STUDIES IN Nonlinear Dynamics and Econometrics

Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
July 1996, Volume 1, Number 2
The MIT Press

Studies in Nonlinear Dynamics and Econometrics (ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, 55 Hayward Street, Cambridge, MA 02142; (617)253-2889; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $130.00. Canadians add additional 7% GST. Prices subject to change without notice.

If Nonlinear Models Cannot Forecast, What Use Are They?

James B. Ramsey[*]

Department of Economics
New York University
New York, NY 10003
ramsey@fasecon.nyu.edu

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 nonisolated systems cannot be forecasted; the extent to which economic systems are closed and isolated provides the true pragmatic limits to forecastability. The reasons why local "overfitting," especially with nonparametric 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.*

1 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. There are many excellent papers devoted to presenting evidence of nonlinearity in economic and financial time series, for example, Ashley et al. (1986), Ashley and Patterson (1989), and Brock et al. (1991). But that investigation is not the subject of this paper. I take it as given that there is abundant evidence of nonlinearity of some sort in almost all economic and financial time series.

The major task of this paper is to evaluate the difficulties faced by inference and forecasting, in particular, if the world is nonlinearly dynamic. The question has been well posed by Diebold and Nason (1990), who state:

> "Why is it that while statistically significant rejections of linearity in exchange rates [and many other economic and financial series] routinely occur, no nonlinear model has been found that can significantly outperform even the simplest linear model in out-of-sample forecasting?"

The case is made in this paper that in the context of nonlinear dynamic 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

[*]The financial and institutional support of the C. V. Starr Center for Applied Economics is gratefully acknowledged. Comments from Rick Ashley were particularly valuable.

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 subcategory 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 their seemingly random time paths, are in fact representations of long-term steady states of the corresponding dynamical system.

A second major strand to 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 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.

2 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 provide 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), Bollerslev et al. (1990), Brock and Potter (1993), and Le Baron (1994).

The initial flurry of activity investigating the role of "chaos" in economics that began in 1986 centered on calculating Lyapunov exponents and dimension numbers. Lyapunov exponents determine the degree of local instability of the time paths of a dynamic system, and dimension indicates the degree of complexity of the system as determined by the minimum number of variables that is needed to model the system. An objective of the early analysis 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 was a problem, albeit not the most important one. There were 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 and Yuan (1989) and Ramsey et al. (1990). Some of the other reasons for the "nondetection" of chaos will be enunciated below when I discuss the concepts of openness and nonisolation 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 borderline 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 and Teräsvirta (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 and 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 and Potter 1993; Le Baron 1994; and Ashley and Patterson 1989).

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 it cannot be accomplished by a straightforward 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 must incorporate the nonlinear processing of shocks to the system that behave like random noise. Depending on the relative size of the noise variance, the presence of such terms on dimension calculations can produce results that are difficult to distinguish from colored noise; that is, error terms that are representable by a general linear model.

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 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 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), Ashley et al. (1986), Ramsey et al. (1990), Granger (1991), Ramsey and Rothman (1995), Lee et al. (1993), Bollerslev et al. (1990), Barnett et al. (1993), Rothman (1994), and Le Baron (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 main debate that arises from this literature is whether there is evidence of nonlinearity net of generalized autoregressive conditional heteroskedastic (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 it should be modeled is still an open question. A recent example of current attempts is provided by the "ceiling-floor" model of U.S. GDP growth rates by Pesaran and Potter (1993). An alternative approach that returns to the analysis of the time variation of the means is the examination of time series for evidence of time irreversibility (Ramsey and Rothman 1995).

The work of Ramsey and Rothman on the existence of time irreversibility in macroeconomic variables provides a litmus test for the relevance of any proposed models of macrobehavior; that is, for a variety of macrovariables, relevant models must be time irreversible. In short, the business cycle is "asymmetric" in that upturns do not have the same shape as downturns, but are inverted. The evidence for asymmetry is extensive in macroeconomic variables. The advantages of the time-irreversibility concept are 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 and Lewis (1988), who examined the relative information content of implied volatilities from option prices and historical volatilities obtained by using both generalized autoregressive conditional heteroskedastic (GARCH) and exponential generalized autoregressive conditional heteroskedastic (EGARCH) processes, concluded that "weekly volatility is difficult to predict." A more informative comment was made by McCurdy and Stengos (1991). 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 later.

Mizrach (1992) used nearest-neighbor procedures to attempt to forecast European Monetary System (EMS) exchange rates. The result is essentially that such procedures do not provide reliably improved forecasts. Algoskoufis and 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 the nonlinearity is unlikely to be "chaotic" in nature. The potential for substantial forecasting gains appears to be slight. Prescott and 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.

However, a number of other studies have found little forecasting benefit using nonlinear models. Swanson and White (1995) used an adaptive neural network approach for a variety of macrovariables, and contrasted the results to those of the Survey of Professional Forecasters. A major conclusion was that while adaptivity was very important, that is, re-estimating the neural network for each forecast horizon, nonlinearity seemed to play a relatively minor role. Two other recent articles obtained disappointing results for the benefits of nonlinear models in the context of foreign exchange markets. Diebold and Nason (1990) found a dramatic difference in the relative performance of their nonparametric predictors for the conditional means in sample and out of sample; the out-of-sample forecast record for the nonparametric predictors was worse than for the random walk models. Subsequent work by Meese and Rose (1991) examined the empirical relationship between nominal exchange rates and macroeconomic fundamentals for five major Organization for European Community Development (OECD) countries, and found that neither time deformation (Stock 1987) nor incorporating nonlinearities into structural models improved the forecasts of the levels of exchange rates.

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.

3 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 even 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 and 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 and 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. In short, the calculated standard errors of forecast were essentially irrelevant to determining the probability bounds for the subsequent path of the forecasted variables.

The Poole analysis (1988) focused on the unanticipated sudden cessation in the steady rate of growth in velocity of 3.1 percent per year that held from the early 1950s to the early 1980s. This surprise occurred despite the widespread econometric attention paid to the velocity measure and the theoretical importance that was attached to it.

While I do not want to over-emphasize these two 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. While the observed degree of variation over any given interval did not change very much over the subsequent period, the actual path of the variable was well outside the predicted bounds; the calculated error bounds were at best misleading. One may usefully speculate that the lack of forecasting performance stems from problems

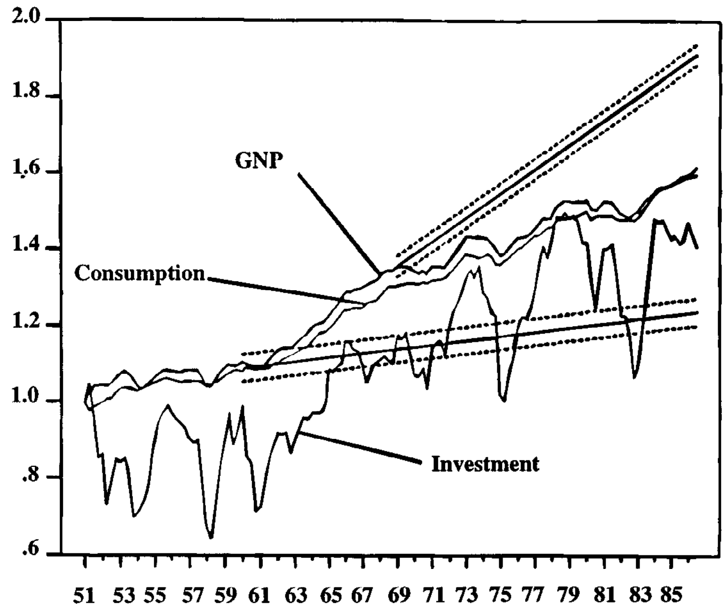

Figure 1
Post-war Real per Capita U.S. GNP, Total Consumption, and Gross Private Domestic Investment (in logarithms).

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 J. W. Stock and M. V. Watson, "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, Summer 1988, 2(3):148.)

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.

3.1 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 or is not nonlinear is often a function of the choice of coordinate system, with respect to which the specific formulation of the model is tied. A suitable change in coordinate system can often redefine the relevant equations so that with respect to the new coordinate 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.

To illustrate, 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 and 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 and 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 information arrival. 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 coordinates is best modeled by nonlinear equations, it is equally plausible that given a suitable choice of coordinates, 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 coordinate 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 potential for nonlinearity is even more common. As illustrated in Granger (1991), if $E(y_t|x_t) = g(x_t)$, where $g(\cdot)$ is linear, but with a nonzero 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_t^2, then in fact the relationship between y_t and x_t is nonlinear. 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 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.

3.2 Open Nonisolated 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 nonisolated. 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 is 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. In equilibrium, there is a material balance in the flows so that flows are steady state, there is no change in the efficiency losses, and there is no accumulation, or decumulation, in inventories. 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 line of causality for the equations of motion need not run from "inputs" to "outputs." In economic markets the direction of causality can be from outputs to inputs, in that the supply side of the market responds to variations in the demand for its products. The direction of causality depends on the structure of the dynamical system, including the presence of feedback loops. Consequently, the identification of both "inputs" and "outputs" imputes to neither that they are exogenous, or "predetermined" variables as these terms are normally used in econometrics. The causal structure of the dynamical system must be analyzed to determine which, if any, of the variables involved in the openness of the system are exogenous, or predetermined, in the usual sense. Engle et al. (1983) clarified the term "exogenous" and delineated the conditions under which the presence of an exogenous variable enabled one to make inferences about subsets of parameters in the absence of information about the joint distribution of the "exogenous" variables. The definition of "openness" used above and the corresponding notions of "inputs" and "outputs" cuts across the distinctions discussed by Engle, Hendry, and Richard, or EHR. A system may be open with a feedback loop from the outputs to the inputs in such a way that there are no "exogenous" variables in the EHR sense. Openness and isolation are concepts in specifying the properties of the dynamical system, irrespective of whether one can estimate and identify from observed data the parameters of a model describing the system. "Exogeneity" can be regarded as a constraint on the properties of joint distribution functions, whereby inferences about a model's parameters are invariant to the parameters of the marginal distribution of the exogenous variable.

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 that 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. Nonisolation means that the relationships within a system are altered by external events. This may mean either that the functional form of the equations modeling the relationship are altered by external events, or, more simply, that the coefficients in the model are altered by changes in external events. For example, the operation of the bond market interacts with the operation of the stock and money markets through variations in returns. Technological change, especially in the intermediate and short run, is the quintessential example of the effects of external events that alter the productive relationships. As another example, legislation on the operation of the market ensures that the market is not isolated. Nonisolation 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 isolated 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, though it is not an "input" into productive processes, has the effect of altering the productive relationships.

Open isolated 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, then the system and its outputs can be forecasted. Alternatively, if the causal flow is from outputs to inputs, that is, for example, the market responds to fluctuations in demand, inputs are perfectly elastic, and demand can be forecasted in its own right, then the system and its inputs can be forecasted. To forecast such a system, open and 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. However, in the former case, as is well understood, the forecasting results usually depend on forecasting without a model the behavior of the variables that are driving the system.

In the latter case, where one extends the system, and using one of the examples in a prior paragraph, we may close the financial markets by examining as one dynamical system the stock, bond, and money markets. Indeed, merely to speak of a "stock," or of a "bond" market, involves the extension of the individual markets for each instrument to a much larger market that includes the interactions between all the components. We are able to speak of a "bond market" or a "stock market" only if there is sufficient comovement between the components of each market that is, in turn, distinguishable from the comovement of the other markets. The approach of incorporating interacting subsystems, however, is rapidly self-defeating in that one soon has too large a problem to solve, with far too many variables to evaluate.

One empirical approach to limit the explosion of systems to evaluate, to obtain forecasts of relatively closed and isolated systems, is to estimate the sensitivity of the estimates of prime interest to the variation of the variables introduced by the 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.

Nonisolated systems cannot usually be forecasted. The main difficulty is that if one is dealing with a nonisolated system, then by "definition" as it were, the system under investigation is being altered by events that are usually neither recorded, nor measured by the researcher. Under certain fortuitous circumstances, it may be the case that the variation in the dynamical system can be represented as a smooth time variation in the values of the coefficients, or the path of the coefficients can be linked to some observable variable. These are the easiest examples of nonisolation with which to deal. However, one must realize that without a supporting theory, one will never know when the fortuitous circumstances cease to apply. Two examples of papers that address aspects of these issues from the econometric view are Ghysels et al. (1995), who provides a test for a structural change at an unknown breakpoint, and Hastie and Tibshirani (1993), who investigate the procedures for estimating coefficients that are themselves functions of other, but observable, variables. In the former example, there is presumed to be a single point in time at which the nonisolated system's dynamics are changed. While the change in coefficient values is recognized as a problem, no attention is paid to the econometric problem of disentangling transient dynamical behavior induced by the shift, from the new steady-state dynamics that will apply once the transients have died out. The Hastie and Tibshirani problem is one in which a system's coefficients change smoothly over time as a function of a variable outside the system; in short, the nonisolation of the system, that is, the shifting in the system's relationships, can be captured by a functional relationship between the coefficients and some other variable.

Clearly, if over time the relationships themselves are changing, then even very good fits to a 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, and the forecast variance formulae based on nontime-varying coefficients will not be relevant. Paradoxically, it is often the case that the better the fit for nonisolated 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 nonisolated. One of the effects of nonisolation is to produce model coefficient estimates that vary through time with subsamples of the data. Estimating the equations in a nonisolated system over a given fixed time period implies that the coefficient estimates, which are adjusted to minimize the squared errors, reflect a trade-off between minimizing the actual errors and the errors induced by the time-varying effects of the nonisolation. 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 subsection.

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 nonisolation 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.

There are three potential ways to deal with the issue of nonisolation, two of which are common to economists. First, one can try to pick markets and subperiods 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 usually predict the dynamics of the events that make the analyzed system nonisolated.

The second approach is to expand the system to include those aspects that account for most of the elements that make the system nonisolated. 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 practical solution.

A third approach for dealing with nonisolation is to rely on the possibility of the separation of time scales to separate out the effects of nonisolation. At very short time scales, some 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, nonisolation may be evidenced by rapidly varying effects that can be averaged out; this is, of course, the econometric assumption underlying random coefficient models (Swamy 1970). An intermediate period case is illustrated by Osborn (1990), who examined models with seasonally varying coefficients. In so far as this model is relevant, this approach is an example of nonisolation that can be modeled in terms of continuous functions of time.

Similarly, it may be determined that nonisolation 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 dynamical system has to be reinitialized. Unfortunately, it is likely that the mean arrival time is also time-varying.

A traditional method in economics to deal with aspects of nonisolation is the use of "dummy" variables, which are often used to allow a different intercept to hold during a specified period, for example, during a war, or a monetary experiment. The concentration on dynamics indicates that modeling external events that alter relationships must go beyond the simple allowance for a shift in mean values. In general, even the long-term steady-state path will be affected, as will its stability properties, so that one should examine effects beyond a change in equilibrium values. At the least, one should consider interaction effects between the externally induced shift and the effect that the other variables have on the system. In the context of linear models using analysis of variance (ANOVA), the effects that should be allowed for include both interaction and higher-order terms. However, the situation is even more complex, in that an externally induced change will induce transients into the system that produce short-run variations in the outputs of the system, in addition to those that will hold in the long-term steady state.

A point of clarification is perhaps needed about the role of "dummies" that are introduced to represent the effects of a shift in relationships. Suppose that one is modeling a productive process, and that the introduction of a dummy variable that takes the value one during a strike and zero elsewhere accurately captures the effect of the strike on the dynamics of the productive process. The dummy variable is clearly needed in the estimation process to estimate the "nonstrike" relationship. When one tries to forecast over a specific time period, two questions arise. First, to the extent that one can forecast "no strike" over the forecast horizon, the estimated model coefficient estimates provide a useful basis for forecasting production, but only given a "no strike" assumption. This is, of course, the problem. Seldom will one have accurate information about the occurrence of a strike, so that the reliability of the forecast depends on the reliability of the no-strike forecast. If there were some statistical regularity in the occurrence of strikes, so that one could estimate the probability of a strike, then one could, in confidence, produce a forecast with relevant standard errors. The second question concerns the relevancy of the estimates of the effects of a strike in that the transients of the reaction may well be state-dependent, and it will be difficult to estimate from one observational period such state dependence.

Consider, for example, the introduction of options markets into the stock market. The long-term steady-state relationship between the trade-off between return and risk was altered, as was the reaction of the stock market to news of earnings changes. In addition, for some time, perhaps several years in this example, the new market was undergoing transient changes as traders learned how to use and price the new instruments. In addition, the system would have had to adjust from the initial conditions that prevailed at the time of the change in the relationships. A similar example is provided by the "floating" of the major currencies in 1974. Reputedly, a few years were needed for the transients induced by this change to work themselves out. The difficulties in determining and estimating the coefficients of models undergoing change from external events is clear from these examples. To consider forecasting any market dependent on the introduction of options, or the floating of exchange rates, one must estimate, contemporaneously to the change in the market, the new dynamical system and one must also isolate the transients. To achieve this result would seem to require a degree of theoretical detail and precision that is far beyond anything available in economics or finance at this time.

A better example of these problems with nonisolated systems is provided by the first oil-price shock. In this case, it is easy to perceive the role of both long-run adjustments to a new dynamical relationship and the role of transients in the dynamics that were generated immediately. Oil firms and academic researchers had considerable difficulty in obtaining reliable estimates of future demand relationships, both to forecast the mean level at given prices and given levels of economic activity, and to forecast the outcome of another shock to the oil market. The adjustment process with its accompanying transient dynamics lasted for years; meanwhile, the oil industry experts who attempted to estimate "new" demand and supply relationships did not take into account the effect on their estimates of any of the dynamics that were involved in reaction to the oil-price shock.

Openness and nonisolation 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. If one cannot forecast the causal components in an open system, one cannot forecast the system outputs, no matter how well the relationship is understood. If the system is nonisolated, in general one cannot forecast at all in that the system's model is changing both in the short run due to transients, and in the long run to reflect the new dynamics. Only in those rare circumstances when one can model the effects of such external changes in the relationships might one contemplate forecasting a nonisolated system's outputs.

The concepts of closure and isolation have an interesting effect on the comparison between the statistical properties of the forecasts from linear and nonlinear models. For nonlinear models, the difficulty induced by the lack of closure and isolation are likely to indicate serious lack of fit very soon, that is, after a relatively short forecast horizon. Linear models, in contrast, as indicated above, may give no indication that there is a difficulty for some time. The problem with the linear models is subtle in that while there may be no obvious specification error to be observed in the residuals, the coefficient estimates are biased in that the estimates collapse into one value (the effect of the variable itself) and the effects of the lack of closure or isolation. While the linear models might appear to produce more stable fits in the short run, inferences drawn from them would nevertheless be in serious error in the presence of undiscovered nonisolation and openness. At least with a nonlinear model, one will usually receive early warning in terms of observable specification errors that a model misspecification has been made.

4 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. 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 nonisolation than is local fitting. One can always find some approximation to a 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 can easily 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 on forecast accuracy of a lack of accuracy in observing, or estimating, initial conditions. Their introductory paragraph presents the problem very well (Yao and 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 one's 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, in Ozaki (1985), the examination of the stochastic dynamics of a ship rolling in waves reveals 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.

Consider an example originally generated by Riggs (1963). The data-generating mechanism is given in Equation 4.1.

$$\begin{aligned} y_t &= A_0 + A_1 Exp(B_1 t) + A_2 Exp(B_2 t) + e_t, \\ A_0 &= 8.0, A_1 = 36.0, B_1 = -0.5, \\ A_2 &= = 22.0, B_2 = -0.1. \end{aligned} \tag{4.1}$$

Figure 2 shows a sample of 20 observations drawn from this model, and the regression fits using a standard nonlinear regression routine for five models: the model shown in Equation 4.1, a model with A_0 set to zero, a model with the second exponential replaced by a linear function in time, t, and finally, two models that were modeled by orthogonal polynomials in "t" of degree three and five, respectively. Only the orthogonal polynomial models provide any obvious evidence of specification error at the ends of the series. If the noise level were any higher than the signal-to-noise ratio of 17:1 that was used in Equation 4.1, the estimation results for the polynomial models would have given no obvious evidence of misspecification either. Even so, the R^2 values for the polynomial models were in excess of 0.98, and all the coefficients were very highly significant. The "exponential" models that were alternatives to Equation 4.1 had even better regression fits, but the "t-ratios" were somewhat lower, although unambiguously highly significant. With the possible exception of the two orthogonal polynomial models, each model gives the appearance of providing an excellent fit to the data.

Consider now Figure 3, in which is shown the projections of each of these models beyond the sample range. To stress that the problem that I am addressing with this example is not a result of "over-fitting" to sampling variation, the extrapolations presented in Figure 3 are based on coefficient estimates obtained from the "no-error" version of the model shown in Equation 4.1, that is, e_t is set identically to zero. All of the other "seemingly correct" models are wildly at variance with the extrapolations of the correct model and any common forecast standard-error calculation would be totally irrelevant.

Using the last few observations to provide a fit by minimum mean-forecast error would not aid the matter in these cases, as the same basic fits would be obtained. The source of the "error" here is that the alternative models do not impose the global constraints that underlie the original model in that as "t" approaches infinity, the system being modeled converges to a lower asymptote of 8.0.

Yet another problem occurs when the system is not well understood and the researcher is searching for a model of the data. Under these conditions, 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

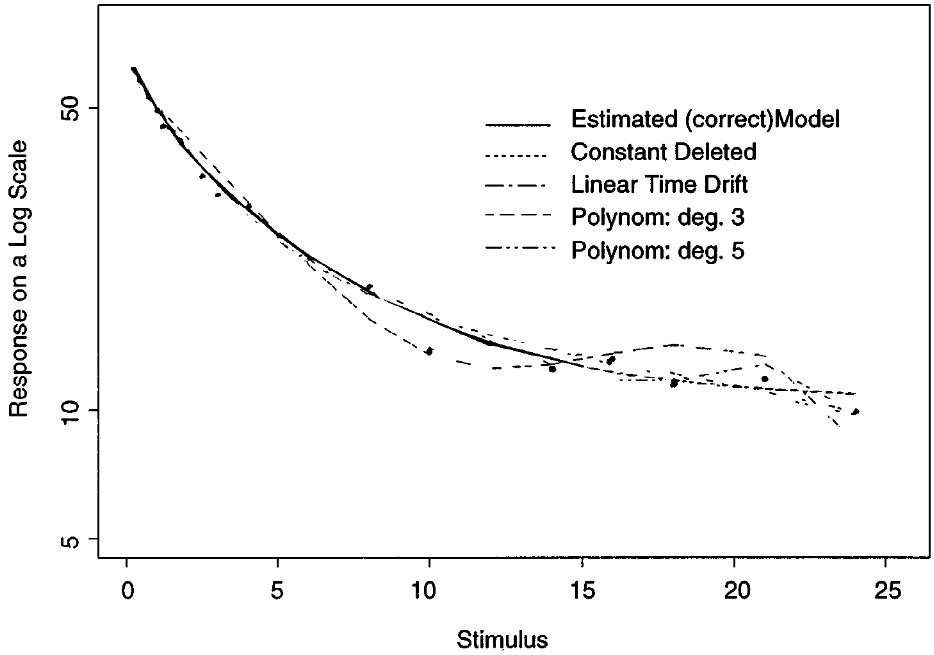

Figure 2
Fits Using Five Alternative Models.

True Model: $8.0 + 36.0^* \exp(-0.5^* t) + 22.0^* \exp(-0.1^* t) + e$; $e \sim N(0, 1)$

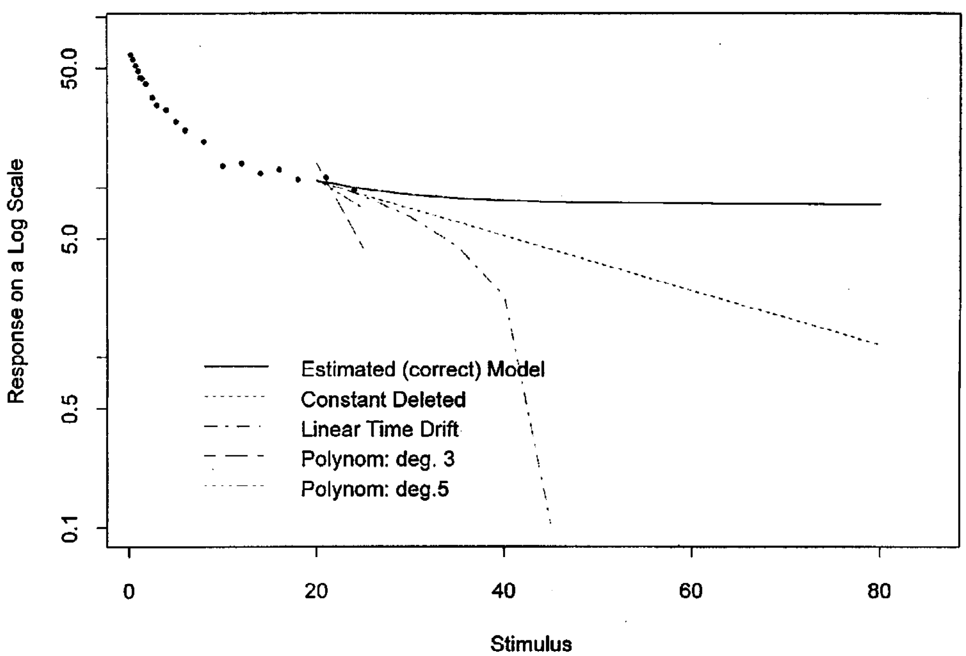

Figure 3
The Dangers of Extrapolation from Estimated Relationships.
Extrapolations based on coefficients estimated without error terms

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 the 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 with 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 10 drawings from [0, 1] with equal probability. If we count the number of outcomes that pattern searchers would regard as "without pattern" in a string of 10 drawings, we get a number on the order of 256, and perhaps considerably less. But there are a total of 2^{10}, or 1,024, drawings, so that the probability of observing some pattern is nearly three-quarters. The general lesson is clear: while the probability of a specific pattern is very low, the probability of observing *some* pattern is very high, even with relatively large numbers of observations.

After comparing various parametric and nonparametric models using simulated data, Granger and Teräsvirta (1992) 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, nonparametric formulations that are usually "representational" lead to local over-fitting in the regression, and therefore to worse forecasts when the "over-fitted" results are extrapolated. Swanson and White (1995) came to similar conclusions; indeed, these authors felt compelled to design what they called their "insanity filter" to prevent the neural net results from producing obvious and egregious errors. There is seldom sufficient attention paid to the properties of the approximating model and whether those properties are reasonable and can be extrapolated.

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 avoiding the tendency to overfit local random patterns. An early example of efforts to provide "global" constraints to improve the estimation of coefficients in a nonlinear model is Diewert and Wales (1987), which followed early work by Lau (1978), Gallant, and Golub (1984). The task that Diewert and Wales set for themselves was to estimate flexible functional forms that would still meet global curvature conditions that were believed to be incontestable. Consider a further 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 were 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 subperiods; that is, the model was not altered over time merely to meet local subperiod "fits." In short, local evidences of fit were exchanged for evidence of global consistency. This approach marks a departure from 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.

4.1 Sampling Rate and Aggregation

Another problem that is aggravated 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 short-run 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. In contrast, 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; see, for example, the discussion on "intermittency" in Lorenz (1989). 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, 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 macrovariables, and is discussed in full in another paper (Ramsey 1994). Nonetheless, it is clear that even C^{∞} functions contain few subclasses 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 with the phase itself varying over time, as compared to aggregating two harmonic series with a fixed phase difference. The former aggregate series is difficult to analyze, while the latter aggregate is simply handled by Fourier analysis.

5 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. Before proceeding, some clarification of the idea of "forecasting" is useful.

5.1 A Few Clarifying Comments on Forecasting in General

One may distinguish one-step-ahead forecasts from multiple-step-ahead forecasts, and whether or not there is contemplated a sequence of such forecasts. The distinction between single- and multiple-step-ahead forecasts mainly involves complications when nonlinear models are used (Granger and Teräsvirta 1992).

The first difficulty that arises with multistep forecasts with nonlinear models can be briefly indicated by a simple example. If an output y_t depends on an input x_t through a nonlinear function $g(\cdot)$, and one wishes to forecast y_t on the basis of forecasts of x_t itself, then the nonlinearity of $g(\cdot)$ complicates the choice of how to forecast the x_t for which, in general, there will be no model. For example, suppose that phenomenologically it is known that:

$$x_t = \alpha x_{t-1} + e_t, \tag{5.1}$$

where e_t is IID. The one-step-ahead forecast for x_t at time $(t-1)$ is simply αx_{t-1}, so that the one-step-ahead forecast for y_t is $g(\alpha x_{t-1})$. But for even a two-step-ahead forecast, we have several choices. The first is the "exact" forecast that is given by averaging $g(\alpha x_t + e)$ with respect to the stationary distribution for $[e]$ conditional on x_{t-1} that was specified in Equation 5.1. We might also consider the "naive" forecast that replaces $\alpha x_t + e_t$ by its mean, αx_t. The difference between the two results will be due solely to the extent of curvature in the function $g(\cdot)$, and therefore to the difference between the expectation of $g(\cdot)$ with respect to e_t, conditional on the information available at time $(t-1)$ and evaluating $g(\cdot)$ at the expectation of e. One may consider other procedures that attempt to address the problem that multistep forecasts of y_{t+h} depend on forecasts of x_{t+h}, which itself can only be forecast with error and usually without an explicit model. The "practical" forecast involves two sources of error: those due to the estimation of the model, and those due to the need to forecast the conditioning events for which there is usually no model. Theoretically, one can consider forecasts based on known values of the conditioning events; such an approach is useful in trying to separate the modeling-induced errors from those due to our lack of knowledge of the time path of the conditioning events.

If we contemplate a sequence of forecasts, one step ahead or multistep, one has a further set of choices: forecasts using estimates based on information only available at the time of the initial estimation of the model, using all information available up to the beginning of each forecast in the sequence, or only using the last T elements of information available before each forecast, where T is the number of observations used in the initial estimation. I do not at this time wish to discuss the strategy that one might use in combining these various forecast procedures to learn more about the structural stability of the model, as that discussion is not germane to the main argument. For the current purposes, the various choices reflect the researchers' beliefs about the constancy of the model over time. If one believes that an appropriate model is being used and that the coefficients are constant over time, then one's forecasts will gain in accuracy by capitalizing on the increased sample size that will come from using all available information. However, if one believes that the coefficients are "slowly varying," then one will be concerned to choose an appropriate window size, say of length T^1, and base one's forecasts on only the last available T^1 observations.

The usual practice is to choose small values for h, the forecast horizon, often just one or two periods. But because of inertia in most economic systems, forecast comparisons and the attempt to choose between models on such short horizons is often unproductive. Compare, for example, the results shown in Figure 3: one-step-ahead forecasts might be within two standard deviations, but three- or four-step-ahead forecasts are certainly not. The real key to discriminating between models is not so much the length of the time interval as it is the magnitude of the change in the signal to the magnitude of the ambient error. Longer time horizons usually ensure that the size of the change in the signal is greater, and therefore the ability to discriminate between models is enhanced.

One of the difficulties with recommending comparisons of forecasts is that the forecast errors across different models are highly correlated (Diebold and Mariano 1994). More recently, Ashley has contributed an alternative resampling-based procedure to deal with this class of problems (Ashley 1995). In particular, Ashley found that considerably more than 5 to 20 periods is required to detect reductions of mean-square error of 20 to 30 percent; this result confirms the comments above.

In the above discussion it was expositionally convenient to restrict attention of the forecast problem to one of forecasting means. While this approach dominates the current literature, one might well contemplate the notion that the entire distribution is evolving over time and that it is the distribution that should be forecasted. In principle, this extension of ideas is most persuasive. Indeed, the references cited above dealing with the ARCH literature indicate an important extension of the idea of forecasting from means to variances. We might well with benefit consider the higher moments, especially the third, and indeed we should consider the time variation of the whole distribution. However, this introduces a substantial research project in itself.

5.2 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 logically link 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 are the distinguishing characteristics 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, 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 1991), wavelets (Ramsey and Zhang 1994, 1995). Yet another representation arises out of neural network theory using logistic functions (Hornik et al. 1989; Swanson and White 1995), 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.

Clearly, representations are used when we do not know or cannot easily discover the model itself. However, the choice of representation implies a given level of information about the system. We may know that a system is the result of a mapping, for a famous example, the logistic map, $X_{t+1} = rX_t(1 - X_t)$; or we may only know that the output is bounded between zero and one, or we may know that the signal is smooth, but is contaminated by noise, or we may only know that the system seems to be ergodic, stationary, and uncorrelated; and so on. Our state of knowledge indicates in part at least the most relevant representation to be used. As we shall see, if we know that we are dealing with some nonlinear map, but do not know what it is, we would be throwing away valuable information if we used a Wold representation. If we know that our signal is smooth and highly differential, we might well find Fourier transforms very useful. However, if we know that the signal is very irregular with a low-level noise floor, we should consider wavelets.

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:

$$X = UDV \tag{5.2}$$

where D is a diagonal matrix with the singular values, or eigenvalue weights, on the diagonal, and U and 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 5.2 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.

Having obtained a representation, one can ask whether there are useful low-order approximations, compare the representations with other representations obtained from similar data, and the best use of all, examine whether the representation provides any guides toward the formulation of a model. In this last use, the usual procedure is to concentrate on major effects and suppress details to aid in the revelation of some structural characteristics that might be useful in modeling the data. The procedure is warranted by the fact that any model of the data must have the same representation. Suppose in this example that there is some nonlinear model for the term structure of interest rates over time that will yield the observed data. Then it is clear that the representation illustrated in Equation 5.2 is a representation of the model as well. However, the nonlinear model that purports to characterize the behavior of the mechanism underlying the generation of the data is very important for interpreting the data and understanding the conditions under which the generating mechanism performs as observed. Further, the nonlinear model will, in general, incorporate far more information about the system than is usually true for any representation. Indeed, the usefulness of representations is that they are representations for broad classes of relationships, or systems, so that necessarily they will incorporate far less *a priori* information about a specific system.

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 5.3,

$$y_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} \tag{5.3}$$

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 5.3 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.

As I indicated above in the context of a singular value decomposition of interest rates, all models themselves have representations. Another example is provided by real business cycles that have the representation shown in Equation 5.3. Even if the data-generating mechanism that underlies real business cycles is nonlinear in structure, but produces data that are stationary, then that nonlinear model will always produce data that are representable by a Wold decomposition. 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 autoregressive integrated moving average (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, $j = 1, 2, \ldots, k$.

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. 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. To begin, we should recognize that when comparing forecasting capabilities, we should remember the degree of information that is incorporated in the choice of representation. Different representations incorporate different levels of *a priori* information, as I indicated above. For example, 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 the model and its representation depend on the relative degrees of information that are incorporated in the two. Using only the information that the series is stationary, a Wold representation is a reasonable choice. But if we know that the series is generated by some map, such as the logistic mentioned above, then it is immediately clear that we will gain in forecasting capabilities by utilizing such information. The difficult choices between representations and purported nonlinear models occur when the disparities in information are not so great and the representation, perhaps with the use of more degrees of freedom, provides equally good forecasts.

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, which 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). Nevertheless, 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 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 nonstationary. 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 and Zhang 1994, 1995). Waveform dictionaries therefore can provide a very wide range of structures to represent data, including 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.

To strengthen the argument, assume that whatever mechanism generated the data during the estimation period continues to apply during the forecast period. First, insofar 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, insofar 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 anything, 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 a 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 a 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, useful forecasts in which some reliability can be placed because the "causal mechanism" is at least partially understood.

6 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 modeling 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. Another key point is that open and nonisolated systems 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 isolation are good approximations.

The conventional rule of thumb that "good fits do not give good forecasts" is justified by recognizing that while fitting is local, forecasting is an essentially global idea. Even good modeling 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.

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, forecasts are more sensitive to the errors in evaluating initial conditions than is true for regression 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. The forecasting capabilities of representations depend in part on the generality of the supposed class underlying the representation; alternatively expressed, the forecasting capabilities depend on the level of *a priori* information that is included in the choice of representation. 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.

References

Alogoskoufis, George S., and Thanasis Stengos (1991). "Testing for Nonlinear Dynamics in Historical Unemployment Series." European University Institute Working Paper ECO 91/38, Badia Fiesolana, Italy.

Ashley, Richard (1995). "Non-Nested Model Selection/Validation: Making Credible Post-Sample Inference Feasible." Economics Dept. Working Paper E95-07, Blacksburg, Virginia: Virginia Tech.

Ashley, Richard A., Douglas A. Patterson, and Melvin J. Hinich (1986). "A Diagnostic Test for Nonlinear Serial Dependence in Time Series Fitting Errors." *Journal of Time Series Analysis*, 7(3):165–178.

Ashley, Richard A., and Douglas M. Patterson (1989). "Linear Versus Nonlinear Macroeconomics: A Statistical Test." *International Economic Review*, 30(3):685–704.

Barnett, William A., A. Ronald Gallant, and Melvin Hinich (1993). "Detection of Nonlinearity and Chaos: Application to Money Stock." In Dept. of Economics Working Paper Series 178, Washington University, St. Louis, Missouri.

Bollerslev, T., R. Y. 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, Massachusetts: MIT Press.

Brock, William A., and Simon M. Potter (1993). "Nonlinear Time Series and Macroeconomics." In G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Handbook of Statistics*, vol. 11. New York: Elsevier Science, pp. 195–229.

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

Casdagli, Martin (1992). "A Dynamical Systems Approach to Modeling Input-Output Systems." In Martin Casdagli and Stephen Eubank, eds., *Nonlinear Modeling and Forecasting, Diagnostic Testing for Nonlinearity, Chaos, and General Dependence in Time Series Data*. Redwood City, California: Addison-Wesley, pp. 265–282.

Dacorogna, M. M., et al. (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, Theodore E., and Craig M. Lewis (1988). "Stock Market Volatility and the Information Content of Stock Index Options." *Journal of Econometrics*, (April/May) 52:267–287.

Diebold, Francis X., and Roberto S. Mariano (1994). "Comparing Predictive Accuracy." Technical Working Paper Series 169, Cambridge, Massachusetts: National Bureau of Economic Research.

Diebold, Francis X., and James A. Nason (1990). "Nonparametric Exchange Rate Prediction?" *Journal International Economics*, 28:315–332.

Diewert, W. E., and T. J. Wales (1987). "Flexible Functional Forms and Global Curvature Conditions." *Econometrica*, 55(1):43–68.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard (1983). "Exogeneity." *Econometrica*, 51(2):277–304.

Gallant, A. R., and E. H. Golub (1984). "Imposing Curvature Restrictions on Flexible Functional Forms." *Journal of Econometrics*, 26:295–321.

Ghysels, Eric, Alain Guay, and Alastair Hall (1995). "Predictive Tests for Structural Change with Unknown Breakpoint." *CIRANO Scientific Series*, (March) 95s-20:1–27.

Granger, C. W. J., and Timo Teräsvirta (1992). "Experiments in Modeling Nonlinear Relationships Between Time Series." In Martin Casdagli and Stephen Eubank, eds., *Nonlinear Modeling and Forecasting, Diagnostic Testing for Nonlinearity, Chaos, and General Dependence in Time Series Data*. Redwood City, California: Addison-Wesley, pp. 189–198.

Granger, Clive W. J. (1991). "Developments in the Nonlinear Analysis of Economic Series." *Scandinavian Journal of Economics*, 93(2):263–276.

Hastie, Trevor, and Robert Tibshirani (1993). "Varying-Coefficient Models." *Journal of the Royal Statistical Soc. Series B*, 55(4):757–779.

Hornik, K., M. Stinchcombe, and H. White (1989). "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks." University of California, San Diego DP 89-34. La Jolla: Economics Dept.

Hunter, Norman F., Jr. (1992). "Application of Nonlinear Time Series Models to Driven Systems." In Martin Casdagli and Stephen Eubank, eds., *Nonlinear Modeling and Forecasting, Diagnostic Testing for Nonlinearity, Chaos, and General Dependence in Time Series Data*. Redwood City, California: Addison-Wesley, pp. 467–492.

Jaditz, Ted, and Chera Sayers (1993). "Are Chaotic Attractors Generic in Economic Data?" *International Journal of Bifurcation and Chaos*, 3:745–755.

Lau, L. J. (1978). "Testing and Imposing Monotonicity, Convexity and Quasiconvexity Constraints." In M. Fuss and D. McFadden, eds., *Production Economics: A Dual Approach to Theory and Applications*, vol. 1. Amsterdam: North Holland, pp. 409–453.

LeBaron, Blake (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, Hans-Walter (1989). "Nonlinear Dynamical Economics and Chaotic Motion." *Lecture Notes in Economics and Mathematical Systems*, no. 334. New York: Springer-Verlag.

Lorenz, Hans-Walter (1993). "Complexity in Deterministic, Nonlinear Business Cycle Models—Foundations, Empirical Evidence, and Predictability." In F. Gori, L. Geronazzo, and M. Galeotti, eds., *Nonlinear Dynamics in Economics and Social Sciences*. Berlin: Springer-Verlag, pp. 17–52.

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

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

McCurdy, Thomas H., and Thanasis Stengos (1988). "A Comparison of Risk-Premium Forecasts Implied by Parametric versus Nonparametric Conditional Estimators." *Journal of Econometrics*, 52:225–244.

Meese, Richard A., and Andrew K. Rose (1991). "An Empirical Assessment of Nonlinearities in Models of Exchange Rate Determination." *Review of Economic Studies*, 58:603–619.

Mittnik, Stefan (1991). "State Space Modeling of Nonlinear Time Series." Stony Brook, NY: State University of New York, Dept. of Economics. Economics Working Paper 90-336.

Mizrach, Bruce (1992). "Multivariate Nearest Neighbor Forecasts of EMS Exchange Rates." *Journal of Applied Econometrics*, Supplement 7:S151–S164.

Osborn, D. R. (1990). "The Implications of Periodically Varying Coefficients for Seasonal Time-Series Processes." *Journal of Econometrics*, 48:373–384.

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

Ozaki, Tohru (1985). "Non-Linear Time Series Models and Dynamical Systems." In E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, eds., *Handbook of Statistics*, vol. 5. New York: Elsevier Science, pp. 25–83.

Pesaran, M. Hashem, and Simon M. Potter (1993). "A Floor and Ceiling Model of U.S. Output." Los Angeles, California: University of California, Los Angeles, Dept. of Economics, Working Paper 9407.

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

Prescott, David M., and Thanasis Stengos (1991). "Testing for Forecastable Nonlinear Dependence in Weekly Gold Rates of Return." European University Institute Working Paper Series ECO 91/31, Badia Fiesolana, Italy.

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

Ramsey, James B. (forthcoming). "On the Existence of Macro Variables and of Macro Relationships." *Journal of Economic Behavior and Organization*.

Ramsey, James B. (1992). "Seasonal Economic Data as Approximate Harmonic Oscillators." New York: C. V. Starr Center for Applied Economics, New York University, Working Paper 92-16.

Ramsey, James B., and Philip Rothman (forthcoming). "Time Irreversibility and Business Cycle Asymmetry." *Journal of Money, Credit, and Banking*.

Ramsey, James B., and Zhifeng Zhang (1994). "The Application of Wave Form Dictionaries to Stock Market Data." New York: C. V. Starr Center for Applied Economics, New York University, Working Paper series 90-05.

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

Ramsey, James B., and Zhifeng Zhang (1995). "The Analysis of Foreign Exchange Data Using Waveform Dictionaries." New York: C. V. Starr Center for Applied Economics RR#95-03, New York University.

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

Riggs, Douglas S. (1963). *The Mathematical Approach to Physiological Problems*. Baltimore, Maryland: Williams and Wilkins.

Rothman, Philip (1994). "Forecasting Asymmetric Unemployment Data." Working Paper, Greenville, North Carolina: East Carolina University, Dept. of Economics, Working Paper 93-39.

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

Stock, James H. (1987). "Measuring Business Cycle Time." *Journal of Political Economy*, 95:1240–1261.

Stock, James K., and Mark W. Watson (1988). “Variable Trends in Economic Time Series.” *Journal of Economic Perspectives*, 2(3):147–174.

Sugihara, G., B. T. Grenfell, and R. M. May (1990). “Distinguishing Error from Chaos in Ecological Time Series.” *Philosophical Transactions of the Royal Society, London, Series B*, 330:235–251.

Swamy, P. A. V. B. (1970). “Efficient Inference in a Random Coefficient Regression Model.” *Econometrica*, 38(2):311–323.

Swanson, Norman R., and Halbert White (in press). “A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks.” *The Review of Economics and Statistics*.

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.

Editorial Policy

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826