ELSEVIER International Journal of Forecasting 10 (1994) 191–207
international journal of forecasting

A nearest neighbor model for forecasting market response

Francis J. Mulhern[a], Robert J. Caprara[b]
[a]Department of Marketing, Pennsylvania State University, University Park, PA 16802, USA
[b]Raptor Consulting Group, 203 Brooke Drive, West Chester, PA 19380, USA

Abstract

Researchers in marketing often are interested in modeling time series and causal relationships simultaneously. The prevailing approach to doing so is a transfer function model that combines a Box–Jenkins model with regression analysis. The Box–Jenkins component assumes that a stationary, stochastic process generates each data point in the time series. We introduce a multivariate methodology that uses a nearest neighbor technique to represent time series behavior that is complex and nonstationary. This methodology represents a deterministic approach to modeling a time series as a discrete dynamic system. In this paper we describe how a time series may exhibit chaotic behavior, and present a multivariate nearest neighbor method capable of representing such behavior. We provide an empirical demonstration using store scanner data for a consumer packaged good.

Keywords: Forecasting; Nearest neighbors; Chaos

1. Introduction

Time series data is commonly analyzed in marketing to ascertain causal relationships and make forecasts. Often researchers want to jointly model time series patterns and causal relationships. The most prevalent approach for doing so is transfer function modeling (Leone, 1983). The time series portion of a transfer function model consists of the autoregressive and moving average components popularized by Box and Jenkins (1976). Transfer function models work by placing a fixed structure on the behavior of variables over time. However, in some situations a time series may exhibit a pattern that cannot be represented with a Box–Jenkins formulation. One such situation occurs when chaos exists in a time series.

The portfolio of forecasting techniques available for marketing analysis should include an approach that accommodates the complex patterns exhibited by chaos and other nonlinear phenomena. In this paper, we present an approach to jointly modeling time series and explanatory relationships that, unlike Box–Jenkins based approaches, is capable of representing chaotic time series patterns. We present a model that combines a univariate nearest neighbor model with regression analysis to forecast market response. We apply this model to the problem of forecasting weekly brand sales for a consumer packaged good in a retail store.

The need for such a methodology is based on the possibility that marketing time series may exhibit complex behavior. Therefore, we begin with a brief overview of chaos theory.

SSDI 0169-2070(94)09000-U

2. Chaos theory

Recently, several researchers have considered the prospects of chaotic time series patterns in marketing (McQuitty, 1992; Henson and Rigdon, 1992). Donthu and Rust (1986) suggest that incorporating chaotic behavior into marketing models may be useful because marketing phenomena need not exhibit simple, equilibrium behavior.

Chaos is characterized by simple deterministic systems that exhibit complex behavior in a time series. The term chaos is appropriate because it implies images of complex and changing patterns. Those patterns tend to be so numerous and intertwined that they appear to have no order at all. More importantly, chaotic patterns strongly resist quantification with traditional methodologies. General introductions to chaos theory are provided by Gleick (1987) and Stewart (1989).

The discovery of chaos confounds the meaning of deterministic and stochastic systems. Deterministic systems contain fixed rules that involve no elements of chance. Stochastic systems draw on probability theory to accommodate random behavior. Chaos clouds the distinction between deterministic and stochastic behavior because it embodies fully deterministic behavior that appears random to the human eye. Research in physical and biological systems has discovered a family of deterministic equations that produce seemingly random time series patterns that are actually deterministic.

Chaos can be modeled with nonlinear dynamic systems. A dynamic system consists of two parts: a state—the essential information about the system—and a dynamic—the rule that describes how the state evolves over time. Consider the example of a pendulum. The state of a pendulum consists of its position and velocity. The dynamic is a differential equation that describes how position and velocity change over time. In Box–Jenkins ARIMA models used in marketing, the state consists of present and lagged values of a time series that represent all the essential information about the time series pattern. The dynamic is the difference equation that expresses a value in the time series as a function of lagged values and error terms. Dynamic systems can be modeled either in continuous time, with differential equations, or discrete time, with difference equations. In marketing, time series models consist of difference equations since marketing time series such as product sales and market share are measured at discrete time intervals.

A primary concept in modeling dynamic systems is an attractor. An attractor is a value, or set of values, to which a system eventually settles. For example, the equation $y_t = 0.5y_{t-1}$ asymptotically converges towards the equilibrium value of zero. That is an example of a point attractor. A second type of attractor is a periodic, or limit, attractor. Here a time series converges to an oscillatory state, continuously repeating the same pattern. These types of attractors are simple and orderly and usually can be identified by observing a time plot.

Chaotic behavior introduces a third type of attractor that is disorderly, complex and frequently unidentifiable by visual inspection of a time plot. When a chaotic attractor is present, behavior can appear to be totally random but actually may be deterministic. A chaotic (strange) attractor is a bounded path which, while having a fully determined path, never recurs. Over the duration of a time series, various attractors may operate. For instance, a series may exhibit periodic behavior in one portion of the time series, and chaotic behavior in another.

We demonstrate some properties of chaos with the two plots shown in Fig. 1 which have been generated from the equation $y_t = 2t_{t-1}^2 - 1$. Fig. 1(a) shows a chaotic time series pattern that occurs with a starting value of 0.54321. An important property of a chaotic time series is that the behavior of the series frequently *resembles* past behavior of the series. Inspection of Fig. 1(a) clearly shows similar, though not identical, patterns of behavior at various points in the time series. Identical patterns of behavior at different points in time characterize a periodic attractor, while non-identical but similar patterns of behavior often represent the presence of chaos. Fig. 1(b) shows another time plot gener-

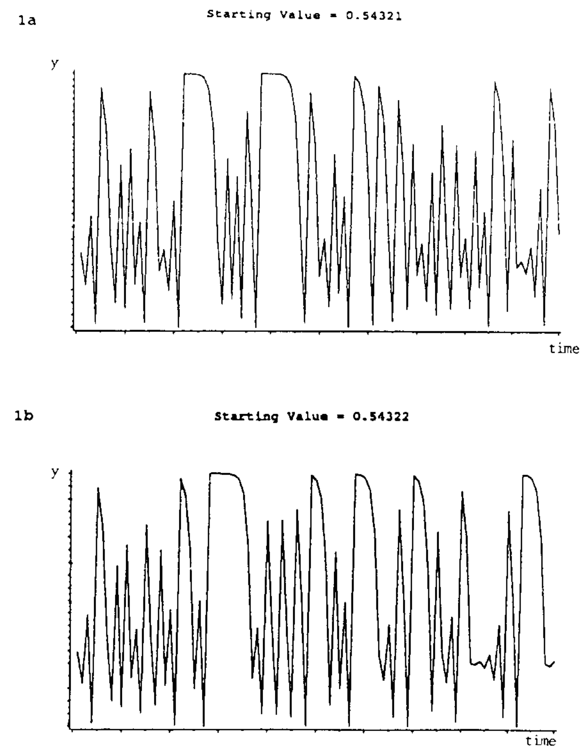

Fig. 1.

ated from the same equation with the only difference being the slightly higher starting value of 0.54322. Note that while the overall patterns of Figs. 1(a) and (b) are similar, the actual values for specific points in time are quite different. An extremely small change in the starting value results in a substantially different time series. That property is known as sensitivity to initial conditions, and is one of the defining characteristics of chaos.

The prospect of chaotic behavior has serious implications for time series modeling. A fundamental part of building time series models with traditional methodologies is the discrimination between random and nonrandom behavior. For example, in building a Box–Jenkins model, a researcher specifies autoregressive and moving average components that represent time related effects such as trend and periodic behavior. The resulting Box–Jenkins model filters the time series into a white noise (i.e. random) process. Elements of a time series that cannot be decomposed into a combination of autoregressive and moving average components are treated as random processes. When chaos exists in a time series, methods such as Box–Jenkins analysis erroneously treat the chaotic behavior as random because the patterns of chaos pass the normal tests of randomness. The possibility of chaos means that time series techniques must be able to discriminate between truly random behavior and chaotic behavior in order to perform as desired.

In this paper, we address the problem of jointly modeling time series and causal phenomena. The most prevalent approach for doing so is to combine a Box–Jenkins model with a regression model to form a transfer function model. The approach presented in this paper likewise represents a combination of regression with time series analysis. It differs in that the time series component consists of a deterministic nearest neighbor model capable of representing chaos rather than the autoregressive, moving average elements of a transfer function model.

3. Time series analysis in marketing

Time series analysis relies on the dependence of observations at different points in time. Unlike econometric and multivariate models that are theoretically specified, time series models are specified to maximally predict a time series. The best time series model is the one with the highest predictive accuracy.

Time series models serve two general purposes in marketing: forecasting values of a certain variable (e.g. Moriarty and Adams, 1984), and removing time related effects (trend, seasonality, etc.) from a time series so that causal relationships can be explored (Hanssens, 1980; Leone, 1983). The earliest forms of dynamic models consisted of simple deterministic equations to make short term forecasts. Examples of these include simple trend models (e.g. $y_t = \alpha y_{t-1}$), moving average models (e.g. $y_t = 1/3(y_{t-1} + y_{t-2} + y_{t-3})$), and exponential smoothing. Such deterministic formulations were supplanted by stochastic models popularized by Box and Jenkins (1976). The commonality of all these approaches is that short term forecasts are made by extrapolating lagged values of a time series.

In a Box–Jenkins model, a stationary time series is decomposed into a stochastic process consisting of autoregressive and moving average polynomials and random error. Structurally, Box–Jenkins modeling has been largely linear, as correlation coefficients provide the foundation for the model building process. The only nonlinear aspects have been the procedures required to estimate models containing moving average components. Linear models are a special case of the more general class of nonlinear models which are becoming more prevalent in time series analysis (Priestley, 1988).

An alternative stochastic method of modeling time series data is spectral analysis (Priestley, 1981). Spectral analysis evolved from periodogram analysis in which a time series is assumed to consist of a superposition of sine and cosine waves of different phases and frequencies and a stochastic noise component. Spectral analysis differs from periodogram analysis in that the entire time series, not just the noise component, is treated as stochastic. While spectral analysis differs procedurally from Box–Jenkins modeling, it is philosophically similar in that a model is developed to separate stochastic, time related processes from the random components in the data. What's more, the two procedures, under certain circumstances, are equivalent (Pollock, 1987, p. 133).

Box–Jenkins and spectral analysis diagnostic procedures evaluate a time series to determine if nonrandom behavior exists. One criteria for a good Box–Jenkins model is that the residuals of the modeled series are Gaussian white noise. When chaotic behavior exists in a time series, Box–Jenkins procedures falter because the behavior generated by the chaotic process passes the tests of randomness. The result is a time series model that, while satisfying the diagnostic checking criteria in a Box–Jenkins analysis, is unable to make accurate forecasts.

While existing methods of time series analysis are appropriate in many situations, there may be times (e.g. when chaos is present) that such methods are unable to adequately represent a time series. Further, in Box–Jenkins analysis, it is not difficult to select an inappropriate model. For example, an overparametized model can pass diagnostic tests but yield poor forecasts. Such problems are particularly acute when explanatory variables are added to form a transfer function model. Harvey (1989, p. 371) notes that transfer function models become cumbersome when two or more explanatory variables are included, and such models do not fare well in systematic model selection procedures.

The approach we describe fundamentally differs from the prevailing Box–Jenkins and spectral analysis procedures in that it is highly nonlinear and deterministic. Our method incorporates techniques from a relatively recent class of methodologies for dealing with chaotic behavior in dynamic systems. As we describe below, this method deterministically establishes a nearest neighbor representation of a time series without specifying any functional relationship or estimating any parameters. This representation creates a subset of data most applicable to the point in time under scrutiny. A regression analysis is performed on this subset to estimate causal relationships and make a forecast.

4. Applications of chaotic models to time series

Much of the literature on chaos theory focuses on illustrating the complexity of simple nonlinear equations. Chaotic dynamics typically are introduced with the following nonlinear function used to describe the fluctuations of a population over time (May, 1976).

$$y_t = \alpha y_{t-1}(1 - y_{t-1}) \qquad (1)$$

where y_t is the value of a time series in period t and α is a constant.

Depending on the value of the parameter α, Eq. (1) can demonstrate a wide range of behaviors, including periodic and chaotic behavior. The path of a time series represented by Eq. (1) converges to an attractor (point, periodic or chaotic) as a function of α. Detailed discussions of the behavior of this equation can be found in Baumol and Benhabib (1989, p. 84), May (1976, p. 460) and Stewart (1989, p. 155).

Several studies in the physical and social sciences have explored the application of chaotic dynamics to a wide variety of topics. Some of the more notable examples of applications of chaotic behavior include shifts in biological populations (May, 1976), weather patterns (Lorenz, 1963), brain waves (Skarda and Freeman, 1987), strategic decision making (Richards, 1990) and rational choice behavior (Benhabib and Day, 1981). The possibility for chaos in marketing exists as researchers regularly use the logistic equation to represent new product growth (Bass, 1969) and competitive market evolution (Lambkin and Day, 1989). Such dynamic growth models can exhibit chaotic behavior for certain parameter ranges, a possibility not considered in the many published papers in marketing that use the logistic equation.

Most existing empirical studies of chaos consist of applications to a time series of one of the discrete dynamic models (such as Eq. (1)) that can exhibit chaotic behavior. There is evidence that the behavior of biological and physical systems can be represented by equations capable of exhibiting chaos. So too there is some evidence of chaos in economic and social sciences. For example, Frank and Stegnos (1989) find evidence of chaos in economic time series of gold and silver rates of return.

Many of the existing applications of chaos to the social sciences have consisted of ad hoc applications of equations capable of representing chaotic behavior to a time series (e.g. Baumol and Benhabib, 1989, p. 84). In describing the shortcomings of such applications in economics, Mirowski (1990, p. 300) notes,

> While this is responsible mathematical pedagogy, it has been disastrous for economists, because it gives the impression that somehow this is what chaos is all about, not to mention giving rise to the temptation to generate all sorts of single-variable matchbook models where a single difference equation is somehow supposed to "account" for the apparently stochastic behavior of stock prices, macro fluctuations, monetary disturbances and every other scourge known to mankind.

Chaos may or may not exist in the behavior of marketing time series. Rather than apply one of the equations known to exhibit chaotic behavior to a marketing time series, we develop a multivariate model that can represent chaos if it exists in a time series. Traditional Box–Jenkins and spectral analysis procedures are unable to handle chaotic behavior, as they require a time series to be stationary (or convertible to stationarity through differencing). A time series is stationary when it has a constant mean, a finite variance, and the autocovariances depend only on the length of the lag (see Brockwell and Davis, 1991, p. 12). Since chaotic behavior can be nonstationary, it cannot be represented with traditional time series modeling. An alternative approach to time series analysis is a class of methodologies known as nearest neighbor techniques that do not require the statistical equilibrium of stationarity. Rather than relying on any statistical property, these techniques rely on a time series resembling itself; a property evident in the chaotic patterns shown in Fig. 1.

5. Nearest neighbor models

A general approach to modeling chaotic behavior is the concept of nearest neighbor analysis (Stone, 1977; Packard et al., 1980). Nearest neighbor models use lagged values of the time series to form vectors in a state space. The vectors are used to reconstruct the attractor of the time series. A prediction is based on the historical paths of the vectors around the attractor.

Nearest neighbor models do not assume a functional relationship. Instead, they geometrically attempt to reconstruct whatever attractor is operating in a time series. To date, nearest neighbor models have been limited to univariate time series such as population levels in biology (May, 1976) and liquid flows in physics (Miles, 1984). We have found that the applicability of such univariate models to marketing phenomena is very limited because few marketing time series can be predicted without consideration of causal variables. For example, the time series behavior of sales and market share cannot be represented very well without the inclusion of effects of advertising, price and promotion. To overcome that limitation, we develop a multivariate nearest neighbor methodology that incorporates the effects of causal variables in making forecasts.

5.1. A multivariate nearest neighbor model

We present a discrete dynamic model that forecasts the values of a time series subject to the effects of explanatory variables. Our model relies on the concept of a geometric analysis of state space projections described by Sugihara and May (1990), and is philosophically similar to a method developed by Farmer and Sidorowich (1988). The methodology provides a geometric solution to the problem of describing time series behavior that may be chaotic. Such a geometric representation circumvents the need to specify an explicit mathematical equation to represent a time series pattern. The method uses a nearest neighbor model to represent time effects and includes a regression analysis to represent the effects of causal variables.

Our approach falls into a general class of models known as robust regression (Stone, 1977). Cleveland (1979) describes a robust weighted regression method that combines regression with nearest neighbor analysis. Rust and Bornman (1982) apply Cleveland's approach to calibrate the relationship between coffee consumption and age. Rust (1988) develops a comparable distribution free method for doing regression.

The method we describe differs from that of Cleveland in purpose and technique. Our model is designed as a forecasting technique whereas Cleveland's model is set up to estimate regression coefficients. Our approach is a nearest neighbor forecasting method in which the forecasts are adjusted based on regression coefficients. In contrast, Cleveland provides a regression model in which the observations are weighted according to nearest neighbor distances. Conceptually, though not computationally, our model is a special case of Cleveland's model in which some of the observations are given a zero weight (i.e. they are excluded from the regression analysis).

The nearest neighbor model we present, much like other time series models, relies on the premise that short term predictions can be made based on past patterns of the time series. However, nearest neighbor approaches, rather than extrapolating past values into the immediate future, work by a curve fitting process that matches geometric sections of a time series to similar geometric sections that occurred in the past. The model then uses the time series values that immediately follow the identified nearest neighbors to formulate a forecast. By doing so, forecasts are realized by focusing on the pertinent area of the attractor.

This approach is philosophically very different from Box–Jenkins models which rely on correlations among lagged observations and error terms. A Box–Jenkins model can easily represent spikes in a time series that occurs, say, every fourth period. However, if the spikes occur sporadically, a Box–Jenkins model would have difficulty representing them. In contrast, a nearest neighbor model can represent such behavior because the methodology selects the relevant prior observations based on their levels and their geometric trajectories, not their locations in the time series.

We utilize the nearest neighbor methodology to represent the time component in a multivariate time series model. The nearest neighbors are identified in the same fashion as done by Sugihara and May (1990). We generalize their model to allow for the influence of causal variables by using the nearest neighbors to identify a subset of time series observations to be used in a regression analysis. The premise of our methodology is that, for a given observation in a time series, a carefully chosen subset of observations more accurately characterizes the relationship between the dependent variable and the causal variables than the entire dataset. Thus the model allows the relationships between the dependent variables and the independent variables to vary across different regions of the attractor. Importantly, this introduces the assumption that the causal relationships systematically relate to the regions of the attractor.

The regression results formulate a forecast based on the projected levels of the independent variables and the time series pattern identified with the nearest neighbor methodology. If the regression results show that the dependent variable is not significantly affected by the independent variables, the model collapses to a univariate nearest neighbor approach, much like that developed by Sugihara and May (1990).

Model description

The first step in the model is the creation of a state space library that contains the geometric sections that have occurred in the past. The model reviews the time series iteratively using one, two, and on up to ten 'embedding dimensions', E. As the number of embedding dimensions increases, the model becomes capable of picking up more complicated attractors. However, higher dimensionality requires more computing time and larger datasets. We limit our analysis to ten embedding dimensions as a reasonable trade-off between computing time and modeling power.

The model uses lagged coordinates to represent each sequence of data points as a vector, $\mathbf{y}_E$, in E-dimensional state space such that

$$\mathbf{y}_E = (y_{t-(E-1)}, y_{t-(E-2)}, \ldots, y_t) \qquad (2)$$

Thus a one-dimensional representation consists of the original time series. A two-dimensional representation consists of a series of *(1 x 2)* vectors representing each pair of adjacent time series observations. For example, consider the time series y_1, y_2, y_3, y_4. A two dimensional state space, often called a phase space, consists of the vectors $[y_1, y_2]$, $[y_2, y_3]$, $[y_3, y_4]$. A three-dimensional model consists of the vectors $[y_1, y_2, y_3]$, $[y_2, y_3, y_4]$. All ten embedding dimensions are evaluated, and the one that yields the highest predictive accuracy is used for forecasting. For each embedding dimension, the model retains the time series values of the independent and dependent variables for the next four time periods associated with that E-dimensional vector. We refer to the data for these next four time periods as the *outcome data* associated with a particular vector. Note that the outcome data are not future values; they are the historical time series values that come in the periods immediately following each vector formed by the model. The outcome data represent the historical response of the time series when a particular pattern of behavior, represented by a state space vector, is observed. We limit our forecast to four periods because the reconstructing of an attractor in state space allows forecasts to be made only within a narrow range of the attractor. Forecasts more than a few periods into the future become very inaccurate because the curve fitting process breaks down as the portion of the attractor included in the forecast region lengthens. Results of several simulations have revealed that forecasts for more than a few periods are beyond the range of the technique for most types of time series.

To demonstrate the basic structure of the nearest neighbor approach, we describe a two dimensional example and a single period forecast. Consider the following simulated time series for a single variable:

Period	$t-7$	$t-6$	$t-5$	$t-4$	$t-3$	$t-2$	$t-1$	t	$t+1$
Value	2.10	4.10	1.90	4.00	2.05	3.95	1.96	4.05	?

We want to forecast the value for period $t+1$. For a two dimensional model, the state space consists of the vectors representing each pair of adjacent points. These vectors, along with the associated outcome data, for one step ahead are

Vector	Outcome data
[2.10, 4.10]	1.90
[4.10, 1.90]	4.00
[1.90, 4.00]	2.05
[4.00, 2.05]	3.95
[2.05, 3.95]	1.96
[3.95, 1.96]	4.05
[1.96, 4.05]	?

The vectors are shown in the two-dimensional state space in Fig. 2. The last pair of time series values, 1.96 and 4.05, is represented by the star in Fig. 2. Surrounding this star are a set of points (vectors) that represent sequences of observations that exhibit patterns that resemble the last pair of time series values (i.e. 1.96 followed by 4.05). These vectors are the nearest neighbors, and here can be visually identified. The philosophy of the technique is that the time series values that historically follow these points can be used to make a forecast for period $t + 1$. The nearest neighbors are the two dimensional vectors that represent similar sequences of time series values that occurred in the past (in this case an approximate two followed by an approximate four). The outcome data for the identified nearest neighbor vectors are 1.90, 2.05 and 1.96. A forecast for period $t + 1$ is made on the basis of those outcome data. For example, we could use the mean of the outcome data (1.97) as our forecast.

While the library of E-dimensional vectors of a time series may be described as vectors in state space, we prefer to think of them as trajectories. A vector consisting of up to E lagged values of a time series represents the trajectory of an attractor, be it a point, periodic or strange attractor. While the trajectories of a chaotic attractor never recur, they often approach each other very closely. By fitting sections of curves (E-dimensional vectors) to a library of similar curves, the behavior of the attractor can be represented. When chaos exists, the trajectories in the library contain the information necessary to quantify the relationship between independent and dependent variables at a given point in time. These trajectories can account for a multitude of influences such as seasonality and long term cycles in the time series.

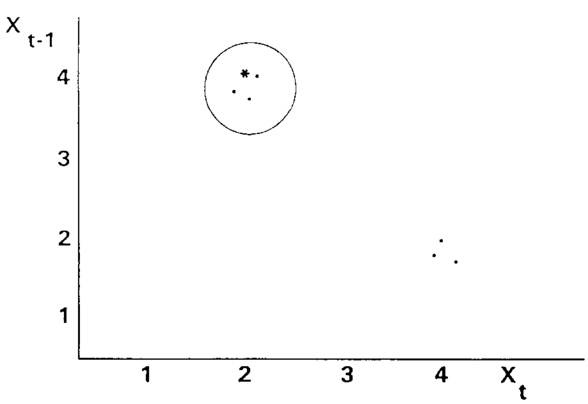

Fig. 2.

An important issue is how the nearest neighbors are identified. There are a variety of approaches for doing so. In our formulation, the model calculates the Euclidean distance between the E-dimensional vector and each vector in the state space library and retains a specified number of vectors with the smallest distances. The distance, d, between two vectors in E-dimensional space, often known as a norm, is calculated as

$$d = \left(\sum_{i=0}^{z} (y_{t-1} - p_{t-i})^2 \right)^{1/2} \tag{3}$$

where y_{t-i} is the value of the vectors of the state space library, p_{t-i} is the value of the for which a prediction is desired, and z is the number of nearest neighbors.

After the computation of the distances, the model constructs a matrix in which the rows represent the computed distances along with the values of the independent and dependent variables associated with the next four time periods (the outcome data) of each nearest neighbor. Symbolically, the matrix is as follows:

$$\begin{matrix} d_1 & (y, \boldsymbol{x})_{1,t+1} & (y, \boldsymbol{x})_{1,t+2} & \cdots & (y, \boldsymbol{x})_{1,t+4} \\ d_2 & (y, \boldsymbol{x})_{2,t+1} & (y, \boldsymbol{x})_{2,t+2} & \cdots & (y, \boldsymbol{x})_{2,t+4} \\ \vdots & \vdots & \vdots & & \vdots \\ d_z & (y, \boldsymbol{x})_{z,t+1} & (y, \boldsymbol{x})_{z,t+2} & \cdots & (y, \boldsymbol{x})_{z,t+4} \end{matrix} \tag{4}$$

where y is the dependent variable value, $\mathbf{x}$ is a vector of independent variable values and z is the number of nearest neighbors.

Sugihara and May (1990) formulate their model to retain the $E+1$ nearest neighbors. Additional vectors, beyond $E+1$, often improve the accuracy of forecasts especially for low dimensionality situations. There is no theoretical basis for restricting the model to $E+1$ vectors. However, increasing the number of vectors can compromise the nearest neighbor methodology by including sections of the attractor that are further away from the vector being analyzed.

Our formulation allows the researcher to specify the number of nearest neighbors used in the analysis. There are several factors to consider in specifying the number of nearest neighbors, z. Since we only estimate the regression model on a portion of the data, we may encounter unstable parameter estimates if the number of nearest neighbors is small. In selecting the number of nearest neighbors, one must trade-off the gain in parameter estimate stability achieved by using more observations with the loss in uniqueness of the region of the attractor brought on by using fewer nearest neighbors. The appropriate number of nearest neighbors to retain depends on the complexity of the attractor and the number of explanatory variables in the regression model.

It is preferable to have a sufficiently long time series so that all areas of the attractor appear several times in the time series. This allows the model to select nearest neighbors from many locations in the time series. When a time series is short, the number of vectors in any one area of the attractor may be relatively small. As a result, the selected nearest neighbors may drift away from the location of interest on the attractor, thereby dampening the ability of the model to represent the time series pattern.

We have found that a weighing algorithm that regulates the contribution of each vector by a function of its distance from the original vector reduces the sensitivity of the forecasts to the number of nearest neighbors. Before a predicted value is generated, each outcome data value is factored into the prediction based on a weighing algorithm. That algorithm uses a researcher supplied power factor applied to the distance between a vector and the E-dimensional vector for which a forecast is desired such that the distance, d, becomes d^p where p is the selected power factor.

We prefer this approach to that of Sugihara and May (1990) because no single exponential factor is optimal for every time series. The role of the power factor can be interpreted as a calibrating parameter that improves the performance of the model by weighing the nearest neighbors. We initiate the model by setting p to be -1 which weighs the nearest neighbors proportionally to the inverse of their distances. A negative exponent assures that the greatest weight is given to the neighbor that is closest to the vector in question. It appears that the length of the time series affects the suitability of one power factor over another. The smaller the dataset forming the library of nearest neighbors, the greater the need for weighing. Note that there is no reason for this power factor to be an integer. In fact, a major concept of chaos and fractal geometry is noninteger dimensionality. The evaluation of noninteger values for the power factor may be a fruitful topic for future research.

To incorporate the effects of explanatory variables, the model regresses the dependent variable on the causal variables using only the observations identified as the outcome data of the nearest neighbors. Doing so concentrates the analysis on the data associated with the most relevant section of the attractor. The estimated regression parameters are incorporated into the model to forecast future values. The model compares the actual time series values with the single period forecasts for all but the first 15 observations in the time series.[1] The forecasting accuracy for each dimension is assessed to determine the most appropriate dimensionality for forecasting.

Once the regression parameters are calculated, the projected independent variables can be used to forecast the levels of the dependent variable in future time periods. Forecasts for future

*[1] The first 15 observations are used to create the original library of nearest neighbors. No forecasted values are made for these observations.

values of the dependent variable are obtained by inserting projected price levels into the model. In the event that none of the independent variables are significantly related to the dependent variable, the model reduces to a univariate nearest neighbor approach to make a prediction. In that case, the model makes a prediction using a weighted average of the outcome data for the nearest neighbors. The form of the univariate equation is:

$$\hat{y}_{t+1} = (d^p_{1,t+1}/w)^* y_{1,t+1} + \dots + (d^p_z/w)^* y_{z,t+1} \quad (5)$$

where $\hat{y}_{t+1}$ is the forecast for period $t+1$, $y_1 \dots y_z$ are the outcome data for the z nearest neighbors of the dependent variable, and w is a weighing factor such that

$$w = d^p_1 + d^p_2 + \dots + d^p_z \quad (6)$$

A similar set of equations is used for the other three prediction intervals, $(y_{t+2},\ y_{t+3},\ y_{t+4})$. In theory, given sufficient data to form the library of past behavior, all types of behavioral patterns such as trend, periodic and chaotic influences can be represented by the model.

An important consideration is the data interval used in the analysis. Research in marketing has shown that time series models produce varying results depending on the data interval (Bass and Leone, 1983). With the method described here, the data interval can present a problem if it has a much lower frequency (e.g. monthly instead of weekly) than that of the time series. In the empirical demonstration that follows, we utilize weekly sales data. We have found that aggregation of the data to the monthly level loses much of the resolution in the data, and weakens the ability of the technique.

6. Empirical demonstration

In this section, we demonstrate the forecasting ability of the multivariate nearest neighbor model and compare the results to a transfer function model and a univariate nearest neighbor model. We analyze store level scanner data for weekly sales of a single brand of cake mix in a single grocery store. The data span a period of nearly 2 years (101 weeks). The data feature a few large spikes in the time series attributable to promotional price cuts, a pattern well documented in the marketing literature (e.g. Blattberg and Wisniewski, 1989). Forecasting brand sales represents an attractive application of the nearest neighbor model because brand sales are affected by both time series and explanatory factors. Further, McQuitty (1992, p. 481) notes that scanner data may be one of the most attractive types of data for modeling chaos because such data provides long time series of precise measurements on unit sales.

We calibrate the model on the first 86 observations, and use the remaining 15 observations as a holdout sample. All forecasts are one step ahead forecasts. We employ several diagnostic measures as recommended by Steece (1982). We evaluate the following diagnostic measures of forecast accuracy:

1. Correlation = Pearson product moment correlation

2. Root mean square error (RMSE) $= \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$

3. Mean absolute error (MAE) $= \frac{1}{n}\sum_{t=1}^{n} |e_t|$

4. Mean absolute percentage error (MAPE) $= \frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right|$

5. Akaike information criteria (AIC) $= \ln\sigma^2 + \frac{2k}{n}$ where σ^2 is the error variance of the model, k is the number of parameters estimated and n is the number of observations. Note that the AIC is based on the error plus a penalty for the number of parameters.

6. Theil's U $= \sqrt{\dfrac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} (y_t - y_{t-1})}}$

A U value of 1.0 indicates a forecasting accuracy equivalent to a random walk. Improvements over a random walk are indicated by U values below 1.0.

For these measures, e_t, is the one step ahead forecast error for week t, y_t is the time series value in week t and n is the number of observations. Since the nearest neighbor models do not provide one step forecasts for the first 15 observations, forecast evaluations are based on observations 16 through 86. Forecast evaluations for the holdout sample are based on the one step forecasts for observations 87 through 101.

6.1. Transfer function model

A transfer function model consists of a regression model with the appropriate autoregressive and moving average elements operating on the error term. For weekly sales of cake mix, the time series component consists of AR(1) and AR(3) coefficients. That representation is similar to time series models for brand level sales of consumer packaged goods that have appeared in other studies (Walters and MacKenzie, 1988; Mulhern and Leone, 1991). We specify a multiplicative (double logarithm) functional form where the parameter estimates are equivalent to price elasticities. The following transfer function model was estimated (standard errors in parenthesis):

$$\ln Q_{1t} = \underset{(1.6)}{9.98} - \underset{(0.34)}{1.33} \ln P_{1t} + \underset{(0.33)}{0.34} \ln P_{2t} + \underset{(0.26)}{1.96} X_{1t} + \frac{1}{1 - \underset{(0.09)}{0.41} B^1 - \underset{(0.10)}{0.30} B^3} \epsilon_{1t} \tag{7}$$

where Q_{1t} is the quantity sold in week t, P_{1t} and P_{2t} are the prices for brands 1 and 2 in week t, $X_{1t} = 1$ when brand 1 has an advertised promotion price, but is 0 otherwise, and B is the backshift operator such that $B^k \epsilon_t = \epsilon_{t-k}$.

The direct price elasticity is −1.33. The cross-price elasticity is insignificant for the one competitive brand in the model. Fig. 3 provides graphs of the actual and predicted values for the three models used to make forecasts. Table 1 shows the actual and forecasted sales for the transfer function model and the two nearest neighbor models discussed below. The exhibit shows how well each methodology predicts the promotional spikes in sales.

For the observations used to construct the transfer function model, one step ahead forecast accuracy is quite good according to the diagnostic criteria in Table 2. However, for the holdout observations, the forecasting accuracy is poor. Theil's U coefficient is above 1.0, indicating that the forecasts are no better than a random walk. As can be seen in Fig. 3(b), the model does not precisely represent all of the promotional spikes. This could be a result of the promotional dummy coefficient representing an average effect across several promotional periods. Forecasting accuracy might be improved by including additional independent variables to represent seasonality, holidays, promotional versus non-promotional prices and so forth. While that is standard econometric practice, it can get quite cumbersome; and it cannot account for chaotic behavior when it is present. An alternative approach, which we use in the multivariate nearest neighbor model, is to do the regression analysis only on time series observations selected because of their location on the attractors in the time series.

6.2. Univariate nearest neighbor model

In this section, we apply a methodology similar to that developed by Sugihara and May (1990) to determine the ability of a univariate nearest neighbor model to model brand level sales. This model is a special case of the multivariate model described above when explanatory variables are excluded. For the brand we analyze, the nearest neighbor model that produces the most accurate predictions has six embedding dimensions and a power factor, p, of −2. A plot of the predicted weekly sales levels is shown on Fig. 3(c). The correlation between actual and predicted values is 0.25 for the model observations and 0.17 for the holdout sample. This model provides the poorest forecasting performance for both the model observations and the holdout sample

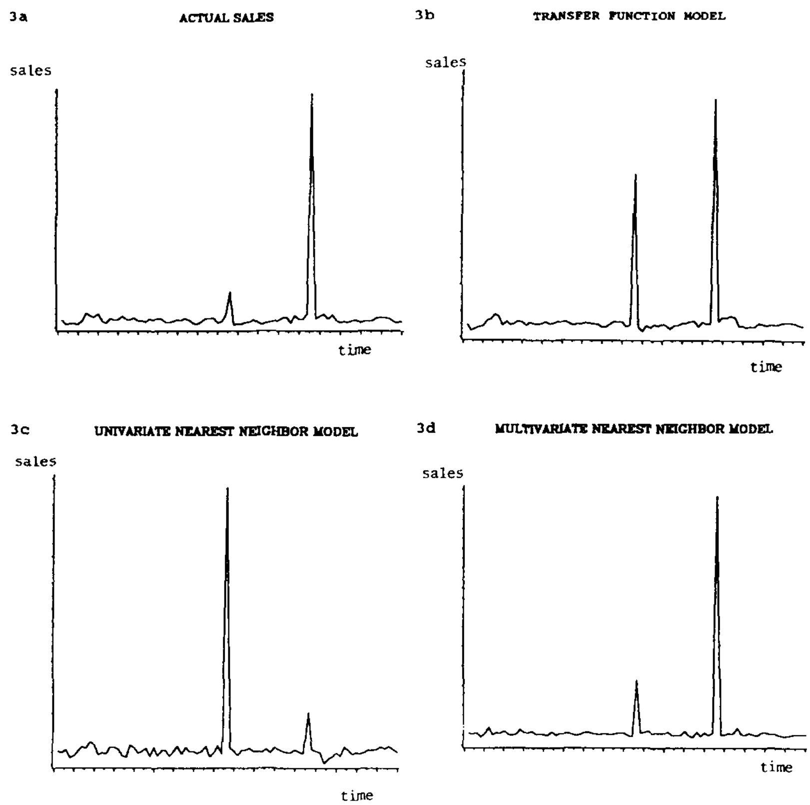

Fig. 3.

according to all but one of the diagnostic measures. Such poor predictive accuracy reflects the inability of the univariate model to account for the effects of the price discounts in two ways.

(1) Since the state space methodology is univariate, it is unable to represent the spikes in sales volume attributable to price discounts. The model attempts to incorporate the spikes into the time series framework. Even if such a representation could be achieved, the model would be incorrect because the huge jumps in sales are not a time dependent factor.

(2) A second problem lies in the construction of the library of state space vectors. When the time series is broken into E-dimensional vectors, every spike in the time series appears as outcome data for at least one vector in the library. If that vector is chosen as a nearest neighbor, its value contributes to the prediction whether or not price is discounted for that week. This explains why spikes appear in Fig. 3(c) which do not correspond to price discounts.

6.3. Multivariate nearest neighbor model

Finally, we generate forecasts with the multivariate nearest neighbor model. That model consists of performing the following regression analysis on the selected nearest neighbor outcome data:

$$\ln Q_{1j} = \alpha + \beta_1 \ln P_{1j} + \beta_2 \ln P_{2j} + \gamma X_{1j} + \epsilon_j \tag{8}$$

where Q_{1j} is the quantity sold of brand 1 in week j, P_{ij} is the price of brand i in week j, X_{1j} is the promotional dummy variable described for Eq. (5), α, β_1, β_2 and γ are parameters to be estimated, and $j = 1$ through z nearest neighbors.

Eq. (5) is equivalent to the regression component of the transfer function model (Eq. (7))

Table 1
Comparison of one step ahead forecasts

Week	Actual Sales	Transfer Function	Univariate N.N. Model	Multivariate N.N. Model
1	23	45.11		
2	32	31.89		
3	42	35.37		
4	12	35.83		
5	19	23.77		
6	29	31.07		
7	36	25.37		
8	41	31.78		
9	27	38.02		
10	30	34.23		
11	34	37.14		
12	642	561.27		
13	42	37.81		
14	45	40.27		
15	50	43.30		
16	56	46.04	52.59	59.42
17	22	49.22	47.19	54.64
18	31	30.29	60.11	58.01
19	35	38.08	26.18	37.68
20	26	30.93	35.41	54.72
21	54	53.76	56.93	87.32
22	113	60.27	71.87	47.37
23	90	73.95	65.03	60.79
24	82	71.40	90.02	56.83
25	105	50.02	79.14	67.98
26	49	64.32	37.00	52.61
27	34	45.89	48.34	47.95
28	71	63.74	43.48	53.28
29	59	62.04	70.08	77.45
30	60	51.59	68.40	57.71
31	89	47.27	60.69	61.59
32	66	57.92	35.91	65.94
33	59	51.56	30.42	51.78
34	74	59.58	78.34	53.54
35	49	59.74	67.44	53.22
36	55	63.08	47.66	47.43
37	41	55.44	51.86	65.20
38	66	52.96	66.01	49.12
39	60	58.22	31.14	58.46
40	66	53.49	71.10	57.08
41	65	61.48	34.53	59.62
42	42	51.63	56.28	54.16
43	38	47.09	56.77	48.88
44	48	51.77	33.40	51.09
45	60	49.21	53.24	60.50
46	55	52.29	73.08	65.18
47	69	51.91	40.45	58.09
48	56	52.90	68.99	54.80
49	33	50.12	51.35	52.33

Table 1
Continued

Week	Actual Sales	Transfer Function	Univariate N.N. Model	Multivariate N.N. Model
57	95	49.05	41.85	41.86
58	272	603.11	1164.96	329.66
59	18	34.80	66.56	58.83
60	28	38.80	55.09	63.67
61	25	30.87	36.61	70.36
62	37	26.96	57.87	54.01
63	42	41.51	58.65	59.82
64	60	39.91	64.10	59.39
65	39	51.87	52.18	57.72
66	21	43.37	55.22	62.29
67	36	32.62	49.01	45.41
68	41	37.79	60.78	47.90
69	48	33.11	46.77	49.09
70	53	49.74	67.14	68.41
71	73	50.86	61.18	47.92
72	79	60.72	52.53	37.66
73	35	61.94	49.62	65.14
74	92	42.57	46.69	42.53
75	64	68.34	58.96	61.19
76	63	46.22	52.29	49.10
77	108	63.94	43.77	58.47
78	1781	1188.94	201.51	1264.91
79	67	70.80	61.78	51.49
80	84	105.04	54.79	50.80
81	100	96.67	49.67	57.66
82	66	90.22	6.21	56.84
83	96	71.84	22.22	86.42
84	56	46.24	31.73	53.40
85	37	42.63	49.04	50.86
86	48	40.30	34.22	58.77
Holdout Observation Begin				
87	41	46.23	71.82	52.90
88	52	38.31	59.23	45.88
89	58	53.05	40.84	60.43
90	35	49.75	49.65	35.13
91	50	43.51	45.57	52.98
92	48	49.67	50.52	53.65
93	51	42.00	59.74	51.31
94	71	47.89	57.28	65.71
95	75	54.12	61.75	72.51
96	76	56.35	71.13	76.66
97	69	62.55	61.90	71.19
98	47	61.13	71.43	52.87
99	35	52.51	60.30	51.67
100	48	45.26	50.21	52.03
101	58	45.87	51.39	49.77

Table 2
Measures of forecast accuracy

	Transfer function	Univariate nearest neighbor	Multivariate nearest neighbor
Model observation (16–86)[a]			
Correlation	0.94	0.25	0.98
RMSE	82.71	211.30	66.02
MAE	26.92	53.66	26.33
MAPE	0.28	0.46	0.40
AIC	8.97	10.71	8.38
Theil's U	0.29	0.73	0.23
Holdout observation (87–101)			
Correlation	0.32	0.17	0.87
RMSE	13.27	14.91	6.64
MAE	11.49	12.20	5.00
MAPE	0.22	0.25	0.11
AIC	5.84	5.40	4.19
Theil's U	1.07	1.21	0.54

[a] Since the nearest neighbor models do not generate forecasts for the first 15 observations, diagnostic measures are based on the forecasts for observations 16 through 86.

except that Eq. (8) is applied to the outcome data for the $j = 1 \ldots z$ nearest neighbors rather than the $t = 1 \ldots n$ observations.

As with the univariate nearest neighbor model, the most accurate forecasts were made using six embedded dimensions and a power factor of -2. The predicted values of this analysis are presented in Fig. 3(d). For nearly all forecast evaluations, the multivariate nearest neighbor model provides the most accurate forecasts. The correlations between actual and predicted values for the multivariate nearest neighbor model are highest for both the model observations and the holdout sample. Similarly, all error measures for the holdout forecasts are lower for this technique than for the other two techniques. In particular, note that the Theil's U measure for the holdout sample is 0.54 which represents a substantial improvement over a random walk. In contrast, the Theil's U for the holdout sample forecasts from the other techniques does not indicate an improvement over a random walk.

The incorporation of the effects of the independent variables in the model substantially improves the ability of the model to represent the attractor in the time series. That is, the inclusion of a causal variable allows the model to account for changes in the attractor due to differences in price and promotional activity. In a sense the attractor is 'deformed' by the influences of the explanatory variables. This methodology attempts to compensate for these deformations by representing the causal effects of prices while maintaining the integrity of the attractor.

Since the model retains outcome data for the four time series values that follow each nearest neighbor, we are able to make one, two, three and four step ahead forecasts. For each time series observation (weeks 16 through 101), we generate forecasts for the following four periods based on the model constructed for weeks 1 through 86. Table 3 shows the actual and forecasted values for the one, two, three and four step ahead forecasts with the multivariate nearest neighbor model. As indicated by the diagnostic measures, the forecasting accuracy breaks down after one period into the future. That is consistent with the philosophy of nearest neighbor approaches, and results from the vectors in the state space diverging from each other as longer forecasts are made. Nearest neighbor models are intended to represent complex time series patterns. In order to do so, the ability to make long term forecasts is sacrificed.

Table 3
Multivariate nearest neighbor model forecast accuracy (all forecast observations (16–101))

	One step forecasts	Two step forecasts	Three step forecasts	Four step forecasts
Correlation	0.99	0.29	0.27	0.06
RMSE	60.06	183.59	184.07	189.44
MAE	22.62	42.77	41.60	39.84
MAPE	0.35	0.45	0.41	0.38
AIC	8.23	10.46	10.47	10.52
Theil's U	0.23	0.70	0.70	0.72

7. Discussion

The multivariate nearest neighbor model provides an alternative to a transfer function model for jointly modeling both time series and causal factors. An important advantage of this approach is the capability of representing chaotic time series patterns which cannot be picked up by traditional techniques. Questions regarding the pervasiveness of chaos in marketing time series remain. A variety of techniques have been developed to test for the presence of chaotic behavior (Wolf et al., 1985). Most of these techniques require large amounts of data and are difficult to implement with datasets of the size typically found in marketing. While it is desirable to have a means of uncovering the presence of chaos, it is more important to have a methodology, such as the model we introduce, that can represent chaos along with other types of behavior. An additional advantage is that the model can be implemented without prespecifying a time series formulation as required in Box–Jenkins analysis.

The model works by developing a library of state space vectors, and making a prediction for each vector in the library based on the nearest neighbor vectors identified in the library. The model is calibrated to identify the embedding dimension, weighing power factor, and the number of nearest neighbors which produce the most accurate predictions. While the nearest neighbor portion of the model is somewhat complicated, it can be efficiently programmed in a matrix language such as SAS/IML. We envision the model being used by managers as a profit maximizing tool to forecast sales on a real time basis. For consumer packaged goods, this could be accomplished by integrating the model into an expert system that accesses scanner data and generates forecasts under various scenarios. The model would periodically assemble the space state library based on historical data. It would then take the most recent E intervals of data to form a prediction vector, locate the nearest neighbors, and compute causal relationships to make a short term forecast. A marketing manager could use this method to forecast sales in future time periods under several different pricing scenarios.

Our intention has been to introduce a new methodology for time series modeling which accounts for the time effects often missed by traditional techniques. This is a more general solution to the problem of modeling chaotic behavior than applying one of the equations capable of exhibiting chaos. The method allows for chaotic behavior in a time series but does not assume the presence of chaos, nor does it force a specific functional form on the data. We expect that the predictive capabilities of the model can be improved with additional research. An important issue for further study is the influence of data transformations such as differencing and principal components analysis on forecasting ability.

The methodology also offers an alternative to traditional modeling procedures which do not fully exploit the information available in time series data. Time, and time related factors, are typically treated as a nuisance. In a transfer function model, the time effects are filtered out with an ARIMA specification, rather than being integrated into the causal relationships. The multivariate nearest neighbor model fully incorporated the information on temporal order into the analysis rather than removing it. Technological developments in retailing and other areas are making available large datasets of precise measurements of consumer behavior over time, best exemplified by UPC scanner data. Such time series data are superior to the snapshot measures provided by cross-sectional data sources for inferring causal relationships between marketing mix variables and dependent measures of performance. Future research should develop additional robust models to take advantage of the high quality data now available for marketing analysis.

Acknowledgments

The authors express gratitude to Jeff Inman, Gary Lilien and Robert Leone for helpful comments on an earlier version of this paper.

References

Bass, F.M., 1969, A new product growth model for consumer durables, *Management Science*, 15, 215–227.

Bass, F.M. and R.P. Leone, 1983, Temporal aggregation and data interval bias, and empirical estimation of bimonthly relations from annual data, *Management Science*, 31, 1–11.

Baumol, W.J. and J. Benhabib, 1989, Chaos: significance, mechanism, and economic applications, *Journal of Economic Perspectives*, 3, 77–106.

Benhabib, J. and R.H. Day, 1981, Rational choice and erratic behavior, *Review of Economic Studies*, 48, 459–472.

Blattberg, R.A., and K.J. Wisniewski, 1989, Price-induced patterns of competition, *Marketing Science*, 8, 291–309.

Box, G.E.P. and G.M. Jenkins, 1976, *Time Series Analysis: Forecasting and Control* (Holden-Day, Inc., San Francisco).

Brockwell, P.J. and R.A. Davis, 1991, *Time Series: Theory and Methods* (Springer, New York).

Cleveland, W.S. 1979, Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, 74, 829–836.

Donthu, N. and R.T. Rust, 1986, Patterns of dynamic disequilibrium: A new paradigm for marketing thought, in *Proceedings of the Summer Educator's Conference*, American Marketing Association.

Farmer, J.D. and J.J. Sidorowich, 1988, Can new approaches to nonlinear modeling improve economic forecasts, in P.W. Anderson, K.J. Arrow and D. Pines, editors, *The Economy as an Evolving Complex System* (Addison-Wesley, New York), pp. 99–115.

Frank, M. and T. Stegnos, 1989, Measuring the stangeness of gold and silver rates of return, *Review of Economic Studies*, 56, 553–568.

Gleick, J., 1987, *Chaos: Making a New Science* (Viking Press, New York).

Hanssens, D.M., 1980, Market response, competitive behavior, and time series analysis, *Journal of Marketing Research*, 17, 470–485.

Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, New York).

Henson, S.W. and E.E. Rigdon, 1992, Chaos theory: A demonstration, recent developments, and implications for marketing scientists, in C.T. Allen and T.J. Madden, editors, *Marketing Theory and Applications*, AMA Winter Educators' Proceedings, pp. 484–494.

Lambkin, M. and G.S. Day, 1989, Evolutionary processes in competitive markets: beyond the product life cycle, *Journal of Marketing*, 53, 4–20.

Leone, R.P., 1983, Modeling sales-advertising relationships: an integrated time series-econometric approach, *Journal of Marketing Research*, 20, 291–296.

Lorenz, E.N., 1963, Deterministic non-periodic flow, *Journal of Atmospheric Science*, 20, 130–141.

May, R.M., 1976, Simple mathematical models with very complicated dynamics, *Nature*, 261 (June 10), 459–467.

Miles, J., 1984, Strange attractors in fluid dynamics, *Advances in Applied Mechanics*, 24, 189–213.

McQuitty, S., 1992, An examination of chaos theory and its relation to marketing, in C.T. Allen and T.J. Madden, editors, *Marketing Theory and Applications*, AMA Winter Educators' Proceedings, pp. 474–483.

Mirowski, P., 1990, From Mandelbrot to chaos in economic theory, *Southern Economic Journal*, 57, 289–307.

Moriarty, M.M. and A.J. Adams, 1984, Managerial judgments forecasts, composite forecasting models, and conditional efficiency, *Journal of Marketing Research*, 21, 239–250.

Mulhern, F.J. and R.P. Leone, 1991, Implicit price bundling of retail products: A multiproduct approach to maximizing store profitability, *Journal of Marketing*, 55, 63–76.

Packard, N.H., J.P. Crutchfield, J.D. Farmer and R.S. Shaw, 1980, Geometry from a time series, *Physical Letters Review*, 45, 712–716.

Pollock, R.S.G., 1987, The methods of time series analysis, *Interdisciplinary Science Reviews*, 12, 128–135.

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

Priestley, M.B., 1988, Current developments in time series modeling, *Journal of Econometrics*, 37, 67–86.

Richards, D., 1990, Is strategic decision making chaotic?, *Behavioral Science*, 35, 219–232.

Rust, R., 1988, Flexible regression, *Journal of Marketing Research*, 25, 10–25.

Rust, R. and E.O. Bornman, 1982, Distribution-free methods of approximating nonlinear relationships, *Journal of Marketing Research*, 19, 372–374.

Skarda, C.A. and W.J. Freeman, 1987, How brains make chaos in order to make sense of the world, *Behavioral and Brain Sciences*, 10, 161–195.

Steece, B.M., 1982, Evaluations of forecasts, in S. Makridakis and S.C. Wheelright, eds., *The Handbook of Forecasting: A Manager's Guide* (Wiley, New York), 457–468.

Stewart, I., 1989, *Does God Play Dice? The Mathematics of Chaos* (Basil Blackwell, Cambridge, MA).

Stone, C.J., 1977, Consistent nonparametric regression, *Annals of Statistics*, 5, 595–620.

Sugihara, G. and R.M. May, 1990, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, 344, 734–741.

Walters, R.G. and S.B. MacKenzie, 1988, A structural equations analysis of the impact of price promotions on store performance, *Journal of Marketing Research*, 25, 51–63.

Wolf, A., J.B. Swift, H.L. Swinney and J.A. Vastano, 1985, Determining lyapunov exponents from a time series, *Physica*, 16D, 285–317.

Biographies: Francis J. MULHERN is Assistant Professor of Marketing at the Pennsylvania State University. His research focuses on new methodologies for analyzing large databases in marketing. Much of his research measuring causal relationships between marketing variables and retail sales as well as forecasting retail sales. Research on these areas has appeared in the *Journal of Marketing*, the *Journal of Business Research* and the *Journal of the Academy of Marketing Science*.

Robert J. CAPRARA is a Principal Consultant for Raptor Consulting Group in West Chester, PA. He specializes in the development of computer algorithms for Geographic Information Systems, neural networks and forecasting models. He has developed models in several application areas including environmental assessments, marketing models and financial consulting.