MODELING MULTIPLE NONLINEAR TIME SERIES: A GRAPHICAL APPROACH TO THE TRANSFER FUNCTION

Raymond G. Hoffmann, hoffmann@biostat.mcw.edu, Biostatistics Division
The Medical College of Wisconsin, Milwaukee, WI 53226

KEY WORDS: SETAR time series, Ecological time series, Case study

ABSTRACT

A transfer function model for the interrelationship of multiple nonlinear time series is developed for the case of an explanatory nonlinear time series. Directed graphs and cross-directed-graphs are used to explore the common trajectory of the interrelated time series.

A Case Study of the monthly U.S. Stroke mortality, U. S. Respiratory Mortality, and the monthly mean U.S. temperature time series from 1938 to 1989. Influenza epidemics may affect stroke mortality, since the stroke patients form a pool of susceptibles that will have stroke as the primary cause of death. Elevated summer temperatures have been hypothesized as accelerating stroke mortality.

INTRODUCTION

The detection of nonlinearity in a univariate time series can be performed

- graphically with directed scattergrams and directed multigrams. A directed graph plots each point versus the previous point in time and collects the points in order of occurrence. A cross-directed-graph follows the trajectory of the two different time series over time. The subsequence directed graph partitions the series up into different segments to test whether the character of the series evolves over time.
- on a global basis with tests based on comparing the autocorrelation properties of the residuals squared with the autocorrelation properties of the squared residuals (Box-Pierce and Ljung-Box Portmanteau tests). The residuals from the seasonal effects are used to test for nonlinearity of the transfer function between the two time series (Brockwell). The residuals from the ARIMA fitting of Y_t on itself, can be used directly to estimate the transfer function (Enders) provided the non-X component is linear.
- with tests based on the homogeneity of the series over time (runs tests or chi-square tests of heterogeneity).

This paper takes the problem of identification of nonlinearity of multiple time series through a transfer function approach. Each series is tested for nonlinearity itself, and if nonlinearity is present, then the relationship between the two series is tested for linearity. Both graphical and quantitative methods for identification of nonlinearity will be explored.

THE MODEL

A linear time series is usually expressed in terms of an autoregressive and a moving average component (Box and Jenkins). The autoregressive component, e.g., a first order autoregressive series, AR(α), can be written as

$$X_t = \alpha X_{t-1} + \varepsilon_{t-1}$$

The simplest nonlinear time series is a segmented time series which switches between two different AR(α) processes on a regular or "seasonal" basis. The next step in complexity is the SETAR(α_1, α_2 ; τ) which switches from one AR(α) to another whenever the observed value of the series passes a threshold τ.

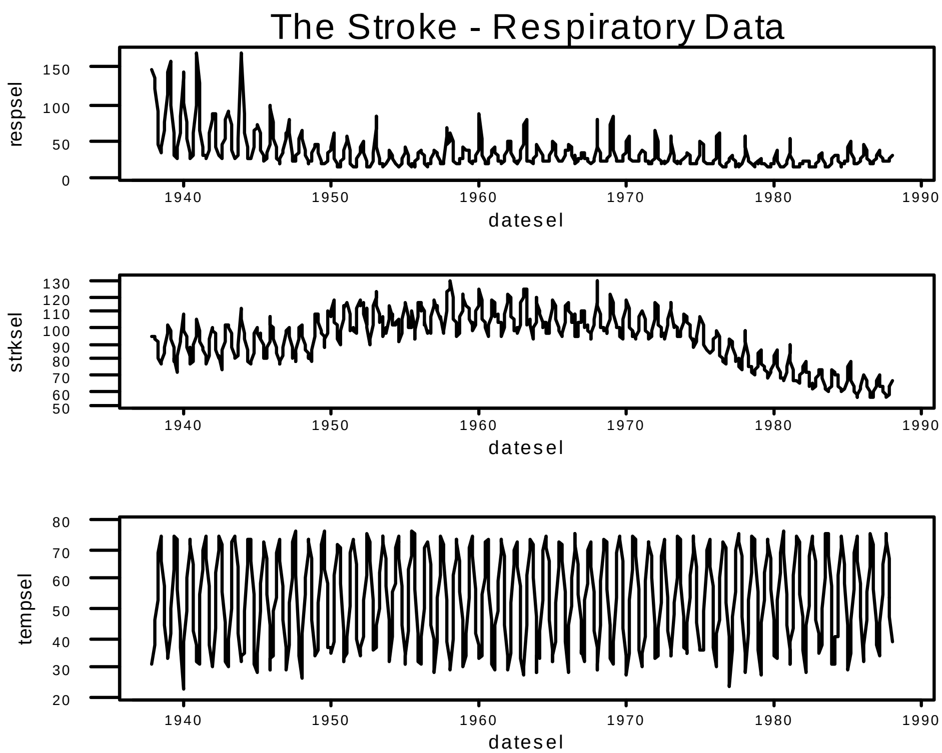
The transfer function model assumes that there is an underlying nonlinear time series

$$X_t = \alpha_1 X_{t-1} + \varepsilon_{t-1} \text{ if } X_t > \tau.$$

$$X_t = \alpha_2 X_{t-1} + \varepsilon_{t-1} \text{ if } X_t \leq \tau.$$

and that the outcome series, Y_t, depends on X_t through a linear or nonlinear function F(.)., as well as on previous values of Y_t. The non-X portion of Y can also be nonlinear; however, for this model it will be assumed that the nonlinearity is induced through the covariate series X_t. Consequently the model for Y_t is

$$Y_t = \beta Y_{t-1} + F(X_t) + \xi_t$$

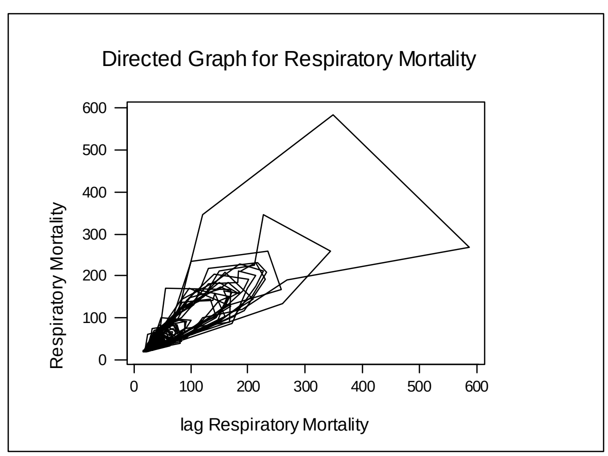


APPLICATION: STROKE AND INFLUENZA MORTALITY

The data set is the monthly United States stroke and respiratory mortality and mean U.S. monthly temperature (Lanska) over the years 1920-1988. The subset from 1938 to 1988 was actually used since the data is stationary (with a log transform of the respiratory data) and since the 1937 monthly stroke data is not available. Figure 1 shows the three data sets, which have a common seasonal effect as well as evidence of heterogeneity over time and evidence of a point process influenza epidemic effect. The respiratory mortality curve is hypothesized to drive the stroke mortality curve; persons who have had strokes -- especially very serious ones -- form a pool of susceptibles that are at a high risk of dying with the stroke as the primary cause of death. The aggregate number of strokes and the subsequent mortality should be independent from month to month, if no outside forces are acting. On the other hand, respiratory mortality clearly does have a dependency (representing the dependency of incidence on contagion). Summer heat waves represent another potential effector of stroke mortality.

GRAPHICAL METHODS

A directed graph plots each point versus the previous point and connects the points in order of occurrence. The directed graph is often used to identify differential cyclical behavior in a time series. Figure 2 shows the directed graph (lag 1) for the respiratory mortality. Since the data has a strong yearly seasonal effect, the lag 12 directed graph is displayed in figure 3.



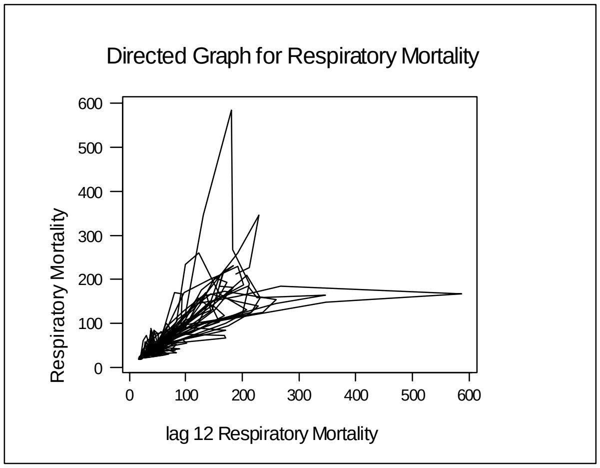
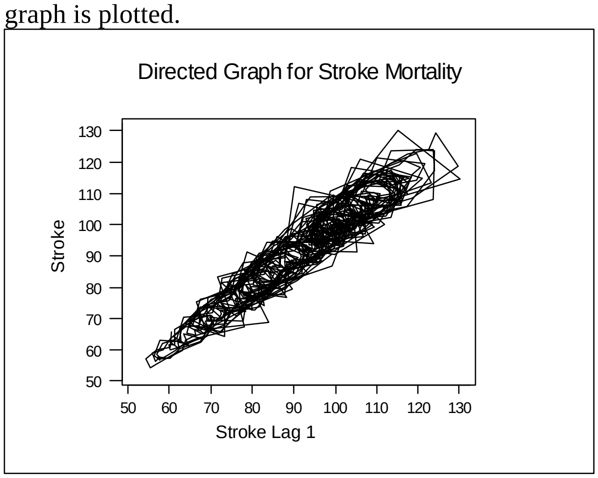
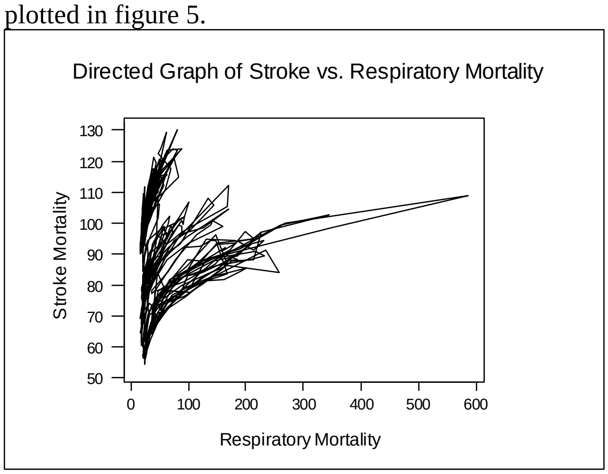


Figure 4 shows a directed graph of the lag 1 stroke mortality; the yearly (lag 12) curve looks essentially the same because of the time trend. Stroke mortality needs to be detrended before the directed graph is plotted.

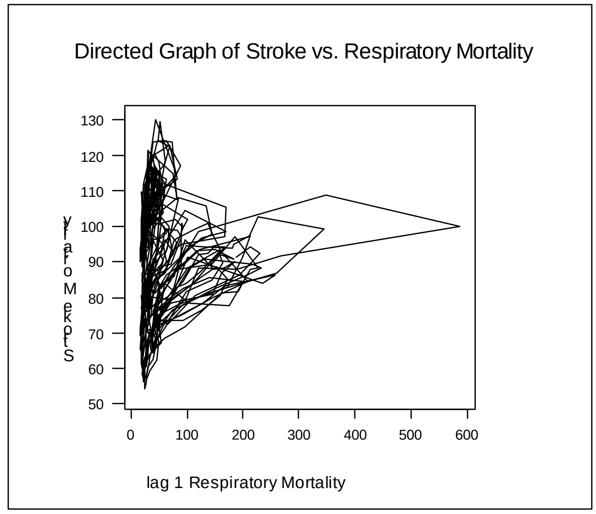


To examine the relationship between the two series, a cross-directed graph of the two series is plotted in figure 5.

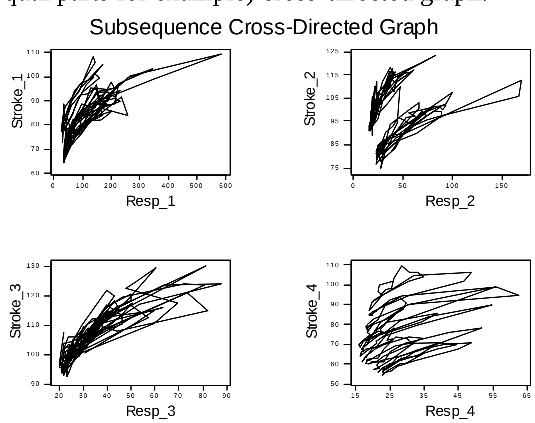


Notice the three cycles corresponding to the three different types of influenza epidemic: weak, strong and extremely strong influenza peaks

The series effects on each other may not be synchronous, so the lag one cross-directed graph is also shown in figure 6. The three types of cycles are overlapped here, suggesting a lead by one month of the influenza over the stroke.

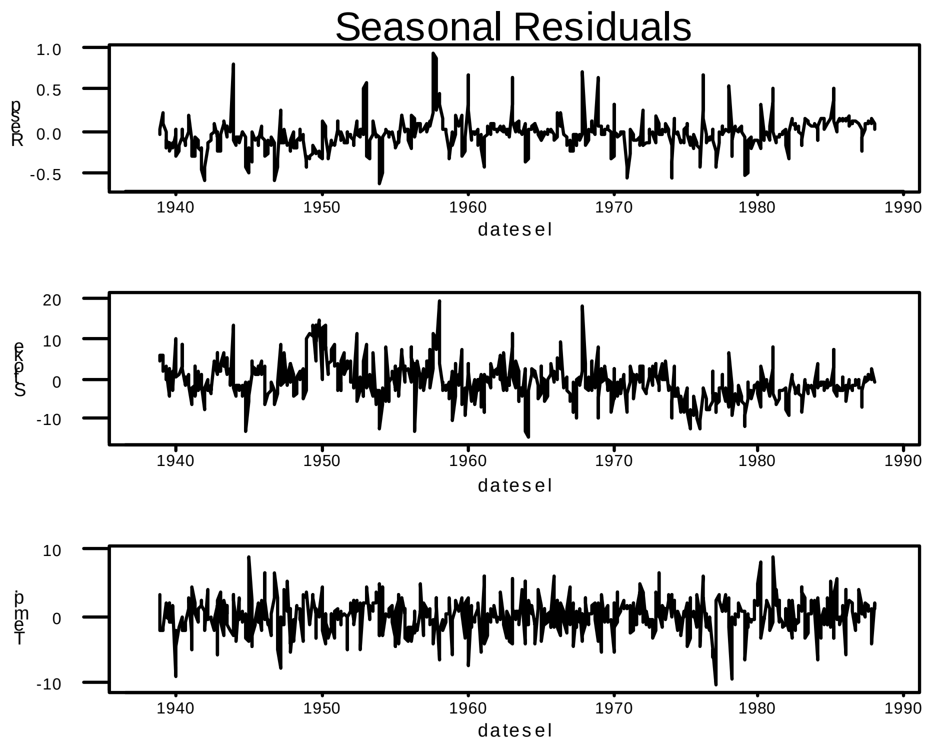


This is an excellent graphical method to identify the relationships among two series. The question of the consistency of the relationship over time can be answered with a subsequence (dividing the series into 4 equal parts for example) cross-directed graph.

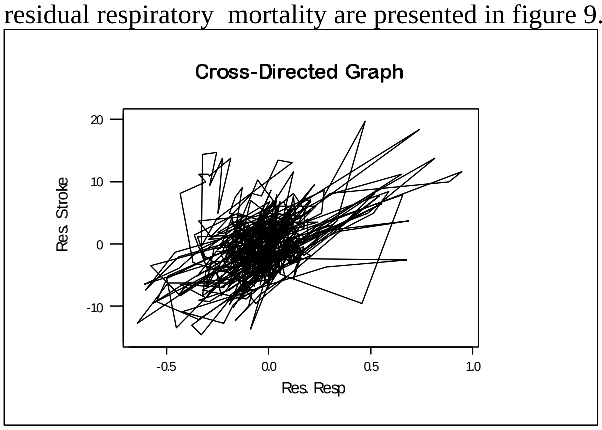


The relationship is clearly changing over time and appears to be quite nonlinear in character, e.g. the separate cycles observed in the second and the fourth subsequences.

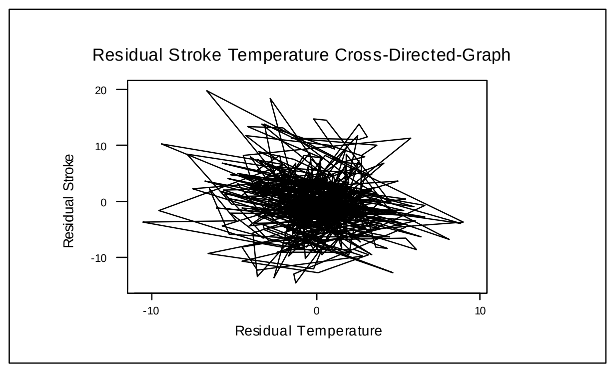
Nevertheless, it is also possible that the relationship is only due to a (non-linked) seasonal effect, and thus the next area to examine is the residuals from the seasonal effect. The residuals from a Box-Jenkins seasonal ARIMA model (MINITAB) applied to each series individually are displayed in figure 8 for each series.



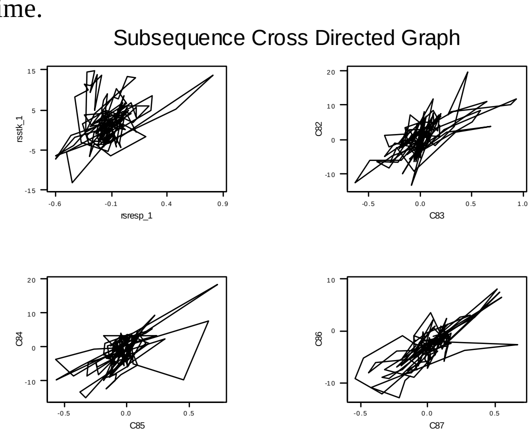
The cross-directed graph for the residual stroke and residual respiratory mortality are presented in figure 9.



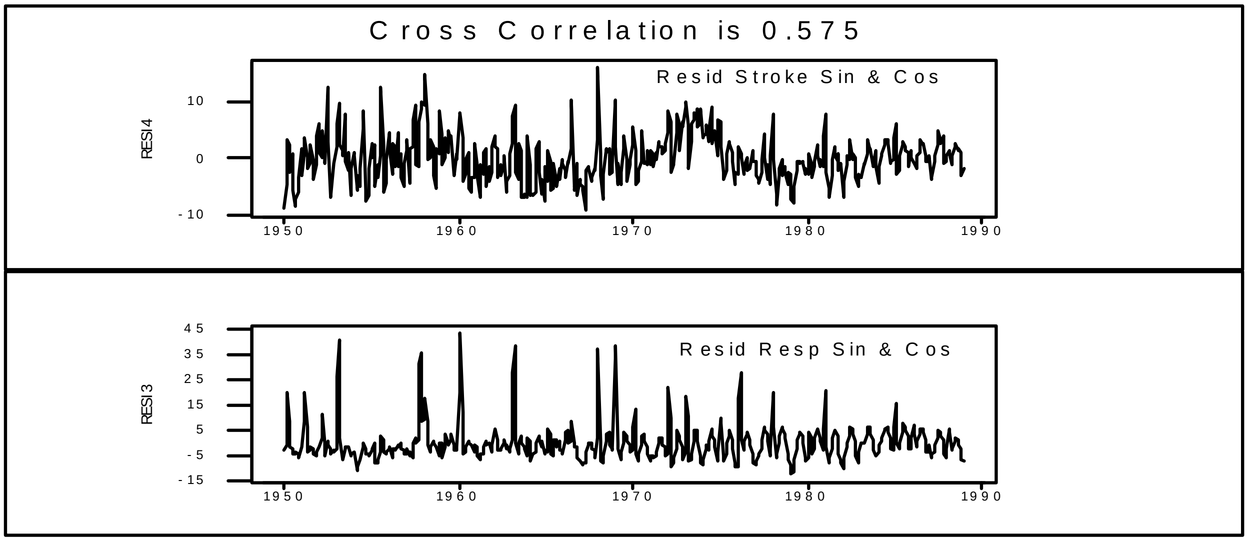
There are synchronous peaks in figure 9, suggesting a linear transfer function from the point process in the influenza mortality to the stroke mortality. Figure 10 displays the cross-directed graph for stroke and temperature. This graph suggests that there is little relationship between the residual stroke and the residual temperature. The time scale may be too gross; it may be that the relationship is between the mean *maximum* monthly temperatures; or the relationship is only geographically local.



The cross subsequence graph, figure 11, shows a similar stroke-respiratory residual (point process) over time.



An alternative method for removing the seasonality is to use a sin and cosine to represent the seasonality, and to fit this to respiratory and stroke mortality separately. In this case the cross-correlation between the two curves is 0.575. The cross-correlation between the ARIMA residual series was 0.433.



Several types of models are possible for the transfer function (Bendat). The following hierarchy of nonlinearity can be tested using portmanteau tests of nonlinearity:

- No relationship between the two series
- A linear transfer function for the relationship
- A cubic (symmetric) nonlinear transfer function
- A linear plus saturated transfer function
- A threshold transfer function

The last two figures suggest a linear function with respiratory mortality and no relation with temperature.

REFERENCES

Box GEP and Jenkins GW, Time Series Analysis, 3rd edition, 1991, Academic Press, NY, NY

Brockwell and Davis Time Series: Theory and Methods, Springer-Verlag, 1991, NY, NY.

Bendat JS, Nonlinear System Analysis and Identification, John Wiley & Sons, Inc. 1990.

Harvey A and Koopman SJ, “Structural time series models in medicine”, Stat Meth in Med Res.5::23-49, 1996.

Hoffmann RG, “Identification of Multivariate Binary Time Series”, Proceedings of the Statistical Computing Section, ASA, 1983.

Lanska DJ, “Geographic distribution of stroke mortality in the United States: 1939-1941 to 1979-1981”, Neurology 43:1839-51, 1993.

Tong H, Nonlinear Time Series, 1990, Oxford University Press, New York, NY