

# The Use of Transfer Function Models, Intervention Analysis and Related Time Series Methods in Epidemiology

ULRICH HELFENSTEIN

Helfenstein U (Biostatistical Center, Institute of Social and Preventive Medicine, University of Zurich, Sumatrastrasse 30, 8006, Zurich, Switzerland). The use of transfer function models, intervention analysis and related time series methods in epidemiology. *International Journal of Epidemiology* 1991, **20**: 808–815.

In epidemiology, data often arise in the form of time series e.g. notifications of diseases, entries to a hospital, mortality rates etc. are frequently collected at weekly or monthly intervals. Usual statistical methods assume that the observed data are realizations of independent random variables. However, if data which arise in a time sequence have to be analysed, it is possible that consecutive observations are dependent. In environmental epidemiology, where series such as daily concentrations of pollutants were collected and analysed, it became clear that stochastic dependence of consecutive measurements may be important. A high concentration of a pollutant today e.g. has a certain inertia i.e. a tendency to be high tomorrow as well.

Since the early 1970s, time series methods, in particular ARIMA models (autoregressive integrated moving average models) which have the ability to cope with stochastic dependence of consecutive data, have become well established in such fields as industry and economics. Recently, time series methods are of increasing interest in epidemiology.

Since these methods are not generally familiar to epidemiologists this article presents their basic concepts in a condensed form. This may encourage readers to consider the methods described and enable them to avoid pitfalls inherent in time series data. In particular, the following topics are discussed: Assessment of relations between time series (transfer function models). Assessment of changes of time series (intervention analysis), forecasting and some related time series methods.

In epidemiology data often arise in the form of 'time series': Notifications of diseases, entries in a hospital, mortality rates etc. are frequently collected at weekly or monthly intervals. Usual statistical methods assume that the observed data are realizations of independent random variables. However, if data which arise in a time sequence have to be analysed, it is possible that consecutive observations are dependent. Even though time series models have been studied for many years, they have belonged to the domain of theoretical statistics. This situation changed when Box and Jenkins provided a method for actually constructing time series models in practice.<sup>1</sup> Their method of identification, estimation and checking is now often referred to as 'the Box-Jenkins approach' and the (autoregressive integrated moving average) ARIMA models are also called Box-Jenkins models. Since the first edition of

their book in 1970, many time series data arising in such fields as industry and economics have been analysed by this method.

Armitage was probably the first to suggest the application of Box-Jenkins models to epidemiological time series data.<sup>2</sup> Recently, there has been increasing interest in time series methods in epidemiology. This may be reflected in a recent issue of the journal *Statistics in Medicine* (March 1989) which contained three articles<sup>3–5</sup> concerned with epidemiological time series analysis.

In environmental epidemiology where series such as daily concentrations of pollutants were analysed, it became evident that stochastic dependence of consecutive measurements may not be disregarded. A high concentration of a pollutant today e.g. has a certain inertia i.e. a tendency to be high tomorrow as well (positive autocorrelation). Observations often show a stronger correlation when the time interval between them becomes shorter. This appears to be plausible

Biostatistical Center, Institute of Social and Preventive Medicine, University of Zurich, Sumatrastrasse 30, 8006 Zurich, Switzerland.

when considering e.g. an air pollutant or a climatic variable. If the second measurement is taken one minute after the first it is likely that the concentration of the pollutant does not change notably. Thus, this second observation may not add much information to what is already known from the first one. It is well known that the variance of the mean  $\bar{y}$  of  $n$  uncorrelated random variables  $y_i$  ( $i = 0, 1, \dots, n$ ) with equal variances  $\text{Var}(y_i) = \sigma^2$  is given by:

$$\text{Var}(\bar{y}) = \sigma^2/n \quad (1)$$

However, if the variables are correlated it may be shown that the variance of the mean is given by:

$$\text{Var}(\bar{y}) = \sigma^2/n \left( 1 + 2/n \sum_{i=1}^n \sum_{j=i+1}^n \rho_{ij} \right) \quad (2)$$

where  $\rho_{ij}$  is the correlation between  $y_i$  and  $y_j$ . The two formula above show how the variance may be inflated if data are positively correlated. In the (theoretical) extreme case where all  $\rho_{ij}$  are equal to 1 the term in brackets equals  $n$  and the variance of the mean is  $\text{Var}(\bar{y}) = \sigma^2$ , i.e. the same as the variance of a single observation. Box, Hunter and Hunter<sup>6</sup> present an example (with  $n = 10$ ) which illustrates the effect of positive or negative correlation when observations follow moving average processes of first order (compare next section); in this case the term in brackets in equation (2) may vary by a factor 19. In case of autoregressive processes of first order, the effect of dependency of the observations on the variance of the mean may be even larger. Similar deliberations apply if other standard methods (comparison of means, regressions etc.) designed to analyse independent observations are applied to dependent time series data.

Since time series methods are not generally familiar to epidemiologists this article presents their basic concept in a condensed form. This presentation has necessarily to be restricted to a reasonable length; therefore it is not possible to provide an exhaustive discussion of details. However, it is thought that a summary may encourage readers to consider the methods described and also enable them to avoid pitfalls inherent in time series data. The amount of calculation involved is no longer a drawback to these methods since efficient computer programs are available (SAS,<sup>7</sup> BMDP<sup>8</sup>).

Many situations which may be adequately represented by the so called 'transfer function model' arise in epidemiology. One is often interested in the assessment of relations between a target or output series such as daily number of patients coming to a clinic, and explanatory or input series such as daily concentrations of a pollutant, daily mean temperature or other climatic series.

Other frequently arising questions are concerned with changes of time series. Changes may be man-made or they may arise naturally: How efficient was a preventive programme to decrease the monthly number of accidents? How did the pattern of morbidity in a population change after an environmental accident? These and many other related questions may be investigated by the so-called intervention analysis proposed by Box and Tiao.<sup>9</sup>

Forecasts of epidemiological time series are needed for many reasons: For public health organizations, it is of interest to know what frequencies of diseases have to be expected in the future in order to plan for the distribution of resources. Forecasting can also be used as a complementary method to intervention analysis.<sup>10,11</sup> A forecast obtained from data before the intervention may be compared with actual data obtained after the intervention.

In the following section the so-called ARIMA model is briefly presented. It serves as the framework for what follows. Thereafter, the transfer function model, intervention analysis and some additional related time series methods are discussed. Finally suggestions for further reading are given which appear to be of particular interest to epidemiologists and which may allow a deeper insight into time series methods.

### THE ARIMA MODEL

In order to give precise meaning to concepts and examples presented in the sequel, let  $\dots, z_{t-1}, z_t, z_{t+1}, \dots$  denote the observations (number of entries to a hospital, concentration of a pollutant, etc.) at times  $\dots, t-1, t, t+1, \dots$  (e.g. yesterday, today, tomorrow etc.). For simplicity assume that the mean value of  $z_t$  is zero (otherwise the  $z_t$  may be considered as deviations from their mean). Let  $a_{t-1}, a_t, a_{t+1}, \dots$  be a white noise series consisting of independent identically distributed random variables with mean zero and variance  $\sigma_a^2$ . It may be helpful to think of them as random shocks.

To begin, assume that the present observation  $z_t$  is linearly dependent on the previous observation  $z_{t-1}$  and on (the random shock)  $a_t$ :

$$z_t = \phi z_{t-1} + a_t, \text{ where } \phi \text{ is a parameter.} \quad (3)$$

The expression resembles an ordinary regression equation. Since  $z_t$  is regressed on  $z_{t-1}$  it is called an autoregressive model of first order, abbreviated AR(1) model.

Alternatively one may express  $z_t$  as a linear combination of the present and the previous random shock:

$$z_t = a_t - \theta a_{t-1}, \text{ where } \theta \text{ is a parameter.} \quad (4)$$

This expression is called a moving average model of first order, abbreviated MA(1) model.

The above two basic models are special cases of two more general models:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t \quad (5)$$

called an autoregressive model of order  $p$  (AR( $p$ ) model) and

$$z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (6)$$

called a moving average model of order  $q$  (MA( $q$ ) model).

By combining (5) and (6) one obtains what is called the autoregressive moving average model of order  $p$  and  $q$  (ARMA( $p, q$ ) model):

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (7)$$

An important concept is that of *stationarity*, which implies that the probability structure of the series does not change with time. In particular, a stationary series has a constant mean. It is well known that many epidemiological time series are not stationary; they often exhibit trends. However, experience in economics and more recently in epidemiology has shown that the series of differences  $\nabla z_t = z_t - z_{t-1}$  is often stationary. In this case the original series  $z_t$  is called an integrated ARMA or an ARIMA model. Box and Jenkins have extended the above concepts to cope with series containing seasonal variations. In particular, a monthly series with seasonal non-stationarity may be transformed to stationarity by taking seasonal differences:

$$\nabla_{12} z_t = z_t - z_{t-12}.$$

The dependence structure of a stationary time series  $z_t$  is described by the autocorrelation function (ACF). This function determines the correlation between  $z_t$  (the present value of the series) and  $z_{t+k}$  (the value  $k$  time units later):

$$\rho_k = \text{correlation}(z_t, z_{t+k}). \quad (8)$$

$k$  is called the time lag. The empirical ACF is the main tool for the identification of the model. The ACF of an AR(1) process e.g. follows an exponential curve, whereas the ACF of a MA(1) process shows a single peak at time lag  $k = 1$ .

In order to identify an ARIMA model Box and Jenkins<sup>1</sup> have suggested the following iterative procedure: One starts with a provisional model which may be chosen by looking at the autocorrelation function. Then the parameters of the model are estimated and the adequacy of the model is checked. If the model does not fit the data adequately<sup>12</sup> one goes back to the start and chooses an improved model. Among different models which represent the data equally well one chooses the simplest one, i.e. the model which contains

the least amount of parameters (principle of parsimony).

ARIMA-models are needed for several reasons: They provide e.g. a means to identify transfer function models (compare next section). In addition, the resulting residual variance  $\sigma_a^2$  provides a yardstick which may be compared with residual variances of more elaborate models.<sup>13</sup> As will be demonstrated, it is easy to calculate forecasts from ARIMA models.

## THE TRANSFER FUNCTION MODEL

One is often interested in the assessment of relations between a target or output series (e.g. daily number of patients with a specified disease coming to a clinic) and one or several explanatory or input series (e.g. daily concentrations of pollutants, daily mean temperature etc). Figure 1 gives a schematic diagram of the transfer function model which represents such a situation.

In Figure 1,  $y_t, y_{t+1}, \dots$  denote the observations (number of entries, death rates etc) at times  $t, t+1, \dots$ . The output series  $y_t$  is considered to be composed of two parts:

$$y_t = u_t + n_t \quad (9)$$

$u_t$  is the part which may be explained in terms of the input variable  $x_t$  (concentration of a pollutant, etc).  $n_t$  is an error or noise process, which describes the unexplained part of  $y_t$ .

It is assumed that the explained part  $u_t$  is given by a weighted sum of the present and of past values of  $x_t$ :

$$u_t = v_0 x_t + v_1 x_{t-1} + \dots, \quad (10)$$

and  $n_t$  is an ARIMA process as described in the previous section.  $v_0, v_1, \dots$  are called transfer function weights.

The relation between two time series  $x_t$  and  $y_t$  is determined by the cross-correlation function (CCF):

$$\rho_{xy}(k) = \text{correlation}(x_t, y_{t+k}), \quad k = 0, \pm 1, \pm 2, \dots \quad (11)$$

This function determines the correlation between the two series as a function of the time lag  $k$ . The main tool to identify a transfer function model is the empirical CCF  $r_{xy}(k)$ . However, a basic difficulty arises in the interpretation of the empirical CCF. As shown by Bartlett<sup>14</sup> and by Box and Newbold<sup>15</sup> the empirical CCF between two completely unrelated time series which are themselves autocorrelated can be very large due to chance alone. In addition, the cross-correlation estimates at different lags may be correlated. This is due to the autocorrelation within each individual series.

Box and Jenkins<sup>1</sup> proposed a way out of this difficulty called 'prewhitening'. The ARIMA model for the

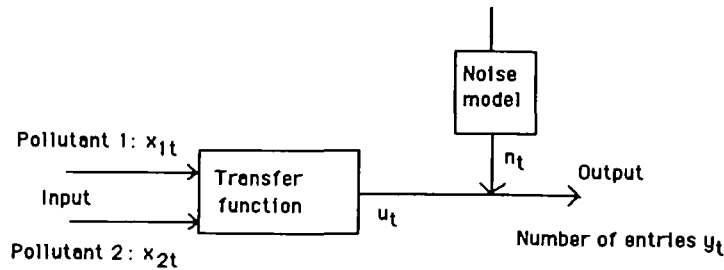


FIGURE 1 Schematic representation of a transfer function model with two inputs.

input series converts the correlated series  $x_t$  into an approximately independent series  $\alpha_t$ . Applying the identical operation to the output series  $y_t$  produces a new series  $\beta_t$ . The CCF between  $\alpha_t$  and  $\beta_t$  (the prewhitened CCF) shows at which lags input and output are related. Programs such as SAS<sup>7</sup> and BMDP<sup>8</sup> allow to calculate the prewhitened CCF easily. An alternative way of prewhitening has been described by Haugh<sup>16</sup> and by Haugh and Box.<sup>17</sup>

Because prewhitening is considered to be of principal importance in the assessment of relations between time series, the following epidemiological example is presented to demonstrate its effect: During the winter period daily concentrations of sulphur dioxide (SO<sub>2</sub>) (and of many other environmental time series) were collected in Basle.<sup>18</sup> Simultaneously, the number of respiratory symptoms in a randomly selected group of preschool children was recorded. Figure 2(a) shows the CCF between the series SO<sub>2</sub> (logarithmically transformed) and the series of symptoms before prewhitening. This CCF is not interpretable: Significant coefficients are smeared over a large range of time lags and there are typical nonsense coefficients: Since, as may easily be shown,  $r_{xy}(k) = r_{yx}(-k)$ , the large positive coefficients at negative time lags signify that a high number of respiratory symptoms today is expected to be followed by high pollution during the next ten days!

Figure 2(b) shows the prewhitened CCF. After prewhitening, a marked peak is found at time lag 0 while all other coefficients are not significantly different from zero. This result suggested the following transfer function model for  $y_t$  and  $x_t$ :

$$y_t = v_0 x_t + n_t \quad (12)$$

where the differenced noise  $\nabla n_t = n_t - n_{t-1}$  was found to be adequately represented by a MA(1) model. Thus, the transfer function model revealed that input is related instantaneously with output; in particular, there is no delayed effect of the pollutant (no transfer function weight different from zero at non-zero time lags). Since the noise  $n_t$  is non-stationary, large variations

in the number of symptoms are to be expected even if no changes in the level of SO<sub>2</sub> would arise. It is interesting to note that in this study climatic variables (e.g. temperature) did not contribute to the explanation of the output series.<sup>18</sup>

### INTERVENTION ANALYSIS

Other frequently arising problems concern 'changes' of time series. Changes may be expected after a 'man-made intervention' e.g. the introduction of a preventive programme. Alternatively changes may occur e.g. after an environmental accident. Characteristic properties of changes of series may be investigated by means of intervention analysis.<sup>9</sup> The following basic model has been used to represent the possible effect of a preventive measure on the number of injuries:

$$y_t = \omega_0 s_t + n_t, \text{ with} \quad (13)$$

$$s_t = \begin{cases} 1 & \text{for } t \geq T \\ 0 & \text{for } t < T. \end{cases} \quad (14)$$

The step function  $s_t$  represents e.g. a preventive measure starting at time  $T$ .  $\omega_0$  is a parameter (the height of the step etc.) which measures the size or strength of the effect.  $n_t$  represents an ARIMA process as described above. Since  $n_t$  may be non-stationary, large changes of the series could occur even when no preventive measure takes place.

Alternatively the dummy input variable may be a pulse function  $p_t$ :

$$y_t = \omega_0 p_t + n_t, \text{ where} \quad (15)$$

$$p_t = \begin{cases} 1 & \text{for } t = T \\ 0 & \text{else.} \end{cases} \quad (16)$$

The pulse function  $p_t$  may represent e.g. an unusual event which acts only at time  $T$ .

Figure 3 shows basic intervention models. In the first line the two dummy input variables  $s_t$  and  $p_t$  are depicted. The lines (a), (b) and (c) below show responses corresponding to the models. Line (a) of

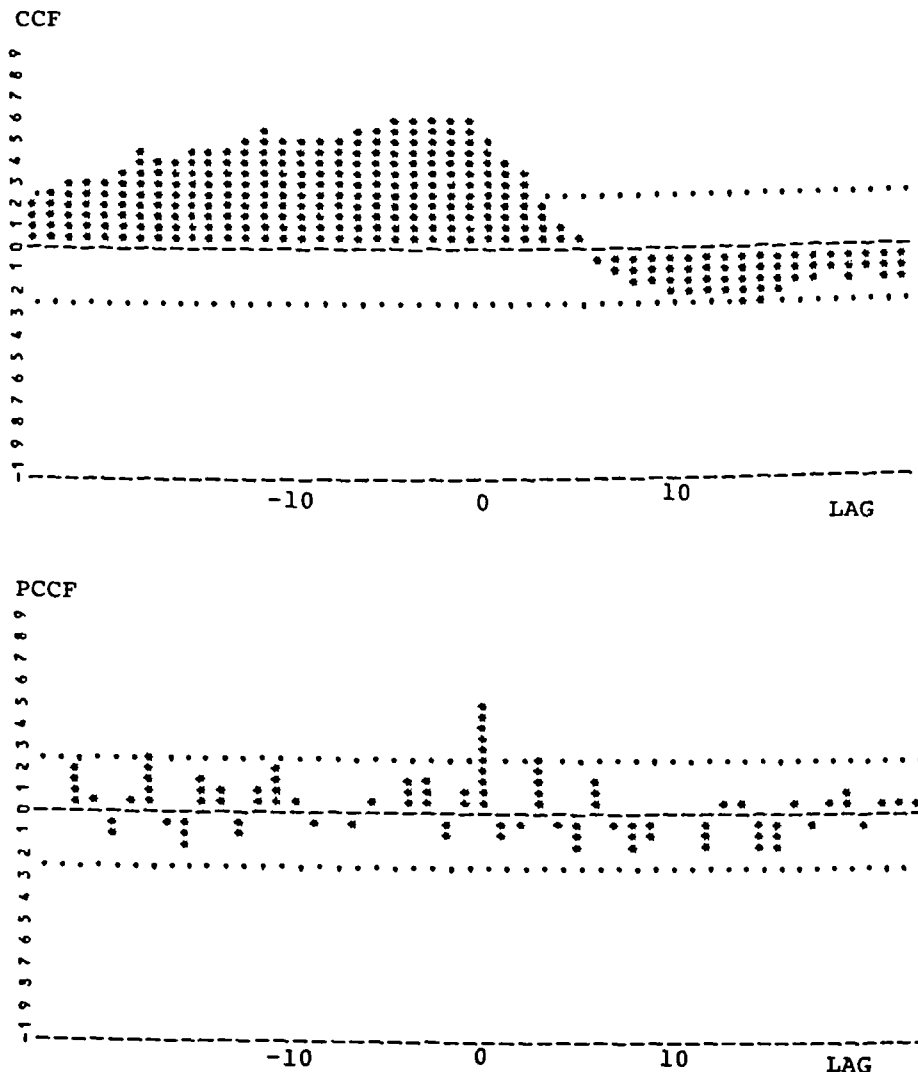


FIGURE 2 Cross-correlation function between the series sulphur dioxide ( $\text{SO}_2$ ) (logarithmically transformed) and the series of symptoms. (a) Before prewhitening (CCF). (b) After prewhitening (PCCF).

Figure 3 depicts the response corresponding to the above two models. In this case  $p_t$  (respectively  $s_t$ ) is simply multiplied by  $\omega_0$ .

It is possible that model (a) does not fully represent the characteristic properties of the series: with a step input the model predicts that the final level is reached immediately. However, it may be more realistic to consider the model in line (b) of Figure 3.

$$y_t = \omega_0 s_t - \omega_1 s_{t-1} + n_t \quad (17)$$

Here the final level is reached in two steps. This model may give a better fit to the data (i.e. a smaller residual variance) than model (a). Figure 3(c) shows another possible type of response. The final level is reached only gradually. This model is represented by:

$$y_t = \delta y_{t-1} + \omega_0 \xi_t + n_t. \quad (18)$$

Extensions of intervention models in many different directions are straightforward (compare last section). Of course it is desirable to include a comparison group (or area) with no intervention in the study design. Such studies have been performed e.g. in assessing the impact of preventive programmes (enforcement and media information campaign etc.) on the course of alcohol-related accidents.<sup>19,20</sup>

#### FURTHER RELATED TIME SERIES METHODS

The following example illustrates how forecasts may be calculated in a simple manner when an ARIMA model has been identified.

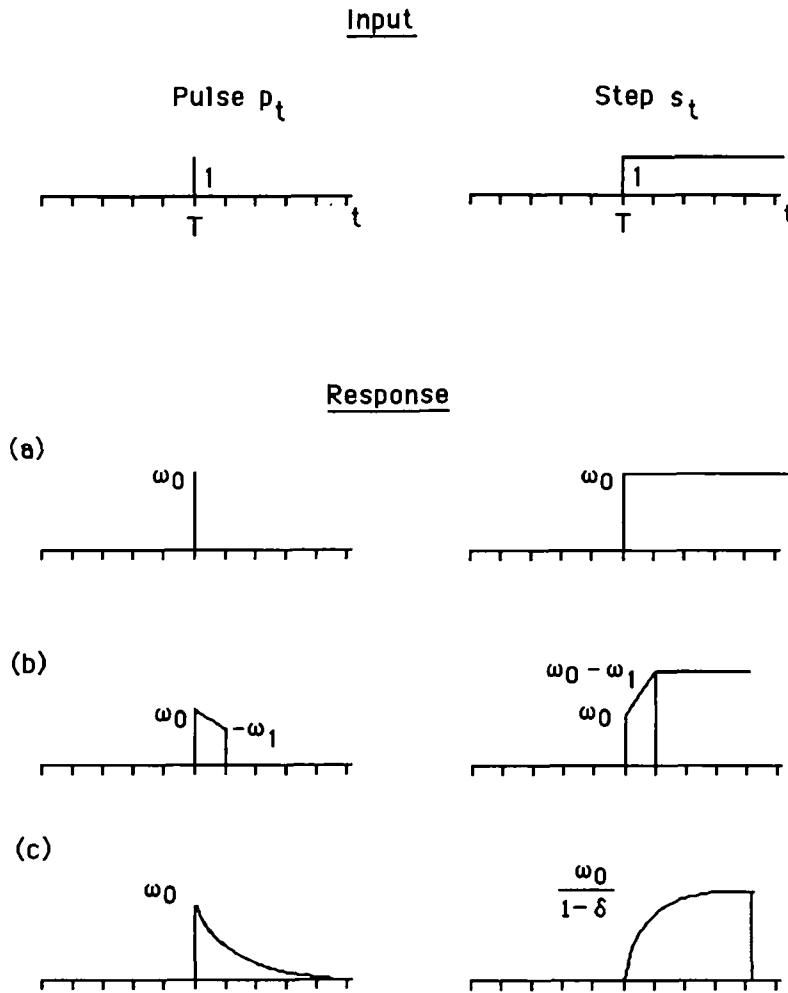


FIGURE 3 Responses to a unit pulse and step input. (a), (b), (c) Basic patterns of response.

To obtain forecasts  $\hat{z}_{t+1}$  for  $l$  time units (days, months, etc.) ahead one has only to write the corresponding model equation by replacing (a) future values of the random shocks  $a$  by zero, (b) future values of  $z$  by the corresponding forecasts and (c) past values of  $z$  by their observed values. The following example may illustrate the simplicity of obtaining forecasts e.g. for an AR(1) model:

$$l = 1: z_{t+1} = \phi z_t + a_{t+1}$$

$$\hat{z}_{t+1} = \phi z_t$$

$$l = 2: z_{t+2} = \phi z_{t+1} + a_{t+2}$$

$$\hat{z}_{t+2} = \phi \hat{z}_{t+1} = \phi(\phi z_t) = \phi^2 z_t \text{ etc.}$$

By continuing this procedure one may see that the forecasts corresponding to the AR(1) model follow an exponential curve.

Forecasts are not only of interest 'by itself'. In addition,

they provide a complementary way to investigate a possible change of a series.<sup>10</sup> One may compare forecasts arising from the pre-intervention series with actual data of the post-intervention series.

Both methods, intervention analysis and comparison of forecasts with actual data assume that the time of the change is well known. However, when a preventive programme is started e.g. it is often not clear when it starts to show an effect. A modification of the intervention analysis allows the most likely time of change to be found. For each possible time an intervention analysis is performed (Box-Tiao stepping) and the corresponding residual variance is calculated. The most likely time is then the time instant which provides the best model i.e. the model with the smallest residual variance.<sup>21</sup> A more advanced approach to identify such a 'change-point' may be found in Broemeling.<sup>22</sup> In



studying the possible effect of preventive programmes it has been observed that the expected effect may precede the actual time of intervention.<sup>21</sup> It is possible as well that the effect shows up only after a certain delay time.

#### CONCLUDING REMARKS AND FURTHER READING SUGGESTIONS

The elaboration of models described in the preceding sections in many different directions is straightforward. Intervention analysis with input consisting of multiple pulses for example may be used to analyse questions such as: Are there more deaths due to infarction in years with influenza A than in years without? The sequence of pulses then represents years with influenza A.

Responses to an intervention need not to occur instantaneously. A preventive programme may show an effect eventually after a delay. Occasionally such programmes may show an effect even before the official programme is started. In such cases one may try to estimate a possible pre-intervention effect.<sup>21</sup> In addition, effects of preventive programmes need not to show a permanent effect. Decomposing the dummy input in a short-term and a long-term component may help to decide if an effect is only transient.

Transfer function models and intervention analysis may be combined: Inversion of temperature for example may be represented by a dummy variable consisting of a sequence of 'ones' during the inversion preceded and followed by 'zeros'. This 0/1-input may be used simultaneously with continuous inputs representing e.g. concentrations of pollutants.

ARIMA models (and transfer function models) may be very useful for forecasting non-stationary series containing ordinary or seasonal trends. In particular, it may be possible to distinguish between models representing deterministic and stochastic polynomial trends. In a deterministic trend model the coefficients of the polynomial are constant over time. In a stochastic trend model the coefficients are subject to random variation and the trend changes stochastically according to random shocks that enter the system.<sup>23</sup>

Thus, time series methods are suited to represent a large amount of relevant epidemiological problems. However, one should be aware of limitations which are naturally inherent to these methods. Gruchow *et al.*<sup>24</sup> studied relations between alcohol consumption and ischaemic heart disease by calculating CCFs. They came to the conclusion that 'time series correlations of aggregated data are not useful for the study of latency periods . . .'. This conclusion was drawn after calculating CCFs between series containing trends, i.e. non-

stationary series. In order to get an interpretable CCF, the series have to be transformed to stationarity, i.e. the differencing operation has to be applied. Differencing eliminates trends; it has the effect of a high pass filter<sup>25</sup> and the information contained in the low frequency range is lost. However the presence of dominating low frequencies may make it impossible to detect short-term correlations. Differencing thus may allow the detection of hidden relations in the high frequency range.<sup>26</sup> It is also possible that the relation between two time series and the corresponding CCF change with time. In a study concerned with the assessment of relations between air pollutants and respiratory symptoms it was found that the CCF may change with season. During the winter period an instantaneous relation between the concentration of SO<sub>2</sub> and the number of symptoms was identified but no such relation was found for the other seasons. Thus, it may be advisable to fit separate models to different sub-series (time splitting<sup>13</sup>).

It is important to mention that an alternative way of analysing time series data is based on the auto-spectrum and on the cross-spectrum instead of the ACF and the CCF. The auto-spectrum and the ACF (and the cross-spectrum and the CCF) are mathematically equivalent since they constitute a Fourier transform pair.<sup>1</sup> However, they shed light on different complementary aspects of time series data. The auto-spectrum reveals which frequencies are present in an individual series and the cross-spectrum shows at which frequencies two series are related.<sup>27,28</sup> Box and Jenkins<sup>1</sup> use the ACF and not the auto-spectrum for the purpose of model identification because the parsimonious models they found to be useful in practice could be simply described in terms of the ACF.

Since the presentation in this article is necessarily short, the following suggestions for further reading may be helpful.

An extensive introduction to ARIMA models and their identification may be found in 'The Analysis of Time Series: An Introduction'.<sup>25</sup> A thorough presentation of transfer function models may be found in Chapter 8 of 'Statistical Methods of Forecasting',<sup>23</sup> where forecasting is also discussed in depth.

For a deeper study of specific epidemiological time series problems the following articles are thought to be of particular interest. Analysis of relations between seasonal series is discussed in.<sup>29,30</sup> Identification of seasonal ARIMA models representing infectious diseases is presented in some detail in.<sup>31</sup> Advantages of forecasting epidemiological series by means of ARIMA models over traditional methods are described in.<sup>32</sup>

Several investigators have used intervention analysis to assess changes of epidemiological series, in particular to analyse the efficiency of preventive measures.<sup>33-35</sup> A case study using intervention models to assess possible health effects of an environmental disaster may be found in.<sup>36</sup> Comparison of forecasts with actual data has been used to study changes in the frequency of traditional neurological diagnostic methods after the introduction of computer tomography.<sup>37</sup>

## REFERENCES

- <sup>1</sup> Box G E P, Jenkins G M. *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco: Holden Day, 1976.
- <sup>2</sup> Armitage P. *Statistical Methods in Medical Research*, Oxford and Edinburgh: Blackwell Scientific Publications, 1971; 347-48.
- <sup>3</sup> Katzoff M. The application of time series forecasting methods to an estimation problem using provisional mortality statistics. *Stat Med* 1989; 8: 335-41.
- <sup>4</sup> Schnell D, Zaidi A, Reynolds G. A time series analysis of gonorrhea surveillance data. *Stat Med* 1989; 8: 343-52.
- <sup>5</sup> Zaidi A, Schnell D, Reynolds G. Time series analysis of syphilis surveillance data. *Stat Med* 1989; 8: 353-62.
- <sup>6</sup> Box G E P, Hunter W G, Hunter J S. *Statistics for experimenters: An introduction to design, data analysis and model building*, New York: Wiley, 1978; pp 88-89.
- <sup>7</sup> Dixon W J (ed). *BMDP Statistical Software*, Berkeley: University of California Press, 1988.
- <sup>8</sup> SAS Institute Inc. *SAS/ETS User's guide, version 5 edition*. Cary, NC: SAS Institute Inc, 1984.
- <sup>9</sup> Box G E P, Tiao G C. Intervention analysis with applications to economic and environmental problems. *J Am Stat Assoc* 1975; 70: 70-9.
- <sup>10</sup> Box G E P, Tiao G C. Comparison of forecast and actuality. *Appl Stat* 1976; 25: 195-200.
- <sup>11</sup> Tiao G C, Box G E P, Hamming W J. Analysis of Los Angeles photochemical smog data: A statistical overview. *J Air Pollution Control Ass* 1975; 25: 260-8.
- <sup>12</sup> Ljung G M, Box G E P. On a measure of lack of fit in time series models. *Biometrika* 1978; 65: 297-303.
- <sup>13</sup> Jenkins G M. *Practical Experiences with Modelling and Forecasting Time Series*. St Helier: Gwilym Jenkins & Partners (Overseas) Ltd, 1979; p 14.
- <sup>14</sup> Bartlett M S. Some aspects of the time-correlation problem in regard to tests of significance. *J R Stat Soc* 1935; 98: 536-43.
- <sup>15</sup> Box G E P, Newbold P. Some comments on a paper of Cohen, Gomme and Kendall. *J R Stat Soc A* 1971; 134: 229-40.
- <sup>16</sup> Haugh L D. Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach. *J Am Stat Assoc* 1976; 71: 378-85.
- <sup>17</sup> Haugh L D, Box G E P. Identification of dynamic regression (distributed lag) models connecting two time series. *J Am Stat Assoc* 1977; 72: 121-30.
- <sup>18</sup> Helfenstein U, Ackermann-Lieblich U, Braun-Fahrlander Ch, Wanner H U. Air pollution and diseases of the respiratory tracts in pre-school children: A transfer function model. *J Environ Monitor Assess* 1991; 17: (in press).
- <sup>19</sup> Sali G J. Evaluation of Boise selective traffic enforcement project. *Transportation Res Rep* 1983; 910: 68-74.
- <sup>20</sup> Amick D R, Marshall P B. Evaluation of the Bonneville County, Idaho, DUI accident prevention program. *Transportation Res Rep* 1983; 910: 81-93.
- <sup>21</sup> Helfenstein U. When did a reduced speed limit show an effect? Exploratory identification of an intervention time. *Accident Analysis & Prevention* 1990; 22: 79-87.
- <sup>22</sup> Broemeling L D. *Bayesian Analysis of Linear Models*. New York: Marcel Dekker, 1985; pp 308-12.
- <sup>23</sup> Abraham B, Ledolter J. *Statistical Methods for Forecasting*. New York: Wiley, 1983; pp 225-29, 336-55.
- <sup>24</sup> Gruchow H W, Rimm A A and Hoffmann R G. Alcohol consumption and ischemic heart disease mortality: Are time series correlations meaningful? *Am J Epidemiol* 1983; 118: 641-50.
- <sup>25</sup> Chatfield C. *The Analysis of Time Series: An Introduction*. London: Chapman and Hall, 1989; pp 27-65, 168-169.
- <sup>26</sup> Helfenstein U. Detecting hidden relations between time series of mortality rates. *Methods of Information in Medicine* 1990; 29: 57-60.
- <sup>27</sup> Hoppenbrouwers T, Calub M, Arakawa K, Hodgman J E. Seasonal relationship of sudden infant death syndrome and environmental pollutants. *Am J Epidemiol* 1981; 113: 623-35.
- <sup>28</sup> Lebowitz M D, Collins L, Holberg C J. Time series analyses of respiratory responses to indoor and outdoor environmental phenomena. *Environ Res* 1987; 43: 332-41.
- <sup>29</sup> Bowie C, Prothero D. Finding causes of seasonal diseases using time series analysis. *Int J Epidemiol* 1981; 10: 87-92.
- <sup>30</sup> Murphy M F G, Campbell M J. Sudden infant death syndrome and environmental temperature: an analysis using vital statistics. *J Epidemiol Community Health* 1987; 41: 63-71.
- <sup>31</sup> Helfenstein U. Box-Jenkins modelling of some viral infectious diseases. *Statistics in Medicine* 1986; 5: 37-47.
- <sup>32</sup> Choi K, Thacker S B. An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza deaths. *Am J Epidemiol* 1981; 113: 215-26.
- <sup>33</sup> Bhattacharyya M N, Layton A P. Effectiveness of seat belt legislation on the Queensland Road Toll. An Australian case study in intervention analysis. *J Am Stat Assoc* 1979; 74: 596-603.
- <sup>34</sup> Lassare S, Tan S H. Evaluation of safety measures on the frequency and gravity of traffic accidents in France by means of intervention analysis. In: Anderson O D (ed) *Time series Analysis, Theory and Practice I*, North-Holland Publishing Company, 1982; pp 297-306.
- <sup>35</sup> Martinez-Schnell B, Zaidi A. Time series analysis of injuries. *Statistics in Medicine* 1989; 8: 1497-508.
- <sup>36</sup> Helfenstein U, Ackermann-Lieblich U, Braun-Fahrlander Ch, Wanner H U. The environmental accident of 'Schweizerhalle' and respiratory diseases: A time series analysis. 1991; *Statistics in Medicine* 1991; 10 (In Press).
- <sup>37</sup> Tsouros A D, Young R J. Applications of time series analysis: A case study on the impact of computer tomography. *Statistics in Medicine* 1986; 5: 593-606.

(Revised version received February 1991)