lambda)} = N_t$. Hence, we can use the univariate model (A.15) for $Y_t^{(\lambda)}$ as a first guess of the structure of N_t, that is

$$\nabla\nabla_{12} N_t = (1 - \theta B)(1 - \Theta B^{12}) a_t. \tag{A.17}$$

Combining (A.17) with (A.16), we obtain the overall intervention model

$$\nabla\nabla_{12} Y_t^{(\lambda)} = (\omega_0 - \omega_1 B)\nabla\nabla_{12}\xi_t + (1 - \theta B)(1 - \Theta B^{12}) a_t.$$

In contrast, the intervention model corresponding to the mechanism (A.13) could be formulated as

$$\nabla Y_t^{(\lambda)} = \omega_0 \xi_t + N'_t. \tag{A.18}$$

Again, setting $\omega_0 = 0$, the initial structure for N'_t could be set equal to the structure of $\nabla Y_t^{(\lambda)}$. Using (A.15), this is

$$\nabla_{12} N'_t = (1 - \theta B)(1 - \Theta B^{12}) a_t. \tag{A.19}$$

Combining (A.18) with (A.19), we obtain the overall intervention model

$$\nabla\nabla_{12} Y_t^{(\lambda)} = \omega_0 \nabla_{12}\xi_t + (1 - \theta B)(1 - \Theta B^{12}) a_t.$$

Further examples of identifying intervention models are given in Section 3.3.

In general, several terms of the form (A.12) may need to be introduced into one of the models described in these Appendices. For example, a pulse may be needed at several points to allow for the effect of different holidays on energy consumption.

A.5. Multivariate stochastic models

Pioneering work in the area of multivariate stochastic models was carried out by Bartlett (1950, 1953) and Quenouille (1957). Quenouille's approach was to generalise the univariate autoregressive-moving average models of Appendix A.2 to

$$\boldsymbol{\phi}_0 z_t = \boldsymbol{\phi}_1 z_{t-1} + \cdots + \boldsymbol{\phi}_p z_{t-p} + \boldsymbol{\theta}_0 a_t - \boldsymbol{\theta}_1 a_{t-1} - \cdots - \boldsymbol{\theta}_q a_{t-q} \qquad (A.20)$$

where z_t is a column vector whose transpose $z_t' = (z_{1t}, z_{2t}, \ldots, z_{mt})$ is a row vector of m time series, the $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_j$ are $m \times m$ autoregressive and moving average matrices respectively and the elements a_{it} of the vector a_t are mutually uncorrelated at all times. Whereas, the Quenouille model (A.20) provides a useful starting point, it suffers from two practical disadvantages:

(i) The formulation in (A.20) constrains the univariate models for the individual time series z_{it} to have autoregressive operators which

(a) have the same order,

(b) have the same parameter values.

This constraint is highly undesirable and can be removed by writing the model in the form (Alavi, 1973)

$$\begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) \cdots \phi_{1m}(B) \\ \phi_{21}(B) & \phi_{22}(B) \cdots \phi_{2m}(B) \\ \cdot & \cdot \quad \cdots \quad \cdot \\ \cdot & \cdot \quad \cdots \quad \cdot \\ \cdot & \cdot \quad \cdots \quad \cdot \\ \phi_{m1}(B) & \phi_{m2}(B) \quad\; \phi_{mm}(B) \end{bmatrix} \begin{bmatrix} z_{1t} - c_1 \\ z_{2t} - c_2 \\ \cdot \\ \cdot \\ \cdot \\ z_{mt} - c_m \end{bmatrix} = \begin{bmatrix} \theta_{11}(B) \; \theta_{12}(B) \cdots \theta_{1m}(B) \\ \theta_{21}(B) \; \theta_{22}(B) \cdots \theta_{2m}(B) \\ \cdot \quad \cdot \quad \cdots \quad \cdot \\ \cdot \quad \cdot \quad \cdots \quad \cdot \\ \cdot \quad \cdot \quad \cdots \quad \cdot \\ \theta_{m1}(B) \; \theta_{m2}(B) \ldots \theta_{mm}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \\ \cdot \\ \cdot \\ \cdot \\ a_{mt} \end{bmatrix}$$

or

$$\boldsymbol{\phi}(B)(z_t - c) = \boldsymbol{\theta}(B) a_t \qquad (A.21)$$

where the autoregressive operator $\phi_{ij}(B)$ is a polynomial of degree p_{ij} in the backward shift operator B and the moving average operator $\theta_{ij}(B)$ is a polynomial of degree q_{ij} in B. As in previous models, the time series z_{it} in the vector z_t in (A.21) may need to be transformed to $z_{it}^{(\lambda_i)}$ before analysis.

In (A.21) the polynomials in the diagonal positions start with unity while the polynomials in the off-diagonal positions start with a power of B. With this formulation, the a_{it} are the one-step-ahead forecast errors which then must be allowed to have a covariance matrix at simultaneous times

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \cdots \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 \cdots \sigma_{2m} \\ \sigma_{m1} & \sigma_{m2} \ldots \sigma_m^2 \end{bmatrix} \qquad (A.22)$$

with $\sigma_{ij} = \sigma_{ji}$, but are otherwise mutually uncorrelated at non-simultaneous times.

The model (A.21) will be referred to as a multivariate autoregressive-moving average model or ARMA $(\boldsymbol{P}, \boldsymbol{Q})$ model. During the model building process the degrees of the operators $\phi_{ij}(B)$ and $\theta_{ij}(B)$ can be adjusted so that the univariate models for the individual time series accurately describe the behaviour of each series and are not automatically constrained as in the model (A.20).

(ii) A further disadvantage of the model (A.20) is that it assumes stationarity, that is the m time series are in statistical equilibrium about fixed means. To describe series which have stochastic trends, the model (A.21) may be generalised to

$$\boldsymbol{\phi}(B)(w_t - c) = \boldsymbol{\theta}(B) a_t$$

where

$$w_t' = (\nabla^{d_1} z_{1t}^{(\lambda_1)}, \ldots, \nabla^{d_m} z_{mt}^{(\lambda_m)}) \qquad (A.23)$$

with $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ as defined in (A.21) and c a vector of constants. Such a model will be referred to as a *multivariate* ARIMA $(\boldsymbol{P}, \boldsymbol{d}, \boldsymbol{Q})$ model, where the matrices $\boldsymbol{P} = (p_{ij})$, $\boldsymbol{Q} = (q_{ij})$ determine the degrees of the polynomials in the autoregressive and moving average matrices and the row vector $\boldsymbol{d}' = (d_1, d_2, \ldots, d_m)$ has elements corresponding to the degrees of differencing required to induce stationarity in each of the individual time series.

Seasonal multivariate stochastic models. As for univariate and transfer function models it may be necessary to difference a series both seasonally as well as non-seasonally to induce stationarity. Thus the elements w_{it} of the vector w_t in (A.23) will need to be defined as

$$w_{it} = \nabla^{d_i} \nabla_s^{D_i} z_{it}^{(\lambda_i)}. \qquad (A.24)$$

One possible way to introduce seasonality into the main part of the multivariate stochastic model (A.23) is to write the autoregressive and moving average matrices as products of non-seasonal and seasonal matrices, by analogy with the univariate case (A.4). However, certain theoretical difficulties have been experienced with this formulation. The way in which this problem has been resolved in practice is to allow the autoregressive and moving average operators in the matrices in (A.23) to be non-multiplicative. Thus, the computer programs necessary for building multivariate stochastic models have been written to allow an autoregressive operator $\phi_{ij}(B)$ in (A.21) to be written, for example, as

$$\phi_{ij}(B) = 1 - \phi_{ij,1} B - \phi_{ij,12} B^{12} - \phi_{ij,13} B^{13}$$