Significant breakout detection and projection from a neutral corridor during time series analysis
by Weidong Weng

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

Abstract:
The neutral corridor is an important pattern formed within a time series. By studying neutral corridors, we can obtain two pieces of valuable information for decision making by engineering managers. One is when the time series has broken out of the neutral corridor, the other is to determine how far the time series will move upward or downward after it breaks out.

A pattern comparison method has been developed for breakout signal detection. The neutral corridor can be regarded as a group of observations with a similar pattern. Once the pattern is violated, the neutral corridor will not last, and the time series breaks out of the corridor. By using ModelWare software, a similarity value was computed to quantitatively measure the similarity between new observation and the original data called Reference Data. This similarity value was then compared with a predetermined threshold value. If the similarity value exceeded the threshold value, a breakout signal was detected. A total of 20 neutral corridors were detected for the breakout signal and correct and timely signals were obtained for all of them.

To project the movement out of the neutral corridor, the author has derived a mathematical model that reveals the cause-and-effect relationship between the projection and its previous consolidation. A neutral corridor can be viewed as an energy building zone. Once the time series breaks out, the energy will be transferred and drive the time series to move. Further research concerning this mechanism needs to be conducted. However, the model was validated by 26 neutral corridors formed by historical market security data. It was observed that 23 out of the 26 obtained less than 7% percent error for the projection.

SIGNIFICANT BREAKOUT DETECTION AND PROJECTION FROM A NEUTRAL CORRIDOR DURING TIME SERIES ANALYSIS

by

Weidong Weng

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Industrial & Management Engineering

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

June 1996

APPROVAL

of a thesis submitted by

Weidong Weng

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 Jun '96

(Signature) Date

Approved for the Department of Mechanical and Industrial Engineering

Dr. Victor Cundy 6/28/96

(Signature) Date

Approved for the College of Graduate Studies

Dr. Robert Brown

(Signature) Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a masters degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying 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 July 8, 1996

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

List of Tables

List of Tables - Continued

List of Tables - Continued

List of Figures

List of Figures - Continued

ABSTRACT

The neutral corridor is an important pattern formed within a time series. By studying neutral corridors, we can obtain two pieces of valuable information for decision making by engineering managers. One is when the time series has broken out of the neutral corridor, the other is to determine how far the time series will move upward or downward after it breaks out.

A pattern comparison method has been developed for breakout signal detection. The neutral corridor can be regarded as a group of observations with a similar pattern. Once the pattern is violated, the neutral corridor will not last, and the time series breaks out of the corridor. By using ModelWare software, a similarity value was computed to quantitatively measure the similarity between new observation and the original data called Reference Data. This similarity value was then compared with a predetermined threshold value. If the similarity value exceeded the threshold value, a breakout signal was detected. A total of 20 neutral corridors were detected for the breakout signal and correct and timely signals were obtained for all of them.

To project the movement out of the neutral corridor, the author has derived a mathematical model that reveals the cause-and-effect relationship between the projection and its previous consolidation. A neutral corridor can be viewed as an energy building zone. Once the time series breaks out, the energy will be transferred and drive the time series to move. Further research concerning this mechanism needs to be conducted. However, the model was validated by 26 neutral corridors formed by historical market security data. It was observed that 23 out of the 26 obtained less than 7% percent error for the projection.

CHAPTER 1

INTRODUCTION

In recent years, it has become increasingly clear that pattern analysis has gained a lot of power in the time series forecasting area. A neutral corridor is a pattern representing a range wherein supply and demand are at a standoff. The study of neutral corridors in this paper results in two pieces of valuable information.

1. Detection of significant breakout signals out of the neutral corridor. This will answer the question when and in which direction the time series will move out of the consolidation area and that nontrivial movement will ensue.
2. Projection of movements. This will answer how far the time series will move resulting from the previous neutral corridor.

Neutral Corridor

Although different methods are used to defined a neutral corridor, only two horizontal lines are needed to construct a neutral corridor. The top line is called "upper boundary" or "resistance line", and the bottom one is called "lower boundary" or "support line".

To draw these two horizontal lines, two points are needed to be identified, which are called "valid top" and "valid bottom" respectively. These two points are usually determined in the earlier period of a consolation. They represent the highest and lowest value that the time series has reached. See Figure 1 on the following page for a neutral corridor in which A stands for the "valid top" and B represents the "valid bottom".

A neutral corridor can last for one week, several months, or several years. Relatively long consolidations will eventually result in spectacular breakouts and create pleasant opportunities. Only neutral corridors of at least 50 points were of interest.

Breakout versus Shakeout

Breakout means that the time series moves out of either boundary of the neutral corridor significantly, and it will sustain this movement later on. This movement will keep going, until at least, it is away from the broken boundary at a distance of more than the width of the neutral corridor. Sometimes it may turn back toward the boundary instantly, but very soon, it will take off again without oscillating within the neutral corridor.

Shakeout means that the time series moves out of the neutral corridor insignificantly, and will turn back to the neutral corridor before it is away from the broken boundary at a distance of the width of the neutral corridor, and continue to oscillate within the neutral corridor. Figure 1 also shows this kind of shakeout.

Energy Accumulation and Transfer

When time series oscillate within neutral corridors, battles are being waged between sellers and purchasers, as for example, producers and consumers. The buyers move the prices to the upper boundary of a neutral corridor but fail to keep the rally going. The sellers then

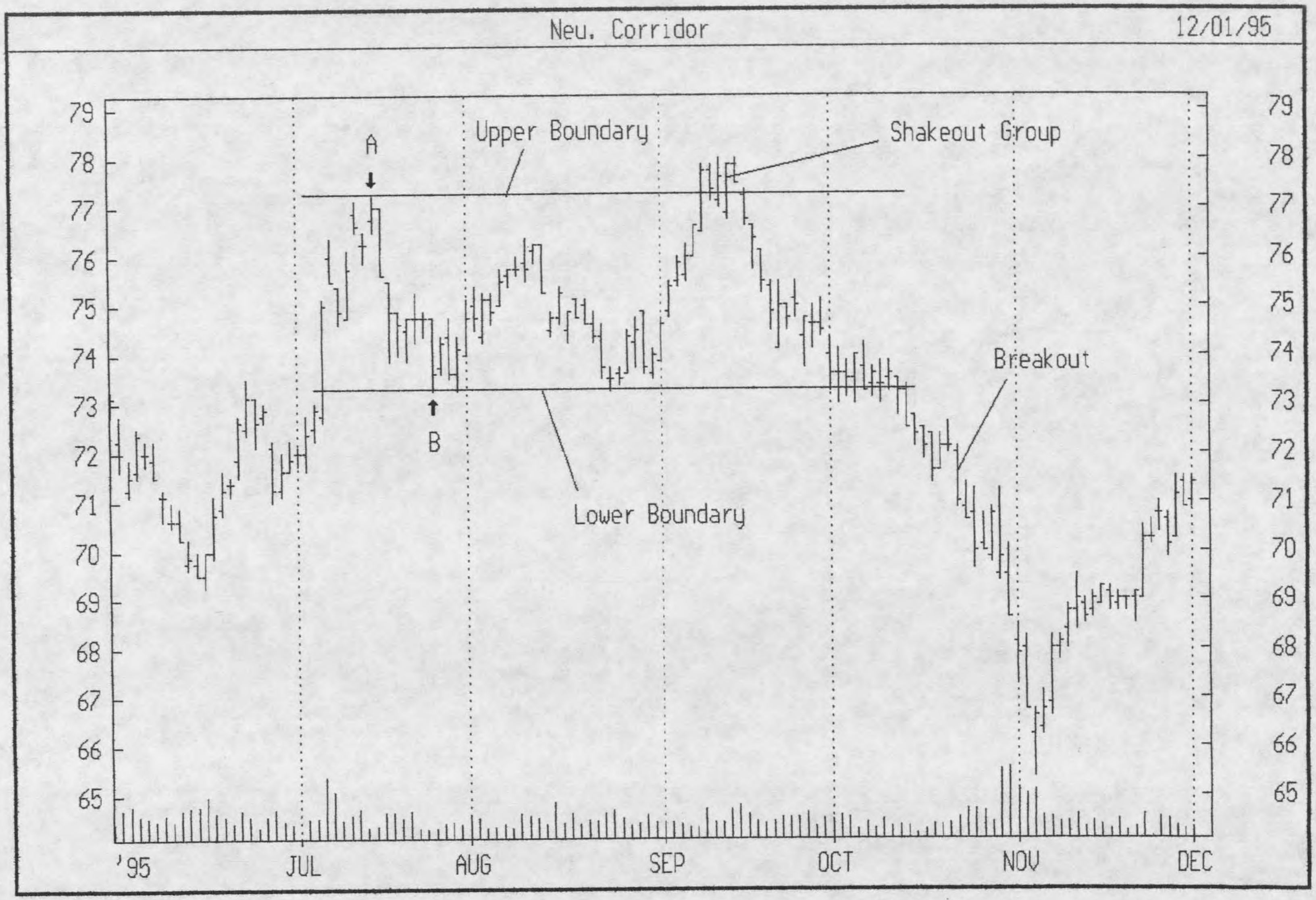

Figure 1. An Example of a Neutral Corridor

take over and move prices back down to the lower boundary of the neutral corridor. This phenomenon is analogous to twisting a spring. The more twisting, the more energy will be accumulated. Finally, one group or the other dominates and succeeds in moving the series out of the neutral corridor. The energy therefore starts to dissipate. The series is driven to move upwards or downwards until the energy is totally exhausted. At the same time, the time series has reached a new top or bottom. Price moves are usually proportional to their preceding consolidation area. There is a cause-and-effect relationship existing, which will be discussed further in Chapter 4.

Pattern and Pattern Violation

Twins resemble each other, since they have similar patterns in appearance, which is exhibited by the eyes, nose, face, and so on. It is not only for these organs to be resembled, but also they should combine in the similar way.

This is exactly what happens in the neutral corridor. Instead of the eyes, nose, and face, some specific variables were chosen to reflect the pattern of a neutral corridor. The data of these variables within a neutral corridor were stored as a reference library. When new data became available, they were compared to the reference data regarding the pattern. The *similarity* value will be computed which gives the information about how similar the pattern of the current observation is compared to the original pattern.

If the similarity is less than a predetermined threshold value, the original pattern has been violated, the time series has broken out of the neutral corridor. Otherwise there is no pattern violation, and the neutral corridor will continue.

Data for Analysis

The author tried to find neutral corridors formed by engineering data through various ways such as searching in the library and the Internet. Unfortunately, this effort was finally given up because of the extreme difficulties to access the data sources. Even some production data were available, but they failed to form neutral corridors due to their short time periods. In view of this, data from market securities were used for analysis [9]. Using data from market securities has some other advantages:

1. Many neutral corridors can be identified.
2. Accurate, timely, and readily available.
3. Ease of verification.
4. Possibility to utilize some commercial software for data analysis.

CHAPTER 2

LITERATURE REVIEW

A neutral corridor was first named by Schillings [7] for a pattern that represents supply and demand at a standoff.

Some other analysts call it *consolation area* [1] or *rectangles* [5]. "Every price trend, takes a breather from time to time. It is a period of indecision when the pressure of buyers and sellers balance each other out" [2]. A neutral corridor itself doesn't provide an opportunity for the decision making. "Once it comes out of the consolidation, an opportunity does exist, since a direction has been indicated". Unfortunately, "there is no way to know how long the trading range is going to last."

Currently, the author has not found that any quantitative analysis has been done in detecting significant breakout signals from a neutral corridor. However, Pring [5] found some heuristic rules were used by some forecasters.

The same situation is found for the Breakout Projection. The energy building in a neutral corridor has been observed by forecaster Arms. "Price move is usually proportional to their preceding consolidation area" [1]. However, not any mathematical model has been developed to reveal the cause-and-effect relationship between the neutral corridor and its following breakout movement.

The author was then encouraged by Dr. Schillings to do some basic research on both of the breakout detection and projection using quantitative techniques; hence, this thesis.

CHAPTER 3

TOOLS FOR ANALYSIS

ModelWare

Background

ModelWare is a generic modeling tool developed by Teranet IA Incorporated, Vancouver, BC, Canada, which can be applied for signal processing, prediction, decision/control and diagnostic/classification problems [5]. At the heart of modelware is Teranet's Universal Process Modeling (UPM) algorithm.

How ModelWare works

Figure 2 shows the mechanism of how ModelWare works. First, variables to be used to model the system are selected. Then data are collected for these variables from the system and a Reference Library is built. This Reference Library can be thought of as the "captured knowledge" of how the system or process behaves under known conditions. ModelWare requires no equations to describe how variables of the system relate to one another. The system data itself contains all the information about how the system works.

Once the Reference Library has been built, ModelWare is ready to predict system behavior for new input data known as Observed Data. Modelware makes a prediction for all variables for each observation in turn. The prediction then is compared to the observation.

If the observation deviates from the prediction, the system is no longer operating within its range of historical performance.

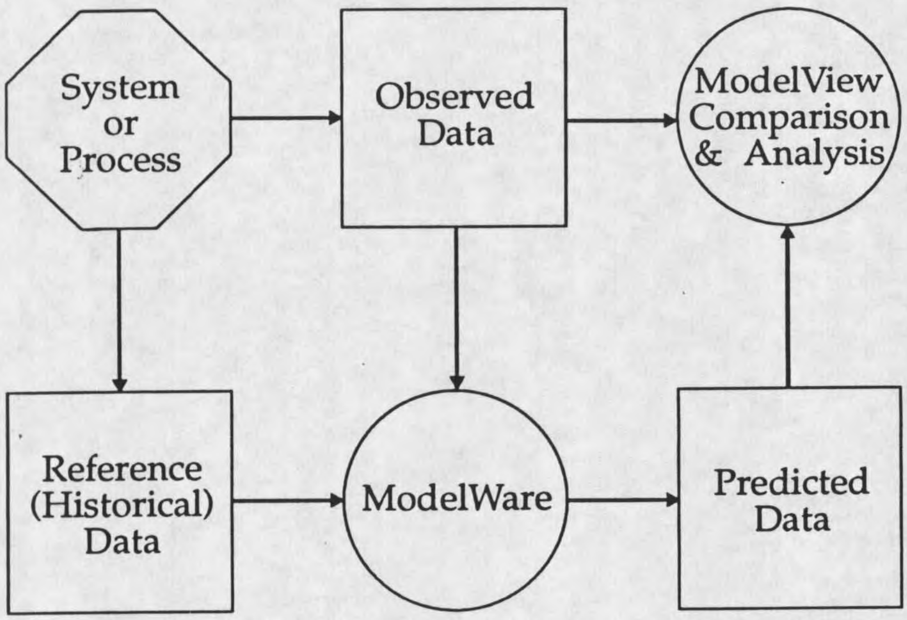

Figure 2. How ModelWare Works

The UPM Algorithm

The UPM algorithm consists of six basic steps:

1. Get the next observation to model from the input file and compute the similarity between this observation and all the examples in the Reference Library file. The measure of similarity is a single number between 0.0 and 1.0 which indicates how mathematically similar one set of numbers are with another. A value of 1.0 means that the two sets of data numbers are identical.

2. Pull out from the Reference Library those examples that are most similar to the input observation. The minimum and maximum number of examples pulled out can be set by the user.
3. Compute a single weighting coefficient for each example pulled out from the Reference Library. This coefficient is a measure of the contribution the example makes to the final prediction.
4. Compute the predicted value for each data variable in turn, by summing, in turn, all the products for a particular variable's value in each of the most similar Reference Library examples with the corresponding coefficient associated with each most similar example. Figure 3 shows an example of this procedure.

	Ref. Image 1	Ref. Image 2	Ref. Image 3
Element 1	3.0	6.5	4.0
Element 2	2.5	1.9	2.7
Element 3	4.0	4.8	5.6
Element 4	1.6	2.1	2.6
Element 5	7.1	9.0	9.9

●

0.6	Coefficient of Ref. Image 1
0.3	Coefficient of Ref. Image 2
0.1	Coefficient of Ref. Image 3

Prediction for element 1 = (3.0 x 0.6) + (6.5 x 0.3) + (4.0 x 0.1) = 21.50
Prediction for element 2 = (2.5 x 0.6) + (1.9 x 0.3) +(2.7 x 0.1) = 7.47
Prediction for element 3 = (4.0 x 0.6) + (4.8 x 0.3) + (5.6 x 0.1) = 17.36
Prediction for element 4 = (1.6 x 0.6) + (2.1 x 0.3) + (2.6 x 0.1) = 16.16
Prediction for element 5 = (7.1 x 0.6) + (9.0 x 0.3) + (9.9 x 0.1) = 32.25

Figure 3. Calculation of Predicted Value

5. Validate the prediction by computing the similarity of the prediction with the input observation.

6. Repeat steps 1 through 5 until all input observations have been modeled.

Several Important Terms

Eight important terms are used in the paper.

1. System

   System refers to any object, or collecton of objects, that can be represented by sets of numbers. Industrial processes naturally fall into this category because the parameters that control the process are usually monitored by physical sensors that provide numerical outputs as a function of time. As a less obvious example, a diverse group of financial indicators could also be taken as a system. Notice that this definition of a system does not imply that there is any causal or dependent relationships among any of the measured parameters.

2. Element (variable)

   Element refers to a measured parameter within the system.

3. Image

   Image refers to a set of element measurements taken at a single moment in time. Given simultaneous values for all elements, it is possible to completely specify the state of a system at that moment. System behavior over time can then be represented as a time-stamped series of images.

4. Reference Library

Reference library refers to a collection of images that describes the system as it operates under known and proper conditions. This library will typically consists of ten to hundreds of images completely encompassing all of the extreme values of the system's operation in each element. Any subsequent images taken of the system in operation should be similar to one or more images in the reference library, providing the system is still operating normally, or at least in a previously observed fashion.

5. Model

A mathematical formula created by ModelWare for each image of the observed data file. Each formula produces a predicted, or modeled, data image. Running a model file produces the formula and the modeled data. The formula is derived by analyzing the reference data of the system being modeled. This formula is not available to the ModelWare user.

6. Model file

A file that contains information necessary to create a model. After a model file is run, it also contains the modeled, or predicted, data created by the model.

7. Modeling

Modeling means the action of synthetically generating values for the elements in a "predicted image", based on some combination of an observed image and the reference library. This definition does not necessarily require any understanding of the system itself and there are no equations that represent the nature of the system mathematically.

8. ID file

The identity file is used to create a model. This file contains information about the data elements. ID file can be edited to determine whether or not an element will be used for modeling. For example, Element 1 can be in the Reference Data, but it can be changed to "passive" by editing the ID file, and Element 1 will then not be used to make a model.

Application

ModelWare was used to detect significant breakout signals. First, the similarity value of a new image was computed with respect to the reference data. Then, this similarity value was compared to the threshold value. If the threshold value was exceeded, a breakout signal was detected.

Shewhart's Control Charts

Background

Shewhart's control charts were first developed by Walter Shewhat in 1931 for process quality control [3]. As a Bell Labs statistician, Shewhart established from the very beginning the overarching philosophy that drives the control chart concept.

There are two process characteristics that are of general interest in the statistical process control scenario, the mean level of the process and the amount of variation in the process.

Construction of Control Charts

The following simple steps define the procedures necessary to set up X-bar and R control charts.

1. Calculation of sample means:

$$\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n ,$$

where X_{ij} is the jth measurement, j = 1, ..., n, in the ith sample and where n = sample size.

2. Calculation of the spread or dispersion within the ith sample R_i:

$$R_i = X_{largest} - X_{smallest} .$$

Because of its simplicity and ease of calculation, the range, R_i, is commonly employed as a measure of within-sample variability.

3. Calculation of the grand sample average which is the arithmetic average of all the available sample averages. This grand sample average is an estimate of the process mean μ_x and becomes the centerline of the control chart: