OVERVIEW OF NONLINEAR TIME SERIES SPECIFICATIONS IN ECONOMICS

Clive W.J. Granger
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
USA

First Draft
May 1998

1. Definition and Introduction

It is worth starting with a definition of linearity (and thus non-linearity) to attempt to reduce a variety of misunderstandings an disagreements in discussions of nonlinearity. Consider initially just a single series x_t which is to be explained by a vector $\underline{y}_t$, which may include lagged (linear) values of x_t. If the conditional mean of x_t given $\underline{y}_t$ takes the form

$$E[x_t|\underline{y}_t] = \underline{\beta}'\underline{y}_t + \varphi(\underline{y}_t) \tag{1.1}$$

then the relationship may be called linear in mean if $\varphi(\underline{y}_t) = 0$ all t, otherwise the relationship is nonlinear. In (1.1) $\underline{\beta}'$ is considered to be chosen to maximize the (linear) relatedness between x_t and linear combinations of $\underline{y}_t$. This definition was proposed by Hal White.

There are a number of features of the definition that need emphasis.

(i) The linearity is in <u>variables</u>, not coefficients. Thus $\underline{\beta}'y_t$ could take the form $\beta y_{1t} + \beta^2 y_{2t}$ and then be linear in variables but not in coefficients. This second form of nonlinearity is of considerable importance when discussing estimation questions and might be relevant when considering linear simplifications of the properties of a process but is of little relevance for nonlinearity properties.

(ii) The definition relates to a particular form of the process x_t, in this case the mean. One could apply the same definition to some function of x_t, such as $k(x_t)$, then

$$E[k(x_t)|y_t] = \gamma_k' \underline{k(y_t)} + \varphi_k(\underline{y}_t)$$

where $\underline{k(y_t)}$ means that each component $\underline{y}_t$ is replaced by k(-) of that component; and again, if $\varphi_k(\underline{y}_t)$ is zero, the relationship between $k(x_t)$ and $\underline{k(y_t)}$ is linear. It will be seen later that this applies to some ARCH models.

The definition (1.1) has to be specified rather carefully in that the components of $\underline{y}_t$ should be defined individually rather than jointly, apart from lags. Thus $\underline{y}_t$ can include w_t and w_t^2, for example. However, if $\underline{y}_t$ include w_t but $w_t = \log r_t$, say, then the relationship between x_t and w_t is linear but not between x_t and r_t.

A rather more modern version of the definition (1) would allow coefficients to be time-varying, so that now

$$E[x_t|\underline{y}_t] = \underline{\beta}'_t \underline{y}_t + \varphi_t(\underline{y}_t).$$

If $\underline{\beta}_t \equiv \underline{\beta}(t)$, a vector of deterministic functions, then the relationship is still linear if $\varphi_t(\underline{y}_t) \equiv 0$. If $\underline{\beta}_t$ is a function of components of $\underline{y}_t$, the specification is nonlinear. If $\beta_t \equiv \beta(\underline{Z}_t)$ where $\underline{Z}_t$ is a "state-space" variable then if $\underline{Z}_t$ is i.i.d. one still has linearity with the usual conditional, but otherwise if $\beta(\underline{Z}_t)$ can be forecast from previous $\underline{y}_t$, by use of the Kalman filter, for instance, one will get nonlinearity.

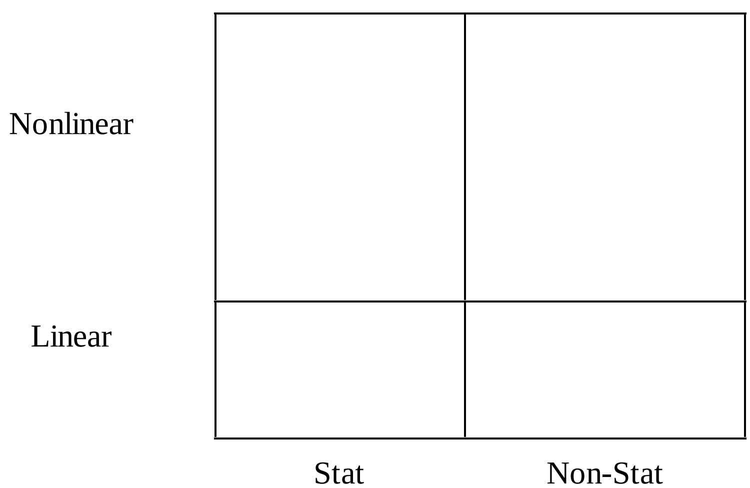
For ease of exposition, just univariate or bivariate series will be considered. In most cases, the multivariate generalizations are obvious but some models only exist in a bivariate case, as will be seen. One problem that has to be faced is that there is no generally agreed method of measuring the basic properties of nonlinear processes. There is nothing compared to the linear properties, the autocorrelations ρ_k and the spectrum. Naturally, these quantities can be deduced for a nonlinear process but at best they can only partly capture the major properties of the process while at worst they can be mis-leading as shown in an example given later.

In many specifications of nonlinear models, there is included as an input an unobserved set of stochastic shocks ε_t. One might think of a specification

$$x_t = f(\underline{y}_t, \varepsilon_t)$$

for example. The process is said to be invertible if, given $x_{t-j}, \underline{y}_{t-j}, j = 0, 1, \ldots$, it is possible to estimate ε_t, with no error, asymptotically. In some specifications there are more than one stochastic impulse used in the generation of a variable, and these shocks may be independent. An example is the model for variance called "stochastic volatility." For this example the shocks are not invertible, but there are cases in which there are two shocks used to generate a single variable but they are both potentially invertible, see later.

The diagram illustrates the inter-relationship between stationarity or not and linearity or not.



The majority of the classical time series models were stationary and linear, then Box and Jenkins introduced the unit root models which included one aspect of non-stationarity in the variance. When classes of nonlinear models are considered, below, they all, with just one notable exception, include linearity as a special case. The exceptions are the chaos (deterministic) models. I view this use of the linear model as a kind of anchor as an example of the difficulty in escaping the linear formulation which has been the basis of time series modeling for so long. A different example is the continued use of a normality assumption for the distribution of errors or shocks. Similarly, the attention paid just to means and variances, possibly conditional, pertains to normal distributions and thus linear relationships.

For ease of discussion, deterministic components, such as polynomial or exponential trends or trigonometrical seasonal components, will not be considered.

2. Nonlinear Autoregressive Models

A widely investigated class of models take the form

$$x_t = f(x_{t-1}) + h(x_{t-1})\varepsilon_t \qquad (2.1)$$

where ε_t are zero mean i.i.d. They can also be called *NL* scale models as they involve just a mean *f* and a scale *h* functions. More generally they will involve more lags of x_t and, in multivariate cases, also other variables. It is well known that a necessary and sufficient condition for stability, provided $h(x)$ is bounded, is:

$$\frac{|f(x)|}{|x|} < 1 \quad \text{all } |x| > |x_0| \qquad (2.2)$$

see Tweedie (1975), Meyn and Tweedie (1993). Stability means that asymptotically the marginal distribution of x_t tends to a constant distribution. Provided coefficients in the model are constant, the condition will also give stationarity. Effectively it means that for x_t far enough away from zero

$$f(x_t) \simeq x_t g(x_t) \quad \text{where } |g(x_t)| < 1.$$

A simple example is the threshold autoregressive model, for instance

$$\begin{aligned} x_t &= \alpha_1 x_{t-1} + \varepsilon_t \quad \text{if } x_{t-1} < 2 \\ &= \alpha_2 x_{t-1} + \varepsilon_t \quad \text{if } x_t > 2 \end{aligned}$$

with $0 < \alpha_1 < \alpha_2 < 1$ which will obey the stability condition (2.2). More complicated forms of threshold models have been considered by Tong (1983) and others, usually involving more lags. Smoothed versions replaced the step function by a logistic function

$$x_t = x_{t-1}\, g(x_{t-1}) + \varepsilon_t$$

$$g(x_{t-1}) = \frac{\alpha}{1 + \exp(-x_{t-1} + k)}, \qquad |\alpha| < 1$$

giving an example of a smooth transition regression model discussed in Granger and Teräsvirta (1993).

A special class of *NLAR* models are those that produce chaos. Essentially they are a map

$$x_t = f(x_{t-1})$$

involving no stochastic shock which, on occasion, can produce <u>some</u> properties of an independent sequence. An example is the tent map

$$x_t = 4x_{t-1}(1 - x_{t-1}) \tag{2.3}$$

which, when started at an appropriate place, will generate a series which visually is similar to a <u>bounded</u> white noise and which empirically has all autocorrelations zero and thus a flat spectrum. It should be noted that the chaotic process does not have the property of an i.i.d. process, as sometimes claimed, as usually corr $[g(x_t)\, g(x_{t-k})] \neq 0$. If the specification is known completely then x_{t-1} is forecastable exactly from x_t at least within computer round-off error.

In the physical sciences, a nonlinear model is almost invariably a chaos model, but in economics, for example, there is no evidence that the data is satisfactorily fitted by such a model, possibly because series are relatively short and naturally very noisy.

If one adds a noise to the map (2.3) one gets a non-stable *NLAR* process; in fact an explosive process, and this is generally true for chaos generators. To add noise one has to change the model to a form such as

$$x_t = 4\lambda x_{t-1}(1 - x_{t-1}) + \varepsilon_t, \qquad 0 < \lambda < 1$$

where ε_t is i.i.d. and $0 \leq \varepsilon_t < 1 - \lambda$, so that $0 \leq x_t \leq 1$.

It is not clear to me if tests for chaos exist. Those most used on economic data involve estimates of the Lyapunov exponent which considers generating sequences using the same map but nearby starting points, and then measuring the speed at which the two series so obtained diverge apart. It is useful when comparing series that are deterministic but some stochastic series have a positive Lyapunov exponent - which is taken to be the distinguishing feature of a chaotic process. An example is the linear explosive AR(1) with coefficient greater than one, or perhaps such a series reduced modulo (*k*), so that if x_t is subtracted from it to return it to the region (0, *k*). Some people would define such a process to be chaotic, even though stochastic, but this merely illustrates the difficulty with definitions in this area when it interacts with statistics.

There are many specific *NLAR* models that have been discussed to various extent, including:[1]

1. Threshold AR(1) [Tong (1983)].

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-1} \vee 0 + \varepsilon_t$$

and so equally could be written

$$X_t = \theta_1 X_{t-1}^{+} + \theta_2 X_{t-1}^{-} + \varepsilon_t$$

where

$$X_t^{+} = X_t \text{ if } X_t > 0, \text{ is zero otherwise}$$

$$X_t^{-} = X_t \text{ if } X_t < 0, \text{ is zero otherwise.}$$

The process has been called "double threshold" if both the conditional mean and variance change with thresholds.

2. Multivariate Adaptive Regression Splines (MARS)-AR (Lewis and Stevens, 1991)[2].

---

[1] It is useful to use the following maximum notation: $A \vee B \equiv \max(A, B)$ notes that $A \vee B \vee C$ is max(*A*,*B*,*C*). Similarly $A \wedge B$ is min (*A*,*B*). Throughout ε_t is a white noise, uncorrelated series, possibly heteroskedastic.

[2] *JASA* 87, 864-877.

$$X_t = a_{11}[(X_{t-1} + c_1) \vee 0] + a_{12}[(X_{t-1} + c_2 \vee 0]$$
$$+ a_{21}[(X_{t-2} + c_3) \vee 0] + \cdots + \varepsilon_t.$$

3. $X_t = -\alpha |X_{t-1}| + \varepsilon_t$

4. $X_t = \text{sign}|X_{t-1}| + \varepsilon_t$, no coefficient needed, scale change in *X* just changes variance of ε_t.

5. *Doubly Stochastic AR(1)* (Tjøstheim, 1986).

$$X_t = A_t X_{t-1} + \varepsilon_t$$

where A_t is stochastic, possibly function of another variable.

6. *MAX-ARMA* (Davis and Resnick, 1989)[3]

$$x_t = \varphi_1 x_{t-1} \vee \varphi_2 x_{t-2} \vee \varepsilon_t \vee \theta_1 \varepsilon_{t-1}$$

could involve more lags of x_t, ε_t. It usually requires all φ's, θ's to be positive.

An interesting property of the max AR(1) model is

$$x_t = \bigvee_{j=0}^{\infty} \varphi^j \varepsilon_{t-j} \text{ using the obvious notation.}$$

7. *M-M Model* (Granger and Hyung, 1998)

$$X_t = (aX_{t-1} + \alpha) \vee (bY_{t-1} + \beta) + \varepsilon_{1t}$$
$$Y_t = (cX_{t-1} + \lambda) \wedge (dY_{t-1} + \delta) + \varepsilon_{2t}$$

so the first equation involves a maximum, the second a minimum. An interesting integrated-type process occurs if $a = b = c = d = 1$. This model is unusual in that it is bivariate but there are no interesting univariate equivalents or components.

8. *MOD-AR* (Hildebrand, 1983)[4]

$$X_t = a_t X_{t-1} + \varepsilon_t \, (\text{mod } \rho)$$

where X mod $\rho = X - \rho$ if $X > \rho$.

---

[3] *Adv. Appl. Prob.* 21, 781-803.

[4] *Annals of Probability* 21, ??-??

9. Nonparametric AR (Teräsvirta, Tjøstheim, and Granger, 1994)

A fairly obvious class is obtained by plotting X_t against X_{t-1}, say, and then a nonparametric smoother used to get a diagrammatic relationship. Such models may be difficult to use for forecasting.

10. Generalized Autoregressions (GAR, Mittnick, 1991)

A generalized autoregressive process is defined to contain products of the form $X_{t-1}^{\alpha} X_{t-2}^{\beta} \ldots X_{t-p}^{\gamma}$ and to be GAR (*r*, *p*) if the maximum power used on any term is *r* and if *p* lags are used. Thus, for example, the G(2.2) model would be

$$
\begin{aligned}
X_t = a_1 X_{t-1} + a_2 X_{t-2} + b_1 X_{t-1} X_{t-2} + c_1 X_{t-1}^2 X_{t-2} \\
+ c_2 X_{t-1} X_{t-2}^2 + d_1 X_{t-1}^2 X_{t-2}^2 \\
+ e_1 X_{t-1}^2 + e_2 X_{t-2}^2 + \varepsilon_t .
\end{aligned}
$$

Without shocks, some chaotic models fall into this class. It is also clear that many of these models do not obey the stability criterion and so will produce explosive series. That this does not always occur is found in the relationship with the bilinear model, as discussed below. Mittnick also shows that there is a state space representation possible for the $G(r,p)$ model, although it is complicated.

I am sure that there are many other specific models that I have missed.

3. Moving Averages and Bilinear Models

There are many simple NLMA models such as

$$X_t = \varepsilon_t + \beta \varepsilon_{t-1} \varepsilon_{t-2}$$

and

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} \vee 0 + \beta_2 \varepsilon_{t-1} \wedge 0$$

which is a threshold form. These models are not invertible. The Volterra representation is a nonlinear generalization of the Wold representation for stationary, and used products of all lags, so includes ε_{t-j}, $\varepsilon_{t-j}\,\varepsilon_{t-k}$, $\varepsilon_{t-j}\,\varepsilon_{t-k}\,\varepsilon_{t-i}$, ... all *j, k, i* ... values, with coefficients. Although of possible theoretical interest it has proved of virtually no practical relevance.

Bilinear models add terms like $X_{t-j}\varepsilon_{t-k}$ to ARMA models. A simple example is

$$X_t = \alpha X_{t-1} + \beta X_{t-2}\,\varepsilon_{t-1} + \varepsilon_t.$$

If $\alpha = 0$, this model has the interesting property that all theoretical autocorrelations are zero, even though X_{t-1} can be forecast nonlinearly from $X_{t-j}\quad j \geq 0$. Results for stationarity and invertibility are known in some special cases, but are unlikely to be used in practice. Using economic data, bilinear models have not been found to be very relevant. However, many tests of linearity are found to have low power against bilinear data.

Mittnick (1991) shows that the invertible, stationary, bilinear model

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

has a generalized autoregressive GAR(2,∞) representation

$$X_{t+1} = \varepsilon_t + \sum_{k=1}^{\infty} \beta_{1,2,\ldots,k}\, X_{t-k} \prod_{i=1}^{k} X_{t-i}.$$

4. State Space Models

There seems to be considerable disagreement in the literature about how state space models should be defined. A general form is

$$X_t = f(X_{t-1}, S_t, \varepsilon_t, \theta) \tag{4.1}$$

where S_t is the state variable, ε_t are a series of shocks, and θ is a vector of coefficients. More lags of X_t could be included and (4.1) has to be augmented by a generating mechanism for S_t, usually Markovian. However, Meyn and Tweedie (1993) call the univariate version of (4.1) without S_t a “scalar nonlinear state space model.”

A form closer to the familiar linear state space model would be

$$X_t = m(S_t) + g(S_t) X_{t-1} + h(S_t)\varepsilon_t \tag{4.2}$$

$$S_t = \theta(S_{t-1}) + \eta_t \tag{4.3}$$

where (4.2) is the observation equation and (4.3) the state variable generating equation. However, S_t could be replaced by some other variable in (4.2), such as a measure as the state of the economy (unemployment level, capacity utilization for example), which are directly observable.

A different form has

$$X_t = A_t S_t + \beta_t \varepsilon_t, \text{ the observation equation} \tag{4.4}$$

$$S_t = \alpha_t S_{t-1} + \beta_t \eta_t, \text{ the state-variable equation} \tag{4.5}$$

where in the standard linear form the coefficients (or matrices in the multivariate case) A_t, B_t, α_t, β_t are know deterministic functions of times. In the generalization form considered by Abraham and Thavaneswaran (1991) these coefficients are all functions of past observed values, X_{t-j}, $j \geq 0$, and thus gives a nonlinear form. They give prediction and fixed point smoothing algorithms for the general model.

It is well known that the Kalman filter can be used to investigate the time-varying parameter form of the model (4.4). J. Hamilton (1994) provides a comprehensive discussion of these models, "Handbook of Econometrics, Volume IV," Chapter 50, edited by R. F. Engle and D.L. McFadden.

5. Growth Series, Stochastic Trends, Cointegration

A growth series might be defined as one having the property $Prob(x_t \to \infty) = 1$ as $t \to \infty$. This can be stated more formally, but is not a useful definition is practice. However, it does illustrate the fact that a series can have a growing mean but not be a growth series. A simple linear illustration uses the random walk with drift

$$x_t = x_{t-1} + g + \varepsilon_t$$

with $var(\varepsilon_t) = \sigma^2(t)$ and $g > 0$. Then $E[x_t] = gt$ and $var(x_t) \equiv v_t = \sum_{j=0}^{t} \sigma^2(j)$ and for t large $x_t \sim N(gt, v_t)$. The lower end of any confidence interval will be of the form

$$\underline{A}_\alpha = gt - b\sqrt{v_t}$$

for some appropriate b and this is not growing if $b(v_t)^{1/2} > gt$. Thus, the likely range of the process includes a non-growth region and so $\text{prob}\,(x_t = \infty) \neq 1$ as $t \uparrow \infty$. As this example shows, after subtracting a possibly growing mean, the volatility of the series $|x_t|$ can also be growing. A series with a trending mean, or having a growing volatility has become known in the economic literature as a "stochastic trend."

A nonlinear class of stochastic trends has been considered by Granger, Inoue, and Morin (1997)

$$X_{t+1} = X_t + g(X_t) + \varepsilon_{t+1} \tag{5.1}$$

where $g(x_t) > 0$, X_0 is given and

$$E[\varepsilon_t^2 | X_{t-j}, j > 1] = \sigma^2(X_{t-1}) > 0.$$

Kershing (1986) proves some growth or not results depending on whether $g(x)$ is $o(x)$ or $O(x)$ and the balance between the sizes of $g(x)$ and $\sigma^2(x)$ as x becomes large. It is found that a wide variety of trends can be generated by (5.1) and they are related to, but dominate on occasion, the deterministic trend generated by (5.1) but with all $\varepsilon_t \equiv 0$.

Most of the discussion of stochastic trends in economics are in connection with unit root processes

$$(1-B)X_t = \varepsilon_t \tag{5.2}$$

where B is the lag operator, so that $B^k x_t = x_{t-k}$, so that (5.2) is the standard random walk, without drift. A model that is less discussed by is relevant is

$$(1-\lambda B)x_t = \varepsilon_t,\ \lambda > 1 \tag{5.3}$$

which produces an explosive autoregressive series. A realistic generalization are the stochastic root models

$$(1-\exp(\alpha_t)B)x_t = \varepsilon_t \tag{5.4}$$

where α_t is a stationary series, with no causality from x_t to α_t. The stochastic unit root process (STUR) has $E[\exp(\alpha_t)] = 1$. If $E[e^{\alpha_t}] = \theta > 1$ it gives an "explosive stochastic root," or *XSR*. This last process has explosive phases, followed by returns to the mean if θ is not too large. It has the appearance of a hyper inflation, for example.

A pair of series X_t, Y_t might be called "co-growing" if they each contain growing components but some linear combination is not growing, $Z_t = X_t - AY_t$. This will only occur if X_t, Y_t have the same growth component and so may be written

$$X_t = AW_t + \tilde{X}_t, \qquad Y_t = W_t + \hat{Y}_t \tag{5.5}$$

where W_t is some growth process, $\tilde{X}_t, \hat{Y}_t$ are not growing. This can all be generalized so that Y_t contains a function of W_t, Z_t is a function of X_t, Y_t but this case is not well developed. The earliest version considered had W_t a random walk, so that $(1-B)W_t$ was stationary, in which case X_t, Y_t are generated by an "error-correction model"

$$(1-B)X_t = \gamma_1 Z_{t-1} + \text{lags } \Delta X_t, \Delta Y_t + \varepsilon_{1t}$$
$$(1-B)Y_t = \gamma_2 Z_{t-1} + \text{lags } \Delta X_t, \Delta Y_t + \varepsilon_{2t}$$

where $\Delta = (1-B)$ and $|\gamma_1| + |\gamma_2|$ is non-zero, so that at least one of γ_1, γ_2 is non-zero. Nonlinear versions of these equations have been considered with Z_t being replaced by $g_1(Z_t)$,

$g_2(Z_t)$ in the two equations. Sometimes, to explore nonlinear possibilities, Z_t is replaced by Z_t^+, Z_t^- components, where $Z^+ = Z$ if $Z > 0$, $= 0$ otherwise.

6. Regime Switching Models

Suppose that in every time period there are two independently generated series W_{1t}, W_{2t}, so that the value taken by W_{1t} may depend on lagged W_{1t} but not on current and lagged W_{2t} and vice-versa and that the input shocks or stochastic components, if any, are independent across and within series. Let the observed series be

$$\begin{aligned} X_t &= W_{1t} \text{ at times } t = 1, 2, \ldots, S_1 \\ &= W_{2t} \text{ at times } S_1 + 1, S_1 + 2, \ldots, S_2 \\ &= W_{1t} \text{ at times } S_2 + 1, + \ldots \\ &\text{etc.} \end{aligned}$$

where the times of the switches are generated by some mechanism which may be deterministic or stochastic and can depend on the value of X_t and its lags. This could be thought of as a "pure switching model" as the two processes that one switches between are unrelated. This formulation is little used. More commonly the underlying processes W_{1t}, W_{2t} are inter-related so that in a linear framework they may be generated by a vector autoregression (VAR). A yet more common form of model has (in the univariate case)

$$X_t = f(X_{t-1}, \underline{\theta}_t) + \varepsilon_t \tag{6.1}$$

where the Markovian, homoskedastic case is shown for simplicity, and the vector of parameters is itself following a "pure switching model" usually deterministically generated but stochastically switching. A successful example is the Markov Switching Model, in which parameters in (6.1) are fixed, but different in different regimes leading to switching specifications, the switches occurring purely stochastically according to a Markov chain process (see Hamilton 1989, 1994 for discussions). Here the number of states is fixed at two in the example above but it clearly can be larger. An example is:

$$X_t = \alpha_1 X_{t-1} + \varepsilon_t \quad \text{when in state 1}$$
$$X_t = \beta_2 X_{t-2} + \varepsilon_t \quad \text{when in state 2}$$

which is just the specification

$$X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$$

with $\alpha = \alpha_1, \beta = 0$ in state 1, $\alpha = 0, \beta = \beta_2$ in state 2. Possibly ways to generalize, or complicate, such models are clear. Rather than having the switch occur exogenously through the Markov chain one could introduce exogenous variables, involving the state of the business cycle, for example. The switch could also occur deterministically. An example has, with monthly data,

$$X_t = \alpha_S X_{t-1} + \varepsilon_t$$

where the value of α_S depends on the month, so that α_S is different for January than for September but repeats cyclically year by year. This class of seasonal series has been extensively considered by Franses (1996). They have the property that the standard concept of seasonal adjustment does not seem to apply as seasonality is deeply imbedded in the fabric of the model, rather being an additive or multiplicative component.

All of these models involve sharp switches from one regime to another. Although one can argue that an individual decision matter may use decisions that induce such breaks, if investigating aggregate data where individuals have breaks at different points, the outcome might be a smooth transition from one regime to another. In the previous notation, this would suggest models of the form

$$X_t = \alpha_t f(X_{t-1}, \theta_{1t}) + (1 - \alpha_t) f(X_{t-1}, \theta_{2t}) + \varepsilon_t$$

again in the univariate, homostochastic case, where $0 \leq \alpha_t \leq 1$ and α_t is some smooth switching variable, either deterministic or stochastic, θ_{1t} are the parameter values in regime 1, θ_{2t} the values in regime 2. An example would have α_t a function of capacity utilization, so that α_t near 1 would correspond to the peak of the business cycle and then most of the weight is on regime 1, but α_t would be near 0 when utilization is low, corresponding to a trough in the cycle, and so X_t would be in regime 2. For other values of α_t, X_t would be in transition. These models have been discussed in Granger and Teräsvirta (1993) with α_t being specified as a logistic function of driving variables, either lagged X_t or causal variables, called smooth transition autoregressive (STAR) and smooth transition regression (STR) models.

A situation that is frequently being considered in recent research is where there are many possible regimes and the parameter values corresponding to each are unknown. Occasionally, due to a “structural break” the economy may switch to a new regime with unknown specification, so that a learning process has to occur. Adaptive, unit root processes perform relatively well directly after the break and a time-varying parameter model, based on a state-space formalization and the Kalman filter, will do relatively well over a somewhat longer period, and, if there is stability in a regime long enough, eventually a more complicated, even nonlinear specification can be used.

One special class of models that has been frequently discussed but probably less frequently used, at least in economics, are those that are designed to have specific, simple, linear temporal properties whilst at the same time producing a series with a pre-specified marginal distribution. A normal distribution is easy, of course, but various tricks, including switching regimes, have been used for other marginal distributions. For example, Garver and Lewis (1980) suggested the “exponential autoregressive” (EAR(1)) model

$$X_t = \alpha X_{t-1} + \begin{matrix} O & \text{with probability } \alpha \\ \varepsilon_t & \text{with probability } 1-\alpha \end{matrix}$$

where ε_t is iid, exponentially distribution. X_t has the autocorrelations $\rho_k = \alpha^k$ of an AR(1) process and has a marginal (not conditionally marginal) exponential distribution. Various other specifications produce processes with the same properties.

A few other specific marginal distributions have been considered including the Pareto, logistic, and gamma.

The basic regime switching model has also been adapted for the unit root/cointegration/error-correction models, with switching cointegrations or time-varying parameters in error correction models, for example Siklos and Granger (1997). There is no particular difficulty in extending the specification of any nonlinear model in the same way.

7. Models of Volatility

Before discussing the topic of this section it is perhaps worth returning to the topic of the definition of linearity and thus of nonlinearity. Suppose that a positive series X_t has the property

$$E[X_t | X_{t-1}] = \alpha X_{t-1} \qquad (7.1)$$

so that the expected mean is linear in the explanatory variable. Define $Y_t = \log_e X_t$, then (7.1) can be written

$$E[\exp(Y_t) | Y_{t-1}] = \alpha \exp(Y_{t-1}) \qquad (7.2)$$

which at first sight may seem to be nonlinear, but according to the definition used in Section 1, it is linear in the variable $\exp(Y_t)$, i.e. X_t. This rather trivial discussion has some relevance when X_t is a series having zero conditional mean, so that $\mu_t = E[X_t | I_{t-1}] = 0$, $I_{t-1} : X_{t-j}, j > 0$, and consider the conditional variance

$$h_t = E[X_t^2 | I_{t-1}]. \qquad (7.3)$$

Suppose that

$$h_t = w + aX_{t-1}^2 \tag{7.4}$$

so that the specification considered is

$$X_t^2 = w + \alpha X_{t-1}^2 + \eta_t$$

where

$$E[\eta_t | X_{t-j}, j \geq 1] = 0$$

and so the model is linear in X_t^2. (7.4) is know as the ARCH(1) specification, for Autoregressive Conditional Heteroskedasticity of order 1, as only one lag is used. Clearly more lags can be considered. A straightforward extension is the generalized ARCH, or GARCH, which in its (1.1) form is

$$h_t = w + \alpha X_{t-1}^2 + \beta h_{t-1}$$

although more lags can be included. ARCH was introduced by Engle (1982) and GARCH by Bollerslev (1986). Both are linear in X_t^2, or $(X_t - \mu_t)^2$. However, in most accounts of nonlinear processes they are classified as nonlinear.

Since their introduction a large variety of extensions of the basic ARCH/GARCH model have been proposed, many of these models are nonlinear. These extensions are often designed to capture particular properties of processes that might be found in practice, such as nonsymmetry. A survey of some of these models is given by Bollerslev, Engle, and Nelson (1994)
A particularly important and useful form, due to Nelson (1991) is the exponential ARCH (EARCH) which takes the form

$$\log h_t = w + (1 + \alpha(B)) \{ \theta Z_{t-1} + \gamma [| Z_{t-1} | - E [| Z_{t-1} | \}$$

where $Z_t = X_t / h_t^{½}$ and $\alpha(B)$ is a polynomial of order p in the lag operator. One can both get nonsymmetry and there are no difficulties in ensuring that the estimated value of h_t is positive.

As a final example of the many possible models in this area is one that has been found useful with financial data, introduced by Ding, Granger, and Engle (1993). It is known as the Asymmetric Power ARCH) (or A-PARCH) with specification

$$h_t^{\delta} = w + \sum_{i=1}^{p} \alpha_i (|X_{t-1}| - \gamma_i X_{t-i})^{\delta} + \sum_{j=1}^{q} \beta_j h_{t-j}^{\delta}$$

with $w > 0$, $\delta > 0$, $\alpha_i \beta_i \geq 0$ for all i.

A stochastic volatility model (SV) will take the form

$$X_t = \mu_t + \sigma_t \varepsilon_t$$

where μ_t is the conditional mean and σ_t is a changing volatility. However, unlike ARCH models where σ_t is modeling in term of potentially observed components, such as ε_{t-j}^2, is stochastic volatility models only unobserved components are used, so that one may have

$$\log \sigma_t = \alpha + \varphi [\log \sigma_{t-1} - \alpha] + \theta \eta_t$$

for example, with η_t iid, zero mean. Alternative specifications are used, sometimes $\varphi = 1$ is assumed but if the distribution of η_t is assumed to be Gaussian, say, the parameters of the model for $\log \sigma_t$ can be estimated, if the specification is correct. It should be noted that the SV and ARCH models are not nested. Further, SV models have two sets of input stochastic shocks, ε_t and η_t usually assumed to be independent, which are not invertible, so that observing X_t will not allow one to observe (i.e. estimate) either ε_t and η_t unless possibly an earlier value of these series is known without error.

8. Empirical Models

Under this heading I will consider a sampling of models or techniques that do not start with a specific specification but are allowed to evolve in the modeling process to provide a satisfactory in-sample data-fit.

A neural network model, in which X_t is explained in terms $\underline{Y}_t$, takes the form

$$X_t = \sum_{j=1}^{q} \theta_j \varphi(\underline{\beta}'_j \underline{Y}_t) + \varepsilon_t$$

where φ ( ) is some bounded function. Thus one has a linear combination of a nonlinear function of linear combinations of the explanatory variables. A popular choice for φ is the logistic function but the specific choice is not critical. If the true relationship is

$$X_t = F(\underline{Y}_t) + e_t$$

then a neural network model can provide a close approximation if F( ) is bounded for q large enough and possible also if F ( ) is not bounded in some directions. A successful use of neural nets is discovering very good approximations to maps of chaotic processes based on fairly short series. A discussion of these models and their estimation can be found in White (1989).

Nonparametric modeling of time series does not require an explicit model, which is the strength of the procedure. Essentially a smoothing or averaging filter is applied to $\underline{Y}_t$ associated with each X_t and thus a diagrammatic, possibly multi-dimensional, is obtained. References are Ullah (1989) and Härdle (1990).

An alternative, intermediate model is "projection pursuit" where initially

$$X_t = S(\underline{\beta}'_1 \underline{Y}_t) + \varepsilon_{1t}.$$

S is a nonparametric smoother and the β's are chosen as an optimum projection, a second dependent variable is formed as $X_t - S(\beta'_1 Y_t)$ and is modeled as $S(\underline{\beta}'_2 \underline{Y}_t) + \varepsilon_{2t}$ and so forth sequentially until the variance of the residual is not reduced significantly by adding further terms.

These, and other techniques that fall in the same category, such as splines, have similar strengths and weaknesses. They are flexible and will search over many possibilities to find a good fit. On the other hand they are inclined to overfit and need to be checked by cross-validation and by post-sample forecastability tests, the models are usually invalid for trending or nonstationary series and the individual parameter values are not interpretable in economic terms in most cases. It is also true that sometimes the methods do not provide satisfactory forecasts if the explanatory variables take post-sample values outside of the range observed in sample.

9. Some Final Comments

For an example of having two stochastic shocks which are potentially observable, suppose that we have a pair of invertible processes Y_{1t}, Y_{2t} with

$$Y_{1t} = g_1(Y_{1,t-1}, \varepsilon_{1t})$$
$$Y_{2t} = g_2(Y_{2,t-1}, \varepsilon_{2t}) \qquad \text{say}$$

and X_t, the directly observable series being decomposed as $X_t = \text{sign}\,(X_t)|X_t|$ with $\text{sign}\,(X_t) = Y_{1t}$ $|X_t| = Y_{2t}$. This may be called the multiplicate observable decomposition.

An example of a nonlinear process which has been found to have misleading linear properties is the simple stationary NLAR (1)

$$X_t = \text{sign}\,(X_{t-1}) + \varepsilon_t$$

where ε_t is Gaussian iid, $N(0, \sigma^2)$. If σ is small compared to 1, the process will be essentially regime switching. Let $p = \text{Prob}(1+\varepsilon)<0 = \text{Prob}(\varepsilon-1)>0$ then the theoretical autocorrelations for X_t are

$$\rho_k = (1-2p)^k$$

and so are short memory. However, simulations show that if p is small, the process will switch only occasionally and the process is a sample of 2 thousand to 20 thousand terms will appear to be I(d), a factional integrated (a form of $1/\omega$) process, sometimes called long-memory.

This is just a brief over-view and many aspects of nonlinearity have not been covered. There is no claim that it is comprehensive.

References

Abraham, B. and A. Thavaneswaran (1991): "A Nonlinear Time Series Model and Estimation of Mission Observations." *Ann. Inst. Statist. Math.* 43, 453-504.

Bollerslev, T., R.F. Engle, and D.B. Nelson: "ARCH Models." in "Handbook of Econometrics, Volume IV", ed. R.F. Engle and D.L. McFadden. Elsevier, Amsterdam.

Ding, Z., C.W.J. Granger, and R.F. Engle (1993): "Long Memory Properties of Stock Market Returns and a New Model." *Journal of Empirical Finance* 1, 83-106.

Franses, P.H. (1996): "Perodicity and Stochastic Trends in Economic Time Series." Oxford University Press.

Granger, C.W.J., T. Inoue, and N. Morin (1997): "Nonlinear Stochastic Trends." *Journal of Econometrics* 81, 65-92.

Granger, C.W.J. and N. Hyung (1998): "Introduction to M-M Processes." UCSD Department of Economics Working Paper.

Granger, C.W.J. and T. Teräsvirta (1993): "Modeling Nonlinear Economic Relationships." Oxford University Press.

Hamilton, J.D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57, 357-384.

Hamilton, J.D. (1994): "Time Series Analysis." Princeton University Press.

Härdle, W. (1998): "Applied Nonparametric Regression." Oxford University Press.

Kershing, E. (1988): "On Recurrence and Transience of Growth Models." *Journal of Applied Probability* 22, 614-625.

Meyn, S.E. and R.L. Tweedie (1993): "Markov Chains and Stochastic Stability." Springer-Verlag: Berlin.

Mittnik, S. (1991): "Nonlinear Time Series Analysis With Generalized Autoregressions: A State Space Approach." Working paper, Department of Economics, State University of New York, Stoneybrook.

Nelson, D. B. (1991): "Conditional Heterskedasticity in Asset Returns: A New Approach." *Econometrica* 59, 347-370.

Siklos, P. and C.W.J. Granger (1997): “Regime-Sensitive Cointegration With An Application to Interest Rate Parity.” *Macroeconomics Journal* 1, 640-657.

Teräsvirta, T., D. Tjøstheim, and C.W.J. Granger (1994): “Aspects of Modeling Nonlinear Time Series,” in “Handbook of Econometrics, Volume IV”, ed. R.F. Engle and D.L. McFadden. Elsevier, Amsterdam.

Tjøstheim, D. (1986): “Some Doubly Stochastic Time Series Models.” *Journal of Time Series Analysis* 7, 51-72.

Tong, H. (1983): “Threshold Models in Nonlinear Time Series Analysis.” Springer-Verlag, Berlin.

Tweedie, R.L. (1975): “Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space.” *Stochastic Process Applications* 3, 385-407.

Ullah, A. (1989): “Semiparametric and Nonparametric Econometrics.” Physica-Verlag, Heidelberg.

White, H. (1989): “Some Asymptotic Results for Learning in a Single Hidden-Layer Feedforward Network Models.” *Journal of the American Statistical Association* 84, 1003-1013.