NONLINEAR MODELS

So far, we have restricted our attention to linear forecasts, i.e., forecasts which are linear combinations of present and past values of the series. Clearly, not all forecasts are linear. Consider, for example, the forecast \( f_{n+1} = x_n^2 \). There are two main reasons for restricting attention to linear forecasts. First, the theory is relatively simple, so that one can always derive (in principle) the best linear forecast for a stationary time series. Second, if the time series is Gaussian (i.e., normally distributed) then the best linear forecast is in fact the best of all possible forecasts: No nonlinear forecast can do better in terms of mean squared prediction error. Thus, as long as the series is Gaussian, we need look no further than the linear methods (e.g., ARMA forecasting) already presented.

In reality, though, actual time series are often non-Gaussian. In this case, the series may be either "linear" or "nonlinear". These terms are defined below. It can be shown that all Gaussian series are "linear". If the series is linear (whether Gaussian or non-Gaussian), then the best linear forecast is once again the best of all possible forecasts. If the series is nonlinear, however, then nonlinear forecasting methods may work better than linear ones. What is needed, then, is a class of nonlinear models to describe such series (e.g., a nonlinear generalization of the ARMA models), and methods for testing the observed data to determine whether the underlying series is nonlinear. We will discuss some specific nonlinear models below. General tests for linearity are now available, but they involve "frequency domain" ideas, and are too complicated to be discussed here. It is well worth noting, however, that many important series (e.g., the sunspot series) have been determined by this test to be nonlinear.

A stationary time series is **linear** if it can be written in the MA\((\infty)\) form

\[
x_t = \sum_{k=0}^{\infty} d_k e_{t-k}
\]

where \( \{e_t\} \) is a series of strictly independent zero-mean random variables with constant variance. Otherwise the series is **nonlinear**. Note that non-Gaussian random variables may be uncorrelated without being independent. (An example is given below.) Thus, independence is a stronger condition than uncorrelatedness. The best forecast of \( e_{n+1} \) based on any function (linear or nonlinear) of \( e_n, e_{n-1}, \ldots \) is just the expectation \( E[e_{n+1}] = 0 \). Thus, \( e_{n+1} \) is not forecastable by any method, linear or nonlinear. By
contrast, consider a white noise process \( \{ \varepsilon_t \} \) where the \( \varepsilon_t \) are merely uncorrelated. Then \( \varepsilon_t \) is not \textit{linearly} forecastable, but it may be possible to obtain a good forecast of \( \varepsilon_{n+1} \) using a \textit{nonlinear} function of \( \varepsilon_n, \varepsilon_{n-1}, \ldots \). A wide variety of series (even nonlinear ones!) may be written in the \( MA(\infty) \) form

\[
x_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}
\]

where the \( \{ \varepsilon_t \} \) are merely uncorrelated (i.e., ordinary white noise) but the series is not linear unless the \( MA \) representation is with respect to \textit{strict} (i.e., independent) white noise, \( \varepsilon_t \).

For a simple example of a nonlinear process, consider

\[
\eta_t = e_t + \beta e_{t-1} e_{t-2}
\]

Then \( \eta_t \) has zero mean, constant variance, and autocovariance function

\[
E[\eta_t, \eta_{t+s}] = E[e_t e_{t+s} + \beta e_{t-1} e_{t-2} e_{t+s} + \beta^2 e_t e_{t+s-1} e_{t+s-2} + \beta e_{t-2} e_{t+s-1} e_{t+s-2}] = 0.
\]

If \( s \neq 0 \) it follows that \( E[\eta_t, \eta_{t+s}] = 0 \). Thus, \( \{ \eta_t \} \) is an uncorrelated process (white noise!), so it is not linearly predictable. But we can obtain a nonlinear predictor, as follows. Start by writing

\[
\eta_{n+1} = e_{n+1} + \beta e_n e_{n-1}.
\]

Since the \( \{ e_t \} \) are strictly independent, it follows that \( e_{n+1} \) is not forecastable (either linearly or nonlinearly), so we may replace it by its mean (zero) to obtain the best forecast \( f_{n+1} = \beta e_n e_{n-1} \). This forecast is nonlinear, since it cannot be expressed as a linear combination of \( \eta_n, \eta_{n-1}, \ldots \).

**Bilinear Models**

The bilinear models are a very useful class of nonlinear models, and are a direct nonlinear generalization of the ARMA models. The general bilinear model for a series \( \{ x_t \} \) is

\[
x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q} + \sum_{i=1}^{l} \sum_{j=1}^{f} c_{ij} e_{t-i} x_{t-j},
\]

where \( \{ e_t \} \) is strict white noise, and the parameters are \( a_1, \ldots, a_p, b_1, \ldots, b_q \), and the \( \{ c_{ij} \} \). If the \( c_{ij} \) are all zero, then the model reduces to an \( ARMA(p,q) \). The parameters in the bilinear model may
be estimated by least squares methods. The model itself may be identified (i.e., the orders, \( p \), \( q \), \( I \), \( J \) may be selected) using the AIC criterion. Methods of forecasting bilinear models are analogous to those used for ARMA models.

The sunspot data (monthly mean sunspot numbers) was identified by Box and Jenkins as either \( AR(2) \) or \( AR(3) \). When allowed to choose from the class of general bilinear models, the AIC criterion selects a bilinear model with \( p = 3 \), \( q = 0 \), \( I = 4 \), \( J = 3 \). The model has 15 parameters, but their inclusion seems to be warranted. The main feature of the sunspot data is a cycle with nonconstant period ranging from about 9 to 14 years. Another interesting feature is that the rise to a maximum is typically steeper than the fall to the next minimum. This feature suggests that the series may be nonlinear. The fact that AIC selects a nonlinear model seems to support this conclusion.

**Threshold Autoregressive Models**

The threshold autoregressive (TAR) models are due to Tong (1978) and provide another class of finite parameter nonlinear models. The TAR models are autoregressive models with parameters depending on past values. The TAR(1) is

\[
x_t = \begin{cases} 
  a^{(1)}_1 x_{t-1} + \varepsilon^{(1)}_t & \text{if } x_{t-1} < d \\
  a^{(2)}_2 x_{t-1} + \varepsilon^{(2)}_t & \text{if } x_{t-1} \geq d
\end{cases}
\]

Thus, \( x_t \) switches between two autoregressive processes, according to whether the most recent past value was above or below the threshold, \( d \). The AIC criterion can be used to determine an appropriate value for \( d \). The model may be extended to the \( k \)-threshold form,

\[
x_t = a^{(i)}_1 x_{t-1} + \varepsilon^{(i)}_t \quad \text{if } x_{t-1} \text{ lies in } R_i \quad (i = 1, \ldots, k),
\]

where \( R_1, \ldots, R_k \) are given sets of real numbers. We may think of the \( k \)-threshold autoregressive model as an approximation to the general nonlinear first order model

\[
x_t = \lambda(x_{t-1}) + \varepsilon_t,
\]

where \( \lambda(x) \) is some general function of \( x \). The \( p \)'th order model, TAR\( (p) \) is

\[
x_t + a^{(1)}_1 x_{t-1} + \cdots + a^{(p)}_p x_{t-p} = \varepsilon^{(i)}_t
\]
if \((x_{t-1}, \ldots, x_{t-p})\) lies in \(\mathbb{R}^4\), and \(\mathbb{R}^6, i = 1, \ldots, k\) is a given region of \(p\)-dimensional space. AIC may be used to select the value of \(p\). An important property of the TAR models is that they can produce "limit cycles": If we set \(\varepsilon_t = 0\) for all \(t \geq t_o\) in the TAR\((p)\) model, then \(x_t\) will settle down eventually to a periodic behavior. (In ordinary ARMA models, if the driving term \(\varepsilon_t\) is removed, the process will eventually decay to zero.) This limit cycle behavior means that the long term forecasts will be cyclical. Thus, the model may be useful for describing periodic (or quasi-periodic) phenomena. For the sunspot data, a TAR model is able to capture both the periodicity and the asymmetric rise and fall of the data.