COMPARISONS OF NEURAL NETWORKS, SHEWHART $\bar{x}$, AND CUSUM CONTROL CHARTS UNDER THE CONDITION OF NONNORMALITY

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Junsub Yi, B.S., M.B.A.

Denton, Texas

August, 1997

Yi, Junsub, Comparisons of Neural Networks with Shewhart $\overline{x}$ and CUSUM Control Charts under The Condition of Nonnormality. Doctor of Philosophy (Management Science), August, 1997, 180 pp., 51 tables, 23 figures, references, 74 titles.

As the marketplace becomes more competitive, companies are increasingly focusing on quality to maintain a competitive advantage. The use of statistical methods leads to improvements in quality and an associated increase in productivity. One of the most widely used statistical methods to improve quality and productivity is control charts. A fundamental assumption in the development of control charts is that the underlying distribution of the monitored quality characteristic is normal. However, in many industrial situations, the process means of the quality characteristics may not be normally distributed, in which different combinations of skewness and kurtosis of the quality data create an infinite number of potentially complicated nonnormal situations. If a standard symmetrical $\overline{x}$ control chart is used when the normality assumption is violated, the error rates can be significantly affected. The need to better define approaches for handling nonnormal data is a partial motivation for this research.

In this study, neural networks are developed under conditions of nonnormality as alternatives to standard control charts, and their performance is compared with those of standard $\overline{x}$ and CUSUM control charts. The performance of $\overline{x}$ control charts is also compared with that of CUSUM charts. The study examines the effects of nonnormality, mean shifts, and sample size on the performance of the three methods to detect out-of-control

states. A heuristic procedure for specifying the parameters used in the neural network configurations is also discussed in detail. These problem specific parameters include learning rate, momentum, and number of hidden layers and neurons. The neural network approach presented in this study offers a competitive alternative to the existing control schemes. Extensive comparisons showed that the neural networks appeared to be a better control procedure for detecting sudden changes in the process mean.

COMPARISONS OF NEURAL NETWORKS, SHEWHART $\bar{x}$, AND CUSUM CONTROL CHARTS UNDER THE CONDITION OF NONNORMALITY

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Junsub Yi, B.S., M.B.A.
Denton, Texas

August, 1997

ACKNOWLEDGMENTS

I wish to express a debt of gratitude to my advisor, Victor R. Prybutok, for his academic guidance, advice, and patience. Additionally, my thanks are extended to the other members of the committee, Drs. Howard R.Clayton, Maliyakal D. Jayakumar, and Ian Parberry for their constructive comments and encouragement throughout the period of this research. Thanks should also go to Mr. David L. Mitchell for helpful discussions in improving the readability of the dissertation.

Special thanks are due my wife, Taelim, my parents, family, and friends for their support and love.

Finally, I must acknowledge University of North Texas for the environment and resources that made this work possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In 1925, Shewhart introduced the control chart in the *Journal of the American Statistical Association.* Since then, control charts have become a powerful tool to monitor process means and variability (Keats and Hubele 1989). Control charts, when used appropriately, help the production manager monitor processes to determine whether they are "in control" or "out of control." When a process being monitored is found to be "out of control," an improvement process can begin, reducing the variability of the process means (Aft 1988).

Among the most widely used control charts are the $\overline{x}$ charts. A fundamental assumption in the development of $\overline{x}$ control charts is that the underlying distribution of the monitored quality characteristic is normal. However, in many industrial situations, the process means of the quality characteristics may not be normally distributed random variables. For example, we may know that the distribution of the process mean is nonnormal due to the small sample size used, coupled with the fact that the underlying distribution of the data is known to be exponential, Weibull, or some other nonnormal distribution. Knowing the underlying distribution, we can derive the sampling distributions of $\overline{x}$ and obtain exact probability control limits for the control charts. However, this approach could be difficult in cases in which the underlying distribution changes over time. In other cases, we may not know the form of the underlying distribution, and then our only choice may be to use the usual control charts based on normality. In such cases, the nonnormality of the process means can have a significant effect on the performance of the $\overline{x}$ charts (Ryan 1989). If the standard symmetrical $\overline{x}$ charts are used when the normality assumption is violated, the error probabilities for the procedure can be significantly larger than expected. Therefore, design considerations for the $\overline{x}$ chart must include recognition of the degree of nonnormality of the underlying data (Ramsey and Ramsey 1990; Yourstone and Zimmer 1992).

When the normality assumption is not valid, control chart users may choose from several different courses of action. Rigdon, Cruthis, and Champ (1994) suggested a transformation of the original data that yields an approximately normal distribution. They recommended the use of one family of transformations, the Box-Cox power transformations, which are often used in regression and analysis of variance models. The transformed data might be more closely modeled by the normal distribution, and the standard control charts could then be used. There are two primary problems with the transformation approach. First, there are difficulties in determining the appropriate transformation, and second, when a transformation has been chosen, there are difficulties in implementing the procedure and justifying the transformation (Montgomery 1991a; Yourstone and Zimmer 1992). Another means of dealing with nonnormality is to increase the sample size until the distribution of the data approaches normality. Some research supports this course of action in handling nonnormal situations. Schilling and Nelson (1976) concluded that it would be reasonable to use the standard control charts with four or more observations in each sample if the normality assumption is not severely violated. Their conclusion is supported by other research dealing with the design of control charts under nonnormality (Montgomery 1980). However, the use of larger samples to overcome severe violations of the normality assumption may not be operationally feasible and is obviously costly.

A third approach to dealing with nonnormal data is to modify the $\overline{x}$ chart. One way to modify the chart is to employ asymmetric control limits. Tagaras (1989) and Yourstone and Zimmer (1992) developed asymmetric $\overline{x}$ control charts with unequal distances from the center line as alternatives to the traditional $\overline{x}$ control chart. However, utilization of this approach requires either knowledge of the underlying distribution of the data or a good approximation. Finally, the cumulative sum (CUSUM) control chart is an alternative to the $\overline{x}$ chart for controlling nonnormal data (Ho and Case 1994). Lashkari and Rahim (1982) developed an economic design of the CUSUM chart which assumes that the observations obtained from the process are nonnormally distributed. Although many researchers have developed remedies for dealing with nonnormality, including those mentioned above, none of these is completely satisfactory. In this study, neural networks are developed as alternatives to the standard control charts to handle the nonnormality problem.

This study is organized as shown in figure 1. In this chapter, the purpose of the study, the research problem, and the significance of the study are addressed. Chapter 2 addresses the background of the $\overline{x}$, R, and cumulative sum control charts. Some important neural network concepts such as processing elements (PEs), feed-forward networks, training, testing, and backpropagation are also discussed. Chapter 3 is a discussion of the reported findings, limitations, and significance of the prior studies. Chapter 4 presents data generation and the methodologies used in comparing the performance of $\overline{x}$ and CUSUM control charts and neural networks for dealing with nonnormal processes. In chapter 5, results of extensive simulation experiments are reported. Finally, in chapter 6, results and limitations of the study are discussed.

Purpose of the Study

In many manufacturing situations, the quality characteristic of interest is a random variable whose density function depends upon one or more parameters of the product quality and often has a nonnormal distribution. In such situations, the use of standard control charts, which are theoretically based on the normality assumption, could have a significant effect on the probabilities (Type I and II errors) associated with the control limits and may incorrectly indicate the process control state (Rahim 1985). In this "contaminated" nonnormal situation, neural networks might be good alternatives to the standard control charts for detecting out of control situations. As a first step, this study develops neural networks as alternatives to the standard $\overline{x}$ control charts. A heuristic procedure for specifying the parameters used in the neural network configuration is discussed in detail. These problem specific parameters include learning rate, momentum, and number of hidden layers and neurons. This study also examines and compares the performance of standard $\overline{x}$ and CUSUM control charts and neural networks for process means when dealing with nonnormal data.

Research Problem

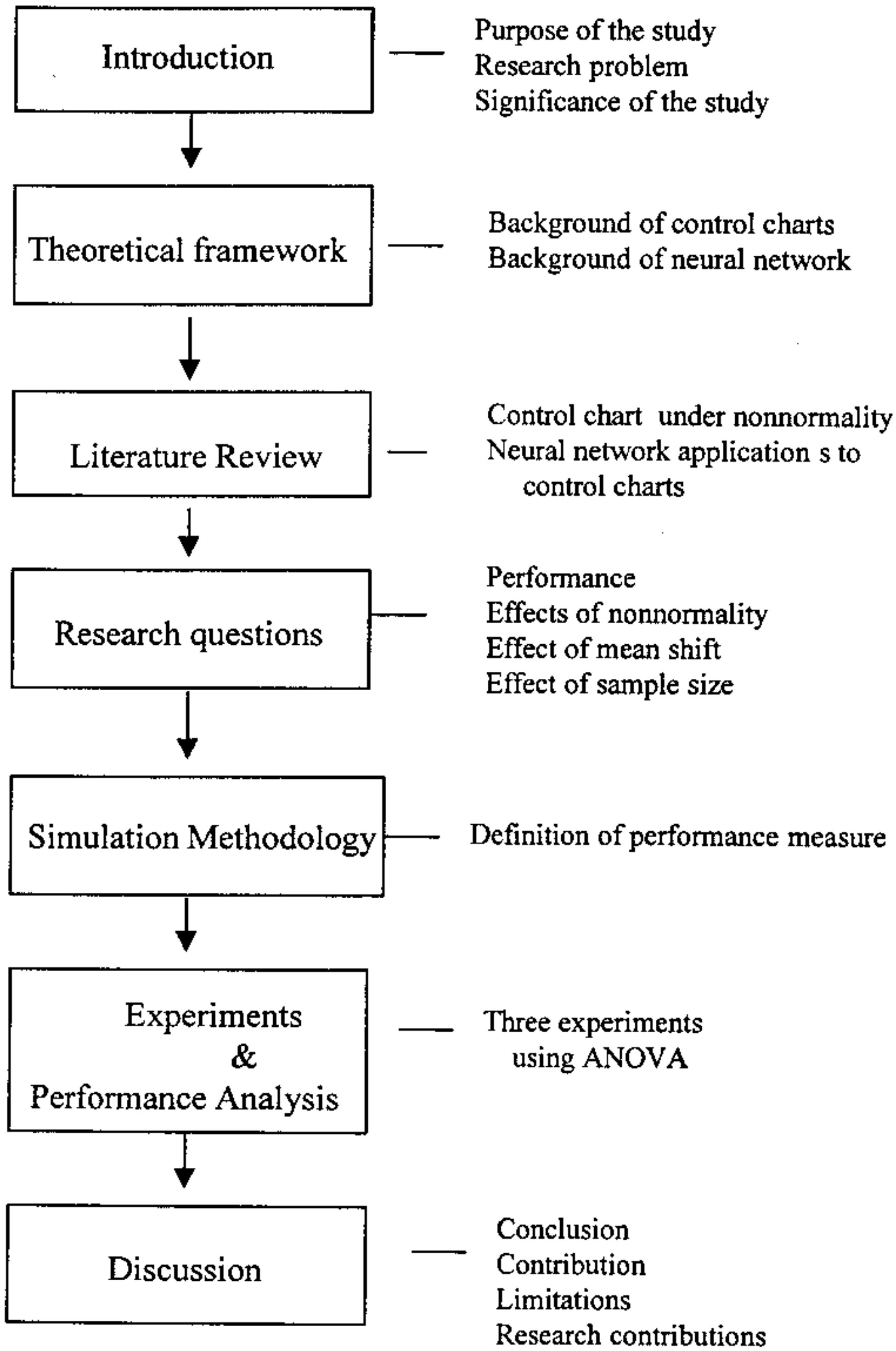
When setting the upper and lower control limits for a $\overline{x}$ chart, the normality assumption should be justified by the central limit theorem, which essentially states that $\overline{x}$ will tend toward a normal random variable regardless of the underlying distribution of data as the sample size becomes large. However, there are cases where either the underlying distribution of the original data is nonnormal, or the sample size is not large enough to apply the central limit theorem. In these situations, the problem is to find an economic design of the $\bar{x}$ chart. The economic design of a $\bar{x}$ chart involves determining the optimal parameters for the chart -- the sample size, the sampling interval, and control limit coefficients that minimize the costs associated with designing the chart (Rahim 1985; Saniga 1989; Ho and Case 1994). A number of researchers have developed economic designs for $\bar{x}$, CUSUM, and asymmetric control charts for dealing with the nonnormality problem. However, none of these designs is uniformly satisfactory. One of the main reasons is that numerous forms of nonnormality are possible for real-world quality characteristics. For example, a process may have been recorded only during a certain interval, resulting in a truncated distribution. Another example of nonnormality is a mixture-distribution, which occurs when products from two or more separate sources are mixed (Rahim 1985).

Significance of the Study

Neural networks are computer-based simulations of living nervous systems that have a mathematical basis. Recently, these computer-dependent neural networks have emerged as alternatives to traditional statistical techniques as the availability and capability of computer systems have increased. The application of neural networks to the quality control area is relatively new and is limited in scope. Although several researchers (Davis and Illingworth 1989; Pugh 1989, 1991; Velasco 1993; Hwarng and Hubele 1993a, 1993b; Prybutok, Sanford, and Nam 1995; Hwarng 1995) have developed neural networks as alternatives to the standard $\bar{x}$ control charts for monitoring process means, these studies have been limited in number and coverage. All of the previous research has involved developing neural networks for monitoring only process means, which is equivalent to the standard $\overline{x}$ control chart. Also, the previous research was done only under the condition of normality.

In a manufacturing environment, however, the characteristics of the process may not follow a normal distribution. Under nonnormality, different combinations of skewness and kurtosis of the quality data create an infinite number of potential complicated nonnormal situations. These rather significant limitations on developing generalizable approaches for nonnormal data are a partial motivation for this research into the application of neural networks to control charts. In this study, neural network alternatives to the standard control charts are developed under conditions of nonnormality. The performance of these neural networks is compared with those of the corresponding $\overline{x}$ and CUSUM control charts. The development of a new and better technique for the analysis of quality characteristics will allow companies to enhance product quality and reduce internal and external failure costs. If the neural networks show potential for improving product quality and/or reducing costs, the results will provide guidance as to the practical use of the neural network as an alternative to or a supplementary technique for standard control charts.

figure 1. Contents of study



CHAPTER II

THEORETICAL FRAMEWORK

Background of Control Charts

Introduction

As the marketplace becomes more competitive, companies increasingly focus on quality for competitive advantage. From the customers' point of view, a quality product or service is one that meets or surpasses their needs and expectations at a price they are willing to pay. Quality characteristics for a product or service include such things as strength, appearance, durability, reliability, and maintainability. However, producers are also concerned that quality of product or service meets their prespecified process requirements. Scientific and systematic methods to monitor, analyze, and control processes allow simultaneously the meeting of the customers' need and expectations and the process quality requirements.

Variation exists in every process and every sample of measurements. That is, the measured quality of a product or service is subject to process variation as a result of chance and assignable causes. We refer to the situation related only to chance causes of variation, as "in-control." Variation within this stable pattern is inevitable. Variation outside of this chance variability is called an "out-of-control" situation. Non-chance variation is attributable to assignable causes that may be the result of such events as an improperly adjusted machine, an operator error, or defective raw material.

Statistical methods can be employed to identify and pinpoint the causes of such unusual variation, whether it is for a manufacturing or general business process. The use of statistical methods leads to improvements in quality and an associated increase in productivity. Statistical process control (SPC) includes all the statistical methods for monitoring, analyzing, and controlling manufacturing and manufacturing related processes. SPC is the application of statistical methods to the measurement and analysis of variation in any process (Juran and Gryna 1993). SPC, taken in a narrow definition, constitutes only those statistical procedures that are used in the actual control of a process. However, in a wider sense, SPC includes the entire body of knowledge and the entire toolbox of statistical procedures required to measure, analyze, and characterize a process (Evans and Lindsay 1989).

A statistical control chart is a graphical method used to monitor various aspects of a process and determine whether the process is in-control or out-of-control. Ideally, an out-of-control state is detected immediately after its occurrence. Conversely, we would like to have as a low false alarm rate as possible. A false alarm rate is the probability that the monitoring process detects an out-of-control state when the process is in-control. The use of statistics allows the establishing of an appropriate balance between the two. When control charts are implemented, it is important that they be used where the potential for cost reduction is substantial. Control charts should also be used where trouble is likely to occur (Ryan 1989). Modern computer technology makes it easy to implement control charts, in real time, for numerous processes. The general form of several control charts ($\bar{x}$, R, c, and p) contains a center line (CL), which represents the average value of the quality characteristic corresponding to the in-control state. The other two horizontal lines are the upper control limit (UCL) and the lower control limit (LCL). A point that should be emphasized is that there is no connection or relationship between the control limits for the control charts and the specification limits for the process. The control limits are driven by the natural variability of the process and are based on past process data. The specification limits are the process requirements specified by the producers. Therefore, it is possible that the process might be in-control based on control charts but not be capable of meeting the process requirements (Ryan 1989).

There are various types of control charts, including $\overline{x}$, CUSUM, R, p, c, moving average, multivariate, and others. The type of control chart that is appropriate for a situation depends on the quality characteristic that is measured and the type of data being collected. Quality characteristics such as the weight, diameter, or concentration of a product yield variables in the form of continuous data. Quality defects such as scratches or blemishes on a product result in what are referred to as attribute (discrete) data. The $\overline{x}$ and R charts are used for measuring variables data, whereas p and c charts are used for measuring attribute data. The multivariate control chart is used when the characteristics of the process consist simultaneously of more than one variable (Tracy, Young, and Mason 1992; Hayter and Tsui 1994). The $\overline{x}$ and R charts are used most often in practice, whereas multivariate and CUSUM charts are among the least used.

Control charts are based on the assumption that data observations are randomly and independently collected. Another important assumption underlying the use of a $\overline{x}$ control chart is that the sample mean is normally distributed. If a control chart is constructed based on samples of at least four or five and the distribution of x is reasonably symmetric and bell shaped, the central limit theorem states that distribution of $\bar{x}$ will not differ greatly from a normal distribution and therefore the normality assumption is appropriate. However, if the underlying distribution of the data is severely nonnormal, the effectiveness and efficiency of the control charts are subject to question. In this chapter, the background of the $\bar{x}$, R, and cumulative sum (CUSUM) control charts is addressed.

$\bar{x}$ Control Charts

The $\bar{x}$ chart is the most widely used among the control charts for monitoring the central tendency of a process quality characteristic. Its relative ease of computation and application and its excellent performance in recognizing out-of-control conditions account for a portion of its popularity (Wadsworth et al. 1986; Ryan 1989). The $\bar{x}$ chart is used to monitor variation in the process mean and to indicate whether shifts in the mean occur. As noted before, the general form of a $\bar{x}$ control chart contains the CL, UCL, and LCL. The center line and two control limits are defined as follows (Grant and Leavenworth 1988; Montgomery 1991b):

$$\text{Center Line} = \mu$$

$$\text{Upper Control Limit} = \mu + z_{\alpha/2}\sigma_{\bar{x}} = \mu + z_{\alpha/2}\sigma/\sqrt{n}$$

$$\text{Lower Control Limit} = \mu - z_{\alpha/2}\sigma_{\bar{x}} = \mu - z_{\alpha/2}\sigma/\sqrt{n}$$

where the process characteristic $\bar{x} = (x_1 + x_2 + \ldots + x_n)/n$ has a normal distribution with mean μ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. In practice, μ and σ are unknown and are estimated from preliminary in-control samples. Suppose that m samples are obtained, each of size n. Generally, m ranges from 20 to 30, and the number of observations in each sample is recommended to be 4 to 6, if available (Quesenberry 1993). Let $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$ be the averages for each sample, where $\bar{x}_i$, i=1,2, ..., m, is the mean of the process characteristic for each of the m samples. The best estimator of μ, the true mean of the process, is $\bar{\bar{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m) / m$. It is very common to use 3 for the $z_{\alpha/2}$ value, and the resulting control limits are called 3-sigma limits. It should be noted that the value 3 can be changed to fit the quality requirement for the process. The UCL and LCL representing 3-sigma limits for sample data are as follows (Ryan 1989):

$$\mathrm{CL} = \bar{\bar{x}}$$

$$\mathrm{UCL} = \bar{\bar{x}} + 3\hat{\sigma}/\sqrt{n}$$

$$\mathrm{LCL} = \bar{\bar{x}} - 3\hat{\sigma}/\sqrt{n}$$

where $\hat{\sigma} = \bar{R}/d_2$.

We note that the quantity

$$A_2 = \frac{3}{d_2\sqrt{n}}$$

is a constant that depends on the sample size, so it is possible to rewrite the equations above as

$$\mathrm{CL} = \bar{\bar{x}}$$

$$\mathrm{UCL} = \bar{\bar{x}} + A_2 R$$

$$\mathrm{LCL} = \bar{\bar{x}} - A_2 R.$$

If a computed sample mean falls between the LCL and UCL, the process is considered to be in-control. If not, it is considered to be out-of-control.

In general, averages are used to develop a $\bar{x}$ control chart because individual observations are not as sensitive as an $\bar{x}$ or CUSUM control chart in detecting mean shifts

(Juran and Gryna 1988). However, using means is not always possible or practical. If items selected from an assembly line are produced at such a slow rate as to preclude the forming of subgroups, individual observation control charts without subgrouping are more appropriate (Ryan 1989).

R Control Charts

R stands for range, the difference between the maximum and the minimum values in a sample. Before interpreting patterns on a $\overline{x}$ chart, it is important first to determine whether or not the corresponding R chart is in-control because, when an R chart indicates an out-of-control condition, interpretation of the $\overline{x}$ chart is not meaningful. However, some assignable causes show up on both the $\overline{x}$ and R charts. If both $\overline{x}$ and R charts exhibit out-of-control conditions, then the best strategy is to eliminate the assignable causes in the R chart first. In many cases, this will also eliminate the out-of-control condition on the $\overline{x}$ chart (Montgomery 1991b). The range (R) is easier to conduct than the sample standard deviation and provides nearly as much information about the sample variation for sample sizes less than 10. Since control charts are typically used to monitor small samples, the sample range is generally used to measure sample variation. An R control chart is used to monitor the range of the characteristic of interest in the process. In general, the narrower the range, the better the process. Let $R_1, R_2, \ldots, R_m$ be the ranges of the m samples. The average range is $\overline{R} = (R_1 + R_2 + \ldots + R_m) / m$.

The established relationship between the range of a sample from a normal distribution and the standard deviation of that distribution is $W = R/\sigma$, where W is called the relative range, with an expected value of d_2. An estimator of σ is $\hat{\sigma} = R/d_2$. The standard deviation of W, say d_3, is a known function of the sample size n. Thus, since R = Wσ, the standard deviation of R is $\sigma_R = d_3\sigma$. Since σ is unknown, we may estimate σ_R by $\hat{\sigma}_R = d_3\frac{\overline{R}}{d_2}$. The center line and control limits for sample data are defined as follows (Montgomery 1991b):

$$\text{UCL} = \overline{R} + 3\hat{\sigma}_R = \overline{R} + 3d_3\frac{\overline{R}}{d_2}$$

$$\text{CL} = R$$

$$\text{LCL} = \overline{R} - 3\hat{\sigma}_R = \overline{R} - 3d_3\frac{\overline{R}}{d_2}.$$

If we let $D_3 = 1 - 3\frac{d_3}{d_2}$ and $D_4 = 1 + 3\frac{d_3}{d_2}$,

then

$$\text{UCL} = \overline{R}D_4$$

$$\text{CL} = R$$

$$\text{LCL} = \overline{R}D_3$$

Cumulative Sum (CUSUM) Control Charts

A major disadvantage of all Shewhart control charts is that they only use the information about the process contained in each of the plotted points separately, and they ignore any information given by the entire sequence of points. As a result, Shewhart control charts are relatively insensitive to small shifts in process characteristics. It is also suggested that other criteria be applied to Shewhart control charts, such as runs rules, which attempt to incorporate information from the entire set of points into the decision procedure (Hwarng and Hubele 1993a; Prybutok, Sanford, and Nam 1995). However, the use of supplementary runs rules requires experience to identify patterns in the process and reduces the simplicity and ease of interpretation of the Shewhart control charts. The CUSUM control chart is an effective alternative to the Shewhart control chart for detecting small shifts (0.5σ to 1.5σ) in process means (Ryan 1989; Montgomery 1991b).

The CUSUM control chart is a chronological plot of the cumulative sum of deviations of process means. It is well known that the CUSUM chart identifies small shifts in a process much faster than the standard $\overline{x}$ control chart (Juran and Gryna 1993). The CUSUM chart focuses on a target value rather than the actual average of process data. Its limits are typically displayed using a v-shaped mask that is constructed and moved as a new point is plotted. The control limits are neither parallel nor fixed. CUSUM charts directly incorporate all the information in the sequence of sample values by plotting the cumulative sums of the deviations of the sample values from a target value. The steps in constructing the CUSUM chart are as follows (Ryan 1989: Montgomenry 1991b).

First, find $S_i = \sum_{j=1}^{i} (\overline{x}_j - \mu_0)$

where S_i = the cumulative sum up to the ith sample

$\overline{x}_j$ = the average of the jth sample

μ_0 = the target value for the process mean.

If the process remains at the target value μ_0, the cumulative sum defined above should fluctuate stochastically around zero. In fact, the cumulative sum is a random walk process with mean zero. However, if the mean shifts upward to some value $\mu_1 > \mu_0$, then an upward or positive drift will develop in the S_i. Conversely, if the mean shifts downward, then a downward or negative drift will develop in the S_i. Using a V mask to monitor a process signals an out-of-control condition when all of the cumulative points are not within the arms of the V mask. The mask is moved backward after each point is plotted, with the slope of the arms of the mask remaining constant.

The main step in the design procedure for a CUSUM control chart is to determine the lead distance d and the angle θ of the V mask.

$$d = \left(\frac{2}{\delta^2}\right) \ln\left(\frac{1-\beta}{\alpha}\right) \quad \text{and}$$

$$\theta = \tan^{-1}\left(\frac{\Delta}{2} A\right)$$

where α = the probability of a false alarm

β = the probability of failing to detect a shift in the mean

Δ = the shift in the process mean that it is desired to detect

$$\delta = \frac{\Delta}{\sigma_{\bar{x}}}$$

A = a scale factor relating the vertical scale unit to the horizontal scale unit

CUSUM control charts have several advantages relative to Shewhart control charts. First, as mentioned, CUSUM charts are more effective in detecting relatively small shifts in the process mean. CUSUM charts will detect small shifts approximately twice as fast as a Shewhart chart. Second, it is relatively easy to detect the starting point of a process shift by examining the plotted data. Third, the CUSUM chart has proved to be a good alternative to the $\bar{x}$ chart when dealing with nonnormal data (Lashkari and Rahim 1982). On the other hand, some disadvantages of the CUSUM charts should be noted. First, the CUSUM control chart is not as good as Shewhart control charts in detecting large shifts in process mean (more than 2σ). Second, diagnosis of patterns on CUSUM charts is difficult since the assumption in pattern recognition is uncorrelated sequence of points.

Background on Neural Networks

Introduction

For neural network researchers in the quality control area, the objective is to develop practical neural network models that can stand as alternatives to or as supplements to standard control charts. In a computer integrated manufacturing (CIM) environment, the neural network approach can be implemented in software and applied to the same data that would normally be used for a standard control chart. In this situation, an operator may not routinely monitor the control chart but, rather, would be alerted by an out-of-control signal generated by the neural network pattern recognition algorithm. Eventually, real time feedback to the system and automated process control will be used to increase the efficiency and productivity of a process (Hwarng and Hubele 1993b; Velasco and Rowe 1993).

A neural network is a computing system consisting of two primary components: (1) a set of processing elements (PEs: the neural network equivalent of a biological neuron) and (2) the interconnections among the PEs. There are three types of PEs in a neural network: input, hidden, and output PEs. These PEs can be organized into many different types of neural networks. One of these, the feed-forward neural network, is especially suited for classification applications and is used in this study. A typical feed-forward neural network is composed of an input layer, one hidden layer, and an output layer.

To better understand this work, it is important to provide background on neural networks. This section contains a summary of the relevant theory on neural networks. The general descriptions provided within the subsections below (Neural Network Models, Training and Testing, Backpropagation Network, Proposed Neural Network Model, Advantages of Neural Networks, and Limitations of Neural Networks) represent a synthesis and summary of material from numerous fundamental references (Rumelhart, McClelland, and PDP Research Group 1986; Wasserman 1989; Caudill 1990; NeuralWare 1991; Burke and Ignizio 1992; Nam 1993).

Artificial Neurons and Architecture

The human brain is the most complex computing device known to man. The brain consists of tens of billions of neurons densely interconnected. In an artificial neural network, the unit analogous to this biological neuron is referred to as a processing element. A neural network is defined as a computing system made up of a number of simple, highly interconnected processing elements, which processes information by its dynamic state response to external inputs. Processing elements, the neural network equivalent of neurons, are generally simple devices that receive a number of input signals, and based on those inputs, either generate a single output signal or do not. Figure 2 shows the activities that take place at the processing element. The process of computing that takes place in a neural network can best be understood by examining these activities.

The output signal of an individual processing element is sent to many other processing elements (and possibly back to itself) as input signals via the interconnections between processing elements. A processing element is generally an extremely simple device that has a number of input signals and a single output signal. Each input signal (x_i) is assigned a relative weight (w_i), so the effective input to the processing element is the weighted total input, or $I = \sum (w_i x_i)$ for all input signals. The simplest kind of processing element compares this weighted sum input to an arbitrary threshold. If the input is greater than the threshold value, the processing element fires or generates an input signal. If the input is less than the threshold, the processing element does not fire and no output is generated.

Neural Network Models

Figure 3 shows a fully connected feed-forward neural network in which all processing elements in any one layer are connected to all processing elements in the next higher layer. The arrows indicate the flow of information. A typical feed-forward neural network has three layers of neurons: input, hidden, and output layer neurons. Input values in the input layer are weighted and sent to the hidden layer. Neurons in the hidden layer produce outputs based on the sum of the weighted values sent to them. In the same way, the values in the hidden layer are weighted and sent to the output layer to produce the final results.

Training and Testing

There are two main phases in the operation of a neural network: training (learning) and testing. Training is the process of adjusting the connection weights in response to the amount of error that the network is generating. Unlike traditional expert systems in which knowledge is made explicit in the form of rules, neural networks tend to generate their own rules by learning. Learning is achieved through a learning rule that adapts or changes the connection weights of the network in response to the inputs and optionally, the desired outputs of those inputs. Testing is the process that examines whether the neural network has generalized over the data set rather than memorizing.

There are two kinds of learning: unsupervised learning and supervised learning. In unsupervised learning, only input stimuli are shown to the network, and the network organizes itself internally so that each hidden processing element responds strongly to a different set of input stimuli or closely related group of stimuli. In supervised learning, for each input stimulus, a desired output stimulus is presented to the system, and the network gradually configures itself to achieve that desired input/output mapping.

Backpropagation Network

Currently the most widely used technique to train the feed-forward neural networks is backpropagation which uses the generalized delta rule. As shown in figure 4, the typical backpropagation neural network always has an input layer, an output layer, and at least one middle (hidden) layer. Each layer is fully connected to the succeeding layer.

The generalized delta rule procedure involves two phases: a forward phase and a backward phase. During the forward phase, the input is presented and propagated forward through the hidden layer to an output value, which is compared with the desired output to compute the error. After the appropriate weight changes are made, the backward phase propagates the errors back to the input layer from the output layer through the hidden layer to update the connection weights.

Perhaps most unique to neural networks, and an increasingly fruitful research area, is the ability of the system to generalize. If a neural system has been trained on a set of data which consists of examples of object attributes and associated evaluations, its ability to generalize is measured by its performance on objects that lie outside the training set. Obviously, generalization ability depends on the adequacy of the training set. The algorithm for backpropagation takes the following steps:

1. Apply the input vector to the network input.

2. Compute the output of the network.

3. Compare the output to the desired output.

4. Compute the error between the network output and the desired output.

5. Adjust the weights of the network in a way that minimizes the error.

6. Repeat steps 1 through 5 for each vector in the training set until the error for the entire set is acceptably low.

Proposed Neural Network Model

The following notation will be used to describe the backpropagation least mean square (BPLMS) algorithm with the generalized delta rule applied to the feed-forward neural network used in this study (Archer and Wang 1993; Philipoom, Rees, and Wiegmann 1994):

x = the [(5+1) x 1] vector consisting of a bias PE ($x_0=1$) and the five input PEs;

$w_{i,r}$ = the connection weight from the ith input PE to the rth hidden PE, r (number of hidden neurons) = 1, ..., h, i (number of input neurons) = 0, ..., 5 and $w_{i,0}$ is the weight from the bias PE;

v_{rk} = the connection weight from the ith hidden PE to the kth output PE, i = 1, ..., h, k (number of output neurons) =1, 2, 3 and v_0 is the weight from the bias PE to the output PE;

$NET = w_{i,0} + \sum(w_{i,r}x_i)$ at each hidden layer PE;

z_r = $[1+\exp(-NET)]^{-1}$, r = 1, ..., h, is the output value of the ith hidden layer PE;

$NET = v_0 + \sum(v_{rk}x_r)$, r = 1, ..., h, at the output PE;

$F(NET) = [1+\exp(-NET)]^{-1}$, is the output value of the output node.

y_j = desired output for each input vector, j=1, ..., n

n = number of training set pairs

$E = (½) \sum [y_j\text{-}F(NET)]^2$, j = 1, ..., n, is the global error function.

The proposed neural network model (figure 5) consists of an input layer with five input PEs, one hidden layer consisting of h PEs, and an output layer with two output PEs for "in-control" and "out-of-control." The inputs to the neural network include one bias value and the five observations in each subgroup created by Monte Carlo simulation. During training, a set of n pairs of input vectors and corresponding outputs, {[**X**(1), y(1)], [**X**(2), y(2)], ..., [**X**(n), y(n)]} is presented to the network one pair at a time. A weighted sum of the inputs, $NET = w_{i,0} + \sum(w_{i,r}x_i)$, i = 1,...,h, is calculated at each hidden layer PE. Each hidden layer PE then uses a sigmoid transfer function to generate an output, $z_i = [1+\exp(-NET)]^{-1}$, i = 1, ..., h, between 0 and 1. The outputs from each of the h hidden-layer PEs, along with the bias input, are then sent to the output PE, and again a weighted sum is calculated, $NET = v_0 + \sum(v_i x_i)$, i = 1, ..., h, k = 1, 2 at each output layer neuron. The weighted sum becomes the inputs to the sigmoid transfer function of the output PEs. The resulting output, $F(NET)= [1+\exp(-NET)]^{-1}$, is then scaled to determine whether the processes are "in-control" or "out-of-control." At this point, the second phase of the BPLMS algorithm, adjustment of the connection weights, begins.

The weights are adjusted so as to minimize the global error, $E=(½) * \sum[y_j\text{-}F(NET)]^2$. There are four basic steps in the weight change procedure. They are as follows:

1. Compute how fast the global error, E, changes as the activity of the output unit [F(NET)] is changed:

$$\delta E/\delta F(NET) = F(NET) - y = e.$$

2. Compute how fast the error, E, changes as the total input (NET) received by an output unit is changed:

$$\delta E/\delta NET=[\delta E/\delta F(NET)][\delta F(NET)/\delta NET]$$

$$=e*F(NET)*[1-F(NET)]$$

3. Compute how fast the error E changes as a weight v_{ik} on the connection into an output unit is changed:

$$\delta E/\delta v_{ik}=(\delta E/\delta NET)(\delta NET/\delta v_{ik})$$

$$=e*F(NET)*[1-F(NET)]*z_i$$

4. Finally, compute how fast the error, E, changes as the activity of a unit in the hidden layer (z_i) is changed:

$$\delta E/\delta z_i=(\delta E/\delta NET)(\delta NET/\delta z_i)$$

$$=e*F(NET)*[1-F(NET)]*v_{ik}$$

Once the difference between actual and desired outputs is determined from step 1, step 3 is used to determine how much to change each weight v_{ik} in the output layer. Then step 4 is used to determine how much to change each weight $w_{i,r}$ in the hidden layer. The process of forward propagation of inputs and backpropagation of errors with corresponding updates of weights continues until training of the neural network is considered complete.

The second step in the development of a neural network model is testing, which is the process that examines whether the neural network has generalized over the data set rather than memorizing. It is important that a neural network be able to generalize beyond the data set with which it was trained. Testing consists of presenting new data pairs to the network and measuring the model's performance. In this study, the performance of the standard $\overline{x}$

and CUSUM control charts and neural networks is measured by the average run length (ARL).

Advantages of Neural Networks

There are a number of reasons why the most often used parametric classification and prediction tools do not always provide satisfactory results. Parametric statistical techniques require that the functional form of the relationship among variables be known and specified. In most real world cases, the functional forms are not known, and in very many cases, no existing form can be determined (Dutta and Shekhar 1988). In addition, parametric techniques require that a variety of assumptions concerning the data be satisfied. These assumptions, which often go untested and are often violated, include such items as data normality, error independence, and homoscedasticity (Tam and Kiang 1992). This leads to abuse of these techniques and erroneous inferences (Marquez et al. 1991). Another reason for a lack of traditional statistical model success rests on the fact that many real-world problems require complex statistical models and sophisticated data transformations, which are often beyond the expertise of the practitioner to execute successfully.

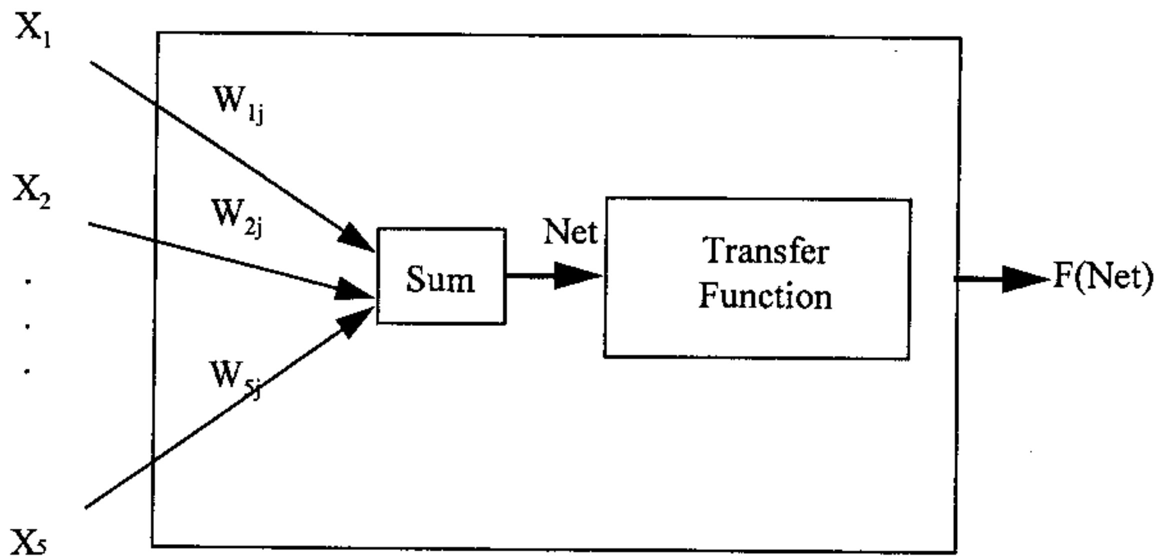
Neural networks, unlike parametric statistical techniques, do not require that a list of assumptions be satisfied. For example, it is not necessary that the functional relationships between independent and dependent variables be specified to a neural network. In fact, it is not necessary that a domain specific model exist for a neural network to discern relationships and build its own model (Dutta and Shekhar 1988). Neural networks are able to learn complex linear and nonlinear relationships from data. As a result of this ability, they tend to be suitable for two basic types of tasks: recognition and generalization (Duliba 1991).

Included in the generalization category are both classification and prediction tasks (Dutta and Shekhar 1988). Both of these tasks can involve forecasting the future state of one or more dependent variables, based on the current and past values of one or more independent variables. Neural networks are also relatively immune to such problems as multicollinearity, heteroscedasticity, and outliers (Raghupathi, Schkade, and Raju 1991). These characteristics of neural networks make them suitable for classification and prediction tasks (Raghupathi, Schkade, and Raju 1991; Salchenberger, Cinar, and Lash 1992). This suitability has led to a number of studies that have tested the utility of neural networks, often contrasting them with one or more statistical methods, in classification and prediction problems.

Limitations of Neural Networks

A disadvantage of neural networks is that they have the potential problem of converging to a local minimum rather than to a global minimum (Archer and Wang, 1993). The avoidance of such convergence is dependent on the model building procedure. For example, one of the important questions is when to stop the learning process to obtain the best results and how to determine the neural network parameters to obtain the best possible convergence of the learning process. Therefore, like all quantitative methodologies, neural networks require an experienced or informed practitioner. Suggestions for avoiding the local minimum problem include: (1) training the neural network many times, each with different initial random weights; (2) increasing the momentum; (3) lowering the learning rate; and (4) adding extra hidden neurons (Wong 1991). The longer the neural network learns the training data, the closer it comes to memorizing the training data. If the neural networks memorize the training data, they may give the perfect prediction for the training data. However, they may not be able to generalize well on the testing data. One method for preventing memorization is to add a small amount of random noise to the training data set. The neural network design is more of an art than a science. It is often a matter of trying several transfer functions, learning rates, and momentum with a different number of hidden layers and/or hidden neurons.

figure 2. Artificial processing element



1. X_i = input variables
2. W_{ij} = weight
3. $Net = (X_1 W_{1j}) + (X_2 W_{2j}) + \ldots + (X_k W_{kj})$
4. $F(Net) = 1/(1+e^{-Net})$
   for Sigmoid transfer function

Where:
$i = 1,2,\ldots 5$ and $j = 1,2, \ldots, 5$

figure 3. Feed-forward neural network

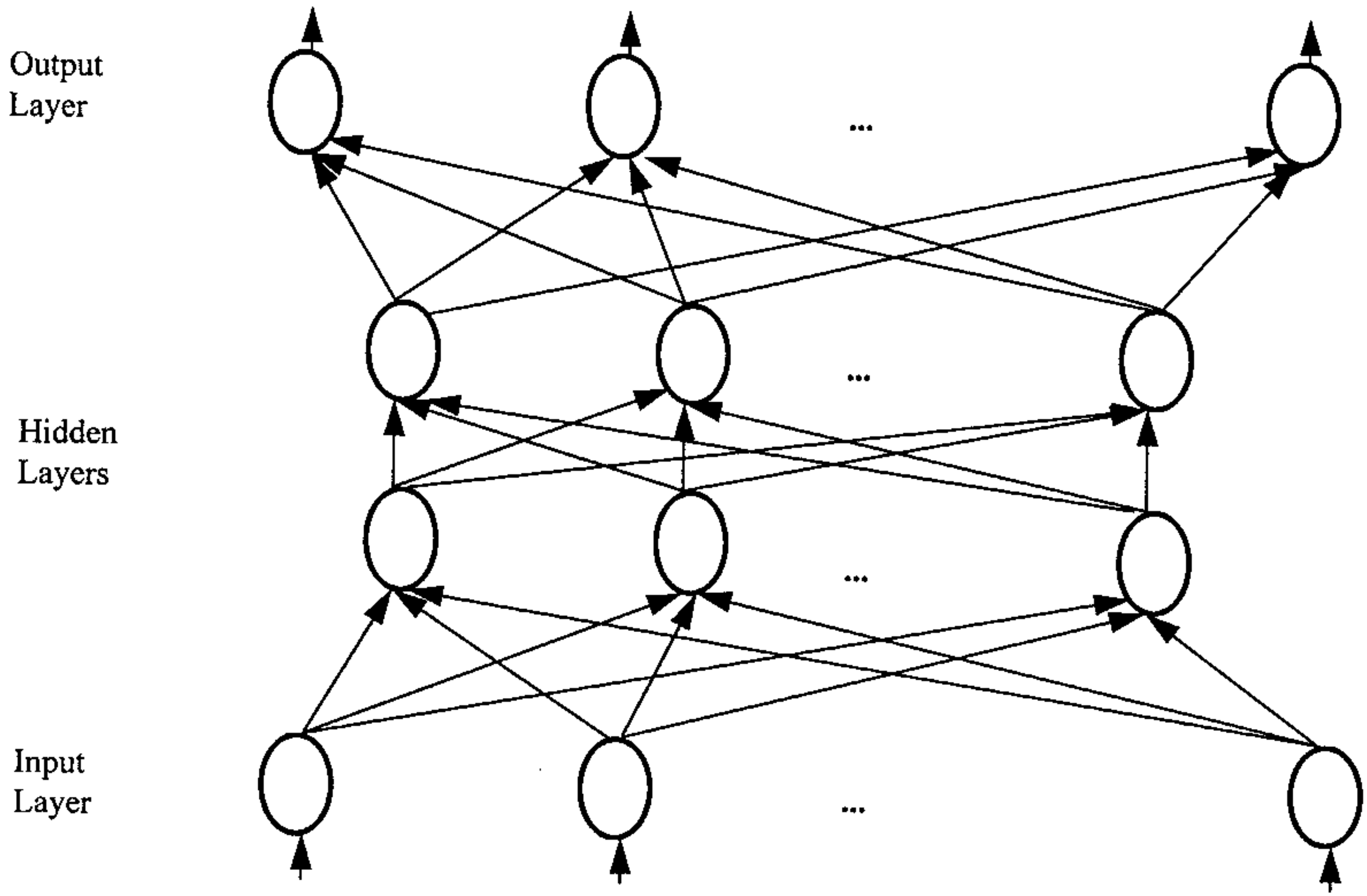


figure 4. Backpropagation neural network

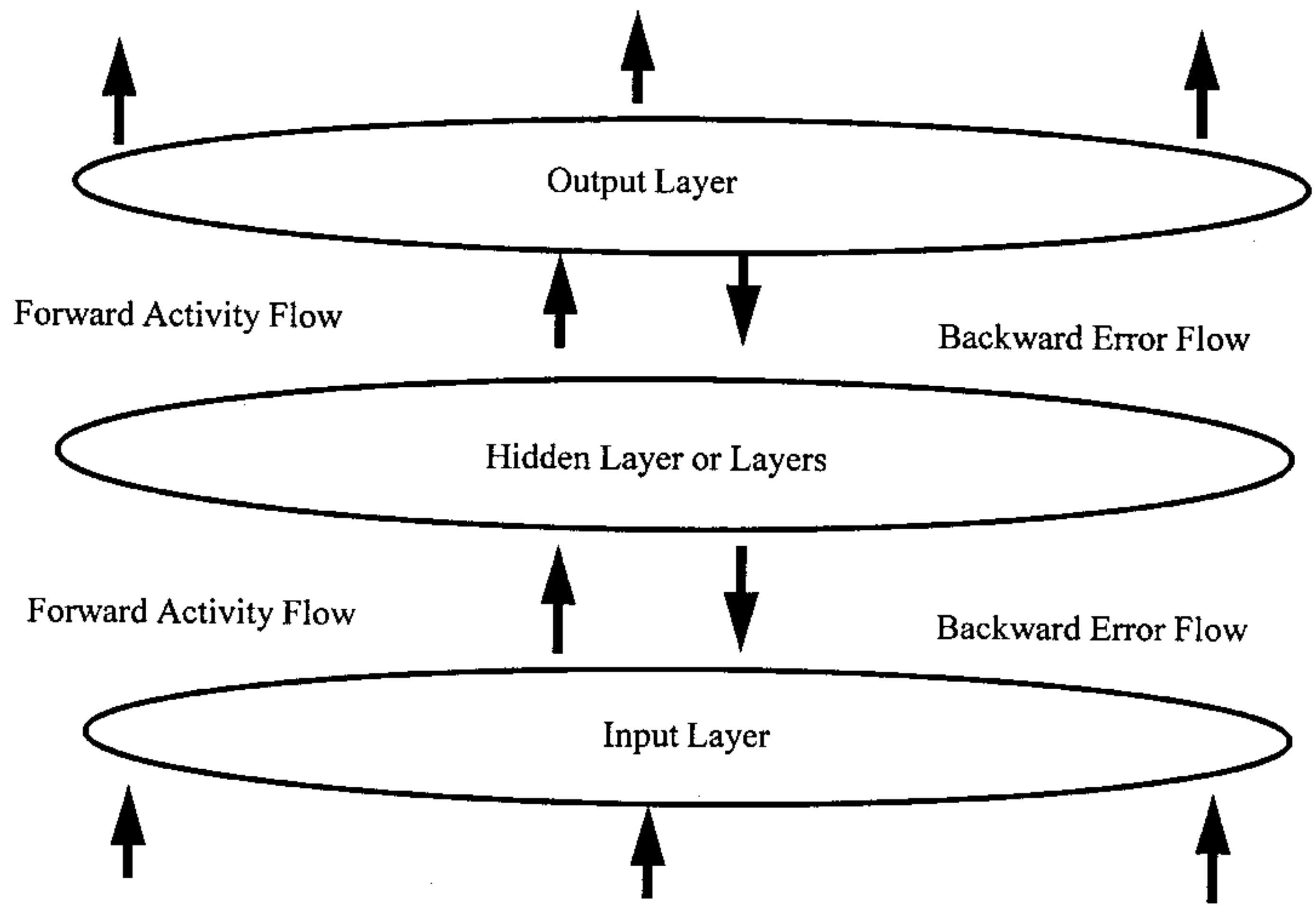
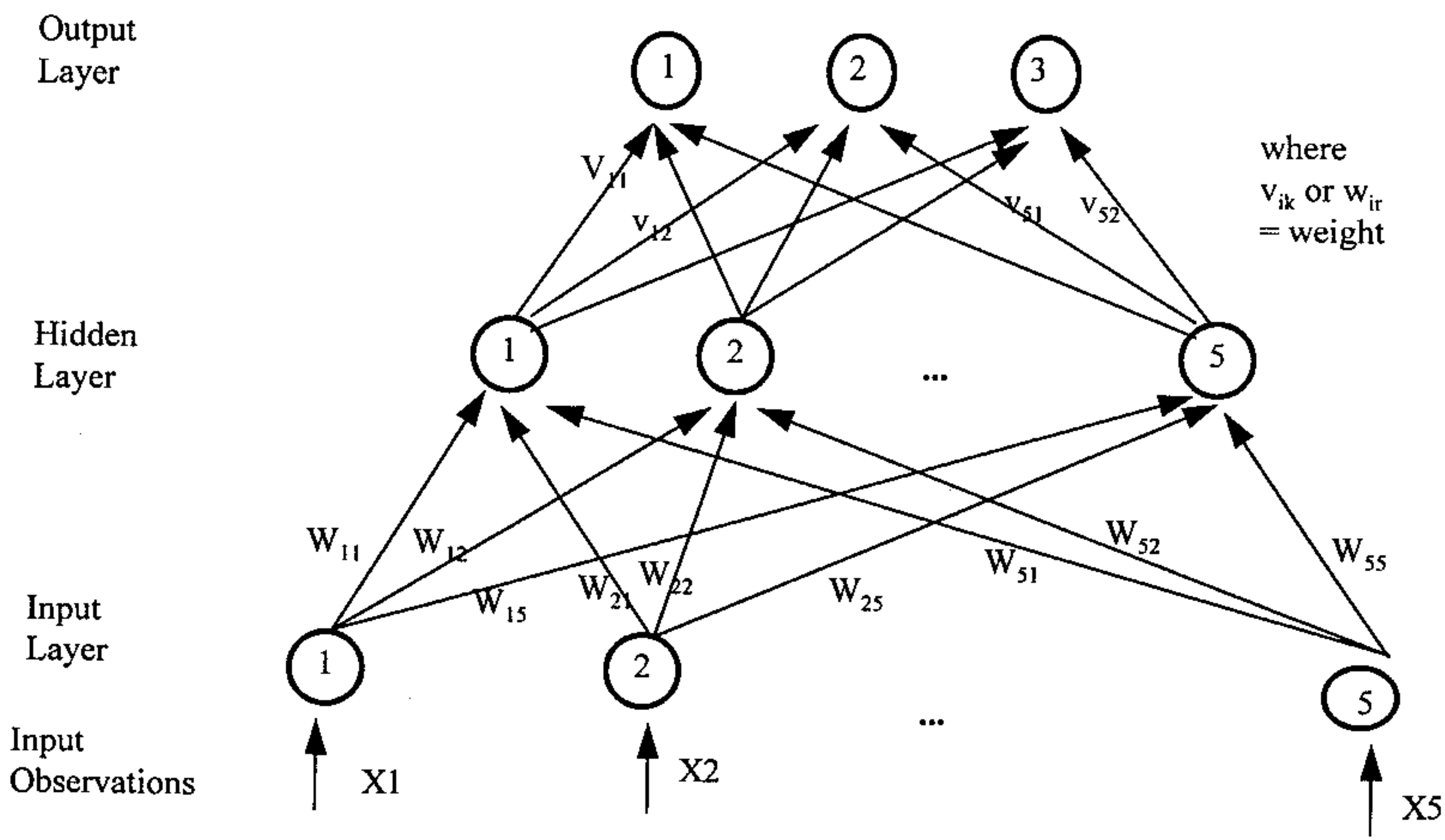


Figure 5. Proposed neural network



CHAPTER III

LITERATURE REVIEW

Introduction

There are two main approaches to control chart design (Saniga 1989). The statistical approach, primarily attributable to Shewhart, prespecifies the Type I error rate and power (which determine control limits). Type I error rate is the probability of the chart indicating that a shift in mean occurs when there is actually no shift. Alternatively, sample size and sampling interval are determined using the Type I error rate and power. This approach is used in the United Kingdom and parts of Western Europe, whereas in the United States the control limits are specified as a multiple of the standard deviation (usually three) of a quality characteristic (Montgomery 1991b). The U.S. type heuristic version of the statistical design of control charts frequently specifies a sample of four or five to satisfy the normality assumption of control charts. The main criticism of the statistical design approach is that it ignores economic performance.

The other approach to control chart design is based on economic criteria. Since Duncan (1956) pioneered his work on the economic design of control charts, numerous new approaches and applications have been developed (Chiu and Cheung 1977; Montgomery 1980; Vance 1983; Woodall 1986; Collani 1986; Rahim 1989). The economic design of control charts involves optimal determination of the design parameters; the sample size, the sampling interval, and the control limits so as to minimize the expected total cost. The main criticism of this approach is that it ignores the statistical performance of the control charts. The basic criticisms of the statistical and economic approaches to control chart design lead directly to a third synthetic approach called the economic statistical design of control charts. Saniga (1989) developed an economic statistical design methodology in which an economic loss function is minimized subject to minimum value of power, maximum value of Type I error probability, and average time to signal at a specified shift in quality characteristics. His approach is a compromise of the statistical and economic design approaches and is readily adapted to any Shewhart control chart. This approach yields both good statistical and economic results, which is something neither approach alone provides.

Both the statistical and economic approaches assume that the underlying distribution of the quality characteristic is normal. However, in many real world manufacturing situations, the normality assumption may not hold. When normality is not a given, the use of standard control charts could have a significant effect on the Type I and Type II errors associated with the control limits and may incorrectly indicate the in or out-of-control state of a process. Type II error rate is the probability of not detecting a given shift in a process. This nonnormality problem has been investigated in a number of studies. In this section, previous research investigating $\overline{x}$, CUSUM, and asymmetric control charts under nonnormality is discussed. Also, several neural network applications to control charts under normality are examined.

Previous Research on Control Charts under Nonnormality

As early as 1967, Burr studied the effects of nonnormality on control limits for $\overline{x}$ and R control charts. When a $\overline{x}$ control chart employs three σ limits under conditions of normality, the Type I error associated with these control limits is .0027. According to Burr, this is also frequently appropriate when the underlying distribution is nonnormal. He presented a set of tables of three σ limits for various nonnormal distributions and concluded that the ordinary control charts based on the normality assumption can be used unless the population is markedly nonnormal. Burr (1967) noted that the standard control charts are robust to the normality assumption and can be used if the assumption is not severely violated.

Since Burr (1967) investigated the effect of nonnormality on control charts, several researchers have investigated the effect of nonnormality on the control limits of the control chart. Schilling and Nelson (1976) presented a numerical method for determining the cumulative probabilities of the distribution of sample means by inverting its characteristic function. By numerically inverting the appropriate characteristic functions, they provided the tables which show the approach to normality for various underlying distributions and samples. The authors examined the uniform, right triangular, gamma, and bimodal distributions. The bimodal distributions of interest to Schilling and Nelson were those that resulted from the mixtures of two normal distributions. Because characteristic functions of mixtures are linear combinations of characteristic functions, this method allowed the authors to investigate such things as contaminated normal populations, previously studied using only Monte Carlo techniques. The authors used the characteristic functions to verify Shewhart's results and to investigate the shape of the distribution of sample means for several underlying populations not easily simulated by Monte Carlo methods (i.e., those not having a finite domain). Their results showed that, in most cases, the normality assumption, based on the central limit theorem, holds with samples of four or more.

Rahim (1985) investigated the effects of nonnormality and measurement error on the design of the $\overline{x}$ control chart. This study successfully incorporated measurement errors (i.e., human errors) in designing an economic model of the $\overline{x}$ control chart under nonnormality conditions. The author developed an economic model of a $\overline{x}$ control chart in which the sample size, sampling interval, and control limits were determined based on minimizing the expected total cost. The optimal value of the design parameters was obtained using a computerized search technique. The author showed that conventional control plans would result in misleading values of the design parameters and the loss-cost if the quality characteristic is not normally distributed and/or subject to measurement errors.

Tagaras (1989) compared the performance of $\overline{x}$ control charts with those of asymmetric control charts. He examined the effect of violations of three common assumptions concerning process and measurement characteristics: (1) equal probabilities of upward and downward shifts in the process mean; (2) constant process variability, independent of the process mean; and (3) perfect measurement of the quality characteristic without errors. He performed a sensitivity analysis regarding the effects of misspecification of the cost and the operating and model parameters (probability of shift, variance, and error of measurement) on the optimal design parameters and the resulting operating loss. His results showed that the probability of a shift in the process mean and the accuracy of measurement have significant effects on the optimal design and the resulting loss. The assumption of constant variance was, however, shown to be relatively unimportant.

Yourstone and Zimmer (1992) examined the effect of nonnormality, as measured by skewness and kurtosis, on the performance of a $\overline{x}$ chart and demonstrated an alternative method under nonnormality, which they called an asymmetric $\overline{x}$ control chart. They employed the generalized Burr distribution to determine the control limits for the asymmetric $\overline{x}$ control chart. The first detailed study of the application of the generalized Burr distribution to control charts was done in 1973 (Austin 1973). The family of Burr distributions covers a wide range of the standardized third and fourth central moments and allows fitting a wide variety of practical data distributions. For example, certain gamma distributions are approximated with Burr distribution. Yourstone and Zimmer (1992) carried out a detailed simulation experiment that compared the average run length (ARL) for symmetric $\overline{x}$ control charts and asymmetric control charts for more than twenty-seven different combinations of skewness and kurtosis. The results showed that the impact of skewness and kurtosis on the performance of a $\overline{x}$ control chart can be substantial. The results also showed that both skewness and kurtosis are important in the consideration of the effect of nonnormality on the operating characteristics of control charts. As the authors noted, the major criticism of their study relates to the applicability of the Burr distribution. The authors recognized, as pointed out by Austin (1973), that the Burr distribution is applicable only to the distributions whose coefficients of skewness and kurtosis are within a particular range of values.

A cumulative sum control chart (CUSUM) is an alternative to the $\overline{x}$ control chart for monitoring process means. Lashkari and Rahim (1982) developed an economic design of the CUSUM chart, assuming that the observations obtained from the process were nonnormally distributed. The nonnormal situation was measured by the skewness and kurtosis of the underlying distribution of the process. The economic design of the CUSUM chart involved the determination of the design parameters that minimize a relevant cost function. The design parameters considered in their study include sample size, sampling interval, reference value, and decision interval. The authors introduced an iterative optimization algorithm for determining the design parameters for the CUSUM chart that minimizes loss for nonnormal processes. The algorithm was also used to determine ARLs and to conduct a sensitivity analysis regarding the effects of nonnormality on the minimum loss for a given set of cost and operating parameters. They concluded that the economic design of CUSUM charts would serve as an alternative to $\overline{x}$ control charts under conditions of nonnormality.

Neural Network Applications for Monitoring Process Means under Normality

Numerous authors have used neural networks to monitor process means when data are normally distributed (Davis and Illingworth 1989; Pugh 1989,1991; Hwarng and Hubele 1993a, 1993b; Velasco 1993; Cheng 1995; Hwarng 1995; Prybutok, Sanford, and Nam 1995). Davis and Illingworth(1989) applied neural networks in a real-time manufacturing environment. The authors tested and evaluated three different neural network paradigms, including Madaline, Hopfield, and backpropagation. The results showed that the backpropagation paradigm was best suited for their study. They used actual historical process control data from a thin-film manufacturing process. Their results demonstrated that neural networks show promise as an alternative to control charts. They also explained how to select, evaluate, and verify the optimal architecture for a neural network model.

Pugh (1989) compared the performance of neural networks to those of several standard control charts in a series of experiments under normality. The data for this experiment were generated by Monte Carlo simulation. The input nodes correspond to observations in a subgroup, and a single output represents either an in-control state or an out-of-control state. The neural networks were trained with both in-control and out-of-control data. The results of this preliminary study showed that a backpropagation neural network could be designed which approximately equals the performance of standard $\overline{x}$ control charts in terms of Type I and Type II errors. Pugh performed a similar experiment in 1991 that showed improvement in the performance of neural networks but failed to demonstrate their superiority to standard control charts for detecting shifts in process means.

Hwarng and Hubele (1993a) presented a backpropagation pattern recognition algorithm to detect patterns of a process such as trends, cycles, systematic patterns, stratification, mixtures, and sudden shifts. This classification algorithm is suitable for real-time statistical process control. The foundation of the algorithm is based on the concepts of constructing and training a neural network for recognition and detection of patterns. The backpropagation pattern recognizer (BPPR) and Boltzmann machine pattern recognizer (BMPR) are trained and tested with consideration of Type I and Type II errors. They divided the range of data into seven zones. Each observation was first normalized into a standard normal variate and mapped into one of the seven zones. A series of normalized and scaled observations constituted the input pattern presented to the neural network. Since each zone is represented by a binary digit, seven digits were required to represent one observation. This input scheme resulted in a considerably large network. The authors used performance measures such as average run length and rate of target, which are sensitive to these types of patterns, to evaluate the algorithm's performance. The performance evaluation was carried out in a comprehensive manner based on an extensive series of simulated patterns of data.

Hwarng and Hubele conducted an experimental design to examine the effects of the number of hidden nodes, the number of scale zones, and their interaction effects on the performance of neural networks. Their results indicated that the number of scale zones has significant effect on the pattern recognition performance of neural networks. The results showed that the BPPR successfully detected target patterns while maintaining reasonable Type I and Type II error rates. However, there is one important issue that should be noted. The neural network was trained and evaluated with low levels of random noises. The results indicated that the performance varied quite dramatically depending on the random noises. As Cheng(1995) mentioned, this implies that Hwarng and Hubele's approach is applicable only to the situations in which the signal-to-noise ratio of the target patterns is very high.

Training a neural network for detecting assignable causes of a process based on neural network backpropagation algorithms is neither easy nor straightforward. In order to test and demonstrate the classification capability of their proposed BPPR, Hwarng and Hubele (1993b) extended their research (1993a) and focused on the efficiency of off-line neural network training. They investigated the training and learning speed of the BPPR using simulated data. Hwarng (1995) also investigated the proper and effective training of the BPPR as an alternative to the standard $\overline{x}$ control charts for detecting cyclic patterns of data. This work showed that building an effective cyclic pattern recognition, a neural network requires proper training strategies to improve its performance. Hwarng conducted a series of experiments to study the effects of the number of output pattern classes and noise on neural network training and performance. The results showed that reducing the number of output pattern classes does not guarantee effective learning and that noise added to the training data should be maintained at a reasonable level to achieve a balanced performance. These results are supported by reference to general characteristics of neural networks.

Prybutok, Sanford, and Nam (1995) evaluated the ability of neural networks to detect shifts in a process mean. They provided a heuristic procedure for specifying parameters such as learning rate, momentum, and the number of hidden neurons in a hidden layer for neural network configurations. The effects of the number of hidden neurons of a neural network, its training time, learning rate, and momentum on the performance of the neural networks were examined. They also compared the performance of neural network backpropagation models to those of $\overline{x}$ control charts with and without supplemental run rules. Data from four normal populations, one with no shift and three with different shifts of the process mean, were used to train the neural networks. The authors found that neural networks are more sensitive to small and large shifts of process means as compared to the $\overline{x}$ control charts without supplementary runs rules. However, neural networks perform less effectively than $\overline{x}$ control charts with supplementary runs rules to detect small shifts. These results indicated that neural networks might be an alternative to standard $\overline{x}$ control charts when there are large shifts in process mean.

Cheng (1995) proposed a different neural network approach as an alternative means of process control. The neural network he developed is concerned with the detection of two different types of out-of-control situations, linear trends and sudden shifts in the process mean. A sudden shift of mean could be caused by the introduction of new workers, new raw materials, a change in the measurement system, and so on. The neural network chosen is a three-layer fully connected feed-forward network with a backpropagation training algorithm.

His research is unique in terms of data representation and features selection. A coding scheme was proposed to characterize the important features contained in the input data. This approach divides the range of input data into zones of width 0.5 $\sigma_{\bar{x}}$. The data were represented by the midpoint of the zone interval. This data smoothing scheme was used to discover the important features in the data. To further improve the sensitivity of the neural network, a supplementary index was proposed. The index is set to one if any observation is beyond 3 $\sigma_{\bar{x}}$ limits. The neural network was set to analyze a window of sixteen observations plus one supplementary index. The primary focus is on the problem of detecting both small and large changes in the process mean.

In Cheng's (1995) research, the performance of the neural network was evaluated on the basis of ARL. Since the described neural network is designed to detect both small and large changes in the process mean, he compared the performance of the neural network with that of the combined Shewhart-CUSUM scheme. To obtain a fair comparison between the combined Shewhart-CUSUM schemes and the neural network, the cutoff values of the neural network were chosen so that the in-control ARLs would match those of the combined Shewhart-CUSUM schemes. The results indicated that the proposed neural network was superior to the combined Shewhart-CUSUM schemes for detecting both sudden shifts and linear trends of the process mean. The proposed neural network has 20-30 percent faster detection of small trends than the combined Shewhart-CUSUM schemes.

Research Questions

As mentioned in the previous chapters, a fundamental assumption in the development of $\overline{x}$ control charts is that the underlying distribution of the monitored quality characteristic is normal. However, in many industrial situations, the process means of the quality characteristics may not be normally distributed. If a standard symmetrical $\overline{x}$ control chart is used when the normality assumption is violated, the Type I and Type II errors can be significantly larger than expected. Therefore, design considerations for the $\overline{x}$ chart must include recognition of the degree of nonnormality of the underlying data (Ramsey and Ramsey 1990; Yourstone and Zimmer 1992). To date, however, no research related to the application of neural networks to control charts dealing with nonnormality has been published, and no research related to control charts under nonnormality has yielded generally satisfactory results.

In this study, neural networks are developed as an alternative to standard control charts under conditions of nonnormality, and their performance is compared with that of standard $\overline{x}$ and CUSUM control charts. The $\overline{x}$ chart is the most widely used among the quality control charts for monitoring the central tendency of a process quality characteristic. Its relative ease of computation and application and its excellent performance in recognizing out-of-control conditions account for a portion of its popularity (Wadsworth et al. 1986; Ryan 1989). The $\overline{x}$ chart is used to monitor variation in the process mean and to indicate whether shifts in the mean occur. The standard $\overline{x}$ control chart is less efficient than CUSUM control charts in detecting small shifts in process means, but it is more efficient for detecting large shifts. However, neither of these two popular control charts is able to handle the nonnormality problem. This fact leads to the development of a third method, neural networks, to deal with this problem.

As noted above, the main purpose of this study is to develop neural networks as alternatives to control charts under conditions of nonnormality and to compare their performance to those of standard $\bar{x}$ and CUSUM control charts. Thus, the following research questions are addressed.

1. What are the differences in the performance of the three methods when the underlying distributions of quality characteristics are normally distributed? The following three detailed research questions are delineated to further address this main research question.
   a. Do the neural networks perform better than $\bar{x}$ control charts?
   b. Do the neural networks perform better than CUSUM control charts?
   c. Do the $\bar{x}$ control charts perform better than CUSUM control charts?
2. What are the differences in the performance of the three methods when the underlying distributions of quality characteristics are not normally distributed? The following three detailed research questions are delineated to further address this main research question.
   a. Do the neural networks perform better than $\bar{x}$ control charts?
   b. Do the neural networks perform better than CUSUM control charts?
   c. Do the $\bar{x}$ control charts perform better than CUSUM control charts?
3. What are the effects of degrees of nonnormality measured by skewness and kurtosis on the performance of the three methods when a population is nonnormal?

4. What are the effects of mean shifts on the performance of the three methods when a population is nonnormal?
5. What is the effect of sample size on the performance of the three methods when a population is normal?
6. What is the effect of sample size on the performance of the three methods when a population is nonnormal?

CHAPTER IV

SIMULATION METHODOLOGY

Introduction

Chapter 3 was devoted to a literature review of standard control charts and neural network applications to control charts. In this chapter, the simulation methodologies used in this study are discussed. This chapter is organized as follows: The introduction contains the definition of a sudden shift, the shift pattern of major interest in this study, and the assumptions used for the simulation experiments are described. This is followed by a description of the Burr distribution, skewness, and kurtosis. In the next section, the simulation methodology is discussed. This discussion includes method of data generation; building of $\bar{x}$ and CUSUM control charts; data preparation, training, and testing for neural networks; and definition of the performance measure.

Sudden Shifts

The experiment focuses on detection of out of control conditions caused by sudden shifts. A sudden shift is defined as a series of points with a sudden change in a quality characteristic. The occurrence of a sudden shift may result from the introduction of new workers, methods, raw materials, or machines; a change in the inspection method or standards; or a change in either the skill or motivation of the operators (Montgomery 1991b). Some potential causes of a sudden shift on $\bar{x}$, R, and p charts are shown in table 1 (Western Electric 1956).

Assumptions

The assumptions for this simulation experiment are listed below.

1. The type of the nonnormal distribution measured by skewness and kurtosis is known in advance.
2. The process mean and standard deviation of each nonnormal distribution are known.
3. The type of process change (a sudden shift in the mean) to be detected is known in advance.
4. All the process changes, regardless of the magnitude of the mean shift, are to be detected at the same rate.
5. No in control data precede the out of control data in the test data sets.

Table 1. Potential causes of a sudden shift on various charts

$\bar{x}$ chart	R chart	p chart
New operator	New operators	New lot of material
New inspector	New equipment	Change from one machine or operator to another
New machine	Change in motivation of operators	Change in methods
New machine setting	Change of raw material	Change in standards
Change in set-up or method	Change in maintenance schedule	Change in a calibration of a test set
Change of raw material		

Assumption 1 is required for developing a neural network as an alternative to a conventional control chart. However, this assumption should not be considered as advantageous only to the neural network, because this information can be easily obtained while constructing any conventional control chart. The general form of a $\overline{x}$ control chart contains a center line which represents the mean of the quality characteristic corresponding to the in control state and two other horizontal lines, upper and lower control limits. The upper control limit and the lower control limit are equidistant (the same number of standard deviations) from the center line. These limits are chosen, so that if the process is in control, nearly all of the plotted points fall between the two control limits. When constructing a control chart for a process, a number of preliminary samples (usually 20 to 25) are obtained, each of size 4, 5, or 6, when the process is thought to be in control. Therefore, assumption 2 is required for developing both control charts and neural networks. Based on this information, the distribution of the measure of the quality characteristic can also be examined.

The third assumption, that the type of process change (a sudden shift in the mean) to be detected is known in advance, allows the neural network to detect either positive or negative shifts. This results in a neural network that works well for detecting only sudden shifts. In this study, only sudden shifts are investigated. The fourth assumption, that any process change, regardless of the magnitude of the mean shift, is to be detected as quickly as possible, implies that equal size of training samples for each magnitude of mean shifts is appropriate to develop the neural networks.

Burr Distribution

Burr (1942) introduced a system of distributions that resulted from distribution functions, F(x), that satisfy the differential equation dF/dx = A(F)g(x). The differential equation is a generalization of the Pearson equation: the objective was to fit a distribution function, rather than a density function, to data and then obtain the density function by differentiation (Rodriguez 1977). The cumulative distribution function of the Burr distribution for a random variable x is (Burr 1967; Austin 1973; Yourstone and Zimmer 1992):

$$FX(x) = 0 \qquad x<0 \tag{1}$$

$$= 1 - \frac{1}{(1+x^{c})^{K}} \qquad x \geq 0 \qquad c, k \geq 1$$

where c and k are positive numbers greater than or equal to 1. The c and k values specify a member of the family of Burr distributions that is used to approximate an empirical distribution.

The Burr family covers a wide and important range of standardized third and fourth central moments and is used to fit a wide variety of practical data distributions. The distribution given by (1) covers a region or area of points on the skewness and kurtosis plane, whereas the normal curve (skewness=0, kurtosis=3) and the uniform distribution (skewness=0, kurtosis=1.8) occupy the respective single points. The gamma distribution covers only a curve on the plane. Burr also found that by using the reciprocal transformation x = 1/y, yielding the distribution function shown below,

$$G(y) = 0 \qquad y<0 \qquad (2)$$

$$= 1 - \frac{1}{(1+y^{-c})^{K}} \qquad y \geq 0 \qquad c, k \geq 1$$

the region of coverage is greatly extended toward low kurtosis for a given skewness. This extension of the region covers virtually all of the bell-shaped and J-shaped regions of the beta region. Thus, by using either (1) or (2), one can match up the first four moments for almost any ordinary distribution likely to occur in practical situations (Burr 1973).

The first four moments, or the first two moments and the skewness and kurtosis of an empirical distribution, may be used to determine values of c and k. These c and k values then specify a member of the Burr family that approximates the observed empirical distribution. Therefore, if the first four moments of a distribution are in the range of those covered by the Burr family, the empirical distribution can be easily approximated by a combination of c and k. Two distributions with the same first four moments will not necessarily be identical, but it is reasonable to assume that they will not differ practically. With c= 4.874 and k=6.158, the Burr distribution approximates a normal distribution. For other values of c and k, the Burr distribution becomes distinctly nonnormal (Yourstone and Zimmer 1992). Distribution systems (1) and (2) have far greater coverage of skewness and kurtosis than other functions such as beta, gamma, and so on. The coverage of the Burr distribution includes up to the skewness equal to 2 and kurtosis equal to 9, which are the moments of the exponential (Burr 1973).

Skewness and Kurtosis

To measure the degree of departure from normality of a distribution, skewness and kurtosis are usually used. When data tend toward a nonsymmetric distribution, the distribution is referred to as skewed. Skewness measures the tendency of a distribution to stretch in a particular direction. The sign of skewness shows the direction of skewness, + (positive) if above, - (negative) if below. The magnitude of the skewness measure is the relative extent of the lack of symmetry. If a distribution is symmetric, skewness is 0. The other measure of shape of a distribution, referred to as the kurtosis, measures the peakedness of a distribution. Kurtosis measures the relative rate at which the frequency curve approaches zero frequency in the two directions. For the normal curve, kurtosis is 3, while for a uniform distribution, it is 1.8. The minimum possible value of kurtosis is 1. The quantity (α_4 - 3) is referred to as excess and is sometimes used as the measure of kurtosis (Burr 1967).

$$\text{skewness} = \alpha_3 = \frac{\Sigma(x_i - \bar{x})^3}{[\Sigma(x_i - \bar{x})^2]^{3/2}}$$

$$\text{kurtosis} = \alpha_4 = \frac{\Sigma(x_i - \bar{x})^4}{[\Sigma(x_i - \bar{x})^2]^2}$$

Notice that these estimates are based on individual data values x_i. After the sample size, n, is determined for the samples in the control process, the estimates for the skewness and kutosis of the sample averages are:

$$\text{Skewness} = \alpha_{3\bar{x}} = \frac{\alpha_3}{n^{1/2}} \tag{3}$$

$$\text{Kurtosis} = \alpha_{4\bar{x}} = \frac{\alpha_4 - 3}{3} + 3 \qquad (4)$$

Burr (1967, 1973) tabulated c and k values giving combinations of skewness and kurtosis for various Burr distributions. The table was used in this study to obtain the c and k parameters of the Burr distribution as well as the mean and standard deviation for a given skewness and kurtosis. Thus, the Burr distribution to approximate the distribution of a quality characteristic is generally applicable to any distributions whose coefficients of skewness and kurtosis are in the Burr distribution range of values.

Simulation Methodology

The main purpose of this study is to explore the viability of neural networks as alternatives to standard $\bar{x}$ and CUSUM control charts under conditions of nonnormality. This study compares the performance of neural networks to those of standard $\bar{x}$ and CUSUM control charts for monitoring process means. To accomplish the objectives of the study, numerous experiments and simulations were conducted. The simulations in this study consist of four stages: the generation stage, the training stage, the testing stage, and the comparison stage. In the generation stage, the data sets for neural network training and testing were generated from the Burr distribution for nine combinations of skewness and kurtosis. In the training stage, neural networks with different numbers of hidden layers and neurons, different values of momentum, learning rate, and different learning rules, and transfer functions were tested to obtain better architecture for each one of the nine nonnormal distributions. In the testing stage, FORTRAN programs were used to calculate average run lengths (ARLs) for the three methods for each of the nine nonnormal distributions. ARL is the average number of subgroups that are observed from a process before an out of control signal is detected. In the comparison stage, the three major experiments were carried out as noted earlier.

Data Generation

Monte Carlo simulation was used to generate training and testing data for each of the nine nonnormal distributions examined in this study. The nine nonnormal distributions with different skewness and kurtosis were created from the generalized Burr distribution. The nine combinations of skewness and kurtosis yield distributions that vary from one similar to a uniform distribution to one that resembles an exponential distribution. Control chart schemes with exponentially distributed observations arise naturally in the context of monitoring the occurrence rate of rare events, since interarrival times for a homogeneous Poisson process are independent and identically distributed exponential random variables (Vardeman and Ray 1985). Vardeman and Ray (1985) examined ARLs for CUSUM schemes when observations are exponentially distributed. They provided some tables of ARLs for an exponential case and commented on an application of exponential CUSUM charts to controlling the intensity of a Poisson process. There are many possible cases in which observations might be exponentially distributed. Consider a manufacturing environment (such as the manufacture of plastic film or copper wire) with a long continuous stream of a product that can develop flaws along its extent. Suppose further that there is a mechanism, human or automatic, for detecting such flaws as the line of product passes by on its way to being cut into sheets, rolls, or other more manageable units. Suppose finally that the rate at which the continuous stream of product passes the inspection mechanism is uniform. Then, it is natural to use the occurrence times of a Poisson process as a model of the occurrence of defects (Vardeman and Ray 1985).

Table 2. Parameters of the nine nonnormal distributions generated from the Burr distribution

Skewness	Kurtosis	c	k	μ	σ
0.00	2.00	-18.148445	0.062932	0.538666	0.290634
0.00	3.00	-11.251863	0.146295	0.641872	0.283920
0.00	4.00	27.068908	1.325754	0.985829	0.059999
1.00	3.00	-13.068026	0.030050	0.285705	0.300232
1.00	4.00	-7.176003	0.078658	0.378436	0.339226
1.00	5.00	2.347094	4.428629	0.506046	0.262382
2.00	6.20	-21.416286	0.007433	0.137895	0.236343
2.00	7.20	-7.507666	0.027031	0.174654	0.272662
2.00	8.80	-5.864337	0.044223	0.218459	0.308432

* c and k - two parameters of the Burr distribution

Eighteen different mean shifts (±.25, ±.50, ±.75, ±1.0, ±1.5, ±2, ±3, ±4, ±5) were examined for each of the combinations of skewness and kurtosis, and the performance of the neural networks was compared with those of the $\overline{x}$ and CUSUM control charts. Seventeen hundred samples (see Table 3) were generated from each of the nine combinations of skewness and kurtosis. A sample size of five in each sample was used, because the control chart procedure, based on sample averages of five observations, is believed to be approximately normal due to the central limit theorem. Another reason that five was selected as the sample size is that it is a common value used in practice. The parameters of the nine combinations of skewness and kurtosis generated from the Burr distribution are shown in table 2.

Development of $\overline{x}$ Control Charts

A $\overline{x}$ control chart consists of a center line and two control limits. To construct the control limits, an estimator of a standard deviation is needed. The estimator may be obtained from either the standard deviations or the ranges of preliminary samples. If a sample size is relatively small, the range method yields almost as good an estimator of the variance as does the estimator from the standard deviation (when $n = 3$, the relative efficiency is 0.992). However, the range method loses efficiency rapidly as the sample size becomes larger. When $n = 5$, the relative efficiency is 0.955. In this study, the standard deviation method is used to obtain more accuracy.

Development of CUSUM Control Charts

An effective alternative to Shewhart $\overline{x}$ control charts when small shifts are of interest is the CUSUM procedures. It is well known that CUSUM procedures perform better than Shewhart charts in detecting small process shifts. Cumulative sum charts were first proposed by Page (1954), and a number of modifications have resulted since then. The CUSUM is so named because an observation (which may be a single reading or a statistic obtained from a sample) is accumulated if it exceeds a prespecified goal value by more than k units.

A relatively small percentage of firms in the U.S. uses CUSUM charts. Workers engaged in quality improvement are still primarily interested in learning about $\overline{x}$ and R control charts (Ryan 1989). Some CUSUM procedures are easy to learn (and to program) and are also quite intuitive, while others are somewhat more involved. If quality control personnel are to make the transition from $\overline{x}$ charts to CUSUM control charts, it seems reasonable to assume that they will be drawn toward procedures that are easy to understand and that bear some relationship to a $\overline{x}$ chart. Therefore, in this study, the method proposed by Lucas (1982) and Lucus and Crosier (1982) was used because it meets both requirements above and its performance is very competitive.

This CUSUM procedure is an algorithmic version based on the recursive statistics:

$$S_{Hi} = \max[0, (Z_i - k) + S_{Hi-1}]$$

$$S_{Li} = \max[0, (-Z_i - k) + S_{Li-1}]$$

where $z_i = (\overline{x}_i - \mu)/\sigma_{\overline{x}}$. The first equation is for detecting positive mean shifts, and the second equation is for detecting negative mean shifts. The starting values are $S_{H0} = S_{L0} = 0$. The value of k is usually selected as one-half of the mean shift (in z units) that one wishes to detect. If the process remains in control at the target mean, the cumulative sum defined above should fluctuate stochastically around zero because the CUSUM is a random walk with mean zero. However, if the mean shifts upward, then an upward or positive drift will develop in the S_{Hi}. Conversely, if the mean shifts downward, then a downward or negative drift in S_{Li} will develop. CUSUM procedures are alternatives to $\overline{x}$ charts when small shifts are of interest. The performance of the CUSUM control chart is determined by two parameters, k and h. The usual choice is k = 0.5, which is the appropriate choice for detecting a 1 sigma mean shift. Two sums, S_{Hi} and S_{Li}, are computed for each z_i; i designates the ith subgroup. There is a minus sign in front of z_i in the formula for computing S_{Li}. This simply ensures that $-z_i$ will be positive whenever there is a negative mean shift. Neither S_{Hi} nor S_{Li} will ever be negative since they will always be the maximum of 0 and $(z_i-k)+S_{Hi-1}$ and $(-z_i-k)+S_{Li-1}$, respectively. The sums are reset after an out of control signal is issued. That is, whenever the absolute value of z_i is greater than k, S_{Hi} or S_{Li} are zero and accumulation starts again (Ryan 1989).

The process is considered out of control, and action should be taken whenever S_{Hi} or S_{Li} is greater than h. The threshold, h, is chosen to be either 4 or 5. Different choices of h change the performance of the CUSUM procedures. When a large h value is used, the Type I error rate decreases (i.e., in control ARL increases), while Type II error increases (i.e., out of control ARL increases). On the other hand, when a small h value is used, Type I error rate increases (i.e., in control ARL decreases), while Type II error decreases (i.e., out of control ARL decreases). For example, the in control ARL value when h=4 is used is much smaller than that when h=5 is used. However, this is offset to some extent by the fact that the out of control ARL values are larger using h=5 for every mean shift. Thus, the CUSUM procedure with h=5 delays the detection of an out of control state because an out of control state is detected when the accumulated sum is 5 or larger. If the shift could cause a considerable increase in the number of defects, the extra delay might be critical. Conversely, if the cost of searching for assignable causes when none exists is substantial, the difference between the in control ARLs might be of critical importance. Thus, the choice of h should be made with careful consideration of all the relevant factors, both statistical and nonstatistical. In practice, the CUSUM procedure with h=5 is frequently used. In this study, to obtain a fair comparison among $\overline{x}$ and CUSUM control charts and neural networks, the h value was chosen so that the in control ARLs of the CUSUM procedure would match those of the $\overline{x}$ control chart and the neural network for each one of the nine different combinations of skewness and kutosis.

Data Preparation for Neural Network

For neural network control chart applications, the data may require conversion into another form to be meaningful to a neural network. Regardless of which network model is used, the general approach to solving the data representation problem will be the same. Understanding the problem at hand well enough to know what kind of information is relevant is one key to the successful design of a neural network. How data are represented and translated also plays an important role in designing the neural network. When data represent naturally occurring groups, binary categories are often the best method for making correlations. When the data are continuous, artificially breaking them up into groups can be a mistake. It is difficult for a neural network to learn from examples that have values on or near the border between two groups (Lawrence 1991).

Often, however, the choice between binary and continuous representations is not so simple. In this study, a continuous data entry scheme was used for the five observations in each sample subgroup because a measure of a characteristic of a product such as length, weight, or width is considered to be continuous data. On the other hand, because this study can be considered to be three group discriminant problems, a binary data entry scheme was used for the output of each subgroup. The three groups include an in control state and positively and negatively shifted out of control states. That is, an output of [0 0 1] represents a subgroup from a population with no mean shift. An output of [0 1 0] represents a subgroup from a population with a positive mean shift, and an output of [1 0 0] represents a subgroup from a population with a negative mean shift.

Neural Network Training

The construction of a training set for the present study was complicated because many magnitudes of process changes are considered at the same time. There are a number of issues that require addressing to obtain an adequate training set (Cheng 1995). First, the neural network should be trained across a wide spectrum of possible mean shifts in order to yield satisfactory performance over the entire range of possible process changes. In other words, the range of mean shifts included in the training set should be representative of the whole spectrum of the shifts of interest. However, the determination of the range of mean shifts is subject to the required recognition capability and the expected training time. A neural network trained at a single magnitude might not work well for other magnitudes of mean change. In essence, while the network can generalize to some extent, the network can only identify patterns that resemble those it has seen during its training. Therefore, care in selecting the range of mean shifts is required.

Extensive preliminary experiments were conducted to determine the range of mean shifts to be included in the main study. The experiments showed that when small magnitudes of process changes (.25, .5, and .75 sigma) are included in the training set, the resulting performance is poor as measured by the Type I error rate (the rate of false alarms). When data examples with small mean shifts were included, the overlap between the in control data examples and out of control examples was too severe for the neural network to distinguish one from the other. The results of these experiments were consistent with the general rule that the training data should include both in control data and out of control data, but these out of control data should be readily distinguishable from in control data.

The second consideration is the number of training examples to be used for each magnitude of a process mean shift. In this study, it is assumed that any process change, regardless of the magnitude of the mean shift, is to be detected as quickly as possible. Accordingly, an equal number of training examples was used for each magnitude of a process mean shift. The third issue concerns the size of the training set. It is well known that a neural network performs best when trained with a sufficient number of examples to achieve good generalization. Also, during training, the best results are obtained when training examples are presented to the neural network in a random fashion. A neural network trained with an insufficient set of examples often leads to poor performance in terms of Type I and Type II errors.

Table 3 outlines the contents of the training data sets. The magnitude of the mean shifts is expressed by the standardized shift parameter δ defined as follows:

$$\delta = |\mu - \mu_0| / (\sigma/\sqrt{n})$$

where

μ_0 = process mean before a shift,

μ = process mean after a shift,

σ = process standard deviation,

n = subgroup size.

A target mean (μ_0) of zero is used in this study, which is consistent with the values used in several other control chart simulation studies (Crowder 1987; Runger and Pignatiello 1991;

Walker, Philpot, and Clement 1991). The number of training examples used in this study was determined through experimentation. There were 1,700 examples in each of nine training sets, corresponding to the nine combinations of skewness and kurtosis. It is important to note that each of the training sets includes both in control data and out of control data. This application requires that the neural network distinguish between in control patterns

Table 3. Training data

Shift (δ)	Training sample size
0	500
±1	200
±2	200
±3	200
±3	200
±4	200
±5	200
±6	200
total	1700

and out of control patterns. The training method of the neural networks for control chart applications is the same as that used in discriminant analysis. As shown in table 3, each data set consists of three basic groups. One group with no mean shift consists of 500 in control data examples. The other two groups consist of 600 data examples each and represent positively shifted (+1δ to +6δ) and negatively shifted (-1δ to -6δ) out of control data .

The simulation experiments in this study used the back-propagation learning algorithm to train the neural networks. The back-propagation algorithm is widely used, not only because it is based on solid mathematical principles, but also because it has produced a number of successful solutions to various business statistical problems. The speed of learning is one of the issues related to neural network training and has been one of the major concerns in real-time applications. Although the slowness of learning is often a concern when using backpropagation networks, the learning speed with the training sets used in this study was fast enough. In this study, only five input neurons, five hidden neurons, and three output neurons were used. The simplicity of the neural network architecture used guaranteed a reasonable training time. In general, the somewhat lengthy learning process can be done prior to using the neural network online.

Another concern related to training time is the determination of the optimal stopping point. The performance of the neural network is affected by the number of presentations of the training examples. During the training process, the performance of the neural network on the training data set improves continuously. However, the performance on the testing data set may become worse because the neural network has the potential to over learn or memorize the training data set. In other words, the longer the neural network learns the training data set, the closer the neural network comes to memorizing it. Thus, the neural network can make very accurate predictions for the training data set. However, it may not be able to generalize well to the testing data set. The memorization problem causes the nongeneralization of predictive ability of the neural network to the testing data set. There currently exists no method of optimizing the selection of a stopping point or the selection of the various neural network parameters. Therefore, any neural network solution is potentially suboptimal.

The selection of the neural network training architecture is crucial to the performance of the trained neural network. The selection includes for use the best combination of learning rate, momentum, number of hidden layers, number of hidden layer neurons, learning rule, and transfer function. In this study, a learning rate of 0.1 and a momentum of 0.6 were used after testing more than thirty different combinations of the parameters for each of the nine different nonnormal distributions examined in this study. The architecture also affects one of the major issues in neural network training, the convergence of the learning algorithm. To ensure efficient convergence and the desired performance of the trained network, the training parameters, including the learning coefficient, momentum, number of hidden layers, and the number of hidden neurons, must be selected with care. More important, proper selection of the activation function is required to ensure better performance of the neural network. In this study, a sigmoid activation function is used. This function is frequently used in neural network applications in the business area primarily because of its desirable function features and its relative insensitivity to noise. The sigmoid transfer function is monotonic, differentiable, and semilinear.

A number of experiments were performed to determine the best combination of learning rate, momentum, number of hidden layers, number of hidden layer neurons, learning rule, and transfer function. The two most popular learning rules, the generalized delta and cumulative delta, were tested. A fully connected feed-forward network with five input neurons was employed. Each of the first five neurons represents an observation in a subgroup (n=5). The desired output was set to be [0 0 1] if the data were from the target process with mean equal to zero and [0 1 0] or [1 0 0] if the data were from distributions with positive mean shifts or negative mean shifts, respectively. There are some general guidelines for determining the number of neurons to be placed in the hidden layer of a fully connected feed-forward network (NeuralWare 1993). One such guideline specifies that H = (number of training cases)/5*(I + O), where H, I, and O are numbers of neurons in the hidden, input, and output layers, respectively. This guideline was followed in this study. A single hidden layer was used because neural networks tested with more than one hidden layer did not produce any significant performance improvement. To examine the performance of neural networks with different numbers of hidden layer neurons, the experiments were conducted with 3, 4, 5, 6, 7, 10, 15 neurons in a single hidden layer. Although the neural networks for some of the cases of the nine different skewness/kurtosis combinations with four or six hidden neurons worked almost the same as or a little better than the neural networks with 5 hidden neurons, the differences were marginal. Therefore, five neurons in a single hidden layer were used for all the nine different nonnormal distributions examined in this study. Several other rules of thumb are suggested in the literature. They include the following:

1. The number of hidden neurons is equal to

   $2 * \sqrt{number\ of\ input\ neurons + number\ of\ output\ neurons}$

   rounded down to the nearest integer (Neuroshell 1991).

2. The number of hidden neurons falls between the total number of input layer and output layer neurons (Neuralware 1991).

3. The number of hidden neurons should be 75percent of the number of input layer neurons (Salchenberger, Cinar, and Lash 1992).

A common feature of the rules of thumb above is that the numbers of input and/or output neurons are used to determine the number of hidden neurons. Three output neurons, as mentioned above, are used in this study. In neural network training, the connection weights are adjusted after every presentation of a training example. Throughout the training, the NeuralWare utility, SAVEBEST was used to monitor and save the connection weights that yielded the lowest root mean square error (RMS). SAVEBEST provides a means of dealing with the problem of when to stop training the network. The fixing of the neural network connection weights is similar to establishing the three sigma control limits for a $\overline{x}$ control chart. It is the configuration of connection weights that leads to an observation being classified as either in or out of control.

Since the pairing of each input vector with a target vector representing the desired output was available, a supervised training neural network was used. The generalized delta rule was used for error correction. The backpropagation neural network in this study was trained using NeuralWorks Professional II/PLUS software by NeuralWare Inc. The performance of the neural networks was evaluated on the basis of the number of misclassifications observed on testing data set. The misclassification rate of in control data is analogous to the Type I error rate (i.e., the rate of false alarms; concluding the process is out of control when it is really in control). The misclassification rate of out of control data is simply another way of stating the Type II error rate (i.e., concluding the process is in control when it is really out of control). In general, a preferable control scheme is supposed to maintain a balanced performance with acceptable classification accuracy (1-probability[Type II error]) and a reasonably low Type I error rate. It is believed that a low

Type I error rate is as important as high classification accuracy. Therefore, a balance between Type I and Type II error rates is desirable.

Neural Network Testing

After the neural networks were trained, they were used to identify in control and out of control states. Each output value from the testing data consists of three numbers, [a, b, c], one from each of the three output neurons. The values from each of the output neurons is between 0 and 1 because the sigmoid tranfer function was used to train the neural networks. The neural networks were trained using [0 0 1] to represent a subgroup from a population without a mean shift, [0 1 0] to represent a subgroup from a population with a positive mean shift, and [1 0 0] to represent a subgroup from a population with a negative mean shift. Therefore, if the value c is close to one, then the subgroup in the testing data set is considered to be in control. One the other hand, if the value a or b is close to one, then the subgroup is considered to be out of control.

There are several options for setting the cutoff value for the in control state. One possibility is to increase or decrease the cutoff value for the last output neuron, c. Lowering the cutoff point for c increases the probability of a Type I error, but reduces the probability of a Type II error. On the other hand, increasing the cutoff point increases the probability of Type II error, but reduces the probability of a Type I error. Another option is to increase and/or decrease the cutoff values for a and/or b. In this study, nonnormal distributions, including various nonsymmetric distributions, are investigated. Except for the symmetric cases, nonnormal distributions do not have mirror images. Therefore, the second method described above was used because the cutoff value 'a' can be used to handle positive mean shifts and the cutoff value 'b' can be used to handle negative mean shifts. Tests of the two methods showed that neural networks with two cutoff points, one for 'a' and one for 'b', performed better than those with only one cutoff point for 'c'. In this study, to obtain fair comparisons among the $\overline{x}$ chart and CUSUM charts, and neural networks, the cutoff values for 'a' and 'b' were chosen by trial and error simulations so that the in-control ARLs of the neural networks would match those of the $\overline{x}$ and CUSUM charts for each one of the nine different combinations of skewness and kurtosis. The cutoff points for neural networks and the 'h' values for the CUSUM charts used in this study are shown in table 4.

Table 4. Cutoff points for neural networks and CUSUM charts

Case	Neural network cutoff point (a)	Neural network cutoff point (b)	CUSUM (h)
1	0.8236	0.9224	5.465
2	0.7675	0.8738	4.785
3	0.5235	0.8578	4.485
4	0.7001	0.8718	4.450
5	.8700	.4850	4.375
6	0.8375	0.5170	4.202
7	0.5185	0.6370	3.815
8	0.6400	0.5330	3.780
9	0.8400	0.3820	3.790

Definition of Performance Measure

The performance of each of the three methods discussed above is evaluated using average run length (ARL). Conventionally, the average run length serves as a useful and standard caliber for measuring the effectiveness of a control chart scheme. ARL is defined as the average number of subgroups that are observed from a process before an out-of-control signal is detected. When the process is in-control (shift parameter $\delta=0$), it is desirable to have a large ARL. However, if the mean of the process has shifted (shift parameter $\delta \neq 0$), it is desirable to have a small ARL. In other words, the monitoring process should generate signals as quickly as possible (i.e., shorter ARLs) if the production process is out-of-control and as late as possible (i.e., longer ARLs) if the production process is in-control. The problem associated with nonnormal data is that both the in-control ARL and the out-of-control ARL are impacted. For a $\overline{x}$ control chart, the in-control ARL is calculated as

$$\frac{1}{\alpha}$$

where α is the probability of the chart indicating that a shift in mean occurs when there is actually no shift. To calculate the ARL for the out-of-control situation, β values are calculated for the 135 combinations of determined mean shifts, skewness, and kurtosis. Once these values are determined, the out-of-control ARL is calculated as

$$\text{ARL} = \frac{1}{1-\beta}$$

where β is the probability of not detecting a given shift in a process. Calculation of ARLs for CUSUM control charts and neural networks is not as straightforward as it is for $\overline{x}$ control charts. There are several ways to calculate the ARLs for CUSUM charts. Champ and Woodall (1987) used Markov chains to obtain the ARL for CUSUM control charts. Another method employed in this study is the use of simulation. The ARLs for neural networks are calculated by using a simulation as well.

For calculating ARLs for neural networks, the basic procedure is as follows. After a neural network is trained, it classifies observations as either in-control or out-of-control. Control status is determined by the value of the networks output neurons. The number of subgroups created since the previous action signal is recorded and the number of subgroups between two action signals is called the run length. This procedure is repeated until the 10,000th action signal occurs. The average of these 10,000 run lengths is the ARL for the given combination of skewness and kurtosis.

Since the proposed neural networks are designed to detect both small and large mean shifts in a process, they can be compared with a combined Shewhart-CUSUM scheme, which, is a combination of $\overline{x}$ and CUSUM control charts. This combined scheme is designed to take advantage of the respective strengths of both $\overline{x}$ and CUSUM control charts so that it detects small shifts as quickly as a CUSUM chart and large shifts as quickly as a $\overline{x}$ chart. In this Shewhart-CUSUM scheme, an out-of-control signal would be received if either the absolute value of z exceeds three or either S_H or S_L exceeds h (Ryan 1989). From a practical point of view, however, it is better to compare the neural network with the $\overline{x}$ and CUSUM control charts separately. The reason is that $\overline{x}$ control charts are usually used for detecting mean shifts in the real world (although some use is made of the CUSUM scheme).

The main experiment in this study involves comparing the performance of neural networks to those of standard $\overline{x}$ and CUSUM control charts for monitoring process means. The three methods are compared for nine different combinations of skewness and kurtosis generated from Burr distribution by Monte Carlo simulation. The hypotheses for comparing the ARLs of the methods for each of the nine combinations are as follows:

$$H_0 : \mu_{D_{ij}} \leq 0$$

$$H_a : \mu_{D_{ij}} > 0 \qquad i = 1,2,3 \text{ and } j = 1,2,3 \quad i \neq j$$

where D_{ij} is the difference between each pair of ARLs generated by the methods. Each of the three levels of skewness is represented by the value i, and each of the three levels of kurtosis is represented by the value j. The parameters used in this study are specified in table 5.

Table 5. Parameter specification in the simulation

Methods	Mean shifts (δ)	Skewness	Kurtosis	Sample size
$\bar{x}$ chart	0	normal	low	3*
CUSUM	±.25	skewed (+)	medium	5
Neural network	±.50	skewed (++)	high	
	±.75			
	± 1			
	± 1.5			
	±2			
	±3			
	± 4			
	± 5			

* Sample size of 3 is examined only for the selective cases, 2, 4, and 7.

CHAPTER V

EXPERIMENT AND PERFORMANCE ANALYSES

Introduction

Chapter 4 was devoted to a discussion of simulation methodologies used in this study. In this chapter, the results of three extensive simulation experiments are reported. The first experiment compares the performance of the neural networks with those of the $\overline{x}$ and CUSUM control charts for each of the nine normal and nonnormal distributions with different degrees of skewness and kurtosis. This experiment also examines the effects of nonnormality measured by different skewness and kurtosis on the ability of the three methods to detect out-of-control states. The second experiment investigates the effects of mean shifts on the performance of the three methods. The third experiment examines the effects of sample size on the performance of the three methods.

Comparisons of the Three Methods

A randomized block design allows comparison of the performance of the three methods. The randomized block design is an experimental design for comparing two or more treatment means when an extraneous source of variability (blocks) is present (Ott 1988). By choosing mean shifts to represent the blocks, the variability among the mean shifts is effectively filtered out, allowing more precise comparisons among the three methods. After running the block design, Tukey's multiple comparisons technique allows identification of which methods differ from each other.

Case 1 (skewness=0, kurtosis=2)

In case 1, nonnormal data with skewness of 0 and kurtosis of 2 were examined. A skewness of 0 and kurtosis of 2 indicate that the underlying distribution is symmetric, but has lighter tails than the normal distribution (kurtosis for the normal distribution = 3) (figure 6--all the figures are located in Appendix B to avoid complexity of the manuscript). Actually, the shape of the distribution is close to a uniform distribution, which is why the in-control ARL (784) of this distribution is much higher than the ARL (371) of the normal distribution. With symmetric distributions, process changes in positive and negative directions will be detected equally well because of the directional invariance property. As the mean shift becomes larger, the out-of-control ARL becomes dramatically smaller, as shown in table 6 (All the tables are located at the end of the chapter to avoid complexity of the manuscript) due to the fact that, with heavy tails, mean shifts result in more area under the distribution exceeding the control limits. The slight differences in ARLs between positive shifts and negative shifts can be attributed to randomness in the simulation. The Type I error rates (represented by in-control ARL, the ARLs corresponding to a mean shift of zero) of the $\overline{x}$ charts, CUSUM charts, and neural networks were kept the same to maintain fairness in each of the nine cases examined in this study. The out-of-control ARLs characterize the speed with which a control procedure is capable of detecting a process change in a sequence of data. Therefore, a small ARL is desired when the process is out-of-control. On the other hand, a large ARL is desired when the process is in-control because the in-control ARL provides the rate of false alarms.

Table 6 contains the ratios, ρ_1, ρ_2, and ρ_3. ρ_1 and ρ_2 represent the ratio of the ARL values for the neural networks to the corresponding ARL values for the $\bar{x}$ and CUSUM control charts, respectively. ρ_3 represents the ratio of the ARL values for the $\bar{x}$ charts to the corresponding ARL values for the CUSUM control charts. These three ratios are used because they provide concise and convenient comparisons of the performance of the three methods. It can be concluded that the proposed neural networks perform better than the corresponding $\bar{x}$ control charts when ρ_1 is less than one and perform better than the corresponding CUSUM control charts when ρ_2 is less than one. For example, the ρ_1 value for a mean shift of -0.25δ is 0.94. This implies that the proposed neural network detected the out-of-control state 6 percent faster than the $\bar{x}$ chart. It can also be concluded that the proposed $\bar{x}$ charts perform better than the CUSUM control charts when ρ_3 is less than one. As shown in table 6, the values of ρ_1 for both positive and negative shifts in the mean are approximately one. This implies that the neural network performs as well as the $\bar{x}$ control chart. The values of ρ_2 and ρ_3 for large mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one. This implies that the $\bar{x}$ control chart and the neural network perform better than the corresponding CUSUM chart to detect out-of-control states. However, the values of ρ_2 and ρ_3 for small mean shifts ($\pm 0.25\delta$ to $\pm 2\delta$) are much larger than one. This implies that the $\bar{x}$ control chart and the neural network perform worse than the corresponding CUSUM control chart.

These results are supported statistically by the results from the analysis of variance employed in this study. The results from the analysis of variance show that there is a significant difference among the performances of the three methods (p-value = .0022) (table

7). Furthermore, the multiple comparisons technique indicates that the CUSUM chart performs better than the $\overline{x}$ control chart and the neural network for the symmetric nonnormal distribution with kurtosis = 2 at $\alpha = .05$. The multiple comparisons also indicate that there is no significant difference between the $\overline{x}$ control chart and the neural network at $\alpha = .05$ as shown in Table 8 (Table 8 contains all the results from the multiple comparisons for the nine cases examined in this study). The comparison of the ARLs of the three methods is shown in figure 7 in Appendix B. Notice that the values on the vertical axis are ARLs for the $\overline{x}$ chart, CUSUM chart, and neural network and the values on the horizontal axis are the magnitudes of the mean shifts in terms of δ. The resulting ARL curves are symmetric because the distribution examined here is symmetric.

Case 2 (skewness = 0, kurtosis = 3)

In case 2, the normal distribution (skewness = 0 and kurtosis = 3) was examined (figure 8). With a normal distribution, process changes in positive and negative directions will be detected equally well because of the directional invariance property. As shown in table 9, the ratios, ρ_1, ρ_2, and ρ_3 in case 2 are similar to those in case 1. This implies that the relative performance of the three methods in case 2 is about the same as in case 1.

These results are supported statistically by the results from the analysis of variance, which shows that there is a significant difference among the performances of the three methods (p-value = .0004) (table 10). In addition, the multiple comparisons technique indicates that the CUSUM chart performs better than the $\overline{x}$ control chart and the neural network at $\alpha = .05$ when the distribution is normal. The multiple comparisons also show that there is no significant difference between the $\overline{x}$ control chart and the neural network at $\alpha =$

.05. Basically, there is no difference in the performance results between cases 1 and 2 due to the symmetry of the distributions. The resulting ARL curves are symmetric because the distribution examined here is normal, as shown in figure 9.

Case 3 (skewness = 0, kurtosis = 4)

In case 3, nonnormal data with skewness of 0 and kurtosis of 4 were examined. A skewness of 0 indicates that underlying distribution is not skewed, similar to the cases 1 and 2, while a kurtosis of 4 indicates that the peakedness of the distribution is higher than for the normal distribution (figure 10). As in cases 1 and 2, process changes in positive and negative directions are detected equally well because of the directional invariance property. As shown in table 11, the ratios, ρ_1, ρ_2, and ρ_3 in case 3 are similar to those in cases 1 and 2. This implies that the relative performance of the three methods in case 3 is about the same as in cases 1 and 2. This result is mainly due to the symmetry of the distributions in cases 1-3.

These results are supported statistically by the results from the analysis of variance, which shows that there is a significant difference among the performances of the three methods (p-value = .0001) (table 12). The multiple comparisons technique indicates that the CUSUM chart performs better than the $\bar{x}$ control chart and the neural network at $\alpha = .05$ for the nonnormal distribution. The multiple comparisons also show that there is no significant difference between the $\bar{x}$ control chart and the neural network at $\alpha = .05$. The analysis of variance indicates that there is a significant effect due to mean shift (p-value = .0001). Basically, the results from this case are the same as those from cases 1 and 2, which again is due to the symmetry of the distribution. The resulting ARL curves are symmetric because the distribution examined here is symmetric, as shown in figure 11.

Case 4 (skewness = 1.0, kurtosis = 3.0)

In case 4, nonnormal data with skewness of 1 and kurtosis of 3 were examined. A skewness of 1 indicates that the underlying distribution is positively skewed, while a kurtosis of 3 indicates that the peakedness of the distribution is the same as for the normal distribution (figure 12). As shown in table 13, the values of ρ_1 for positive mean shifts are approximately one. This implies that the neural network performs as well as the $\overline{x}$ control chart for positive mean shifts. On the other hand, the values of ρ_1 for negative mean shifts are much smaller than one. This means that the performance of the neural network is considerably superior to that of the $\overline{x}$ control chart for negative mean shifts.

The values of ρ_2 for positive large mean shifts (3δ to 5δ) and all negative mean shifts are much smaller than one. This implies that the neural network performs better than the corresponding CUSUM chart for these mean shifts. The values of ρ_2 for small positive mean shifts ($.25\delta$ to 2δ) are larger than one. This implies that the neural network performs worse than the corresponding CUSUM chart for these mean shifts. The values of ρ_2 for large positive and negative mean shifts are smaller than one. This indicates that $\overline{x}$ control chart performs better that the corresponding CUSUM chart. The values of ρ_3 for small positive and negative mean shifts ($\pm.25\delta$ to $\pm 2\delta$) are larger than one. Specifically, the values of ρ_3 for the negative mean shifts (-0.25δ to -1δ) are 30.07, 128.81, and 43. This implies that the $\overline{x}$ control chart performs much worse than the corresponding CUSUM chart for small mean shifts. The values of ρ_3 for the large negative mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one. This suggests that the $\overline{x}$ control chart performs better than the corresponding CUSUM chart for these mean shifts.

The analysis of variance shows that there is no statistical significant difference among the performances of the three methods at $\alpha = .05$ (p-value = .0521) (table 14). However, the means of the ARLs of the three methods indicate that the neural network performs better than the corresponding $\bar{x}$ and CUSUM control charts (the mean ARL of the $\bar{x}$ chart = 224.36; the mean ARL of the CUSUM chart = 26.87; and the mean ARL of the neural network = 18.99). The results from the analysis of variance indicate that there is no statistically significant effect due to mean shift (p-value = .4136).

This result is also shown in figure 13. The ARL curve of the neural network matches that of the $\bar{x}$ control chart for positive mean shifts, but the out-of-control ARLs of the neural network for negative mean shifts are considerably smaller, as seen in the figure. The out-of-control ARLs of the $\bar{x}$ chart become larger as the mean shifts approach -.75 δ (ARL = 1876) and become smaller again after that point. This result can be explained by examining the shape of the distribution (figure 12). The nonnormal distribution with skewness = 1 and kurtosis = 3 has a long tail on the right-hand side and short but heavy tail on the left-hand side. When the distribution is shifted to the right, the light long tail on the right-hand side of the distribution moves beyond the upper control limit of the $\bar{x}$ control chart. This implies that, when the distribution is shifted to the right, the probability for the data observations to be beyond the upper control limit gradually increases just as in a normal case. On the other hand, when the distribution is shifted to the left, the probability for data observations to be beyond the lower control limit increases dramatically due to the heavy tail on the left-hand side of the positively skewed distribution.

Case 5 (skewness = 1.0, kurtosis = 4.0)

In case 5, nonnormal data with skewness of 1 and kurtosis of 4 were examined. A skewness of 1 indicates that the underlying distribution is positively skewed, and a kurtosis of 4 indicates that the peakedness of the distribution is higher than for the normal distribution. The shape of the distribution is similar to that of case 4, but is more peaked, as shown in figure 14. As a result of the high kurtosis, the distribution of the averages is closer to that of the normal distribution than in case 4. As shown in table 15, the values of ρ_1, ρ_2, and ρ_3 are about the same as those in case 4 except for the values of ρ_2 for the small negative mean shifts (-.25δ to -1.5δ). This result is mainly due to the similar shapes of the distributions in these two cases.

The values of ρ_1 for positive mean shifts are approximately one. This implies that the neural network performs as well as the $\overline{x}$ control chart for positive mean shifts. The values of ρ_1 for negative mean shifts are much smaller than one. This means that the performance of the neural network is considerably superior to that of the $\overline{x}$ control chart for negative mean shifts. Furthermore, the values of ρ_2 for the large mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one. This implies that the neural network performs better than the corresponding CUSUM chart in detecting out-of-control states for these mean shifts. However, the values of ρ_2 for the small mean shifts ($\pm 0.25\delta$ to $\pm 2\delta$) are much larger than one. This implies that the neural network performs worse than the corresponding CUSUM control chart. On the other hand, the values of ρ_3 for the small mean shifts ($\pm 0.25\delta$ to $\pm 2\delta$) are much larger than one. This implies that the $\overline{x}$ control chart performs worse than the corresponding CUSUM control chart for small mean shifts. In contrast, the values of ρ_3 for large mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one which implies that the $\overline{x}$ control chart performs better than the corresponding CUSUM control chart for large mean shifts.

These results are supported statistically by the results from the analysis of variance, which shows that there is a significant difference among the performances of the three methods at $\alpha = .05$ (p-value = .0243) (table 16). The multiple comparisons indicate that the CUSUM chart and the neural network perform better than the $\overline{x}$ control chart at $\alpha = .05$ for this positively skewed and kurtotic distribution. The multiple comparisons also show that there is no significant difference between the CUSUM chart and the neural network at $\alpha = .05$. The results from the analysis of variance indicate that there is no significant effect due to mean shifts (p-value = .0516).

This result is also shown in figure 15. The ARL curve of the neural network matches that of the $\overline{x}$ control chart for the positive mean shifts, but the out-of-control ARLs of the neural network for the negative mean shifts are considerably smaller, as seen in the figure. The out-of-control ARLs of the $\overline{x}$ chart become larger as the mean shifts approach $-.50\delta$ (ARL = 664) and become smaller again after that point. This result can be explained by examining the shape of the distribution as described in case 4.

Case 6 (skewness = 1.0, kurtosis = 5.0)

In case 6, nonnormal data with skewness of 1 and kurtosis of 5 were examined. A skewness of 1 indicates that the underlying distribution is positively skewed, while a kurtosis of 5 indicates that the peakedness of the distribution is much higher than that of the normal distribution. The shape of the distribution is similar to those of cases 4 and 5, but is much more peaked, as shown in figure 16. As shown in table 17, the values of ρ_1 for positive mean shifts are approximately one. This implies that the neural network performs as well as the $\bar{x}$ control chart for positive mean shifts. The values of ρ_1 for negative mean shifts are smaller than one. However, the values of ρ_1 for negative mean shifts are much larger than those in cases 4 and 5. This means that the performance of the neural network is still superior to that of the $\bar{x}$ control chart for negative mean shifts, but not nearly as much so as in cases 4.

This result can be attributed to the distribution of the sample averages, which is much closer to symmetric than in cases 4 and 5 as seen in figure 16. The performance of the neural networks proposed in this study seems to be equivalent to that of $\bar{x}$ control charts when the distribution is symmetric, as demonstrated in the results from cases 1, 2, and 3. On the other hand, the values of ρ_2 for large mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one. This implies that the neural network performs better than the corresponding CUSUM chart in detecting out-of-control states for these mean shifts. However, the values of ρ_2 for small mean shifts ($\pm 0.25\delta$ to $\pm 2\delta$) are much larger than one. This implies that the neural network performs worse than the corresponding CUSUM control chart for these mean shifts. The values of ρ_3 for small mean shifts ($\pm 0.25\delta$ to $\pm 2\delta$) are much larger than one which implies that for these mean shifts the $\bar{x}$ control chart performs worse than the corresponding CUSUM control chart. The values of ρ_3 for large mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one. This implies that the $\bar{x}$ control chart performs better than the corresponding CUSUM control chart for these mean shifts. Overall, the ρ_1, ρ_2, and ρ_3 ratios in this case 6 are similar to those in case 5. This implies that the relative performance of the three methods in case 6 is about the same as in case 5.

The results above are statistically supported by the results from the analysis of variance. The analysis of variance shows that there is a significant difference among the performances of the three methods at $\alpha = .05$ (p-value = .0044) (table 18). The multiple comparisons indicate that the CUSUM chart performs better than the $\overline{x}$ control chart and the neural network at $\alpha = .05$. The multiple comparisons also show that there is no significant difference between the $\overline{x}$ chart and the neural network at $\alpha = .05$. The results also indicate that there is significant effect due to mean shifts (p-value = .0001). This result is also shown in figure 17. The ARL curve for the neural network matches that for the $\overline{x}$ control chart for positive mean shifts, while smaller out-of-control ARLs for the neural network for negative mean shifts can be seen in the figure. The ARL curves for the three methods are much more symmetric than those for cases 4 and 5 due to the fact that the shape of the distribution of sample averages is much closer to being symmetric.

Case 7 (skewness = 2.0, kurtosis = 6.2)

In case 7, nonnormal data with skewness of 2 and kurtosis of 6.2 were examined. A skewness of 2 indicates that the underlying distribution is severely skewed to the right, while a kurtosis of 6.2 indicates that the distribution is highly kurtotic (figure 18). As shown in table 19, the values of ρ_1, ρ_2, and ρ_3 for positive mean shifts are about the same as those in the previous cases, 1-6. This implies that the relative performance of the three methods is about the same in detecting positive out-of-control states in all the cases 1-7. However, notice that the values of ρ_2 and ρ_3 tend to be smaller as the distributions become more skewed and/or more kurtotic. For example, the value of ρ_3 for a mean shift of .75 is 5.12 in case 2 (normal distribution), but becomes smaller, 5.0, 3.16, and 2.53 in cases 3, 4, and 7, respectively. On the other hand, the values of ρ_1 and ρ_2 for all the negative mean shifts are much smaller than one. The values are also much smaller than those in the previous cases 1-6. This means that the performance of the neural network is considerably superior to that of the $\bar{x}$ and CUSUM control charts for negative mean shifts. The values of ρ_3 for small negative mean shifts (-.25δ to -2δ) are larger than one. Specifically, the values of ρ_3 for negative mean shifts (-0.5δ to -1.5δ) are 14.04, 45.04, 118.93, and 556.74, respectively. This implies that $\bar{x}$ control chart performs much worse that the corresponding CUSUM chart for small mean shifts. The values of ρ_3 for the large negative mean shifts ($\pm 3\delta$ to $\pm 5\delta$) are smaller than one which suggests that the $\bar{x}$ control chart performs better than the corresponding CUSUM chart for these mean shifts.

The results from the analysis of variance show that there is no statistically significant difference at $\alpha = .05$ among the performances of the three methods (p-value = .0662) (table 20). However, the means of the ARLs of the three methods indicate that the neural network and the CUSUM chart perform much better than the corresponding $\bar{x}$ chart (the mean ARL of the $\bar{x}$ chart = 255.6; the mean ARL of the CUSUM chart = 13.4; and the mean ARL of the neural network = 13.2). The results also indicate that there is no statistically significant effect due to the mean shifts (p-value = .5028).

This result is also shown in the figure 19. The ARL curve of the neural network matches that of the $\bar{x}$ control chart for positive mean shifts, but the considerably smaller out-of-control ARLs of the neural network are shown in the figure for negative mean shifts. The out-of-control ARLs of the $\bar{x}$ chart become larger as mean shifts approach -1.5δ (ARL = 2449) and become dramatically smaller after that point. This result can be explained by

Table 6. ARLs of the three methods, Case 1 (skewness = 0, kurtosis = 2, n =5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3
0	787.35	787.62	786.85				0						
0.25	479.05	185.67	514.86	1.07	2.77	2.58	-0.25	480.8	183.75	451.93	0.94	2.46	0.26
0.5	202.53	43.93	218.27	1.08	4.97	4.61	-0.5	209.44	44.15	197.37	0.94	4.47	4.74
0.75	93.04	18.73	98.71	1.06	5.27	4.97	-0.75	94.58	18.74	89.38	0.95	4.77	5.05
1	46.11	11.26	48.56	1.05	4.31	4.10	-1	46.59	11.33	44.35	0.95	3.91	4.11
1.5	14.72	6.23	15.36	1.04	2.47	2.36	-1.5	14.53	6.22	14.06	0.97	2.26	2.34
2	6.15	4.34	6.37	1.04	1.47	1.42	-2	6.06	4.32	5.86	0.97	1.36	1.40
3	2.02	2.76	2.06	1.02	0.75	0.73	-3	1.98	2.75	1.97	0.99	0.72	0.72
4	1.2	2.12	1.21	1.01	0.57	0.57	-4	1.19	2.11	1.19	1.00	0.56	0.56
5	1.02	1.83	1.02	1.00	0.56	0.56	-5	1.02	1.83	1.02	1.00	0.56	0.56

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 7-a & b. Analysis of variance
case 1 (skewness = 0, kurtosis = 2, n=5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	479.05	185.67	514.86
0.5	202.53	43.93	218.27
0.75	93.04	18.73	98.71
1	46.11	11.26	48.56
1.5	14.72	6.23	15.36
2	6.15	4.34	6.37
3	2.02	2.76	2.06
4	1.2	2.12	1.21
5	1.02	1.83	1.02
-0.25	480.8	183.75	451.93
-0.5	209.44	44.15	197.37
-0.75	94.58	18.74	89.38
-1	46.59	11.33	44.35
-1.5	14.53	6.22	14.06
-2	6.06	4.32	5.86
-3	1.98	2.75	1.97
-4	1.19	2.11	1.19
-5	1.02	1.83	1.02

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	49373	24736	7.36	.0022
block	17	764338	44961	13.37	.0001
error	34	114336	3362		
total	53	928148			

Table 8. Performance ranking from multiple comparisons for nonnormality (n = 5) (Tukey tests)

Case	$\bar{x}$ chart	CUSUM chart	Neural network
1	2	1	2
2	2	1	2
3	2	1	2
4	1 (3)	1 (1)	1 (2)
5	2	1 (1)	1 (2)
6	2	1	2
7	1 (2)	1 (1)	1 (1)
8	1 (2)	1 (1)	1 (1)
9	2	1	1

* The numbers in parentheses represent the order of means that are not statistically significant.

Table 9. ARLs of the three methods, Case 2 (skewness = 0, kurtosis = 3, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3
0	372.28	371.87	371.96				0						
0.25	280.16	122.54	272.85	0.97	2.23	2.29	-0.25	282.48	122.7	292.57	1.04	2.38	2.30
0.5	157.1	35.28	151.09	0.96	4.28	4.45	-0.5	153.81	35.53	163.16	1.06	4.59	4.33
0.75	83.27	16.27	79.48	0.95	4.89	5.12	-0.75	81.06	16.11	85.7	1.06	5.32	5.03
1	44.9	9.93	43.48	0.97	4.38	4.52	-1	43.94	9.93	45.87	1.04	4.62	4.42
1.5	15.01	5.57	14.73	0.98	2.64	2.69	-1.5	14.99	5.53	15.68	1.05	2.84	2.71
2	6.28	3.86	6.17	0.98	1.60	1.63	-2	6.31	3.88	6.51	1.03	1.68	1.63
3	1.98	2.48	1.97	0.99	0.79	0.80	-3	2.02	2.49	2.05	1.01	0.82	0.81
4	1.19	1.96	1.19	1.00	0.61	0.61	-4	1.18	1.96	1.19	1.01	0.61	0.60
5	1.02	1.6	1.02	1.00	0.64	0.64	-5	1.02	1.61	1.02	1.00	0.63	0.63

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 10-a & b. Analysis of variance
case 2 (skewness = 0, kurtosis = 3, n=5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	280.16	122.54	272.85
0.5	157.1	35.28	151.09
0.75	83.27	16.27	79.48
1	44.9	9.93	43.48
1.5	15.01	5.57	14.73
2	6.28	3.86	6.17
3	1.98	2.48	1.97
4	1.19	1.96	1.19
5	1.02	1.6	1.02
-0.25	282.48	122.7	292.57
-0.5	153.81	35.53	163.16
-0.75	81.06	16.11	85.7
-1	43.94	9.93	45.87
-1.5	14.99	5.53	15.68
-2	6.31	3.88	6.51
-3	2.02	2.49	2.05
-4	1.18	1.96	1.19
-5	1.02	1.61	1.02

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	22679	11339	10.03	.0004
block	17	283260	16662	14.73	.0001
error	34	38457	1131		
total	53	344397			

Table 11. ARLs of the three methods, Case 3 (skewness = 0, kurtosis = 4, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3
0	260.27	259.65	260.42				0						
0.25	205.27	101.01	215.31	1.05	2.13	2.03	-0.25	208.63	101.36	202.11	0.97	1.99	2.06
0.5	131.06	32.17	137.85	1.05	4.29	4.07	-0.5	127.6	31.5	121.93	0.96	3.87	4.05
0.75	75.6	15.12	80.22	1.06	5.31	5.00	-0.75	73.82	14.96	70.25	0.95	4.70	4.93
1	43.15	9.33	45.98	1.07	4.93	4.62	-1	41.75	9.36	40.33	0.97	4.31	4.46
1.5	15.27	5.26	16.24	1.06	3.09	2.90	-1.5	15.35	5.22	14.81	0.96	2.84	2.94
2	6.41	3.65	6.77	1.06	1.85	1.76	-2	6.41	3.67	6.3	0.98	1.72	1.75
3	1.98	2.37	2.04	1.03	0.86	0.84	-3	2.02	2.38	2	0.99	0.84	0.85
4	1.18	1.87	1.19	1.01	0.64	0.63	-4	1.18	1.89	1.17	0.99	0.62	0.62
5	1.02	1.48	1.03	1.01	0.70	0.69	-5	1.02	1.5	1.02	1.00	0.68	0.68

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 12-a & b. Analysis of variance
case 3 (skewness = 0, kurtosis = 4, n=5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	205.27	101.01	215.31
0.5	131.06	32.17	137.85
0.75	75.6	15.12	80.22
1	43.15	9.33	45.98
1.5	15.27	5.26	16.24
2	6.41	3.65	6.77
3	1.98	2.37	2.04
4	1.18	1.87	1.19
5	1.02	1.48	1.03
-0.25	208.63	101.36	202.11
-0.5	127.6	31.5	121.93
-0.75	73.82	14.96	70.25
-1	41.75	9.36	40.33
-1.5	15.35	5.22	14.81
-2	6.41	3.67	6.3
-3	2.02	2.38	2
-4	1.18	1.89	1.17
-5	1.02	1.5	1.02

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	14171	7085	11.84	.0001
block	17	164560	9680	16.17	.0001
error	34	20355	598		
total	53	199087			

Table 13. ARLs of the three methods, Case 4 (skewness = 1, kurtosis = 3, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3
0	256.7	255.86	256.38				0						
0.25	142.29	90.04	142.80	1.00	1.59	1.58	-0.25	477.9	109.72	96.25	0.20	0.88	4.36
0.5	81.2	31.36	81.71	1.01	2.61	2.59	-0.5	936.22	31.13	29.52	0.03	0.95	30.07
0.75	48.7	15.4	49.01	1.01	3.18	3.16	-0.75	1876.82	14.57	13.7	0.01	0.94	128.81
1	30.23	9.49	30.79	1.02	3.24	3.19	-1	392.54	9.13	7.73	0.02	0.85	43.0
1.5	12.74	5.28	12.79	1.00	2.42	2.41	-1.5	19.17	5.12	3.46	0.18	0.68	3.74
2	6.13	3.68	6.22	1.01	1.69	1.67	-2	5.99	3.6	2.07	0.35	0.58	1.66
3	2.15	2.37	2.16	1.00	0.91	0.91	-3	1.88	2.38	1.25	0.66	0.53	0.79
4	1.2	1.85	1.21	1.01	0.65	0.65	-4	1.21	1.89	1.05	0.87	0.56	0.64
5	1	1.51	1	1.00	0.66	0.66	-5	1.03	1.45	1.01	0.98	0.70	0.71

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 14-a & b. Analysis of variance
case 4 (skewness = 1, kurtosis = 3, n=5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	142.29	90.04	142.80
0.5	81.2	31.36	81.71
0.75	48.7	15.4	49.01
1	30.23	9.49	30.79
1.5	12.74	5.28	12.79
2	6.13	3.68	6.22
3	2.15	2.37	2.16
4	1.20	1.85	1.21
5	1	1.51	1
-0.25	477.9	109.72	96.25
-0.5	936.22	31.13	29.52
-0.75	1876.82	14.57	13.7
-1	392.54	9.13	7.73
-1.5	19.17	5.12	3.46
-2	5.99	3.6	2.07
-3	1.88	2.38	1.25
-4	1.21	1.89	1.05
-5	1.03	1.45	1.01

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	487680	243840	3.23	.0521
block	17	1381181	81245	1.08	.4136
error	34	2569393	75570		
total	53	4438255			

Table 15. ARLs of the three methods, Case 5 (skewness = 1, kurtosis = 4, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusu	Neural	ρ_1	ρ_2	ρ_3
0	226.55	225.75	225.99				0						
0.25	131.65	86.3	132.36	1.01	1.53	1.53	-0.25	389.85	104.36	208.32	0.53	2.00	3.74
0.5	78.49	30.56	79.08	1.01	2.59	2.57	-0.5	664.78	30.37	79.52	0.12	2.62	21.89
0.75	48.59	15.12	48.81	1.00	3.23	3.21	-0.75	541.05	14.25	31.48	0.06	2.21	38.0
1	31.05	9.4	31.26	1.01	3.33	3.30	-1	131.67	9.05	15.36	0.12	1.70	14.55
1.5	13.17	5.2	13.32	1.01	2.56	2.53	-1.5	19.1	5.04	5.38	0.28	1.07	3.79
2	6.22	3.62	6.32	1.02	1.75	1.72	-2	6.13	3.56	2.73	0.45	0.77	1.72
3	2.15	2.34	2.16	1.00	0.92	0.92	-3	1.88	2.34	1.37	0.73	0.59	0.80
4	1.19	1.84	1.20	1.01	0.65	0.65	-4	1.18	1.84	1.07	0.91	0.58	0.64
5	1.01	1.48	1.01	1.00	0.68	0.68	-5	1.03	1.42	1.01	0.98	0.71	0.73

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 16-a & b. Analysis of variance
case 5 (skewness = 1, kurtosis = 4, n = 5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	131.65	86.3	132.36
0.5	78.49	30.56	79.08
0.75	48.59	15.12	48.81
1	31.05	9.4	31.26
1.5	13.17	5.2	13.32
2	6.22	3.62	6.32
3	2.15	2.34	2.16
4	1.19	1.84	1.20
5	1.01	1.48	1.01
-0.25	389.85	104.36	208.32
-0.5	664.78	30.37	79.52
-0.75	541.05	14.25	31.48
-1	131.67	9.05	15.36
-1.5	19.1	5.04	5.38
-2	6.13	3.56	2.73
-3	1.88	2.34	1.37
-4	1.18	1.84	1.07
-5	1.03	1.42	1.01

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	94998	47499	4.15	.0243
block	17	373486	21969	1.92	.0516
error	34	388785	11434		
total	53	857270			

Table 17. ARLs of the three methods, Case 6 (skewness = 1, kurtosis = 5, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3
0	193.76	194.28	193.81				0						
0.25	121.69	76.87	123.69	1.02	1.61	1.58	-0.25	297.93	90.83	267.52	0.90	2.95	3.28
0.5	76.02	29.18	77.54	1.02	2.66	2.61	-0.5	336.35	28.46	223.12	0.66	7.84	11.82
0.75	48.06	14.37	48.73	1.01	3.39	3.34	-0.75	203.6	13.71	112	0.55	8.17	14.85
1	30.6	8.97	31.45	1.03	3.51	3.41	-1	88.08	8.6	49.66	0.56	5.77	10.24
1.5	13.23	5.02	13.66	1.03	2.72	2.64	-1.5	19.19	4.9	12.77	0.67	2.61	3.92
2	6.41	3.5	6.57	1.02	1.88	1.83	-2	6.46	3.46	4.85	0.75	1.40	1.87
3	2.11	2.26	2.15	1.02	0.95	0.93	-3	1.9	2.28	1.69	0.89	0.74	0.83
4	1.18	1.78	1.19	1.01	0.67	0.66	-4	1.18	1.78	1.14	0.97	0.64	0.66
5	1.01	1.4	1.01	1.00	0.72	0.72	-5	1.03	1.36	1.02	0.99	0.75	0.76

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 18-a & b. Analysis of variance
case 6 (skewness = 1, kurtosis = 5, n = 5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	121.69	76.87	123.69
0.5	76.02	29.18	77.54
0.75	48.06	14.37	48.73
1	30.6	8.97	31.45
1.5	13.23	5.02	13.66
2	6.41	3.5	6.57
3	2.11	2.26	2.15
4	1.18	1.78	1.19
5	1.01	1.4	1.01
-0.25	297.93	90.83	267.52
-0.5	336.35	28.46	223.12
-0.75	203.6	13.71	112
-1	88.08	8.6	49.66
-1.5	19.19	4.9	12.77
-2	6.46	3.46	4.85
-3	1.9	2.28	1.69
-4	1.18	1.78	1.14
-5	1.03	1.36	1.02

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	26973	13486	6.38	.0044
block	17	235662	13862	6.55	.0001
error	34	71909	2114		
total	53	334542			

Table 19. ARLs of the three methods, Case 7 (skewness = 2, kurtosis = 6.2, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusu	Neur	ρ_1	ρ_2	ρ_3
0	125.58	126.17	125.78				0						
0.25	81.04	57.08	81.07	1.00	1.42	1.42	-0.25	195.17	68.66	6.62	0.03	0.10	2.84
0.5	52.76	26.04	53.30	1.01	2.05	2.03	-0.5	319.48	22.75	3.84	0.01	0.17	14.04
0.75	34.83	13.76	35.22	1.01	2.56	2.53	-0.75	523.42	11.62	2.78	0.01	0.24	45.04
1	23.66	8.54	23.68	1.00	2.77	2.77	-1	888.38	7.47	2.24	0.00	0.30	118.93
1.5	11.77	4.76	11.91	1.01	2.50	2.47	-1.5	2449.67	4.4	1.61	0.00	0.37	556.74
2	6.28	3.29	6.36	1.01	1.93	1.91	-2	5.48	3.13	1.3	0.24	0.42	1.75
3	2.27	2.12	2.29	1.01	1.08	1.07	-3	1.78	2.22	1.07	0.60	0.48	0.80
4	1.22	1.67	1.23	1.01	0.74	0.73	-4	1.18	1.57	1.02	0.86	0.65	0.75
5	1	1.32	1	1.00	0.76	0.76	-5	1.04	1.22	1	0.96	0.82	0.85

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 20-a & b. Analysis of variance
case 7 (skewness = 2, kurtosis = 6.2, n = 5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	81.04	57.08	81.07
0.5	52.76	26.04	53.30
0.75	34.83	13.76	35.22
1	23.66	8.54	23.68
1.5	11.77	4.76	11.91
2	6.28	3.29	6.36
3	2.27	2.12	2.29
4	1.22	1.67	1.23
5	1	1.32	1
-0.25	195.17	68.66	6.62
-0.5	319.48	22.75	3.84
-0.75	523.42	11.62	2.78
-1	888.38	7.47	2.24
-1.5	2449.67	4.4	1.61
-2	5.48	3.13	1.3
-3	1.78	2.22	1.07
-4	1.18	1.57	1.02
-5	1.04	1.22	1

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	704334	352167	2.94	.0662
block	17	1987518	116912	.98	.5028
error	34	4067087	119620		
total	53	6758940			

Table 21. ARLs of the three methods, Case 8 (skewness = 2, kurtosis = 7.2, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusu	Neural	ρ_1	ρ_2	ρ_3
0	119.09	119.87	119.42				0						
0.25	77.6	54.79	78.96	1.02	1.44	1.42	-0.25	182.48	70.4	10.29	0.06	0.15	2.59
0.5	52.09	24.96	52.72	1.01	2.11	2.09	-0.5	285.78	23.51	5.17	0.02	0.22	12.16
0.75	34.77	13.35	35.26	1.01	2.64	2.60	-0.75	449.97	11.84	3.37	0.01	0.28	38.00
1	24.11	8.38	24.19	1.00	2.89	2.88	-1	713.79	7.57	2.53	0.00	0.33	94.29
1.5	11.95	4.66	12.10	1.01	2.60	2.56	-1.5	1600	4.42	1.75	0.00	0.40	361.99
2	6.24	3.25	6.32	1.01	1.94	1.92	-2	5.91	3.13	1.35	0.23	0.43	1.89
3	2.27	2.1	2.30	1.01	1.10	1.08	-3	1.77	2.14	1.08	0.61	0.50	0.83
4	1.2	1.66	1.21	1.01	0.73	0.72	-4	1.19	1.57	1.02	0.86	0.65	0.76
5	1	1.28	1	1.00	0.78	0.78	-5	1.04	1.21	1	0.96	0.83	0.86

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 22-a & b. Analysis of variance
case 8 (skewness = 2, kurtosis = 7.2, n = 5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	77.6	54.79	78.96
0.5	52.09	24.96	52.72
0.75	34.77	13.35	35.26
1	24.11	8.38	24.19
1.5	11.95	4.66	12.10
2	6.24	3.25	6.32
3	2.27	2.1	2.30
4	1.2	1.66	1.21
5	1	1.28	1
-0.25	182.48	70.4	10.29
-0.5	285.78	23.51	5.17
-0.75	449.97	11.84	3.37
-1	713.79	7.57	2.53
-1.5	1600	4.42	1.75
-2	5.91	3.13	1.35
-3	1.77	2.14	1.08
-4	1.19	1.57	1.02
-5	1.04	1.21	1

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	382166	191083	3.51	.0410
block	17	900359	52962	.97	.5062
error	34	1849277	54390		
total	53	3131802			

Table 23. ARLs of the three methods, Case 9 (skewness = 2, kurtosis = 8.8, n = 5)

Shift	$\bar{x}$ chart	Cusum	Neural	ρ_1	ρ_2	ρ_3	Shift	$\bar{x}$ chart	Cusu	Neural	ρ_1	ρ_2	ρ_3
0	116.86	117.14	117.04				0						
0.25	77.69	55.15	77.55	1.00	1.41	1.41	-0.25	175.65	69.13	57.31	0.33	0.83	2.54
0.5	52.81	25.16	52.43	0.99	2.08	2.10	-0.5	265.88	23.51	12.72	0.05	0.54	11.31
0.75	36.03	13.36	35.74	0.99	2.68	2.70	-0.75	400.02	11.85	6.31	0.02	0.53	33.76
1	24.84	8.33	24.72	1.00	2.97	2.98	-1	597.13	7.59	3.9	0.01	0.51	78.67
1.5	12.28	4.64	12.32	1.00	2.66	2.65	-1.5	74.26	4.43	2.2	0.03	0.50	16.76
2	6.36	3.24	6.36	1.00	1.96	1.96	-2	6.16	3.13	1.54	0.25	0.49	1.97
3	2.27	2.1	2.27	1.00	1.08	1.08	-3	1.77	2.13	1.13	0.64	0.53	0.83
4	1.2	1.66	1.20	1.00	0.72	0.72	-4	1.18	1.57	1.03	0.87	0.66	0.75
5	1	1.27	1	1.00	0.79	0.79	-5	1.04	1.21	1.01	0.97	0.83	0.86

* ρ_1 = the ratio, ARL of neural network / ARL of $\bar{x}$ chart
* ρ_2 = the ratio, ARL of neural network / ARL of cusum chart
* ρ_3 = the ratio, ARL of $\bar{x}$ chart / ARL of cusum chart

Table 24-a & b. Analysis of variance
case 9 (skewness = 2, kurtosis = 8.8, n = 5)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	77.69	55.15	77.55
0.5	52.81	25.16	52.43
0.75	36.03	13.36	35.74
1	24.84	8.33	24.72
1.5	12.28	4.64	12.32
2	6.36	3.24	6.36
3	2.27	2.1	2.27
4	1.2	1.66	1.20
5	1	1.27	1
-0.25	175.65	69.13	57.31
-0.5	265.88	23.51	12.72
-0.75	400.02	11.85	6.31
-1	597.13	7.59	3.9
-1.5	74.26	4.43	2.2
-2	6.16	3.13	1.54
-3	1.77	2.13	1.13
-4	1.18	1.57	1.03
-5	1.04	1.21	1.01

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	79862	39931	4.39	.0202
block	17	173114	10183	1.12	.3771
error	34	309460	9101		
total	53	562437			

examining the shape of the distribution (figure 18). The nonnormal distribution with skewness = 2 and kurtosis = 6.2 has a long tail on the right side and a short but heavy tail on the left side. When the distribution is shifted to the right, the light long right tail of the distribution moves beyond the upper control limit of the $\overline{x}$ control chart. This implies that when the distribution is shifted to the right, the probability for the data observations to be beyond the upper control limit gradually increases just as in a normal case. Therefore, the out-of-control ARLs for positive mean shifts become gradually shorter. On the other hand, when the distribution is shifted to the left, the probability for the data observations to be beyond the both upper and lower control limits gets smaller. The heavy negative tail of the distribution lies within the lower control limit with small negative mean shifts (up to -1.5δ), while the positive tail beyond the upper control limit becomes smaller. However, the heavy negative tail of the distribution moves beyond the lower control limit after that point of the mean shift. Therefore, the probability of the distribution to be beyond the lower control limit dramatically increases, resulting in a dramatic decrease of the out-of-control ARLs. On the other hand, the out-of-control ARLs of the CUSUM chart and the neural network decrease gradually and consistently as the magnitude of the mean shifts increases for both positive and negative mean shifts.

<u>Case 8 (skewness = 2.0, kurtosis = 7.2)</u>

In case 8, nonnormal data with skewness of 2 and kurtosis of 7.2 were examined. A skewness of 2 indicates that the underlying distribution is severely skewed to the right, while a kurtosis of 7.2 indicates that the distribution is highly kurtotic (figure 20). Basically, there are no differences in the performance comparison results between cases 7 and 8, as can be seen from tables 19 and 21.

The results from the analysis of variance show that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ (p-value = .041) (table 22). However, the multiple comparisons indicate that there is no pairwise difference among the methods. The means of the ARLs of the three methods indicate that the neural network and the CUSUM chart perform much better than the corresponding $\bar{x}$ chart (the mean ARL of the $\bar{x}$ chart = 191.84; the mean ARL of the CUSUM chart = 13.35; and the mean ARL of the neural network = 13.42). The results also indicate that there is no statistically significant effect due to mean shift (p-value = .5062).

These results are also shown in figure 21. The ARL curve of the neural network matches that of the $\bar{x}$ control chart for positive mean shifts, but the out-of-control ARLs of the neural network are considerably smaller than those of the $\bar{x}$ chart for negative mean shifts, as can be seen in the figure. The out-of-control ARLs of the $\bar{x}$ chart become larger as the mean shifts approach -1.5 δ (ARL = 1600) and become dramatically smaller after that point. One main difference between cases 7 and 8 is that the out-of-control ARLs of case 8 are smaller than those of case 7 for negative mean shifts. This result can be attributed to the more kurtotic shape of the distribution, which results in a distribution of sample averages that is more symmetric.

Case 9 (skewness = 2.0, kurtosis = 8.8)

In case 9, nonnormal data with skewness of 2 and kurtosis of 8.8 were examined. This distribution is very close to an exponential distribution (skewness = 2 and kurtosis = 9) (figure 22). As mentioned before, exponentially distributed observations arise naturally in the context of monitoring the occurrence rate of rare events. Generally, the results for this case do not differ from those of cases 7 and 8, as can be seen by a comparison of table 19, 21, and 23.

The results from the analysis of variance show that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ (p-value = .0202) (table 24). The multiple comparisons indicate that the CUSUM chart and the neural network perform better than the $\overline{x}$ control chart at $\alpha = .05$. The multiple comparisons also show that there is no significant difference between the CUSUM chart and the neural network at $\alpha = .05$. The results also indicate that there is no statistically significant effect due to the mean shifts (p-value = .3771).

These results are also shown in figure 23. The ARL curve of the neural network matches that of the $\overline{x}$ control chart for positive mean shifts, but the considerably smaller out-of-control ARLs of the neural network are shown in the figure for negative mean shifts. The out-of-control ARLs of the $\overline{x}$ chart become larger as the mean shifts approach -1.0δ (ARL = 597) and become dramatically smaller after that point of the mean shift. One main difference between this case and the previous two cases, 7 and 8, is that the out-of-control ARLs of case 9 are smaller than those of cases 7 and 8 for the negative mean shifts. This result can be attributed to the more kurtotic shape of the distribution, which results in a distribution of sample averages that is more symmetric.

Effects of Degrees of Nonnormality

Clearly, the effects of skewness and kurtosis on the performance of a control chart for averages can be substantial because nonnormality can dramatically affect the in-control ARLs as well as the out-of-control ARLs for standard control charts. The more skewed and/or the kurtotic, the more nonnormal the distribution of the quality characteristic. Both of these aspects of nonnormality are important in the consideration of the effects of nonnormality on the operating characteristics of control charts. The quality management decision maker is well advised to consider neural networks when nonnormality is present in the underlying distribution of a quality characteristic. In the previous section, the effects of degrees of nonnormality on out of control ARLs were examined for each of the nine nonnormal cases. In this section, the effects of degrees of nonnormality on in-control ARLs are examined. Controlling Type I error rates as measured by in-control ARLs is as important as controlling Type II error rates as measured by out-of-control ARLs because responding to false alarms are sometimes more expensive than taking no action.

The effect of kurtosis on in-control ARLs is significant. As shown in table 25, the in-control ARLs for cases, 1, 2, and 3 become dramatically smaller going from 787 to 372 and then to 260. Notice that the kurtosis of the cases becomes larger, starting at 2, then moving to 3 and 4, while their skewness is the same. The in-control ARLs for cases, 4, 5, and 6 become smaller going from 257 to 224 and then to 194. Notice that the kurtosis for these cases becomes larger from 3 to 4, whereas their skewness is the same. The same happens for the last group of cases, 7, 8, and 9. Based on the three groups of cases above, it can be concluded that higher kurtosis causes higher false alarm rates.

The effect of skewness is important because the effects of skewness on the in-control ARLs are significant. The in-control ARLs of cases 2 and 4 get smaller, going from 372 to 257 (table 25). Notice that the skewness of these cases becomes larger, going from 0 to 1, while their kurtosis is the same. The in-control ARLs of cases 3 and 5 also become smaller, going from 260 to 227. Notice that the skewness of these cases becomes larger, going from 0 to 1, while their kurtosis is the same. Based on the two groups of cases above, it can be concluded that higher skewness causes higher false alarm rates. Just as with the effect of kurtosis, the effect of large skewness values is an inappropriately large value of false alarm rates. In summary, in-control ARLs become smaller as the degree of nonnormality as measured by skewness and kurtosis, become higher.

Table 25. The effects of degrees of nonnormality on in-control ARLs

Case	Skewness	Kurtosis	ARLs	Notes
1	0	2	787	Almost uniform distribution
2	0	3	372	Normal
3	0	4	260	Symmetric and peaked
4	1	3	257	Positively skewness
5	1	4	227	Positively skewed and peaked
6	1	5	194	Positively skewed and peaked
7	2	6.2	126	Highly positively skewed and highly eaked
8	2	7.2	119	Highly positively skewed and severely peaked
9	2	8.8	116	Almost exponential distribution

Effects of Mean Shifts

The eighteen different mean shifts are divided into four categories: small and intermediate mean shifts and large mean shifts for both the positive and negative directions. This categorization was selected because it allows a concise and comprehensive examination of the effects of the mean shifts on the performance of the three methods.

Small and intermediate positive mean shifts (0.25δ to 2δ)

The analysis of variance shows that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ for each of the mean shifts. The analysis of variance results for the mean shifts are shown in tables26 - 31, respectively. The multiple comparisons indicate that the CUSUM charts perform better than the corresponding $\bar{x}$ control charts and neural networks at $\alpha = .05$ for each of the mean shifts. The multiple comparisons also show that there is no significant difference between the $\bar{x}$ charts and the neural networks at $\alpha = .05$ for any of the mean shifts. The multiple comparisons for all four categories are shown in table 32.

Small and intermediate negative mean shifts (-0.25δ to -2δ)

The analysis of variance shows that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ for each of the mean shifts. The resulting analysis of variance tables for the mean shifts are shown in tables 33-38, respectively. The multiple comparisons (table 32) indicate that CUSUM charts and neural networks perform better than the corresponding $\bar{x}$ charts at $\alpha = .05$ for any of the mean shifts. The multiple comparisons also show that there is no significant difference between the CUSUM charts and the neural networks at $\alpha = .05$ for each of the mean shifts. It is well known that a CUSUM chart performs better than a $\bar{x}$ chart for small and intermediate mean shifts. These findings from the study shows that the performance of the three methods when distributions are nonnormal is the same as when the distribution is normal for small and intermediate negative mean shifts .

Large positive mean shifts (3.0δ to 5.0δ)

The analysis of variance shows that there is no significant statistical difference among the performances of the three methods at $\alpha = .05$ for the 3.0δ mean shifts. On the other hand, the results show that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ for mean shifts of 4.0δ and 5.0δ. The resulting analysis of variance tables for the mean shifts are shown in tables 39 - 41. The multiple comparisons indicate that $\bar{x}$ charts and neural networks perform better than the corresponding CUSUM charts at $\alpha = .05$ for the 4.0δ and 5.0δ mean shifts. The multiple comparisons also show that there is no significant difference between the $\bar{x}$ charts and the neural networks at $\alpha = .05$ for these mean shifts.

Large negative mean shifts (-3.0δ to -5.0δ)

The analysis of variance shows that there is a statistically significant difference among the performances of the three methods at $\alpha = .05$ for all of the large negative mean shifts. The resulting analysis of variance tables for the mean shifts are shown in tables 42 - 44, respectively. The multiple comparisons for the -3δ mean shift indicate that neural networks perform better than the corresponding $\bar{x}$ charts and the CUSUM charts at $\alpha = .05$. The multiple comparisons also show that the $\bar{x}$ charts perform better than the CUSUM charts at $\alpha = .05$ for these mean shifts.

In addition, the multiple comparisons for mean shifts of -4δ and -5δ indicate that neural networks and $\bar{x}$ charts perform better than the corresponding CUSUM charts at $\alpha =$ .05. It is well known that a $\bar{x}$ chart performs better than CUSUM chart for large mean shifts when a distribution is normal. The experiment indicates that the neural networks proposed in this study perform statistically as well as the corresponding $\bar{x}$ charts for large negative mean shifts. However, the means of ARLs of the two methods indicate that the neural networks perform better than the corresponding $\bar{x}$ charts for most of the eight different types of nonnormal cases for mean shifts of -4δ and -5δ.

Effects of Sample Size

Several authors have investigated the effects of departures from normality on control charts. Burr (1967) noted that the standard control charts are very robust to the normality assumption and can be employed unless the population is extremely nonnormal. Schilling and Nelson (1976) also studied the effect of nonnormality on the control limits of a $\bar{x}$ control chart. Their studies indicate that, in most cases, sample size 4 or 5 is sufficient to ensure reasonable robustness to the normality assumption. However, if we need a larger sample size to obtain the distribution of sample averages closer to normality, such an increase would be wasteful when other methods are available and easily implemented. Furthermore, in some industries, it is too costly or impossible to use large samples. In these situations, as Yourstone and Zimmer (1992) suggest, it is not reasonable to use large sample sizes to get the distribution of sample averages closer to normality. On the other hand, with the small samples typically used, the traditional control charts might perform poorly because of the potential nonnormality of production processes.

Large sample sizes can mitigate some of the effects of nonnormality in the distribution. However, it is difficult to specify a sample size which would guarantee that the normality assumption of the sample averages is satisfied. The degree of nonnormality in sample averages depends on the amount of nonnormality in the individual observations. If the amount of nonnormality is measured by a positive skewness, the skewness in the distribution of the sample averages becomes small at the rate of $1/\sqrt{n}$. That is, if the distribution of the individual observations has a skewness of 2 and a sample size as large as 5 is used, the distribution of sample averages has a skewness of .894. With a sample of 3, the skewness of the distribution of sample averages is 1.154. In this study, the effects of sample size are examined with selective combinations of the skewness and kurtosis (Yourstone and Zimmer 1992).

In this section, the effects of sample size on the performance of $\overline{x}$ charts, CUSUM charts, and neural networks are examined. First, the analyses of the performance of the three methods are conducted for some selective cases with a sample size of three. These cases include case 2 (a normal distribution), case 4 (a nonnormal distribution with skewness = 1 and kurtosis = 3), and case 7 (a nonnormal distribution with skewness = 2 and kurtosis = 6.2). A randomized block design allows comparison of the performance of the three methods. Tukey's multiple comparisons are also used, as in the previous section. The ARLs of the three methods with sample size of three for each of the three cases are shown in tables 45 - 47, respectively. The orders of the performances of the three methods for each of the three cases are held the same as in the cases with sample size 5. The results from the analysis of variance are shown in tables 48-50, respectively. The results from the multiple comparisons are shown in table 51.

Second, each of the three methods with sample size 3 is compared to those with sample size 5 to investigate the effects of sample size on the performance of each of the methods. However, it is inappropriate to compare the three methods directly because their in-control ARLs are different. Therefore, the effects of sample size on the performance of each of the three methods in detecting out-of-control states are examined taking into consideration the changes of in-control ARLs. The out-of-control ARLs characterize the speed with which a control procedure is capable of detecting a process change in a sequence of data. Therefore, a small ARL is desired when the process is out-of-control. On the other hand, a large ARL is desired when the process is in-control. The in-control ARL represents the rate of false alarms.

Table 45 - 47 contains the ratios, ψ_1, ψ_2, and ψ_3, of the ARL values for $\overline{x}$ charts, CUSUM charts, and neural networks with sample size of three to those with sample size of five, respectively. The ratios, ψ_1, ψ_2, and ψ_3, are used because they provide more concise comparisons of the performance of the methods. Assuming that controlling Type I error rates is as important as controlling Type II error rates, it can be concluded that $\overline{x}$ charts with sample size three perform better than those with sample size five, if the values of ψ_1 when the means are shifted are smaller than the values of ψ_1 when the means are not shifted. For example, the ψ_1 value for no mean shift is 0.99, and the ψ_1 value for a mean shift of $0.5\,\delta$ is 0.98 for case 2. This implies that the Type II error rate for the $\overline{x}$ chart is decreased by about 2 percent, whereas its false alarm rate is increased by about 1percent. The same

conclusion can be made for CUSUM charts and neural networks. On the other hand, the values of ψ_1 for mean shifts are 0.99, .86, and .74 for cases 2, 4, and 7, respectively. These results support the conclusion in the previous section that higher nonnormality causes higher false alarm rates.

Case 2 (skewness = 0, kurtosis = 3)

As seen in table 45, the values of ψ_1, ψ_2, and ψ_3 for both no mean shift and mean shifts are approximately one for all three methods. This implies that there is no significant effect of sample size on the performance of the three methods. The value of ψ_1, .99, when the mean is not shifted, implies that there is no big difference between in-control ARLs of $\overline{x}$ charts for sample sizes of three and five. This is because the underlying distribution is normal.

Case 4 (skewness = 1, kurtosis = 3)

As seen in the table 46, the value of ψ_1, 0.86, for no mean shift, implies that the false alarm rate of the $\overline{x}$ chart with sample size three has increased by about 14 percent. The values of ψ_3 for the mean shifts of -0.25 δ to -2 δ are 0.18, 0.26, 0.35, 0.27, 0.61, and 0.73, respectively, which are significantly smaller than the ψ_1 values for no mean shift. This implies that the neural network with sample size of three performs better than that with a sample size of five when the means are shifted. However, most of the values of ψ_1, ψ_2, and ψ_3 for mean shift conditions are higher than those for the no mean shift condition. This implies that each of the three methods with a sample size 3 performs worse than the corresponding method with sample size 5.

Case 7 (skewness = 2, kurtosis = 6.2)

As seen in table 47, the value of ψ_1, 0.74, for no mean shift, implies that the false alarm rate of the $\overline{x}$ chart with a sample size of three has increased by about 26 percent. Most of the values of ψ_1, ψ_2, and ψ_3 for mean shift conditions are significantly higher than those for the no mean shift condition. This implies that each of the three methods with a sample size of three performs worse than the corresponding method with a sample size of five.

Table 45. Effects of sample size, Case 2 (skewness = 0, kurtosis =3)

Mean shifts	$\bar{x}$ chart n = 3	$\bar{x}$ chart n = 5	ψ_1	CUSUM n = 3	CUSUM n = 5	ψ_2	Neural n = 3	Neural n = 5	ψ_3
0	368.33	372.28	0.99	368.46	371.87	0.99	369.82	371.96	0.99
0.25	280.06	280.16	1.00	121.25	122.54	0.99	277.96	272.85	1.02
0.5	153.34	157.1	0.98	34.88	35.28	0.99	153.89	151.09	1.02
0.75	81.14	83.27	0.97	16.17	16.27	0.99	81.80	79.48	1.03
1	44.65	44.9	0.99	9.9	9.93	1.00	45.10	43.48	1.04
1.5	15.16	15.01	1.01	5.5	5.57	0.99	15.33	14.73	1.04
2	6.37	6.28	1.01	3.86	3.86	1.00	6.47	6.17	1.05
3	1.99	1.98	1.01	2.48	2.48	1.00	2.01	1.97	1.02
4	1.18	1.19	0.99	1.95	1.96	0.99	1.19	1.19	1.00
5	1.02	1.02	1.00	1.6	1.6	1.00	1.02	1.02	1.00
-0.25	286.93	282.48	1.02	119.89	122.7	0.98	295.73	292.57	1.01
-0.5	156.32	153.81	1.02	34.82	35.53	0.98	162.98	163.16	1.00
-0.75	81.96	81.06	1.01	16	16.11	0.99	84.8	85.7	0.99
-1	44.38	43.94	1.01	9.91	9.93	1.00	45.77	45.87	1.00
-1.5	14.94	14.99	1.00	5.51	5.53	1.00	15.37	15.68	0.98
-2	6.33	6.31	1.00	3.83	3.88	0.99	6.49	6.51	1.00
-3	2	2.02	0.99	2.47	2.49	0.99	2.04	2.05	1.00
-4	1.19	1.18	1.01	1.96	1.2	1.00	1.20	1.19	1.01
-5	1.02	1.02	1.00	1.61	1.03	1.00	1.03	1.02	1.01

* ψ_1 = (ARL of $\bar{x}$ chart with n=3) / (ARL of $\bar{x}$ chart with n=5)

* ψ_2 = (ARL of CUSUM chart with n=3) / (ARL of CUSUM chart with n=5)

* ψ_3 = (ARL of neural network with n=3) / (ARL of neural network with n=5)

Table 46. Effects of sample size, Case 4 (skewness = 1.0, kurtosis = 3.0)

Mean shifts	$\overline{x}$ chart n = 3	$\overline{x}$ chart n = 5	ψ_1	CUSUM n = 3	CUSUM n = 5	ψ_2	Neural n = 3	Neural n = 5	ψ_3
	220.21	256.7	0.86	222.29	255.86	0.87	222.46	256.38	0.87
0.25	124.29	142.29	0.87	78.03	90.04	0.87	124.02	142.80	0.87
0.5	73.55	81.2	0.91	29.87	31.36	0.95	73.41	81.71	0.90
0.75	44.70	48.7	0.92	14.80	15.4	0.96	44.70	49.01	0.91
1	27.80	30.23	0.92	9.24	9.49	0.97	27.69	30.79	0.90
1.5	12.04	12.74	0.95	5.15	5.28	0.98	12.23	12.79	0.96
2	6.13	6.13	1.00	3.59	3.68	0.98	6.24	6.22	1.00
3	2.17	2.15	1.01	2.31	2.37	0.97	2.26	2.16	1.05
4	1.22	1.20	1.02	1.81	1.85	0.98	1.24	1.21	1.02
5	1	1	1.00	1.46	1.51	0.97	1	1	1.00
-0.25	412.16	477.9	0.86	102.56	109.72	0.93	17.17	96.25	0.18
-0.5	812.31	936.22	0.87	29.29	31.13	0.94	7.76	29.52	0.26
-0.75	1691.28	1876.82	0.90	13.84	14.57	0.95	4.85	13.7	0.35
-1	31.38	392.54	0.08	8.74	9.13	0.96	3.44	7.73	0.45
-1.5	3257.13	19.17	8.30	4.97	5.12	0.97	2.11	3.46	0.61
-2	5.57	5.99	0.93	3.50	3.6	0.97	1.52	2.07	0.73
-3	1.86	1.88	0.99	2.33	2.38	0.98	1.11	1.25	0.89
-4	1.19	1.21	0.98	1.79	1.89	0.95	1.02	1.05	0.97
-5	1.04	1.03	1.01	1.38	1.45	0.95	1	1.01	0.99

* ψ_1 = (ARL of $\overline{x}$ chart with n=3) / (ARL of $\overline{x}$ chart with n=5)

* ψ_2 = (ARL of CUSUM chart with n=3) / (ARL of CUSUM chart with n=5)

* ψ_3 = (ARL of neural network with n=3) / (ARL of neural network with n=5)

Table 47. Effects of sample size, Case 7 (skewness = 2.0, kurtosis = 6.2)

Mean shifts	$\overline{x}$ chart n = 3	$\overline{x}$ chart n = 5	ψ_1	CUSUM n = 3	CUSUM n = 5	ψ_2	Neural n = 3	Neural n = 5	ψ_3
0	93.08	125.58	0.74	93.45	126.17	0.74	93.51	125.78	0.74
0.25	62.40	81.04	0.77	46.29	57.08	0.81	63.97	81.07	0.79
0.5	43.69	52.76	0.83	23.00	26.04	0.88	44.27	53.30	0.83
0.75	30.63	34.83	0.88	12.76	13.76	0.93	30.95	35.22	0.88
1	22.06	23.66	0.93	8.19	8.54	0.96	21.83	23.68	0.92
1.5	11.23	11.77	0.95	4.60	4.76	0.97	11.37	11.91	0.95
2	5.87	6.28	0.93	3.13	3.29	0.95	6.42	6.36	1.01
3	2.52	2.27	1.11	2.02	2.12	0.95	2.65	2.29	1.16
4	1.16	1.22	0.95	1.63	1.67	0.98	1.23	1.23	1.00
5	1	1	1.00	1.19	1.32	0.90	1	1	1.00
-0.25	143.30	195.17	0.73	58.55	68.66	0.85	3.41	6.62	0.52
-0.5	232.28	319.48	0.73	20.04	22.75	0.88	2.43	3.84	0.63
-0.75	389.64	523.42	0.74	10.50	11.62	0.90	1.97	2.78	0.71
-1	642.03	888.38	0.72	6.86	7.47	0.92	1.71	2.24	0.76
-1.5	1678.85	2449.67	0.69	4.10	4.4	0.93	1.41	1.61	0.88
-2	7.22	5.48	1.32	2.98	3.13	0.95	1.23	1.3	0.95
-3	1.65	1.78	0.93	2.17	2.22	0.98	1.05	1.07	0.98
-4	1.20	1.18	1.02	1.44	1.57	0.92	1.01	1.02	0.99
-5	1.05	1.04	1.01	1.18	1.22	0.97	1	1	1.00

* ψ_1 = (ARL of $\overline{x}$ chart with n=3) / (ARL of $\overline{x}$ chart with n=5)

* ψ_2 = (ARL of CUSUM chart with n=3) / (ARL of CUSUM chart with n=5)

* ψ_3 = (ARL of neural network with n=3) / (ARL of neural network with n=5)

Table 48-a & b. Analysis of variance
Case 2 (skewness = 0, kurtosis = 3, n = 3)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	280.06	121.25	277.96
0.5	153.34	34.88	153.89
0.75	81.14	16.17	81.80
1	44.65	9.9	45.10
1.5	15.16	5.5	15.33
2	6.37	3.86	6.47
3	1.99	2.48	2.01
4	1.18	1.95	1.19
5	1.02	1.6	1.02
-0.25	286.93	119.89	295.73
-0.5	156.32	34.82	162.98
-0.75	81.96	16	84.8
-1	44.38	9.91	45.77
-1.5	14.94	5.51	15.37
-2	6.33	3.83	6.49
-3	2	2.47	2.04
-4	1.19	1.96	1.2
-5	1.02	1.61	1.03

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	23507	11753	9.95	.0004
block	17	286234	16837	14.25	.0001
error	34	40179	1181		
total	53	349921			

Table 49-a & b. Analysis of variance
Case 4 (skewness = 1, kurtosis = 3, n = 3)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	124.29	78.03	124.02
0.5	73.55	29.87	73.41
0.75	44.70	14.80	44.70
1	27.80	9.24	27.69
1.5	12.04	5.15	12.23
2	6.13	3.59	6.24
3	2.17	2.31	2.26
4	1.22	1.81	1.24
5	1	1.46	1
-0.25	412.16	102.56	17.17
-0.5	812.31	29.29	7.76
-0.75	1691.28	13.84	4.85
-1	31.38	4.97	2.11
-1.5	31.38	4.97	2.11
-2	5.57	3.50	1.52
-3	1.86	2.33	1.11
-4	1.19	1.79	1.02
-5	1.04	1.38	1

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	324494	162247	2.66	.0844
block	17	1076760	63338	1.04	.4454
error	34	2072763	60963		
total	53	3474018			

Table 50-a & b. Analysis of variance
Case 7 (skewness = 2, kurtosis = 6.2, n= 3)

(a)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	62.40	46.29	63.97
0.5	43.69	23.00	44.27
0.75	30.63	12.76	30.95
1	22.06	8.19	21.83
1.5	11.23	4.60	11.37
2	5.87	3.13	6.42
3	2.52	2.02	2.65
4	1.16	1.63	1.23
5	1	1.19	1
-0.25	143.30	58.55	3.41
-0.5	232.28	20.04	2.43
-0.75	389.64	10.50	1.97
-1	642.03	6.86	1.71
-1.5	1678.85	4.10	1.41
-2	7.22	2.98	1.23
-3	1.65	2.17	1.05
-4	1.20	1.44	1.01
-5	1.05	1.18	1

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	349759	174879	3.07	.0594
block	17	941771	55398	.97	.5070
error	34	1935846	56936		
total	53	3227377			

Table 51. Performance ranking from multiple comparisons for nonnormality (n =3) (Tukey tests)

Case	$\bar{x}$ chart	CUSUM chart	Neural network
2	2	1	2
4	1	1 (1)	1 (1)
7	1	1 (1)	1 (1)

* The numbers in parentheses represent the order of means that are not statistically significant.

CHAPTER VI

CONCLUSION AND DISCUSSION

Chapter 5 was devoted to a discussion of the results of three simulation experiments. In this chapter, the results from the experiments are summarized and the six research questions defined in Chapter 3 are addressed. This is followed by discussions of the contributions of the study. Finally, this chapter discusses the study limitations and future research, detailing where attention must be directed in order to improve the performance of the neural networks and compare them with other possible quality control methods.

Conclusion

The first experiment in this study compares the performance of $\bar{x}$ and CUSUM control charts, and neural networks for the normal distribution and eight nonnormal distributions with different degrees of skewness and kurtosis. This experiment also examines the effects of nonnormality on the ability of the three methods to detect out-of-control states. The second experiment investigates the effects of mean shifts on the performance of the three methods for some selected cases. The third experiment examines the effects of sample size on the performance of the three methods. In this section, the results from the experiments are summarized to address the six research questions of the study.

The first and second research questions are about the differences in the performance of the three methods when the underlying distribution is normal and nonnormal, respectively. Overall, for both the normal and symmetric nonnormal distributions, the CUSUM charts perform better than the corresponding $\overline{x}$ charts and neural networks for small mean shifts. For large mean shifts, on the other hand, the CUSUM charts perform worse than the corresponding $\overline{x}$ charts and neural networks. This result indicates that the performance of the $\overline{x}$ and CUSUM charts is the same for symmetric nonnormal distributions as for the normal distribution. While it is well known that a CUSUM chart performs better than a $\overline{x}$ chart for small mean shifts and performs worse than a $\overline{x}$ chart for large mean shifts when the distribution is normal, this study has shown that this remains true for nonnormal distributions. In addition, the results of this study show that neural networks perform as well as the corresponding $\overline{x}$ charts for both normal and symmetric nonnormal distributions.

For skewed and/or kurtotic distributions, neural networks generally perform better than the corresponding $\overline{x}$ and CUSUM charts. While for some cases with the skewed and/or kurtotic distributions CUSUM charts seem to outperform neural networks, neural networks appear advantageous even for small mean shifts as the degree of skewness and kurtosis increases. Just as for the normal distribution, CUSUM charts perform better than $\overline{x}$ charts for small mean shifts and worse than $\overline{x}$ charts for large mean shifts when the underlying distributions are skewed and/or kurtotic. These results also addressed the fourth research question concerning the effects of mean shifts on the performance of the three methods.

The third research question is about the effects of degrees of nonnormality of a population, as measured by skewness and kurtosis, on the performance of the three methods. Clearly, the effects of skewness and kurtosis on the performance of the methods were substantial because nonnormality dramatically affected the in-control ARLs as well as the out-of-control ARLs for the methods. For example, consistent with our expectation, in-control ARLs become smaller as the degree of nonnormality, as measured by skewness and kurtosis, become higher.

The fifth and sixth research questions are about the effect of sample size on performance of the three methods when the underlying distribution is normal and nonnormal, respectively. It can be concluded that there is no significant performance effect due to sample size when the distribution is normal. However, consistent with the Central Limit Theorem, each of the three methods when using a sample size of three performs worse than the corresponding method with a sample size of five for skewed and/or kurtotic distributions.

Discussion

Contribution

In the last decade, neural networks have received a great deal of attention and are being explored extensively to examine their appropriateness for statistical applications. Such applications include discriminant analysis, regression analysis, time series forecasting, production operations management, and quality control charts. Numerous authors (Archer and Wang 1993; Hill et al. 1994; Prybutok and Nam 1995) suggest that, under some conditions, neural networks offer an advantageous alternative to traditional statistical methods. Among these advantages is the employment of a pattern recognition approach rather than a complex statistical approach. The pattern recognition approach of neural networks can be used as an alternative to conventional control charts for detecting out-of-control conditions.

In this study, neural network alternatives to standard control charts are developed under conditions of nonnormality. The performance of these neural networks is compared with those of corresponding $\overline{x}$ and CUSUM control charts. The development of a new and better technique for the analysis of quality characteristics will allow companies to enhance product quality and reduce internal and external failure costs. This study shows that neural networks have potential for improving quality control monitoring systems by effectively detecting defects when the underlying distribution is not normally distributed. The results provide guidance concerning the practical use of neural networks as alternatives or as supplements to standard control charts. The major contributions of this study are summarized as follow:

1. Neural networks are developed as alternatives to standard control charts under conditions of nonnormality.
2. This study pursued the most extensive range of nonnormality of any current research investigation into the effects of nonnormality on control chart performance. The nonnormality encompasses nine different combinations of skewness and kurtosis from the generalized Burr distribution that vary from one similar to a uniform distribution to one that resembles an exponential distribution.
3. This is the first study to compare the performance of neural network models with both $\overline{x}$ and CUSUM control charts.
4. This study also shows the effects of nonnormality, as measured by different skewness and kurtosis, on the ability of the three methods to detect out-of-control states.
5. This study investigates the effects of mean shifts and sample size on the performance of the three methods.

6. This study utilizes a unique methodology for comparing the relative performance of the three methods. A randomized block design is used to compare the performance of the three methods. By choosing mean shifts to represent the blocks, the variability among the mean shifts is effectively filtered out, allowing more precise comparisons among the three methods.
7. This study is the first to utilize the method of determining cutoff values for neural networks by performing simulations, so that the in-control ARLs of the neural networks match those of the $\bar{x}$ and CUSUM charts.
8. This study is the first to adjust the cutoff value, h, for CUSUM charts for each of the different cases. Developing this method was necessary to obtain fair comparisons among the $\bar{x}$ and CUSUM charts, and neural networks.

Limitations and Future Research

The limitations and future research of this study are listed below.

1. The parameters of the neural networks, such as number of hidden layers, number of hidden neurons, momentum, and learning rate, are determined through experimentation. The chosen parameters do not guarantee optimal performance of the neural networks.
2. Backpropagation used to train the neural networks in this study has the potential to converge to a local minimum because it employs a gradient decent algorithm. Various neural network paradigms have been developed such as genetic algorithms, probability neural networks, and adaptive resonance theory. Different paradigms are effective for solving different types of problems. When a user selects a neural network to solve a given problem, the user needs to decide which is the correct paradigm. Thus, further research

is needed to compare backpropagation with other paradigms to find the strengths, limitations, and applications for each paradigm.

3. This study is exploratory, not only because it is the first study that develops neural networks for detecting out-of-control states for nonnormal distributions, but also because it covers only a small portion of all possible nonnormal distributions. In this study, nine different nonnormal distributions are examined to compare neural networks with two conventional quality control charts, $\overline{x}$ and CUSUM. There exist an infinite number of possible nonnormal distributions, including a variety of bimodal distributions and many unimodal distributions with combinations of skewness and kurtosis different from those used in this study. Therefore, other types of nonnormal distributions should be examined in the future to better evaluate the performance of the three methods.
4. The neural networks developed in this study are designed to detect a sudden shift in the process mean. There are many other patterns that may exist in the process indicating out of control situations such as trends, cycles, systematic and stratification patterns, and mixtures. Each one of these patterns should be investigated in future research. However, the characteristics of these out-of-control patterns should be considered and might need to be redefined under nonnormality.
5. The neural networks developed in this study are compared with the two conventional symmetric quality control charts. Because nonnormality has a significant effect on the performance of control charts, design considerations for these charts must include recognition of the degree of nonnormality of the underlying distribution. Some previous research has developed asymmetric $\overline{x}$ control charts and asymmetric CUSUM charts to

deal with the nonnormal data. The neural networks should be compared with these asymmetric control charts.

6. The neural networks can be trained with additional information such as a particular range of previous data points, the ranges of each subgroup, the mean of each subgroup, and standardized values of the means, as discussed earlier. Similar to $\bar{x}$ control charts, a major disadvantage of the neural networks developed in this study is that they utilize only the information about the process contained in the last plotted point and ignore any information provided by the entire sequence of points. This feature might make the neural network relatively insensitive to small mean shifts. This might be remedied by including a certain range of previous points. Additional information might also be provided by including the ranges of each subgroup, the means of the subgroups, and/or standardized values of the means in addition to the last input point. Although the use of these supplementary inputs reduces the simplicity and ease of interpretation of the neural networks, it might substantially improve their performance.

7. In this study, one neural network is used to detect both positive and negative mean shifts. Instead of incorporating one neural network for both mean shifts, two parallel dedicated neural networks could be utilized in future research, one for detecting positive mean shifts and the other for detecting negative mean shifts. Training a neural network to distinguish two groups is easier and generally yields better results than training it to distinguish three groups as done in this study.

8. The simulation study performed here is restricted to only two sample sizes (n = 3 and 5). Further research with different numbers of sample sizes is needed to examine the effects

of sample size on the performance of neural networks, $\overline{x}$ charts, and CUSUM charts.

9. In this study, the performance of the neural networks was evaluated by estimating the ARLs. The neural network approach presented in this study offers a competitive alternative to existing control schemes. The performance of the neural network depends heavily on the selection of training samples. Hence, the results reported in this study do not offer a universally optimal solution. Rather, this study demonstrates the feasibility of applying neural network models to process monitoring.

APPENDIX A

TABLES

Table 26-a & b. Analysis of variance
(mean shift = .25)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	479.05	185.67	514.86
normal (0 / 3)	280.16	122.54	272.85
symmetric/peaked (0 / 4)	205.27	101.01	215.31
skewed right (1 / 3)	142.29	90.04	142.80
skewed right / peaked (1 / 4)	131.65	86.3	132.36
skewed right / peaked (1 / 5)	121.69	76.87	123.69
severely skewed right / highly picked (2 / 6.2)	81.04	57.08	81.07
severely skewed right / highly picked (2 / 7.2)	77.6	54.79	78.96
severely skewed right / highly picked (2 / 8.8)	77.69	55.15	77.55

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	46156	23078	7.59	.0048
block	8	262587	32823	10.8	.0001
error	16	48646	3040		
total	26	357390			

Table 27-a & b. Analysis of variance
(mean shift = .50)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	202.53	43.93	218.27
normal (0 / 3)	157.1	35.28	151.09
symmetric/peaked (0 / 4)	131.06	32.17	137.85
skewed right (1 / 3)	81.2	31.36	81.71
skewed right / peaked (1 / 4)	78.49	30.56	79.08
skewed right / peaked (1 / 5)	76.02	29.18	77.54
severely skewed right / highly picked (2 / 6.2)	52.76	26.04	53.03
severely skewed right / highly picked (2 / 7.2)	52.09	24.96	52.72
severely skewed right / highly picked (2 / 8.8)	52.81	25.16	52.43

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	28073	14036	17.04	.0001
block	8	35946	4493	5.46	.002
error	16	13178	823		
total	26	77198			

Table 28-a & b. Analysis of variance
(mean shift = .75)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	93.04	18.73	98.71
normal (0 / 3)	83.27	16.27	79.48
symmetric/peaked (0 / 4)	75.6	15.12	80.22
skewed right (1 / 3)	48.7	15.4	49.01
skewed right / peaked (1 / 4)	48.59	15.12	48.81
skewed right / peaked (1 / 5)	48.06	14.37	48.73
severely skewed right / highly picked (2 / 6.2)	34.83	13.76	35.22
severely skewed right / highly picked (2 / 7.2)	34.77	13.35	35.26
severely skewed right / highly picked (2 / 8.8)	36.03	13.36	35.74

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	10229	5114	33.2	.0001
block	8	5925	740	4.84	.0036
error	16	2448	153		
total	26	18603			

Table 29-a & b. Analysis of variance
(mean shift = 1.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	46.11	11.26	48.56
normal (0 / 3)	44.90	9.93	43.48
symmetric/peaked (0 / 4)	43.15	9.33	45.98
skewed right (1 / 3)	30.23	9.49	30.79
skewed right / peaked (1 / 4)	31.05	9.40	31.26
skewed right / peaked (1 / 5)	30.60	8.97	31.45
severely skewed right / highly picked (2 / 6.2)	23.66	8.54	23.68
severely skewed right / highly picked (2 / 7.2)	24.11	8.38	24.19
severely skewed right / highly picked (2 / 8.8)	24.84	8.33	24.72

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	3514	1757	69.27	.0001
block	8	1026	128	5.06	.0029
error	16	405	25		
total	26	4946			

Table 30-a & b. Analysis of variance
(mean shift = 1.5)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	14.72	6.23	15.36
normal (0 / 3)	15.01	5.57	14.73
symmetric/peaked (0 / 4)	15.27	5.26	16.24
skewed right (1 / 3)	12.74	5.28	12.79
skewed right / peaked (1 / 4)	13.17	5.2	13.32
skewed right / peaked (1 / 5)	13.23	5.02	13.66
severely skewed right / highly picked (2 / 6.2)	11.77	4.76	11.91
severely skewed right / highly picked (2 / 7.2)	11.95	4.66	12.10
severely skewed right / highly picked (2 / 8.8)	12.28	4.64	12.32

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	413.24	206.62	475.91	.0001
block	8	28.25	3.53	8.13	.0002
error	16	6.95	.43		
total	26	448.45			

Table 31-a & b. Analysis of variance
(mean shift = 2.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	6.15	4.34	6.37
normal (0 / 3)	6.28	3.86	6.17
symmetric/peaked (0 / 4)	6.41	3.65	6.77
skewed right (1 / 3)	6.13	3.68	6.22
skewed right / peaked (1 / 4)	6.22	3.62	6.32
skewed right / peaked (1 / 5)	6.41	3.5	6.57
severely skewed right / highly picked (2 / 6.2)	6.28	3.29	6.36
severely skewed right / highly picked (2 / 7.2)	6.24	3.25	6.32
severely skewed right / highly picked (2 / 8.8)	6.36	3.24	6.36

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	44.66	22.33	382.93	.0001
block	8	.40	.050	.86	.5694
error	16	.93	.058		
total	26	45.99			

Table 32. Performance ranking from multiple comparisons for mean shift (n = 5) (Tukey tests)

Mean shifts	$\bar{x}$ chart	CUSUM chart	Neural network
0.25	2	1	2
0.5	2	1	2
0.75	2	1	2
1	2	1	2
1.5	2	1	2
2	2	1	2
3	1	1	1
4	1	2	1
5	1	2	1
-0.25	2	1(1)	1(2)
-0.5	2	1(1)	1(2)
-0.75	2	1(1)	1(2)
-1	2	1(1)	1(2)
-1.5	1(3)	1(1)	1(2)
-2	2	1(1)	1(2)
-3	2	3	1
-4	1	2	1(1)
-5	1	2	1(1)

* The numbers in parentheses represent the order of means that are not statistically significant.

Table 33-a & b. Analysis of variance
(mean shift = -.25)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	480.80	183.75	451.93
normal (0 / 3)	282.48	122.7	292.57
symmetric/peaked (0 / 4)	208.83	101.36	202.11
skewed right (1 / 3)	477.9	109.72	96.25
skewed right / peaked (1 / 4)	389.85	104.36	208.32
skewed right / peaked (1 / 5)	297.93	90.83	267.52
severely skewed right / highly picked (2 / 6.2)	195.17	68.66	6.62
severely skewed right / highly picked (2 / 7.2)	182.48	70.4	10.29
severely skewed right / highly picked (2 / 8.8)	175.65	69.13	57.31

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	177406	88703	13.59	.0004
block	8	203772	25471	3.9	.0099
error	16	104458	6528		
total	26	485637			

Table 34-a & b. Analysis of variance
(mean shift = -.5)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	209.44	44.15	197.37
normal (0 / 3)	153.81	35.53	163.16
symmetric/peaked (0 / 4)	127.6	31.5	121.93
skewed right (1 / 3)	936.22	31.13	29.52
skewed right / peaked (1 / 4)	664.78	30.37	79.52
skewed right / peaked (1 / 5)	336.35	28.46	223.12
severely skewed right / highly picked (2 / 6.2)	319.48	22.75	3.84
severely skewed right / highly picked (2 / 7.2)	285.78	23.51	5.17
severely skewed right / highly picked (2 / 8.8)	265.88	23.51	12.72

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	576201	288100	10.14	.0014
block	8	166158	20769	.73	.6634
error	16	454408	28400		
total	26	1196768			

Table 35-a & b. Analysis of variance
(mean shift = -.75)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	94.58	18.74	89.38
normal (0 / 3)	81.06	16.11	85.70
symmetric/peaked (0 / 4)	73.82	14.96	70.25
skewed right (1 / 3)	1876.82	14.57	13.70
skewed right / peaked (1 / 4)	541.05	14.25	31.48
skewed right / peaked (1 / 5)	203.60	13.71	112.00
severely skewed right / highly picked (2 / 6.2)	523.42	11.62	2.78
severely skewed right / highly picked (2 / 7.2)	449.97	11.84	3.37
severely skewed right / highly picked (2 / 8.8)	400.02	11.85	6.31

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	1173843	586921	5.35	.0166
block	8	772116	96514	.88	.5534
error	16	1755304	109706		
total	26	3701264			

Table 36-a & b. Analysis of variance
(mean shift = -1.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	46.59	11.33	44.35
normal (0 / 3)	43.94	9.93	45.87
symmetric/peaked (0 / 4)	41.75	9.36	40.33
skewed right (1 / 3)	392.54	9.13	7.73
skewed right / peaked (1 / 4)	131.67	9.05	15.36
skewed right / peaked (1 / 5)	88.08	8.60	49.66
severely skewed right / highly picked (2 / 6.2)	888.39	7.47	2.24
severely skewed right / highly picked (2 / 7.2)	713.79	7.57	2.53
severely skewed right / highly picked (2 / 8.8)	597.13	7.59	3.90

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	580824	290412	7.48	.0051
block	8	259908	32488	.84	.5841
error	16	620974	38810		
total	26	1461707			

Table 37-a & b. Analysis of variance
(mean shift = -1.5)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	14.53	6.22	14.06
normal (0 / 3)	14.99	5.53	15.68
symmetric/peaked (0 / 4)	15.35	5.22	14.81
skewed right (1 / 3)	19.17	5.12	3.46
skewed right / peaked (1 / 4)	19.10	5.04	5.38
skewed right / peaked (1 / 5)	19.19	4.90	12.77
severely skewed right / highly picked (2 / 6.2)	2449.67	4.40	1.61
severely skewed right / highly picked (2 / 7.2)	1600.00	4.42	1.75
severely skewed right / highly picked (2 / 8.8)	74.26	4.43	2.20

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	1286722	643361	2.34	.1289
block	8	2175824	271978	.99	.4807
error	16	4408067	275504		
total	26	7870614			

Table 38-a & b. Analysis of variance
(mean shift = -2.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	6.06	4.32	5.86
normal (0 / 3)	6.31	3.88	6.51
symmetric/peaked (0 / 4)	6.41	3.67	6.30
skewed right (1 / 3)	5.99	3.60	2.07
skewed right / peaked (1 / 4)	6.13	3.56	2.73
skewed right / peaked (1 / 5)	6.46	3.46	4.85
severely skewed right / highly picked (2 / 6.2)	5.48	3.13	1.30
severely skewed right / highly picked (2 / 7.2)	5.91	3.13	1.35
severely skewed right / highly picked (2 / 8.8)	6.16	3.13	1.54

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	38.24214	19.12107	14.14	.0003
block	8	20.45274	2.55659	1.89	.1324
error	16	21.63206	1.35200		
total	26	80.32694			

Table 39-a & b. Analysis of variance
(mean shift = 3.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	2.02	2.76	2.06
normal (0 / 3)	1.98	2.48	1.97
symmetric/peaked (0 / 4)	1.98	2.37	2.04
skewed right (1 / 3)	2.15	2.37	2.16
skewed right / peaked (1 / 4)	2.15	2.34	2.16
skewed right / peaked (1 / 5)	2.11	2.26	2.15
severely skewed right / highly picked (2 / 6.2)	2.27	2.12	2.29
severely skewed right / highly picked (2 / 7.2)	2.27	2.10	2.30
severely skewed right / highly picked (2 / 8.8)	2.27	2.1	2.27

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	.19185	.09593	2.85	.0873
block	8	.05256	.00657	.2	.9876
error	16	.53841	.03365		
total	26	.78283			

Table 40-a & b. Analysis of variance
(mean shift =4.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	1.20	2.12	1.21
normal (0 / 3)	1.19	1.96	1.19
symmetric/peaked (0 / 4)	1.18	1.87	1.19
skewed right (1 / 3)	1.20	1.85	1.21
skewed right / peaked (1 / 4)	1.19	1.84	1.20
skewed right / peaked (1 / 5)	1.18	1.78	1.19
severely skewed right / highly picked (2 / 6.2)	1.22	1.67	1.23
severely skewed right / highly picked (2 / 7.2)	1.20	1.66	1.21
severely skewed right / highly picked (2 / 8.8)	1.20	1.66	1.20

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	2.33570	1.16785	140.81	.0001
block	8	.05852	.00731	.88	.5519
error	16	.13270	.00829		
total	26	2.52692			

Table 41-a & b. Analysis of variance
(mean shift = 5.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	1.02	1.83	1.02
normal (0 / 3)	1.02	1.6	1.02
symmetric/peaked (0 / 4)	1.02	1.48	1.03
skewed right (1 / 3)	1.00	1.51	1.00
skewed right / peaked (1 / 4)	1.01	1.48	1.01
skewed right / peaked (1 / 5)	1.01	1.40	1.01
severely skewed right / highly picked (2 / 6.2)	1.00	1.32	1.00
severely skewed right / highly picked (2 / 7.2)	1.00	1.28	1.00
severely skewed right / highly picked (2 / 8.8)	1.00	1.27	1.00

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	1.23610	.61805	63.87	.0001
block	8	.09825	.01228	1.27	.3248
error	16	.15484	.00968		
total	26	1.48919			

Table 42-a & b. Analysis of variance
(mean shift = -3.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	1.98	2.75	1.97
normal (0 / 3)	2.02	2.49	2.05
symmetric/peaked (0 / 4)	2.02	2.38	2.00
skewed right (1 / 3)	1.88	2.38	1.25
skewed right / peaked (1 / 4)	1.88	2.34	1.37
skewed right / peaked (1 / 5)	1.90	2.28	1.69
severely skewed right / highly picked (2 / 6.2)	1.78	2.22	1.07
severely skewed right / highly picked (2 / 7.2)	1.77	2.14	1.08
severely skewed right / highly picked (2 / 8.8)	1.77	2.13	1.13

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	3.13460	1.56730	45.42	.0001
block	8	1.21233	.15154	4.39	.0057
error	16	.55213	.03451		
total	26	4.89907			

Table 43-a & b. Analysis of variance
(mean shift = -4.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	1.19	2.11	1.19
normal (0 / 3)	1.18	1.96	1.19
symmetric/peaked (0 / 4)	1.18	1.89	1.17
skewed right (1 / 3)	1.21	1.89	1.05
skewed right / peaked (1 / 4)	1.18	1.84	1.07
skewed right / peaked (1 / 5)	1.18	1.78	1.14
severely skewed right / highly picked (2 / 6.2)	1.18	1.57	1.02
severely skewed right / highly picked (2 / 7.2)	1.19	1.57	1.02
severely skewed right / highly picked (2 / 8.8)	1.18	1.57	1.03

(b)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	2.61756	1.30878	127.76	.0001
block	8	.17923	.02240	2.19	.087
error	16	.16390	.01024		
total	26	2.96070			

Table 44-a & b. Analysis of variance
(mean shift = -5.0)

(a)

Nonnormality (skewness / kurtosis)	$\bar{x}$ chart	CUSUM chart	Neural network
symmetric/flatter (0 / 2)	1.02	1.83	1.02
normal (0 / 3)	1.02	1.61	1.02
symmetric/peaked (0 / 4)	1.02	1.50	1.02
skewed right (1 / 3)	1.03	1.45	1.01
skewed right / peaked (1 / 4)	1.03	1.42	1.01
skewed right / peaked (1 / 5)	1.03	1.36	1.02
severely skewed right / highly picked (2 / 6.2)	1.04	1.22	1.00
severely skewed right / highly picked (2 / 7.2)	1.04	1.21	1.00
severely skewed right / highly picked (2 / 8.8)	1.04	1.21	1.01

(b)

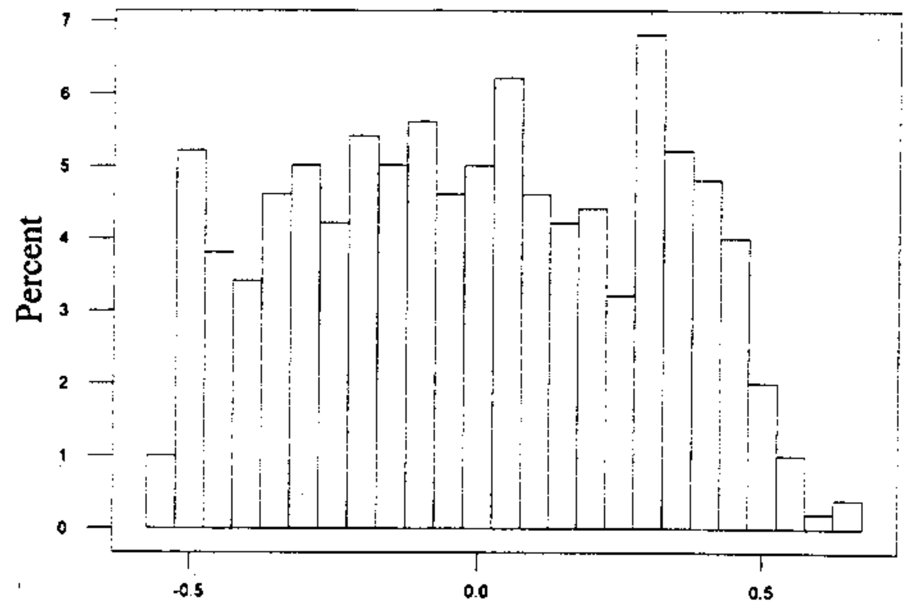
Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F value	P-value
method	2	.97212	.48606	33.56	.0001
block	8	.11261	.01408	.97	.4907
error	16	.23175	.01448		
total	26	1.31647			

APPENDIX B

FIGURES

figure 6. The shape of distribution

case 1 (individual observations)



case 1 (averages, n=5)

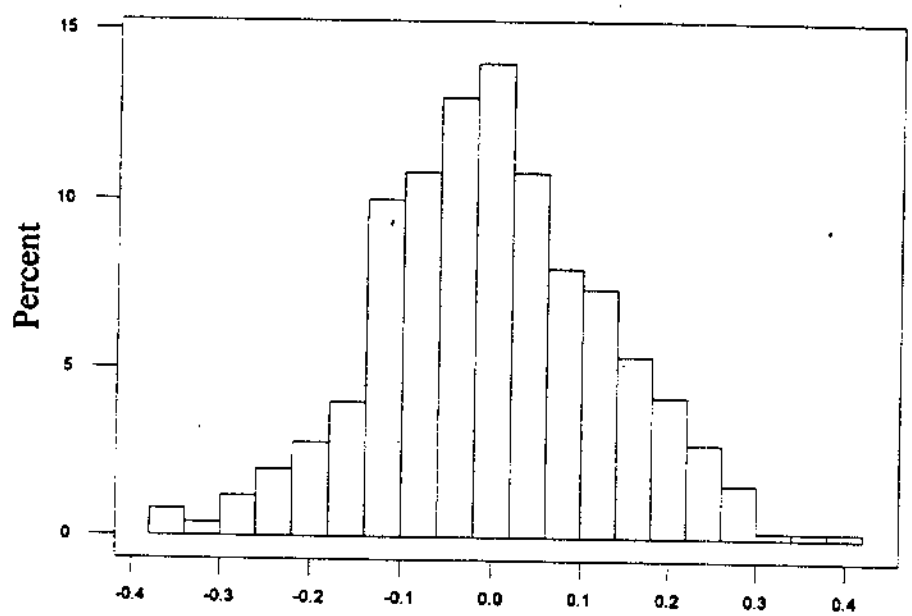


figure 7. Comparison of ARLs of the three methods

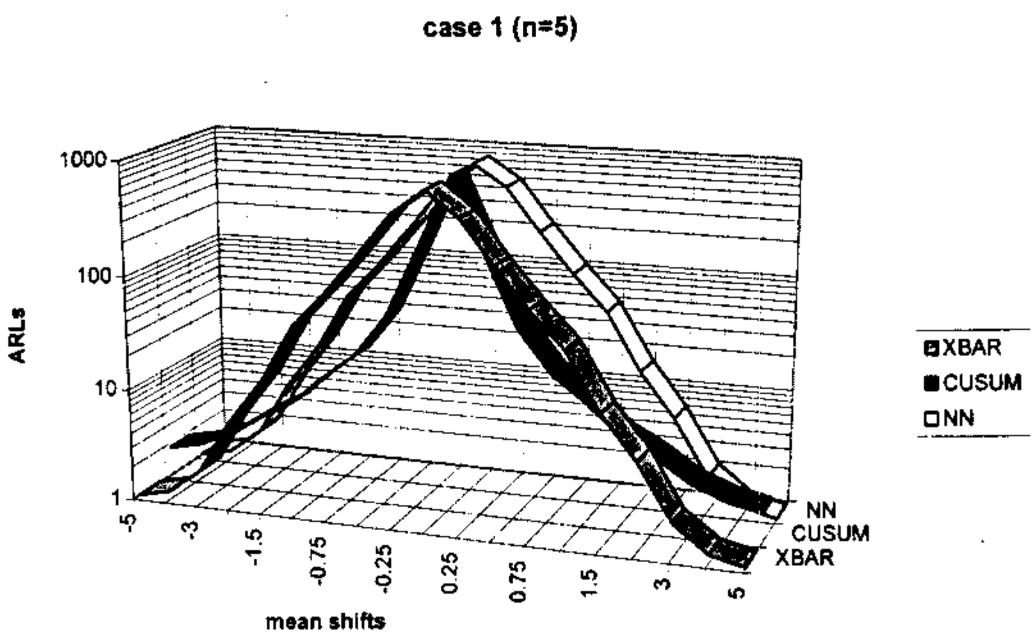
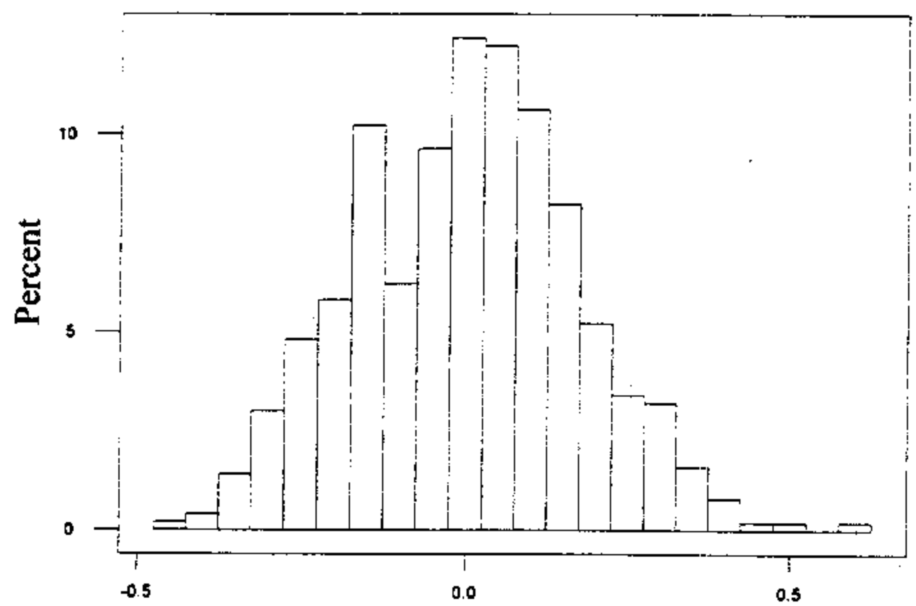


figure 8. The shape of distribution

case 2 (individual observations)



case 2 (averages, n=5)

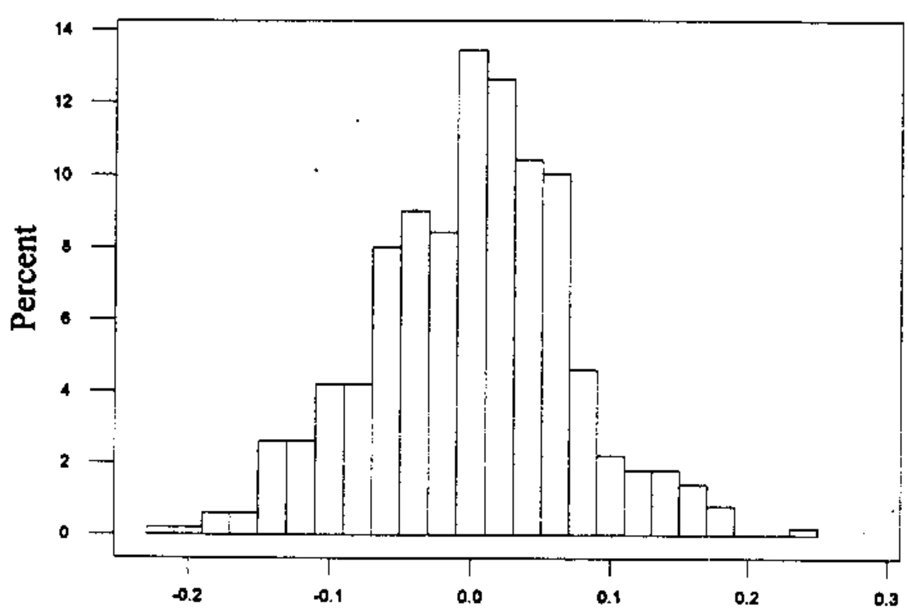


figure 9. Comparison of ARLs of the three methods

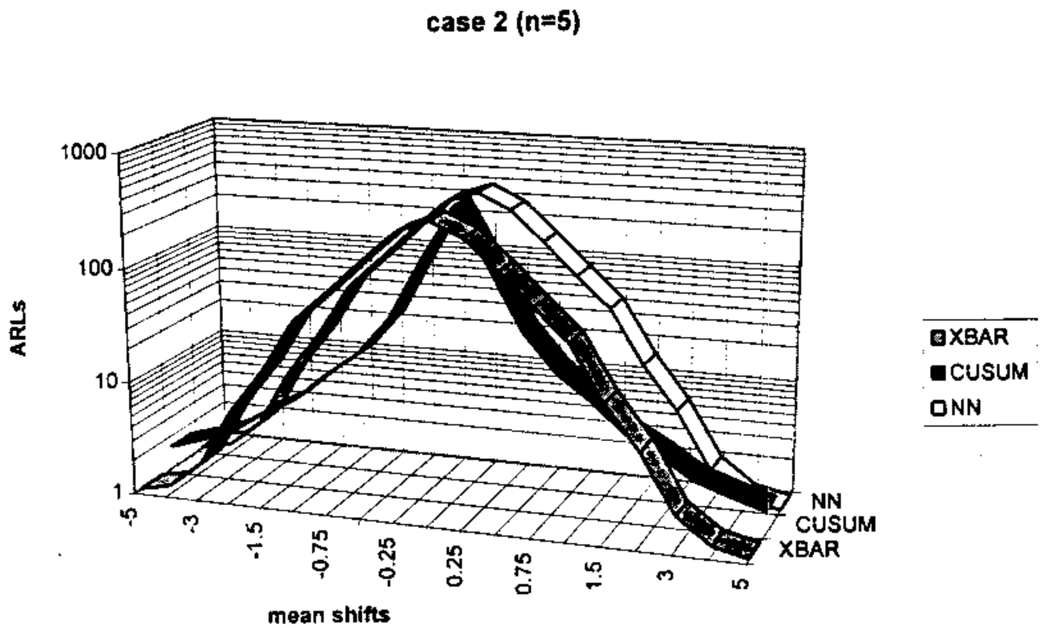
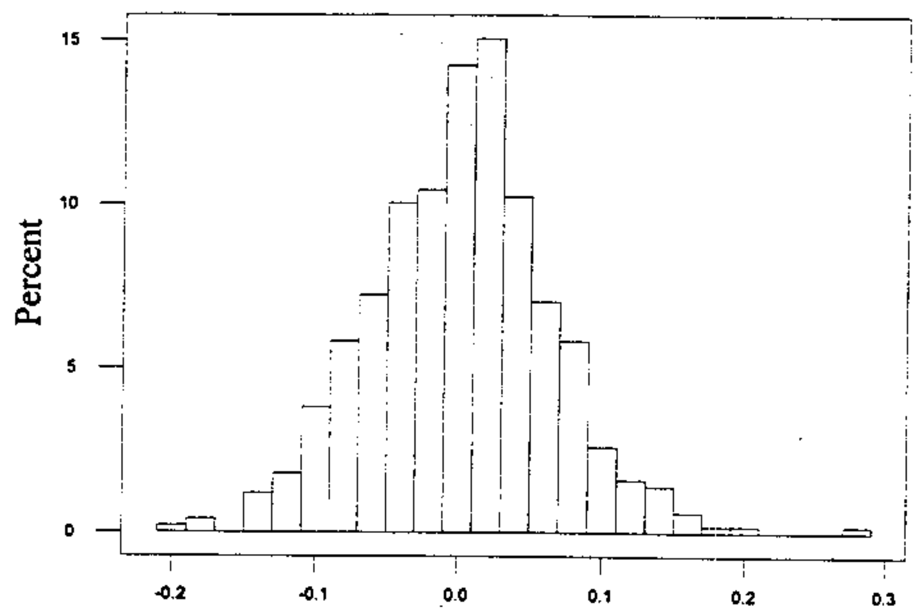


figure 10. The shape of distribution

case 3 (individual observations)



case 3 (averages, n=5)

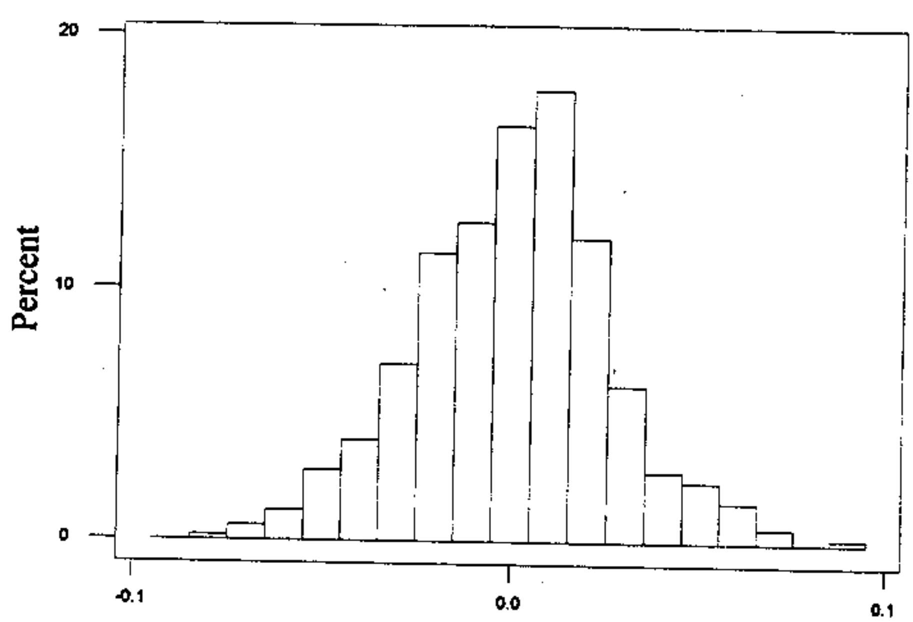


figure 11. Comparison of ARLs of the three methods

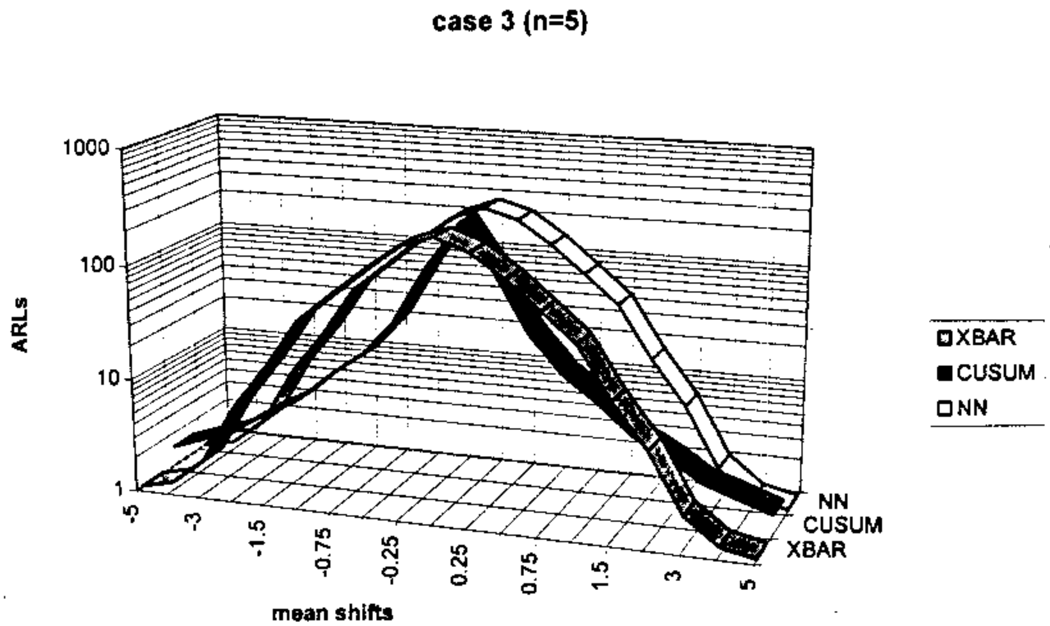
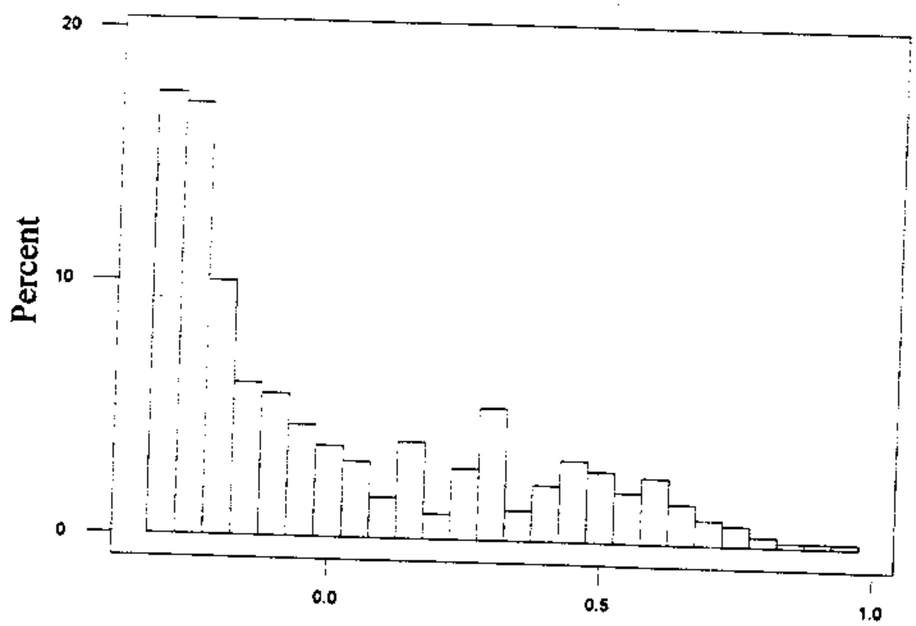


figure 12. The shape of distribution

case 4 (individual observations)



case 4 (averages, n=5)

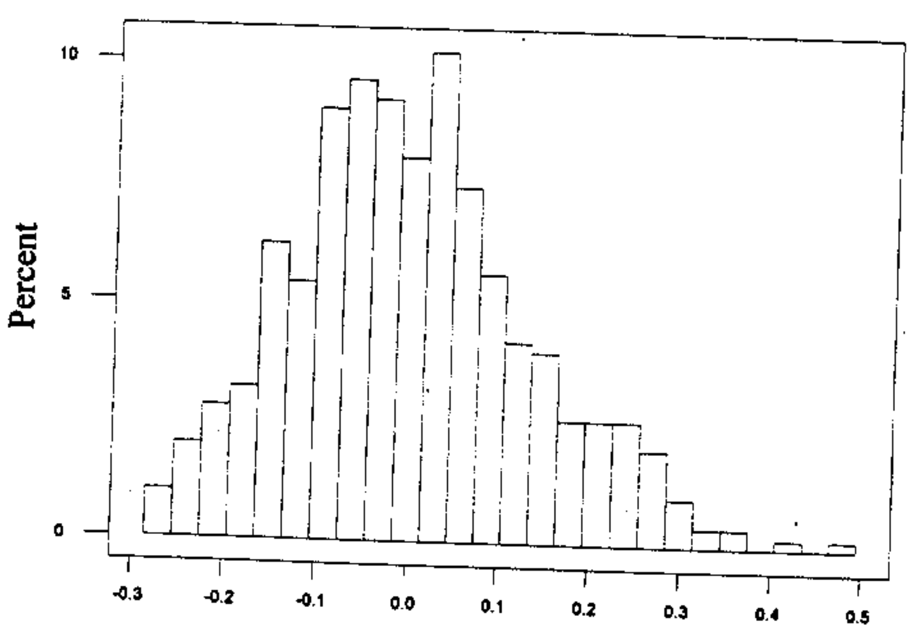


figure 13. Comparison of ARLs of the three methods

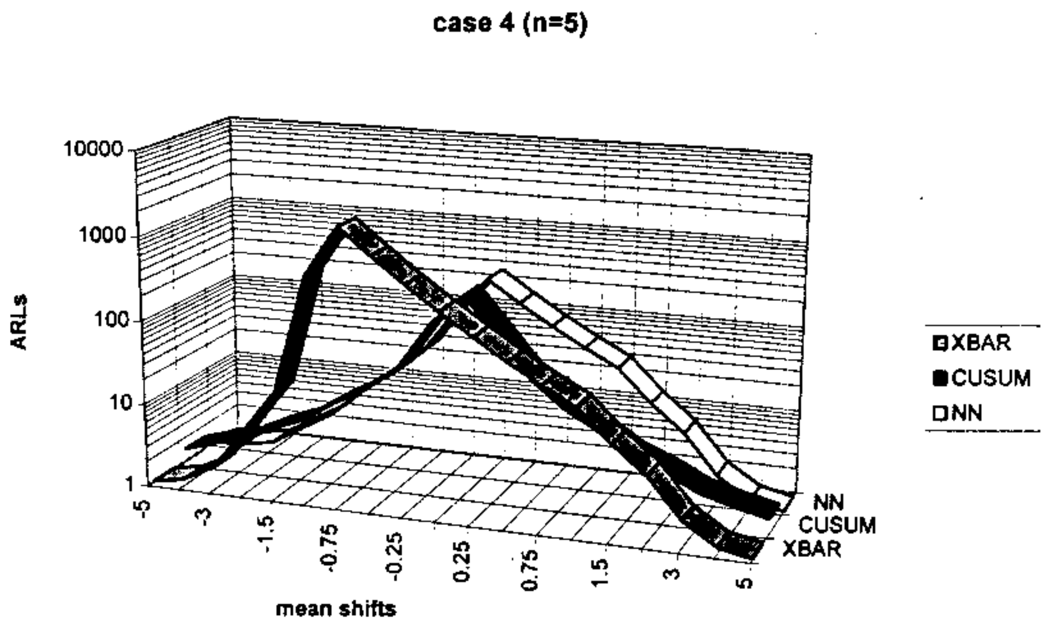
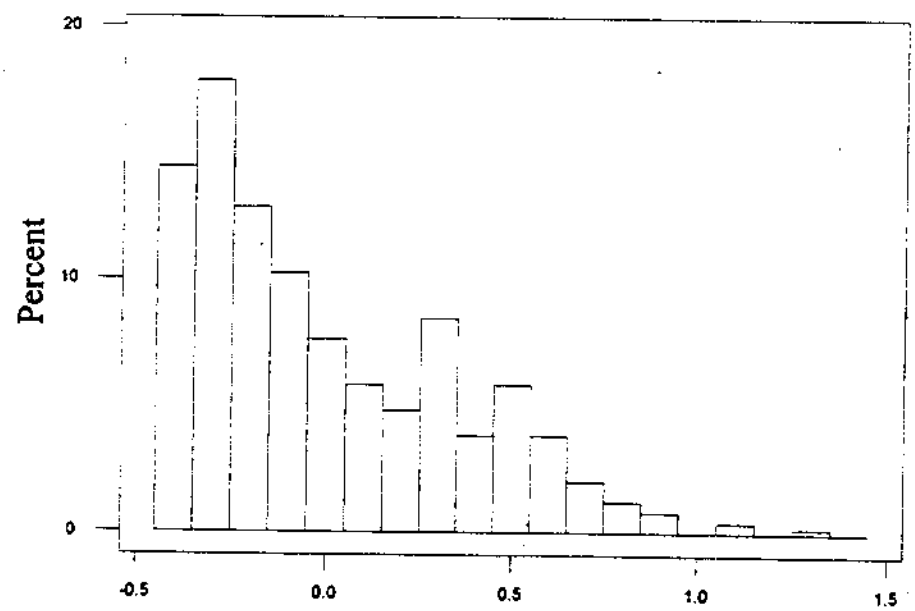


figure 14. The shape of distribution

case 5 (individual observations)



case 5 (averages, n=5)

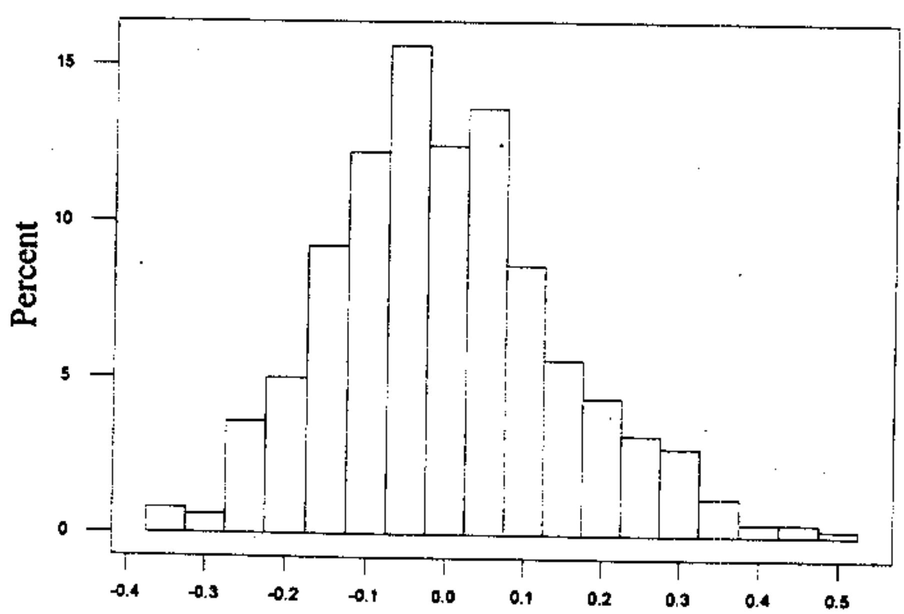


figure 15. Comparison of ARLs of the three methods

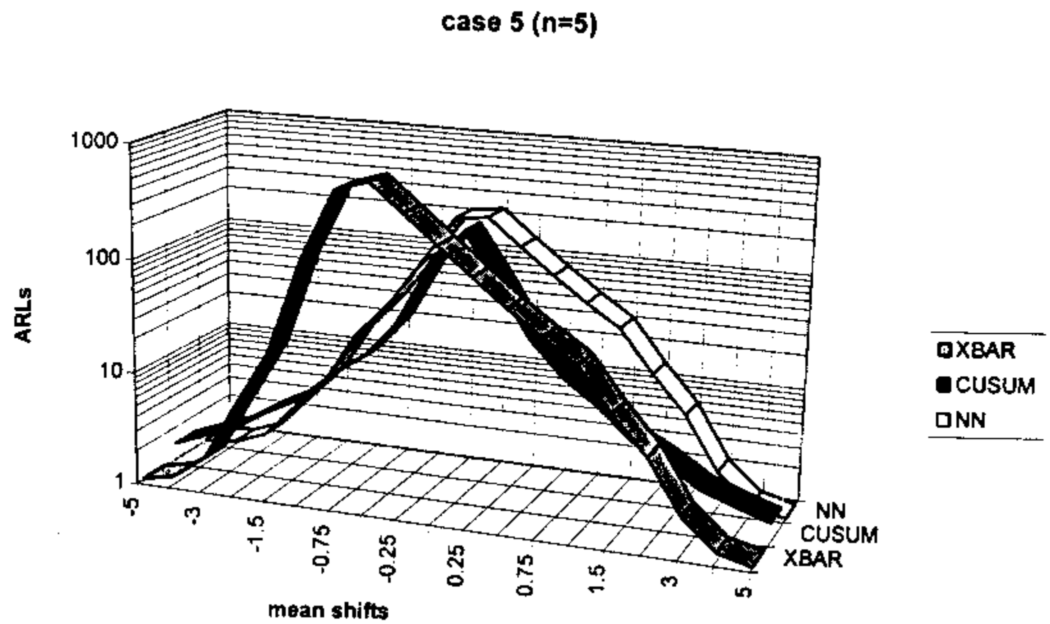
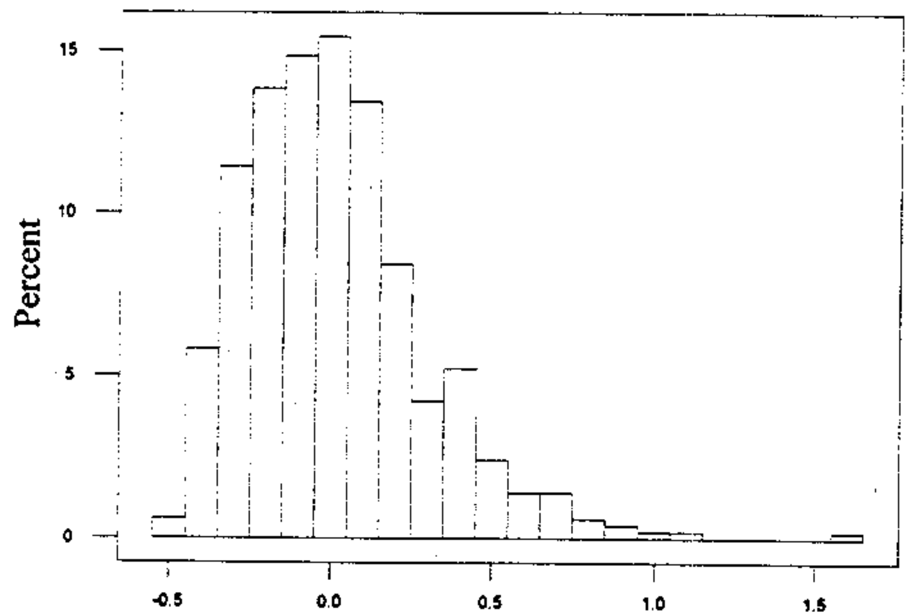


figure 16. The shape of distribution

case 6 (individual observations)



case 6 (averages, n=5)

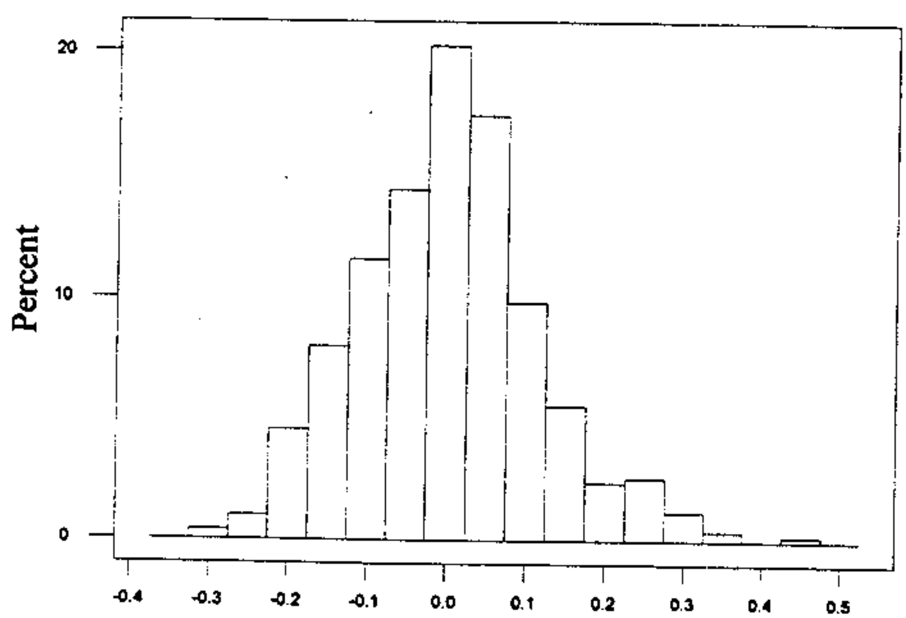


figure 17. Comparison of ARLs of the three methods

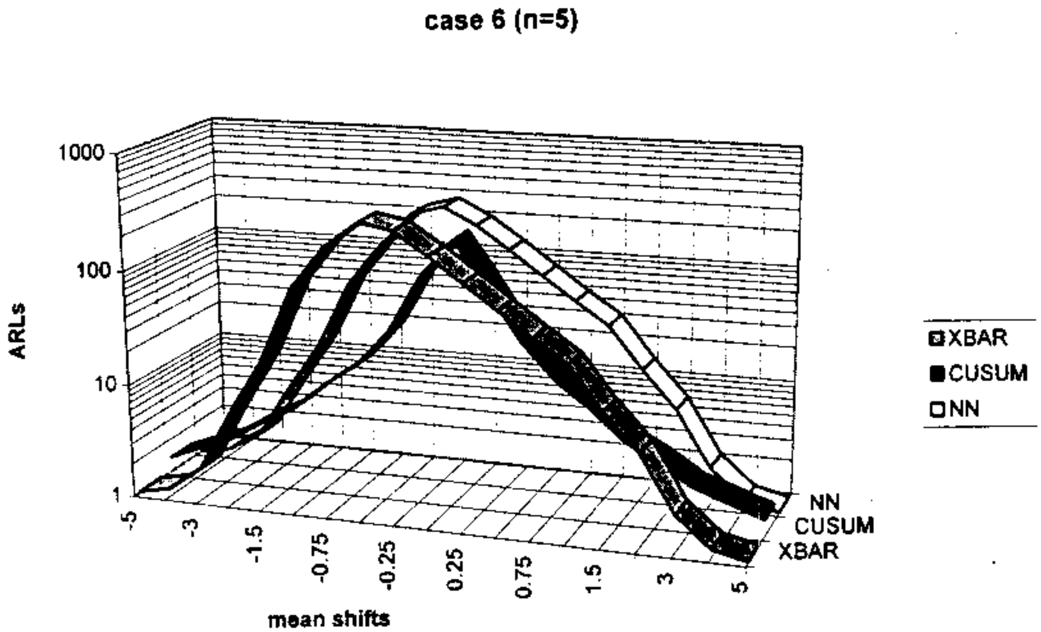
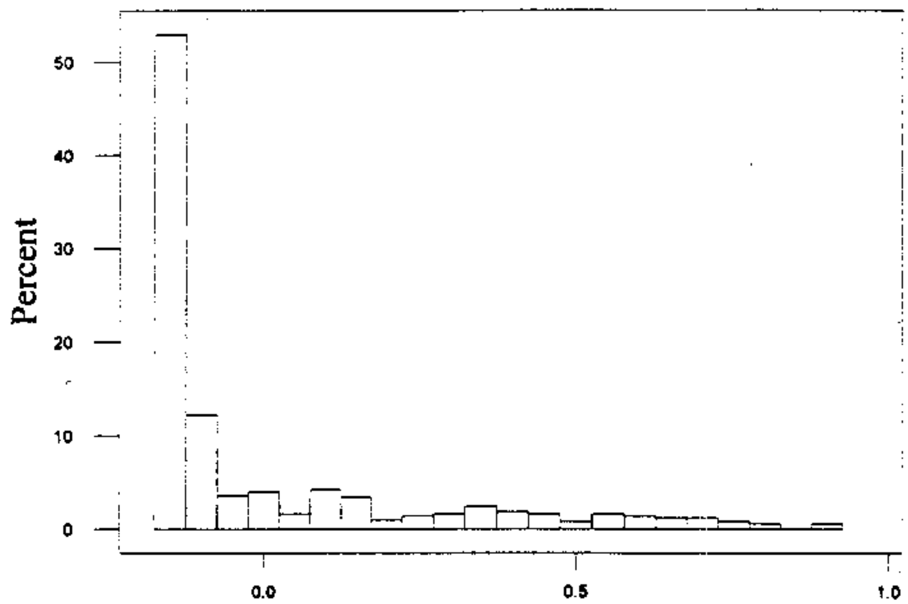


figure 18. The shape of distribution

case 7 (individual observations)



case 7 (averages, n=5)

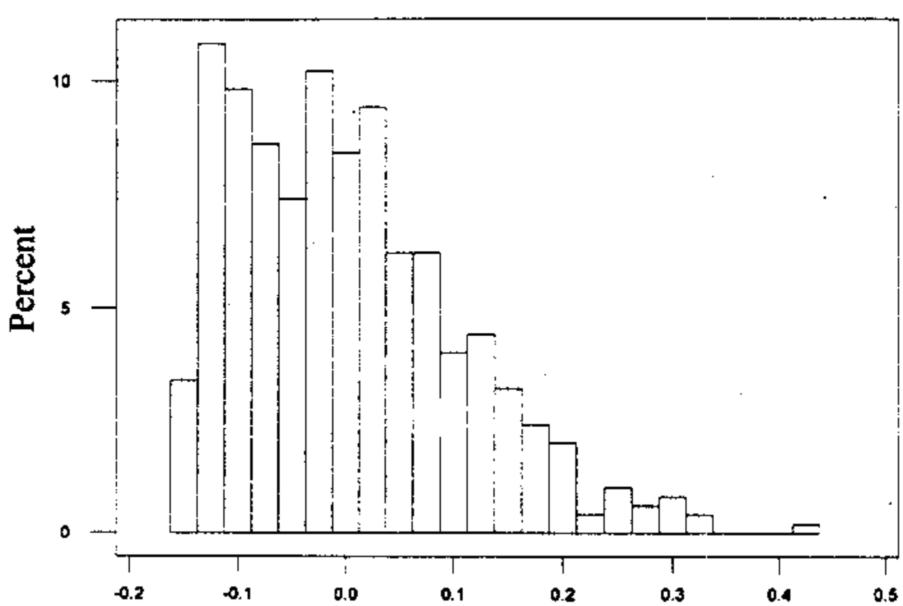


figure 19. Comparison of ARLs of the three methods

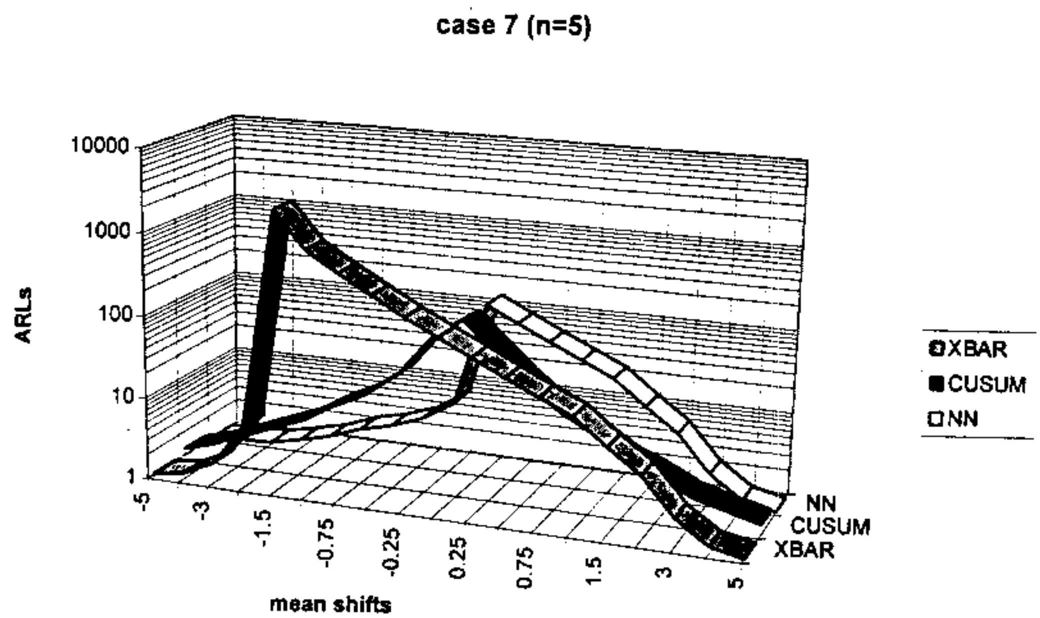
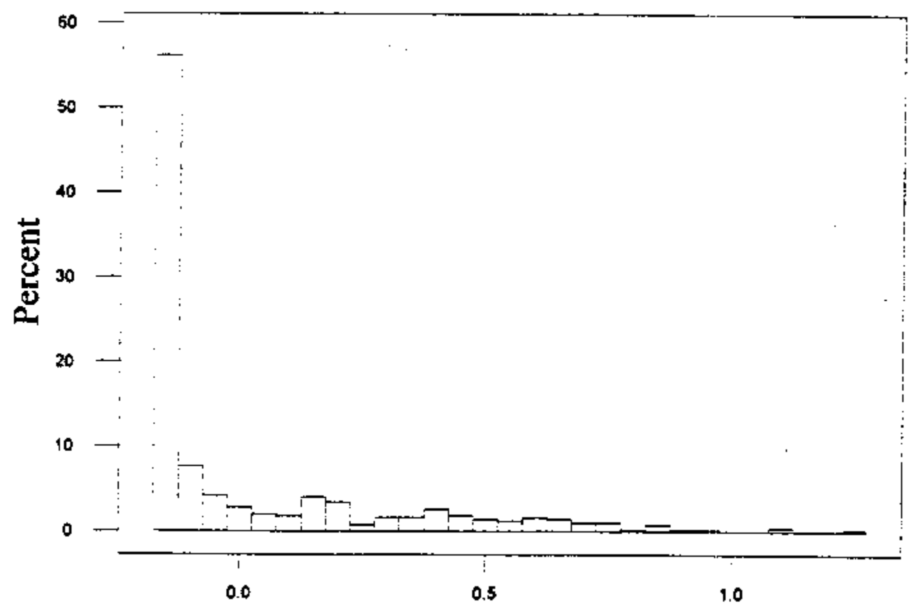


figure 20. The shape of distribution

case 8 (individual observations)



case 8 (averages, n=5)

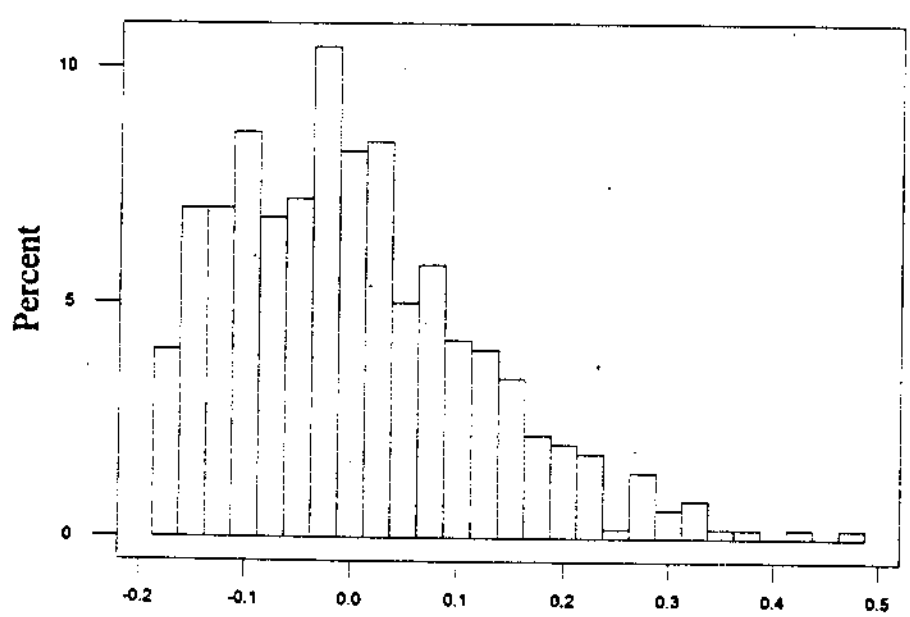


figure 21. Comparison of ARLs of the three methods

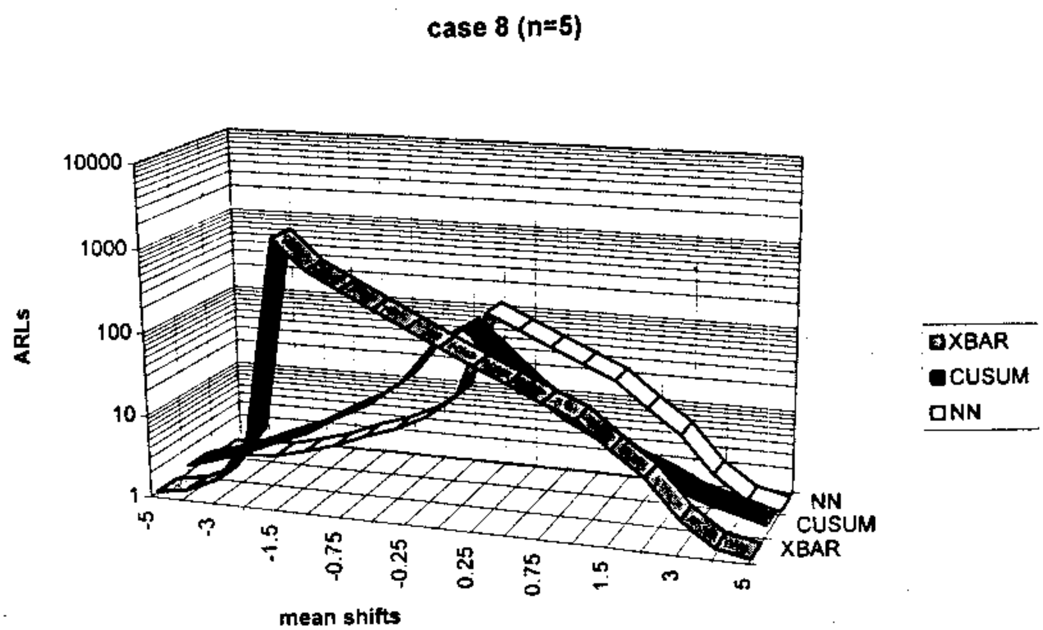
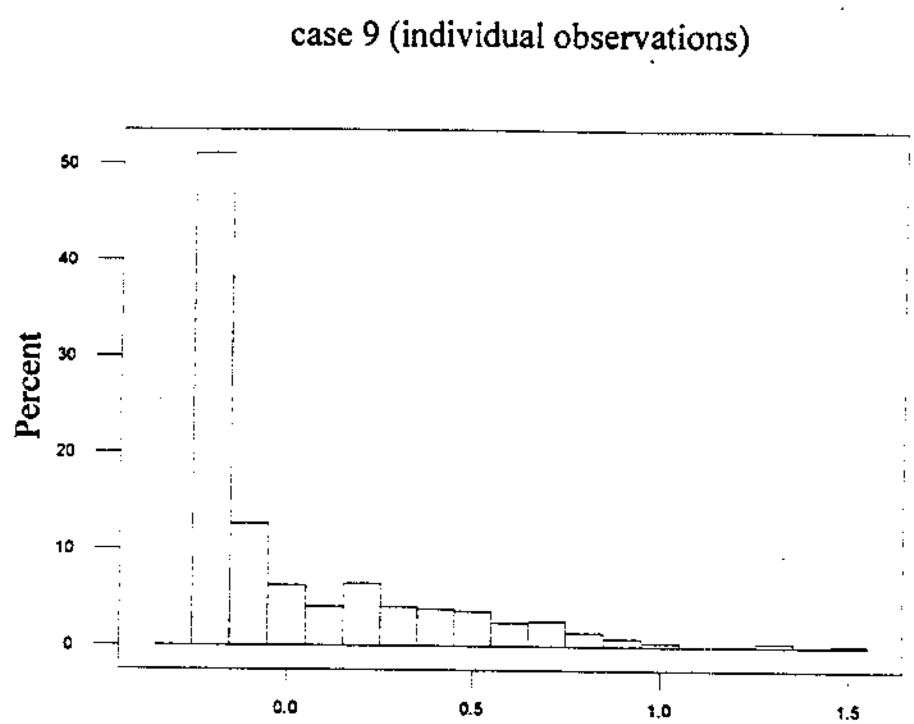


figure 22. The shape of distribution



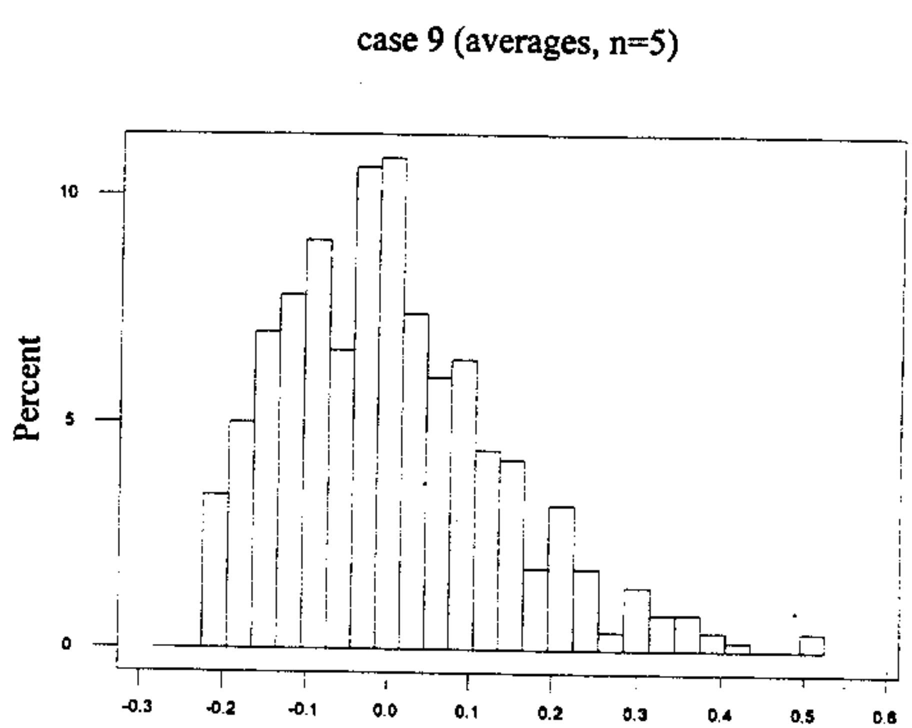
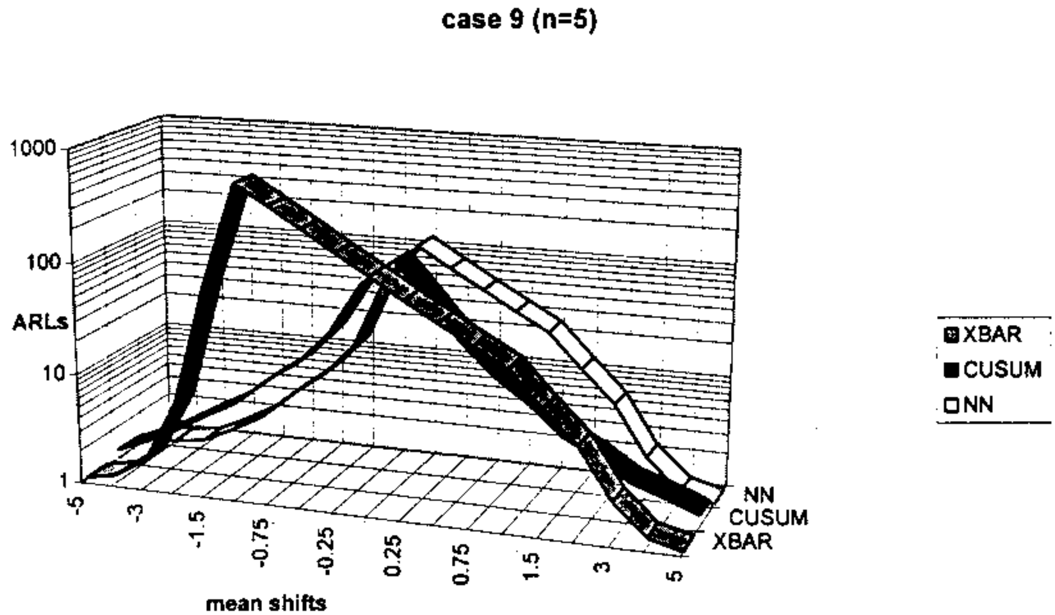


figure 23. Comparison of ARLs of the three methods



APPENDIX C

FORTRAN COMPUTER PROGRAM FOR CALCULATING ARLs

```
      integer inp, replic, loops, trials
      real tval, cutoff1, cutoff2, c, k, mean, stdv
      parameter (inp=5,  replic=1, loops=100,
     +     trials=4000,tval=1.812,cutoff1=0.845,cutoff2=0.36,
     +     cutoff3=3.764, stdv=.308432,
     +     c=-5.864337, k=0.044223, mean=0.218459)
c    5 input values, 1 output value, 11 reps to compute C.I.,
c    1000 loops to compute mean run length, 4000 trials to
c    determine run length for each loop
      integer   i, j, m, q, iseed, runlnt(3), ranout(3)
      real xr(inp),xout(4),shift,ssqrl(3),uconl(3), z, shnew,
     +    slnew,shold,slold,lconl(3),meanrl(3),
     +    stdrl(3), grandm(3)
      logical   cntrl1, cntrl2, cntrl3
c Note: xout(1) will store the sample output for xbar and cusum charts
c while xout(2-4) will store the three node output for neural net
c **********************************************************
c    initialize variables
c **********************************************************
      open(unit=2,file='unsort/all3.out')
      open(unit=3,file='unsort/all3.ind')
c    output file named for normal with shift 1
c    Start out by initializing rn seed, counters, and stats
      iseed   = 123457
      call rnset (iseed)
      do 5 q=1,3
        grandm(q) = 0.0
        ssqrl(q) = 0.0
        ranout(q) = 0
5     continue
      shift = 0.0
c
c    set up for outputting individual results to a file
c
      write(3,*)'    xbar','      cusum ','    neural '
c
      do 200 i=1, replic
        meanrl(1) = 0.0
        meanrl(2) = 0.0
        meanrl(3) = 0.0
        do 180 j = 1, loops
c      Secondary loop for computing run length stats starts here
c      Start in control and having run length value at least 1
```

```
        cntrl1 = .true.
        cntrl2 = .true.
        cntrl3 = .true.
        runlnt(1) = 1
        runlnt(2) = 1
        runlnt(3) = 1
c   Initialize variables for use in cusum & to avoid warning
        shold = 0.0
        slold = 0.0
        shnew = 0.0
        slnew = 0.0
          z = 0.0
c   Now start tertiary loop for computing run length or
c   counting until data runs out
        do 100 m = 1, trials
c   Generate 5 burr random variates using inverse transform
c   & store in array xr
          do 10 q=1,inp
            xr(q)=-mean+((1-rnunf())**(-1/k)-1)**(1/c)
            xr(q) = xr(q)+shift
10         continue
c*nor        call rnnor(inp,xr)
c  Compute new xbar only if xbar or cusum chart is still in
c  control for this loop
c
          if(cntrl1 .or. cntrl2) then
           sumx=0.0
           do 20 q = 1, inp
            sumx=sumx+xr(q)
20          continue
           xout(1)=sumx/real(inp)
c
c  First check if xbar chart is in control before reusing it
c
           if (cntrl1) then
c
c  Check the computed xbar with control limits adjusted for "n"
c  If value is within limits, process is in control, add 1 to counter
c
            if((xout(1) .ge. -0.4138) .and.
    +          (xout(1) .le. 0.4138)) then
             runlnt(1) = runlnt(1) + 1
            else
```

```
c   Else process is out of control, set flag then prepare to jump
c   out of loop for tallying run length
c
              cntrl1 = .false.
            endif
          endif
c
c  Then check if cusum is indeed in control before using this chart
          if (cntrl2) then
c
c Compute zvalue by multiplying xbar with reciprocal of stderr
c
            z = xout(1)*sqrt(real(inp))/stdv
c
c  Next call cusum subroutine to update sh and sl new values
c
            call cusum(z,shold,slold,shnew,slnew)
c
c  Then update the sh and sl old values with the new ones
c
            shold = shnew
            slold = slnew
c
c  Check the computed sh/sl values with control limits
c
            if((shnew.le. cutoff3).and.(slnew.le. cutoff3)) then
c
c  If both are less than 5 then still in control, add 1 to counter
              runlnt(2) = runlnt(2) + 1
            else
c   Else process is out of control, set flag then prepare to jump
c   out of loop for tallying run length
c
              cntrl2 = .false.
            endif
          endif
        endif
c   Now whether or not xbar or cusum is in control, check if Neural
c   net still signals "in control" before calling neural subroutine
c
        if(cntrl3) then
          call neural(xr,xout)
c
```

```
c    If the index value is within a specified interval, process
c    is in control, add 1 to the separate neural net counter
c
       if( (xout(2).le.cutoff1) .and. (xout(3).le.cutoff2)) then
         runlnt(3) = runlnt(3) + 1
         else
c   Else process is out of control, set flag then prepare to jump
c   out of loop for tallying run length
c
          cntrl3 = .false.
         endif
        endif
c
c  Jump out of loop only if all 3 monitors signal "out of control"
c
        if((.not. cntrl1) .and. (.not. cntrl2) .and.
   +       (.not. cntrl3)) then
         goto 120
        endif
c
c       End tertiary (inner loop)
100      continue
c
c       Tally run lengths for the jth loop
c
120      do 140 q = 1,3
        meanrl(q) = meanrl(q) + runlnt(q)
140      continue
c
c       Check if "in-control" flags not switched then tally # times
c       ran out of data
c
       if(cntrl1) ranout(1) = ranout(1) + 1
       if(cntrl2) ranout(2) = ranout(2) + 1
       if(cntrl3) ranout(3) = ranout(3) + 1
c
c      End the secondary "j" loop
180     continue
c      Now compute the mean run lengths over the # of loops
c
      do 190 q = 1,3
       meanrl(q) = meanrl(q)/real(loops)
190     continue
```

```
c
c     write the individual mean values for each loop to a file for
c     ANOVA section of data analysis
c
      write(3,230)meanrl(1),meanrl(2),meanrl(3)
c
c     Then begin to accumulate stats for the grand mean over all
c     the replications
c
      do 195 q = 1,3
        grandm(q) = grandm(q) + meanrl(q)
        ssqrl(q) = ssqrl(q) + meanrl(q)**2
195     continue
200   continue

c   Compute overall mean and standard deviation
    do 210 q = 1, 3
      grandm(q) = grandm(q)/real(replic)
      stdrl(q)=((ssqrl(q)-replic*grandm(q)**2)/real(replic-1))**.5
c
c   Next compute confidence limits for overall mean run length
c
      lconl(q) = grandm(q)-tval*stdrl(q)/sqrt(real(replic))
      uconl(q) = grandm(q)+tval*stdrl(q)/sqrt(real(replic))
210   continue
c
c   Then write output to file
    write(2,*) 'method    meanrl    stdrl   Low      High'
    write(2,220)' xbar',grandm(1),stdrl(1),lconl(1),uconl(1)
    write(2,220)' cusum',grandm(2),stdrl(2),lconl(2),uconl(2)
    write(2,220)'neural',grandm(3),stdrl(3),lconl(3),uconl(3)
    write(2,*)'times xbar ran out of data = ',ranout(1)
    write(2,*)'times cusum ran out of data = ',ranout(2)
    write(2,*)'times neural ran out of data = ',ranout(3)
    call rnget(iseed)
    write(2,250)iseed
220   format(a6,4f10.2)
230   format(3f10.2)
240   format(7f8.6)
250   format(I17)
    close(unit=2,status='keep')
    stop
    end
```

```
c
c
c
      subroutine neural(rayin, rayout)
      real rayin(*), rayout(*)
      real inpsum, a, b, c
      real inpnod(6), hidnod(5)
      real outnod(3)
      integer u,v
c
      real rmin(5) /-1, -1, -1, -1, -1/
c
      real rmax(5) /2.7, 2.7, 2.7, 2.7, 2.7/
c
      real wgt1(6,5) /5.5277, -2.4075, -2.5717, -2.9476, -2.6488,
     +-2.8981, -3.6721, 1.6080, 1.3728, 1.8384, 1.2456, 1.8865,
     +1.3999, -1.9442, -1.7115, -2.1664, -1.8643, -2.0200,
     +-1.7169, 0.1495, 0.1932, 0.2337, 0.5272, -0.3038,
     +2.8305, -2.4705, -2.2854, -2.4717, -3.0897, -2.8317/
c
c
c
      real wgt2(6,3) / -1.1316, 0.8302, -.9469, 1.0797, -.4816,
     +1.9036, .6154, -2.5390, 1.367, -.0156, .3363, -.1345,
     +-.2306, 1.8431, -1.3074, -1.8502, -.3724, -2.1134/

      do 400 u = 1, 5
        a = rayin(u)
        b = rmin(u)
        c = rmax(u)
        inpnod(u+1) = scalin(a,b,c)
400   continue
      inpnod(1) = 1.0
c
      do 430 v = 1,5
        inpsum = 0.0
        do 420 u =1,6
          inpsum = inpsum + inpnod(u)*wgt1(u,v)
420     continue
        hidnod(v) = sgmoid(inpsum)
430   continue
c
      do 450 v = 1,3
```

```
      inpsum = 0.0
      do 440 u =1,5
        inpsum = inpsum + hidnod(u)*wgt2(u+1,v)
440     continue
      inpsum=inpsum+wgt2(1,v)
      outnod(v) = sgmoid(inpsum)
450   continue
     do 460 v = 1,3
      rayout(v+1)=(outnod(v)-.2)/0.6
      if(rayout(v+1) .lt. 0.0) rayout(v+1)=0.0
c Note v+1 is used in rayout array to force 3 output node values into
c the last 3 cells of the array.  The first is being used in the main
c program to store xbar and cusum chart sample output values.
c
460   continue
     return
     end
c
c
c
     real function sgmoid(x)
     real x,k
     k = 1.0/(1.0+exp(- x))
     if (k .lt. 0.01) k = 0.0
     sgmoid = k
     return
     end
c
c
c
     real function scalin(value, r_min, r_max)
     real value, r_min, r_max
     if (value .lt. r_min) value = r_min
     if (value .gt. r_max) value = r_max
     scalin = (value - r_min) /(r_max - r_min)
     return
     end
c
c
c
c      real function scalout(value, r_min,r_max)
c      real value, r_min, r_max, x
c      x = (value - 0.2) / 0.6
```

```
c     if (x .lt. r_min) x = r_min
c     if (x .gt. r_max) x = r_max
c     scalout = x
c     return
c     end
c
c
      subroutine cusum(zval,sho,slo,shn,sln)
      real zval, sho, slo, shn, sln
c
      shn = zval - .5 + sho
      if (shn .lt. 0) then
        shn = 0.0
      endif
c
      sln = -zval - .5 + slo
      if (sln .lt. 0) then
        sln = 0.0
      endif
c
      return
      end
```

REFERENCES

Aft, Lawrence S. 1988. *Quality Improvement Using Statistical Process Control*. San Diego: Harcourt Brace Jovanovich.

Archer, Norman, and Shouhong Wang. 1993. Application of the back propagation neural network algorithm with monotonicity constraints for two-group classification problem.*Decision Science* 24: 60-75.

Austin, John A. Jr.. 1973. Control chart constants for largest and smallest in sampling from a normal distribution using the generalized burr distribution. *Technometrics* 15(4): 931-933.

Burke, Laura I. and James P. Ignizio. 1992. Neural networks and operations research: an overview. *Computers Operations Research* 19 (3/4): 184-185 and 255.

Burr, Irving W. 1942. Cumulative frequency functions. *Annals of Mathematical Statistics*, 13: 215-232

______________. 1967. The effect of non-normality on constants for $\overline{x}$ and R charts. *Industrial Quality Control* May: 563-569

______________. 1973. Parameters for a general system of distributions to match a grid of α_3 and α_4. *Communications in Statistics* 2: 1-21.

Caudill, Maureen. 1990. Neural network Primer Part I. *AI EXPERT* 2 (December): 46-52.

Champ. Charles W. and William H. Woodall. 1987. Exact results for Shewhart control charts with supplementary runs rules. *Technometrics.* 29(4): 393-399.

Cheng, Chuen-Sheng. 1995. A multi-layer neural network model for detecting changes in the process mean. *Computers Industrial Engineering* 28 (1): 51-61.

Chiu, W.K. and K.C. Cheung. 1977. An economic study of $\overline{x}$ charts with warning limits. *Journal of Quality Technology* 9(4): 166-171.

Collani, Elart Von. 1986. A simple procedure to determine the economic design of an $\overline{x}$ control chart. *Journal of Quality Technology* 18(3): 145-151.

Crowder, Stephen V. 1987. A Simple method for studying run length distributions of exponentially weighted moving average charts. *Technometrics* 29: 401-407.

Davis, Stanton, and Bill Illingworth. 1989. *Neural Network Simulation Applied to Statistical Process Control.* Dallas: Texas Instruments Incorporated.

Duliba, Katherrine A. 1991. Contrasting neural nets with regression in predicting performance in the transportations industry. *Proceedings of the Twenty-fourth annual Hawaii International Conference on System Sciences* 4: 163-70.

Duncan, A. J. 1956. The economic design of $\overline{x}$ -charts used to maintain current control of a process. *Journal of American Statistical Association* 51: 254-256.

Dutta, Soumitra, and Shashi Shekhar. 1988. *Bond Rating: A Non-conservative Application of Neural Networks.* International Conference on Neural Networks: 443-450.

Evans, James R., and William M. Lindsay. 1989. *The Management and Control of Quality.* St. Paul: West.

Grant, Eugene L. and Richard S., Leavenworth. 1988. *Statistical Quality Control,* Sixth Edition, New York: McGrow-Hill.

Hayter, Anthony J., and Kwok-Leung Tsui. 1994. Identification and quantification in multivariate quality control problems. *Journal of Quality Technology* 26(3): 197-208.

Hill, Tim, Leorey Marquez, Marcus O'Connor, and William Remus. 1994. Artificial neural network models for forecasting and decision making. *International Journal of Forecasting* 10: 5-15.

Ho, Chuanching, and Kenneth E. Case. 1994. Economic design of control charts: a literature review for 1981-1991. *Journal of Quality Technology* 26(1): 39-53.

Hwarng, H. Brian and Norman Faris Hubele. 1993a. Back-propagation pattern recognizers for $\overline{x}$ Control charts: methodology and performance. *Computers Industrial Engineering* 24(2): 219-235.

Hwarng, H. Brian and Norman Faris Hubele. 1993b. $\overline{x}$ control chart pattern identification through efficient off-line neural network training, *IIE Transactions* 25(3): 27-38.

Hwarng, H. Brian. 1995. Proper and effective training of a pattern recognizer for cyclic data, *IIE Transactions* 27: 746-956.

Juran J,M.,and Frank M. Gryna. 1988. *Juran's Quality Control Handbook,* 4th Ed. New

York: McGraw-Hill.

Juran J,M.,and Frank M. Gryna. 1993. *Quality Planning and Analysis*. 3rd Ed. New York: McGraw-Hill.

Jurik, Mark. 1992. Trading techniques: the care and feeding of a neural network, *Futures: The Magazine of Commodities & Options* 21: 40-44.

Keats, J. Bert, and Norma Faris Hubele. 1989. *Statistical Process Control in Automated Manufacturing*. New York: Marcel Dekker.

Lashkari R.S. and M.A. Rahim. 1982. An economic design of cumulative sum charts to control non-normal process means. *Computers Industrial Engineering* 6 (1): 1-18.

Lawrence Jeannette. 1991. Data preparation for a neural network. *AI EXPERT* November : 34-41.

Lucas, J. M. 1982. Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology* 14: 134-142.

Lucas, J. M. and R. B. Crosier. 1982. Fast initial response for CUSUM quality control schemes. *Technometrics* 24: 171-182.

Lydia Kronsjo, and Dean Shumsheruddin. 1992. *Advances in Parallel Algorithms*. John Wiley & Sons, Inc..

Marquez, Leorey, Tim Hill, Reginald Worthley, and William Remus. 1991. Neural network models as alternative to regression, IEEE, *Proceedings of the Twenty-four Annual Hawaii International Conference on System Sciences* 4: 129-135.

Montgomery, Douglas C. 1980. The economic design of control charts: A review and literature Survey. *Journal of Quality Technology* 12 (2): 75-87.

______________________.1991a. *Design and Analysis of Experiments*. 3rd Ed. New York, Wiley.

______________________. 1991b. *Statistical Quality Control*, 2nd Edition,New York, Wiley.

Nam, Kyungdoo T. 1993. *A Heuristic Procedure For Specifying Parameters in Neural Network Models For Shewhart X-Bar Control Chart Application*. unpublished dissertation: University of North Texas.

Neural Ware. 1991. *Neural Computing: Neural works Professional II/ PLUS and Neuralworks Explorer*. Pittsburgh: Neural Ware.

NeuroShell: *Neural network shell program*. 1991. Frederick, MD: Ward Systems Group.

Ott, Lyman.1988. *An introduction to statistical methods and data analysis*. Boston, PWS-Kent.

Page, E. S. 1954. Continuous inspection schemes. *Biometrics* 41: 342-350.

Philipoom, Patrick R., Rees,Loren Paul, and Wiegmann, Lars. 1994. Using Neural networks to determine internally-set due-date assignments for shop scheduling. *Decision Sciences*25: 825-851.

Prybutok, Victor R., Clive C. Sanford, and Kyungdoo Ted Nam. 1995. A Comparison of Neural network to shewhart applications. Submitted and Accepted to the *Journal of quality Control*.

Prybutok, Victor R. and Kyungdoo Ted Nam. 1995. Application of neural networks to statistical production/inventory forecasting. *Midwestern Business and Economic Review* 21: 33-39.

Pugh, G. Allen. 1989. Synthetic neural networks for process control. *Proceedings of the 13th Annual Conference of Computers Industrial Engineering* 17 (1-4): 24-26.

Pugh, G. Allen, 1991. A Comparison of neural networks to spc charts. *Proceedings of the 13th Annual Conference of Computers Industrial Engineering* 21(1-4): 253-255.

Quesenberry, Charles P. 1993. The effect of sample size on estimated limits for $\overline{x}$ and x control charts, *Journal of Quality Technology* 25(4): 237-247.

Raghupathi, Wullianalur, Lawrence L. Schkade, and Badi B. Raju. 1991. A Neural Network application for bankruptcy prediction, *Proceedings of the Twenty-fourth Annual Hawaii International Conference on System Sciences* 4: 147-155.

Rahim, M. A. 1985. Economic Model of $\overline{x}$ Charts Under non-normality and measurement errors. *Computers and Operations Research* 12 (3): 291-299.

Rahim, M.A. 1989. Determination of optimal design parameters of joint $\overline{x}$ and R charts. *Journal of Quality Technology* 21(1): 65-68.

Ramsey, Patricia P. and Philip H. Ramsey. 1990. Simple Tests of normality in small samples, *Journal of Quality Technology* 22 (4): 299-309.

Rigdon, Steven E., Emma N. Cruthis, and Charles W. Champ. 1994. Design strategies for individuals and moving range control charts. *Journal of Quality Technology* 26(4): 274-287.

Rodriguez, Robert N. 1977. A guide to the Burr type XII distributions. *Biometrica* 64(1): 129-134.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. 1986. *Parallel Distributed Processing*, Cambridge, MA: MIT Press.

Runger, George C. and Hoseph J. Pignatiello, Jr.. 1991. Adaptive sampling for process control. *Journal of Quality Technology* 23: 135-155.

Ryan, Thomas P. 1989. *Statistical Methods for Quality Improvement.* New York: Wiley.

Salchenberger, Linda M., E. Mine Cinar, and Nicholas A. Lash. 1992. Neural networks: A New tool for predicting thrift failures. *Decision Sciences* 23: 899-916.

Saniga Erwin M. 1989. Economic statistical control chart designs with an application to $\overline{x}$ and R charts. *Technometrics* 31 (3): 313-320.

Schilling, Edward G. and Peter R. Nelson. 1976. The effect of non-normality on the control limits of $\overline{x}$ charts. *Journal of Quality Technology* 8 (4): 183-188.

Tagaras, George. 1989. Economic $\overline{x}$ charts with asymmetric control limits. *Journal of QualityTechnology* 21 (3): 147-154.

Tam, Kar Yan, and Melody Y. Kiang. 1992. Managerial applications of neural networks: The case of bank failure predictions. *Management Science* 38 (7): 926-947.

Tracy, Nola D., John C. Young, and Robert L. Mason. 1992. Multivariate control charts for individual observations. *Journal of Quality Technology* 24 (2): 88-95.

Vance, Lonnie C.. 1983. A bibliography of statistical quality control chart techniques, 1970-1080. *Journal of Quality Technology* 15: 59-62.

Vardeman, Stephen and Di-ou Ray. 1985. Average run lengths for CUSUM schemes when observations are exponentially distributed. *Technometrics* 27 (May): 145-150.

Velasco, Tomas and Mark R. Rowe. 1993. Back propagation artificial neural networks for the analysis of quality control charts. *Computers and Industrial Engineering* 25 (1-4): 397-400.

Wadsworth, Harrison M., Kenneth S. Stephens. 1986. *Modern Methods for Quality Control and Improvement*. New York: Wiley.

Walker, Esteban, John W. Philpot, and James Clement. 1991. False Signal rates for the Shewhart control chart with supplementary runs tests. *Journal of Quality Technology* 23 (3): 247-252.

Wasserman, Philip D. 1989. *Neural Computing*. New York: Van Nostrand Reinhold.

Western Electric. 1956. *Statistical Quality Control Handbook*. Western Electric Corporation. Indianapolis, Ind.

Wong, F. S. 1991. A 3D neural network for business forecasting. *Proceedings of the Twenty-fourth annual Hawaii International Conference on System Sciences* 4: 113-23.

Woodall, W. H. 1986. The design of CUSUM quality control charts. *Journal of Quality Technology* 18: 99-102.

Yourstone, Steven A. and William J. Zimmer. 1992. Non-normality and the design of control charts for averages. *Decision Sciences* 23: 1099-1113.