Corridor analysis for detection of significant breakout movements
by Mahadevan Krishnamoorthi

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mechanical & Industrial Engineering
Montana State University

Abstract:
Time series form patterns during their movements and have been continuously analyzed for relevance to forecasting. A neutral corridor is one of many patterns formed by time series. Neutral corridors, when broken, result in significant upward or downward movements of time series. Therefore, it is imperative that a forecaster be able to detect such significant movements. Research was undertaken to identify a quantitative technique for the detection of such significant breakout movements.

Three techniques were analyzed for possible use. The techniques are cusum control charts, nonparametric tests for location and trends, and a new empirical method developed by the author. Neutral corridors were identified from historical data corresponding to market securities. Rules were formed to define significant breakout movements and each test was performed on all sets of data. A signal of positive or negative trend (breakout movements) was verified by comparison to the actual plot of the data. Accuracy of a signal was determined based on the actual direction of movement and the time of movement.

The cusum control charts and the nonparametric tests gave erratic signals of positive and negative trends and were not confirmed by the actual plot of the time series. Hence these tests cannot be recommended for the purpose of detecting significant breakout movements. The new empirical method illustrated a high accuracy of above 90 percent correct breakout calls, and therefore, is highly recommended for detection of significant movements away from neutral corridors.

CORRIDOR ANALYSIS FOR DETECTION OF SIGNIFICANT BREAKOUT MOVEMENTS

by

Mahadevan Krishnamoorthi

A thesis submitted in partial fulfillment

of the requirements for the degree

of

Master of Science

in

Mechanical & Industrial Engineering

MONTANA STATE UNIVERSITY-BOZEMAN

Bozeman, Montana

May 1996

APPROVAL

of a thesis submitted by

Mahadevan Krishnamoorthi

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Paul Schillings	Paul L. Schillings	28 May 1996
	(Signature)	Date

Approved for the Department of Mechanical and Industrial Engineering

Dr. Victor Cundy	Vic A. Cy	5/28/96
	(Signature)	Date

Approved for the College of Graduate Studies

Dr. Robert Brown	R Brown	6/10/96
	(Signature)	Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University - Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair-use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature ____________________

Date May 28, 1996

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF TABLES - Continued

LIST OF TABLES - Continued

LIST OF FIGURES

LIST OF FIGURES - Continued

ABSTRACT

Time series form patterns during their movements and have been continuously analyzed for relevance to forecasting. A neutral corridor is one of many patterns formed by time series. Neutral corridors, when broken, result in significant upward or downward movements of time series. Therefore, it is imperative that a forecaster be able to detect such significant movements. Research was undertaken to identify a quantitative technique for the detection of such significant breakout movements.

Three techniques were analyzed for possible use. The techniques are cusum control charts, nonparametric tests for location and trends, and a new empirical method developed by the author. Neutral corridors were identified from historical data corresponding to market securities. Rules were formed to define significant breakout movements and each test was performed on all sets of data. A signal of positive or negative trend (breakout movements) was verified by comparison to the actual plot of the data. Accuracy of a signal was determined based on the actual direction of movement and the time of movement.

The cusum control charts and the nonparametric tests gave erratic signals of positive and negative trends and were not confirmed by the actual plot of the time series. Hence these tests cannot be recommended for the purpose of detecting significant breakout movements. The new empirical method illustrated a high accuracy of above 90 percent correct breakout calls, and therefore, is highly recommended for detection of significant movements away from neutral corridors.

CHAPTER 1

INTRODUCTION

Time series form patterns during their movements. These patterns have been continuously analyzed for relevance to forecasting. By studying the nature of previous turning points, it is possible to develop some characteristics that can help identify movements of time series. Studies of patterns for purposes of forecasting are based on the assumption that history repeats itself. *The art of forecasting - for it is an art - is to identify trend changes at an early stage and to maintain a posture until the weight of evidence indicates that the trend has reversed* [Pring, 11]. It is worthwhile to note that time series never duplicate their performance exactly, but the recurrence of similar characteristics is sufficient to enable forecast analysts to identify important junctures.

Here, data obtained from market securities have been used for analysis. Reasons for using data from market securities are fourfold:

1. Ease of availability
2. Accurate and timely information
3. Availability of real time data, for verification.
4. Reluctance of business people to release privy information.

Movements of market securities can be classified as long (primary), intermediate, and short term. Primary movements typically work out in a period of 1 to 3 years. Intermediate movements usually develop over a period of 3 weeks to as many months, sometimes longer. Short term movements last less than 3 weeks. Studies presented in this paper are primarily confined to identification of intermediate term movements in the market securities.

As in a court of law or when testing a hypothesis, a trend in presumed innocent (continuing) until proven otherwise. The "evidence" is the objective element in forecasting. The evidence is derived from the use of one or more forecasting techniques. Not all of them work for all situations. The "art" consists of combining forecasting techniques into an overall picture and recognizing the resemblance of that picture to market patterns.

Corridors

Corridors, are one of the most important and often analyzed patterns. Corridors could be classified as neutral, positive or negative corridors depending on the direction of movement of the time series within a corridor. A neutral corridor (or sideways trend) is essentially horizontal or transitional, which usually separates two major market movements. To the forecaster, the neutral corridor has great significance because it marks the turning point between major market movements. The phenomenon of neutral corridors and their formation is described below. Figure 1 illustrates a typical neutral corridor formed by a market security.

Suppose a time series moving in the upward direction reaches the top at point *A* (Figure 1) and then starts moving down. The point A corresponds to the highest value, the market security reached within that trading period (usually a day). Now, if the time series

Figure 1. Neutral Corridor formed by Market Security 8

reaches point *B* and reverses its direction of movement, to go up, a neutral corridor is likely to be formed. The point B corresponds to the lowest value the market security reached in that trading period. The formation of the neutral corridor is confirmed only some time after this picture, including points A and B, is formed. Now, if the time series, while moving up, does not go above point A, but changes direction and moves past the center of the vertical distance between points A and B, point *A* is called a "valid top." A line (1-1 in Figure 1), is drawn parallel to the X axis from point *A* and termed the "upper boundary" [Schillings, 12]. When the time series once again changes direction from a downward to an upward movement moving past the center line, we are permitted to draw a line from point *B*. The second line, (2-2 in Figure 1) is also drawn parallel to the X axis, and is termed the "lower boundary." The point B is now called a "valid bottom." Appendix A illustrates more examples of neutral corridors formed using the above method.

Neutral corridors sometimes include micro corridors within their boundaries. These micro corridors may exhibit positive, negative or sideways trends. However, for purposes of analysis one can assume randomness of the oscillations within the boundaries of a neutral corridor.

Neutral corridors are broken only by significant movements of the time series. A significant movement is defined as "a movement outside the boundary in either direction, culminating at a distance of more than the width of the corridor." The culmination of a movement is observed when the time series retracts by more than half of the total distance moved in the primary direction of movement. Understanding the concept of significant movements is imperative because time series often cross the boundaries of corridors momentarily and then move back into the corridors. Such moves are quite misleading and may result in incorrect decisions. Figure 9 shown in Appendix A illustrates temporary movements out of a corridor.

Research was undertaken to eliminate subjectivity in detecting significant movements away from neutral corridors. Here, an effort has been made to formulate a quantitative technique to consistently identify significant movements. This research may be classified as part of the continuing search for the "holy grail" or perfect technique.

Three techniques have been analyzed for possible use to detect significant movements. The techniques are cusum control charts, nonparametric tests for location and trends, and an empirical method developed by the author.

CHAPTER 2

LITERATURE REVIEW

A neutral corridor is one of many time series patterns that are often analyzed by forecasters. Although patterns resembling neutral corridors have been observed and considered for many years, the author and Schillings [Schillings, 12] found no published material that analyzed quantitative aspects for the detection of significant breakout movements. It is worth mentioning that the patterns studied in this paper were referred to for the first time, as *neutral corridors,* by Schillings. Pring [Pring, 11] calls these patterns *transition zones* or *rectangles*.

Currently, there is one technique used for analyzing patterns resembling neutral corridors which concentrates on the position of the time series values and corridor lengths. If a neutral corridor is formed at a market top, then the time series is usually expected to undergo a reversal and revert to a negative trend [Pring, 11]. If the corridor is formed at the bottom, then the time series is expected to revert to a positive trend. Analysis based on the length of the corridor leads to an estimate of how far the time series will go up or down after the breakout.

Pring finds that some practitioners (references not cited) define a movement away from a rectangle by a three percent penetration of either boundary as "significant." This filters out only some of the misleading moves, but does not ensure consistency in detection of statistically significant breakout movements. A movement of half the width of the corridor away from either boundary is also termed "significant." Both rules are only heuristics followed by some forecasters.

The review of published literature related to corridor analysis, as defined in this paper, does not indicate the availability of a quantitative technique. Hence, research was undertaken to formulate a quantitative technique for the detection of statistically significant breakout movements from time series neutral corridors.

CHAPTER 3

CONCEPTS OF STATISTICAL TESTING

Tests, usually called hypothesis tests or significance tests, often play a major part in statistical investigations. The basic idea of most statistical techniques is to increase our knowledge about populations using information in samples taken from them. In statistical testing, we are concerned with examining the truth, or otherwise, of hypotheses about some feature(s) of one or more populations. A statistical hypothesis is a statement about a population; for example its form or shape, or some aspect thereof; the numerical value of one or more parameters; and so forth [Gibbons, 4]. A hypothesis set consists of two statements, called the null hypothesis, H_0 and the alternative hypothesis, H_1. These statements must be mutually exclusive.

Whether or not particular hypotheses are eventually assessed to be reasonable, will depend on the weight of evidence contained in the sample data taken from population(s). In hypotheses testing, we never claim to *prove* anything completely (beyond doubt) by means of statistical test: we simply pronounce a judgement based on the available evidence, and give an assessment of the strength of that evidence [Neave, 9]. Sometimes, the evidence may be so overwhelming that a hypothesis may be regarded as proved (or disproved) "for all practical purposes."

Approach to Hypothesis Testing

For example, if the null hypothesis is stated as "The populations under consideration do not differ in persistence", then an alternative could be stated as "A motivated population exhibits more persistence for a task than a population with no motivation." As a mathematical statement, if μ_1 and μ_2 are parameters representing average persistence for the populations with motivation and without motivation, respectively, then the null hypothesis is $\mu_1 = \mu_2$ and the alternative is $\mu_1 > \mu_2$. The alternative hypothesis here is called one-sided (or one-tailed or right-tailed), because it states a particular direction of inequality. When the alternative is stated as $\mu_1 < \mu_2$, it is called left-tailed and when it is stated as μ_1 not equal to μ_2, then it is called two sided.

A decision to accept or reject any hypothesis is made on sample evidence according to some *statistical test procedure* and *test statistic*. The test procedure may also be called one-sided or two-sided, according to whether the alternative is one-sided or two-sided. One-sided tests are either left-tailed or right-tailed, depending on the direction of the alternative hypothesis, H_1. The test statistic should be consistent with, and appropriate for, the type of alternative, the data available for analysis, and the assumptions the investigator is willing to make about the population. By central limit theorem, the sampling distribution of the sample mean (test statistic in this case) is a normal distribution.

Once the test statistic is chosen, its value is calculated for the data obtained. Then the investigator can use this value and the sampling distribution of the test statistic to determine a quantity called the *P-value*, or the *associated probability*. The P-value is the probability, when H_0 is true, of obtaining a value of the test statistic which is equal to or "more extreme" in the appropriate direction than its critical value [Gibbons, 4].

When the investigator wishes to make a statistical decision, whether to reject or accept H_0, the decision can be based on the magnitude of the P-value in the following method. Because the P-value is found from the probability distribution of the test statistic under the null hypothesis, a very small P-value implies that a sample result this extreme, when H_0 is true, occurs only very rarely, by chance phenomena. Sampling error is primarily due to two causes: (1) sample does not encompass the entire population, and (2) each member of the population is not equally accessible to the sample. When the P-value is critically small, then the investigator can state that the data do not support the null hypothesis, that rejection is "statistically significant ." When the P-value is less than 0.05 the result is *probably* significant and if it is less than 0.01 then the result is *highly* significant. In both cases the statistical decision is to "Reject the null hypothesis, H_0 ."

Suppose that the P-value is large, then the sample result offers no convincing evidence that the statement made in H_0 about the population is false. In other words, the data do not deny the null hypothesis. The statistical decision is "Fail to reject the null hypothesis, H_0 ."

Errors Associated with Hypothesis Testing

It is imperative that we properly address the question "How rare is the rare event?", while discussing the decision process. The probability value taken as the cutoff between a rare event and a likely occurrence is frequently called the *level of significance* of the test

[Gibbons, 4]. The value for the cutoff is primarily a matter of personal choice, but it should reflect the investigator's feeling about the cost and the consequence of error.

Two types of error may occur. The first error occurs when the null hypothesis is rejected, while it is actually true and is called, Type I error, α. The second error, called Type II error, β occurs, when the null hypothesis is accepted while it is actually false. These definitions are summarized below.

Table 1. Summary of Type I and Type II Errors

Actual Situation

Decision Taken	**H_0 TRUE (Probability)**	**H_0 False (Probability)**
Accept H_0	Correct Acceptance ($1-\alpha$)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Rejection ($1-\beta$)

If rejecting a true hypothesis would be considered a serious mistake, a small value for α would be appropriate. However, a very small value of α may not be good either, because Type I error and Type II errors interact. When the probability of one type of error decreases, the probability of the other increases, but disproportionately because of the non-linearity of the distribution function. This is to be distinguished from the trivial case at μ_0, where $\alpha + \beta = 1.0$. Thus if a Type II error has serious consequences, a larger α value might be advisable. Once the value for α is chosen and the test result gives a P-value, the null hypothesis H_0 is rejected if $P <= \alpha$, but not otherwise. It is important to note that in statistical analysis, P-values can be an aid to making a research conclusion, but only in concert with the judgement and intelligence of the investigator.

CHAPTER 4

INVESTIGATION OF CUSUM CONTROL CHARTS

Process Behavior and Control Charts

A process is set of causes and conditions that repeatedly come together to transform inputs into outcomes [Moen, 8]. The inputs may include people, methods, material, equipment, environment, and information. The outcome is some product or service. Process behaviours are often studied with the use of control charts. Control charts are graphical displays of statistics plotted according to the order of their observation [Devor, 2]. A process is said to be in statistical control if there are only random patterns on a control chart. A succession of data emanating from a process, which is under statistical control, will exhibit variability due to a constant set of causes that are inherent in the process. These causes, usually called *common* causes, can also be thought of as causes leading to variability in the process.

Data observed from processes, often fall into a predictable pattern of variation, such as a normal distribution, easily described by simple statistical measures, namely, a mean and a standard deviation. These measures serve as a model to predict process behavior, when the process is subject only to a set of common causes. When a process is subject only to a set of common causes, data observed from the process will fall within the control limits of the process. Control limits are boundaries constructed on either side of the mean of the process, usually at a distance of three standard deviations.

For process control, data collected over time may be used to develop a statistical model as long as the data are collected while the process is subject to a set of common causes. Hence, if we can develop a model for the process measurements, then, when a major disturbance or an abnormal situation affects the process, the ensuing data will not conform to that model. The data from the abnormal situation will stand out clearly from the common-cause variability pattern. To be able to distinguish abnormal data from data generated during normal operating conditions, we should consider the process as it evolves over time.

Adoption to Forecasting

Data within a neutral corridor can be thought of as data generated by a process. Chi-squared tests for normality, were performed on sets of data and the results show that data within neutral corridors exhibit normality. All chi-square tests were performed at five percent level of significance and the results of some of them are listed in Appendix B. A mean and a standard deviation can now be computed and used to construct control charts for visual analysis of existence of special causes. If nonrandom patterns or the existence of data generated by abnormal situations are observed, then it is concluded that all the data points in the given set no longer fall within a neutral corridor.

A particular control chart that can be effectively used for purposes of predicting movements out of the boundaries of the neutral corridors is the cusum chart, where "cusum" stands for "cumulative sum ."

Cusum Charts

Cusum charts are becoming widely used in the industry because they are powerful, versatile, and have the ability to quickly detect small changes in a process mean [Lucas, 7]. The use of cumulative sums of sample data for on-line process control was developed by E. S. Page [Page, 10] and G. A. Barnard. The principle behind these charts is based on a method using the sequential likelihood ratio test [Wald, 13], where, as each new sample becomes available, a test is conducted to determine whether the process mean deviates by at least a specified amount from the target value.

Advantages of Cusum Charts

The most fundamental advantages of cusum charts are threefold:

i. Ease with which changes in mean level can be detected, either by observing points that lie outside the control limits or by a change in the slope of the chart.

ii. Ability to locate a point of change on a chart as that point at which the change in slope occurs.

iii. Efficiency over the standard control chart (Shewhart control chart) for changes of about 0.5σ and 2.0σ, which means that in this region changes can be detected approximately twice as quickly, when compared with Shewhart control charts [Ewan, 3]. These advantages make the cusum charts suited to analyze processes expected to have small and sustained deviations in the mean.

Traditional Cusum Charts

Common to all cusum control charts that are developed according to the sequential likelihood ratio is the idea of hypothesis testing between two alternative quality levels, one acceptable and the other rejectable. These cusum charts are usually referred to as *traditional cusum charts* [Devor, 2]. To construct traditional cusum charts both acceptable and rejectable quality levels must be specified. The necessity to specify the shifts (acceptable and rejectable quality levels) in advance, presents a serious concern in the application of these charts. However, the success of traditional cusum control charts depends mainly on the accurate estimation of both the target process parameter value and the size of the shift the control chart is designed to detect. Therefore, the use of traditional cusum charts is suited only when accurate estimation of target process parameter and shift are possible. Instead a modified version of the traditional cusum chart called the *standardized cusum chart* can be used. The method of preparation and analysis of the standardized cusum chart will be discussed later.

The traditional cusum control chart is developed according to the concept of hypothesis testing, where the hypotheses, for a one-sided cusum chart, are stated as follows:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1 \qquad (\mu_1 > \mu_0),$$

where μ is the mean of X and both Type I and Type II errors are specified beforehand. H_1 is accepted only if there is a significant rejection of H_0. If rejection is not significant then H_0 stands unrejected.

The likelihood ratio can be expressed in terms of α and β as,

$$\frac{\beta}{1-\alpha} < \text{likelihood ratio of the data} < \frac{1-\beta}{\alpha}$$

The above equation can be used to construct a one-sided cusum control chart for detecting upward shifts in the mean. If we have a time ordered sequence of independent sample observations, $X_1, X_2, \ldots, Xt$, from a normal population with variance σ_x^2 and an uncertain mean μ, it can be shown that the ratio will take the form

$$\frac{\sigma_x^2}{\Delta_1}\ln\frac{\beta}{1-\alpha} < S_t < \frac{\sigma_x^2}{\Delta_1}\ln\frac{1-\beta}{\alpha}$$

where $\Delta_1 = \mu_1 - \mu_0$ is the difference between the two hypotheses means or the shift in the process mean from μ_0, and

$$S_t = \sum \left[X_i - \left(\mu + \frac{\Delta_1}{2}\right)\right]$$

is the cumulative sum of the sample deviations of the data from the average of the two means. However, when it is desirable to detect a process mean shift in either direction from the target, as with predictions of significant movements away from neutral corridors, one could use a pair of one-sided cusum charts to monitor the process for upward and downward shifts, separately. Let the two off-target mean values of interest be denoted by $\mu_1 > \mu_0$ and $\mu_2 < \mu_0$ with a Type II error β and a Type I error 2α. Note, that the Type II errors for the upward and downward shifts can be different, but with the neutral corridors, the selection of two different β values is not meaningful.

By combining two one-sided decision criteria, it can be shown that the upper and lower control limits for the centered cumulative sum, $\Sigma(X_i - \mu)$, are

$$UCL = \frac{\sigma^2_x}{\Delta_1} \ln\left(\frac{1-\beta}{\alpha}\right) + t\left(\frac{\Delta_1}{2}\right)$$

$$LCL = \frac{\sigma^2_x}{\Delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + t\left(\frac{\Delta_2}{2}\right)$$

where $\Delta_1 = \mu_1 - \mu_0$ and $\Delta_2 = \mu_2 - \mu_0$. Because the equations for the control limits are functions of the sample size t, the upper and lower control limits are linear trend lines for the cusum chart. An example of the resulting cusum chart is shown in Figure 3 and the corresponding neutral corridor is shown in Figure 2.

The general model for the cusum chart above is that of a sequential sampling scheme shown in Figure 3. There are two alternative hypotheses for rejection of the null hypothesis:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_0 + \Delta_1,$$

and $$H_2: \mu = \mu_0 + \Delta_2.$$

Only the upper and lower trend lines are used for the detection of shifts in the mean. In every case cusums are cumulative sums of sample deviations from the target process mean μ_0, where μ_0 for a neutral corridor as defined earlier, is the value of the center line of the neutral corridor.

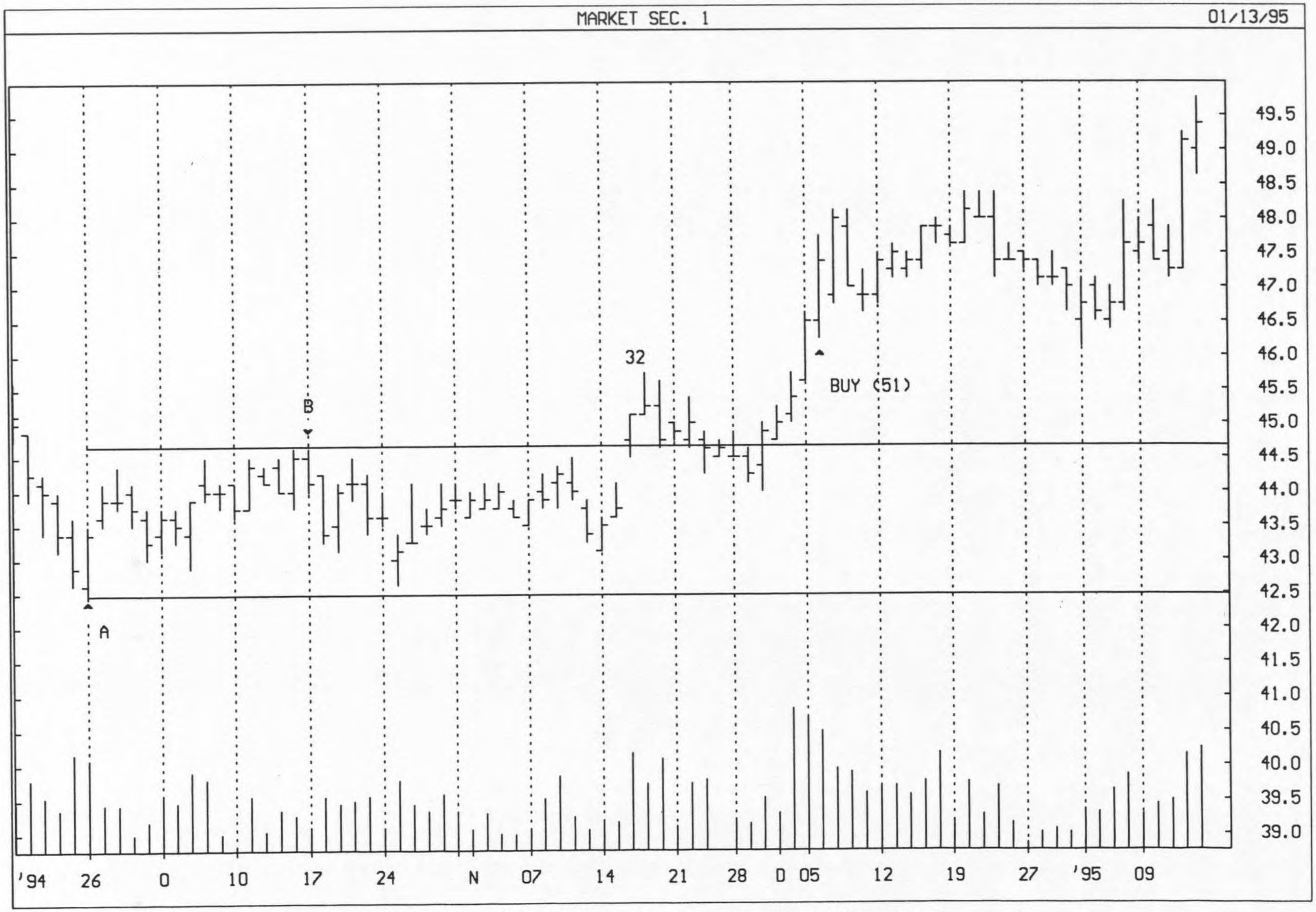

Figure 2. Neutral Corridor formed by Market Security 1

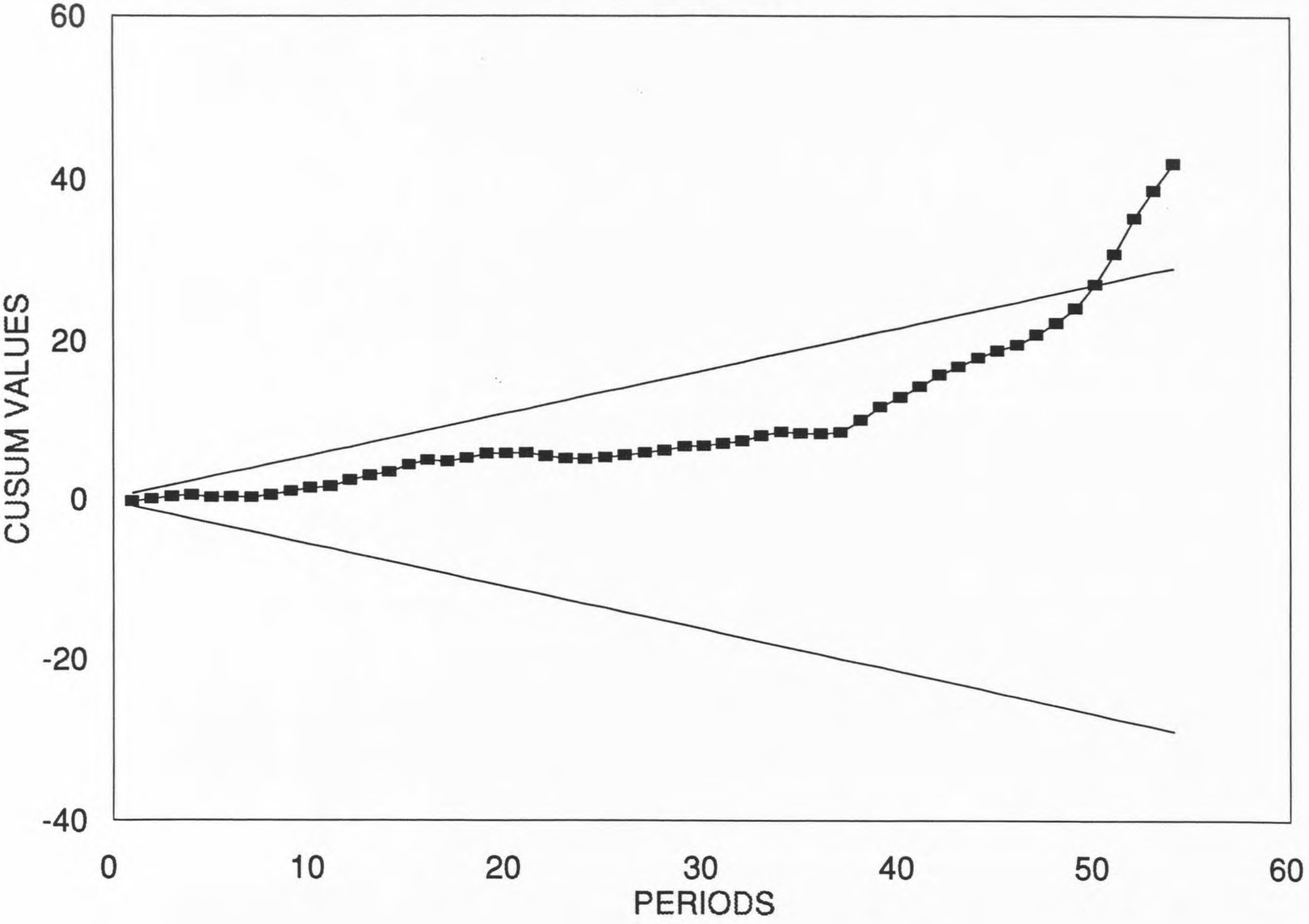

Figure 3. Traditional Cusum Chart for Market Security 1

Standardized Cusum Chart

Cusum charts can be used for individual measurements, when these measurements exhibit a normal distribution with mean μ and standard deviation σ. Given the individual measurements, X_i, $i = 1, 2, 3, \ldots, t$, we can standardize them by applying the transformation

$$Z_i = \frac{X_i - \mu}{\sigma}$$

The sum of Z's will have a normal distribution with mean = 0 and variance = t {since the variance of each Z_i is 1.0, the variance of t independent Z_i's will be (t)(1.0) = t [Devor, 2]}. We can write the cusum as follows:

$$S_t^* = \frac{\Sigma Z_i}{\sqrt{t}}$$

where, S_t^* also follows a normal distribution. This allows us to plot the S_t^* on a control chart with constant control limits (usually ± 3) and a centerline of zero. Thus a control chart simpler than the traditional cusum control chart can be constructed for the cusum defined above. Figure 4 shows a standardized cusum chart which has been plotted on the same neutral corridor data (Figure 2) used for plotting the traditional cusum chart.