# Naive Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of "Technical Analysis"

Salih N. Neftci

*Journal of Business*, Volume 64, Issue 4 (Oct., 1991), 549-571.

**Salih N. Neftci***
*Graduate School, City University of New York*

# Naive Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of "Technical Analysis"

This article attempts a formal study of technical analysis, which is a class of informal prediction rules, often preferred to Wiener-Kolmogorov prediction theory by participants of financial markets. Yet Wiener-Kolmogorov prediction theory provides optimal linear forecasts. This article investigates two issues that may explain this contradiction. First, the article attempts to devise formal algorithms to represent various forms of technical analysis in order to see if these rules are well defined. Second, the article discusses under which conditions (if any) technical analysis might capture those properties of stock prices left unexploited by linear models of Wiener-Kolmogorov theory.

The attention that technical analysis receives from financial markets is somewhat of a puzzle. According to Wiener-Kolmogorov prediction theory, time-varying vector autoregressions (VARs) should yield best forecasts of a stochastic process in the mean square error (MSE) sense. Yet, quasitotality of traders use technical analysis in day to day forecasting although it bears no direct relationship to Wiener-Kolmogorov prediction theory. In fact, technical analysis is a broad class of prediction rules with unknown statistical properties, developed by practitioners without reference to any formalism.

This article investigates statistical properties of technical analysis in order to determine if there is any objective basis to the popularity of its methods. Broadly, there are two issues of interest. First, can one devise formal algorithms that can generate buy and sell signals identical to the ones given by technical analysis—that is, are any of these rules (mathematically) well defined? The second issue is to what extent well-defined rules of technical analysis are useful in prediction over and above the forecasts generated by Wiener-Kolmogorov prediction theory.

* The author is professor of economics at the Graduate School and University Center of the City University of New York. I would like to thank Chris Sims for his comments. I also would like to thank Jonathan Sampson.

(*Journal of Business*, 1991, vol. 64, no. 4)

0021-9398/91/6404-0005$01.50

Normally, just the resources spent on using and developing new forms of technical analysis should provide sufficient motivation for this article. However, a series of interesting papers makes such a study more relevant. For example, Brockett, Hinich, and Patterson (1985) and Hinich and Patterson (1985) have argued that several time series, among them asset prices, are stochastically nonlinear. Thus any method that can capture the nonlinearity of asset prices can potentially improve forecasts generated by the Wiener-Kolmogorov prediction theory. For example, Wiener-Kolmogorov theory will not utilize the information contained in higher-order moments of nonlinear processes. It is possible that, in developing technical analysis, practitioners have informally attempted to use the information contained in higher-order moments of asset prices. In fact, it appears that since the October 19, 1987, crash of financial markets, traders have shown more interest in technical analysis—possibly because a crash of that magnitude is a nonlinear event, and the framework provided by the Wiener-Kolmogorov theory would fail to handle it properly.

In particular, linear models are incapable of describing at least two types of plausible stock market activity that are of interest to participants in financial markets. First is the problem of how to issue sporadic buy and sell signals. By nature, this problem is nonlinear. The decision maker observes some indicators, and at random moments, issues signals. VARs cannot explicitly generate such signals. The second example involves "patterns" that may exist in observed time series. Linear models such as VARs can handle these patterns only if they can be fully characterized by the first- and second-order moments. This basically involves any pattern with smooth curvatures. A speculative bubble, which generates a smooth trend and then ends in a sudden crash, cannot be handled easily by linear models.

This article shows that most patterns used by technical analysts need to be characterized by appropriate sequences of local minima and/or maxima and will lead to nonlinear prediction problems. It is well known that the theory of the minima and maxima of stochastic processes can be very tedious (Leadbetter, Lindgren, and Rootzen 1983). Under these circumstances, technical analysis may serve as a practical way of using the information contained in such statistics. At the least, this is a possibility that needs to be investigated.[1]

To the best of my knowledge, there is no formal study of the pre-

1. The popularity of technical analysis admits a second explanation. If markets are efficient, asset prices would behave (approximately) as Martingales. Then, VARs would yield trivial-looking forecasts, such as $\{X_{t+\tau} = X_t, \tau = 1, 2 \ldots\}$. Finding it unattractive to report such forecasts that remain constant over the forecasting horizon, traders might use (irrationally) techniques that give them nontrivial-looking forecasts, even though they are suboptimal. This interpretation requires that financial markets continue to allocate significant resources on a practice that has negative returns.

dictive power of technical analysis. Existing studies are mostly directed toward practical applications, informal treatments of which Pring (1980) is a good example. One of the first illustrations of technical analysis is the discussion of Dow theory in Rhea (1932). Although not directly related to any form of technical analysis, the survey by Tong (1983) and the pioneering work of Granger and Andersen (1978) provide some of the tools used here.[2]

The article is organized as follows. First, I discuss some reasons behind conducting such a study. In the next section I introduce the notion of Markov times and show that a rule of technical analysis has to generate Markov times in order to be well defined. I then discuss results that can help in deciding whether a rule generates Markov times or not. I show under what conditions well-defined forms of technical analysis can be useful over and above the Wiener-Kolmogorov prediction theory. Finally, I provide examples using the Dow-Jones industrials from 1792 to 1976.

## Can Technical Analysis Be Formalized?

Pring (1980) introduces "technical analysis" and related methods as follows:

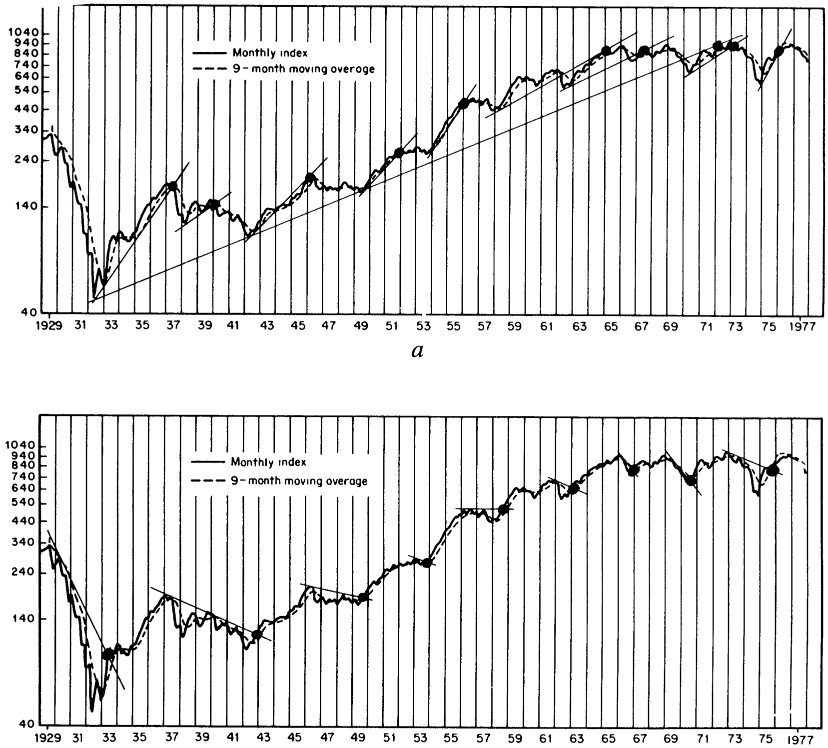
> The technical approach to investment is essentially a reflection of the idea that the stock market moves in trends which are determined by changing attitudes of investors to a variety of economic, monetary, political and psychological forces. The art of technical analysis, for it is an art, is to identify changes in such trends at an early stage and to maintain an investment posture until a reversal of that trend is indicated. . . . By studying the nature of previous market turning points, it is possible to develop some characteristics which can help identify major market tops and bottoms. Technical analysis is therefore based on the assumption that people will continue to make the same mistakes that they made in the past.[3]

Clearly, technical analysis covers a broad category of highly subjective forecasting rules. To simplify the discussion, I first adopt a preliminary classification. A survey of the literature suggests three major classes to group various forms of technical analysis.

Letting $\{X_t, t = 0, 1, \ldots\}$ represent asset prices, the first class of rules issues signals of market turning points using level crossings of the $X_t$ process. The level is almost always defined using various local

2. A recent example to the popularity of technical analysis is the following. "Starting today The *New York Times* will publish a comprehensive three-column market chart every Saturday. . . . History has shown that when the S&P index rises decisively above its (moving) average the market is likely to continue on an upward trend. When it is below the average that is a bearish signal." [*New York Times*, March 11, 1988]

3. Pring (1980), p. 2.




*a*

*b*

FIG. 1.—The trend crossing method. *a*, Long-term Dow showing important bull-market trend lines. *b*, Long-term Dow showing important bear-market trend lines. Source: Pring (1980).

maxima or minima of $\{X_t\}$. It is the choice of the level that differentiates one rule from another. Figures 1*a* and 1*b* illustrate two examples. The bull (bear) markets are signaled as the Dow-Jones industrials cross trend lines determined by appropriate local maxima (minima). We label this class of rules the trend crossing method.

Figure 2 displays a second major category labeled moving average method. Various moving averages of an observed series are obtained and the intersections of these averages are interpreted as buy and sell signals.

The third group consists of various patterns, whose occurrence is claimed to signal particular types of future behavior by $\{X_t\}$. Some of these patterns are shown in figure 3. This article argues that, in principle, all these patterns can be fully characterized using appropriate local minima and maxima. Hence, any pattern can potentially be formalized. However, I show that formal identification of local minima and maxima that can accomplish this is likely to be quite tedious.

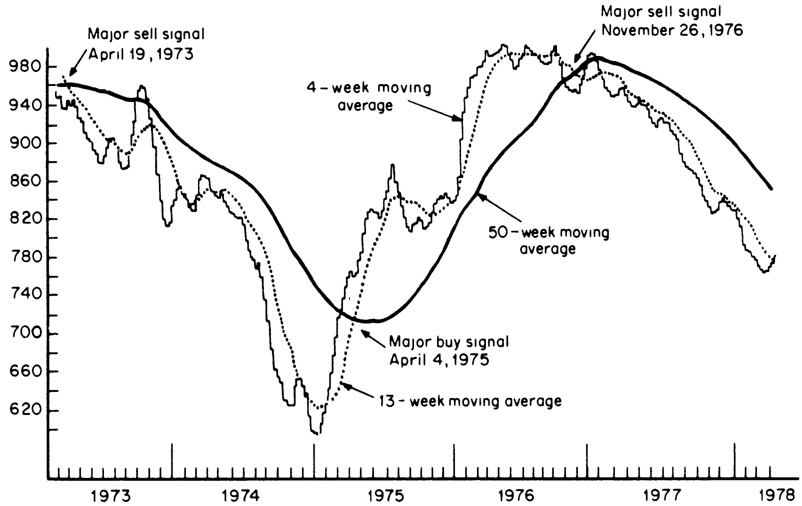



FIG. 2.—The moving average method, Dow-Jones industrials, 1973–79. Source: Pring (1980). Reprinted by permission of Dow Theory, Inc.

Thus, the first step of the analysis is to quantify and formalize, whenever possible, these three categories of technical analysis. I proceed in two stages. First, I prove that any method that relates to crossings of moving averages constitutes a well-defined prediction methodology. Second, I show that patterns or trend crossings used in obtaining market signals are almost always related to some sequences of local minima and maxima, and, more important, are generally ill defined in their current formulation. I discuss these points using the important notion of Markov times. In fact, one contribution this article makes is to recognize the importance of Markov times as a tool to pick well-defined rules for issuing signals at market turning points.

## Markov Times

Let $\{X_t\}$ be an asset price observed by decision makers. Let $\{I_t\}$ be the sequence of information sets (sigma-algebras) generated by the $X_t$ and possibly by other data observed up to time $t$.

DEFINITION. We say that a random variable $\tau$ is a Markov time if the event

$$A_t = \{\tau < t\}$$

is $I_t$-measurable—that is, whether or not $\tau$ is less than $t$ can be decided given $I_t$. According to this definition, Markov times are random time periods, the value of which can be determined by looking at the current information set. Thus, Markov times cannot depend on future information. In order to see the distinction between Markov times and non-

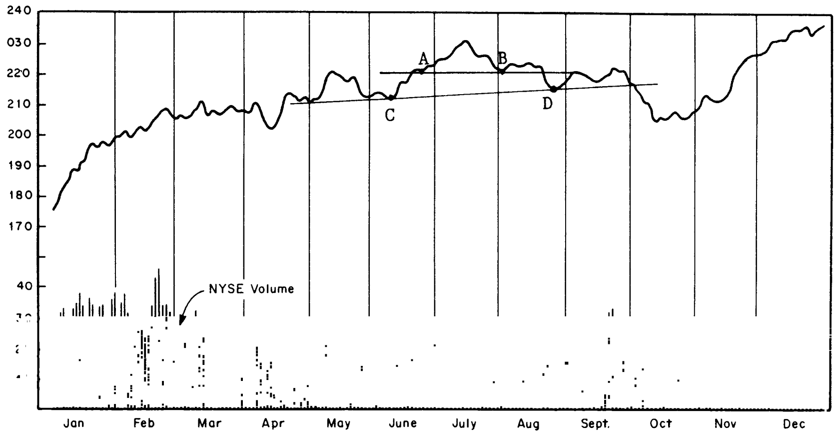



FIG. 3$a$.—Head and shoulders Dow Jones transportation average, 1976

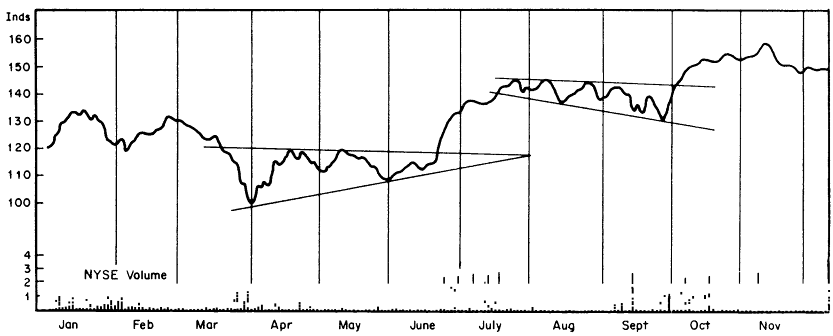



FIG. 3$b$.—Triangles, Dow Jones industrial average, 1938. Source: Pring (1980).

Markov times better, and to emphasize the importance of this concept in studying methods of technical analysis, two examples are discussed.

*Example 1.* Let $\tau_1$ denote the date at which a process $\{X_t\}$, observed continuously, shows a 10% jump for the first time during $t \in [0, \infty)$:

$$\tau_1 = \inf_t \{t \in [0, \infty) : d(\ln X_t)/dt > .1\}. \qquad (1)$$

Then $\tau_1$ is a Markov time since, by looking at the current information set, it is possible to tell whether such a jump in $X_t$ has occurred or not.

*Example 2.* Let $\tau_2$ denote the beginning date of a business cycle or a stock market uptrend. Then, $\tau_2$ is not a Markov time since, in order to know whether $\tau_2 = t$, one needs to have access to $I_{t+s}$, $s > 0$. In fact, suppose one is at time $t$ and that an uptrend started at time $\tau_2 =$

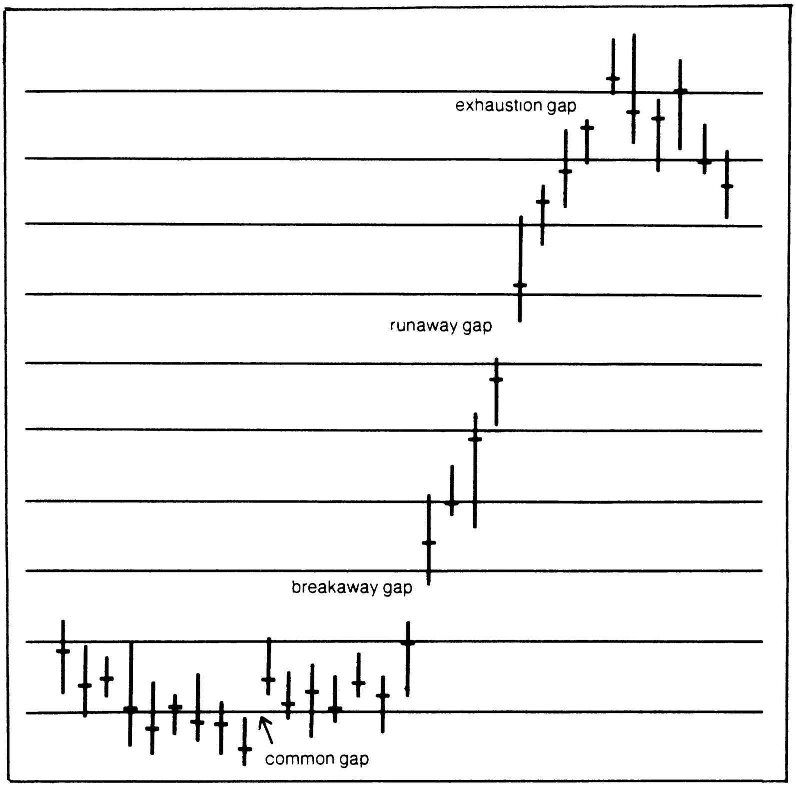



FIG. 3*c*.—Various gaps

$t - 2$. In general, one has to wait more than 2 months to be sure that an upturn is under way. Thus, one needs $I_{t+s}$, 3, 4, . . . before one knows $\{\tau_2 < t\}$—that is, future information is needed before deciding which value $\tau_2$ has assumed.

Clearly, any well-defined technical analysis rule has to pass the test of being a Markov time since any buy or sell signal should, in principle, be an announcement based on data available at time $t$. If a rule generates a sequence of buy and sell orders that fail to be Markov times, then the procedure would be using future information in order to issue such signals. The procedure would anticipate the future. This implies a signaling decision based on considerations that are not part of the available information at time $t$. These are often the subjective feelings of the forecaster or information not available to the general public.

It is surprising that such infeasible rules of technical analysis may look perfectly reasonable when illustrated on a chart displaying past data. In using charts, an investigator may implicitly use "future" information while defining a procedure. For example, note that, on a chart displaying observed data, the beginning dates of any uptrend can easily be identified, yet these dates are not Markov times as example 2 demonstrates. Graphic methods are not the best ways of determining classes of Markov times that are useful in prediction. Yet, more often than not, this is how technical analysis rules are defined. Hence the importance of developing formal algorithms that can duplicate the buy and sell signals given by technicians.

This discussion suggests that any method that exploits the current inflection point of a series will fail to generate Markov times since these latter are not $I_t$-measurable. At the same time, several popular forms of technical analysis use past local maxima (minima) and these are $I_t$-measurable.

We now have a criterion to determine which rules of technical analysis can be quantified. Indeed, if one can show that signals generated by a rule of technical analysis are Markov times, then this would simultaneously imply (1) that the method can be quantified, (2) that it is feasible, and (3) that one can investigate its predictive power using formal statistical models.

The following theorem is important in sorting out Markov times.

THEOREM. Let $\{X_t\}$ be a random process assuming values on the real line $R$. Let $B$ be the set of all intervals belonging to $R$, and $I_t$ be the information set at time $t$. Then the times $\{\tau_A^s\}$,

$$\tau_A^s = \inf_t \{t < s : X_t \in A, A \in B\}, \tag{2}$$

are Markov times (Shiryayev 1985).

Basically, this theorem states that the first entry of $X_t$ in an interval $A$ is always a Markov time. The interval in question can, for example, be $[0, \infty)$ or $(-\infty, 0]$; but it can also depend on the $I_t$ itself since, if $X_t \in A_t$, we can define $Y_t = X_t - f(A_t)$ such that $Y_t \in [0, \infty]$, as long as $A_t$ is $I_t$ measurable. For example, suppose that a forecaster intends to issue a sell signal as soon as observed price $X_t$ crosses, from above, a trend line $f(I_t, t)$:

$$f(I_t, t) = a_t t + b_t,$$

where $a_t$ and $b_t$ are $I_t$-measurable slope and intercepts of the trend line. Then a signal is issued at:

$$\tau = \inf_t \{t : X_t < f(I_t, t)\}.$$

This signal deals with the first entry of $X_t$ in a time-dependent set $A_t = [0, f(I_t, t)]$. The time dependence of $A_t$ can easily be eliminated by redefining

$$Y_t = X_t - f(I_t, t)$$

and issuing a signal at the first entry on $Y_t$ in $(-\infty, 0]$:

$$\tau = \inf_t \{t : Y_t < 0\}.$$

Hence the above theorem can be applied to first entries of $X_t$ in $I_t$-measurable sets as well.

Also, the fact that the theorem deals with the first entry is not a real restriction. The same theorem can be proven for $n$th entry of $X_t$ into $A$. The important restriction is that $n$ be known a priori by the forecaster. In fact, I intend to show below that most methods of technical analysis are ill defined precisely because they do not set this parameter $n$ a priori. Below is shown which of the broad forms of technical analysis can be formulated as Markov times.

### *Characterizing Moving Average Crossings*

Figure 2 illustrated an example of how moving average crossings are used to signal turning points. To formalize these moving average crossings I first define

$$Z_t = \left[(1/n)\sum_{s=0}^{n-1} X_{t-s}\right] - \left[(1/m)\sum_{s=0}^{m-1} X_{t-s}\right]. \tag{3}$$

The "moving average" rule of technical analysis then uses sign changes in $Z_t$ to generate the times $\{\tau_i\}$ sequentially as

$$\tau_i = \inf_t \{t : t > \tau_{i-1}, Z_t Z_{t-1} < 0\}, \tag{4}$$

with $\tau_0$ defined as zero.

Now consider what (3) and (4) say in words. I basically calculate two moving averages of the $X_t$ process. Assuming that $n > m$, the first moving average will be smoother than the second one in the sense of having relatively more power at low frequencies. Then, as soon as $Z_t$, $\tau_{i-1} < t$ changes sign, the rule in (4) will assign the value of $t$ to $\tau_i$. These latter are signals of major market downturns and upturns according to the moving average method of technical analysis.

I now show that the $\{\tau_i\}$ are Markov times, and that they constitute a well-defined method of prediction. Clearly, the product $Z_{t-1}Z_t$ is measurable with respect to $I_t$—that is, given $I_t$, the value of $Z_{t-1}Z_t$ is known. The $\tau_i$ are then defined as the first entry of $Z_{t-1}Z_t$ in the interval $(-\infty, 0) \in R$. Thus the $\tau_i$ are Markov times according to the theorem above. This makes the moving average method a statistically well-defined procedure. We should, in principle, be able to evaluate the contribution of the $\{\tau_i\}$ in predicting market turning points using formal tools.

*Characterizing Trend Crossings*

Methods that use crossings of observed data with trend lines, defined in a variety of ways, constitute the most common form of technical analysis. In contrast to the moving average method, it is not possible to determine a unique definition that would encompass all trend crossing rules. The notion of a moving average immediately suggests a mathematical formulation, whereas trend crossings appear to be based on arbitrary hand-drawn trends in charts illustrating historical data. Figure 1 displays an example. The main idea behind trend crossing methods is to determine two linear trends, one above, the other below, that would envelop the portion of the data observed since the last turning point. Then, upcrossings (downcrossings) of the upper (lower) envelope are taken as signals of market strength (weakness).

It is clear that all trend lines that envelop observed data can be defined by using only two extrema of the portion of the series under consideration. In order to obtain an upper envelope, the two highest local maxima are used. Two lowest local minima define, similarly, a lower envelope. Thus, the theory of local minima (maxima) of time series will play an important role in investigating this type of technical analysis.

I first show that most signals generated using trend lines are not Markov times. Let $t_1$ and $t_0$ be the times of onset of the two lowest (highest) local minima (maxima) of $X_t$ during the period $(\tau_{i-1}, t]$, where $\tau_{i-1}$ is assumed to be known. Let $X_1$ and $X_0$ be the values of these minima (maxima). Consider the trend line $T(t)$,

$$T(t) = [(X_1 - X_0)/(t_1 - t_0)](t) + [(X_0 t_1 - X_1 t_0)/(t_1 - t_0)], \tag{5}$$

for $t_1 > t_0 > \tau_{i-1}$. This function defines a straight line that goes through the two lowest (highest) local minima (maxima) observed during the interval $(\tau_{i-1}, t]$. As in the previous case, we obtain the times $\{\tau_i\}$ using

$$Z_t = X_t - T(t), \tag{6}$$

and

$$\tau_i = \inf_t \{t : t > \tau_{i-1}, Z_t Z_{t-1} < 0\}. \tag{7}$$

I now show that the times $\{\tau_i\}$ generated by this algorithm will not be Markov times.

It is clear that if the $Z_t$ defined by (6) is $I_t$-measurable, and if we adopted the rule in (7) to determine the $\{\tau_i\}$, then this would be the first entry in the interval $[0, \infty)$ by an $I_t$-measurable random variable, and the $\tau_i$ would be Markov times. But it turns out that, in general, $Z_t$ is not $I_t$-measurable since $t_1$ and $t_0$ are never specified as the times of onset of the first two (or the $n$th) local minima (maxima) during $\tau_{i-1} < t$. In practice, the $t_0$ and $t_1$ are simply said to be two lowest (highest)

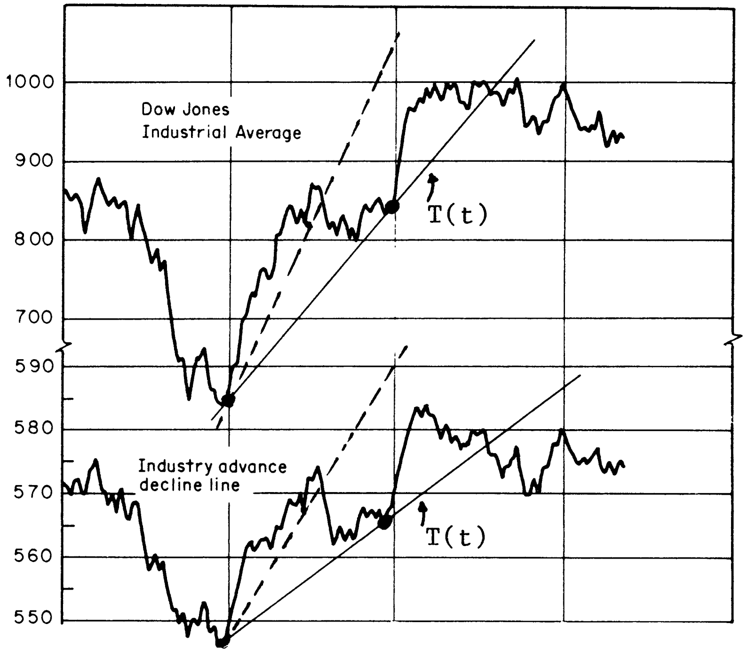



FIG. 4.—An example to non-$I_t$-measurable trend lines

local minima (maxima) that occur after some predetermined time $\tau_{i-1}$. But such an event is not $I_t$-measurable since, before it can be decided whether a local maxima is highest or second highest, one needs to know the levels of subsequent maxima. Figure 4 illustrates this point. None of the trend lines shown here utilize the first two local maxima in determining the $T(t)$. There were several local maxima between the selected $t_1$ and $t_0$, and these were ignored in obtaining the trend lines of figure 4. Two of these are shown on figure 4 as dotted lines.

Thus we see that trend crossing techniques will not generate Markov times unless one specifies an $I_t$-measurable mechanism for ignoring the local minima (maxima) between $t_1$ and $t_0$.

### *Characterization of Patterns*

The third class of procedures used by technical analysts utilizes the occurrence of various patterns to issue signals. Some of these patterns are shown in figures 3*a–c*. The theorem above suggests that, if these patterns are well-defined signals of upcoming events, one should be able to formulate them as first entries of an $I_t$-measurable random process in a set $A \in R$. In this article, the two most popular patterns are considered, namely, "triangles" and "head and shoulders." I first show that, in principle, these patterns can be formally defined using particular sequences of local minima and maxima. Second, I claim that, in their current formulation, these patterns are not $I_t$-measurable events.

An example of head and shoulders is shown in figure 3*a*. According to this figure, a head and shoulders pattern is observed whenever the trend lines that envelop the data behave as a step function: two sets of local minima with similar heights, separated by some higher local minima during the interval $(\tau_{i-1}, t]$. Let the mutually exclusive sets,

$$\{t_0 < \ldots < t_k\}, \quad \{t_{k+1} < \ldots < t_m\}, \quad \{t_{m+1} < \ldots < t_n < t\}, \qquad (8)$$

denote the times of onset of three sets of (lowest) consecutive local minima up to time $t$. To obtain a head and shoulders pattern, the heights of the local minima in the first and third sets must be (approximately) the same, say $M^*$. In addition, the levels of the local minima in the second set must be significantly higher, say $M^{**}$. Then a (sell) signal is issued the first time $X_t$ falls below $M^*$ once such a pattern takes shape. That is to say,

$$\tau_i = \inf_t \{X_t < M^*, \quad t > t_n\}.$$

Since $\{t_i < t, i = 1, \ldots, k\}$, the local minima defined in (8) are $I_t$-measurable. Hence an event describing head and shoulders becomes $I_t$-measurable once a formal way of subdividing the three sets of local extrema shown in (8) is selected. Such a criterion is needed in order to decide when the local minima in the middle exceed significantly the local minima in the first and third sets. This requires an a priori selection of a lower bound on the difference $M^{**} - M^*$, although the levels of $M^*$ and $M^{**}$ need not be specified individually. If all these conditions are met, a head and shoulders pattern becomes $I_t$-measurable.

This construction shows that actual signals generated using observed head and shoulders patterns are not Markov times. For example, in figure 3, the first occurrence of such an event is illustrated by the line *AB* rather than the suggested head and shoulders pattern *CD* selected by technicians. The only way one would select *CD* is if one anticipated that a local minimum such as *D* would occur at time $t$. Accordingly, in this example, the decision of whether $\tau = t$ or not depends on future values of the underlying series. The stopping time illustrated in figure 3*a* cannot be a Markov time.

Further, head and shoulders patterns defined formally as above are likely to be probability-zero events if one insists that the minima in the first and second sets have the same height $M^*$. Conversely, removing this requirement will impose further a priori restrictions on the sets of local minima shown in (8).

Figure 3*b* illustrates a triangle. In principle, this pattern can also be defined using consecutive local minima and maxima. To generate such a triangle, consecutive local maxima have to be in descending, and consecutive local minima in ascending order. Thus let

$$\{t_{\max,1} < \ldots < t_{\max,k}\} \text{ and } \{\mathrm{Max}_1 > \mathrm{Max}_2 > \ldots > \mathrm{Max}_k\} \tag{9}$$

represent the times of onset and the heights of $k$ consecutive local maxima of $X_t$ during $\tau_{i-1} < t$. Similarly, let $\{t_{\min,1} < \ldots < t_{\min,k}\}$ and $\{\mathrm{Min}_1 < \ldots < \mathrm{Min}_k\}$ be the times of onset and heights of $k$ lowest local minima during the same period. Then, a buy or sell signal is generated as soon as observed $X_t$ exceeds the last local maxima or falls below the latest local minima (see fig. 3*b*). More precisely,

$$\tau_i = \inf_t \{\mathrm{Max}_1 > \ldots > \mathrm{Max}_k < X_t \text{ or } \mathrm{Min}_1 < \ldots < \mathrm{Min}_k > X_t\}, \tag{10}$$

for $\tau_{i-1} < t$.

Clearly, this is an $I_t$-measurable event, hence a prediction method using triangles defined this way will generate Markov times. Yet this does not mean that, in practice, the signals generated using triangles are nonanticipatory. In fact, what makes the above signals $I_t$-measurable is the a priori specification of the parameter $k$, namely, the number of minima or maxima that one has to observe before the crossing occurs. If this information is omitted from the definition, the use of triangles will cease to generate Markov times. Without this parameter, even two consecutive local extrema can generate a triangle. Clearly, this is not what technical analysts have in mind, as shown in figure 3*b*. If two extrema are not sufficient, then how many does one need? Obviously, the answer to these questions necessitates a priori selection of some parameter such as $k$.

Finally, in figure 3*c*, we show another pattern, namely, gaps in daily price ranges. In contrast to other patterns, the use of gaps does generate Markov times since a signal is issued the first time three consecutive gaps are observed. This constitutes a first entry into an interval and leads to Markov times.

## A Criterion on Practical Use

Suppose a method of technical analysis is known to generate Markov times $\{\tau_i\}$ as signals of market turning points. A forecaster may, in addition, want to know the size of the probability $p(\tau_i < \infty)$, $i = 1, 2, \ldots$ before investing resources in applying this rule. Indeed, if this probability is less than one, then the rule may never give a signal. This may be uneconomical. Yet, in terms of formal statistical criteria, there is nothing wrong with a Markov time that fails to be finite. In this section, I discuss which categories of technical analysis are likely to yield finite Markov times.

DEFINITION. We say that a Markov time $\tau$ is *finite* if

$$P(\tau < \infty) = 1. \tag{11}$$

It is clear that a Markov time that is not finite may fail to be a financially rewarding method of forecasting since it may never give a positive or negative signal in spite of being well defined.

It turns out that only in very few cases the $\{\tau_i\}$ generated by technical analysis will be finite, hence usable, in the sense above. The major exception is the method of moving averages. I show below the conditions under which the moving average method generates finite Markov times.

PROPOSITION 1. If the observed process $\{X_t\}$ is stationary and $m$-dependent, all moving average methods characterized by (3)–(4) generate finite Markov times.

*Proof.* Let $Z_t$ be given by (3). If $X_t$ is stationary, then $Z_t$ and $Z_tZ_{t-1}$ are stationary (e.g., Breiman 1968, proposition 6.6). Also note that, due to stationarity, $E[Z_t] = 0$, hence $0 < P(Z_t \geq 0) < 1$, unless $Z_t = 0$ almost surely. Let

$$Y_t = Z_tZ_{t-1}, \quad t = 0, 1 \ldots .$$

Clearly, $P(Y_t \geq 0) < 1$. Now, I apply the theorem provided in the Appendix. Consider

$$P(Y_t \leq 0, \text{at least once for } t \leq n) = 1 - P(Y_0 > 0, Y_1 > 0, \ldots Y_n > 0).$$

The theorem in the Appendix requires this probability be one, as $n$ goes to infinity. To show that this is indeed true, note that, if $X_t$ is $m$-dependent, the $Y_t$'s sufficiently apart will also be independent. Thus, select an integer $u$ so that $Y_t$ and $Y_{t+u}$ are independent. We utilize such $Y_t$'s sufficiently apart to write, for large $n$,

$$\begin{aligned} P(Y_0 > 0, \quad Y_1 > 0, \ldots Y_n > 0) &\leq P(Y_0 > 0)P(Y_u > 0) \ldots P(Y_{ku} > 0) \\ &\leq P(Y_0 > 0)^k \end{aligned}$$

by stationarity and $m$-dependence. As we let $k \rightarrow \infty$,

$$P(Y_0 > 0)^k \rightarrow 0,$$

since $P(Y_0 > 0) < 1$, as shown above. Thus,

$$P(Y_n \leq 0 \text{ at least once}) = 1.$$

Hence, all conditions of the theorem supplied in the Appendix are satisfied and Markov times $\tau_1, \tau_2, \ldots, \tau_n$ are finite.

Note that assumptions such as stationarity and mixing, a simple form of which is $m$-dependence, are needed to obtain this result. This might seem unnecessary, but without similar assumptions, one cannot guarantee the finiteness of these Markov times. Indeed, if the process

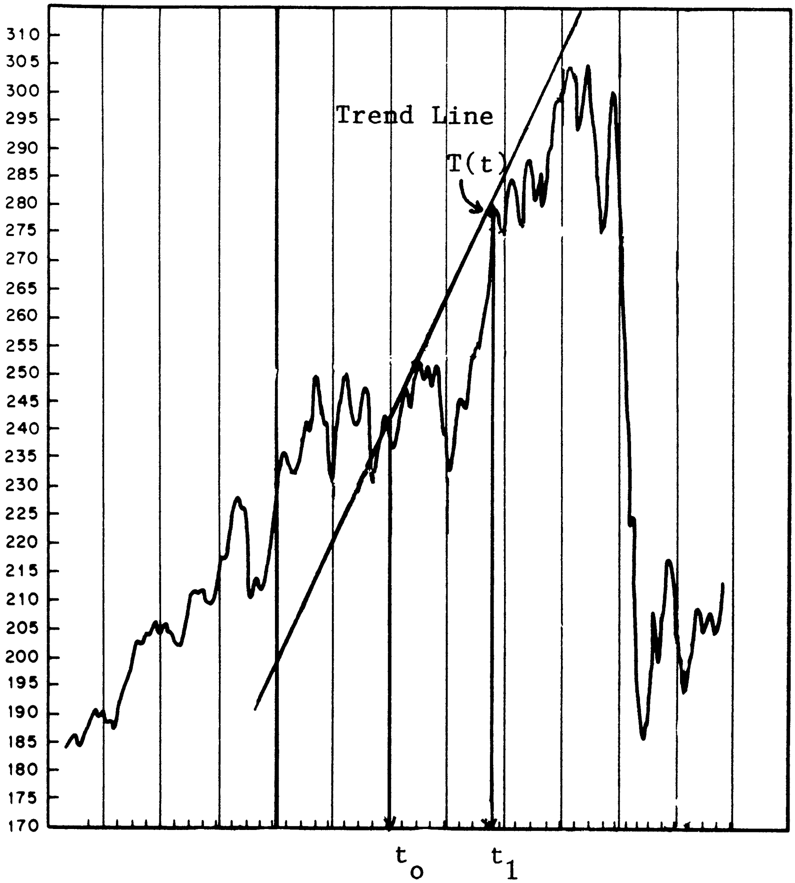



FIG. 5.—An infinite Markov time

$X_t$ is explosive enough, then a moving average method may not generate finite Markov times.[4]

One implication of this is that trend crossing methods of technical analysis might not always yield finite Markov times—even after a pre-

4. Here is an example provided by the referee. Let $X_t$ be generated by an explosive AR(1) model:

$$X_t = \beta X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots$$

where $\beta > 2$, and $\epsilon_t$ are independently and identically distributed random variables with uniform distribution over the interval $[-1, 1]$. Note that the event $E = \{X_t$ is always larger than 1 and tends to $\infty\}$ has positive probability. Now, consider two moving averages with 1 and 2 terms, respectively. Then $Z_t$ defined by formula (3) is equivalent to

$$Z_t = (1/2)[(\beta - 1)X_{t-1} + \epsilon_t].$$

With this $Z_t$, we have $\tau = \infty$ on the event $E$.

cise definition is adopted. To illustrate this, note that a trend line $T(s)$ such as the one shown in figure 5, will admit the representation,

$$T(s) = a_t + b_t s \quad s > t, \tag{12}$$

where $a_t$ and $b_t > 0$ are $I_t$-measurable intercept and slope. Now, the difference,

$$D_s = X_s - a_t - b_t s, \quad s > t,$$

is clearly not stationary. So the assumptions of proposition 1 are not satisfied for $D_s$, $s > t$. Hence, even with stationary and $m$-dependent $\{X_t\}$, as $s$ goes to infinity, $P(X_s - a_t - b_t s \leq 0)$ may equal one, and $X_s$ may never cross the trend line $T(s)$ again. Under these conditions, implied Markov times may be infinite even though they are well defined. For example, this may be the case if $\{X_t\}$ is given by

$$X_t = \epsilon_t + .5\epsilon_{t-1}$$

where the distribution of the independently and identically distributed (i.i.d.) errors $\{\epsilon_t\}$ has finite support:

$$P(\epsilon_t \geq \alpha) = 0 \quad 0 < \alpha < \infty.$$

REMARK. Note that, even if a rule generates signals that are finite with high probability, with, say, $P(\tau_i < \infty) = .9$, this may still create major problems for practical users. In fact, such a probability implies that one out of every 10 signals may be infinite—assuming that the signals are sufficiently apart, and that they are not correlated. For a forecaster working in real time, a long waiting period then implies either a large (but finite) $\tau_i$ or, with smaller probability, an outcome where no signal will be given. In this latter case, the forecaster should switch to other rules. Since technical analysis never specifies how one rule should be abandoned in favor of others, the requirement that $P(\tau_i < \infty) = 1$ is less trivial than it seems at the outset.

## Predictive Power of Technical Analysis

The fact that some methods of technical analysis admit a formal definition is important. Yet well-defined sequences of finite Markov times $\{\tau_i\}$ may still have no predictive value. Thus, the next question is, Under what conditions (if any) would the well-defined procedures of technical analysis be useful in prediction over and above the standard econometric models?

There are two results. The first deals with the usefulness of technical analysis under the assumption that observed data can be characterized as linear processes. I adopt the following definition of linearity.

DEFINITION. A process $\{X_t\}$, $E[X_t] < \infty$ is said to be linear, or has the linear regression property, if, for $s \geq 0$,

$$E[X_{t+s}|X_{t-1}, X_{t-2}, \ldots, X_{t-k}] = \alpha_1 X_{t-1} + \ldots + \alpha_k X_{t-k}.$$

That is, the process is linear if expectations of $X_t$ given finite past $X$'s are linear in the latter. In particular, Gaussian processes are linear. In fact, the class of processes that have the linearity property is identical to the sub-Gaussian processes (Hardin 1982). However, our definition of linearity is not identical to the one given in Hardin (1982), who does not discriminate between past and future $X$'s as conditioning factors.

REMARK. It is interesting to note that the definition of linearity that we have here is not equivalent to $E[X_t|X_{t-1}, X_{t-2}, X_{t-3} \ldots]$ being a linear combination of past $X_t$'s. It is possible to construct finite moving average (MA) processes with infinite autoregressive representations, that are not linear according to the definition used here.[5]

I now show that nonlinearity of asset prices is a necessary condition for the usefulness of technical analysis.

PROPOSITION 2. If the $X_t$ process is linear in the sense above, then no sequence of Markov times obtained from a finite history of $\{X_t\}$ can be useful in prediction over and above (vector) autoregressions.

*Proof.* If $\{\tau_i\}$ are Markov times obtained from a finite history of $\{X_t\}$, they must be measurable with respect to $\{X_t, X_{t-1}, \ldots, X_{t-k}\}$, some finite $k$. This means that

$$\begin{aligned} &E[X_{t+s}|\{X_t, X_{t-1}, \ldots, X_{t-k}\}, \quad \{\tau_i : \tau_i < t\}] \\ &= E[X_{t+s}|\{X_t, X_{t-1}, \ldots, X_{t-k}\}] \\ &= \alpha_0 X_t + \alpha_1 X_{t-1} + \ldots + \alpha_k X_{t-k}, \end{aligned}$$

due to the linearity of $\{X_t\}$.

This proposition may have important implications for technical analysis. First of all, it can be seen that one necessary condition for the usefulness of any technical analysis rule is the requirement that asset prices be nonlinear in the sense of the definition above. For example,

5. An interesting example provided by the referee is the following: Let $\epsilon_t$ be i.i.d. and

$$\epsilon_t = \begin{cases} -1 & \text{with probability } 2/3 \\ 2 & \text{with probability } 1/3. \end{cases}$$

Construct the process $X_t$ using

$$X_t = \epsilon_t + .5\epsilon_{t-1}.$$

Clearly, $X_t$ has an infinite autoregressive representation, hence, is a linear combination of all past $X_t$'s. Yet, $E[X_t|X_{t-1}]$ cannot be linear in $X_{t-1}$. To see this, note that $E[X_t|X_{t-1} = 0]$ can be directly calculated to be $-1/2$, yet if $E[X_t|X_{t-1} = 0]$ were linear in $X_{t-1}$ in the sense of our definition, $E[X_t|X_{t-1} = 0]$ would have to equal zero. This contradiction implies that the $X_t$ process cannot have a linear regression property as defined here.

if a rule is well defined and yet stock prices are Gaussian, then, due to proposition 2, we immediately know that the rule is useless as a prediction technique. If, however, there is some evidence that stock prices are nonlinear, technical analysis may be useful in prediction—that is, it may be a simple way of taking such nonlinearities into account. Under these conditions, the question of whether technical analysis rules have any predictive power becomes an empirical issue.

Recent work such as Hinich and Patterson (1985) and Diebold (1989) provide evidence on the nonlinearity of data from financial markets. Yet because notions of linearity and nonlinearity used in these papers and here are not identical, these empirical results do not necessarily imply that there are forms of technical analysis useful in prediction. For example, Diebold (1989) shows that taking nonlinearities into consideration does not improve forecasts of exchange rates, although there appears to be a great deal of evidence that these latter are nonlinear.

Since Martingales are linear processes, a corollary to proposition 2 is the following:

COROLLARY. If the $X_t$ process is a Martingale, then no sequence of finite Markov times $\{\tau_i\}$, calculated from a finite history of $\{X_t\}$, can be useful in prediction over and above linear regressions.

The point is that, if some technical analysis rules are indeed useful in prediction, then this should rule out a Martingale representation for the series under consideration. Using these propositions we provide some empirical results.

*Empirical Results.* To illustrate how one can test the predictive value of technical analysis, I select the method claimed to work the best according to participants in financial markets. "Although technical analysts caution that investors should consider a variety of factors in trying to discern the market's direction, they say the single, clearest factor is probably the 150-day moving average. History has shown that when the (Dow-Jones) index rises decisively above its moving average the market is likely to continue on an upward trend. When it is below the average it is a bearish signal."[6]

The moving average method was one of the few rules that generated Markov times. Also, these Markov times were easy to quantify. This, plus one other consideration, made me choose stock prices as the $X_t$ in (3), and the 150-day moving average as the $X_t^*$. We then use the algorithm in (4) to obtain a sequence of Markov times $\{\tau_i\}$. The last consideration for making these selections was the availability of a long sample for the Dow-Jones index. In fact, when using Dow-Jones industrials it is possible to go all the way to 1792 and work with almost a 200-year-long monthly data series. This greatly facilitates investigating

6. *New York Times* (March 11, 1988).

the predictive power of technical analysis since on-and-off prediction rules are likely to yield relatively few signals compared to regular monthly data.

Proposition 2 and the corollary that follows it provide the necessary framework to do the empirical work. According to these, we need to show that, given a long autoregression, the addition of Markov times to the right-hand side does not improve forecasts of Dow-Jones industrials. If this is the case, then the rule in question will have no predictive value.

Thus I let

$$X_{t+\mu} = \sum_{i=1}^{k} \alpha_i X_{t-i} + \sum_{i=1}^{n} \beta_i D_{t-i} + \epsilon_{t+\mu} \quad \mu > 0, \quad n < k, \tag{13}$$

where

$$D_t = \begin{cases} 1 & \text{if } X_t > X_t^* \text{ has occurred at } t \text{ given } X_{t-1} < X_{t-1}^*, \\ -1 & \text{if } X_t < X_t^* \text{ has occurred at } t \text{ given } X_{t-1} > X_{t-1}^*, \\ 0 & \text{otherwise,} \end{cases}$$

and where the disturbances $\{\epsilon_t\}$ form an innovation sequence with respect to the finite history of $X_t$,

$$E[\epsilon_{t+\mu} | X_{t-1}, \ldots, X_{t-k}] = 0.$$

According to this, $\epsilon_{t+\mu}$ measures all unpredictable events between $t$ and $t + \mu$. The $\{\beta_i\}$ represents the contribution of the Markov times $\{\tau_i\}$ in explaining $X_{t+\mu}$ over and above the own past of the series. To the extent the $D_t$'s are obtained from $\{X_{t-1}, \ldots, X_{t-k}\}$, they should have no contribution to forecasting $X_{t+\mu}$ beyond the finite history of $X_t$ if this latter is a linear process.

There is an important point that concerns inference with equation (13). Note that (13) requires a sufficiently long autoregressive component (i.e., a large $k$). Otherwise, if $k$ is small, then some $D_{t-i}$'s may become significant simply because they are calculated from a more distant past of $\{X_t\}$.

The parameter $\mu$ in (13) determines how many periods ahead one is forecasting. It captures the claim that the moving average method detects changes in long-run trends, and that it is not necessarily useful for 1-period-ahead forecasts. Hence the value of $\mu$ should be selected as greater than 1 or 2 months. In the empirical work reported below, I selected $\mu$ (arbitrarily) as 12 months. The results remain qualitatively similar for $\mu$ greater than 12. Inference with the equation shown in (13) appears to be straightforward at the outset. However, if $\mu > 1$, the errors of equation (13) will be serially correlated, and this needs

TABLE 1 Dow-Jones Industrials, 1792:1–1851:12

| Label and Lag | Coefficient | $t$-statistic |
|---|---|---|
| Dow Jones industrials: | | |
| 18 | .83 | 3.8 |
| 19 | −.14 | −.45 |
| 20 | .03 | .05 |
| 21 | .02 | .03 |
| 22 | .03 | .05 |
| 23 | −.03 | −.09 |
| 24 | −.05 | −.15 |
| 25 | −.01 | −.04 |
| 26 | .01 | .04 |
| 27 | −.05 | −.15 |
| 28 | .01 | .05 |
| 29 | .01 | .05 |
| 30 | −.05 | −.15 |
| 31 | .01 | .04 |
| 32 | .26 | 1.34 |
| Dummy: | | |
| 18 | .45 | 1.16 |
| 19 | .63 | 1.39 |
| 20 | .67 | 1.35 |
| 21 | .58 | .92 |
| 22 | .61 | 1.06 |
| 23 | .50 | .71 |
| 24 | .32 | .69 |
| Constant | 1.51 | 7.12 |

NOTE.—$R^2$ = .73; sum of square of residuals = 3592; $F$-statistic = .46. 1792:1 = January 1792.

to be taken into account. In fact, the error structure in these equations will always be given by a ($\mu$ − 1)th-order MA process:

$$\epsilon_t = v_t + a_1 v_{t-1} + a_2 v_{t-2} + a_3 v_{t-3} + \ldots + a_{\mu-1} v_{t-\mu-1}, \qquad (14)$$

where the $\{v_t\}$ are the innovations in the $X_t$ process.

I corrected for the serial correlation shown in (14) using Hannan's efficient procedure. In fact, the $\{\epsilon_t\}$ can be consistently estimated by applying ordinary least squares to (13). The periodogram of these (first-stage) residuals is then calculated. The Fourier transform of $X_t$ is divided by the corresponding entries of the square root of the periodogram of residuals. This series is then transformed back to the time domain. Equation (13) is estimated with these transformed data.

Empirical results are provided in tables 1–3. The results are interesting. The $F$-tests on the $D_t$ are insignificant for the subperiods 1795–1851, and 1852–1910. However, they are highly significant for the period 1911–76. Thus the particular moving average rule of 150 days seems to have a significant predictive power for the latter part of the sample. It is interesting to note that any general belief by market parti-

**TABLE 2** **Dow-Jones Industrials, 1852:12–1910:12**

| Label and Lag | Coefficient | $t$-statistic |
|---|---|---|
| Dow Jones industrials: | | |
| 18 | .86 | 4.60 |
| 19 | .00 | .01 |
| 20 | −.08 | −.31 |
| 21 | .04 | .17 |
| 22 | −.21 | −.80 |
| 23 | −.09 | −.34 |
| 24 | .08 | .29 |
| 25 | −.06 | −.20 |
| 26 | −.04 | −.16 |
| 27 | .07 | .27 |
| 28 | −.16 | −.66 |
| 29 | −.06 | −.24 |
| 30 | .06 | .23 |
| 31 | −.01 | −.02 |
| 32 | .55 | 3.15 |
| Dummy: | | |
| 18 | .53 | .67 |
| 19 | .85 | 1.0 |
| 20 | 1.1 | 1.3 |
| 21 | 1.5 | 1.9 |
| 22 | 1.9 | 2.4 |
| 23 | 1.8 | 2.3 |
| 24 | 1.2 | 1.8 |
| Constant | 4.1 | 5.4 |

NOTE.—$R^2$ = .72; sum of square of residuals = 36120.202; $F$-statistic = .050330. 1852:12 = December 1852.

cipants that such a rule is useful would be self-fulfilling and would lead to significant $\{\tau_i\}$.

For this last period, all lags of the dummy variable that indicates buy (+1), sell (−1), and no action (0) signals are significant. Furthermore, the signs are in the right direction, in that they are all positive. It is also interesting to note that the coefficients of the dummy variable have a nice reverse V shape, with the peak occurring at lag 23 (table 1).

Hence, the moving average method does seem to have some predictive value beyond the own lags of Dow-Jones industrials. In fact, the results displayed in these tables remained qualitatively similar when different values were used for $\mu$, except for $\mu = 1$, where the 150-day moving average turned out to be insignificant in all equations.[7]

7. Estimates of the same equation with $\mu = 1$ yields no significant lags for the dummy variable in consideration. This supports the contention that the method predicts longer-run behavior of the $X_t$.

TABLE 3 Dow-Jones Industrials, 1911:1–1976:12

| Label and Lag | Coefficient | $t$-statistic |
|---|---|---|
| Dow Jones industrials: | | |
| 18 | .19 | 2.34 |
| 19 | −.04 | −.39 |
| 20 | .01 | .05 |
| 21 | .18 | 1.36 |
| 22 | .05 | .34 |
| 23 | .04 | .25 |
| 24 | .19 | 1.31 |
| 25 | .12 | .78 |
| 26 | .01 | .04 |
| 27 | .09 | .61 |
| 28 | .02 | .13 |
| 29 | .07 | .48 |
| 30 | .02 | .14 |
| 31 | −.01 | −.05 |
| 32 | .09 | 1.11 |
| Dummy: | | |
| 18 | 26.9 | 2.9 |
| 19 | 28.3 | 2.9 |
| 20 | 28.4 | 2.9 |
| 21 | 28.2 | 2.9 |
| 22 | 30.1 | 3.1 |
| 23 | 30.8 | 3.2 |
| 24 | 21.8 | 2.4 |
| Constant | 19.6 | 4.6 |

NOTE.—$R^2$ = .91; sum of square of residuals = 6521516; $F$-statistic = 3.71. 1911:1 = January 1911.

## Conclusions

This article discussed some criteria that one can apply in evaluating the set of ad hoc prediction rules widely used in financial markets and generally referred to as technical analysis. I showed that a few of these rules generate well-defined techniques of forecasting. Under the hypothesis, economic time series are Gaussian, and even well-defined rules were shown to be useless in prediction.

At the same time, the discussion indicated that if the processes under consideration were nonlinear, then the rules of technical analysis might capture some information ignored by Wiener-Kolmogorov prediction theory.

Tests done using the Dow-Jones industrials for 1911–76 suggested that this may indeed be the case for the moving average rule.

## Appendix

PROPOSITION (Breiman 1968). Let the process $\{Y_t\}$ be stationary, such that

$$P(Y_n \geq 0 \text{ at least once}) = 1;$$

then the $\tau_i$, $I = 1, 2, \ldots$ are finite almost surely and on the sample space $\{w\colon Y_0 \geq 0\}$ they form a stationary sequence under the probability $p(\cdot \mid Y_0 \geq 0)$, and

$$E[\tau_1 \mid Y_0 \geq 0] = 1/p(Y_0 \geq 0).$$

## References

Breiman, L. 1968. *Probability*. Reading, Mass.: Addison-Wesley.

Brockett, P. L.; Hinich, M. L.; and Patterson, D. 1985. Bispectral based tests for the detection of Gaussianity and linearity in time series. Manuscript. University of Texas at Austin.

Diebold, F., and Nason, J. M. 1989. Nonparametric exchange rate prediction? Manuscript. Washington, D.C.: Federal Reserve Board of Governors.

Granger, C. W., and Andersen, A. 1978. *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenhoeck & Ruprecht.

Hardin, C. D. 1982. On the linearity of regressions. *Zeitschrift für Wahrscheinlichkeitstheorie* 61:291–302.

Hinich, M. L., and Patterson, D. M. 1985. Evidence of nonlinearity in daily stock returns. *Journal of Business and Economic Statistics* 3 (January): 69–77.

Leadbetter, M. R.; Lindgren, G.; and Rootzen, H. 1983. *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.

Pring, M. J. 1980. *Technical Analysis Explained*. New York: McGraw-Hill.

Rhea, R. 1932. *Dow Theory*. New York: Barron's.

Shiryayev, I. 1985. *Probability*. New York: Springer-Verlag.

Tong, H. 1983. *Threshold Models in Non-linear Time Series Analysis*. New York: Springer-Verlag.