ECONOMIC RESEARCH REPORTS

IF NONLINEAR MODELS CANNOT FORECAST, WHAT USE ARE THEY?

by

James B. Ramsey

RR # 95-04 February 1995

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

IF NONLINEAR MODELS CANNOT FORECAST, WHAT USE ARE THEY?

by

James B. Ramsey
Dept. of Economics
New York Univ.
February 17, 1995

The financial and institutional support of the C.V. Starr Center for Applied Economics is gratefully acknowledged.

IF NONLINEAR MODELS CANNOT FORECAST, WHAT USE ARE THEY?

Abstract

This paper begins with a brief review of the recent experience using nonlinear models and ideas of chaos to model economic data and to provide forecasts that are better than linear models. The record of improvement is at best meager. The remainder of the paper examines some of the reasons for this lack of improvement. The concepts of "openness" and "isolation" are introduced and a case is made that open and non-isolated systems cannot be forecast; the extent to which economic systems are closed and isolated provides the true pragmatic limits to forecastability. The reasons why local "overfitting", especially with non-parametric models, leads to worse forecasts are discussed. Models and "representations" of data are distinguished and the reliance on minimum mean square forecast error to choose between models and representations is evaluated.

Classification Codes: C14, C22, C32, C43, C53

IF NONLINEAR MODELS CANNOT FORECAST, WHAT USE ARE THEY?

INTRODUCTION

Most experienced econometricians and statisticians are familiar with the forecasting performance of "classical" structural models and with that of the vector autoregressive models to name the current chief protagonists in the forecasting sweepstakes. What is not so familiar is the forecasting performance of nonlinear models.

There are two tasks contributing to the central theme for this paper. The first task is to review very briefly the empirical evidence for the benefits of nonlinear models in forecasting; the review is restricted to obtaining an idea of general performance, but without the details. The major task is to evaluate the difficulties that inference and forecasting, in particular, face if the world is nonlinear dynamic. The case is made that in the context of nonlinear models, the difficulties in forecasting are not only more serious and more prevalent, but they are qualitatively different. A key point is that forecasting involves global properties, whereas fitting is local.

The term "nonlinear" can be confusing in that it is used for a wide variety of models with widely varying assumptions contained within the maintained hypothesis. Further, as will be discussed in a subsequent section, what is nonlinear and what is not, depends as much on the formulation of the problem as on the inherent structure of the system.

I will begin with discussing "deterministic nonlinearity", which is a general term that contains models of "chaos." Nonlinear deterministic models are, as the name implies, "non linear", and are therefore an amorphous collection of models defined by exclusion from the well defined class of linear models. Nonlinear models are usually modeled by nonlinear differential, or difference equations. Models of deterministic chaos are a sub-category of nonlinear models that are characterized by the conjunction of local instability, thereby giving rise to sensitivity to initial conditions, with global stability that effectively restricts the long term domain of the dynamic orbits to a bounded compact set. Chaotic paths, notwithstanding there seemingly random time paths, are in fact representations of long term steady states of the corresponding dynamical system.

A second major strand of use of the term "nonlinear" is in the context of stochastic models. In this situation, the adjective "nonlinear" may refer to properties of models that link the conditional moments of the distributions involved to the time paths of exogenous variables, or to the nature of the time varying path of the distributions themselves. With respect to the former concept, researchers frequently restrict attention to just the first two moments, although this is more a matter of convenience than of theoretical necessity or of empirical fact.

In the subsequent development I will use the term nonlinear in all of its various senses, but will endeavor to be clear which is which at any given point in the text.

FORECASTING AND NONLINEAR MODELS: THE RECORD

The empirical literature on chaotic and nonlinear dynamics began in about 1986. Since that time, there has been an enormous amount of activity. The following comments cannot hope to begin a thorough review of the empirical literature. I will restrict myself to a brief schematic review indicating only the main results, the "bottom line" as it were. In the discussion below I will refer to a few reviews of the literature which do an excellent job of bringing the new reader up to date, see in particular Lorenz 1993, Brock, et al. 1991, Bollerslev, et al. 1990, Brock & Potter 1993, and LeBaron 1994.

The initial flurry of activity investigating the role of "chaos" in economics that began in 1986 was to calculate Lyapunov exponents and dimension numbers. An objective was to discover the presence of a deterministic chaotic structure that would be analogous to the discoveries in the experimental sciences. Unfortunately, the challenge was too great. The sparsity of data is a problem, but not the most important. There are also difficulties caused by aggregation and the fact that the sampling rate is usually far too coarse for detecting fine dynamical structure, see for example, Ramsey & Yuan 1989; Ramsey, et al. 1990. Some of the other reasons for the "non-detection" of chaos will be enunciated below when I discuss the concepts of openness and non-isolation of systems.

A more germane reason for the failure was that in the experimental sciences the empirical discovery of chaos was always achieved by examining systems that were on the boundary of the **transition** from periodicity to chaos. If we were to examine systems that are deeply into the chaotic regime, we would discover that the tools available for the detection of chaos are inadequate even in the experimental sciences, Casdagli 1989. What is needed to discover chaos in economics is a series of experiments that will enable one to evaluate behavior as the system is "forced" to undergo a phase transition from periodicity to chaos. Actual economic systems are not likely to be observed on the border line between periodicity and chaos. Thus, the most likely potential for the discovery of chaos in economics is through experimental economics.

The outcome of the effort to discover chaos in the context of economic and financial data is best summarized in the words of Granger & Terasvirta 1992: "deterministic (chaotic) models are of little relevance in economics and so we will consider only stochastic models." The theme was echoed and amplified by Jaditz & Sayers 1992, who reviewed a wide variety of research to conclude that there was no evidence for chaos, but that was not to deny the indication of nonlinear dynamics of some sort, Brock & Potter 1993; LeBaron 1994.

The more limited objective of finding a data dependent method for determining the "dimension" of an economic system has been an equal failure so far. The importance of this task is clear. If one can obtain an estimate of the number of degrees of freedom within an economic system, then the task of the econometrician is greatly simplified; and if, for reasons to be discussed later, the number of degrees of freedom tends to vary by small amounts over time, the econometric benefits from dimension estimation would be even greater. Unfortunately, the determination of the number of degrees of freedom of an economic system is no easy task and cannot be accomplished by a straight forward calculation of dimension constants, no matter how defined, Casdagli 1992; Hunter 1992.

There are several reasons for this, some of which will be discussed in the sequel, but one of the most important is that appropriate models for economic and financial data involve internal resonating noise in addition to observational noise. In short, one of the difficulties is that economic models have to incorporate the nonlinear processing of shocks to the system that behave like random noise. Depending on the relative size of the noise variance, the effect of the presence of such terms can produce results that are difficult to distinguish from colored noise in terms of the calculation of a useful dimension measure.

The next major branch of the empirical literature involved the detection of "general nonlinearity" in the context of stochastic models. Some researchers distinguish nonlinearity in the mean from nonlinearity in the variances. While this is a common practice, the distinction is not useful when used in an exclusionary sense. As argued elsewhere, Ramsey 1994, it is likely that the analysis of macroeconomic variables must take into account the contemporaneous interaction of both means and variances. Further, the nonlinearity that we may observe will likely affect all the moments of the relevant distributions; the higher moments in particular. Thus, to restrict attention to nonlinearity in the means, or to nonlinearity in the variances, is probably counterproductive.

The results of the search for stochastic, as opposed to deterministic, nonlinearity are summarized in Brock, et al. 1991,Lorenz 1989, Ramsey, et al. 1990, Granger 1991, Ramsey & Rothman 1993, Lee, et al. 1993, Bollerslev, et al. 1990, Barnett, et al. 1993, Rothman 1994, and LeBaron 1994. All these authors testify to one general result. There is abundant evidence that economic and financial data provide many varied indications of widespread stochastic nonlinearity, even though the main effects seem to be exhibited in the variances of the respective distributions.

The debate that arises out of this literature is mainly whether there is evidence of nonlinearity net of generalized ARCH effects. Some of the work of Brock and LeBaron cited in Brock, et al. 1991, as well as a number of other researchers have all indicated that generalized ARCH models still leave some evidence of nonlinearities in the data. However, what that nonlinearity is and how to model it is still a very open question. A recent example of current attempts is provided by the "ceiling-floor" model of U.S. GDP growth rates by Pesaran & Potter 1993.

The work of Ramsey & Rothman 1993 on the existence of time irreversibility in macro economic variables provides a litmus test for the relevance of any proposed models of macro behavior; that is, for a variety of macro variables relevant models must be time irreversible. In short, the business cycle is "asymmetric" in that upturns do not have the same, but inverted, shape as downturns. The evidence for asymmetry is extensive in macro economic variables. The advantage of the time irreversibility concept is that a variety of definitions of asymmetry can be subsumed under one general concept and that time irreversibility is directly related to the dynamics of the system.

The forecasting literature with respect to nonlinear models can be illustrated by only a few papers that represent the plethora of ARCH type models and some of the work that is now being carried out to compare various types of nonlinear dynamic models. With respect to ARCH models of all types, Bollerslev, et al. 1990 have provided a comprehensive review of the use of ARCH models in finance and foreign exchange. The overwhelming conclusion is that ARCH models of some persuasion provide very useful representations of the second moments of the data, but that the evidence for out of sample forecasts is not very strong. Adding to these statements, Day & Lewis 1988 who examined the relative information content of implied volatilities from option prices and historical volatilities obtained by using both GARCH and EGARCH processes, concluded that "weekly volatility is difficult to predict." A more informative comment was made by McCurdy & Stengos 1988. These authors compared parametric estimates of a model of time varying risk premia with kernel estimators of the conditional mean function. The nonparametric approach produced better within sample fits, but the parametric model produced only marginally better out of sample (one period) forecasts. The authors conclude: "the superior in-sample performance of the [nonparametric kernel estimator] may be attributed to overfitting." I shall return to this idea of overfitting and forecasting performance latter.

Mizrach 1990 used nearest neighbor procedures to attempt to forecast EMS exchange rates. The result is essentially that such procedures do not provide reliably improved forecasts. Alogoskoufis & Stengos 1991 concluded that the U.S. unemployment rate was easily described by an AR(2) process with ARCH errors. There was some evidence of nonlinearity in the U.K. unemployment data, but that the nonlinearity is unlikely to be "chaotic" in nature. The potential for substantial forecasting gains appear to be slight. Prescott & Stengos 1991 investigated the time path of gold prices and concluded that there was no forecastable structure in the conditional means of the return series, but that the evidence of nonlinear dependence seemed to be restricted to the higher moments.

Rothman 1994 compared a variety of nonlinear models to a simple linear model. One should recognize that none of the postulated models arise out of any theory, or even from a close examination of the dynamical properties of the data; this is a point that I will return to later in the paper. The series involved are the U. S. unemployment rates that are known to be asymmetric, so that some form of nonlinear model is required. The forecast gain in forecasting conditional means by the nonlinear models depends on prior transformations to stationarity, but within that restriction there are statistically significant gains relative to linear models.

The overall conclusion is that there is so far very little evidence one way or the other for the forecast benefits of nonlinear models over linear ones. This is true even when one extends the variables of interest from means to variances. What evidence there is, is not overwhelming to say the least, although the Rothman results are among the strongest.

The remainder of this paper will concentrate on the difficulties that successful forecasting must face and query the preeminence of the forecasting criterion. The paper ends with a summary of the current status of nonlinear models and their role in forecasting.

SOME RATIONALIZATIONS OF FORECASTING DIFFICULTIES

An aphorism that can readily be applied to economic forecasts is that "good fits, do not necessarily lead to good forecasts." We can be more adventurous and claim that often the situation is that "better fits, lead to worse forecasts." While I will not document this "claim," merely illustrate it, its purported truth is recognizable by anyone who has had a few years experience in examining economic forecasts. An excellent early empirical evaluation of this claim is Makridakis 1986, who documented the lack of correlation between fit and forecast accuracy even for a single period forecast horizon.

There is evidence, for example, Poole 1988, Stock & Watson 1988, that the standard estimates of forecast error as well as the estimates of the variance of coefficient estimates significantly understate the observed variances of the forecasts and the variances of the coefficient estimates in the post fitting stage. For example, the Stock & Watson 1988 graph that is reproduced in this paper as Figure 1 illustrates one aspect of the forecasting problem in a striking manner. In this figure we see that two long term forecasts from the history of consumption and GNP were well outside the actual subsequent paths and are themselves well separated. The Poole 1988 analysis focuses on the unanticipated sudden cessation in the steady rate of growth in velocity of 3.1 percent per year that held from the early 'fifties to the early 'eighties. Both examples vividly illustrate some of the difficulties in forecasting that will be discussed. While I do not want to over emphasize these examples, they do provide a striking illustration of the central idea. The lesson that they illustrate could as easily have been illustrated at many other points in the history of these series. Indeed, the lesson could have been made with almost any other series. The conventional measures of forecast accuracy do not seem to be relevant in actual forecast situations; worse they appear to have a strong downward bias in that actual variances are far greater than those estimated. One may usefully speculate that the lack of forecasting performance stems from problems more severe than merely the underestimation of variances; it is likely that we really do not have a reasonable grasp of the underlying data generating mechanism.

Before proceeding with the main argument, what is meant by "linear" and "nonlinear" should be clarified.

Linear or Nonlinear is in the Equation

Let me begin by clarifying the use of the term "nonlinear." Strictly speaking the distinction between the terms "nonlinear" and "linear" must refer to the equations, or model, that is used to describe, or characterize the phenomenon under investigation. The distinction has no meaning when applied to the observation of the phenomenon itself without a specific representation in mind. Further, what is linear or nonlinear depends on the approach that is taken to provide a mathematical description of the phenomenon. Generally, any linear partial differential equation is mathematically equivalent to the equations of its characteristics which are usually nonlinear; Van Kampen 1981. For example, Newton's equations of motion for the planets are nonlinear, but the Schrödinger equation of the solar system is linear. The Euler equations from economic optimization may be linear, but the corresponding "equations of motion" of the variables themselves may well be nonlinear.

A further consideration is that what is nonlinear or not is often a function of the choice of co-ordinate system with respect to which the specific formulation of the model is tied. A suitable change in co-ordinate system can often redefine the relevant equations so that with respect to the new co-ordinate system the equations are either linear, or easily linearized. For example, dynamical equations expressed in terms of the parameter time, "t," may be highly nonlinear, but there may exist a differential function h(t) such that by redefining the dynamical equations in terms of the "new time" variable $T = h(t)$, the resulting equations are linear.

For example, the arrival of information in the international foreign exchange market in terms of calendar time is very uneven depending on which regional markets are open at any point in time. Some consideration has been given to this problem by Dacorogna, et al. 1993 and an alternative solution was provided by Ramsey & Zhang 1995. In both cases, the underlying concept was that complex variations in the time paths of price and quantities traded could be simplified by using a different definition of "time" than simple calendar time. In Ramsey & Zhang 1995, time was redefined in terms of the rate of arrival of new postings of bid and ask prices, so that the implied sampling rate was proportional to the rate of arrival of information. Previous schemes using nonlinear truncations of calendar time to define the sampling intervals had implicitly over sampled in periods of low information flow and implicitly under sampled during periods of high information flow. There is some evidence that the use of such "artificial units of time" simplified the analysis of the time variation of returns.

While it is plausible that the time variation of economic variables in terms of the usual Euclidean co-ordinates is best modeled by nonlinear equations, it is equally plausible that given a suitable choice of co-ordinates, the differential equations defining the dynamics of the system may well be linear, or sufficiently near linear that perturbation methods are feasible. A very simple example that is familiar to economists is the transformation from measurements on levels of economic variables to measurements on relative rates of growth defined in terms of first differences of the logarithms of the levels. Often, such a transformation linearizes the problem and simplifies the statistical analysis.

The question of a suitable choice of "coordinates" is a key issue in physics, even in the context of Hamiltonian systems. The idea is to convert a Hamiltonian system defined with respect to a Euclidean system of units that may be constrained to lie on a nonlinear manifold into a simpler Hamiltonian system that is defined without nonlinear constraints. Sometimes, such transformations can be achieved by the simple expedient of transforming from Euclidean coordinates to polar, or spherical coordinates. However, the situation is usually not so simple. Often, one must transform the data by more elaborate procedures and in the process, one often defines new variables of theoretical interest.

Thus, one objective of research in economics should be to investigate the extent to which reformulations of economic relationships in terms of differential equations with an appropriate choice of co-ordinate system will produce near linear equations for estimation. In short, instead of trying to impose linearity on the complex paths of the currently observed data, we should seek to reformulate the system in ways that will facilitate deriving linear equations in terms of generalized coordinates.

When one considers stochastic formulations of economic models, the situation is even more common. For example, as illustrated in Granger 1991 if $E(y_t | x_t) = g(x_t)$, where g(.) is linear, but with a non-zero constant term, then the conditional mean of y_t^2 is nonlinear. Further, if the variance of the additive error term depends on the square of the conditioning variable, that is, the error term is heteroskedastic and the heteroskedasticity is a function of x, then in fact the relationship between y_t and x_t is not linear. Thus, even in these very simple cases, whether or not an equation is linear or nonlinear depends on the formulation of the problem. One obvious way to deal with nonlinearity as an initial step is to reformulate the problem to enhance linearity, even if by so doing the variables involved are in the equations to be estimated are difficult to interpret economically. However, as an ancillary benefit, the reformulated problem may well lead to new theoretical insights into the operation of the economic system.

Open Non-isolated Systems Cannot Be Forecasted

There are two classes of problems involved in forecasting. The first class is inherent to the nature of economic data; the second concerns the way in which we approach the modeling of data. The first class is that economic systems are typically open and non-isolated. The second is that regression fitting is inherently local, whereas successful forecasting relies on the global characteristics of the system.

An "open" system is one for which there is exchange of "material" across system boundaries. For a "closed" system, the time path of variation in the values of all variables are determined within the system; there is neither input, nor output. A "Robinson Crusoe" economy before the arrival of "Man Friday" is a typical, but stark, example of a closed system,. Trade is an obvious example of a mechanism that links open systems. The wheat market is open to the extent that the market for wheat depends on variables in the markets for labor, energy, and capital inputs and that its output is the input to other productive processes and consumption. "Material" crosses system boundaries, or the system is characterized by flows of inputs and outputs, although the line of causality for the equations of motion need not run from "inputs" to "outputs". In economic markets in particular, the direction of causality is often from output to inputs in that the market responds to variations in demand for its products. As an example, the Bond and stock markets are closed except for the input or withdrawal of "cash" into the system and the creation, or destruction, of instruments.

The other relevant concept is "isolation." Markets, or economies, are not isolated, if they react to forces, or "energy," from outside the system. An isolated system is one which does not interact with other systems; it is one for which the equations of motion within the system remain invariant to events outside the system. Non-isolation means that the relationships within a system are altered by external pressure. For example, the operation of the bond market interacts with the operation of the stock and money markets. As another example, legislation on market operation ensures that the market is not isolated. Non-isolation means that the "equations of motion" within a system are influenced by activity outside the system; the equations themselves are changed. In contrast, an open system merely implies that while the state of the system depends on the rates of flow of "inputs" and "outputs", the equations of motion within the system remain invariant to action outside the system.

Systems can be open, but isolated; or closed, but not isolated. A stock market without change in the volume of stocks and without change in the available cash into the system is closed, but is not isolated from interest rate changes. A "Robinson Crusoe" economy that begins to launch little boats with messages is no longer closed, but is still isolated. A "Robinson Crusoe" economy that is showered by acid rain is still closed, but is no longer isolated. The key issue in this example is that the acid rain even though it is not an "input" into productive processes, has the effect of altering the productive relationships.

Open systems can be forecasted, but only under strict assumptions about the flows of inputs and outputs. If the causal relationship is from inputs to outputs and the flow of inputs can be forecasted in their own right and if there is no induced feedback effect from the flow of outputs, then the system can be forecasted. Alternatively, if the causal flow is from outputs to inputs, that is, the market responds to fluctuations in demand, inputs are perfectly elastic, and demand can be forecasted in its own right, then the system can be forecasted. To forecast such a system, open, but isolated, requires either knowledge about the time path of inputs or outputs, depending on the direction of causality, or the necessity to extend the system to include the markets, or systems, for the inputs and outputs. Such an approach is rapidly self-defeating in that one soon has too large a problem to solve with far too many variables to evaluate.

Historically, the solution to this problem has been entirely pragmatic. Sometimes the emphasis is on the output side so that the research emphasis is on the market for the output, sometimes the emphasis is on the input side so that the market for labor, or capital, inputs are examined most closely. However, as is well understood, the forecasting results from such an exercise depend on forecasting, without a model, the behavior of the variables that are introduced by linking the market system of interest to those systems that affect the inputs and the outputs. Pragmatically, the usual practice is to ignore any further interactions.

One possible empirical approach to limit the explosion of systems to evaluate in order to obtain forecasts, is to estimate the sensitivity of the estimates of prime interest to the variation of the variables introduced by openness of the original system. One extends the analysis to include all interactive systems that involve variables whose impact on the variables of interest exceeds some researcher assigned tolerance.

Non-isolated systems cannot be forecasted. The main difficulty is that if one is dealing with a non-isolated system, then by "definition" as it were, the system under investigation is being altered by events that are neither recorded, nor measured by the researcher. Clearly, if over time the relationships themselves are changing, then even very good fits to an historically observed series typically will not produce very good forecasts beyond one or two periods; the estimated variances of both forecasts and of coefficient estimates will in general be smaller than the observed variances from actual realizations of the future data. Paradoxically, it is often the case that the better the fit for open non-isolated systems the worse the forecast. An intuitive explanation of this paradox is easily presented.

Suppose that the model of a closed and isolated version of a system is correct, but that the observed system is non-isolated. One of the effects of non-isolation is to produce model coefficient estimates that vary through time. Estimating the equations in a non-isolated system implies that the coefficient estimates, which are adjusted to minimize the squared errors, reflect a trade-off between minimizing the actual errors and those induced by the time varying coefficients. While this procedure can yield good fits, it is unlikely to produce reasonable forecasts after the first few periods. This problem will be more thoroughly explored in the next sub-section.

There is an exceptional case in which the lack of isolation can be benign in its effects. Suppose that the relationships in the system are all linear and the effect of non-isolation is to produce a linear drift in the coefficients that is highly correlated with one or more of the variables of the system. Under these circumstances, while the estimates of the system's relationships are biased, the impact on forecasting is not detectable. This is because the effect of the time varying coefficients has been transferred by the fitting process to the coefficient values imputed by the model to some of the included regressors. If we now consider forecasting with this model, we see that its reliability and the accuracy of the forecast statements depend on the extent to which the fortuitous correlation between the coefficient drift and the time path of some of the included variables continues to hold.

However, there are three potential ways to deal with the issue of non-isolation; two of which are common to economists. First, one can try to pick markets and sub-periods of time so that one can be reasonably certain that within the restricted system, there is reason to believe that the system is approximately isolated. However, one can never be sure that such is the case even during the estimation period and certainly one cannot predict the dynamics of the events that make the analyzed system non-isolated.

The second approach is to expand the system to include those aspects that account for most of the elements that make the system non-isolated. Presumably, this was the motivation, along with the added problem of openness, underlying the creation of "project link" in which national macro models were linked through their trade and financial market interactions. Unfortunately, there are so many theoretical interdependencies that this is not usually a very practical solution.

A third approach to deal with non-isolation is to rely on the possibility of the separation of time scales to separate out the effects of non-isolation. At very long time scales, many effects can be ignored in that they are changing sufficiently slowly relative to the rate of change of the variables of interest. For example, if external events affecting the system can be determined to be varying over decades, whereas the internal variables are varying over several months, then short run solutions to the system dynamics can be achieved by treating the external variables as constants. However, care must be taken for even very slow changes will eventually have a measurable effect; in short, from this we anticipate the idea of "slowly varying coefficients." At the other extreme, non-isolation may be evidenced by rapidly varying effects that can be averaged out; this is, of course, the classic econometric assumption in modelling.

Similarly, it may be determined that non-isolation of the system is such that over normal time scales for the system, we can ignore the effects until they are evidenced and then allow for the change that has been introduced. More precisely, it may be the case that we can model the effects of, say legislation or a major new technology, as a Poisson process with a very long mean duration time between occurrences. At each occurrence, the system has to be re-initialized.

Openness and non-isolation are facts of life that must be faced. To the extent that one can mitigate the effects, reasonable forecasts can be made.To the extent to which one cannot mitigate the effects, one must recognize the true limits to forecasting.

Local Fitting Leads to Global Misfits

Regression fitting, even with nonlinear models, is essentially a local approximation centered at the vector of means of the constituent variables. Even if the approximation is very good and the underlying system is both closed and isolated, the standard results still may not apply and estimates of forecast accuracy may not match expectations.

This is because forecasting inherently relies on global properties of the system. For example, many dynamic systems are characterized by local instability, coupled with global stability. In these circumstances, local regression fits will inevitably lead to inaccurate forecasts. While one period ahead forecasts rely least on global properties, the lesson is still that forecasting requires more constraints on a system than local fitting. Forecasting is much more influenced by openness and non-isolation than is local fitting. One can always find some approximation to an historical sequence of data, but forecasting essentially involves extrapolation. Successful extrapolation requires global constraints to be met.

Thus, nonlinear models lead to a greater sensitivity of forecast errors to the underlying assumptions of the system. Subtle differences over the fitting region lead to rapidly decreasing usefulness of forecasts, especially when the system's models are nonlinear. The variation in forecasts and fitted values of parameters may be greater than is indicated by local approximating fits. These are complications that are heightened by nonlinearities in the equations. Small deviations within the fitting period, that are not easily detected given the inevitable presence of noise, may quickly grow to very large values in the forecast regime. If fitting is over a local instability region in a globally stable system, then even very good fits will lead to very inaccurate forecasts. For example, consider the well known "tent" function as a nonlinear mapping that is piecewise continuous and linear. Local fitting can lead either to explosive growth, or to implosive collapse, whereas the model is globally stable, even if it has a chaotic path. In either event, the forecasts will in general be wildly at variance with actual outcomes.

Recently, Yao and Tong investigated in the context of nonlinear models the effects of a lack of accuracy in observing, or estimating, initial conditions on forecast accuracy. Their introductory paragraph presents the problem very well, Yao & Tong 1994:

"The predominance until quite recently of the assumption of linearity in time series analysis has perpetuated the misconception that the reliability of the prediction is independent of the state. Indeed many standard textbooks in time series analysis have given 'error bounds' for the point forecasts which are *uniform* over the state space. Although uniformity may be true for linear least squares prediction, it is certainly untrue for non-linear prediction."

Here again, we see the key distinction between local fits and global forecasts. In this case, the emphasis is on the state of the state space in the "relatively distant past" and the effect that it has on the accuracy of ones forecasts.

Another source of forecast instability is the choice of approximation to the equations of motion used in model estimation. In general, the use of polynomials is not to be encouraged. For example, Ozaki 1985 in examining the stochastic dynamics of ship rolling in waves indicates that while a multivariate quadratic approximation to the nonlinear model provides good fits to the observed data the attempts to extrapolate the results produced very bad "forecasts"; this result occurred mainly because while the ship rolling dynamics were stable, the quadratic approximation was not. Another example of an inefficient approximating mechanism for producing forecasts is the use of Taylor's series expansions whenever extrapolations away from the point about which the expansion was calculated are needed. If forecasting is to be successful, even with closed and isolated systems, the approximating equations must have the same global properties as the equations themselves.

Yet another problem occurs when the system is not well understood and the researcher is searching for a model of the data, then the use of local fitting is even more prone to reflect the ephemera of local random fluctuations; one tends to "overfit" the local time idiosyncratic random fluctuations of the data at the expense of the underlying structure. The shorter the time period over which such fitting is done the greater the fluctuation in the estimates of the model's parameters and even in the choice of model itself. In part, this is because with small data sets the relative importance of "errors" to variations in the model's variables is greater; that is, with only a few observations it is more difficult to distinguish model fluctuations from errors, especially a few errors that by chance contain some form of regularity.

The temptation to overfit is greater than is usually given recognition because researchers tend to underestimate the extent to which patterns in data can be generated by purely random fluctuations. While the probability of a *specific* pattern being observed at random is very low, and decreases approximately geometrically in the length of the pattern, the probability of observing some pattern at some length size is very high. For example, imagine that the data consist of a string of binary values of length 10; that is, the data represent ten drawings from [0,1] with equal probability.If we count the number of outcomes that would unambiguously be regarded as "without pattern" in a string of ten drawings we get a number on the order of 256 and perhaps considerably less, but there are a total of 2^{10}, or 1024, possible drawings so that the probability of observing *some pattern* is nearly three quarters. The lesson is clear, the probability of observing *some pattern* is very high.

Granger & Terasvirta 1992 after comparing various parametric and non-parametric models using simulated data came to the conclusion that nonparametric models tended to provide better fits, but worse forecasts than the parametric models. Even given that the experimental design included the condition that one of the parametric models was used to generate the data, these results substantiate the discussion above. That is, non-parametric formulations which are usually "representational," lead to local over fitting in the regression and therefore to worse forecasts when the "over fitted" results are extrapolated. Further, with such models there is seldom any attention paid to the properties of the approximating model and whether those properties are reasonable.

The problems alluded to above suggest their own resolution. In the model searching phase at least, attempts should be made to impose global constraints on the model, the approximating equations, and their estimation, thereby offsetting the tendency to fit approximations with very different global characteristics and to avoid the tendency to overfit local random patterns. Consider, one example from Ramsey 1992 in which was reported the effort to find a useful dynamical model for indices of consumer goods production. The basic concept was that a forced oscillator of some type would provide a potentially meaningful model of such indices, especially when coupled with the idea that the parameters of the model are changing adiabatically over time in response to long time scale changes in the structure of the economy. The concept was that by allowing for adiabatic change, one might be able to find a single class of models that would provide a succinct description of the entire data series. If the goal were to be achieved, then events such as the Depression, World War II, the Korean war, and so on, could be assessed in terms of the changes in the coefficients over time.

The model finally chosen was based on evaluating both the local regression fits and respecting the global constraints. The class of approximating models was restricted to those that could match the global characteristics of the data. Some of the global criteria applied included the stability of the parameter estimates over the whole data series when estimated in epochs, the stability of the variance estimates over epochs, the consistency of the estimates with the known dynamical properties of the data, whether there was any evidence of systematic variation in the residuals at the same time scale as the model's dynamics, and symmetry in the model structure across sub-periods; that is, the model was not altered over time merely to meet local sub-period "fits." In short, local evidences of fit were exchanged for evidence of global consistency. This approach marks a departure from current conventional analysis in that a single class of models with appropriate global characteristics was sought that would provide a unified description of all the data, without relying on the elimination of "difficult" periods of time, nor on building "special cases" for the "difficult" periods.

Sampling Rate and Aggregation

Another problem that is introduced, or exacerbated, by the presence of nonlinear models is the choice of sampling rate. As is well known, there is a trade-off between the stochastic continuity of stochastic processes and the relevance of the Markovian assumption. Stochastic continuity of the stochastic process requires a high sampling rate. Whereas for the process to be Markovian, a much lower sampling rate is required so that the processes of adjustment that are path dependent can be ignored. Typically, empirical research relies on both assumptions holding. However, if data are sampled at too wide an interval, then not only are high frequency components lost, as is well known, but also much of the dynamic structure can be missed and attempts to simulate the data using the estimated coefficients can produce simulated paths that are wildly at variance with the sampled data. If the relevant model is linear, or a linear sum of sinusoids, then sampling at a rate that is a multiple of the periods of the higher frequencies merely loses information on the high frequency components. But with nonlinear models, the low sampling rate may also miss evidence that presages large amplitude oscillations that will eventually be recorded by the low frequency sampling rate. When noise is added to the system, the dangers of a low sampling rate are enhanced. However, intermittently the magnitude of the nonlinear oscillations may grow to a size that is large relative to the noise floor, in which case one has the appearance of a sudden shock to the observed system , or one might be forgiven for presuming that the structure of the model had changed, for interesting examples of these types of situations see Ozaki 1981; Ozaki 1985, and Casdagli 1989.

By now it is almost a truism to state that nonlinear models are not easily aggregated in that it is difficult to determine a smooth function of some aggregate from even certain knowledge of the nonlinear functions at the micro level. This problem really concerns the existence of macro variables and is discussed in full in another paper, Ramsey 1994. None the less, it is clear that even C^∞ functions contain few sub-classes of functions that are closed under summation. Worse is that it may be even more difficult to obtain a useful summary function whose derivatives approximate closely the sum of the derivatives of the constituent functions.

An alternative to stringent constraints on the form of the functions to be aggregated is to restrict attention to distributions of variables that are in the exponential family. Even so, one's choice of functions that are closed under summation, that is, can be aggregated, is still limited. Of course, there is still no guarantee that the functions that can be aggregated are relevant to the requirements of economic theory, Stadler 1994.

Further, it is well known that the aggregation of only two nonlinear systems can preclude the discovery of the structure of either, see for example, Sugihara, et al. 1990. A very simple example, but one that is a common occurrence, is the aggregation of two harmonic series which differ in phase which is itself varying over time, as compared to aggregating two harmonic series with a fixed phase difference. The former aggregate series is difficult to analyze and the latter aggregate is simply handled by Fourier analysis.

MINIMUM MEAN SQUARE ERROR FORECASTS SHOULD NOT BE THE SOLE CRITERION

In the debate between those pursuing the conventional approaches to economic analysis and those pursuing a more nonlinear approach, the idea seems to have arisen that forecasting accuracy, or more narrowly, minimum mean square error of forecast, is the sole criterion for deciding between alternative models and alternative methodologies. Recently, other measures for comparing forecast accuracy have been discussed and evaluated empirically, see for example, Makridakis 1993. However, these considerations, while important in their own right, are not germane to the issue at hand.

If a nonlinear model cannot provide more accurate, or less variable forecasts, the conventional wisdom claims there is no need to be bothered by the nonlinear model. I shall argue in the paragraphs to follow that the forecasting benchmark, while useful, should not be the sole criterion for choosing between models.

Models and Representations

At this stage in the discussion, it is useful to distinguish a "model" from a "representation." A model, as is well known, is an attempt to link in logical fashion seemingly related phenomena into an intellectually coherent framework. A model is based on an underlying theory that provides a common framework of analysis for a range of related variables. Models provide a **causal structure** with, or without, a feedback mechanism. Models also provide restrictions on the observable relationships. The specification of a causal mechanism and restrictions on potential relationships is the distinguishing characteristic of a model.

Some models are phenomenologically based, that is "data driven," in the sense that an attempt is made to derive a model from observed data, and to incorporate whatever prior theoretical knowledge that one has. The idea behind a phenomenologically driven model is that if the essential elements of the variation have been captured, the phenomenologically discovered structure will indicate the causal mechanism and will thereby provide links to similar phenomena. If successful, a phenomenological model can prove to be relevant to the analysis of different data sets that are generated by similar experiments. In short, phenomenologically driven models can be useful devices to discover structural relationships in the data, provided one then checks out the model using different experiments. In either case, a model attempts to provide, or at least to reveal in the case of a phenomenologically driven model, a causal mechanism and to specify restrictions on the relationships between classes of variables.

A "representation," however, is a way of characterizing observations for a well defined, albeit broad, class of sequences and is **not** designed to provide a causal mechanism. Representations are "atheoretical". The Wold decomposition theorem, Priestley 1981, or the spectral representation theorem following Wiener-Khintchine, Priestley 1981 are the two most famous examples. Other examples are representations in terms of Volterra expansions, Mittnik Stefan 1991, wavelets, Ramsey & Zhang 1994; Ramsey & Zhang 1995. Yet another representation arises out of neural network theory using logistic functions, Hornik, et al. 1989, or finally a most useful representation for data in two dimensional arrays is that of the singular value decomposition. While representation theorems are useful, indeed very useful; it is important to recognize the limitations of a representation result.

Consider the singular value decomposition of any matrix of observations. For example, let X be a matrix of T observations on L interest rates. The matrix X can be represented as follows:

$$X = UDV \tag{1}$$

where D is a diagonal matrix with the singular values, or eigenvalue weights, on the diagonal and U, V are orthonormal bases for the column and row spaces of X respectively. While much can be learned about the matrix X and useful approximations of X can be obtained from an analysis of the matrices U, V, and D, the solution to equation (1) sheds no light on the mechanism that is supposed to be generating the observed matrix, X. Representations are purely descriptive. In this example, it is important to remember that **any** T by L matrix can be represented by this procedure.

A more statistical example is this. Let y_t designate any stationary time series. Consider the representation by the "general linear model," shown in equation (2).

$$y_t = \sum_{u=0}^{\infty} g_u \epsilon_{t-u} \tag{2}$$

where g_u is l_2 and ϵ_t is a white noise process. Even when the parameters are consistently estimated and the parameters do not vary over time, equation (2) provides only a representation of y_t, but no understanding of the mechanism that might generate the series. To state that *any* stationary process can be factored into a linear causal sequence of orthogonal increments plus a component that is deterministic and perfectly predictable is not directly helpful in understanding the process that is presumably generating the observations. Because we know that for some finite sequence of coefficients we can approximate the arbitrary stochastic sequence as close as we like, the mere fact of "fitting " is itself uninformative.

Representations, within the class for which each is defined; that is, the class of stationary processes for the Wold decomposition, stationary stochastically continuous processes for the Wiener-Khintchine theorem, and so on; hold universally. The only relevant questions with respect to a representation are whether a representation with a small number of terms is useful and if so, what are the coefficient values. Over long periods of time, one can use representations to check whether the values of the coefficients of the approximating representation are constant, or change over time.

All models themselves have representations. This is a truism that is a source of confusion about the respective roles of models and representations. For example, nonlinear models that pass statistical stationarity tests have Wold representations. Most models with noise that are Markovian in nature will have ARIMA representations. As another example, a simple sinusoid has a representation in terms of Fourier coefficients, or if the process has been sampled at a rate of k terms per cycle, in terms of "k" constants, α_j, j = 1,2....k.

In Brock, et al. 1991, it is shown that a GARCH (p,q) process can be regarded as a finite restriction of a Wold type representation of second order moments. While a precise definition of the class of sequences for which such an expansion is a representation has not been stated rigorously, the argument is indicative that with respect to some reasonably broad class of sequences, the ARCH(p,q) formulation is a representation of the time dependencies of second moments. The defining characteristic of a representation is that within the class all sequences can be so represented. ARCH processes have enjoyed a well deserved popularity; perhaps the realization that they are a first effort at providing a representation for second moments explains their success.

Let us now examine the forecasting implications of these remarks. I pointed out that any stationary process generated by a theoretical or a phenomenological model, whether linear, or nonlinear, is **always representable** in terms of a Wold representation; that is, in terms of a weighted sum of lagged unobservable orthogonal "shocks." The forecasting capabilities of both the model and its representation are the same except for the number of parameters that have been used. Consider a simpler example. Suppose that a process is generated by a mechanism that can be modeled as a sinusoid of a single frequency and that the series is observed at a sampling rate that produces "k" observations per cycle. Compare two "models" of this sequence of observations. One is the sinusoidal model, that might have been suggested by some ideas about the dynamics of the underlying physical process. The other is a "seasonal" dummy representation that uses the data to estimate "k" dummies. The seasonal dummy representation cannot be beaten in terms of mean forecasting error criteria by any model of the purported mechanism, even the simple sinusoidal model. Never the less, one would still be advised to consider carefully the theoretically or phenomenologically driven model, rather than the "representational" method. This is because the former is an attempt to relate the data to some notion of the dynamics involved and may stimulate further ideas about the mechanism itself, stimulate comparisons with other mechanisms, and in general provide a basis for theoretical speculation.

Some might complain that the sinusoidal model mentioned above involves fewer parameters and would therefore be chosen on the grounds of parsimony of parameter use. The above argument still holds, perhaps even more clearly, when the number of required terms in the sinusoidal expansion is greater than "k." The seasonal dummy model still cannot be beaten in a mean square error sense and is now supposed to have fewer parameters. Yet I claim that the sophisticated researcher will prefer the sinusoidal model notwithstanding its greater use of degrees of freedom. In this particular example, one could have one's cake and eat it too as it were, by using the sinusoidal model for analysis, understanding, interpretation, stimulating insights, and determining what will happen in response to changes to the system; while reserving the representation that happens in this case to involve fewer parameters for providing forecast numbers.

An even more insightful example is provided by wavelet representations of time series that may well be non-stationary. The types of structures that can be used to represent such data include in the context of a frequency-time decomposition harmonic frequencies, short bursts of a block of frequencies, known as "chirps", and finally Dirac delta functions; Ramsey & Zhang 1994; Ramsey & Zhang 1995. Waveform dictionaries therefore can provide a very wide range of structures to represent data that includes those found in Fourier analysis, in conventional wavelet analysis, and indeed everything in between. Commonly, it is discovered that a relatively few structures are needed to represent even very complex data series; such was the case for the Standard and Poor's 500 index and for foreign exchange rates as discussed in the Ramsey and Zhang papers. In short, the waveform dictionaries provided very good fits to the data using relatively few structures to achieve that result. However, this result is **not sufficient** to provide good forecasts. The reasons why are instructive in this context.

To strengthen the argument assume that what ever mechanism generated the data during the estimation period continues to apply during the forecast period. First, in so far as the energy in the representation is in harmonic frequencies that last during the entire historical period, then to this extent at least, one has the basis for a forecast; one merely predicts that past frequencies continue to apply in the future. Now consider the case where much of the energy in the system is in the occurrence of Dirac delta functions. Unless, one can discover some form of periodicity in their occurrence, then nothing can be forecasted from these structures. Next, in so far as energy is contained in the "chirps", short bursts of groups of frequencies, and the chirps do not have any pattern, then once again there is no basis for forecasting. We have in this example demonstrated a situation in which one can achieve good fits to historical data, but one cannot forecast very much, if at all, using that information.

A similar situation holds for neural network analysis in that one can achieve very good fits to the data and still not be able to improve forecasts very much at all. In the past, the source of this conundrum has been ascribed to "overfitting"; that is, to fitting a model to elements of noise at the expense of the structure of the model itself. The argument here is different in that the lack of forecasting gain is not due to "overfitting" in this sense, but to the fact that the model has little in the way of a forecastable component. This is true even though any historical segment can be represented by a relatively parsimonious set of model structures. The simplest way of describing the situation is that most of the structure observed in any historical segment is unlikely to be repeated, at least in its entirety, in any other segment. A compounding factor is whether the approximations used to fit the historical data can be extrapolated out of the region over which the fits were made. Often this is not the case.

This type of result should not be too surprising. If we had been able to reduce the analysis of the variation in market price to the outcomes of a binary choice, i.e. a "coin toss", we would have very good understanding of the mechanism underlying the generation of the data, but would be unable to improve our forecasts in the slightest. The broader lesson is that frequently in the statistical examination of economic and financial data, we may well be able to "describe" any historical data set very well and to do so with parsimony, but still not be able to improve our forecasts.

One conclusion is that models that are based on a theory, that provide restrictions on the relationships between variables, that purport to characterize the dynamics of a system, are to be preferred to mathematical or statistical **representations** of the data. This is true, I would claim even at some cost in forecast accuracy; although the compromise suggested in the previous paragraph is an option. If some degree of "universality" in the modeling of economic data is ever to be achieved, it will only come from theoretically, or phenomenologically, driven models. Representations have their uses, most especially in the beginning stages of an analysis, but the ultimate goal is to provide understanding of the underlying mechanisms and through that understanding useful forecasts in which some reliability can be placed because the "causal mechanism" is understood, at least in part.

SUMMARY AND CONCLUSIONS

While there is virtually no evidence in favor of chaos, or deterministic complex behavior, in economic or financial data, there is abundant evidence for nonlinear stochastic processes. Much of the nonlinearity is in the second moments that can be successfully represented by some form of ARCH process. There is some evidence that there is still nonlinearity present in the data even after allowing for ARCH effects. There is not a lot of analysis on the forecasting results for nonlinear models, except for the ARCH type of processes. In any event, the results that are available indicate at best there are only modest gains in forecasting accuracy and virtually no gain in reliability in the forecasts.

A theme of the paper is that many of the modelling difficulties that plague all analyses of real data, are not only more worrisome and difficult to deal with in the context of nonlinear models, but are qualitatively different. The first key point is that open and non-isolated system cannot be forecasted, except under the most stringent conditions. Consequently, the ability to forecast reliably depends on the extent to which the assumptions of closure and non-isolation are good approximations. The conventional rule of thumb that good fits do not give good forecasts is justified by recognizing that fitting is local, but that forecasting is an essentially global idea. Even good modelling approximations for closed and isolated systems can lead to very bad forecasts; this is especially true when the approximating functions do not have the same global properties as the mechanism generating the data. Polynomials and extrapolated Taylor's series are good examples of inappropriate approximations that have good local properties, but very bad global properties. Local overfitting also leads to bad forecasts, especially with nonlinear models. But the most difficult problem is that economic data often seem to be comprised of structural components that constantly change over time in at least seemingly unforeseeable ways.

Nonlinear properties of models can be missed in data analysis, because the sampling rate is too coarse, there is too much aggregation, or for some periods the nonlinear fluctuations are below the noise floor. Further more, forecasts are more sensitive to the error in evaluating initial conditions than is true for fits.

Models and representations of data have been distinguished. A case was made that forecasting should not be the sole criterion in choosing between a model and a representation of the data. Models purport to delimit the causality, whereas the representations provide a basis for projections for broad classes of observed sequences. Models are to be preferred to representations for understanding the mechanism and its limitations and in facilitating the links to other classes of phenomena. These advantages outweigh some loss in forecasting accuracy. One could easily use both procedures; the model for understanding and knowing the limits of applicability of the model; the representation for obtaining more accurate forecasts. However, when the analysis of the model indicates that the structure has changed, the representation will need to be re-estimated before being used as a forecast tool.

FIGURE 1

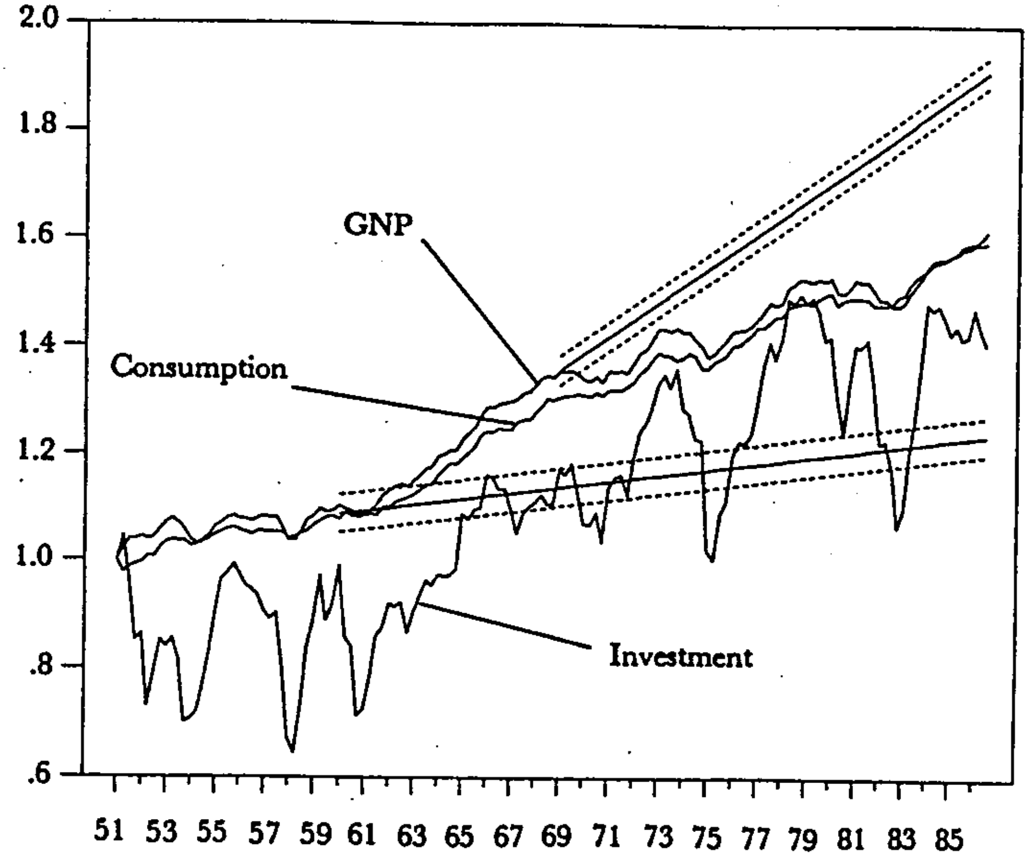

Fig. 1. Postwar real per capita U.S. GNP, total consumption, and gross private domestic investment (in logarithms)

All three series have been arbitrarily set to 1 in 1951:I. The straight solid lines represent two long-run forecasts of GNP, the lower using data from 1951:I–1959:IV and the upper using data from 1960:I–1969:IV. The forecasts were made by extrapolating GNP growth over these periods using a linear deterministic time trend. The dotted lines represent bands of ± two standard deviations of quarterly GNP growth around the long-run forecasts. Were GNP a stationary series about a linear time trend, these bands would provide an approximate long-run 95% confidence interval for the respective GNP forecasts.

Reprinted from Stock, J.W/ and Watson, M.V. "Variable Trends in Economic Times Series," Journal of Economic Perspectives, Summer, 1988, 2, No. 3, pg. 148.

References

Alogoskoufis, G. S. and T. Stengos (1991): *Testing for Nonlinear Dynamics in Historical Unemployment Series.*,European University Institute Working Paper, vol. ECO 91/38, Badia Fiesolana, Italy (25).

Barnett, W. A., A. R. Gallant and M. Hinich (1993): "Detection of Nonlinearity and Chaos: Application to Money Stock," in *Dept. of Economics Working Paper Series*, 178, Washington University-St. Louis, October.

Bollerslev, T., R. Chou, J. Jayaraman and K. F. Kroner (1990): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52(1), 5-60.

Brock, W. A., D. Hsieh and B. LeBaron (1991): *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, Cambridge, Ma.: M.I.T. Press.

Brock, W. A. and S. M. Potter (1993): "Nonlinear Time Series and Macroeconomics," in *Handbook of Statistics, Vol. 11*, ed. by G. Maddala, C. Rao and H. Vinod, New York: Elsevier Science, 195-229.

Casdagli, M. (1989): "Nonlinear Prediction of Chaotic Time Series," *Physica D*, 35, 335-356.

------ (1992): "A Dynamical Systems Approach to Modeling Input-Output Systems," in *Nonlinear Modeling and Forecasting*, ed. by M. Casdagli and S. Eubank.,Diagnostic Testing for Nonlinearity, Chaos, and General Dependence in Time Series Data, Redwood City, Ca.: Addison-Wesley, 265-282.

Dacorogna, M., U. Muller, R. Nagler, R. Olsen and O. Pictet (1993): "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance*, 12, 413-438.

Day, T. E. and C. M. Lewis (1988): "Stock Market Volatility and the Information Content of Stock Index Options," *Journal of Econometrics*, 52 (April/May), 267-287.

Granger, C. and T. Terasvirta (1992): "Experiments in Modeling Nonlinear Relationships Between Time Series," in *Nonlinear Modeling and Forecasting*, ed. by M. Casdagli and S. Eubank.,Diagnostic Testing for Nonlinearity, Chaos, and General Dependence in Time Series Data, Redwood City, Ca.: Addison-Wesley, 189-198.

Granger, C. W. (1991): "Developments in the Nonlinear Analysis of Economic Series," *Scannd. Journal of Economics*, 93(2), 263-276.

Hornik, K., M. Stinchcombe and H. White (1989): *Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks.*,UCSD DP 89-34, La Jolla: Economics Dept. UCSD.

Hunter, N. F., Jr. (1992): "Application of Nonlinear Time Series Models to Driven Systems," in *Nonlinear Modeling and Forecasting*, ed. by M. Casdagli and S. Eubank, Redwood City, Ca.: Addison-Wesley, 467-492.

Jaditz, T. and C. Sayers (1992): "Is Chaos Generic in Economic Data?" *International Journal of Bifurcation and Chaos*, forthcoming.

LeBaron, B. (1994): "Chaos and Forecastability in Economics and Finance," *Phil. Trans. Royal Society,* 348, 397-404.

Lee, T., H. White and C.W.J. Granger (1993): "Testing for Neglected Nonlinearity in Time Series Models.," *Journal of Econometrics*, 56, 269-290.

Lorenz, H. (1989): *Nonlinear Dynamical Economics and Chaotic Motion*, 334.,Lecture Notes in Economics and Mathematical Systems, no. 334, New York: Springer-Verlag.

------ (1993): "Complexity in Deterministic, Nonlinear Business Cycle Models-Foundations, Empirical Evidence, and Predictability," in *Nonlinear Dynamics in Economics and Social Sciences*, ed. by F. Gori, L. Geronazzo and M. Galeotti, Berlin: Springer - Verlag, 17-52.

Makridakis, S. (1986): "The Art and Science of Forecasting: An Assessment and Future Directions," *International Journal of Forecasting*, 2, 15-39.

------ (1993): "Accuracy Measures: Theoretical and Practical Concerns," *International Journal of Forecasting*, 9(4), December, 527-530.

McCurdy, T. H. and T. Stengos (1988): "A Comparison of Risk-premium Forecasts Implied by Parametric Versus Nonparametric Conditional Estimators," *Journal of Econometrics*, 52 (April/May), 225-244.

Mittnik Stefan (1991): "State Space Modeling of Nonlinear Time Series," Stony Brook, NY.: SUNY, Economics, April.

Mizrach, B. (1990): "Multivariate Nearest Neighbor Forecasts of EMS Exchange Rates," Philadelphia, Pa.: The Wharton School, Finance, July.

Ozaki, T. (1981): "NOn-linear Threshold Autoregressive Models for Non-linear Random Vibrations," *Journal of Applied Prob.*, 18, 443-451.

------ (1985): "Non-Linear Time Series Models and Dynamical Systems," in *Handbook of Statistics, Vol. 5*, ed. by E. Hannan, P. Krishnaiah and M. Rao, New York: Elsevier Science, 25-83.

Pesaran, M. H. and S. M. Potter (1993): "A Floor and Ceiling Model of U.S. Output," Los Angles, CA: UCLA, Economics, November.

Poole, W. (1988): "Monetary Lessons of Recent Inflation and Disinflation," *Journal of Economic Perspectives*, 2(3), Summer, 73-100.

Prescott, D. M. and T. Stengos (1991): *Testing for Forecastible Nonlinear Dependence in Weekly Gold Rates of Return.*,European University Institute Working Paper Series, vol. ECO 91/31, Badia Fiesolana, Italy (23).

Priestley, M. (1981): *Spectral Analysis and Time Series*, New York: Academic Press.

Ramsey, J. B. (1992): "Seasonal Economic Data as Approximate Harmonic Oscillators," 92-16.,C.V.Starr Center for Applied Economics Wkg. Papers, no. 92-16, New York, April.

------ (1994): "On the Existence of Macro Variables and of Macro Relationships," Dept. of Economics, New York Univ., October .

------ and P. Rothman (1993): "Time Irreversibility and Business Cycle Asymmetry," New York: New York Univ., Economics, December.

------ and Z. Zhang (1994): "The Application of Wave Form Dictionaries to Stock Market Data," New York: C.V. Starr Center for Applied Economics, New York Univ., February.

Ramsey, J. B., C. L. Sayers and P. Rothman (1990): "The Statistical Properties of Dimension Calculations Using Small Data Sets: Some Economic Applications," *International Economic Review*, 31(4), 991-1020.

Ramsey, J. B. and Z. Zhang (1995): *The Analysis of Foreign Exchange Data Using Waveform Dictionaries.*,C.V.Starr Center for Applied Economics, vol. RR#95-03, New York Univ., New York: Dept. of Economics (22).

Ramsey, J. B. and H. J. Yuan (1989): "Bias and Error Bars in Dimension Calculations and Their Evaluation in Some Simple Models," *Physics Letters-A*, 134 (January), 287-297.

Rothman, P. (1994): "Forecasting Asymmetric Unemployment Data.",Working Paper, Greenville, N.C.: East Carolina Univ.; Economics.

Stadler, G. W. (1994): "Real Business Cycles," *Journal of Economic Literature*, XXXII(4), December, 1750-1783.

Stock, J. K. and M. W. Watson (1988): "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, 2(3), Summer, 147-174.

Sugihara, G., B. T. Grenfell and R. M. May (1990): "Distinguishing Error from Chaos in Ecological Time Series," *Phil. Trans. R. Soc. Lond.*, 330, 235-251.,B.

Van Kampen, N. G. (1981): *Stochastic Processes in Physics and Chemistry*, Amsterdam: North-Holland.

Yao, Q. and H. Tong (1994): "Quantifying the Influence of Initial Values on Non-linear Prediction," *Journal of the Royal Statistical Society; Series B*, 56(4), 701-726.