Nonlinear time series analysis since 1990: some personal reflections

Howell Tong
University of Hong Kong & London School of Economics

Abstract

I reflect upon the development of nonlinear time series analysis since 1990 by focusing on five major areas of development.

Keywords: chaos; common structure; curse of dimensionality; embedding dimension; financial time series; initial value sensitivity; local polynomial smoother; long memory; Markov chain Monte Carlo; nonlinear dynamical systems; nonlinear state space models; nonlinear time series; nonparametrics; panel of time series; projection pursuit; semi-parametrics; test for determinism; thick tail; volatility.

1. Introduction

Nonlinear time series analysis (NTSA) began to attract serious attention in the literature in the early 1980s due to the realisation that linear time series models are incapable of reflecting many important real phenomena, e.g. asymmetric business cycles, sustained animal population cycles, volatility of stock markets, regime switchings and many others. Tong (1990) has given a fairly comprehensive account of major developments of the new subject in the 1980s. It can be said that this period was dominated by the development of parametric models. In particular, two classes of models, namely the threshold autoregressive (TAR) models and the (generalised) autoregressive conditional heteroscedastic ((G)ARCH) models, developed during these early days, seem to have withstood the test of time and continue to attract varying degrees of attention among practitioners in the wider scientific and financial communities, which include biology, dynamical systems, ecology, economics/econometrics, financial engineering and many others.

The 1990s saw movement of NTSA into many important and exciting directions. My personal choice, shared by some other enthusiasts in NTSA, lists the following as the five most promising directions: the interface between

NTSA and chaos, the nonparametric/semiparametric approach, nonlinear state space modelling, financial time series (in both discrete time and continuous time) and nonlinear modelling of panels (e.g. spatially distributed) of time series. Needless to say, some of these are still at an embryonic stage.

I shall give some personal reflections on these five directions. A less than formal style will be adopted with the hope that it will help the overview appealing to a wider audience. In many cases, only representative references are given with no intention of slighting the un-cited contributions.

2. Interface between NTSA and Chaos

Nearly a hundred years ago, Poincar'e (1908) stated explicitly that one fundamental source of randomness is the initial-value sensitivity of a dynamical system. That is to say even a deterministic dynamical system can generate randomness as long as it is sensitive to initial values. Unfortunately, this truth was lost to many traditional educational systems. Then all of a sudden, toy-models such as $x :\rightarrow 4x(1-x), \quad x \in (0,1)$, and the fact that they possess unexpectedly complicated structures began to fascinate even the man in the street.

Chaos reached the street in the 1980s. During this period, enthusiasm for Chaos [1] led some people to think that the days of statistical science were numbered. Rather curiously, many statisticians remained indifferent! However, a small number of researchers started to take up the challenge of exploring the interface between Statistics and Chaos, with NTSA providing a natural platform for this endeavour.

One of the key notions to explore is that of initial-value sensitivity. To explain this, consider a deterministic difference equation

$$y_{t+1} = f(y_t), \quad t = 0, 1, \cdots . \tag{1}$$

Let us consider the simple case of two solutions with infinitesimally different initial conditions $y_0 = y$ and $y_0 = y + dy$ so that $dy_0 = dy$. Then after one

[1] I prefer not to use the term Chaos Theory because Chaos is really part of the theory of dynamical systems

iterate, the infinitesimal difference between the two solutions is clearly, $dy_1 = f'(y_0)dx$, where f' denotes the first derivative of f (assumed to exist). After m iterates, the chain rule of differentiation implies that the infinitesimal difference between the two solutions becomes $dy_m = f'(y_0)f'(y_1)\cdots f'(y_{m-1})dy$. The product $f'(y_0)f'(y_1)\cdots f'(y_{m-1})$ is of fundamental interest. Notice that if it is greater than 1 in modulus, small initial difference, dy, is amplified. Notice further that, for the same difference equation, different initial values can have different amplifications; thus amplification is typically state-dependent.

In the early 1990s, nonlinear time series analysts have succeeded in generalising the above notion to a stochastic difference equation, i.e. a nonlinear autoregressive model of the form

$$Y_{t+1} = f(Y_t) + \epsilon_{t+1}, \tag{2}$$

where $E[\epsilon_t|Y_0, \cdots, Y_{t-1}] = 0, \quad t \geq 1$. Essentially, the product

$$f'(y_0)f'(y_1)\cdots f'(y_{m-1}) \tag{3}$$

has now been generalised to

$$E\{f'(Y_0)f'(Y_1)\cdots f'(Y_{m-1})|Y_0 = y\}, \tag{4}$$

and others. Chan and Tong (2001) has discussed these notions in a stochastic setting.

In equation (2), the argument of f can be extended to include $Y_{t-1}, \cdots, Y_{t-d+1}$ besides Y_t, giving us a nonlinear autoregressive model of order d. This order plays a similar role as the so-called 'embedding dimension' in Chaos. An important statistical challenge is its estimation. Although Chaos specialists have proposed several methods, they tend to be *ad hoc* in that little attention seems to have been paid to the sampling properties of the estimates. Again nonlinear time series analysts have succeeded in producing estimates of d which are consistent and require only realistic sample size. The key is the use of the local constant (also called the Nadaraya-Watson) nonparametric regression (-to be discussed in the next section) coupled with the application of either Akaike's final prediction approach or the cross-validation principle by deleting one observation at a time in the regression function fitting. Chan

and Tong (2001) has given details.

Even with purely deterministic difference equations, nonlinear time series analysts have made refreshing discoveries. For example, An (1996) described a method of generating pseudo-random numbers via a suitable chaotic map with the property that, starting from any initial value, the empirical distribution function of the generated pseudo-random numbers converges *everywhere* to a specified distribution. For other examples, see Chan and Tong (2001).

Nonlinear time series analysts have also enriched and provided theoretical underpinnings for several basic tools in Chaos. Notable ones are the estimation of the correlation dimension and the surrogate method. See Chan and Tong (2001) for details.

To conclude this section, I would like to recall the words of Cox (1997), who, when commenting on nonlinearity as a widely occurring theme in modern statistics as in modern mathematics, said 'Nonlinear time series models provide one important example and ... The link with chaos is clear.' In this context, it is encouraging that significant progress has been made in forging the link. However, many important open problems still remain. For example, as yet I am not aware of any satisfactory test for the null hypothesis that the noise component in equation (2) is negligible-the so-called test for (operational) determinism. Chan and Tong (2001) has listed other open problems.

3. Nonparametric/semiparametric approach

The emergence of the nonparametric approach and later the semiparametric approach in NTSA is a natural development for at least three reasons: (1) Modern computing power has enabled us to experiment with fewer and fewer assumptions on the generating mechanism of the underlying dynamics. (2) Connections with fields such as Chaos have required that estimation of many of the 'invariants' (e.g. initial-value sensitivity) of the generating mechanism be constructed free from the 'prejudice' typically imposed by a parametric approach. (3) Nonparametrics and/or semiparametrics can sometimes aid the selection of a suitable parametric model, especially if data are plentiful.

Tjøstheim (1994) reviewed NTSA with special emphasis on nonparametrics till that time. Fan and Yao (2002), when available, may well provide a systematic account. Despite the comparatively long history of nonparametric regression for independent data (-e.g. regressogram is at least 40 years old), it seems that Robinson (1983) was the first to give a serious theoretical development of the use of the local constant smoother/estimation in nonlinear time series modelling. As far as the testing for nonlinearity is concerned, we had to wait for almost another 10 years before a nonparametric approach first emerged. As far as I am aware, it was An and Cheng (1991) who first developed a Kolmogorov-Smirnov type test for nonlinearity.

To fix ideas, let us consider a stochastic regression formulation:

$$Y_t = f(X_t) + \epsilon, \tag{5}$$

where X_t is a $d \times 1$ vector, called a stochastic regressor or a covariate, and $E[\epsilon_t | Y_0, \cdots, Y_{t-1}] = 0, \quad t \geq 1$. The above model reduces to a nonlinear autoregressive model of order d if $X_t = (Y_{t-1}, \cdots, Y_{t-d})^T$. Suppose that $\{(X_t, Y_t) : 1 \leq t \leq n\}$ denote a sample from a strictly stationary time series $\{(Y_t, X_t)\}$. Let $K(.)$ denote the so-called smoothing kernel typically in the form of a probability density function on R^d. Minimizing

$$\sum_{t=1}^{n} \{Y_t - a\}^2 K(\frac{X_t - x}{h}), \tag{6}$$

with respect to a leads to the so-called Nadaraya-Watson estimate of $f(x)$. Here, h controls the amount of local smoothing and is called the bandwidth. The Nadaraya-Watson estimate is really a local constant smoother, which can be generalised to a local linear smoother by incorporating the linear term in the above sum of squares, namely

$$\sum_{t=1}^{n} \{Y_t - a - b^T (X_t - x)\}^2 K(\frac{X_t - x}{h}). \tag{7}$$

The minimizer, say $(\hat{a}, \hat{b})$, is the local linear estimate of $f(x)$ and of its partial derivatives (with respect to the components of x). In principle, the local linear smoother could be further generalised to local polynomial

smoother/estimate by incorporating higher order terms. However, in practice, it is often the case that local linear smoother will suffice, bearing in mind also the increasing computation with each additional higher order term. It is generally accepted that local polynomial smoothing enjoys a number of good properties in terms of bias and variance, absence of boundary effects and others. (See, e.g. Fan and Gijbels, 1996.) To-date, the local polynomial approach (especially at low degrees) enjoys particular popularity in non-linear time series analysis. See, e.g., Chan and Tong (2001a) and Fan and Yao (2002). Among the other nonparametric approaches which have also featured in the nonlinear time series literature are the spline, the neural network and the wavelet. It seems fair to say that, despite their potential, their combined impact to-date is not as great as the local polynomial approach.

The main disadvantage of the non-parametric approach is the so-called curse of dimensionality. Slightly curious is the fact that this oft-used term is not as often given a precise definition or characterisation. One possible characterisation is the need for the sample size to be of the order of κ^d, where κ is some positive constant and d the dimension of X_t. It is therefore not surprising that the consistency rate for a nonparametric estimate is generally much slower than the typical root-N rate in the parametric approach.

A half-way house is the so-called semi-parametric approach, which begins to feature more prominently in recent non-linear time series literature. The approach has, of course, a much longer history outside time series analysis. For example, one of the most illustrious examples of the approach is the projection pursuit regression proposed by Friedman and Stuetzle (1981). This and other semi-parametric models can be put in the form

$$y_t = g(B^T X_t) + \epsilon_t; \tag{8}$$

Effectively $f(X_t)$ in equation (5) is replaced by $g(B^T X_t)$, where B is a $p \times d$ matrix with $p \leq d$. The basic idea is that in many practical applications, p may be substantially smaller than d. If this is the case, the curse of dimensionality can then be significantly ameliorated. Here, g is typically an unknown (non-parametric) function to be estimated. Often the literature considers the estimation of the parameters B, sometimes called the indices (perhaps following the usage in econometrics) or directions, after 'undersmoothing' the estimate of the unknown function g, in order to achieve

root-N consistency for the parametric estimation. By under-smoothing is meant deliberately choosing a smaller bandwidth. Unfortunately, this seems to me more of a theoretical trick than a genuine practical proposition because I am not aware of any generally accepted recommendation of how to under-smooth in practice. The new millennium has started with exciting new developments. For example, Xia *et al.* (2001) has developed a methodology in which, besides other features, under-smoothing is not necessary to achieve the above-mentioned objective. Li and Chan (2001) has considered the broader issue of a vectorial response time series Y_t in model (8).

4. Nonlinear State Space Modelling

Let $\{Y_t\}$ denote a scalar time series. Suppose that by stacking the Ys to form $X_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p+1})^T$, $\{X_t\}$ is a first-order p-dimensional Markov chain (with state space R^p). We have seen this operation in the last section. Now, one well-known approach to nonlinear time series modelling is via a stochastic difference equation of the following form:

$$X_{t+1} = G(X_t, \epsilon_{t+1}), \tag{9}$$

where G is a vector valued function and $\{\epsilon_t\}$ is a sequence of independent identically distributed p dimensional random vectors. The class of nonlinear autoregressive models is a special case of this approach. Typically modelling is directed at estimating the unknown function G and the noise distribution, namely the distribution of ϵ_t. If the functional form of G is known up to an unknown parameter, then the problem reduces to the estimation of unknown parameters. If G is completely unknown but may be assumed to be sufficiently smooth, then many nonparametric methods are available, e.g. the locally polynomial method described above. Sections 1 and 2 above are cast essentially within this framework.

An alternative and relatively recent approach to nonlinear time series modelling is the so-called nonlinear state space approach. (See, e.g., Durbin and Koopman, 2001.) It focuses directly on the transition probability, namely $P(X_{t+1,j+1} < x_{j+1} | X_t = x, X_{t+1,i} = x_{t+1,i}, i \leq j)$, where $X_{t+1,j+1}$ denotes the $(j+1)$-th component of X_{t+1}. Let the above conditional probability be denoted $F_j(x_{j+1} | x, x_{t+1,i}, i \leq j)$ for short. In the case of Gaussian transition

probabilities, the conditional probabilities are fully specified by the conditional means and conditional variance-covariances and the famous Kalman filter emerges. Since Gaussianity and linearity are intimately related although not equivalent (Tong, 1990), nonlinear state space approach is typically designed to attack non-Gaussian transitional probabilities. It is clear that unless some assumptions are made on the unknown *functions*, namely the $F_j(x_{j+1}|x, x_{t+1,i}, i \leq j)$s, estimating them is an horrendous task. It is pertinent to remember that many of the optimal properties enjoyed by the Kalman filter are intimately bound up with Gaussianity, without which they may no longer be obtained.

An early serious attempt by Kitagawa (1987) approximated the corresponding conditional probability densities by piece-wise linear functions in the manner of linear splines. In the past ten years or so, this approach has attracted the attention of many statisticians and engineers. (Fitzgerald *et al.*, 2000; Kitagawa and Gersch, 1996.) Often some assumptions/approximations are made on the conditional probabilities, e.g. the above piece-wise linearisation, the mixture of normals and others, so as to ease the burden of computation. Consequently, many different nonlinear filters have been developed, each claiming superiority over the previous ones. It seems to me that more attention could be paid to the issue of optimality. In the linear Gaussian case, it is well known that the Kalman filter is optimal (in the least squares sense). The picture with nonlinear filters is far from being clear.

One of the most recent favourite techniques in the area of nonlinear state space is the Monte Carlo Markov Chain (MCMC), which is an important numerical technique for Bayesian statistics (Gelfand and Smith, 1990). Typically rooted in conditional probability computation, it is unquestionable that it is a very powerful numerical technique, which has completely transformed the fortunes of Bayesian statistics in practice and enabled Bayesian statisticians to think the unthinkable. However, it seems to me that stopping rules for checking whether or when a steady state (i.e. stationary distribution) is attained is still an unresolved issue. I wonder if the issue is ever resolvable. Another issue has to do with the use of pseudo-random numbers, which is often an integral part of MCMC. I have already referred to the potential danger of quasi-cycles in §2. Durbin and Koopman (2001) has developed systematically an alternative based on importance sampling.

Last but not least, the lack of interaction between the nonlinear stochastic difference equation approach and the nonlinear state space approach is glaring as well as puzzling. A relevant question is then whether the two formulations are fundamentally equivalent. Chan and Tong (2001b) has given an affirmative answer. It would be interesting to see whether this equivalence will lead to an increase in interaction or simply an excuse for maintaining the status quo.

6. Panels of Time Series

In the absence of substantive or external information, it is normally the case that the information required from the data to reveal any underlying nonlinear (if present) generating mechanism cannot be adequately provided by a single time series unless the latter is sufficiently long. However, in modelling ecological populations, while the ecologists have strong reasons to believe that the generating mechanism should be generally nonlinear (based on considerations such as predator-prey interaction, competition for limited resources and others), their time series data are typically short. On the other hand, it is often possible for them to obtain replicates (by e.g. repeating laboratory experiments) or near replicates (by taking field observations from nearby-sites). Similar situations exist in other applications, e.g. economics/finance (commodity prices of similar products in different economies), epidemiology (incidence rates of the same disease in similar regions) and others. A natural approach is to pool information from a collection of short time series to unravel common underlying nonlinear dynamics. Experience so far suggests that a good balance between model complexity (without forgoing nonlinearity) and data length is crucial for successful modelling of such data.

It is fair to say that NTSA analysis of a collection of typically short time series is very much at its infancy. Although Stenseth *et al.* (1999) and Yao *et al.* (2000) have shown the relevance and vast potential of the area, much remains to be done. For example, how to incorporate spatial correlation in NTSA analysis of short series remains an exciting challenge.

7. Financial Time Series

The availability of extremely high volume data at a rapid rate has given particular impetus to the study of financial time series. These series tend to possess one or more of the following special features: clusters of high volatility (i.e. large conditional variance), heavy tails of marginal distributions and long memory. Modelling such series has led to a rapidly growing field, which has attracted time series analysts, probabilists, financial engineers, risk analysts, econometricians, physicists/ex-physicists and many others. The methodologies employed include stochastic differential equations, conditional heteroscedastic models (e.g. GARCH models, stochastic volatility models), threshold models, extreme value statistics, quantile regression, MCMC and others. Concomitant with the technical diversity is the fact that financial time series can arise in many different fields of application, e.g. equity markets, exchange rates, insurance, banking and so on. This state of affairs may be appreciated even by a quick glance at recent proceedings such as Chan *et al.* (2000). The technical diversities have sometimes bewildered the new-comers but the recent book by Tsay (2002) should help smooth the entry.

To-date, the subject is still at its embryonic stage. For example, it seems to me that even the basic notion of value at risk has not been standardised, some of the commonly used measures not being coherent in a technical sense. If I have to be provocative, I would say that some of the publications are far too esoteric to be truly relevant. However, it is certain that the subject is very important, highly relevant and will lead to exciting developments in NTSA.

8. Conclusion

The history of modern time series analysis is not long; it started about one hundred years ago if Schuster's periodogram is taken the starting point. It started with the pre-modelling periodogram (periodogram and correlogram) at the beginning of the twentieth century, went through the early modelling period (Yule's linear autoregressive models and Slutsky's moving average models) in the late 1920s. Then the period between the 1950s and the 1970s saw the height of activities in spectral analysis (smoothed periodograms initiated independently by Bartlett and Tukey), linear time series modelling for both scalar time series and multiple time series.

The history of nonlinear time series analysis is even shorter; it is only about twenty years old. In the past twenty years, the subject has developed quite fast. I can safely predict that the next twenty years will see even faster and exciting developments.

REFERENCES

An, H. (1996). A note on chaotic maps and time series. In *Athens conference on applied probability and time series*, vol. II, 15-26.

An, H. and Cheng, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Int. Statist. Rev.*, **59**, 287-307.

Chan, K.S. and Tong, H. (2001a) *Chaos: a statistical perspective.* New York: Springer Verlag.

Chan, K.S. and Tong, H. (2001b). A note on the equivalence of two approaches for specifying a Markov process. *Bernoulli*, **7** ???

Chan, W.S., Li, W.K. and Tong, H. (2000). *Proceedings of the Hong Kong International Workshop on Statistics and Finance: An Interface.* London: Imperial College Press.

Cox, D.R. (1997). The current position of statistics: a personal view. *Int. Stat. Rev.*, **65**, 261-276.

Durbin, J. and Koopman, S.J. (2001) *Time series analysis by state space methods.* Oxford: Oxford University Press.

Fan, J. & Gijbels, I. (1996). *Local polynomial modeling and its applications.* London: Chapman and Hall.
Fan, J. & Yao, Q. (2002). Nonlinear time series: Nonparametric and parametric methods. Springer Verlag, New York.

Fitzgerald, W.J., Smith, R.L., Walden, A.T. and Young, P.C. (2000). *Nonlinear and nonstationary signal processing.* Cambridge: Cambridge Univ.

Press.

Friedman, J.H. and Steutzle, W. (1981). Projection pursuit regression. *J. Am. Statist. Ass.*, **76**, 817-823.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling- ased approaches t calculating marginal densities. *J. Amer. Stat. Ass.*, **85**, 398-409.

Kitagawa, G. (1987). Non-Gaussian state space modeling of non-stationary time series. *J. Am. Stat. Assoc.,*, **82**, 1032-1063.

Kitagawa, G. and Gersch, W. *Smoothness priors analysis and time series.* New York: Springer Verlag.

Li, M-C. and Chan, K.S. (2001). Semiparametric reduced-rank regression. Tech. Rep., Department of Statistics, Univ. Iowa, USA.

Poincaré, H. (1908). *Science et Méthode*, Paris: Earnest Flammarion. (English Translation: *Science and Method.* New York: Dover, 1952.)

Robinson, P.M. (1983). Non-parametric estimation for time series models. *J. Time Series Anal.*, **4**, 185-208.

Stenseth, N.C., Chan, K.S., Tong, H., Boonstra, R., Boutin, S. Krebs, C.J., Post, E., O'Donoghue, M., Yoccoz, N.G., Forchhammer, M.C. and Hurrell, J.W. (1999). Common dynamic structure of Canadian lynx populations within three climatic regions. *Science*, **285**, 1071-1073.

Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scan. J. Statist.*, **21**, 97-130.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press, Oxford.

Tsay, R. (2002). *Analysis of financial time series.* New York: J. Wiley.

Xia, Y., Tong, H., Li, W.K. and Zhu, L. (2001). An adaptive estimation of

dimension reduction space. ???????

Yao, Q., Tong, H., Finkenstädt, B. and Stenseth, N.C. (2000). Common structure in panels of short ecological time-series. *Proc. R. Soc. Lond.*, **B 267**, 1-9.