Nonlinear Time Series Analysis

Howell Tong
London School of Economics and Political Science

November 4, 2015

1. Introduction.

A function f from R^p to R is said to be *linear* if for vectors $x, y \in R^p$ and any real scalar α, $f(\alpha x + y) = \alpha f(x) + f(y)$. Any function f that is not linear is said to be *nonlinear.*

In the analysis of stationary time series, the spectral density function, if it exists, is nonlinear under the above definition. However, for reasons to be made clear later, a statistical analysis that is based on it or its equivalents is ordinarily considered a linear analysis. Often, a time series is observed at discrete time intervals. For a discrete-time stationary time series $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$ with finite variance, $corr(X_t, X_{t+s})$ is a function of s only, say $\rho(s)$, and is called the auto-correlation function. The spectral density function is the Fourier transform of $\rho(s)$ if $\sum_{s=-\infty}^{\infty} |\rho(s)| < \infty$. Now, Yule (1927) introduced the celebrated autoregressive model in time series. Typically the model takes the form

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t, \tag{1}$$

where the α's are parameters and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance, or a white noise for short. It is commonly denoted as an $AR(p)$ model. Clearly X_t is a linear function of $X_{t-1}, \ldots, X_{t-p}, \varepsilon_t$. Under the assumption of normality,

the distribution of the time series is completely specified by its constant mean, constant variance and $\rho(s)$'s. Perhaps for the close connection with the analysis of linear models (of which the autoregressive model is one), an analysis based on the autocorrelation function or equivalently the spectral density function is loosely referred to as a linear analysis of the time series. By the same token, an analysis based on higher order moments or their Fourier transforms is loosely called a nonlinear analysis. Broadly speaking, tools based on the Fourier transforms of moments constitute what is called the frequency-domain approach, while those based on the moments constitute the time-domain approach, which often includes building a time series model of the form (1) or its generalizations.

Similar discussion as the above can be extended to cover $\{X(t) : t \in R\}$

2. Can we do without nonlinearity?

A general answer is in the negative simply because the dynamical laws governing Nature or human activities are seldom linear. In the real world, we can see the footprints of nonlinearity everywhere we look. Below are a few examples.

(a) *Phase Transition*

The melting of ice of a glacier will alter fundamentally the amount of water flowed in a river near the glacier. Phase transition (from solid to liquid in the above example) is an importance signature of nonlinearity. Animals behave differently (e.g. hunting effort) during time of short food supply versus time of abundant food supply.

(b) *Saturation*

In economics, diminishing return is a well-known phenomenon: doubling your effort does not necessarily double your reward.

(c) *Synchronization*

The celebrated Dutch scientist, Christiaan Huygens, observed that clocks placed on the same piece of soft timber were synchronized! Biological systems can also exhibit synchronization. It has been noted that girls sharing the same dormitory have higher chance of synchronizing their menstruation. Even female keepers of baboons have been known to have similar experience.

(d) *Chaos*

When we toss a coin to randomize our choice, we are exploiting nonlinearity,

for the dynamical system underlying the tossing is a set of (typically three) nonlinear ordinary differential equations, the solution of which is generally very sensitive to the initial spinning unless we 'cheat'. The system is said to generate chaos in a technical sense. When statisticians generate pseudo-random numbers, they are also generating chaos. One of the most commonly used pseudo-random generator is the linear congruential generator, which is a piecewise linear (i.e. nonlinear) function that does precisely this. It might surprise you that you are actually using nonlinear devices almost daily because encrypting passwords is closely related to pseudo-random number generation.

In the following sections, we focus on the time-domain approach because at the current state of development, this approach tends to admit simpler interpretations in practical applications.

3. What is a nonlinear time series model?

A short answer is that it is not a linear time series model. This raises the need to define a linear model. A fairly commonly adopted definition is as follows. A stationary time series model is called a linear time series model if it is equivalent (for example in the mean-square sense) to

$$X_t = \sum_{s=-\infty}^{\infty} \beta_s \varepsilon_{t-s}, \tag{2}$$

where $\{\varepsilon_t\}$ is a white noise and the summation is assumed to exist in some sense. An alternative definition due to Hannan (1973) is one that requires that the minimizer of $E|X_t - h(X_{t-1}, X_{t-2}, \ldots)|^2$ with respect to h over the space of all measurable functions is the linear function. Here the mean square is assumed to exist.

4. Are linear time series models fit for purpose?

Examples abound of the inability of linear time series models to capture essential features of the underlying dynamics.

Yule (1927) introduced the autoregressive model to model the annual sunspot numbers with a view to capturing the observed 11-year sunspot cycle but noted the inadequacy of his model. He noted the asymmetry of the cycle and attempted to model it with an $AR(4)$ model only to discover that it gave statistically a worse fit than a simpler AR(2) model.

Moran (1953) fitted an $AR(2)$ model to the annual lynx data corresponding to the MacKenzie River region in Canada, with a view to capturing the observed 10-year cycle. He was quick to point out that the fitted residuals were heteroscedastic.

Whittle (1954) analyzed a seiche record from Wellington Bay in New Zealand. He noted that, besides the fundamental frequency of oscillations and a frequency due to the reflection of an island at the bay, there were sub-harmonics bearing an interesting arithmetic relation with the above frequencies. Now, sub-harmonics are one of the signatures of nonlinear oscillations, long known to the physicists and engineers.

5 Examples of nonlinear time series models.

First, we describe parametric models. Due to space limitation, we describe the two most commonly used models. For other models, we refer to Tong (1990). We shall describe (i) the threshold model and (ii) the (generalized) autoregressive conditional heteroscedasticity model, or in short the TAR model and the (G)ARCH model respectively. The former was introduced by Tong in 1977 and developed systematically in Tong and Lim (1980) and Tong (1983, 1990), and the latter by Engle (1982), later generalized by Bollerslev (1986).

There are several different but equivalent ways to express a TAR model. Here is a simple form. Let $\{Z_t\}$ denote an indicator time series that takes positive integer values, say $\{1, 2, \ldots, K\}$. Let $\{\eta_t\}$ denote a white noise with zero mean and unit variance, $\alpha_0^{(j)}, \alpha_i^{(j)}, \beta^{(j)}$ be real constants for $j = 1, 2, \ldots, K$. Then a time series $\{X_t : t = 0, \pm 1, \pm 2, \ldots\}$ is said to follow a *threshold autoregressive model* if it satisfies, when $Z_t = j, \;\; j = 1, \ldots, K$,

$$X_t = \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} X_{t-i} + \beta^{(j)} \eta_t. \tag{3}$$

For the case in which $Z_t = j$ if and only if $X_{t-d} \in R_j$ for some positive integer d and for some partition of R, i.e. $R = \bigcup_{i=1}^{K} R_i$ say, the TAR model is called a *self-exciting threshold autoregressive model*, or SETAR model for short. In this case, given $X_{t-s}, s > 0$, the conditional mean of X_t is piecewise linear, and the conditional variance of X_t piecewise constant.

For the case in which $Z_t = j$ if and only if $Y_{t-d} \in R_j$ for some covariate time

series $\{Y_t\}$, some positive integer d and some partition of R, i.e. $R = \cup_{i=1}^{K} R_i$ say, then we have a TAR model driven by (or excited by) $\{Y_t\}$. Note that the covariate time series $\{Y_t\}$, and thus the indicator time series $\{Z_t\}$, can be observable or hidden. If the indicator time series, whether observable or hidden, forms a Markov chain, then we call $\{X_t\}$ a Markov-chain driven TAR; this model was first introduced by Tong (Tong and Lim, 1980, p.285 and Tong 1982, p.62). In the econometric literature, the sub-class with a hidden Markov chain is commonly called a *Markov switching model*.

The TAR model, especially the SETAR model, has many practical applications in diverse areas/disciplines, including earth sciences, ecology, economics, engineering, environmental science, finance, hydraulics, medical science, water resources and many others.

The nonlinear parametric model that is mostly and widely used in econometrics and finance is the (G)ARCH model. The ARCH model is given by

$$X_t = \eta_t \sigma_t, \tag{4}$$

where $\{\eta_t\}$ is as defined previously but sometimes assumed to be Gaussian, and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \quad \alpha_0 > 0, \alpha_i \geq 0, i = 1, \ldots, p$. Note that the ARCH model differs from the SETAR model in its σ_t being a continuous function instead of a piecewise constant function as in the latter. The GARCH model generalizes σ_t^2 to $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$, where the β_is are usually also assumed to be non-negative, although the non-negativity assumption may be relaxed; see Cryer and Chan (2008, Chapter 12).

One of the limitations of any parametric modelling approach is the subjectivity of selecting a family of possible parametric models. We can sometimes mitigate the situation if a certain parametric family is suggested by subject matter considerations. In the absence of the above, mitigation is weaker even if we are assured that the family is dense in some sufficiently large space of models. It is then tempting to allow the data to suggest the form of F where we are contemplating a model of say

$$X_t = F(X_{t-1}, \ldots, X_{t-p}, \varepsilon_t), \tag{5}$$

F being unknown. This is one of the strengths of the nonparametric modelling approach, which is a vast and rapidly expanding area. A word of caution is

the so-called curse of dimensionality, meaning that when $p > 3$ the estimated F is unlikely to be reliable unless we have a huge sample size. One way to ameliorate the situation is to replace $X_{t-1}, \ldots, X_{t-p}$ by $\xi_{t-1}, \ldots, \xi_{t-q}$ with q much smaller than p, e.g. $q = 1$ or 2. The ξ's are typically suitably chosen but unknown linear functions of X's, sometimes called indices. This is called the semi-parametric modelling approach, which is also a rapidly expanding field. For comprehensive accounts of the above developments, see, e.g., Fan and Yao (2005) and Gao (2007). Another way is to impose some simplifying structure on (5) such as zero interaction as in Chen and Tsay (1993), who gave

$$X_t = F(X_{t-1}) + \cdots + F(X_{t-p}) + \varepsilon_t. \tag{6}$$

ACKNOWLEDGEMENTS

Reprinted with permission from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science. Heidelberg: Springer Science +Business Media, LLC

REFERENCES

1. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics.* vol. 31, 307-327.

2. Chen, R. and Tsay, R.S. (1993). Nonlinear Additive ARX Models. *J. Amer. Stat. Ass.* vol. 88, pp. 955-967.

3. Cryer, J.D. and Chan, K.S. (2008) *Time Series Analysis: With Applications in R.* New York: Springer Verlag.

4. Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica.* vol. 50, 987-1008.

5. Fan, J. & Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods.* New York: Springer-Verlag.

6. Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* London: Chapman and Hall/CRC.

7. Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Applied Probability.* vol. 10, 130-145.

8. Moran, P.A.P. (1953). The statistical analysis of the Canadian lynx cycle. *Aust. J. Zoology.* vol. 1, 163-173.

9. Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis.* New York: Springer Lecture Notes in Statistics.

10. Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

11. Tong, H. & Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Stat. Soc.* series B, vol. 42, 245-292.

12. Whittle, P. (1954). The statistical analysis of a seiche record. *Sears Foundation J. Marine Research.* vol. 13, 76-100.

13. Yule, U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Soc., London*, series A, vol. 226, 267-298.