The ARMA (1,1) Model in the Context of the Joint Estimation Technique and Outlier Detection in Quality Control

A dissertation submitted to the
Kent State University Graduate School of Management
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

by

John K. Mahaney, Jr.

May 2003

UMI Number: 3098515

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Dissertation written by

John K. Mahaney, Jr.

A.B., Catholic University of America, 1964

B.Ch.E., Catholic University of America, 1964

M.S.M.., University of Akron, 1994

Ph.D., Kent State University, 2003

______________________ Chair, Doctoral Dissertation Committee

______________________ Members, Doctoral Dissertation Committee

____

____

____

Accepted by

______________________ Doctoral Director, Graduate School of Management

______________________ Dean, Graduate School of Management

Acknowledgments

One does not finish a dissertation and the other work for the Doctoral degree alone. A lot of people have helped and supported me during this lengthy process.

First, I would like to thank my wife, Jolean, for putting up with my education efforts over the past eleven years. Second, I have appreciated the support of our children, Sean, Kristin, and Brian, who I think feel it is neat to have such a non-traditional student as a father. Third, I have received much support from my business colleagues across the country and I deeply appreciate their support.

I greatly appreciate the help that Christine M. Wright gave me in understanding and using the SCA software we employed in the work reported here.

A project like this can only go forward with the immense amounts of help, advice, and counseling one receives from his dissertation committee. This has been a long and laborious process with many fits and starts. The support and counsel of David Booth, Felix Offodile, and Michael Hu, all of the College of Business Administration at Kent State University, were very necessary to achieving the end of the project. Without their help and direction, I would never have completed this work. I am indebted to them for their support.

Table of Contents

Tables

CHAPTER 1

Introduction

In various industries today, the use of statistical process control charts is common. These charts range from the traditional Shewhart charts through more sophisticated Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts that are considered superior to the conventional charts for determining small movements of the process average. (Hunter, 1986) These various charts have served well and are an important simple tool, as Deming would list them, in the continuing effort to improve processes and products throughout the world.

However, all of the control charting methods make the assumption that the individual data points are independent and that the underlying distribution of each sample is identical to prior samples. (Montgomery & Mastrangelo, 1991; Hamburg, Booth & Weinroth, 1996) This is the IID assumption - independent and identically distributed data. But, many authors, including Box and Jenkins, have noted that the data are probably not IID. Alwan and Roberts (1995) have in fact argued that most of the common data sets used as examples over the years are not IID.

Many authors have indicated that the data are actually time series and have shown that particular data sets are particular time series. However, with the exception of work reported by Prasad, Booth, Hu and Deligonul (1995A), no one has systematically examined a number of data sets to determine the exact time series model that is applicable to that data set.

The IID assumption does make the use of the charts much more simple, than using a time series. The basic rules for dealing with outliers, data points that do not appear representative of the process or product, can be applied and used to differentiate between true outliers and false alarms. Conventionally, outliers were considered to be observations that were more than 3.0 standard deviation units from the process mean. Other rules, such as the rule of runs where seven (7) consecutive points in a row above or below the mean indicates a process change, can also be applied to identify possible outliers (LTV Steel Policy Guidelines for Control Charts, 4th Edition, 1996). However, all these rules are based upon the probability parameters of randomly selected data values from IID data sets. If IID does not apply, these rules lack applicability.

There have been a number of mathematical methods developed and tested to better describe the data in control charts which is non-IID. Various methods such as GM-estimators (Booth, 1984 and Booth, Acar, Isenhour, and Ahkam, 1990) to identify outliers have been used. Polynomial smoothing is another mathematical method reported by Sebastian, Booth and Hu (1994). Grznar, Booth, and Sebastian also reported robust smoothing in 1997. Of current interest is the Joint Estimation Technique of Chen and Liu (1993). This latter method is a time series based approach which identifies types of outliers and does not require the IID assumption. In this work, we will again use outlier identification with time series models to deal with quality control data that is not IID.

There are different types of time series outliers known. Chen and Liu (1993) list four types as an Additive Outlier (AO), an Innovative Outlier (IO), a Level Shift Outlier (LS), and a Temporary Change Outlier (TC). Each of these outlier types has a particular pattern which can make them recognizable when the Chen and Liu model is applied. To better understand the types of outliers we need to define each of the four types.

All outliers are events that impact the time series. An Additive Outlier (AO) affects the time series for only one time period or data point. An Innovative Outlier (IO) affects following values observed after its occurrence However, the effect is temporary as the effect of the IO decays. It is noted that this type of outlier frequently shows the impact of an external cause. A Level Shift Outlier (LS) represents a permanent change to the time series at a given time. Finally, a Temporary Change (TS) is change that has an initial impact but whose influence gradually dies out over time.

1.1 Research Objectives

The process control data is a time series. If it is not IID, what is the proper model and does that model vary for each data set? The answer to the second portion of the question is obviously, yes, the model can vary. Thus, the proper model can be, different for each data set and must be determined on a case by case basis. In today's manufacturing environment, computer support of processes and process control systems can provide the necessary computing power, to allow on-line or real-time application of these techniques. Further, the number of data points necessary to accurately define a model is considered by many to be very large, in excess of 50 data points for example (Box and Jenkins, 1976). Wright (1997), however, has pointed out that one can achieve outlier detection with as few as 9 data points. Small data sets allow for more rapid response to changes in the process and thus less material produced that might not be acceptable in the marketplace. However, we still do not know whether model selection is critical to the determination of such outliers. Thus, we have a reported need for large data sets in order to determine the exact time series model, but the possibility of locating outliers with much smaller data sets. If we need to determine the exact model, large data sets may be needed. Can an approximation perform adequately to find outliers without having to determine the exact time series model? Answering the question is the objective of this research.

The focus of this research is to determine if a simple model can be used as an approximation to the mathematically correct model, and still provide sufficient efficiency in outlier detection as the exact model. This is in the spirit of a phrase attributed by many to Box but by others to Tukey: "All models are wrong but some are useful." Among the simple models that will be discussed are the AR (1), MA (1), ARMA (1,1), and ARIMA (1,1,1) models. Various levels of variability will be used for the outlier detection. In addition, control charts will be generated as a standard method of analyzing the data. This will then allow us to compare a control chart, the proper time series model, and various simple time series models to determine if a simple model can be used to adequately identify how a series of data points happens, as well as locating outliers and identifying them.

The type of control chart to be used as a standard is the chart of individuals with a moving range of two. While we are stating that control charts are not always mathematically correct, they are a well understood method of data presentation and examination and any proposed method should be as effective and efficient as the base charts. We will use limits of 2.0 standard deviations. When one examines charts of individuals, if the limits are set at 3.0 standard deviations, there are very few data points that will fall outside the limits because the limits are so broad. The use of rational subgroups in conventional control charts reduces the allowed range of variability by one over the square root of the sample size.(Anderson, Sweeney & Williams, 2001) Thus, a sample size of four (4) cuts the allowed variation in half, as compared to a sample size of one (1) where it is hard to calculate the standard deviation of the sample.. Thus, the use of 2.0 standard deviations as the limiting value for outlier detection puts that value very close to the value used for the grouped data. In addition, the LTV Steel Policy Guidelines for Control Charts, 4^{th} Edition (1996) suggests the use of 2.0 standard deviations for the control limits for charts of individuals. Thus, this approach conforms to current industrial practice.

Two control chart types, CUMSUM and EWMA were introduced to deal with some shortcomings in the conventional charts. A lengthy discussion of these types of charts is found later in this dissertation. Control charts of EWMA and CUMSUM type are very useful for determination of small shifts in process averages as discussed by Montogomery (1991), Wardell, Moskowitz, and Plante (1992), and Lucas and Saccucci (1990) for example. However, we are not looking for small shifts in process averages alone, but rather all shifts of various types. None of the authors noted above, or others to be found later in the section on CUSUM and EWMA charts, note any particular benefit of those chart types in the identification of outliers charts of individuals. Thus, for the base case we will use the more common charts of individuals with a moving range as the basis of comparison in locating outliers as compared to the ARMA (1,1) time series model. The ARMA (1,1) model is chosen because statistical process control (SPC) data sets are most likely to be one-dependent, if not II(Chernick, Downing, and Pike, 1982).

One must be aware that with time series calculations, autocorrelated data present problems. If the data are autocorrelated, as compared to independent, a data point is dependent upon some point some distance in time prior. A time series recognizes that future values are dependent upon more recent data point or points. The same is true of the EWMA control chart, but Lucas and Saccucci (1990) note the assumption of IID and if the data are autocorrelated, they obviously are not IID. Also, with EWMA charts, weighting factors are calculated and applied to the current and past data points to allow calculation of a future point. When the data are correlated, however, such weighting is often not possible. The next value may not be related to the current one but rather to some value well in the past. When such autocorrelation exists, the calculation of a reasonable time series estimate is difficult, if not impossible. A moving average, on the other hand, is an arithmetic average of the number of actual data points selected and that is the method to be used in this work. One also has to be aware that the outliers themselves impact the identification of the model in a time series and the outliers must be adequately identified and steps taken to decrease their on the preliminary calculations. The Joint Estimation technique of Chen and Liu (1993) does this very well and is the method that will be applied when attempting the application of the simple model to the data sets.

1.2 Research Contributions

The goal of this research is to provide a simple model that can be applied to various time series processes, and thus several forms of data sets, and still be efficient and effective in outlier detection and identification because outliers are the out of control points. If one can conclude that a single simple model is a suitable approximation for essentially any data set, then one can apply that model to small runs, typically start-up situations, and quickly determine whether outliers are present or if the process is in the state of statistical control. The conventional definition of the state of statistical control for grouped data is that state of nature where the system is operating within the 3.0 standard deviations control limits for that control chart with no particular patterns and no points beyond the 3.0 standard deviations levels that have been identified as outliers. For individual data points a different limit, 2.0 standard deviations, is suggested by LTV Steel (1996) and the work here confirmed that 3.0 standard deviations limits for charts of individuals were too broad to allow adequate interpretation of the data. The limits are so broad that many data points of interest in explaining the process are not located as potential outliers. The 2.0 standard deviations limits will be used in this work.

The advantage of applying a simple model is that the number of data points necessary to identify outliers can be very small. If effectiveness is locating the outlier and identifying it properly and efficiency is accomplishing this task with a small number of data points, the simple model is both effective and efficient. From an industrial point of view, the sooner, i.e., the use of a smaller number of data points, that an outlier can be determined, the more useful is the process of making that decision because such rapid outlier identification prevents the production of more items that do not meet specifications or customer requirements. The more quickly a process can be judged to be operating acceptably, the absence of outliers, the sooner production can be moved to the proper levels of output, speed, parts/hour, etc. If there is a problem, i.e. outlier, detected, the sooner the better is also the case as a smaller the number of out of specification items will be produced, and the faster the identification of the reason for the problem, the sooner good material can be produced. Thus, it is important to be able to quickly identify if the process is operating properly, or if it is not, to locate the source of the problem.

Of concern is the identification of too many outliers with the simple model as compared to the best model or a control chart. There is a cost associated with false alarms. A false alarm is the detection of a supposed outlier when, in fact, the data point is not an outlier. As a manufacturing operation must stop and identify the source or sources of such a point, one does not want to look for problems that do not exist. If the standard deviations standard applied is too stringent, less than 2.0 for a possible example, more outliers will be detected but a number of them will be false alarms. At any level, some items detected as possible outliers will be false alarms. Statistical theory would predict that approximately 4.56% of the data points in a normal distribution would be more than 2.0 standard deviations from the mean. Such points would be classed as potential outliers, but are not and thus are classified as false alarms. Of course, if one uses too broad a range for control, say over 3.0 standard deviation units, outliers that should be detected are not and we have the opposite condition where we are not finding potential problems. Thus, there is a balancing act between being too tight and too loose. By comparing the number of outliers detected under each of the models studied for a given data set, the relative efficiency of each method can be calculated and conclusions drawn as to the efficiency of the model and the control parameters. In the work at hand, we will compare the potential outliers determined by a chart of individuals with a moving range and 2.0 standard deviations limits against those potential outliers determined using the ARMA (1,1) model and the joint estimation technique of Chen and Liu (1993). We will look at whether the same data point is found by both methods and whether one method consistently locates more areas of interest, potential outliers, than the other.

Thus, it is our goal to determine if a simple time series model can be applied to process and product control data sets. We will examine the efficiency and effectiveness of each model against the best time series model and against a EWMA control chart. The first efforts will be with well used data sets but we will expand the work to cover certain industrial data sets that have become available for such analysis.

A simple model that works adequately would allow industrial applications to more quickly determine whether a start-up process is performing properly or not in a very short period of time because of the small number of data points or samples needed. This would be very beneficial to those operating such processes as it would save them time and money.

1.3 Research Overview

Chapter 2 of this work is the literature review. The literature will provide the data sets for analysis as well as the reports of other researchers in this area and thus the background for our efforts. Details of the various methods to be used, and those that will be used for comparison, will be noted in that chapter. Such topics as EWMA control charts, outlier types, time-series models, the joint estimation technique, and robust statistical methods will be discussed. Chapter 3 will cover the various techniques that are used as a basis for comparison with the proposed technique. This would include the control charts and various other techniques that have been used and reported in the literature. Chapter 4 examines the methodology and logic for the work and demonstrates the application of the method to data from production processes. Chapter 5 will include the application of the proposed technique to data sets of business data and an analysis of the results of those efforts. Business data will be further defined in that chapter but suffice it to say that such data would include sales reports, material performance reports, and other data commonly reviewed by management of a firm to determine how the firm is performing. Chapter 6 will provide conclusions and future research opportunities for the proposed technique.

CHAPTER 2

Literature Review

2.1 Introduction

To understand and control a process, one must have some measurements of process variables and the ability to track those variables in some manner. The Shewhart control chart methods, introduced in the 1930's (Shewhart, 1931), have been widely used for such purposes and have been very successful. However, there are certain assumptions made by Shewhart which are considered by many authors to be untenable. Much effort has been put into finding other methods that are more mathematically accurate than the Shewhart control charts, due to concerns about the assumptions. However, these alternatives are generally very complex and convoluted methods that do not lend themselves to easy use by the normal practitioner on the shop floor. Deming (1982) has counseled the use of 'simple tools' of statistical process control in order to have more involvement by the shop floor personnel. But, the methods described in the literature, which will be covered further in this chapter, and the proposed model, while not 'simple tools,' can be easily followed on the shop floor using personal computers once a control system has been set up.

The purpose of control charts and other methods of analysis is to provide a better understanding of the system under study. An analysis of charts should provide the expert in the area, the practitioner, with a better understanding of what is happening to the system under consideration. Any other method of analysis should also allow the practitioner to gain a better understanding. The better the method of analysis identifies shifts in the process average, outliers, and other changes in the system, the more useful that method is to the practitioner. At the same time, one does have to be aware of overly sensitive systems where too many signals of anomalies are received and this flooding of the system with areas of concern makes analysis of such a voluminous number of signals impossible. Thus, we are looking for a method that will provide more information to the practitioner than current control charts while identifying only those data points of real interest to the practitioner.

It has been suggested by many, among them Box and Jenkins (1976), that much of the data used in process control does not meet the Shewhart assumption of independent, identical distributions (IID), but rather are in fact time series (Wright, 1997). However, everyone who has examined that concept has been very conscious of determining and using the best model of the system. In most cases, that model definition stage is very difficult and time consuming, if it is possible at all. This again takes the method away from the 'simple tools' approach of Deming. On an industrial basis, the concept of 'simple tools' has a lot of benefit as the control of the process has to rest with the workers actually performing the work. To have trained statisticians or quality control/assurance personnel do the controlling with highly sophisticated charting techniques does not lead to increased control of the process. It is reasonable to have a cadre of highly trained people for problem solving help; design of experiment consultation, etc., but the processes have to be controlled by the people who are running it on a day to day basis. What we need is a 'simple tool' to apply to process control using time series models and analysis techniques, or at least one floor personnel can use.

This chapter will review the basic literature in process control methods and technology and then move to the time series methods. Following a review of the various methods used in the past, a review of the rationale for the use of a specific simple model of any time series will be given. The concept to be tested is that one model, ARMA (1,1) is sufficient to adequately describe and analyze almost all time series models of process control situations. Further, even in those cases where the model is a bit less than ideal, from the point of view of industrial applications, it may serve as a very good first approximation. The test will be that the ARMA (1,1) model is more than sufficient for determination of out of control points in process data.

Care should be taken to understand the importance of correct identification of problems. Acar and Booth (1987) note that one must identify the right problem in order to solve it. If we are incorrectly finding, or not finding problems, due to the method used to study the situation, we are not able to properly attack the real problem and thus it will never be solved and removed from the system. A simple method, such as control charts, while intuitively desired on an industrial basis, lacks some value if one finds points identified as out of control when they are not, false alarms, while still missing some other data points that are in fact representative of a problem.

As noted above, we are looking for a method that will help us to better understand the systems we are studying. In many cases, the current techniques have not located shifts and other outliers that from a practitioner's point of view should exist in a given data set. The failure to locate such shifts and other outliers using such methods means the usefulness of such methods is marginal in terms of fully understanding the process being considered. Methods that don't provide sufficient information are not a lot of use.

2.2 Shewhart

A major contribution to process control and improvement is the work of Shewhart published in 1931. It is interesting that a second book by Shewhart on this subject was published with a forward by Deming in 1939. In this work, Shewhart noted that processes could be considered to have common and special causes of variation. He pointed out that if only common causes of variation exist, the process operates in a predictable manner. The predicted level of operation might not be what the operator or management desired, but the system would consistently perform at level if nothing was done to change the system. On the other hand, if special causes of variation exist, the system would be unpredictable in performance and this is not desired. Thus, one of the first actions anyone has to take is to determine the special causes and to remove them, through process improvement or control techniques, and reduce the system to only common causes to ensure predictability. When a system is functioning with only common causes, the system is said to be in the state of statistical control, or to be statistically stable. In this language, stable means predictable.

Shewhart's definitions of control and common and special causes are the basis for any discussions of the analysis of process control data. The original definition of a special cause, an out of control point to Grant and Leavenworth (1980) and others, is a data point more than ± 3 standard deviations from the average of all the points. By the commonly accepted definitions, such a situation should occur about 17 times in 1000 by chance within a given population but 983 times because that data point was actually from a different population. Thus, it was recognized that some of the points thought to be out of control, more than ± 3 standard deviations from the mean or average, were in fact a chance occurrence. However, almost all of them were in fact something different and thus should be examined carefully.

More rules for out of control points, seven in a row above or below the average, six points in row steadily increasing or decreasing, etc, have been added. Walker, Philpot, and Clement (1991) refer to the Western Electric Handbook of 1956 for eight (8) other rules. Grznar, Booth, and Sebastian (1997B) note that Juran in his 1974 Quality Control Handbook used eight (8) points in a sequence to define a run rather than seven (7). Other rules quoted were two of three successive points at 2 standard deviations or beyond and four of five successive points at 1 standard deviation or beyond. The 1996 Fourth Edition of the LTV Steel "Policy Guidelines for Control Charts" also includes rules on trends, centerline hugging (stratification), non-random patterns, and a rule that deals with the number of runs above and below the center line. These references caution that while these additional tests do increase the ability of the system to detect small changes in the process, with increases in ability come increases in the chance of Type I or α error. Again, it must be noted that the rules being applied statistically fit IID situations and that autocorrelated data would have a tendency to create even more erroneous signals.

For comparison purposes, most authors stick to the point more than ± 3.0 standard deviationss from the mean as an out or control point and compare the new method to the Shewhart findings. Thus, the Shewhart method, for all the mathematical problems that may exist, is still the standard against which other methods are measured and evaluated. There may be problems with the assumptions originally suggested by Shewhart but the method quite often works on a day to day basis and is understood by practitioners at all levels. Any method that would seek to replace Shewhart would, if at all possible, need the same level of ease of use and understanding, along with at least the same perceived level of accuracy in determining problem areas in the data.

The major problem with Shewhart charts is the IID assumption. If that assumption does not hold, the basis for the accuracy of the charts, especially in determining out of control points may be severely compromised. Wardell, Moskowitz, and Plante (1994) specifically describe a number of industrial processes where incoming material causes serial correlation of the output. They note that traditional SPC methods are ineffective, inappropriate actually, for monitoring and improving quality of such processes.

Another criticism of the Shewhart charts is that they are slow to show small shifts in the process average. Two methods often used because of their efficiency in more quickly finding such small shifts are the CUSUM and EWMA charts. Interestingly, these charts do treat the data as a time series rather than IID. The CUSUM is an infinite length time series in that the difference from the long term average value of each point is added algebraically to the cumulative algebraic sum of all previous differences and then plotted.

The technique is somewhat cumbersome as it requires special techniques to determine when a point is out of control. The EWMA uses a weighted series of previous data points, along with the current data point, to determine the expected value of that point. The calculated value is then plotted against standard Shewhart control limits looking for a trend or out of control point. Thus, two methods commonly used to overcome a weakness in the Shewhart method are in fact time series methods and have been clearly shown and accepted by practitioners in the field as better than the conventional charts for detecting small shifts in the process average.

It is important to understand how a Shewhart control chart is constructed because the techniques for dealing with out of control points, special causes, are well defined, but often overlooked. When one begins the calculations for the control chart, all the values of the particular process control point, it could also be product data or any other data, are used in the calculations of the mean, standard deviation, and control limits, the ± 3.0 standard deviations level. Note that in general all Shewhart charts are made with the data in rational subgroups (discussed below), one of the assumptions that is difficult to prove.

Grant and Leavenworth (1980) state that rational subgroups should be selected such that each subgroup is as homogeneous as possible while providing the maximum opportunity for variation to occur between and among one subgroup and other such groups. The most homogeneous subgroup would obviously be that with a size of one (1). There are Shewhart techniques that deal with charts of individuals but such charts are normally considered difficult to use and understand because all the variation of the process is contained in one chart and thus makes interpretation of data difficult. In general, use is made of the statistical rule that the level of variation of group means is a function of the square root of the group size. Groups of size four (4), for example, are shown to have ½ of the variance in the mean as groups of size one (1). Thus, by using grouping, one can contract the range of average values under consideration. The variability of the process is converted to another chart called the range chart that is a chart of the differences between the largest and smallest value in each subgroup. The removal of some of the variability to another chart aids in interpretation of the charted data.

However, what is really a rational subgroup is still always open to question. On an industrial basis, many groups are generated at the level of five (5) because that is a size easily gathered. Likewise, some practitioners recommend a sample size of ten (10) so that the calculations are simplified to moving a decimal point to go from a sum to an average. If the group is too small, it is possible to increase the variation between and among groups. Likewise, if the group is too large, it will contain more variation than desired.

We want to be careful that we do not lose sight of why we are discussing this situation. Whether we have large or small subgroups, we are looking for situations where the data performs in an unexpected or strange manner. We are looking for changes in the process, either shifts to the process average or points some distance from that average that might belong to another distribution. As Grant and Leavenworth (1980) show, if the subgroups are too large, they can contain process shifts within themselves and thus hide such shifts from the observer. Also, if the subgroups are too small, they will influence the amount of variation shown for the process and this will also confuse the ability to determine shifts in the process average because the expected variation will generate such broad control limits that small shifts will not be observed.

Please note that very interestingly Grant and Leavenworth (1980) also state that the groups are taken from material produced in succession. It is also fully understood that Shewhart charts require the plotting of the data points in time sequence. Thus, each point follows in time, at least, the previous data point. I will discuss this point further a bit later in this chapter.

The limits are considered for the groups and not the individual data points. The concept of individuals vs. groups is one of process or product capability and that is beyond the discussions here. Grant and Leavenworth's "Statistical Quality Control" (1980) provides a discussion of this concept for the interested reader.

Once the average and control limit values are determined, they are plotted on the chart and then the individual data points which were used to determine these limits are plotted against the limits. If all the points fall within the limits, the process is considered to be statistically stable and the chart and limits are considered acceptable for use in the process control efforts. On the other hand, if the one or more points fall outside the control limits, such points are investigated for special causes, and if such special cause is found, such points are discarded from the calculations and the entire calculation set is redone with a new average determined and new control limits calculated. Then, the data points are again plotted against these new limits and a determination made if any are out of control. This comparison method with the discarding of data points determined to be special causes is continued until no more such cases are found. At that point, the process is judged to be statistically stable and the charts, and limits, usable for control of production. I would suggest that the joint estimation technique of Chen and Liu (1993), to be discussed in more detail later, is the application of this method to time series data and that that method works in exactly the same manner. Out of control points, commonly called outliers, are determined and excluded from the calculations on an iterative basis because a data point that is in control with broad limits may not be so with tighter limits constructed by excluding data points representing special cause situations.

It should be well understood that Shewhart charts are based upon the assumption of independent and identically distributed (IID) and that the distribution is normal about the mean when the process is in the state of statistical control. Further, the independence means that there is no particular pattern to the data. (Sebastian, Booth, and Hu, 1995)

Another concern with Shewhart charts is the necessity for large numbers of subgroups or data observations. Hillier (1969) notes that at least 25 subgroups are needed to establish a control chart, as a rule of thumb. If some of those points are in fact out of control due to special causes, one might wonder about the quality of chart eventually created. This is also a concern with modern manufacturing where the number of pieces made, or the time of production, is such as to generate only a small number of subgroups or data observations. He suggests using a two stage charting process where one takes into account the errors to be allowed, recognizing that multiple tests result in less error allowed for each test than the total, say 0.05%. He proposes taking a small number of initial lots or subgroups and, using modified control chart constants, develop an initial chart with limits. Then, as history grows, he suggests a recalculation of the charts' limits, he takes both X-bar and R-charts, with tightened constants for the control limits.

Traver (1985), on the other hand, suggests pre-control in such short run situations. He suggests setting up limits based upon target tolerance and then a set of procedures and rules of how to run the appropriate tests and to determine whether the process is operating satisfactorily or not. Wright (1997) was successful in detecting outliers or out or control points with very short length time-series using joint estimation techniques, which will be discussed later in this chapter. If one can adequately determine real limits for the charts, that would seem better than some artificial number that is useful but that needs modified, perhaps more often than desired.

2.3 Deming

While a first rate statistician in his own right, W. Edwards Deming's fame came from the application of 'simple tools' and techniques to real world production problems. He is credited with much of the improvement in Japanese manufacturing techniques in the 1950's and later which have resulted in the recognition of Japanese manufacturing to be among the best in the world in terms of quality and productivity. Among the very 'simple tools' proposed by Deming (1982) were the Shewhart control charts.

Deming's strength was in his understanding of the necessity to control the processes at the shop floor level and that management did not control the processes. The processes controlled themselves but could be forced into tighter control, less variation, by the people working on the manufacturing floor directly with the processes. Deming fully recognized that management did not even have any idea of what was happening with the processes, let alone know how to improve them. Deming and Japanese authors such as Isakawa (1983) understood the importance of statistical tools that were simple enough to be understood, and thus used, by the shop personnel, with some minimal level of training. The background thinking is that sophisticated statistical techniques, while excellent for problem solving use by personnel well schooled in such techniques, will not be readily accepted and used by shop floor personnel. Thus emphasis on 'simple methods' of cause and effect diagrams, also known as fishbone charts and Isakawa diagrams, Pareto charts, and the basic Shewhart control charting techniques. These methods have proven over time their value in industrial process and product control and problem solving. The cause and effect diagrams and Pareto charts are not mathematically based and will not be discussed further. The Shewhart control charts and the assumptions of them that may mean other methods are more accurate and timely in the identification of problems are the concern of these studies.

2.4 Time Series

It has been well recognized that many sets of process control data are actually time series rather than IID. It should be noted that two prominent control charting techniques, CUSUM and EWMA charts, are in fact time series charts because they take into account the previous data points. These two types of charts are considered particularly efficient in finding small shifts from the process average, as compared to conventional Shewhart charts. That is, the CUSUM and EWMA charts will detect small movements from the historic average more quickly than a Shewhart chart.

Jiang, Tsui, and Woodall (2000) have reported the use of an ARMA chart as a method of statistical process control monitoring. They indicate that an ARMA chart can outperform both the CUSUM and EWMA charts if the data are autocorrelated and if the model has the proper parameters. In the work reported, Jiang, et al, did not extend their work to situations where the ARMA (1,1) model was used as a first approximation of the time series model. Rather, Jiang, et al, calculated the exact parameters for the time series models they studied and some were specifically ARMA (1,1). Lu and Reynolds (1999) recognize that many processes are time series, rather than IID. Their efforts dealt with monitoring the mean and variance of such processes. Atienza, Tang, and Ang (1998) noted that when data are collected in rapid sequence, a common situation with computer controlled processes, the IID assumption is frequently violated and that autocorrelation of the data is rather common. They do bring up the point that the identification of outliers, AO, IO, or LS (TC is a special case of LS) can allow more rapid determination of the probable causes of process changes.

If we define a time series as a series of observations generated sequentially through time. Then, the data are ordered in time sequence, Shewhart control charts are similarly time ordered, and the subsequent observations are dependent upon the previous observation or observations. Shewhart assumes that each observation is independent of previous observations. If the observations are in fact a function of the previous observation or observations, clearly Shewhart's assumption is compromised.

Vasilopoulos and Stamboulis (1978) comment on the changes in the distributions of the mean and variance of the sample data when serial correlation exists. They proposed that more observations are necessary in order to reduce the variability to the same level as IID. For the average, the number of samples would have to be increased by the factor $(1 - \alpha)/(1 + \alpha)$ and for the variance the factor is $(1 - \alpha^2)/(1 + \alpha^2)$, where α is the parameter for the AR(1) model. As these authors suggest, for large levels of α, strong correlation with the previous data point, the need for increased sample size can be significant.

Time series can be of many types and they can be stationary or nonstationary. A stationary process is one that has a constant mean value and would describe a process considered to be in the state of statistical control by a Shewhart chart. Wardell, Moskowitz, and Plante (1992) note that many SPC systems are in fact stationary in practice. Thus, while it is possible to have an apparently stationary process that in fact consists of stratified data points, i.e. actually points from two or more different populations, such a situation would not be judged statistically stable by Shewhart's definition as all the points would be out of control from the expected overall average. Other rules for judging the presence of out of control points are noted elsewhere in this chapter and such stratified data would undoubtedly also be identified by those rules. The most powerful rule for examination of control charts is simply 'eyeballing' the data for patterns in the data that don't look 'random'. Such stratified data may also be determined by this method. On the other hand, nonstationary processes have a varying average, this variation can be either short or long term but is clearly noticeable, and would be considered a process that was not in the state of statistical control. Frequently, a time series that is nonstationary is trending upward or downward with time and that is the classic description of such a series. It is possible to deal with such situations by a method known as differencing which has the effect of reducing the trend, the actual goal is to eliminate it, and thus forcing the long term average to remain constant. After differencing, the data can be treated with the same techniques as a stationary process because the differencing technique removes the trend component from the time series. As we are interested in examining process control systems that are nominally in the state of statistical control, we will consider only stationary systems. It is suggested that future work could include the application of the proposed technique to nonstationary situations, but these are less common in Statistical Process Control.

Perhaps the question of how often does one encounter a time series as opposed to the IID approach of Shewhart was answered by Alwan and Roberts (1995) who examined 235 quality control applications and found 85% of them to be time series rather than IID. Alwan and Roberts stated in that 1995 paper: "An emperical study of 235 quality control applications suggests that violations of assumptions are the rule (85% observed) rather then exception in practice...leading to: (a) a false assurance that the process is stable, (b) a false search for special causes, (c) failure to search for special causes which can be seen with better analysis, (d) failure to see and act on systematic variation, such as trends, periodic, and autoregressive variation, and (e) control charts being ignored." They proposed that much of the problem was the result of misplaced control limits. As the primary test for stability is the lack of data points outside the control limits, it is very logical that incorrectly placed control limits will lead to incorrect interpretation of the data from the process.

Runger and Willemain (1995) note that one reason for correlated data is that the observation frequency in today's manufacturing environment has such a short time period between observations and thus there is insufficient time between observations to allow for independence. They also suggest that the time series model selected does not need to be exactly correct following Montgomery and Mastrangelo (1991) who recommended using the EWMA chart as the model and treating the residuals from that model. Comments on residual testing are found later in this chapter.

The question of how quickly observations of manufacturing processes occur is clearly a concern of Gong, Jwo, and Tang (1997) who discuss the use of on-line sensors for data collection. Faltin, Mastrangelo, Runger, and Ryan (1997) also note that autocorrelation is a concern with short sampling intervals. They note that such deviation from IID assumptions will lead to increased numbers of false alarms. There would seem to be little question that many on-line control devices determine and record/report process conditions on an almost continuous basis. One would surely not expect such observations to vary greatly from immediately prior or subsequent observations. Gong, et al (1997) suggest a two phase control system with both automatic instrument control and sampling and inspection.

Reynolds, Arnold, and Baik (1996) suggest the use of sampling plans that vary based upon the status of the system. If the process is operating within relatively narrow limits, there is no signal of a change in the process mean, sampling is done relatively infrequently. If there is sign of a process shift, testing is done much more frequently. Most SPC charts are set up for a given sample size over a set time period and are known as fixed sampling interval (FSI) charts. Reynolds, Arnold, and Baik (1996) suggest variable sampling interval (VSI) charts. One important consideration, however, is that as the frequency increases, i.e. short intervals between samples, the probability of correlation between samples will increase. Thus, one could possibly move from an IID to a correlated situation just by changing the frequency of sampling.

There are some models of time series that should be considered. Those of particular interest are the autoregressive model (AR), the moving average model (MA), the autoregressive moving average (ARMA), and the autoregressive integrated moving average model (ARIMA). By definition, the ARIMA model includes differencing and is applicable to data sets with trends or those lacking stationarity. As we are dealing only with stationary systems, we will exclude ARIMA as not necessary for this work.

Alwan and Roberts (1988) suggest that "...a few simple special cases of ARIMA models, such as the first-order integrated moving average process – ARIMA (0,1,1) – may surve as good approximations for many or even most practical applications. [The EWMA chart is based on ARIMA (0,1,1).]" We will comment further on this suggestion later. These authors also note that the major difference between ARIMA (0,1,1) and ARIMA (1,0,1) is stationarity. The former is nonstationary while the latter is stationary. Thus, the control limits of a stationary system are constant while those for a nonstationary system are continually increasing. If we are interested in control charts of statistically stable systems, they are stationary and the ARIMA (1,0,1) is appropriate. This model, as the differencing is zero (0), reduces to ARMA (1,1) as the model of interest. Wardell, Moskowitz, and Plante (1992) specifically state: "The ARMA (1,1) model was chosen because it is stationary, as many SPC systems are in practice, and because it contains both an autoregressive and a moving-average component; hence the effect of each parameter could be examined." Alwan and Radson (1992) note that an ARMA (1,1) model describes how the subsample means develop in a system that has an underlying AR (1) pattern. They also note that while for stationary systems, the ARMA (1,1) is often adequate, for a nonstationary system the use of an ARIMA (1,1,1) would seem appropriate. However, they find it better to use the ARIMA (0,1,1) as there is no intercept described with that model. While recognizing that nonstationary processes can exist deliberately in industry, the great majority of the processes being controlled have a desired stationary process. If the process is in fact not stationary, the process is not performing in the expected manner and will quickly move out of control. The ARMA (1,1) model has been shown by these authors, and others noted here, to find such out of control situations.

Alwan and Roberts (1988) conclude that "...precise model identification may not be essential to effective process control...". They also suggest that the ARIMA (1,0,1) and ARIMA (0,1,1) seem to offer reasonably good fits for many applications. The rationale for using a single simple model is clearly noted by Wright, Hu, and Booth (1999) in the work on short time series. Too many points are necessary to allow clear definition of the mathematically correct model to allow applications to short industrial series. Short time series tend to produce more errors in identification of the outliers.

The autoregressive model (AR) always carries a value after the letters, as in

AR (1), which means an autoregressive model of the first-order. This means that the current value is a direct function of the immediately previous value. If the model were an AR (2), the current value would be a function of the two immediately previous values. We would consider an AR (1) to exist when we can determine the current value of the time series, z_t, as a linear function of the previous value of the series and a random shock, a_t. Note that the use of AR (1), or any AR if the series is not stationary is not correct and will yield incorrect estimates. Also be aware that autocorrelation and outliers among the data points can lead to problems with model identification and use.

The AR model is noted to have a memory function that decreases slowly over time. This is logical if one considers that the current value is a function of the previous value, in AR (1), but that the previous value was a function of the value prior to it, and so forth. Thus, the AR model is characterized by a long term memory of previous shocks to the system.

The moving average model (MA) also carries nomenclature as to the number of previous data points included. An MA (1) model would include only the immediately previous data point. The MA model includes some measure of the previous errors in the determination or estimation of the current value. The use of an MA procedure is helpful in reducing the number of parameters to be estimated and is considered to be a parsimonious method in statistical terms. The memory of an MA model is different from the AR model because the MA model has a memory that has no effect beyond the values being considered.

The real power of AR and MA is the ability to combine them into ARMA, or an autoregresssive moving average model. While higher orders can be used, it is worth noting that and ARMA (1,1), first-order AR and first-order MA, deals very nicely with all levels or orders of both cases because the AR (1) is the equivalent of an infinite order MA while the MA (1) is the equivalent of an infinite order AR. By combining the two models into the ARMA (1,1) we gain the best of both methods, don't have to determine particular model types, which greatly increases the ease of application, and at the same time have a parsimonious model desired by statistical theory and practice. Note that a one-dependent system is most likely in SPC. The real key is the ability to adequately estimate time series without having to resort to complex computer systems to determine the probable exact model identification. We actually can not be completely confident of the models so identified due to potential severe problems with autocorrelation and partial autocorrelation of the data points.

Lucas and Saccucci (1990) suggest that using the correct model will lead to residuals that are IID with a mean of 0 and a variance of σ^2. They also state that as long as the residuals give no indication of an inadequate model, the model should be used. If the residuals appear to show a problem with the model, one should look for special or assignable causes and take some action on the causes if they exist. If no assignable causes are found, the model should be examined.

Faltin, Mastrangelo, Runger, and Ryan (1997) state that the presence of autocorrelation will increase the frequency of false alarms and other indications of problems with statistical control of a process. They suggest the use of an appropriate time series model with application of a conventional Shewhart control chart to the residuals in opposition to the broadening of the control limits on a conventional chart to avoid so many false alarms. They note this residual chart could accompany a conventional chart with the time series model predictions superimposed on it. Their suggested methods would seem difficult to use realistically in an industrial environment. False alarms are in fact data points identified as outliers of some type when in fact they are from the system under consideration and are thus statistically expected, if rare. The problem facing a user of control charts is how to distinguish true outliers from false alarms and what action must be taken when a potential outlier is identified and it is not known whether the data point is a true outlier or a false alarm. As identification of such observations can be expensive, the accuracy of identification is critical. The expense in identification comes from costs associated with such things as extra testing, possible shut down of the operation, problem solving time by trained personnel, and potential problems if some of the material being produced is so unsuitable as to cause problems in later manufacturing operations, internal or external to the manufacturer.

Crowder, Hawkins, Reynolds, and Yashchin (1997) suggest that anyone faced with autocorrelated data should first try to understand the cause of the autocorrelation. They caution that the conventional control charting techniques should not be used in such situations. They interestingly have a disagreement among themselves relative to the use of the residual method noted by Falten, Mastrangeol, Runger, and Ryan (1997). The question is whether hypothesis testing or estimation and engineering control are the point of interest. If one examines what is happening with a conventional control chart, the control limits are in fact the $\pm$ 3.0 standard deviations limit for the distribution in question. The graphical presentation of the population parameters, as estimated from the calculated statistics, gives a visual presentation of how far a data point is from the group average. In hypothesis testing of the comparison of a sample to a population, the question asked is how far from the mean of population is the mean of the sample, taking into account the variability of the measurements. A control chart is merely a graphical or visual demonstration of that distance. If the data point is more than the allowed distance, generally $\pm$ 3.0 standard deviations from the mean, it is classed as an outlier or from a different population. Everyone recognizes that there is a chance for error, 17 times in 1000 at $\pm$ 3.0 standard deviations, but one would classify the point as an outlier and reject the null hypothesis that the sample is from the population. Thus, a control chart is a hypothesis test with every point plotted(Box, Coleman, and Baxley, 1997). Alwan and Roberts (1988) also clearly state: "Checking for a state of statistical control is usually regarded as a test of a null hypothesis."

Futher, the existence of outliers in the data set also can significantly impact the determination of the exact model because they have an undesirable influence on the correlation calculations, the time series model calculations, and control chart parameters. It has been clearly seen by many practitioners in the field that outliers increase the average range value and thus the control limits, for example. By using one model that covers adequately the whole range of potential models, we escape the need to perform calculations that may not be correct but will be time consuming and not well understood. In order to determine the 'correct' time series model, for example, a larger number of data points are required and the calculations are lengthy. The very presence of outliers in this data can produce an incorrect estimate of the actual or 'correct' time series model. Thus, much time and energy can be used to calculate something that is mathematically incorrect and does adequately represent the data under study. The use of a single common approximation of the time series model, the ARMA (1,1) suggested here, alleviates the unnecessary and frequently incorrect calculations.

Luceno (1998) looks at a robust estimation technique for determining multiple outliers in a time series. He comments specifically: "Detecting the outlying observations is important becaue they can seriously affect the estimates of the model parameters and forecasts based on them." He also makes a most interesting observation that industrial time series should not have a large number of outliers because those are points the model does not explain. If one considers that true outliers are in fact from another distribution than the one under consideration, one would not want the model to explain such points. Rather, one wants the model to find such outliers. If there are a large fraction of the time series that are outliers, one would have to conclude the process under consideration is full of special causes rather than common causes, to use the Shewhart terminology. If this is the case, there should be no attempt to model the system until it can be made to consist of strictly common causes.

It is interesting to consider that the proposed generalized model is a first-order one, ARMA (1,1). Logically, as noted previously, an AR (1) is the equivalent of an infinite length MA, while an MA(1) is the equivalent of an infinite length AR. However, is there something else?

In regression analysis, it is fairly well agreed that first-order interactions are common, but that higher order interactions are very infrequent and very unexpected. Neter, Wasserman, and Kutner (1990) specifically state: "If three factor interactions are difficult to understand, higher –order interactions such as four-factor interactions are yet more abstruse. Fortunately, it is often found in practice that these higher-order interactions are quite small or nonexistent. When this is the case, they can be disregarded in the analysis of factor effects."

It is also interesting in multiple regression that the inclusion of too many independent variables will have a deleterious effect on the model being studied. Detailing a model too much as an explanatory method makes for a very poor predictive model. Very often, less is better in terms of a parsimonious model.

Alwan and Roberts (1988) note that a referee suggested the EWMA model as a good starting point and they note that the EWMA is an ARIMA (0,1,1). This was in regard to their suggestion that simple ARIMA (0,1,1) and ARIMA (1,0,1) models might reasonably deal with many applications.

For some reason, nature likes to keep it simple with minimal higher level interactions and this allows us to use relatively simple tools to estimate various systems, even if the relatively simple models are not completely correct. For a first approximation that is certainly adequate in many industrial applications, the simple model is more than sufficient, as long as it picks out the outliers or out of control points. Failure to find such outliers would mean the simple method is not sufficient. We will attempt to demonstrate that the simple method proposed, ARMA (1,1) is sufficient in finding outliers.

It would not be complete to discuss time series analysis without some attention given to intervention analysis. Chen, Liu, and Hudak (1992) defined intervention analysis as follows: "The essence of intervention analysis is to "isolate" the effect of an intervention from other occurrences and the underlying disturbances present in the series under study." We are thus looking for some additional terms to the model to take into account these situations where something outside of the system under study changes or greatly influences the output of the system. Thus, at a given time, the factor for the intervention is added into the calculation of the output, but prior to that time, and perhaps after than time, the factor has a value of zero (0).

An intervention is clearly an action taken to compensate for an outlier in the model being used to explain a time series and then to forecast from that model. Chen, Liu, and Hudak (1992) emphasize that the proper identification and classification of outliers are critical to intervention analysis to improve confidence that no outliers have been missed and that the proper interventions are shown. To properly conduct intervention analysis, one must properly identify and classify all outliers in order to construct the appropriate intervention model.

Intervention analysis is appropriate for forecasting but not for interpretation of control charts. The concept is introduced because it is another facet of time series analysis that makes use of outliers. The detection and proper classification of outliers are critical to proper use of intervention analysis as well as the interpretation of control charts in process or product control situations.

2.5 Outliers

Of major importance in the application of any method of analysis to process control data is the determination of outliers. An outlier is defined by Aczel (1996) as an extreme observation. The classic definition of an outlier by Shewhart is a point more than ± 3.0 standard deviations from the mean. As such a point occurs only 17 times in 1000 under conditions of normality, the chances that such a point actually belongs to the population under study is slim, but always possible. Outliers are the special causes Shewhart uses in his analysis of the process control situations.

As Shewhart defined the outliers, all efforts to find alternative methods compare themselves to those determined by classic Shewhart charts. However, if we have time series rather than IID data, outliers as determined by Shewhart methods may or may not be correct and the list of those found may not be complete. As pointed out by Chernick, Downing, and Pike (1982) in a time series the detection of outliers may be more difficult than with IID data as the outlier may not be a simple choice of an extremely large or small data observation. Further, Chernick, et al (1982) note that outliers may have 'dramatic effects' on various correlations, particularly in a short series. It is recognized that different levels of the value used to define an outlier, the σ levels, impact the denoting of a data point as an outlier. Tighter limits, standard deviations less than ± 3.0 such as ± 2.5, will lead to more points denoted at outliers by any of the methods suggested. However, when a point is found to be an outlier, on an industrial basis, the process must be closely examined, and possibly even shut down for proper investigation of that data point. Such situations can be costly and the necessity to avoid false alarms, determining a point as an outlier when it is not, must be avoided. Shewhart charts do lack sensitivity to some types of movements and several of the alternative methods have found additional data points that should be considered as outliers. However, much caution is noted in the proper selection of limits to be used in determination of outliers by all authors.

The Shewhart methods make use of rational subgroups. The subgroup, a small sample often on the order of 5 individual samples, is assumed to be IID with the other subgroups and to contain the variation typical of systems where only common causes exist. The control limits of such charts are set using the sample averages, subgroup averages, and use the standard error of the mean which is the standard deviation of the individual values divided by the square root of the sample size. Thus, if the standard error of the mean is ½ of the standard deviation of the individuals if the sample size is four (4). The chart of sample averages is accompanied by a chart of the range for each subgroup. Thus, a certain amount of the variability in the data is shown in the range chart and the amount of variation in the chart of subgroup averages is reduced.

An alternative to subgroups is the use of individual observations for the data points being analyzed. In this method, the individual values are plotted but the limits are set by calculating the difference in value between the data points. A common practice is a moving average of size two (2) where the difference between the current value and the value immediately preceding it are compared and the absolute difference calculated. This value is then used to calculate an average range and then the control limits for the range and the actual data point charts. This is the chart of individuals with a moving range chart. This method leads to control limits that are very broad and in fact such limits will often fail to identify out of control points or outliers when the traditional 3 standard deviation limits are applied.

If one considers the empirical rule, which states that 95% of the observations for a normally distributed data set will fall within approximately +/- 2.0 standard deviations of the mean, one can see that using +/- 3.0 standard deviations as the limit, where almost all of the data is included by the empirical rule, is probably not sufficient in recognizing changes, shifts, and outliers in the data. In fact, the amount of variation typically allowed in such +/- 3.0 standard deviation limit charts clouds the situation to the point that changes a practitioner expects to find are hidden or not observed. Another part of the problem is that the variability that is taken to the range portion of the chart using subgroups is still all contained in the chart dealing with the sample values. This induces more visible variation in the chart and makes the chart more difficult to interpret. Thus, we have a chart that is both capable of hiding the outliers from detection and at the same time showing so much variation that interpretation by other than standard control chart methods to be impossible.

In order to bring the chart into some condition where it can be better interpreted, a reduction in the limits to +/- 2.0 standard deviations is probably necessary, and even then the interpretation of the data is clouded by the variation in the chart of individuals. This is, of course, with the assumption of IID.

Another aspect of control charts that must be considered is the manner in which a control chart judges a data point to be out of control or an outlier. The control chart is historic in that it looks backwards. The question of whether a value is within expectations, not an outlier, is based on all the history used to calculate the control limits and sample average. Only after these calculations have been done is a comparison of individual values made to those limiting numbers. Likewise, new values are compared to the historic expectations with the assumption that the current value is independent of the preceding value. The process of calculating control limits does not take into account changing situations, until after a number of values have been found and one goes back to historical points to determine when the shift occurred, or when less variation was encountered and thus the limits tightened. Thus, conventional control charts tend to respond slowly to changes and to often be slow in identification of such changes. The methods used to improve change identification are charts of the EWMA and CUSUM type, which have been discussed previously and in fact are time series methods of data analysis.

Thus, in many ways, the use of charts of individuals with a moving range does not provide adequate identification of process changes and outliers. If the method is not capable of doing such things, the usefulness of such method is very questionable.

While an outlier may be simply an extreme observation, outliers come in different types and the identification of each specific type is important to fully understanding the process under study. Thus, outlier identification consists of both the classification of a data point as an outlier and the proper classification of that point as to outlier type.

One common definition of time series outliers is the use of four types. These are Innovational Outliers (IO), Additive Outliers (AO), Level Shifts (LS), and Temporary

Changes (TC). According to Chen and Liu (1993), the identification of outliers is critical because such outliers could cause an inappropriate model to be determined. Further, as is well known from basic statistics, the inclusion of outliers in calculations of least squares based sample statistics representing population parameters causes those statistics to be biased toward the outlier. This is one reason certain robust methods use the median rather than the mean in calculations of central tendency because the median is less influenced by the outlier than the mean. Chen and Liu (1993) also note that some outliers may not be identified and that will further confuse the determination of the proper model.

Luceno (1998) looks at outlier types as those existing as isolated outliers, outliers in continuous blocks, reallocation outliers, and scattered outliers. He uses reallocation outliers after Wu, Hosking, and Ravishanker (1993) as similar to the additive outliers listed above. Outliers in continuous blocks would appear to equate to the level shift outlier or temporary change, depending upon the length of it, and the isolated and scattered outliers would equate to the innovational outlier listed above.

The Innovational Outlier produces a temporary effect in a stationary series but it is seen for more than one observation. Atienza, Tang and Ang (1998) note that the IO affects the time series at a point in time, t, and then the effect fades exponentially after that point in time. On the other hand, an Additive Outlier, on the other hand, produces a change at only one point in time and then disappears. Recognize, however, that the use of ARMA (1,1) models does allow some impact of this one time action or happening to be seen in other projected values. A Level Shift is a change to the process that is both abrupt and permanent and would be recognized in the classic shift of the process average for which Shewhart charts are not well regarded when the shift is small. It is more commonly accepted that CUSUM and EWMA charts do a better job with small shifts of the process average. The fourth outlier type is the Temporary Change that is between the AO and the LS in time covered. The TC impacts more than a single data point as is the AO situation, but does have an end point, which is different from the continuing LS situation.

Chen and Liu (1993) are extremely conscious of the impact of outliers in their joint estimation method because the outliers can easily cause poor model identification. Even if the model is not critical, as we are proposing, the proper identification of outliers is critical to understanding and controlling a manufacturing process. Outliers must be identified properly in order to determine their source and thus possible corrective actions. Failure to identify an outlier prevents any reaction to it but improper identification can lead to incorrect reactions. Thus, the proposed generalized model must be shown to find all the outliers, remembering that Shewhart charts are sometimes not sufficient standards of detection, and to properly identify them. Wright (1997) has shown the joint estimation technique will deal well with outliers in very short series. The proposed method would have to perform along the same levels to be worthwhile.

There is a weakness in the outlier detection programs in that outliers contained in data points close to the end of the data set may not be identified properly. In general, such outliers are located and denoted but they are frequently misclassified because of the lack of data points after them. The problem is that the recognition of the type of outlier is related to points prior and then after the presence of the outlier. A TC for example may be an LS if sufficient time had not elapsed for the process to return to the original level. Likewise, an additive outlier and innovative outlier could be confused if close to the end of the data set.

Tsay (1987) used ARMA time series models to examine the types and effects of the different outliers. He defined the mathematics of the various types of outliers using the classic Box Jenkins terms. However, Tsay did specify the use of the appropriate ARMA model, rather than generalizing the approach.

Beckman and Cook (1983) presented a major literature review of outliers going back to comments by Bernoulli in 1777 dealing with the discarding of points that were discordant. The statistical practices developed along the lines of finding rationale for discarding such outlying points rather than identifying why they were different. Beckman and Cook (1983) note that the interest may in fact be in the discordant observation rather than in the population parameters being estimated. They note that in such cases the statistical problem really involves drawing inferences regarding the observation in question. Outliers may in fact be the most important points in the data set. They may provide the clues to why some things are very different from what is expected.

Beckman and Cook (1983) spend some time dealing with the masking and swamping of outliers. An outlier is masked when it is not detected because it, or another outlier, is used in estimating the parameter of interest. Logically, if we calculate the mean and variance of a data set and include all the data available, if there are outliers included we will have relatively large variance values found. If we have several outliers of varying degrees, the greater ones can easily mask the lesser outliers. If the outlier is not severe, it may in fact mask itself in the variance calculation and just seem to broaden the distribution, rather than in fact not being part of it. Swamping, on the other hand, has a point declared an outlier because it is included in a number of actual outliers. We all recognize that in testing a hypothesis we agree to accept some level of error in saying that the point is far enough away from the mean, taking into account the variance, to not be a member of the population under consideration. With an alpha level of 5%, we are wrong 5% of the time and such could be the situation with an individual point when judged with a group of other points that are in fact outliers, or a part of another population. Thus, it is critical that we examine each point individually to try to avoid masking and swamping.

Quesenberry (1986) emphasizes that the inclusion of an outlier in the calculation of parameter estimates will have the effect of modifying those estimates and thus make it difficult to clearly determine subsequent outliers or data trends. As we have noted above, not only is the inclusion of an outlier a serious concern, the proper identification of each one is also critical because of how they influence the model at future times. A Level Shift or Temporary Change has very different effects on the immediate subsequent events than do the Innovative or Additive Outliers. Proper identification of the outlier type can help prevent masking and swamping of other outliers.

Chang, Tiao, and Chen (1988) note that data sets often have "unexpected extraordinary observations" that can be the result of errors but can also be the result of some unexpected changes of the conditions of the system under study. These situations can cause some of the observations to become outliers. These authors speak of "nonrepetitive exogenous interventions" as the sorts of influences that will lead to outliers in a data set. In terms of single data points, this logic would fit other authors noted here. However, in the case of TC or LS outliers, these are not single data points.

The most important facet of the 1988 Chang, Tiao, and Chen paper is the introduction of the "Iterative Outlier Detection Procedure." The authors recognize that removing one, or even a group of identified outliers from the calculations of parameter estimates will change those estimates. Thus, once an outlier, or group of outliers, is determined and removed from the calculations, the new model has to be examined for additional outliers that may now be seen due to the change in model parameters. This action is akin to the standard Shewhart control chart practice where the center line and control limits are calculated, or perhaps better recalculated, after outliers are identified and removed from the initial data set. Each time the calculations are performed, the data that were used in those calculations must be examined for possible outliers due to the revised parameters. While the Shewhart charts are doing this with IID data, Chang, et al, (1988) are showing the application to time-series. They in fact used AR (1) and MA (1) models individually and also varied the level of variability allowed by setting different critical levels of the limits for determining an outlier. They examined standard deviation levels of 3.0, 3.5, and 4.0. They also used sample sizes varying from 50 to 150. They concluded that the ability to correctly detect outliers was a function of sample size and the level of σ used. Large samples helped to better identify outliers while larger values of σ clouded the issue. That is, if the critical value of the standard deviations limit is set at 4.0 vs. 3.0, less outliers will be detected, as would be logically expected.

Chen and Liu (1993) recognized that many time series contain outliers and that some method had to be developed to deal with them as the presence of such outliers impacts the accuracy of the model parameters being estimated. Chen and Liu (1993) proposed that the non-specific treatment probably accounts for other methods being very good in some cases and not too accurate in others. If the outlier is treated incorrectly, the model will not be accurate.

Outliers in control charts are determined as those points which fall beyond historic control limits. Thus, control charts look backwards in comparing the new point with what the system was doing when the limits were determined. On the other hand, in a time series, the process is forward looking. The model calculates where the next point should be and compares the actual value with the predicted value. Thus, an outlier is determined as a data point that is different than expected, and that does take into account shifts in the process average.

Interpretation of control charts for outliers is done according to rules and it can take some time for such rules to be invoked. For example, a frequently used rule is the 'run of eight' where a shift is considered to have occurred if eight points in a row are above or below the average. In a time series, the first point will be located at the time it is measured as an outlier. The determination of the type of outlier does take some additional data points for confirmation, but identification is rapid. One of the criticisms of conventional control charts has been the slow determination of shifts of the process average and EWMA and CUSUM charts have been proved as being more effective in locating such shifts. Clearly, those are time series control charts.

2.6 Past Efforts

There have been many efforts made to better explain process control data recognizing the limitations of Shewhart's assumptions. We will now examine a number of such reported methods because they are the bases for comparison of the proposed method. It is interesting to note that almost universally the highly sophisticated methods compare themselves to the Shewhart charts for the data being analyzed. The various methods will be covered individually.

Polynomial Smoothing

Sebastian, Booth and Hu (1995) used methods of polynomial smoothing and data bounding as nuclear materials accounting methods where they used time series models that smoothed the data to determine the underlying time series. They determined outliers and filtered out those and random errors and noise. These authors also reported applying the same methodology to the chemical process industry when Sebastian, Booth, and Hu (1994) discussed the application of the procedure to the Hussong Die Kettles of the standard Grant and Leavenworth data set. Sebastian, et al (1994) noted they found the same out of control points, outliers, as the standard method but also found several others that would not have been observed using a Shewhart chart.

In addition to the polynomial smoothing, these authors used data bounding to adjust data points lying beyond certain limits to smooth peaks and valleys. This method of identification of points some distance from the mean was also applied to outlier detection. The methods applied to the nuclear materials balances were shown to find and identify some instances as outliers that were not found by the joint estimation technique of Chen and Liu (1993), but to miss others. Polynomial smoothing and data bounding did find significantly more outliers than joint estimation. The level of false alarms might be questioned but was not addressed. However the method did pick up outliers earlier in the series than did the joint estimation techniques.

Residual Analysis

If one considers that the application of ARMA or ARIMA, or any other time series model is in fact the use of a model to explain, and then predict, the way in which a system operates, the rationale for examining the residuals is easily understood. In regression analysis, analysis of residuals is considered by Neter, Wasserman and Kutner (1990) to be "...a highly useful means of examining the aptness of a model." Likewise, Dielman (1991) in discussing residuals comments that "After estimating a sample regression equation it is highly recommended that some sort of analysis be conducted to assess the model assumptions." Thus, there is a basis in regression analysis to examine the residuals of the model and this has been extended to time series modeling.

A residual is defined by Neter, Wasserman, and Kutner (1990) as "...the difference between the observed value and the fitted value." Dielman (1991) refers to residuals as "sample disturbances." A residual is the difference between the actually observed value of the dependent variable at a given level of the independent variable, and the calculated value, or expected value, of that dependent variable based upon the model under consideration at a given level of the independent variable. The method of least squares, a common method of calculating the parameters of the model, minimizes the sum of the squares of such differences or distances for each data point in the data set used to calculate the model. It has long been recognized that actual data points located some distance from the expected value often influence the model described by the calculations. Thus, these influential cases (Neter, Wasserman, and Kutner, 1990) can cause the model to be incorrect and can be examined using various techniques such as Cook's Distance. If the data points are too far out of line, they are concluded to be from another distribution and are set aside for further analysis from the calculations and the balance of the data is then used to determine another model. Alternatively, robust regression can be used with the entire data set.

Alwan and Roberts (1988) looked at using residuals from a time series model to determine common and special causes involved in the system under consideration. They note that in the first order autoregressive model, AR (1), the current observation on the process can be considered the dependent variable and the previous observation the independent one. If such a time-series model fits the data, the residuals will be only random and only common causes would be expected. On the other hand, they note that by isolating departures from control as outliers, these can be considered special causes.

Alwan and Roberts (1988) propose the use of two control charts. The first is the "Common-Cause Chart (a chart of fitted values based on ARIMA models)." The second is the "Special-Cause Chart [or chart of residuals (or one-step prediction errors) from fitted ARIMA models]." These authors note that they are suggesting the use of greater statistical skills than the Shewhart charts, but that if the Shewhart charts are not correct, what choice does one have.

There have been a number of methods suggested to deal with these outliers or influential cases in time series models that will be discussed shortly.

Zhang (1998) notes that the residual charting approach is good with a nonstationary process, even autocorrelated data. However, with a stationary process, the residual charts do not have good detection capability for small shifts in the process mean He suggests an alternative, the EWMAST chart. For this chart, Zhang calculates the standard deviation taking into account the autocorrelations of the samples. If the EWMAST chart goes out of control, the process average has shifted and needs to be recalculated. He concludes that the EWMAST chart works better than the X chart when the shifts to the process mean are $\leq 2\sigma_X$. Additional comments on CUSUM and EWMA charts are to be found later in this chapter.

Montgomery and Mastrangelo (1991) note that unless the ARIMA model is of some use in explaining the process, the application of the residual technique will often be of more effort than it is worth. They suggest that the technique to be applied should be the EWMA chart.

Grzner, Booth, and Sebastian (1997B) used a robust smoothing technique with the expectation that residuals would increase in size and thus outliers would be easier to detect and identify. The authors considered smoothing a technique that would filter out random noise and other data irregularities. They assumed that an in control process would generate residuals close to zero while large residuals would be identified as outliers. Any moving average process will smooth out the curve.

The Grzner, Booth, and Sebastian (1997B) work does point out that there are two ways a process is judged out of control. In the first instance, the presence of an outlier is a defacto out of control situation because the outlier comes from another distribution. The second instance is a bit subtler in that the process model can forecast that with the next unit produced, the process will be out of control. They note that many of the common rules applied to control charts deal with anticipated loss of control rather than outliers.

Grznar, Booth, and Sebastian (1997B), in further work using robust smoothing looked at Running Median Smoothing for equally spaced or almost equally spaced data points in a time series. Points closer together give a smoother curve than those some distance apart. As in the earlier work, the idea was to generate a robust method that would allow outliers to be more easily detected and identified because such points would appear as an aberration to the smooth curve developed by this method. These authors clearly defined four criteria against which any suggested method should be judged. First, the method must provide early detection of outliers or other situations where the process is out of control. It is emphasized that the earlier, more quickly, such detection is made, the better. Second, any new method must minimize Type I errors – false alarms. Incorrectly set control limits, for example, would impact the false alarm rate and using an incorrect model would lead to such errors. Third, the new method must result in less Type II errors – failing to detect when the process has changed. And, fourth, the new method must have the ability to define the type of outlier involved in the problem.

Joint Estimation Technique

Chen and Liu (1993) recognized the problems with outliers and the impact such extreme data points have on model identification. Their point was that the identification and choice of an improper model would adversely impact the decision making process. Their method, as noted previously, is the application of standard control charting techniques to time series. The complexity of what they did is far beyond the classic control chart method of determining outliers during the initial calculation of control limits, however. Chen and Liu (1993) noted that most authors had suggested merely recognizing the outliers and then accommodating them somehow. The suggested method identifies those outliers and by the use of iterative calculation rounds eliminates the identified outliers from the calculations of the model. This is done until no more outliers are identified against the level determined by the researcher. Many researchers would use levels of standard deviations on the order of 2.5 to 3.0 in order to more fully understand the process. One wants to find all the outliers but to avoid false alarms. If the level is too tight, more false alarms will be found, but a better understanding of the process may also be gained.

Please note that the false alarms level is in fact a very qualitative measure. If a process is so critical to a manufacturing operation that any output not well within specified limits is not acceptable, closer control is necessary and that means tighter limits than normally considered and consequently more false alarms. However, such practices do minimize the inclusion of less than desired product in the output of the process. Part of this issue is also a process capability issue as improper material is found on an individual basis rather than as a group. Thus, different outlier detection limits would be used based upon the process capability and the output requirements. We are looking for a method that will maximize finding problems at any limit level, but still not sending false alarms if nothing is wrong.

Prasad, Booth, Hu, and Deligonul (1995A) used this joint estimation technique on the classic 'Sheet-Like Process' from Johnson and Bagshaw (1974) and found the same outliers as a CUSUM chart but also found some others. Unlike the CUSUM chart, the joint estimation technique identified the type of outliers present. For the study of transmissions, the joint estimation technique was noted to have found one outlier that the CUSUM method did not. Both methods found one other outlier. Likewise, the study of bore holes showed the joint estimation technique to be superior to CUSUM in outlier identification.

Chen, Liu, and Hudak (1992) provide the exact instructions for performing this joint estimation technique and examples of what pattern residuals will take for different outlier types when analyzed using the Scientific Computing Associates Corp. (SCA) package. It is important to recognize that different outlier types do in fact generate different residual patterns and that this pattern is the basis for detection of such outliers using Joint Estimation. As noted previously, this is an iterative method and does require the setting of the critical limits, standard deviation limits, beyond which a data point will be considered an outlier and be excluded from the parameter estimation process. The residual patterns clearly show how the innovative outlier, IO, impacts only one data point while the additive outlier, AO, impacts some limited number of subsequent data points.

In the case of temporary changes, a number of data points are impacted to differing degrees while the level shift, LS, type of outlier clearly shows a new level for the process average, based upon residual levels.

Prasad, Booth, Hu, and Deligonul (1995B) applied the joint estimation techniques to nuclear material losses and reported the method to be superior to standard control charts, CUSUM, ARMA control charts, ARMA CUSUM, and the Generalized M procedure in detecting such losses.

CUSUM and EWMA Charts

Two methods used to deal with autocorrelation and at the same time have better response to small shifts in the process mean, as compared to X charts, are the Cumulative Sum (CUSUM) and the Exponentially Weighted Moving Average (EWMA). (Montgomery, 1991) Note that each of these charts is in fact a time series chart because each of them takes into account the previous data. The EWMA is a moving average (MA) by definition. The level of involvement of the previous data point or points is a function of the system being considered. The CUSUM is the algebraic sum of the differences between each data point and the long-term group average. If there is shift in the process average, the CUSUM chart will move out of control. However, the determination of control involves a particular technique using a mask. Montogomery (1991, p 283) describes this mask technique.

Johnson and Bagshaw (1974) concluded that the CUSUM test is not robust when autocorrelation, lack of independence, is present in the data set. They emphasize that in production processes items sequentially produced may well not be independent. One measure of the ability of a charting technique to perform, in addition to the determination of shifts in the process mean, is the Average Run Length (ARL). These authors note that with lack of independence, the CUSUM ARL decreases in length markedly. This means that additional false alarms are encountered. The deleterious effect of false alarms has been previously discussed as time consuming and costly. The authors conclude with the comments, and an example, that demonstrates that the importance of understanding the underlying correlation structure of the system prior to determining the appropriate control system to implement.

Montgomery and Mastrangelo (1991) have also commented that autocorrelation, when independence is assumed, causes a higher frequency of false alarms. They suggest the use of the EWMA chart, certainly a time-series model by definition, to approximate the process. They note that it is often very time consuming and "awkward" in the SPC environment to seek the absolute model. Rather, they suggest the EWMA method, with residual analysis. They are specific that the residuals are not correlated and as IID data can be properly treated with the Shewhart control charting methods.

Wardell, Moskowitz, and Plante (1992) compared EWMA charts to ARMA(1,1) models using Special Cause and Common Cause Charts to work with the residuals in a Shewhart method. While they note that often better performance can be obtained from Shewhart charts, in terms of detection of shifts of the mean, by simply increasing the sample size, they also note that if autocorrelation exists, there should be more evidence of runs in the Shewhart charts than if IID conditions hold. They suggest that the ARMA(1,1) model is of individuals rather than groups while Shewhart is usually grouped data. The authors expect the same results with groups as with individuals although they note that the variance observed is now of the subgroup means and not of the process mean.

Roberts (1959) described EWMA charts as Geometric Moving Average Charts as the most recent data point has the greatest weight with the weight decreasing for observations further back in a geometric progression. Note, however, that the EWMA always includes all the data observations ever made, as there is a tie back to the first data point from each subsequent value determined. The author also points out that the control limits for the EWMA chart are ½ of those for the standard control chart.

Lucas and Saccucci (1990) note that EWMA charting methods can be designed to detect small shifts of the mean much more quickly than Shewhart charts. However, Shewhart charts are "superior" in detecting large shifts of the process average than the EWMA method. These authors concluded that the two methods, CUSUM and EWMA are very close in their performance. Woodall and Maragah (1990) in a discussion of the Lucas and Saccucci paper noted that the mask used for CUSUM charts ignores many past observations and concentrates only on the most current ones. Woodall and Maragah (1990) believe the EWMA chart to be much easier to use than the CUSUM for the same level of results, both determination of process shifts and the Average Run Length (ARL) as a measure of false alarms.

An interesting aspect of the Lucas and Saccucci (1990) proposal is the use of fast initial response (FIR) control limits for EWMA charts. One criticism of control charts is the large number of data points necessary to construct a chart. The idea of being able to estimate very early control limits is beneficial to industrial practices, especially as the economic production quantities are continually forced to shorter and shorter production runs.

Yashchin (1993) looked at the use of data transformations of correlated data and then a CUSUM charting method of that data. He notes that such transformations do not eliminate serial correlation in most cases. Rather, the transformations reduce the magnitude of the correlation. Guerrero (1993) states that transformations also have the problem of converting back to the original data form. He notes that such transformed series do not retain the optimal properties when brought back to the original data scheme but rather the estimated mean becomes an estimated median after the application of the inverse transformation.

The Lu and Reynolds (1999) paper "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes" note that the EWMA method is most useful in dealing with the mean of the process. They do note that the determination of the appropriate parameters of the model may be difficult to estimate accurately because of lack of independence of the data points and because the appropriate model may not be clear. They also make the point that the EWMA chart is used to determine small shifts in the process average.

Zhang (1998) looked at four different charting methods and found the EWMAST method the best for determining small shifts in the process mean when the autocorrelation is not particularly strong. He recommends residual charts when the process has strong positive autocorrelation but notes that the residual chart does not well identify small shifts in the process mean. Zhang does note that the use of either an EWMAST or EWMA chart does not require time series modeling, but that one would need to implement a time series modeling method if the residual method is used.

Neural Networks

Hamburg, Booth, and Weinroth (1996) used a neural network approach with an AR(1) model to treat data. The AR(1) was selected because much of the data does in fact follow this model. Neural networks are powerful computing tools but the algorithms they follow are determined by the programmers. Thus, if one determines that the network will use an AR(1) approach to examine the data, it is that model that will be used. It would seem that one could apply the generalized model suggested ARMA(1,1) easily via a neural network and achieve the same type of outlier identification as found using the joint estimation technique discussed above.

While neural networks do give some indication of being very useful tools in statistical process control, they are still so new that they will not be considered further in this chapter. We will now go on to consider in detail the research methods to be used in this dissertation.

CHAPTER 3

Methodology

3.1 Introduction

The purpose of this research is to determine if one simple model, ARMA (1,1) can in fact be used to approximate many different time series models of various processes. In order to adequately approximate the mathematically best time series model, the proposed simple model has to sufficiently accurately represent the actual happenings in the process. Thus, the ARMA (1,1) model has to locate and discriminate between types of outliers. In process control, it is the outliers that are of interest. If the process is continuing to function in the expected manner, no outliers are present in the data. On the other hand, if the process is changing through internal or external actions upon it, outliers can be expected as signs of such changes. Further, as proper practice is to spend time examining any outliers for cause, the identification of a point as an outlier when that point is in fact not an outlier is a false alarm and can be costly on an industrial basis. Thus, the proposed method must first find the outliers, and not identify points that are not outliers, and then the proposed system must properly label those outliers as an important part of the determination of the cause for such outliers.

Chen and Liu (1993) noted four forms of outliers as innovative outliers (IO), additive outliers (AO), level shift outliers (LS), and temporary change outliers (TC). They defined these outliers mathematically as follows:

$$Y_t = \{\theta(B)/[\alpha(B)\phi(B)]\}a_t, \qquad t = 1, 2, 3, \ldots n$$

where n is the number of observations for the series, $\theta(B)$, $\phi(B)$, and $\alpha(B)$ are polynomials of B, the backshift operator, ($\alpha(B)$ is the nonstationary operator) and the following model of a time series influenced by a nonrepeating event is suggested:

$$Y_t^* = Y_t + \{\varpi\{A(B)/[G(B)H(B)]\}I_t(t_1)$$

where Y_t is the original series and $I_t(t_1) = 1$ if $t = t_1$, and 0 otherwise. $I_t(t_1)$ is a function for the occurrence of the outlier impact. The t_1 is the location of the outlier, and this is possibly unknown. The parameter ϖ and $A(B)/\{G(B)H(B)\}$ denote the magnitude and dynamic pattern of the outlier effect. Under these qualifications, the definition of each outlier type is as follows:

IO: $\{A(B)/[G(B)H(B)]\} = \{\theta(B)/[\alpha(B)\phi(B)]\}$

AO: $\{A(B)/[G(B)H(B)]\} = 1$

TC: $\{A(B)/[G(B)H(B)]\} = [1/(1 - \delta B)]$

LS: $\{A(B)/[G(B)H(B)]\} = [1/(1 - B)]$

From these equations we can see that if the dampening or decay rate, δ, is zero (0), a temporary change becomes an additive outlier and if $\delta = 1$, the temporary change does not go away and a level shift results. Chen and Liu (1993) suggest that the AO and the LS are the two boundary cases for a temporary change (TC). They also note that only the innovative outlier (IO) is dependent upon the model. It is important to understand this concept because there is an impact of the model on the identification of innovative outliers but not on the other three types. Likewise, the exact identification of other three types may not be absolutely critical to the process. If we can identify the outlier as existing as either an AO, TC, or LS, we can recognize that the presence of the outlier is the critical issue but that the identification by type is a function of the calculated decay rate by the model being studied.

This concept of decay rate is critical because the combination of the AR (1) and the MA (1) models provides a model that is different from either of the original ones. An AR (1) is actually equivalent to an infinite length MA and an MA (1) is equivalent to an infinite length AR. And, while the MA (1) model does not have a memory the AR (1) does. By definition, the MA (1) model uses only one data value in the calculation and thus there is no effect of previous data values. On the other hand, the AR (1) model includes some effect of all previous data values in the current value being calculated or forecasted. We might also use the notation that an AR (1) = ARMA (1,0) and an MA (1) = ARMA (0,1). We will use the combined ARMA (1,1) for this work because it combines characteristics from both individual models.

If one examines, the responses to an action taken upon a system, one can see that the response is somewhat different for the AR vs. the MA model. In many ways, the two models react just opposite to each other. We can see that the *acf* (autocorrelation function) of an MA process behaves like the *pacf* (partial autocorrelation function) of an AR process; and the *pacf* of an MA process behaves like the *acf* of an AR process. Thus, where one expects the impact of an action to decay quickly, the decay can take a longer time than anticipated and this can result in the identification of different outliers than would be anticipated because the model being used is not exact to the data set. The question we are asking is: "Is the assumed model, ARMA (1,1), a good enough approximation to make up for this short coming."

Alwan and Roberts (1988) have suggested that an ARIMA (1,0,1) or ARIMA (0,1,1) might be good choices as basic models for general application. Part of the basis for that suggestion is the generally recognized success of EWMA models in determining small shifts in process averages. Such a shift would be a Level Shift (LS) and determined as an outlier by the time series analysis of Chen and Liu (1993). The first model has a differencing value of zero (0), indicating a stationary process. The second model has a differencing of one (1), a nonstationary process, but no value for the AR portion of the model. Thus, the second suggested model is a moving average MA (1) with a difference of one (1) time period. In reality, the second model is an EWMA model with the current factor being influenced only by the immediately preceding data point. This method does take into account trends very nicely but we have argued previously that stable process systems are stationary. If the process departs from the expected average, that should be detected as an outlier. The type of outlier might be open to question depending upon the degree and length of the excursion, but such a situation would clearly be understood to be an outlier. Please note that it is well understood that some variation does exist in every process and that we are talking about variations beyond the norm as being outliers.

If the ARIMA (1,1,1) is discounted for this work because we argue that a process must be stationary to be considered statistically stable and in the state of control, that leaves us with the suggestion of the ARIMA (1,0,1) as a general form. As the difference is zero (0), this does become an ARMA (1,1) in reality. Thus, for stationary systems, or those assumed to be stationary because we will test for that property, the development leads to the use of ARMA (1,1) as the general model to be tested for effectiveness and efficiency. Further, this is reasonable since in SPC situations a one-dependent process is reasonable, Chernick, Downing, and Pike (1982).

3.2 Methodology

We will test the hypothesis of this dissertation that one simple model, the ARMA (1,1), can be used to adequately approximate many other time series models present in process data in terms of the location and identification of outliers in the data. We will use the joint estimation technique of Chen and Liu (1993) in the work. We will consider first some well known data sets. Such data sets have been well studied by various practitioners and reported in the literature. For example, the work of Prasad, Booth, Hu, and Deligonul (1995A) using the joint estimation technique on the classic 'Sheet-Like Process' from Johnson and Bagshaw (1974) provides a study with results showing outliers and the identification of them. Other work by Prasad, Booth, Hu, and Deligonul (1995B) on nuclear material losses also, using joint estimation, could provide another basis against which to judge the effectiveness and efficiency of the proposed model.

For the purposes of this work, effectiveness is defined as the ability to find the outliers that are present and to adequately identify them. Likewise, efficiency is defined as the ratio of outliers located compared to those located using the 'correct' time series model. It should be understood that while we have the ability to mathematically calculate and determine the underlying time series model for a data set, the presence of outliers can influence the mathematics and provide an incorrect identification in terms of reality. Thus, we do have to recognize that the data itself has a very important impact on the determination of the model thought to exist. That is also an interesting argument for the value of using one simple model because that model is known and is not being incorrectly identified due to data considerations.

Recognizing that there is a weakness in our determination of efficiency, due primarily to a lack of accuracy in the original model determined, by using various methods as a basis, we will try to provide as reasonable an estimate of the efficiency of the proposed method as possible.

Once the method has been adequately tested on known data sets, we will apply it to some new industrial data that have become available. These data have not been examined by other than conventional Shewhart methods, In that case, we will compare the results of the suggested model to both the Shewhart charts and to a time series model as suggested by the Scientific Computing Associates Statistical package.

There is often much discussion as to the level of difference between existing group averages and outliers. Should one consider as an outlier only those data points 3.0 or more standard deviation units from the average or is some lesser level of difference a better choice. The conventional ± 3.0 standard deviations distance minimizes false alarms but does require a substantial difference in the data values for a point to become an outlier. On the other hand, the use of ±2.0 standard deviations, while sure to identify more data points as outliers will also provide many more false alarms. From basic statistics we recognize that there are many points between 2.0 and 3.0 standard deviations that belong to the in control distribution. There are many less such data points more than 3.0 standard deviations from the group average.

It should be understood that the classic 3.0 standard deviations level was chosen by Shewhart (1933) for grouped data. We recognize that the variance in grouped data, as opposed to individual data points, is reduced by the square root of the size of the group. Thus, while a group may be at the limits of variability due to group variance, the individual data points could be well within specification limits. That is in fact one of the powerful things that the Shewhart charts of group data does – it allows the determination of a concern before the process produces material that is out of specification, under the usual circumstances where the process is considered statistically stable and statistically capable. By capable we mean that the product produced at ±3.0 standard deviations from the group average falls within the specification limits.

Process capability is a function of the individual values while the process is said to be in the state of statistical control when the sample averages operate within the control limits, when dealing with small groups in the conventional Shewhart charts. Thus, the limits for the sample averages are affected by the sample size. If the sample size is only four (4), the standard deviation for the distribution of sample averages is only 50% of that of the individual values. When charts of individuals are used, the actual data values are plotted, not the averages of the subgroups. The limits for such charts are determined by averaging the differences between adjacent individual values.

However, when we are dealing with individuals, that grouping concept does not protect us. We need to consider a more conservative limit. Various authors have suggested $\pm$2.0 standard deviations or $\pm$ 2.5 standard deviations. We will examine each level, as well as $\pm$3.0 standard deviations as a control, in each base model studied.

Algorithim

The testing process consisted of treating the individual data points with a program from Quantum Improvement. This program is an attachment to Microsoft Excel and generates control charts of various types. Control Charts of Individuals with a moving range of 2 were generated for each data set and the number of out of control points were determined for both the Chart of Individuals and the Moving Range Chart at the +/- 2 standard deviation level.

The data were then entered into the Scientific Computing Associates (SCA) Statistical System to be treated as a time series. The computer was first asked to model the data into an ARMA (1,1) time series and then to use this model as a forecasting tool to determine the expected value of the data for each successive data point. The actual and forecasted values were compared at the +/- 2.0 standard deviation level to determine whether the actual value fell within expectations. If the measured value met expectations, the process was repeated for the next observed value. If the observed or measured value did not meet expectations, the value was tested using the definitions for outliers noted previously, and the type of outlier was determined. The mathematics of this follows.

The model of the ARMA (p,q) per SCA is:

$$Z_t = \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \ldots - \phi_p Z_{t-p} = C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \quad - \theta_p a_{t-q}$$

where $\{a_t\}$ is a sequence of random errors that are independently and identically distributed with a normal distribution. The backshift operator, B, can be introduced:

$$BZ_t = Z_{t-1}; \quad B^2Zt = B(Z_t) = Z_{t-2}, \text{ etc.}$$

The model then becomes:

$$Z_t - \phi_1 BZ_t - \phi_2 B^2 Z_t - \ldots - \phi_p B^p Z_t = C + a_t - \theta_1 Ba_t - \theta_2 B^2 a_t - \ldots - \theta_q B^q a_t$$

which can be reduced to:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)Z_t = C + (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t, \text{ or}$$

$$\phi(B)Z_t = C + \theta(B)a_t, \text{ when}$$

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p), \text{ and}$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$$

The ARMA (p,q) model has p as the order of the autoregressive operator $\phi(B)$ and q as the order of the moving average operator $\theta(B)$. One can also represent the ARMA (p,q) model as:

$$Z_t = \mu + [\theta(B)/\phi(B)]a_t$$

where μ is the mean of the stationary time series.

Returning to the Chen and Liu (1993) equations,

let $\pi(B) = \{\phi(B)\alpha(B)/\theta(B))] = 1 - \pi_1 B - \pi_2 B - \ldots,$

and the error term, $e_t = \pi(B)Y_t^*$,

then the four outlier types become:

IO: $e_t = \varpi I_t(t_1) + a_t$,

AO: $e_t = \varpi\pi(B)I_t(t_1) + a_t$,

TC: $e_t = \varpi\{\pi(B)/(1-\delta B)\}I_t(t_1) + a_t$,

LS: $e_\tau = \varpi\{\pi(B)/(1 - B)\}I_t(t_1) + a_t$.

We can now determine the type of outlier based on the error or residual term. We set the limit at +/- 2.0 standard deviations, derived from the variance calculated from the residuals determined from the model and parameter estimates, as an error larger than allowed to compare with +/- 2.0 standard deviations limits set for the control chart of individuals with a moving range. The SCA program reports both the occurrence of the outlier as well as the identification of it.

3.3 Preliminary Results

In an effort to understand the work being proposed, we have taken some conventional data sets and estimated the best time series model using the SCA package and then treated the data using the suggested ARMA(1,1) model. We have looked for outliers at 2.0, 2.5, and 3.0 standard deviations levels. All the data sets were of individuals rather than groups. The results of those efforts follow.

An AR(1) Data Set

A file of Yields from a Batch Chemical Process data shown as Series F in "Time Series Analysis, Forecasting and Control," Third Edition, by Box, Jenkins, and Reinsel (1994) was treated using the SCA package which makes use of the Chen and Liu Joint

Estimation Technique to determine the model type and the location and type of outliers present under that model. The pertinent output is found in the Appendix on page 136 under the title of "Chemyld Data." The data set has 70 data points and was determined to be an AR (1) time series using the SCA system of analysis. Outliers were noted at 2.0 and 2.5 standard deviations but none at 3.0 standard deviations. The outlier types are shown in the Appendix on page 136 as noted. The data set was then examined using the proposed ARMA (1,1) model and various outliers were detected at 2.0 and 2.5 standard deviations but none at 3.0 standard deviations. In examining the output, as shown on page 135 in the Appendix, we can see that many more outliers are found for 2.0 as compared to 2.5 standard deviations. While it might appear that we would see too many false alarms at the 2.0 level, as we do not know what is a true outlier and what is not, we will use the 2.0 standard deviations limits for our studies. As has been noted earlier, the 2.0 standard deviation level for individuals is still a conservative level when compared to the actual value found if one would be using groups as small as four. At that level, we find eight (8) outliers from the AR (1) and the ARMA (1,1) models. Six (6) of the eight (8) are the same in both location and identification as to outlier type, IO, AO, etc. Of the other two (2) the ARMA (1,1) identifies as a TC (temporary change) point 17 that is noted as an AO (additive outlier) by the AR (1) model. Interestingly, the AR (1) model sees point 20 as the TC but this point is not considered an outlier by the ARMA (1,1) model because it sees the change as already haven taken place. Also, the ARMA (1,1) model finds and IO, innovative outlier, at observation 19 but no comparable outlier is noted by the AR (1). It appears that in this case, the two models perform fairly equivalently with the ARMA (1,1) anticipating or discovering somewhat more quickly a temporary change in the process. The ARMA (1,1) appears to be a reasonable approximation of the correct AR (1) model.

A 1/AR (1) Data Set

The well-known Sheet-Like Material data set from Johnson and Bagshaw (1974) was examined using the techniques noted above. The original model was identified as a 1/AR (1) by the SCA method of analysis of the data set for the type of time series, although we recognize that the mere presence of outliers can cause problems with identification of the type of time series. However, based on the calculations performed, this model is an inverse of the typical AR (1) model. If one plotted the response variable, the shape of the curve would be the opposite of an AR (1) model. I examined the outliers treating the data as the 1/AR (1), an AR (1), and the proposed ARMA (1,1). The output from those efforts is shown in Appendix on page 137 under the title of "Sheet-Like Material Data." This data set has 100 observations. The data were examined under three time series models – the AR (1), 1/AR (1), and ARMA (1,1) with the outliers again examined at 2.0, 2.5 and 3.0 standard deviations.

The models all found a significant number of outliers at 2.0 and 2.5 standard deviations and some at 3.0 standard deviations as well. Very good agreement is seen at 3.0 standard deviations among the three models used. At the 2.0 standard deviations limit, the AR (1) model found 29 outliers as compared to 16 for the 1/AR (1) model, and 27 for the ARMA (1,1). Of those outliers identified, only 2 of those found by the 1/AR (1) model, were not located exactly or as immediately prior or subsequent data points by the AR (1) and only 3 were missed by the ARMA (1,1) model. On other hand, both the AR (1) and ARMA (1,1) identified many additional areas of interest. There are some slight differences in the outlier type identified and I would refer the reader to the discussion at the beginning of this chapter as to the mathematical differences among AO, IO, and TC type outliers. All three of the models do identify an outlier early in the data set at observation 6, using the 2.0 standard deviations limit, although the type of outlier does vary by model used. On the whole, the three models agree fairly well on the outliers detected. There are some differences in the identification of the outlier in some instances and it appears that the decay rate as noted in the mathematical explanation earlier in the chapter is significant. However, the model does an adequate job of approximating the originally suggested model. It appears that the proposed model works adequately for this 1/AR (1) model as well as for the AR (1) type discussed previously.

Prasad, Booth, Hu, and Deligonul (1995A) examined this data set using robust and semi-robust time series modeling and found outliers at observations 29, 65, 69, and 83, apparently at the 3.0 standard deviations level. All these observations were likewise noted as outliers by the ARMA (1,1) model, as well as the AR (1) and 1/AR (1) models. Their identification of the outlier type was also consistent with the ARMA (1,1) and the other models used in this work.

An MA (1) Data Set

A data set of observations of the motor cortex neuron interspike intervals for an unstimulated monkey from “A Handbook of Small Data Sets” edited by Hand, Daly, Lunn, McConway, and Ostrowski (1994) was identified using the SCA program as a

MA(1) time series. The pertinent output is found in the Appendix on page 139, titled "Monkey Neuron Interspike Data." There are 100 observations in this data set. The data set was again treated as the MA (1) time series and as the proposed ARMA (1,1) time series looking for outliers and outlier types. Again, the data set was treated at 2.0, 2.5 and 3.0 standard deviations.

At the 2.0 standard deviations level, the MA (1) model found twenty (20) outliers while the ARMA (1,1) model found only ten (10) with seven (7) of them exact matches for location or either immediately prior or following the location identified with the MA (1) model. Nine (9) data points noted as Innovative Outliers (IO) by the MA (1) model early in the data set were missed by the ARMA (1,1) model. It is interesting to observe that the missed outliers were consecutive data points and that the ARMA (1,1) model did locate an outlier immediately prior to these strings of outliers. The ARMA (1,1) notes an Additive Outlier (AO) as the observation immediately prior to a reported Level Shift (LS) noted by the MA (1) model at the 2.0 standard deviations limit. The ARMA (1,1) model finds a Temporary Change (TC) at observation 25 at 2.0 standard deviations, but there is nothing seen in the MA (1) model. The ARMA (1,1) model also finds a LS at observation 92 but nothing is seen in the MA (1) model for that point. The MA (1) model finds an AO at observation 92 but there is nothing reported by the ARMA (1,1). Of course, once the level shift was detected, the additive outlier is undoubtedly part of the new process level.

Conclusions on Models

The proposed ARMA (1,1) model does a more than adequate job of approximating the AR (1) and 1/AR (1) models in terms of outlier detection and identification. There are some differences in identification of outlier type due to the mathematics of the identification process. But, the ARMA (1,1) model is a good approximation and would thus mean one does not have to identify the model type prior to examination for outliers. One can apply the ARMA (1,1) model and expect to find the appropriate outlying observations so they can be studied and the causes for them can be determined. The ARMA (1,1) model, due to the memory effect of the AR portion of the model, is not as good with the MA (1) time series data set studied. The results are still adequate as a first approximation, however, and that is the purpose of the proposal. The ability to work with the data without the need to determine the exact model is what is sought. While one can never expect an approximation to always be exactly correct, it appears that the use of the ARMA (1,1) as an approximation is valuable.

3.4 Further Efforts

The initial experimental work clearly shows that the ARMA (1,1) is a very good estimator of AR (1) and 1/AR (1) time series models. The ARMA (1,1) is also shown to be at least capable of an adequate estimation of MA (1) time series models. On this basis, additional historic or well known data sets are analyzed using the suggested technique and compared to the computer suggested model and other published examinations of the data set. Then, new data sets are also examined using ARMA (1,1) as the approximation model to determine outlier locations and types.

Chapter 4

Results of Studies of Process and Product Data

4.1 Introduction

For the purposes of this study, a product is defined as the result of a process. Product parameters are those characteristics of the product measured at some time during the process, often at the conclusion of it. Some product parameters could be strength level, final concentration of some element or other chemical characteristic, or size, to list but a few examples. A process parameter is facet of the process that is measured and controlled during that process. Process parameters could include temperature as a measure of heat, concentration of reactants or catalysts, or processing speed, to name but a few common ones. This chapter will cover results of our studies of actual industrial data in the process and product areas using the ARMA (1,1) estimator to determine outliers. The advantage to the ARMA (1,1) for a stationary system is that the model includes both the AR (1), and the MA (1) time series models. As has been noted previously, the AR (1) is equivalent to an infinite length moving average (MA) model while the MA (1) model matches an infinite length autoregressive (AR) series. Also as noted previously, we are concentrating on stationary series or those processes that would be considered statistically stable. If one were dealing with a non-stationary process, the ARIMA (1,1,1), for example, would probably be more useful.

The data sets used for analysis were obtained from both historic sources and from some newly available data from a steel mill. The historic data, the sheet-like process for example, can be found in the references cited. The new data sets are extensive and several years old and do not represent processes and products which are still in operation or production. The data sets can be obtained from the author if one is interested in further analyses of that information.

4.2 Data Analyses Methods

To establish a base point for moving forward with improved data analyses methods, the data were first studied using a classic control chart for individuals with a moving range of 2. While the use of subgroups would reduce the amount of variation determined between two plotted data points, we recognized that a rational subgroup may not be possible. Thus, charts of individuals were used but the wide variation inherent in such charts did not allow the data to be well analyzed because a chart of individuals with a moving range provides control limits that are very wide in the portion of the chart where the actual data values are plotted. This portion of the chart is often known as the "X" portion where "X" is the measurements of interest. Because the variation in the data is not moved to the range chart, as is the case when using subgroups, the limits for determining outliers are much wider in the case of charts of individuals than the classic Shewhart charts using rational subgroups. Outliers were determined at 2.0 standard deviation units rather than the more conventional 3.0 standard deviation units. Although we do question the validity of such control charts, as has been previously stated, it is necessary to establish a base point for comparison of outlier detection methods.

After the base case was established for comparison, the data were then treated to the proposed ARMA (1,1) analysis using the joint estimation technique of Chen and Liu (1993). While there are many concerns about the conventional control chart of individuals with a moving range, the method is established and does provide a basis for comparison of outlier detection. The conventional 3.0 standard deviations limits are arguably too broad and this tends to make outliers more difficult to determine. Thus, using the arguments put forth earlier, we will use a 2.0 standard deviation limits for the base case as well as the ARMA (1,1) model.

As the control chart and time series methods mechanically look at the individual data point differently, the control chart is historical and the time series anticipatory, the identification of a data point as an outlier by either method may not exactly match the other method. The point of interest is whether the model is locating a change in the process. Therefore, for the purposes of comparison, if the two methods find the same point to be an outlier, with a tolerance of +/- 1 position, a match is deemed to have occurred. It is clear that both models are finding a change in that area of the data and to say that one missed the change would not be justified if the match is close. Thus, the settlement on the +/- 1 rule for a match.

Another interesting aspect of the study is that because control charts are historically based, shifts in the process average can lead to many additional out of control points after the shift has taken place, if the shift is not recognized and the limits recalculated. In the ARMA (1,1) model, if the process average shifts, the outlier causing that change is identified as a Level Shift (LS) type outlier by the joint estimation technique and later data values are then compared to the expectations of the new process average. However, this is done automatically by SCA rather than requiring intervention by the operator and the recalculation of control chart limits.

Using the argument made previously that limits of 2.0 standard deviations are more reasonable than the 3.0 standard deviations limits, outliers at 2.0 standard deviations limits were determined and compared to those found using the classical control chart method. As noted above, we are interested in changes and shifts in the process. As we are interested in the identification of shifts and other changes in the process, the identification of the exact data point by two methods which mechanically examine the data differently may not lead to exact matches in outlier identification. Thus, we compared outliers found using ARMA (1, 1) to those found in the base case using the criteria of +/- 1 position from the base case. Also, we shortened the data set to the last outlier found and again checked the data for outliers using the ARMA (1,1) model. A comparison of the second run's findings with the complete data set was then made in terms of both the ability of the model to determine the same outliers in the shortened data set, and the type of outlier determined.

Additionally, the moving range data were treated with the ARMA (1,1) model and outliers and outlier types determined. These data sets were also treated a second time after reducing them to make the last data point the last outlier originally determined. As with the individual data values portion of the charted data, 2.0 standard deviations were used as the criteria for determination of an outlier.

4.3 Process Data

The original Shewhart control chart methods involve rational subgroups of samples taken from a manufacturing process over a short period of time such that no machine adjustments are made during the production of the subgroup. Also, there is the assumption of independence of each sample from previous and following samples in the subgroup. I submit that intuitively these assumptions do not hold for process control data and thus the application of standard Shewhart methods can lead to erroneous conclusions about the state of the process.

In any process control situation, there is an operating level for each parameter. If the parameter is checked and the level found meets the desired operating level, no action is taken. If, however, the level determined is outside of the desired level, action is taken to either increase or decrease the parameter under study. For example, if temperature is considered a critical factor for a chemical reaction, there is some range of temperature over which the process is considered to be operating effectively. If the temperature drops, heat would be added to the system to raise the temperature. Likewise, if the measured temperature was beyond the operating goals, less heat would be added to the system, or it might actually be cooled, to reduce the temperature to within the desired operating range. However, as the process is either absorbing or generating heat, if one does nothing because the system is within the operating range specified, will the temperature increase or decrease over time? One would surely expect something to happen. Thus, within a process, the measurement of a parameter causes some response, even no action is a response. Thus, later checks of temperature, in our example, are influenced by the previous check. Thus, the later measurements are not independent of the previous checks and one of the underlying assumptions of Shewhart charting is not met. However, as there is dependence upon prior events, the measurements clearly reflect a time series and establish analysis using time series methods as the appropriate method to treat the data.

An example of process data is a data set dealing with the percent oil concentration in an oil/water mixture used in a rolling operation in a steel mill. The rolling solution acts both as a lubricant and a coolant during the rolling operation. The oil concentration is important in reducing frictional effects and thus heat generation in this process and thus is an important process characteristic that is measured and controlled. These data are identified as the "Oil Concentration" data set on page 140 in the Appendix. A summary table is shown below as Table 1.

Table 1 – Process Data

Data Set	Control Chart Outliers	ARMA (1,1) Matches +/- 1	Additional Outliers From ARMA (1,1)
Oil Concentration	9	8	10

Referring to Table 1, a Shewhart control chart of individuals was constructed as a base case. The data were then treated using the ARMA (1,1) method suggested previously. The results are as shown with the number of outliers determined in the control chart of individuals with a moving range and the number of outliers from the ARMA (1,1) model which match the control chart outliers, using the +/- 1 convention, as well as the additional outliers detected at the 2.0 standard deviations level. As noted previously, because we are comparing two different methods, an historic view in the case of control charts and a forward looking method in the case of a time series, exact matches of out of control points are probably not to be expected. Rather, we are looking for a change in the system that is identified by the method under study. Thus, we have adopted a convention that a match is found when the two methods identify a data point either exactly or the data point on either side of the item in question. Thus, if the control chart method determines that point #39 is an out of control point, a match is judged to have been found if the ARMA (1,1) method determines an outlier as data point #8, #39, or #40.

Oil Concentration

Please refer to Table 1 on page 78 and to page 140 in the Appendix for the data set titled: "Oil Concentration. The control chart of individuals with a moving range located nine (9) out of control points at the 2.0 standard deviations level. The ARMA (1,1) time series found eight of the nine at the same limit of 2.0 standard deviations, but the ARMA (1,1) method detected a level shift prior to the one missed. However, as the model was then operating at the new level the outlier in question would not be considered an outlier as the process was varying in the usual manner except at a different process average. It should be noted that one process average was determined for the control chart as the assumption was made, as noted previously, that the system was statistically stable.

Also, all outliers were judged at the 2.0 standard deviations level in all cases. If one wished to recalculate system averages and control limits for changing systems, that could be done. However, the time series method determines those changes without additional input and thus saves considerable time and effort.

Of even more interest is that the ARMA (1,1) method determined an additional ten (10) areas of possible interest over the period of time studied. Analysis of control charts for runs and other signs of non-random data is very time consuming. The time series method does that as part of the investigation with the location and identification of temporary change and level shift outliers and thus provides superior information about the data under consideration. A run, a series of data points located either above or below the process average is really either a temporary change or a level shift. The early identification of such situations is desired in process control and the ARMA (1,1) model does that well.

When the data were shortened to make the last data point the last outlier found in the original search, all of the outliers originally found were again found and 17 of 18 were identified as the same type. The last outlier found was originally determined as a Temporary Change but with the balance of the data removed was listed as an Additive Outlier during the second run. Chen and Liu (1993) have suggested that AO and LS outliers are the boundary conditions for TC. Thus, it is not strange that when the data points after a reported TC are removed, the data point itself would be determined as an AO type in the second run against the ARMA (1,1) model.

The range data were also treated with the ARMA (1,1) model and outlier types determined. The data are found in the Appendix on page 140. Twenty-one (21) outliers were found in the run using all the range values with the ARMA (1,1) model and detection limits of 2.0 standard deviations for outliers. The shortened run, where the data after the final outlier were removed, still found 18 outliers at the 2.0 standard deviations level. Of those not found with the second run, one was between two others in consecutive data points and the other two outliers were found in both runs. The other two outliers were two consecutive data points classified as IO during the first run. One of the values is exceptionally high and probably should have been detected. The second one is a return to more regular value levels. Five of the outliers detected during the second run were of a different type than suggested initially. A TC was reclassified as an IO, an IO as an AO, an AO as an IO, a LS as a TC, and the final value was a change from IO to AO. It has been noted previously that AO, TC, and LS are of the same family and differ on the decay rate or dampening rate of the model. However, IO is reported by Chen and Liu (1993) to be model dependent. Thus, it is possible that while the ARMA (1,1) approximation may still find the outliers, and that is the principle interest, identification may be of some question in the IO case.

Process data are intuitively a time series and should be treated as such. For the example shown, the ARMA (1,1) estimation technique worked well. Outliers found using the conventional method were matched and several other interesting data points were found and identified as outliers of particular types. The analysis of the range data also found some outliers but did not add greatly to the understanding of the data set as all outliers in question were found by the ARMA (1,1) analysis of the individual values themselves.

4.4 Previously Examined Product Data

The results of the various tests on product data are summarized in Table 2 on page 84. The original output is shown in the Appendix beginning on page 144.

The Shewhart control chart method of analysis of product data has the same problems as process data. Rational subgroups may not exist in many manufacturing activities. Shewhart's method clearly applies to repetitive discrete part manufacturing where time is limited between samples and no machine adjustments are made during the period of time over which the subgroup is selected. However, many manufacturing processes are continuous or batch rather than discrete. Continuous processes make the development of a rational subgroup impossible, because the process does not stop to allow samples to be taken. The output of a continuous process at time t+1 is surely related to the output at time t, or there would be no need to monitor and control the process variables and one would just hope the product characteristics were satisfactory. The previous discussion of process data and lack of independence applies here as well. And, if the rule of independence is not met with continuous processes, the direct application of Shewhart control chart theory is flawed.

A second type of process with problems meeting Shewhart criteria is the batch process. Or, as is often the case in industry, a succession of batch processes such that the original batch of material is neither processed in the same batch nor at the same time nor in the same sequence in various processes. In terms of the sequence of the finished product, as compared to the original sequence, a comparison of shuffling and reshuffling a deck of cards provides the same sort of reordering of the sequence of the parts or products. While the original master batch may be uniform, material from that batch is subjected to various treatments, in various groupings, and at various times. As is well known, the variations stack up in the final product under such circumstances. Further, in terms of measurements of product parameters at the end of the processing cycle, what would be a rational subgroup? The response is that there is no rational grouping possible while maintaining the idea of a relatively short time span over the production of the material. Thus, the Shewhart control charts do not adequately meet the needs for analysis of product manufactured under such circumstances. On the other hand, time series model might have the potential to analyze such product parameters and the ARMA (1,1) is suggested as a first approximation.

Other researchers have investigated some of the data sets available and Prasad, Booth, Hu, and Deligonul (1995A) reported on studies of three product characteristics using robust and semi-robust methods. These authors also used the joint estimation techniques of Chen & Liu (1993) for data analysis in this work, but actually tried to determine the best fitting time series model and then determined the outliers from that model. The method suggested here, the use of ARMA (1,1) as an estimator, was applied to those data sets and some new ones. The results are as follows with the outliers determined at 2.0 standard deviations for the control chart of individuals with a moving range of 2 compared to ARMA (1,1) outliers found at 2.0 standard deviations using the +/- 1 allowance for matches. The additional outliers detected by the ARMA (1,1) model are also shown. As has been noted previously, as we are using two difference mathematical models, one taking an historic view and the other a forward looking view, exact matches of out of control points or outliers are probably not to be expected. Thus, a match of an out of control point is judged to exist if there is an exact match from the control chart and the ARMA (1,1) method if the same point is found or if the identified data point is the one on either side of the control chart identified point. Thus, if the control chart identifies point #26 as an out of control point, a match will be recorded if the ARMA (1,1) method finds point #26 or #25 or #27 as an outlier. The complete data sets are found in the Appendix on pages nnn through nnn. The outputs are summarized below in Table 2.

Table 2 – Product Data

Data Set	Control Chart Outliers	ARMA (1,1) Matches +/- 1	Additional Outliers From ARMA (1,1)
Sheet-like Process	20	16	11
Automatic Transmission Parts	3	2	2
Bore Hole Location	4	4	4
Yield Strength 1	5	5	25
Yield Strength 2	9	8	12
Silicon Content	35	27	19
Ash Percent	5	5	18

Sheet-Like Process

In the Sheet-like Process from Johnson and Bagshaw (1974), Prasad, Booth, Hu, and Deligonul (1995A), found four (4) outliers using the exact time series model and limits of 3.0 standard deviations. As shown in Table 2 on page 84, and pages 144 through 147 in the Appendix, the ARMA (1,1) estimation model found all of those outliers and identified 12 others found by the 2.0 standard deviation control chart technique as well as 11 additional areas of interest. Of those data values shown as outliers by the control chart but not identified by the time series estimation, three (3) of them were very early in the data set and the other one was part of a temporary change in the product average of the model prior to that data point. As the level had already changed, the data point would not have been considered different from the new product average and would not be considered an outlier. I would suggest that the first three data points were found to be outliers using conventional control chart techniques because the average and control limits are based on the entire data set. As these data points were higher than the long-term average, they were noted as being outliers. Had they been anywhere else in the data set, they would have been identified as outliers by all methods used.

There has been some question relative to the effectiveness of the Joint Estimation Technique on very short data sets. There has been much attention paid to short data sets in general and how to determine appropriate limits and outliers for such data sets. It appears that the use of the ARMA (1,1) approximation method does not solve this short data set concern, although the suggested ARMA (1,1) method does very well at determining outliers in longer data sets. Wright (1997) has noted that the joint estimation technique works well with time series well less than 20 data points in length. However, when the data points fall within the first five (5) values, only an historic model such as a control chart will locate such points, and then only as an afterthought as the method would not have designated the values as outliers at the time they were measured. A forward looking method, such as a time series model, will have insufficient time to determine expectations in such early data point cases. It should be noted that this problem with very short data sets is well known and has been addressed by others. No simple method has been found to deal consistently with this situation, although a number of short term solutions have been suggested. When a process first starts up, the variation is expected to be significant but not representative of the long term variability when the process has come to a steady state. While the joint estimation technique appears to fail in this situation, no other method has been found that clearly seems to do better. If the first few data points are not representative of the process as they are startup situations, such situations have to be considered as separate cases apart from any analysis of the longer run. It is not an uncommon industry practice to discard the first few items produced to allow the process to come to equilibrium before attempting to determine conformance to some specification or standard. Thus, in many cases, one would discard such early data from consideration and the significance of the first data points in a data set could be discarded from the analysis.

Interestingly, the 2.0 sigma limits on the control chart identified 20 outliers. Prasad, Booth, Hu, and Deligonul (1995A), reported four (4) of those data values at the

3.0 standard deviations limit. The ARMA (1,1) method found the same outliers as Prasad, et al (1995A), although at the 2.0 standard deviations limit, but also identified them as the same outlier type as found by Prasad, et al (1995A). However, the ARMA (1,1) method found 15 of the additional control chart suggested outliers using the +/- 1 data point criteria. Noting that the first 3 data points were missed, the ARMA (1, 1) method missed only two supposed outliers but had reported a TC just earlier in the data set. When such temporary change takes place, the process average is temporarily shifted and the outliers from the control chart may well not be classified as different from the expected product average at that time using the ARMA (1,1) model.

The truncated or shortened data set, reduced to the last data point identified as an outlier in the first run, found all of the original outliers except the last one, but did also identify three new outliers of interest. The last data point was classified as an AO by the original run but a data point two (2) points earlier was classified as a LS in the second run and the final value not determined to be an outlier.

Twenty-five (25) outliers were suggested by application of the ARMA (1,1) model to the range data. When the data were shortened, all 25 suggested outliers were again found by the ARMA (1,1) model. Sixteen IO types from the first run were reclassified as AO types in the shortened run. All the IO type outliers were reclassified as AO with the changed data set. As noted previously, IO types appear more model dependent than the other types and the change in the data set could have some impact.

Automatic Transmission

With the automatic transmission part data from Quesenberry (1990), Prasad, Booth, Hu, and Deligonul (1995A), found two (2) outliers using their technique at the 3.0 standard deviations level. This data set consists of 45 consecutive measurements of a critical diameter of an automatic transmission part. The data are shown on pages 148 through 149 of the Appendix and summarized in Table 2 on page 84. Referring to Table 2, the control chart of individuals found three (3) outliers at the 2.0 standard deviations level. The ARMA (1,1) method found two of three control chart outliers, missing an early one, actually data point 2, as well as finding all the Prasad outliers and also identifying two (2) additional areas of interest. The ARMA (1, 1) model also classified the outliers detected exactly as the Prasad method had in both the full length and shortened data sets. The last outlier was suggested to be an AO by the full data set and the shortened version also noted it to be an AO type. Again we have the problem of early data points that are found to be outliers based on long run averages using all the data. The time series approach is not as long-term in the value anticipated and that has an impact on the detection of outliers very early in a series. Examination of the data would not suggest the second data point is very different from the first and third in actual value. Only when the comparison is based on many subsequent data points is that difference suggested.

A review of the range data findings shows that full length data set found only the range values associated with the data points detected by Prasad, Booth, Hu, and Deligonul (1995A) outliers and identified the range component to have IO type outliers at those positions. The shortened data set found four (4) additional outliers, three (3) new areas of interest, and changed the final data point from an IO to an AO type. Again, the significance of the model itself on IO vs. AO type outlier determination is seen.

Bore Hole Location

As noted in Table 2, page 84, the Bore Hole data set from Quesenberry (1986) had four (4) outliers by the control chart and Prasad, Booth, Hu, and Deligonul (1995A), using 3.0 standard deviation limits, also found those four (4) outliers, as well as suggesting one new one using the Robust Method. The full data set is found in the Appendix on pages 150 through 151. This data set deals with the location of injection pump bore holes in automotive engine blocks. The Semi-Robust Method also found five (5) outliers, including all those from the control chart. However, the additional outlier identified was different from the extra one found using the Robust Method. The ARMA (1, 1) method found all those outliers as well as three new areas of interest operating at the 2.0 standard deviation limits used in this work. It should be noted that the ARMA (1,1) method found a total of 11 possible outliers, including the actual last data point as was found by the control chart and the Prasad, et al (1995A) methods. As discussed previously, when an outlier is identified by one method is within +/- 1 data points of an outlier identified by another method, the same situation is deemed to have been identified. When the suggested outlier is not adjacent to any previously identified locations, a new area of interest is said to have been found..

The range data found three (3) outliers in the first 19 data range values. By definition, this 19^{th} point would be the difference between the 19^{th} and 20^{th} actual data values. When the data set was shortened to 19 points, a TC in place 7 was not found and the last point which was classified as a LS, was reclassified as an AO. In addition, the 17^{th} value was classified as an AO in the initial run but as an IO in the shortened study. Certainly the difference in the last data point is due to the missing values that make a LS. If the extra values are not there, the mathematics would not determine the LS type so the AO is reasonable. The change just prior to that, from AO to IO, would seem due to the model itself, as noted by Chen and Liu (1993).

4.5 New Product Data

Clearly, the ARMA (1, 1) method identifies the same potential outliers as other methods which have been applied to previously studied data sets. A series of new data sets of product information have also been studied using the same control chart and ARMA (1,1) methods and those results are now reported. The data sets are found in the Appendix on pages 152 through 172 and summarized in Table 2 on page 84.

Yield Strength 1

Referring to Table 2, page 84, and pages 152 through 157 in the Appendix, a series of data values on the yield strength of cold rolled steel coils was reviewed using the control chart of individuals with a moving range of size 2 technique and 2.0 standard deviations limits. The variability within the data set provided limits so broad as to produce only five (5) outliers. Recognizing that this data set is a classic example of multiple batch processes where the material from the original batch is mixed repeatedly in later processing stages as to time and sequence of processing, one wonders whether the data used in testing sequence, which was the only rational data order known, is sufficient to allow detection of material and processing differences that could be significant. Obviously, the control chart method is marginal in such studies, based on the very low number of potential outliers identified.

The analysis of the data using the ARMA (1,1) technique matched all the control chart suggested outliers and noted 25 additional locations of potential outliers. This would be more realistic to a practitioner in this field in terms of locating groupings of processing within the final mix of test pieces. When the data set was shortened, three areas of suggested outliers were not found, and the final value was changed from LS to AO. As noted previously, if the values that make a LS are removed, the LS would have to become something else and an AO is probably reasonable. Some slight changes in outlier type were noted, TC to AO for example, but the same locations for the outliers were identified for further study and understanding.

Interestingly, only three (3) outliers were suggested from the moving range data using the ARMA (1,1) model and when the data set was shortened, the last point, an LS, was not located, although a previous point was noted as an AO type at a point where a TC type was previously suggested.

Yield Strength 2

A second set of yield strength data was acquired and examined and nine (9) outliers were suggested by the control chart of individuals method, as reported in Table 2 on page 84 and on pages 158 through 162 in the Appendix. The ARMA (1,1) model matched eight (8) of those data points as potential outliers. The point missed does not appear to be very different from those on either side of it and is probably noted on the control chart because of long-term averages rather than short-term averages. The ARMA method can provide a changing average for comparison while the control chart has a fixed average. Thus, changes seen in the control chart may not be found important with the ARMA method and as the underlying assumptions of the control chart are questioned and a time series is suggested, the ARMA method is probably superior.

There were a total of 20 outliers suggested by the ARMA (1,1) analysis, including the eight (8) matches with the control chart study. As was noted in previous work on yield strength data, please refer to the comments on Yield Strength 1 above, the outlier types suggested match technical understandings within that industry as to the effects of batch processing and mixing the batches during processing as to sequence and actual process application.

When the data set was shortened, 14 of the data points suggested as outliers were also found. Five (5) data points originally classified as outliers were not found the second time and the last data point noted as an outlier in the first run was not found in the second run. There were also several differences in outlier types suggested between the two runs. The actual data effects the mathematics of model development and is one of the strong reasons for using a single model for this analysis, rather than trying to estimate an exact model, which may not be correct in the first place.

Fourteen (14) outliers were suggested in the range data analysis and when the data set was shortened, thirteen (13) of those were again located. Two new areas of interest were also noted but the final outlier from the original range data was not located with the shortened data set. Agreement of outlier type was good, however, with 11 of 13 items being of the same type in both analyses.

Silicon Content

A data set of the silicon content of cast iron, see Table 2 on page 84 and pages 163 through 168 in the Appendix, was examined and we found 35 outliers suggested by the 2.0 standard deviations limits of a control chart of individuals with a moving range of size 2. The ARMA (1,1) method matched 27 of those locations and provided an additional 19 suggested outliers. Four (4) of the outliers suggested by the control chart but not by ARMA were the first four (4) values in the data set. All the other data points suggested as outliers by the control chart but not ARMA were in areas of the data set that the time series analysis had noted as having a level shift prior to those points. Thus, the data values while outliers based on long-term averages would not be so classified in the short-term case as would be found with an ARMA (1,1) analysis. Again, we see the well recognized problems with early points in a data series. That is an area that is of continuing interest, but beyond the scope of this work.

In addition to the potential outliers noted in both control chart and ARMA (1,1) studies, the additional 19 potential outliers found by ARMA add significantly to the understanding of the product characteristic under study. When the data set was shortened, all the original outliers were again found and no new areas of interest were located, although a couple new data points adjacent to other previously identified outliers were noted. The last data point was classified as a LS in the original work and as an AO in the shortened set, as would be expected.

In the range data, the original ARMA (1,1) run found 20 potential outliers. When the data set was shortened, five (5) data points originally detected were not detected as potential outliers. Nine of the matched data points had the same type of outlier identified and the final point was found as an AO in both runs.

Ash Percent

A data set of ash percent of coke as studied using the control chart of individuals with a moving range. The data are summarized in Table 2 on page 84 and are found in detail in the Appendix on pages 169 through 172. As with all the studies, the moving range is the difference between successive data points and a moving range of 2, the method used here, takes the absolute difference between these two successive data points as a measure of the variability of the process. This information on the process variability is then used to calculate the control limits for the chart of individuals and is the basis for the 2 standard deviations used as the limits throughout this work. Under that study, five (5) points were found to be out of control, as shown in Table 2. When the same data were treated to an ARMA (1,1) analysis, all of the outliers from the control chart were matched on a +/- 1 basis by the Joint Estimation Method and an additional 18 data points were identified as potential outliers. It is interesting to note that all the control chart outliers were noted as either LS or TC types by the ARMA (1,1) method.

When the data set was shortened, only two (2) outlier areas were not identified as the same type of outlier as the original data set. The final outlier was a LS in the complete set but an AO in the shortened version, as has been found to be common.

The range data showed 28 potential outliers, including the very last data point so the shortened set was the same length and the same data points were identified as potential outliers.

4.6 Summary

Product data, as the results shown in Table 2, page 84, indicate, respond nicely to the ARMA (1,1) method suggested, when compared to previously conducted work by Prasad, Booth, Hu, and Deligonul (1995A) on some older data sets. The analysis of the new data sets shows the ARMA (1,1) method presents a better understanding of the data itself and more clearly notes potential outliers than the control chart method. As it is the purpose of an analysis of this type to locate potential problem areas, the method that does a more thorough job is probably to be preferred.

There is some concern with false alarms with overly sensitive analysis methods. We have to be cautious that the additional data points suggested as outliers are in fact actual concerns and not false alarms. The use of wider limits, 2.5 or 3.0 standard deviations for example, would provide fewer outlier indications. For ongoing operations, perhaps that would be reasonable. If one wants to better understand the process under study, however, one is interested identifying and understanding as many anomalies as possible. The question of false alarms can only be based on what level of variation is considered acceptable for the process or product under study. One must first fully understand the process before making that decision. Thus, as the tighter limits aid in an understanding of the process they are valuable for the analysis and understanding of the process. And, I must emphasize that the analysis of the data using the methods described above more clearly reveals to a practitioner in the field what is happening with these product characteristics than previous methods. Previous work had not supplied the depth of analysis presented by the ARMA (1,1) method. Each data set is clearly a time series of some sort. The use of a time series method to analyze the data would seem logically to be superior to other methods that do not take the time series relationships into account.

The analysis of product characteristics using time-series methods is appropriate. The application of ARMA (1,1) as an estimator of the outliers using the Joint Estimation Method is shown to clearly identify many more potential outliers than control charts of individuals with a moving range of 2 at the 2.0 standard deviation limits in the cases of systems which are considered statistically stable and in some systems that operate at different levels but do not have trend patterns. If trend patterns existed, an ARIMA of some sort would be more desirable with a differencing factor determined by the time series in question.

Chapter 5

Results of Studies of Business Data

5.1 Introduction

While product and process data can lend themselves to rational subgroups and traditional Shewhart analysis, business or administrative data do not. Most efforts in control chart studies have been devoted to process and product data. However, there is nothing to make one think that information from the daily conduct of the business can not also be treated in the same statistical manner. Business data is administrative information on measurements such as sales, orders entered, employees, etc. The data are clearly quantitative and ratio in nature as they are generally counts of some sort. Thus, we should be able to analyze business data using the same tools as for product and process data. As the application of such techniques to situations not on the plant floor has not been well reported, this work exposes some new areas of interest for future research. Business data, for the purpose of this work are administrative details, which do not typically occur in groups. Rather, business data are found as individual values only. Also, one can logically argue that an action taken today that yields a business data point has been influenced by previous events. Many of the business situations are heavily influenced by people and their opinions, biases, etc. Business data is not the same as measurements of a part being stamped or machined on a repetitive basis. Rather, business data is a single situation influenced by previous occurrences, but not part of a rational subgroup. Business systems react quickly to events, almost immediately in many cases, as can be seen in situations such as the movement of the stock market relative to changes in rates by the Federal Reserve, unless the market has previously taken some action based upon anticipation of such a change. As with product and process data, we are looking for stationary systems. Thus, we will use the ARMA (1,1) model for the type of data being considered. By using ARMA (1,1), we include the power of both the AR (1) and the MA (1) in the model. Also, the AR (1) is equivalent to an infinite length moving average and the MA (1) is equivalent to an infinite length autoregressive series.

5.2 Data Analyses Methods

Summaries of the business data test results are found in Table 3 on page 99 with more extensive information in the Appendix on pages 173 through 205. The business data used covers rejection rates, fraction of material not useable (usable product divided by total product consumed); rejected tons on a given application by day; and tons of orders entered for one location by day. In each case, the data were treated as a control chart of individuals with a moving range. Outliers were noted at 2.0 standard deviations from the mean. The control limits, of course, were calculated using the moving range data.

The control charts of individuals, even allowing for concern about the IID characteristics of the data, serve as an understood basis of comparison for the effectiveness of the proposed ARMA (1,1) method.

5.3 Business Data

Nine data sets were also examined using the ARMA (1,1) method previously proposed. The results of those trials are summarized below in Table 3, page 99, and found in detail n the Appendix on pages 173 through 205. Each data set will be discussed individually. In the same manner as the process and product information, the number of outliers at the 2.0 standard deviations limits was determined for a chart of individuals with a moving range of 2. As the thought process is different from the two methods, historic control charts and forward looking time series models, it is possible that certain data values will be noted by both of the methods but one method will find the outlier offset slightly from the first method. Thus, the comparisons were made with a tolerance of +/- 1 data value being considered a match. With this matching practice, if the first method found an outlier at position 34, if the second method reported an outlier at any of positions 33, 34, or 35, a match would be deemed to have occurred. Something had been identified as different in both situations and that is the important aspect of the study. In addition, the number of additional outliers found by the ARMA (1,1) model are also listed.

Table 3 – Business Data

Data Set	Control Chart Outliers	ARMA (1, 1) Matches +/- 1	Additional Outliers From ARMA (1, 1)
Rejection Rate 1	0	0	4
Rejection Rate 2	1	1	0

Rejection Rate 3	2	2	6
Rejection Rate 4	1	1	3
Rejection Rate Total	7	5	5
Rejected Tons	3	3	12
Orders 1	7	7	16
Orders 2	7	7	39
Orders 3	9	4	6

Rejection Rate 1

Please refer to Table 3 on page 99 and page 173 in the Appendix. The rejection rate data is a decimal fraction and is the ratio of weight rejected to the weight used on a particular part. As the total weight used included the rejected weight, the value of the fraction always has to be less than 1. There were 22 data points in this data set with the control chart of individuals with a moving range of 2 reporting none out of control at 2.0 standard deviations, as noted in Table 3. The ARMA (1,1) method found 4 outliers, one IO and 3 AO in the original data set. As a check of the method, for each data set we shortened each data set to the point where the last data point was the last outlier detected. Then, the ARMA (1,1) model was again applied to determine if the shortened data set would respond with the same outliers as previously found. When shortened to 19 data points to make the last point a reported outlier, all four points were again identified as outliers of the same type as the original data set. The first point found to be a potential outlier was point 9. This confirms the work of Wright (1997) where short data sets were shown to respond well to the joint estimation technique in outlier detection. Interestingly, there were no points considered an outlier in the range chart, either from control chart efforts or the application of the ARMA (1,1) method.

Visual examination of the data set would lead one to believe that there were changes that took place that were not picked up the standard control chart of individuals. There are instances where the value of a data point was over twice that of the adjacent point, but this was not detected by the control chart. The ARMA (1,1) method picked up those points as outliers.

The definition of an outlier is a data point that has a t-value greater than 2.0 in comparing the proposed and actual value. Each comparison of a data point with either the long term system average in a control chart, or the proposed value in the case of the time series, is in fact a t-test. The control chart performs the test visually by use of the control limits. The time series does the test mathematically. In each case, the outlier has a value of over 2.0.

Rejection Rate 2

A second set of rejection rate data for another period of time was analyzed as Rejection Rate 2 with the output of the tests summarized in Table 3 on page 99 and in the Appendix on page 174. The control chart of individuals found one (1) outlier at the 2.0 standard deviation level. The ARMA (1,1) model found that point as an IO. No additional data points were identified by the ARMA (1,1) method on the original data set of 21 values. The shortened data set of 13 values found the same outlier as the last value and classified it as an AO. Interestingly, the shortened data set also found a level shift at position 9 which again emphasizes the effect of the data set itself on the time series model calculated by the SCA package and other methods.

Examination of the entire range data showed two IO's as positions 13 and 14, which corresponds well to the examination of individual values. When shortened to 14 data points, the range data identified point 13 as a level shift and ignored position 14 on that basis. The shortened version also determined position 10 to be and AO and this matches well with the LS reported in the shortened version of the base data set.

Rejection Rate 3

Rejection Rate 3 is a third time period studied and the control chart of individuals found two (2) out of control points or potential outliers. The data are shown in Table 3 on page 99 and in the Appendix on page 175 and 176. The outliers identified were at positions 6 and 13. The ARMA (1,1) method found the same data points, and 6 additional ones at the 2.0 standard deviations level. There were several TC's reported and the 22nd value was noted as a LS. When the data set was shortened to 22 data points from the original 25, seven of the eight data values identified as outliers were again found. There were some shifts in outlier type with point 6 changing from IO to AO and point 22 changing from LS to AO, which would not be surprising because the supporting data points for the LS judgement had been eliminated. Point 12 was classified as an IO with the shortened data set, although not classified as an outlier with the full data set. The 15th and 16th data values were considered TC's in the original data set but only the 16th value was determined to be a potential outlier in the shortened data set. The 16th value was classified as a TC in the shortened set as well as the original data set.

Using ARMA (1,1), only 2 potential outliers were noted in the range data and those were the 13th and 18th points. Both those matched items determined by ARMA (1,1) analysis of the original data as potential outliers, however neither item 6, nor any value near it, was determined as an outlier by the range data, although we often see an outlier identified in the range data at the same time an outlier is determined for the individual data values. The value for point 18 was noted as a TC with the entire range data set, but as an AO with the shortened set. Naturally, with the data after point 18 removed, the logic for a temporary change was also removed.

Rejection Rate 4

Rejection Rate 4, for a fourth time period, had one out of control point reported, as noted in Table 3 on page 99 and in the Appendix on page 177 and 178, while the ARMA (1,1) model found four (4) additional items to the 19th point as found by the control chart. When the entire 25 point data set was used, point 7 was found as a TC and the 19th point as an AO while the 20th and 21st points were IO's. Value 23 was determined to be a LS. When the data set was shortened to 23 points, values at the 19th and 20th points matched the original work. The 7th point and the 21st point were not classified as outliers, but the 22 point was noted as a LS, rather than the 23rd point.

The range data found two (2) potential outliers with the entire data set using the ARMA (1,1) model and one of them matched the out of control point from the control chart. When the data set was shortened to 20 data points, a new TC was suggested at point 9, while the last value as classified as an AO rather than and IO. This is common using the joint estimation technique of Chin and Liu and was also noted by Wright

(1997). The TC originally suggested at value 19 was found one value later in the shortened version.

Rejection Rate Total

While systems may perform in a statistically stable manner over short periods of time, it seemed interesting to look at something over a much longer period. Thus, after looking at four sets of data dealing with rejection rates over four different, but successive time periods, the four data sets were put together in time sequence to form Rejection Rate Total. The results of the tests are summarized in Table 3 on page 99 with more detail in the Appendix on pages 179 through 183. We were looking for shifts in the process average that would either result in out of control points in the control chart of individuals with the moving range of 2 or as Level Shift outliers in the ARMA (1,1) time series model. The control chart method found seven (7) out of control points. The ARMA (1,1) model matched five (5) of those points. Point #4 was considered out of control by the control chart, but the first point determined as a potential outlier by the ARMA (1,1) method was point 9, which was the second point noted as out of control by the control chart of individuals. Again, the historic model has the ability to go back to the early points and make judgments on them. The forward looking time series does not deal well with such early data values. Wright (1997) has shown that short time series will respond to the joint estimation technique and find outliers, but not as early as we are finding 'missed' in this work. Finding the outlier at point 9 is completely within expectations based on Wright's work. The other outliers from the control chart were identified by the ARMA (1,1) model exactly, except for point 87 which was bracketed as a TC at point 86 and an AO at point 88. The situation was, however, identified by the model. That is what an estimation process should do – make an approximation that causes one to evaluate situations. The control chart found the last value, #93, to be out of control, but the ARMA (1,1) method did not. The data would seem to be trending downward at this point, and the two previous values are more similar to the last value than many previous values so the miss may not be unexpected.

When the data set was shortened to 88 values, the ARMA (1,1) model picked up all 14 suggested outliers, within +/- 1 position. Two new outliers were suggested and the last two data points originally a TC at point 86 and an AO at point 88 were classified as an IO at point 87.

There were 14 range data points classified as outliers by the ARMA (1,1) model using the entire data set. Five of those trailed the out of control point from the control chart by one place, 10 vs. 9, for example. The very early data point listed as out of control, #4, was also not found in the range chart analysis as an out of control point. At the other end of the data set, a LS was determined at point 91 and nothing after that. When the data set was shortened to that point, only 5 of the original 14 points were classified at outliers, although three new areas of interest were found. Again, the importance of the data set itself to time series model identification is seen.

Rejected Tons

Another way to look at rejections is the total weight rejected and a data set of that type of information was also studied. The results of the analyses of this data set are found in Table 3 on page 99 and on pages 184 through 187 in the Appendix. There is an underlying assumption that the weight used in the various time periods is relatively constant. Thus, changes in weight rejected would demonstrate a change in the quality level, or a change in the product being processed or the weight being processed. Shifts in rejected amounts would allow some understanding of how different products effect performance. Outliers would be expected in such instances because something would be different or have changed. While the data sets used for the rejection rate studies had both the rejected rate and processed weight known, at times a good divisor, the amount processed, is often not known and only the rejected weight is reported. In this case, the data contained in this data set is completely separate from the Rejection Rate information. With 68 data points, a control chart of individuals with a moving range found 3 out of control points as shown in Table 3. The ARMA (1,1) method found the exact same points, and an additional 12 potential outliers. The last value found to be a potential outlier was #67, and when the data set was shortened, all the data points suggested as outliers were again located and the identification of them was the same for both the full length and the shortened data set.

There were 11 outliers suggested in the range data at two standard deviations using the ARMA (1,1) model. Three of them matched exactly or to the +/- 1 criteria used in all the studies comparing control chart out of control points and potential outliers by the ARMA (1,1) model. The last point found as a potential outlier was #67 and when the data set was shortened, all the same data points were identified with the shortened set as with the original set. The last point was noted to be an AO with the shortened set as compared to a TC with the complete data set.

Orders 1

Three data sets were available of tons of steel ordered through an office. The data set of Ordered Tons 1 had seven (7) data points classified as out of control in the control chart of individuals with a moving range. Please refer to Table 3 on page 99 for a summary and pages 188 through 193 of the Appendix for more details. All seven (7) points were also found by the ARMA (1,1) model which also found an additional sixteen (16) potential outliers. When the data set was shortened from 122 to 121 data points, all of the original potential outliers were matched within the +/-1 criteria, and eleven (11) new areas of interest were also noted. The final value, #121, as suggested to be a LS with the original data but an AO in the shortened set. As noted previously, the data that makes the LS disappears with the shortening of the data set and the last value is frequently denoted as an AO by the joint estimation technique.

Orders 2

As summarized in Table 3 on page 99 and pages 194 through 199 of the Appendix, Ordered Tons 2 had seven (7) out of control point by the control chart and all of those were matched by the ARMA (1,1) model. An additional 39 potential outliers were suggested by the ARMA (1,1) method. When the data set was shortened from 118 to 117 points, three (3) potential outliers were not located with the ARMA (1,1) model. Those outliers identified had the same outlier type suggested by both analyses. All the out of control points from the control chart were also found with the shortened data set.

Twenty-four potential outliers were detected in the range data with ARMA (1,1) model and 2.0 standard deviations limits. Out of control points from the chart of individuals were found in the range chart either as exact matches or within the +/- 1 limit. When the data set was shortened to 116, all but one previously identified outlier were found and seven new areas of interest were noted. The last outlier in the original set was denoted as a TC and it was reclassified as an AO in the shortened set.

Orders 3

From Table 3 on page 99 and pages 200 through 205 in the Appendix, one can see that Ordered Tons 3 had nine (9) out of control points noted at 2.0 standard deviations in the chart of individuals. The ARMA (1,1) model found only four (4) of those but also six (6) additional areas of interest. The very first data point was considered an out of control point by the control chart but this point was not located by the ARMA (1,1) model. The 14^{th} point was also considered an out of control point. The ARMA (1,1) model suggested 2 TC's prior to this value, the second point was point 12 or just two prior the out of control point. While the ARMA (1,1) model matched point 18 as an outlier with the control chart, point 20 was not matched and point 30 was not matched either. Points 38 and 43 were found in both methods, but point 64 was noted as out of control but not located by the ARMA (1,1) model. The final out of control point, 106, was found by the ARMA (1,1) model as well. When the series was shortened to 106 data points, the first data value was again missed, but all but point 64 were found by the ARMA (1,1) model. As for point 64, a Level Shift (LS) outlier at point 59 probably impacted the location of the out of control point as a potential outlier.

5.4 Summary

The point should be made that we are using a known method as the base, although we have presented information that suggests that the standard method, the control chart, may not be mathematically correct. Thus, it is possible that the out of control points reported by the control chart of individuals with a moving range may not be correctly selected. Rather, the time series may in fact be more accurate in the estimation of out of control points. Certainly more are found with the ARMA (1,1) method than the control chart using the same limit of detection, 2.0 standard deviations. Thus, the information presented by the ARMA (1,1) model is much richer in details with many more areas of interest. For developing a better understanding of a process, more detail is desired. However, for ongoing control, wider limits with less potential outliers found may be a better method. Charts of individuals often don't allow problem identification due to the large limits. The ARMA (1,1) model is probably superior as it better fits the data and provides more opportunities for judgements about the process, product, or business information.

The analyses of the various business data cases clearly show that the use of a time series to determine outliers is superior to the traditional control chart of individuals with a moving range of 2 because the information provided by the ARMA (1,1) method is richer in the identification of areas where the process has changed. More explanation of the data is provided in each case as data values ignored by the control chart are located by the use of the ARMA (1,1) time series. In addition, the use of the ARMA (1,1) model as an estimator very adequately locates shifts and other changes in the data well in excess of the control chart of individuals.

What is sought with any analysis is a better understanding of the system being studied. The identification of outliers or out of control points is critical to understanding a system because the outliers are potentially the special cause situations that need to be identified and understood if one is to better understand the system. Common causes are those causes of variation that exist at all times in the system under study. Special causes are those causes that are not to be expected in the system. The first thing an analyst must do to understand a system is to identify the special causes so they can be removed from the system. The ARMA (1,1) method identifies many more outliers than the control chart of individuals with a moving range method. While we must recognize that some of the outliers identified may in fact be a common cause effect rather than a special cause effect at the alpha level error of approximately 5%, the ARMA (1,1) method identified so many more potential outliers than the control chart method used as a standard that the analysis of the system using the time series data provides a much deeper understanding of the system because more of the special causes are located.

It has been suggested that various weighted averages and cumulative sum methods are better ways to locate out of control points than conventional control charts and thus possibly better understand the system being studied. The use of the ARMA (1,1), which includes the moving average, clearly takes into account the weighted and moving average methods. The CUSUM method was used by Prasad, Booth, Hu, and Deligonul (1995A), along with the joint estimation technique and the estimated time series model. The ARMA (1,1) method found the same data points, as both the CUSUM and estimated actual time series model found. While this data was product information, as the process and product data and business data have been shown here to react similarly in the analysis using ARMA (1,1), the time series, ARMA (1,1) can replace weighted average and CUSUM methods in analyses of data sets for business data, as well as for process and product data as shown in Chapter 4.

Chapter 6

Conclusions and Future Research Opportunities

6.1 Introduction

The original work by Shewhart in the 1920's, and then that work as promulgated by W. Edwards Deming until his death in the 1990's, was developed for discrete manufacturing situations where certain parts were repeatedly formed, machined, etc. In summary, the control chart concept was developed and applied to a manufacturing process where individual parts could be evaluated through measurements of some type. The data were arranged in rational subgroups that were small enough to have some variation, but not so large as to have excessive variation. The charts developed were for the averages of the groups and the ranges of those groups. The control limits were set at 3.0 standard deviations of the sample means so that standard errors were reduced by the square root of the sample size, as compared to limits for individual values. In addition to being developed for grouped data, Shewhart charts assume independence of one group from any others before or after it. The method proposed by Shewhart has worked well in industry for discrete mass manufactured products that could be treated as small groups for analysis. The method has not proved as reliable for other situations.

One major concern has been the independence assumption. Various authors have shown that that assumption is not correct for many data sets but that the data are actually not independent and thus form a time series. Further, as the conventional control charts were developed using small rational subgroups in the analysis of variation, problems have been found with treatment of data from processes where rational subgroups can not be reasonably developed, or where the data are actually measurements of individual parts or individual test results. Various methods have been suggested to better analyze the data under such circumstances, including exponentially weighted moving average (EWMA) control charts. By definition, a moving average is a time series. Thus, one of the more prominent control chart methods in use, and reported as excellent for determining small shifts of the measurement average is by definition a type of time series. If independence does not hold, the application of other time series methods would seem obvious.

Data can be of three basic types: product, process, or business. Product data is information collected by some testing, inspection, or measurement of some characteristic of a product. Such data could include length, weight, strength, temperature, or chemical composition, for example. Sometimes, this data can be collected in rational subgroups from samples taken from a production line. However, often the data are actually individual data points and are not reasonably grouped. Control charts of individuals with a moving range are commonly used in such cases but fail to adequately explain the data under consideration as the control limits are exceptionally broad because the distribution is for individuals rather than groups. Further, data on individual tests or measurements may clearly not be independent of previous or subsequent tests as some underlying characteristic of the system may be influential. In the case of steel coils, for example, the strength level is highly dependent on the chemistry of the material and that chemistry is a function of the original melting practices. In such cases, there is a lack of independence in product testing which clouds efforts to determine the effects of other process variations on the results of such tests.

Process data is generally not independent by definition because the next data point is clearly a function of the value measured at the previous point in time. There are only three decisions that can be made on the basis of any given measurement. One can do nothing; one can take steps to increase a process characteristic; or, one can take steps to decrease a process characteristic. If a measurement of some process characteristic is taken, examples being temperature, chemical concentration, etc., the line operator, or often the computer process control system reacts. In general, if the measurement indicates the process characteristic of interest is within the desired engineering ranges, no action is taken. On the other hand, if the characteristic is found to have measurements outside the engineering limits, the system will be adjusted to bring the process characteristic back into the desired engineering limits. Thus, as there is action taken on a process characteristic based upon the measurement of that characteristic at a given time (t), the process characteristic will be affected and the measurement of that characteristic at the next time (t+1) be affected and thus not independent of the earlier measurement. As the subsequent measurements are a function of previous measurements, a time series results.

In addition, most measurements of process characteristics are individual data points and not reasonably combined into rational subgroups. In such situations, the practitioner relies on the control charts of individuals, often using the exponentially weighted moving average chart, to determine the control of the process. Again, we have large control limits and the resulting slow response to changes in the process characteristics, even with the use of a time series model.

Statistics and statistical methods are not ends in themselves. Such methods are tools for data analysis and thus have to provide some reasonable return on the effort of applying them. The currently existing tools for analysis of much of the data used in the industrial and business worlds lack the ability to determine the changes that are constantly taking place in any system. Of particular interest as examples are the data sets of product data identified as Yield Strength 1 and Yield Strength 2.

The yield strength of a material as a finished product is a function of the chemical analysis and all the processing steps through which the material passed during production. One would expect to find some identifiable variation due to chemistry changes as well as different processing lots. The data sets included many different chemistries and processing lots, some of which repeated at later times in the study, but the control charts of individuals with a moving range of 2 identified very few data points as being of concern as out of control points or outliers. The ARMA (1,1) method, however, identified many changes in process average, as one would expect with a change in chemistry and processing lot. From the point of view of a practitioner in the field of science where such information is commonly reviewed, the control charts failed to locate data points that should have been found. The ARMA (1,1) method, however, provided more insight into the system because that method identified as outliers changes in the system that would be expected. To understand a system, we have to find the special causes and all the shifts and changes. The conventional control chart method does not adequately locate such changes and shifts as compared to the time series model proposed.

A third type of data is business data. Such data has seldom been examined because almost without exception, business data are individual data points that are not rationally combined into subgroups and thus fall into the control charts of individuals with the wide control limits and poor ability to distinguish changes in the process. Also, such data is seldom seen in the public domain so that analysis by independent researchers is seldom possible. Within many firms, as the data are seldom found to be informative enough to allow decisions to be made on the basis of the control chart analysis, use of the charting methods noted above are also seldom found.

What we have seen over the years are unsuccessful attempts to use the conventional Shewhart control charting methods with process, product, or business data presented as individual data points. Various methods have been suggested, but the application of statistical quality or process control methodology to such data has not proven as successful as practitioners need. Thus, while the Shewhart methods work well with discrete product data that can be organized into rational subgroups, the application of such statistical methods to individual types of data has not worked well. However, the method suggested here has successfully treated data presented as individual data points in the process, product, and business application areas. The use of a time series model, ARMA (1,1) being suggested as a first approximation, has proved to more than adequately investigate and provide information on system changes and data points which do not belong to the base case. Based on the work reported here, the ARMA (1,1) model allows a practitioner in the field of study to better understand the system under consideration as compared with a conventional control chart of individuals with a moving range, using the same criteria to identify data points which are different.

The work reported here has been done on data presented as individual data points and which were not suitable for combination into rational subgroups. The method has not been applied to grouped data. When the data are grouped, the conventional Shewhart method, the first value of interest is the subgroup average and it is these subgroup averages which are plotted and on which further calculations are made. The variability present is reduced in the subgroup averages because the variance of subgroups is related to the variance of individuals by dividing the group variance by sample size. This reduction in variance leads to an appearance of less variation in the control chart of the subgroup averages. The variability factor is transferred to the range portion of the control chart. We do not know how this seeming reduction in variability affects the determination of outliers. The Hussong Kettle study reported by Grant and Leavenworth (1984) and studied by Sebastian, Booth and Hu (1994) is grouped data, and the Joint Estimation Technique of Chen and Liu (1993) does locate the outliers found by the conventional Shewhart method. Some limited trials, not reported here, with the individual data studied in this work placed in subgroups, did not provide sufficient explanation of the systems under consideration, probably because of the reduced variance in the subgroup average portion of the control charts. We can not say with certainty, however, that the time series method suggested will or will not provide sufficient explanation of grouped data. We suggest that for future work.

Statistics and statistical methods are not considered capable of determining causality. While we may say in regression analysis that the dependent variable is some function of the independent variable, we are also careful to acknowledge that the relationship shown by regression is not proof of causality. Causality is determined by practitioners of a particular science. Statistical methods are used to confirm the hypotheses of other science disciplines. The results of the statistical analysis should either confirm an existing hypothesis, including causality, or disprove it and thus lead to the formation of new hypotheses. Thus, statistical methods should confirm or refute hypotheses, but only the practitioner in the area of science under study will clearly understand the conclusions of the statistical test. Further, a statistical test or tests should confirm or refute the expectations of the practitioner as to the impact of various changes in the manufacturing process on the process and product characteristics being measured. These expectations would include the presence of outliers and shifts in the process. If the method fails to identify such outliers to confirm or refute the expectations or assumptions of the practitioner, one can draw erroneous conclusions that the items of interest have no effect. Such incorrect conclusions can lead to further lack of control of material or processes as the necessary changes to compensate for shifts in the process and other outliers will not be made and problems can continue. One would be classifying everything as a common cause, rather than recognizing special as well as common causes of variation. The inability to make such determination leads to failure to understand the system. The use of extremely broad control limits and the inability to recognize small changes in the process or product characteristic average are the downfall of the control charts of individuals because the method does not locate many potential outliers and/or process shifts. Exponentially weighted moving average charts have some advantages and are recognized for their ability to detect small shifts in the sample average. However, what if the sample average is not shifting?

If the sample average is not shifting, many changes in the item being measured will go unrecognized by control charts of individuals with a moving range. The observer will fail to identify many changes that should have been found. Thus, the control chart method will be judged unsatisfactory because it does not provide sufficient understanding of the system. What is needed is a method that will identify the small changes that can take place without a process average shift. Of course, small shifts in the process average must also be found. This research has proposed the use of a specific time series method in conjunction with the method of Chen and Liu for outlier identification to perform such a task. The proposal was also that rather than seeking to determine the particular time series of the elements under study, one would use the ARMA (1,1) time series model as an approximation.

6.2 Results of the Research

As have been clearly shown in Chapters 4 and 5, the use of ARMA (1,1) as the model time series in conjunction with the Chen and Liu (1993) method for outlier identification finds many potential outliers. In general, the proposed method identifies either exactly or within +/- 1 observation almost everything found with 2.0 standard deviations limits on a control chart of individuals with a moving range of 2. The only times that there were problems was with very early data points, generally within the first five observations. The control chart is backward looking and does not identify these observations as out of control until everything after them has been included in the mathematical calculations. The time series is forward looking and compares the actual value with an expected value. Thus, the ARMA (1,1), or any time series model, will not go back and look at the early observations. As it is fairly commonly recognized that such early observations are often not characteristic of the process, such situations should be discarded in terms of viewing the general worth of the proposed method.

The proposed method, ARMA (1,1) as the approximation of the time series of the data, showed itself capable of finding many potential outliers in process data, product data, and business data. The application of statistical techniques to business data has been low because of the lack of response from the traditional models. The use of a time series approach allows more information to be found in the data and thus makes it a more valuable tool in the analysis of such business data.

In terms of the product data, the proposed time series approach found many more changes in the system than were ever identified with a control chart. The use of control charts for such product data has not met with a lot of success over the years because one could not identify changes and shifts in the system that a practitioner knew were there but could not clearly identify. Thus, the proposed method clearly aids in the understanding of many product systems that could not be understood well in the past.

It can be suggested that the many potential outliers identified by the ARMA (1,1) method are in fact false alarms. A false alarm is an alpha error. The limit of 2.0 standard deviations has an alpha error rate of less than 5%. A limit of 3.0 standard deviations has an alpha error rate of well less than 1%. On the other hand, the beta error for the 2.0 standard deviations limit will be less than that for the 3.0 standard deviation limits. In general, statistical methods commonly use the alpha level of 5% as a reasonable balance of alpha and beta errors. Thus, the choice of a limit of 2.0 standard deviations is appropriate as a generally acceptable statistical limit

As noted previously, one major concern with the chart of individuals with a moving range is the overestimation of the variation in the process due to the comparison of a single value with only those values prior to it and after it. This method of calculating the variation results in larger than expected values for the range and the resulting control limits where are calculated based on the average range value. With wider limits, fewer data points will be found to be out of control or beyond the limits. This inability to locate data points probably representing special causes is a weakness in this method of control charting. In addition, the process average is fixed in the control chart calculations and so does not take into account the continuing fluctuations, even if small, of the process with time.

On the other hand, the ARMA (1,1) model, as with any time series, is predicting where the next data value will fall, based upon previous values, and is therefore comparing a moving process average with the new data point rather than using a fixed process average as the basis of comparison. The joint estimation technique of Chen and

Liu (1993) is iterative in that the limits are calculated from the original data set, time series, and outliers detected from these limits. Then, the series is adjusted for the effects of the outliers and the limits again calculated. The data are then treated until no outliers are determined. The final adjusted series is then used to examine the data and locate and identify the outliers. At the same limits of detection, the number of standard deviations used to define an outlier, the ARMA (1,1) model finds more changes in the process because the comparison is made to a moving process average instead of a constant process average with the control chart method. It is also well to remember that the process average predicted for the ARMA (1,1) depends upon prior data values. The process average for the control chart depends on both prior and future data values because all of the values are used in the calculations, not just those prior to the event.

Thus, because the time series model more accurately predicts the expected process average for a given point in time, the time series model can more accurately identify outliers and provide a richer study of the process at hand.

If more outliers are detected, however, are they real or false alarms? One of the problems we have is the lack of a standard method that is mathematically appropriate and against which we can compare other methods of analysis. Thus, while the comparison of the out of control points from a control chart of individuals and the ARMA (1,1) model outliers is made, there is not proof that the control chart method is mathematically correct and the outliers or out of control points a reasonable basis for comparison. Rather, we would assume that it is not mathematically appropriate, based on previously reported work that showed most data sets are actually time series rather than IID data sets. Thus, while we can compare what is found with one method to what is located with the other method, we do not have a firm basis for concluding either is completely correct mathematically. The problems with the control charts and the IID assumption have been discussed previously. We have also noted that the ARMA (1,1) is being used as an approximation of the actual time series model which remains undefined. We are probably comparing the results from two approximation techniques – the control chart which is known to be mathematically flawed and the ARMA (1,1) which is stated as an approximation

In the final analysis, one is looking for a method that will locate and identify changes to a process, product, or business situation. Methods that fail to identify changes that would be expected to be found, based on the knowledge of practitioners in the field, are not well received and will not be used. The time series approach, using the ARMA (1,1) model as the first approximation, clearly locates and identifies outliers using the Chen and Liu method of Joint Estimation. The proposed method is clearly superior to other methods that are not time series based because in every data set studied, the proposed method locates more shifts and changes in the system than the control chart of individuals with a moving range method. It should be understood that it is the identification of a change which is important rather than the absolute identity of the type of outlier.

Compared to the control chart of individuals with a moving range of 2, several more data points were identified as outliers using the ARMA (1,1) method than the control chart of individuals, once one got beyond the 5th data point. The Yield Strength data sets would be expected to show many changes and shifts in the data because of the numerous chemistry and processing groups involved. The control charts of individuals with a moving range does not locate such changes but the ARMA (1,1) method does. This pattern is recognized in all of the data sets examined by this work. Thus, it is reasonable to conclude that the time series method suggested is superior to the conventional control chart of individual method because more shifts and changes in the data are identified by the ARMA (1,1) method.

It was also possible to compare results from the specific time series with the ARMA (1,1) model approximation by using the work of Prasad, Booth, Hu, and Degligonul (1995A). The approximation method found the same outliers as reported in the previous work, but also identified several others that should be investigated, although the original work was done at 3.0 standard deviations limits rather than the 2.0 standard deviations limits used in this work. However, the approximation method is shown to be comparable in this one case, at least, to other methods presented in the past.

One of the problems with outlier, or out of control point, identification, is the use of 3.0 standard deviations limits on control charts of individuals. There is some literature that notes the use of 2.5 standard deviations limits, but this work is based on 2.0 standard deviations limits. There is recognition of the impact of 'false alarms' where one discovers too many potential problems. That situation has to be weighed against the inability to find small changes without the use of tighter limits. If one moves away from the 2.0 standard deviations limits, many slight shifts might be missed. The strength of the system is in the location and identification of changes in the system. Moving to too wide a control limit would cause less location and identification to take place and thus less understanding of the system to be had. As noted previously, charting methods can be used for analysis of a process or system, or for control of the process or system. One can clearly make the argument that for analysis tighter limits are needed to provide a better understanding of the process or system under consideration because one wants to identify everything that could be a special cause so that cause can be eliminated. On the other hand, once a process or system is released for production, wider limits may be valid if the amount of variation hidden by such limits is not detrimental to the product in the market. Thus, we may in fact want to use different levels of detection for different stages in a process or system life cycle.

As an additional argument on the use of 2.0 standard deviations limits, if one considers that a rational subgroup under Shewhart methods is around a size of 5, one should recognize the change in the distribution of such data. If one considers a population with a mean of μ and a standard deviation of σ for individual values, if samples of size 4 are taken and the sample averages are plotted, the mean will still be at μ. However, the standard error of the means will be only ½ of the original σ because the standard error of the means is σ/ square root of n, or σ/square root of 4. Thus, with control charts of small groups, we are using much tighter control limits than we are with control charts of individuals. Therefore, to achieve the same limits of detection, it is reasonable to use limits of the same order with charts of individuals and that is with limits at the 2.0 standard deviations level.

The data analyzed in this work was real world data taken from actual plant operating reports. It has all the flaws and errors one would expect to find with such real world data. However, the response of the data to the ARMA (1,1) model approximation clearly is to locate and identify many instances of changes to the system under consideration. The proposed method works well with real data and that is the true test of a method – will it work in the real world. The approximation suggested clearly does work well in the real world and locates and identifies changes in process, product, and business systems that were not be found with conventional control charting methods.

6.3 Recommendations for Future Research

The suggested process works well at 2.0 standard deviations limits. However, no work was done on the level of 'false alarms' as one would need a base point for the comparison and it was judged that conventional control charts are not really an accurate method and thus would not be appropriate as a base point against to rate the potential for 'false alarms'. Additional work on a more exacting limit might be worthwhile, however. That would allow practitioners in the field to avoid potential 'false alarms' as much as possible, while still maximizing the ability to locate and identify outliers and system changes. Such work would necessitate the careful review of current data and the application of the method to it and then a review of each identified outlier as to whether such observation was really different or not. If that could be done with multiple data sets, one could determine whether the limit of 2.0 standard deviations is appropriate or if another level is better suited to the analysis of the data.

The other area of concern is the failure of the method to identify early observations that are strange. Wright (1997) has clearly demonstrated that even at the 10^{th} observation, the method of Chen and Liu (1993) finds and identifies the outliers. It is only in the first 5 observations, or less, where the system has not 'settled down' that one finds observations not characterized by the ARMA (1,1) approximation as outliers per Chen and Liu. The observations were not judged as out of control when they occurred, but rather as history. On the other hand, the time series looks forward and the model had not been sufficiently developed to make judgments on the early observations and thus ignored them. Additional work in this area would also be helpful to practitioners in the business and manufacturing sectors.

The work reported here was of individual data values or measurements. The data itself, as previously noted, did not lend itself to the formation of rational subgroups. Thus, all the conclusions in this work apply only to such individual data point situations. No work was done on product data from discrete manufacturing processes and which can be formed into rational subgroups. Data of this form is the type conventionally examined by Shewhart control chart methods. It would seem worthwhile to compare such data using the time series method and the conventional control charts.

6.4 Conclusions

The proposed method, using ARMA (1,1) model as an approximation of the actual time series model for process, product, and business data, clearly locates and identifies outliers using the Joint Estimation Technique of Chen and Liu (1993). In this work, the method has been tested on real world data sets rather than on computer generated data and found to clearly locate and identify more observations as potentially of interest than a control chart of individuals with a moving range of 2 and limits of 2.0 standard deviations in both methods. The method has also been compared to work done by others, e.g., Prasad, Booth, Hu, Deligonul (1995A), and found to locate and identify the same observations as outliers, and find the same type of outlier. However, the proposed method also found other observations that could be of interest as well, although at 2.0 rather than 3.0 standard deviations. Clearly the proposed method locates and identifies the same or more observations of interest than other methods previously reported, and than control charts of individuals with a moving range of 2.

From the point of view of a practitioner in the field, the proposed method clearly locates and identifies many observations as of interest as potential outliers or special causes that were not found using conventional control charting techniques. The data sets studied should have shown many potential changes and shifts in the data, based upon experience with the systems being considered. The failure of conventional methods to identify such observations left one unable to fully understand the system. The proposed method clearly aids in the understanding of each of the systems studied be locating and identifying many potential outliers that should have been found. The results of the ARMA (1,1) analysis of the data sets provided information that was expected, but could not be previously confirmed. The value of any method is in whether it will explain a system to the practitioner in the field. Conventional control chart methods often fail to do so in the process, product, and business areas. The proposed method does a superior job of locating and identifying potential outliers and is thus much more valuable to the practitioner in the field.

The contribution to the field is that this method makes it more likely that a practitioner can more easily identify a potential problem so that a solution to the problem can be found if necessary.

Bibliography

Acar, W. and D. E. Booth. "Easier Quality Control: Combining Problem classification with Time Series Analysis," *Production and Inventory Management, Third Quarter*, 1987: 53-57.

Aczel, A. D. *Complete Business Statistics, Third Edition*, Chicago: Richard E. Irwin,Inc., 1996.

Alwan, L. C. and D. L. Radson. "Time-Series Investigation of Subsample Mean Charts," *IIE Transactions*, 24, No. 5 (1992): 66-80.

Alwan, L. C. and H. V. Roberts. "Time-Series Modeling for Statistical Process Control," *Journal of Business & Economics Statistics*, 6, no. 1 (1988): 87-95.

Alwan, L. C. and H. V. Roberts. "The Problem of Misplaced Control Limits," *Applied Statistics*, 44, no. 3 (1995): 269-278.

Anderson, D. R., D. J. Sweeney, and T. A. Williams. *Statistics for Business and Economics, Eighth Edition*, Cincinnati: South-Western, 2002.

Atienza, O. O., L. C. Tang, and B. W. Ang. "A SPC Procedure for Detecting Level Shifts of Autocorrelated Processes," *Journal of Quality Technology*, 30, no. 4 (1998): 340-351.

Beckman, R. J. and R. D. Cook. "Outlier..........s," *Technometrics*, 25, no. 2 (1983): 119-145.

Booth, D. E. Some Applications of Robust Statistical Methods to Analytical Chemistry, Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1984.

Booth, D. E., Acar, W., Isenhour, T. L., and Ahkam, S. "Robust Time Series Models and Statistical Process Control," Industrial Mathematics, 40 (Part 1) (1990: 73-91.

Box, G. E. P., D. E. Coleman, and R. V. Baxley, Jr. "A Comparison of Statistical Process Control and Engineering Process Control." *Journal of Quality Technology*, 29, no. 2 (1997): 128-130.

Box, G. E. P. and G. M. Jenkins. *Time Series Analysis forecasting and control, Revised Edition*, San Francisco: Holden-Day, 1976.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis, Forecasting and Control, Third Edition*, Englewood Cliffs: Prentice-Hall, 1994.

Chang, I., G. C. Tiao, and C. Chen. "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, no. 2 (1988): 193-204.

Chen, C. and L. Liu. "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, no. 421 (1993): 284-297.

Chen, C., L. Liu, and G. B. Hudak. "*Modeling and Forecasting Time Series in the Presence of Outliers and Missing Data*," Oak Brook: Working Paper No., 128, Scientific Computing Associates Corp., 1992.

Chernick, M. R., D. J. Downing, and D. H. Pike. "Detecting Outliers in Time Series Data," *Journal of the American Statistical Association*, 77, no. 380 (1982): 743-747.

Crowder, S. V., D. M. Hawkins, M. R. Reynolds, Jr., and E. Yashchin. "Process Control and Statistical Inference," *Journal of Quality Technology*, 29, no. 2 (1997): 134-139.

Deming, W. E. *Quality, Productivity, and Competitive Position*, Cambridge: Massachusetts Institute of Technology, 1982.

Dielman, T. E. *Applied Regression Analysis for Business and Economics*, Boston: PWS-Kent, 1991.

Faltin, F. W., C. M. Mastrangelo, G. C. Runger, and T. P. Ryan. "Considerations in the Monitoring of Autocorrelated and Independent Data," *Journal of Quality Technology*, 29, no. 2 (1997): 131-133.

Gong, L., J. Wushong, and K. Tang. "Using On-Line Sensors in Statistical Process Control," *Management Science*, 43, no, 7 (1997): 1017-1028.

Grant, E. L. and R. S. Leavenworth. *Statistical Quality Control, Fifth Edition*, New York: McGraw-Hill, 1980.

Grznar, J., D. Booth, and P. Sebastian. "The Use of Robust Smoothers in Nuclear Material Safeguards," *Journal of Chemical Information and Computer Science*, 37 (1997A): 236-240.

Grznar, J., D. E. Booth, and P. Sebastian. "A Robust Smoothing Approach to Statistical Process Control," *Journal of Chemical Information and Computer Science*, 37 (1997B): 241-248.

Guerrero, V. M. "Time-series analysis supported by Power Transformations," *Journal of Forecasting*, 12 (1993): 34-48.

Hamburg, J. H., D. E. Booth, and G. J. Weinroth. "A Neural Network Approach to Monitoring the Quality of Production Processes," Submitted for publication in the *International Journal of Applied Quality Management* (1997).

Hamburg, J. H., D. E. Booth, and G. J. Weinroth. "A Neural Network Approach to the Detection of Nuclear Material Losses," *Journal of Chemical Information and Computer Science*, 36 (1996): 544-553.

Hand, D. J., F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski. *A Handbook of Small Data Sets*, London: Chapman & Hall, 1994.

Hillier, F. S. "X-Bar and R-Chart Control Limits Based on a Small Number of Subgroups," *Journal of Quality Technology*, 1, no. 1 (1969): 17-26.

Hunter, J. S. "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, 18, no. 4 (1986): 203-210.

Ishikawa, K. *Guide to Quality Control, Second Revised Edition, Eighth Printing*, Tokyo: Asian Productivity Organization, 1983.

Jiang, W., K-L. Tsui, and W. H. Woodall. "A New SPC Monitoring Method: The ARMA Chart," *Technometrics*, 42, no. 4 (2000): 399-410.

Johnson, R. A. and M. Bagshaw. "The Effect of Serial Correlation on the Performance of CUSUM Tests," Technometrics, 16, no. 1 (1994): 103-112.

LTV Steel. *Policy Guidelines for Control Charts, Fourth Edition*, 1996.

Lu, C-W. and M. R. Reynolds, Jr. "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology*, 31, no. 2 (1999): 166-188.

Lu, C-W. and M. R. Reynolds, Jr. "Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes," *Journal of Quality Technology*, 31, no. 3 (1999): 259-274.

Lucas, J. M. and M. S. Saccucci. "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, no. 1 (1990): 1-16.

Luceno, A. "Detecting possibly non-consecutive outliers in industrial time series," *Journal of the Royal Statistical Society*, 9, part 2 (1998): 295-310.

Montgomery, D. C. *Introduction to Statistical Quality Control, Second Edition*, New York: John Wiley & Sons, 1991

Montgomery, D. C. and C. M. Mastrangelo. "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology*, 23, no. 3 (1991): 179-193.

Neter, J., W. Wasserman, and M. H. Kutner. *Applied Linear Statistical Models, Third Edition*, Homewood, IL: Irwin, 1990.

Prasad, S., D. Booth, M. Hu, and S. Deligonul. "Modeling Autocorrelated Production Processes," *The Journal of the Industrial Mathematics Society*, 45, Part 2 (1995A): 105-122.

Prasad, S., D. Booth, M. Y. Hu, and S. Deligonul. "The Detection of Nuclear Materials Losses," *Decision Sciences*, 26, no. 2 (1995B): 265-281.

Quesenberry, C. P. "Screening Outliers in Normal Process Control Data with Uniform Residuals," *Journal of Quality Technology*, 18, no. 4 (1986): 226-233.

Quesenberry, C. P. "Screening Outliers in Process Control Regression Data with Uniform Residuals, II," Journal of Quality Technology, 18, no. 4 (1990): 226-233.

Reynolds, M. R., Jr., J. C. Arnold, and J. W. Baik. "Variable Sampling Interval X Bar Charts in the Presence of Correlation," *Journal of Quality Technology*, 28, no. 1 (1996):12-30.

Roberts, S. W. "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, no. 3 (1959): 239-250.

Runger, G. C. and T. R. Willemain. "Model-Based and Model-Free Control of Autocorrelated Processes," *Journal of Quality Technology*, 27, no. 4 (1995): 283-292.

Sebastian, P. R., D. E. Booth, and M. Y. Hu. "Using Polynomial Smoothing and Data Bounding for the Detection of Adverse Process Changes in a Chemical Process," *Journal of Chemical Information and Computer Science*, 34 (1994): 881-889.

Sebastian, P. R., D. E. Booth, and M. Y. Hu. "Using Polynomial Smoothing and Data Bounding for the Detection of Nuclear Material Diversions and Losses," *Journal of Chemical Information and Computer Science*, 35 (1995): 442-450.

Liu, L-M and G. B. Hudak. *Forecasting and Time Series Analysis Using the SCA Statistical System, Volume I*, Oak Brook, IL: Scientific Computing Associates Corp., 1994.

Shewhart, W. A. *Economic Control of Quality of Manufactured Product*, New York: D. Van Nostrand, 1931.

Shewhart, W. A. *From the Viewpoint of Quality Control*, Washington: The Graduate School, The Department of Agriculture, 1939.

Traver, R. W. "Pre-Control – A Good Alternative to X Bar – R Charts," *Quality Progress*, September 1985: 11-14.

Tsay, R. S. "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7 (1988): 1-20.

Vasilopoulos, A. V. and A. P. Stamboulis. "Modification of Control Chart Limits In the Presence of Data Correlation," *Journal of Quality Technology*, 10, no. 1 (1978): 20-30.

Wardell, D. G., H. Moskowitz, and R. D. Plante. "Control Charts in the Presence of Data Correlation," *Management Science*, 38, no. 8 (1992): 1084-1105.

Wardell, D. G., H. Moskowitz, and R. D. Plante. "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes," *Technometrics*, 36, no. 1 (1994): 3-17.

Woodall, W. H. and H. D. Maragah. "Discussion of Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements by Lucas, J. M. and M. S. Saccucci," *Technometrics*, 32, no. 1 (1990): 17-18.

Walker, E., J. W. Philpot, and J. Clement. "False Signal Rates for the Shewhart Control Chart with Supplementary Runs Tests," *Journal of Quality Technology*, 23, no. 3 (1991): 247-252.

Wright, C. M. Effectiveness of Joint Estimation Outlier Detection Method for Short Time Series with Quality Control Applications, PhD Dissertation, Kent State University, 1997.

Wright, C. M., M. Y. Hu, and D. E. Booth. "Effectiveness of Joint Estimation when The Outlier is the Last Observation in an Autocorrelated Short Time Series," Submitted for publication (1999).

Yashchin, E. "Performance of CUSUM Control Schemes for Serially Correlated Observations," *Technometrics*, 35, no. 1 (1993): 37-52.

Zhang, N. F. "A Statistical Control Chart for Stationary Process Data," *Technometrics*, 40, no. 1 (1998): 24-38.

Appendices

Chemyld Data

Observation	AR(1)			ARMA(1,1)		
	Sigma 2.0	Sigma 2.5	Sigma 3.0	Sigma 2.0	Sigma 2.5	Sigma 3.0
7	IO					
11	IO	IO		IO	IO	
17	AO	AO		AO	TC	
19					IO	
20	TC	TC				
29	AO	AO		AO	AO	
32	TC	TC		TC	TC	
34	AO					
39	AO	AO		AO	AO	
45	TC					
51	IO					
57	IO	AO		AO	AO	
58	AO			AO		
61	IO					
70	AO	AO		AO	AO	

Sheet-Like Material Data

OBS.	AR(1)			1/AR(1)			ARMA(1,1)		
	S. 2.0	S. 2.5	S. 3.0	S. 2.0	S. 2.5	S. 3.0	S 2.0	S. 2.5	S. 3.0
6	TC	IO		AO			IO		
7	AO	AO		AO	AO				
8							TC		
10	AO								
13	TC			TC					
16							IO		
17	AO	AO					AO	AO	
24	AO								
25							IO		
26	AO	AO		AO					
27							IO		
29	TC	IO	IO	TC	TC	TC	TC	IO	TC
31	LS						IO		
32				TC					
34	AO								
35							IO		
36	IO	TC		TC			AO	TC	
37							IO		
38	AO	AO							
43	AO						AO		
47				TC					
50							AO		
51	IO								
53	AO	AO					AO		
57	IO			TC			AO		
60	AO						AO		
64	TC						IO		
65	AO	AO	AO	TC	TC		AO	AO	AO
69	IO	TC	TC	TC	TC	TC	TC	TC	TC

Sheet-Like Material Data

Continued

OBS.	AR(1)			1/AR(1)			ARMA(1,1)		
	S. 2.0	S. 2.5	S. 3.0	S. 2.0	S. 2.5	S. 3.0	S 2.0	S. 2.5	S. 3.0
74				TC					
76	AO						AO		
80	IO			TC			TC		
82	IO						IO		
83	TC	IO	TC	TC	TC	TC	IO	IO	TC
84	LS	AO						AO	
85	IO			TC			TC		
89	TC						TC		
92							IO		
93	IO			TC					
95							AO		
96	TC								

Monkey Neuron Interspike Data

OBS.	MA(1) Sigma 2.0	MA(1) Sigma 2.5	MA(1) Sigma 3.0	ARMA(1,1) Sigma 2.0	ARMA(1,1) Sigma 2.5	ARMA(1,1) Sigma 3.0
5	IO	IO	IO	IO		
6	IO	IO				
7	IO	IO	IO			
8	IO	IO	IO			
9	IO	IO	IO		IO	
11	IO	IO	IO			
13				AO	AO	AO
14	LS	LS	LS			
15	IO	LS				
25				TC	TC	
27	IO					
28	IO			IO	IO	IO
29	IO			IO	IO	IO
33				IO		
34	IO					
36				AO	AO	
37	IO					
38	IO					
39	IO					
40	IO					
41	IO					
47	TC	LS		TC	TC	
60				TC		
68	TC					
92				LS	LS	LS
97	AO	AO				

Data Type **Process** **Data Set** **Oil Concentration**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	3.2						
2	3.0				0.2		
3	2.4				0.6		
4	2.6				0.2		
5	2.5				0.1		
6	2.8				0.3		
7	2.6				0.2		
8	3.0		IO	IO	0.4		
9	2.6				0.4		
10	2.8				0.2		
11	3.0		AO	AO	0.2		
12	2.6				0.4		
13	3.4		AO	AO	0.8	IO	IO
14	2.8				0.6	AO	
15	2.8				0.0	TC	TC
16	2.8				0.0		
17	2.0		IO	IO	0.8	TC	IO
18	2.4				0.4		
19	1.6	X	AO	AO	0.8	IO	IO
20	2.4				0.8	IO	IO
21	2.4				0.0	TC	TC

Data Type **Process** Data Set **Oil Concentration**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	2.4				0.0		
23	2.4				0.0		
24	2.2	X			0.2		
25	2.8		IO	IO	0.6	IO	IO
26	2.4				0.4		
27	3.4		LS	LS	1.0	IO	
28	3.1		IO	IO	0.3	IO	
29	3.4				0.3		
30	3.4				0.0	IO	AO
31	3.2				0.2		
32	3.5	X			0.3		
33	3.2				0.3		
34	2.8		LS	LS	0.4		
35	3.0				0.2		
36	2.8				0.2		
37	2.6				0.2		
38	3.1		IO	IO	0.5		
39	2.8				0.3		
40	2.8				0.0	TC	TC
41	3.6	X	TC	TC	0.8	IO	IO
42	3.6	X			0.0		

Data Type **Process** **Data Set** **Oil Concentration**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	3.6	X	AO	AO	0.0		
44	3.2				0.4		
45	2.8		AO	AO	0.4		
46	2.2	X	AO	AO	0.6	AO	IO
47	2.9				0.7	IO	IO
48	3.0				0.1		
49	3.0				0.0	TC	TC
50	3.0				0.0		
51	3.0				0.0		
52	2.8				0.2		
53	3.0				0.2		
54	3.0				0.0	LS	TC
55	3.0				0.0		
56	2.8				0.2		
57	3.0				0.2		
58	3.0				0.0		
59	3.4		AO	AO	0.4		
60	2.8				0.6	IO	IO
61	2.7				0.1		
62	3.6	X	AO	AO	0.9	IO	IO
63	2.6				1.0	IO	IO

Data Type **Process** **Data Set** **Oil Concentration**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	2.6				0.0		
65	2.8				0.2		
66	3.0		LS	LS	0.2		
67	3.0				0.0		
68	3.6	X	TC	AO	0.6	IO	AO
69	3.4				0.2		

Data Type **Product** **Data Set** **Sheet-Like Process**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	0.06	X							
2	0.09	X					0.03		
3	0.07	X					0.02		
4	0.02						0.05		
5	-0.01						0.03		
6	-0.06	X			IO	IO	0.05		
7	-0.08	X					0.02		
8	-0.02				TC	TC	0.06	IO	AO
9	0.01						0.03		
10	0.03						0.02		
11	0.00					AO	0.03		
12	0.01						0.01	TC	TC
13	0.04					TC	0.03		
14	0.03						0.01		
15	0.03						0.00	IO	AO
16	0.00				IO		0.03		
17	0.04				AO	AO	0.04		
18	-0.02						0.06	IO	AO
19	-0.03						0.01	TC	TC
20	-0.01						0.02		
21	0.01						0.02		
22	0.02						0.01		
23	0.02						0.00	AO	AO
24	-0.01						0.03		
25	0.03				IO	IO	0.04		

Data Type **Product** **Data Set** **Sheet-Like Process**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
26	0.06	X					0.03		
27	0.00				IO	IO	0.06	AO	AO
28	-0.02						0.02		
29	-0.13	X	IO	IO	TC	TC	0.11	IO	AO
30	-0.10	X					0.03		
31	-0.04				IO	IO	0.06	IO	AO
32	0.00						0.04		
33	0.00						0.00	AO	AO
34	-0.02						0.02		
35	0.02				IO	IO	0.04		
36	-0.05	X			AO	AO	0.07	AO	AO
37	-0.02				IO	IO	0.03		
38	-0.06	X					0.04		
39	-0.01						0.05	IO	AO
40	0.00						0.01	LS	LS
41	-0.02						0.02		
42	-0.02						0.00		
43	-0.04				AO	AO	0.02		
44	0.00						0.04		
45	0.01						0.01		
46	0.02						0.01		
47	0.03						0.01		
48	0.04						0.01		
49	0.02						0.02		
50	0.03				AO	AO	0.01		

Data Type **Product** **Data Set** **Sheet-Like Process**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
51	-0.02						0.05	IO	AO
52	-0.02						0.00		
53	0.03				AO	AO	0.05	IO	AO
54	-0.01						0.04	AO	AO
55	-0.01						0.00		
56	-0.02						0.01		
57	0.03				AO	AO	0.05	IO	AO
58	0.02					LS	0.01		
59	0.03						0.01		
60	0.00				AO	AO	0.03		
61	0.02						0.02		
62	0.02						0.00		
63	0.03						0.01		
64	-0.02				IO	IO	0.05	IO	AO
65	0.07	X	AO	AO	AO	AO	0.09	IO	AO
66	0.01						0.06	IO	AO
67	0.03						0.02		
68	0.03						0.00		
69	0.10	X	TC	AO	TC	TC	0.07	IO	AO
70	0.08	X					0.02		
71	0.06	X					0.02		
72	0.04						0.02		
73	0.04						0.00		
74	0.05	X					0.01		
75	0.03						0.02		

Data Type **Product** **Data Set** **Sheet-Like Process**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
76	0.05	X			AO	AO	0.02		
77	0.02						0.03		
78	0.01						0.01		
79	-0.01						0.02		
80	-0.04				TC	TC	0.03		
81	-0.02						0.02		
82	-0.06	X			IO	IO	0.04	AO	AO
83	-0.16	X	TC	IO	IO	IO	0.10	IO	AO
84	-0.14	X					0.02		
85	-0.03				TC	TC	0.11	IO	AO
86	-0.02						0.01		
87	-0.02						0.00		
88	-0.01						0.01		
89	0.02				TC	AO	0.03		
90	0.01						0.01		
91	0.01						0.00		
92	-0.02				IO	IO	0.03		
93	-0.05	X				LS	0.03		
94	-0.04						0.01		
95	-0.04				AO		0.00		
96	0.00						0.04	IO	AO
97	0.00						0.00		
98	-0.01						0.01		
99	-0.01						0.00		
100	-0.01						0.00		

Data Type Product **Data Set** Automatic Transmission

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	27.187								
2	27.200	X					0.013		
3	27.196						0.004		
4	27.192						0.004		
5	27.191						0.001		
6	27.194						0.003		
7	27.194						0.000		
8	27.192						0.002		
9	27.191						0.001		
10	27.189						0.002		
11	27.192						0.003		
12	27.190						0.002		
13	27.192						0.002		
14	27.190						0.002		
15	27.190						0.000		
16	27.195						0.005		TC
17	27.191						0.004		
18	27.189						0.002		
19	27.176	X	AO	AO	AO	AO	0.013	IO	AO
20	27.191						0.015		AO
21	27.192						0.001		
22	27.189						0.003		
23	27.193						0.004		
24	27.190						0.003		
25	27.191						0.001		

Data Type **Product** **Data Set** **Automatic Transmission**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
26	27.189						0.002		
27	27.190				AO	AO	0.001		
28	27.184						0.006		TC
29	27.191						0.007		
30	27.188						0.003		
31	27.193						0.005		
32	27.187						0.006		
33	27.194						0.007		AO
34	27.185				AO	AO	0.009		
35	27.189						0.004		
36	27.194						0.005		
37	27.190						0.004		
38	27.191						0.001		
39	27.192						0.001		
40	27.189						0.003		
41	27.188						0.001		
42	27.188						0.000		
43	27.201	X	IO	IO	AO	AO	0.013	IO	AO
44	27.191						0.010		
45	27.193						0.002		

Data Type **Product** **Data Set** **Bore Hole Location**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	-37								
2	-12						25		
3	7						19		
4	-12						19		
5	9						21		
6	0						9		
7	-19				LS	LS	19	TC	
8	1						20		
9	-16			LS			17		
10	-28				TC	TC	12		
11	-25						3		
12	-25						0		
13	-10						15		
14	-14						4		
15	-6						8		
16	-9						3		
17	-21						12		
18	144	X	IO	IO	IO	IO	165	AO	IO
19	184	X	IO	IO	IO	IO	40		
20	30				AO	AO	154	LS	AO
21	-13						43		
22	-8						5		
23	-20						12		
24	-28				TC	TC	8		
25	-32						4		

Data Type **Product** **Data Set** **Bore Hole Location**

Item Number	Item Value	2s Outlier Control Chart	Prasad et al Robust Method	Prasad et al Semi Robust Method	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
26	-13						19		
27	-31						18		
28	-44				TC	TC	13		
29	-38						6		
30	-25						13		
31	-2				TC	TC	23		
32	-87	X	TC	AO	AO	AO	85		
33	-26		IO		LS	LS	61		
34	-21						5		
35	-28						7		
36	-45						17		
37	-37						8		
38	-22						15		
39	-19						3		
40	-62	X	AO	AO	AO	AO	43		

Data Type Product **Data Set** Yield Strength 1

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	37.2						
2	39.8				2.6		
3	39.8				0.0		
4	38.4				1.4		
5	42.0				3.6		
6	42.4		TC	TC	0.4		
7	42.5				0.1	TC	
8	37.8		AO	AO	4.7		
9	40.8				3.0		
10	42.4				1.6		
11	38.1		AO		4.3		
12	41.5				3.4		
13	40.0				1.5		
14	37.1		AO	AO	2.9		
15	41.1				4.0		
16	41.0				0.1		
17	38.8				2.2		
18	40.2				1.4		
19	39.0				1.2		
20	38.2				0.8		
21	38.6				0.4		

Data Type **Product** **Data Set** **Yield Strength 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	37.4		TC	TC	1.2		
23	37.4				0.0		
24	37.1				0.3		
25	38.3				1.2		
26	40.7				2.4		
27	47.2	X	AO	AO	6.5		
28	40.3				6.9		
29	37.1		AO		3.2		
30	43.1	X	AO	AO	6.0		
31	34.4		AO	AO	8.7		
32	37.9				3.5		
33	39.5				1.6		
34	38.0				1.5		
35	39.0				1.0		
36	40.4				1.4		
37	38.5				1.9		
38	37.9				0.6		
39	37.0		TC	TC	0.9		
40	37.4				0.4		
41	38.0				0.6		
42	39.0				1.0		

Data Type **Product** **Data Set** **Yield Strength 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	36.0		AO	AO	3.0		
44	41.0				5.0		
45	36.9				4.1		
46	39.0				2.1		
47	36.7		TC	TC	2.3		
48	37.1				0.4		
49	41.0		TC	TC	3.9		
50	40.0				1.0		
51	40.8				0.8		
52	40.3				0.5		
53	37.9		TC	TC	2.4		
54	37.9				0.0		
55	37.0				0.9		
56	42.5		AO	AO	5.5	TC	AO
57	40.0				2.5		
58	38.0				2.0		
59	40.0				2.0		
60	41.5		AO		1.5		
61	40.0				1.5		
62	38.5				1.5		
63	38.8				0.3		

Data Type **Product** **Data Set** **Yield Strength 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	39.5				0.7		
65	39.8				0.3		
66	39.2				0.6		
67	40.0				0.8		
68	39.4				0.6		
69	39.3				0.1		
70	38.5				0.8		
71	38.1				0.4		
72	38.2				0.1		
73	39.3				1.1		
74	39.5				0.2	LS	
75	38.9				0.6		
76	39.3				0.4		
77	39.5				0.2		
78	41.8		AO	AO	2.3		
79	39.8				2		
80	36.2		AO	AO	3.6		
81	40.8				4.6		
82	40.0				0.8		
83	37.0		TC		3		
84	37.7				0.7		

Data Type **Product** **Data Set** **Yield Strength 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	41.1		TC		3.4		
86	36.7		AO	AO	4.4		
87	40.0				3.3		
88	40.3				0.3		
89	40.3				0		
90	36.7		AO	AO	3.6		
91	40.0				3.3		
92	42.1		AO	AO	2.1		
93	38.3				3.8		
94	38.5				0.2		
95	41.0		TC		2.5		
96	36.5		AO	AO	4.5		
97	40.2				3.7		
98	40.2				0		
99	40.9		TC	TC	0.7		
100	41.3				0.4		
101	39.3				2		
102	42.1				2.8		
103	42.9	X	AO	AO	0.8		
104	42.9	X			0		
105	42.5				0.4		

Data Type **Product** Data Set **Yield Strength 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
106	41.0				1.5		
107	42.0		AO	AO	1		
108	40.5				1.5		
109	39.0			LS	1.5		
110	37.9				1.1		
111	42.0		TC	TC	4.1		
112	40.0				2		
113	40.9				0.9		
114	38.2				2.7		
115	40.9				2.7		
116	38.7				2.2		
117	39.1				0.4		
118	37.8				1.3		
119	36.4		AO		1.4		
120	41.1		AO	AO	4.7		
121	32.0	X	AO	AO	9.1		
122	37.7		LS		5.7		
123	36.9				0.8		
124	37.3				0.4		
125	37.3				0		

Data Type **Product** **Data Set** **Yield Strength 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	61.2						
2	61.5				0.3		
3	61.1				0.4		
4	63.3				2.2		
5	60.8				2.5		
6	63.7		TC		2.9		
7	62.3				1.4	TC	
8	63.5				1.2		TC
9	64.3		TC		0.8		
10	63.4				0.9		
11	61.6				1.8		
12	62.1				0.5		
13	61.5				0.6		
14	61.8				0.3		
15	61.4			TC	0.4		
16	60.7				0.7		
17	60.8				0.1	TC	TC
18	61.6				0.8		
19	60.3				1.3		
20	59.5	X			0.8		
21	60.5				1.0		

Data Type **Product** **Data Set** **Yield Strength 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	62.7		LS		2.2		
23	61.5				1.2		
24	62.8				1.3		
25	61.4				1.4		
26	61.7				0.3	TC	TC
27	62.6				0.9		
28	66.3	X	TC	TC	3.7	IO	IO
29	65.1				1.2		
30	64.9				0.2	TC	TC
31	64.7				0.2		
32	61.1		IO	TC	3.6	IO	IO
33	62.7				1.6		
34	62.2				0.5		
35	60.6		IO	TC	1.6		
36	61.5				0.9		
37	61.0				0.5		TC
38	60.5				0.5		
39	63.8		AO		3.3	TC	IO
40	59.1	X	AO	AO	4.7	AO	AO
41	62.6				3.5		IO
42	61.7				0.9		

Data Type **Product** **Data Set** **Yield Strength 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	62.8				1.1		
44	63.6		IO		0.8		
45	66.4	X	IO	AO	2.8		IO
46	64.1				2.3		
47	63.3				0.8		
48	61.1		TC	TC	2.2		
49	59.7	X	IO		1.4		
50	61.3				1.6		
51	62.6				1.3		
52	62.1				0.5		
53	63.3				1.2		
54	65.6		AO	AO	2.3		
55	63.1				2.5		
56	62.3				0.8		
57	61.6				0.7		
58	63.1				1.5		
59	65.0		TC		1.9		
60	65.7			AO	0.7		
61	61.4		TC		4.3	AO	AO
62	61.6				0.2		TC
63	62.5				0.9		

Data Type **Product** **Data Set** **Yield Strength 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	62.7				0.2		
65	62.3				0.4		
66	63.5				1.2		
67	63.2				0.3		
68	61.4		AO		1.8		
69	63.8				2.4	TC	TC
70	66.6	X	TC	AO	2.8		
71	65.4				1.2		
72	66.9	X	AO	AO	1.5		
73	63.5				3.4	AO	IO
74	63.1				0.4		
75	65.0		IO		1.9		
76	66.7	X	IO	AO	1.7		
77	62.8				3.9	AO	IO
78	68.9	X	IO	AO	6.1	TC	TC
79	64.5				4.4		
80	67.2		AO	AO	2.7		
81	64.0				3.2		
82	65.6				1.6		
83	63.9				1.7		
84	65.2				1.3		

Data Type **Product** **Data Set** **Yield Strength 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	62.3		AO		2.9	AO	
86	64.1				1.8		
87	65.0				0.9		
88	63.6				1.4		
89	64.3				0.7		
90	64.9				0.6		

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	1.11	X					
2	1.05	X			0.06		
3	1.02	X			0.03		
4	1.00	X			0.02		
5	0.83				0.17		
6	0.71		IO	IO	0.12		
7	0.66				0.05		
8	0.66				0.00	IO	IO
9	0.52	X	TC	TC	0.14		
10	0.51	X			0.01		
11	0.63		IO	IO	0.12		
12	0.76		TC	TC	0.13		
13	0.94		LS	LS	0.18		
14	1.08	X	AO	AO	0.14		
15	0.93				0.15		
16	1.04	X	TC	TC	0.11		
17	0.99	X			0.05		
18	0.89				0.10		
19	0.85				0.04		
20	0.72			IO	0.13		
21	0.92		LS	LS	0.20		

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	0.89				0.03		
23	0.79				0.10		
24	0.56	X	TC	TC	0.23	AO	
25	0.63				0.07		
26	0.81		TC	TC	0.18		
27	0.75				0.06		
28	0.94		AO	AO	0.19		
29	0.76				0.18		
30	0.66		TC	TC	0.10		
31	0.65				0.01	TC	TC
32	0.74				0.09		
33	0.70				0.04		
34	0.73				0.03		
35	0.67				0.06		
36	0.56	X	TC	TC	0.11		
37	0.61				0.05		
38	0.72		IO	TC	0.11		
39	0.87		LS	LS	0.15		
40	1.01	X	IO	LS	0.14		
41	1.08	X			0.07		
42	1.03	X			0.05		

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	0.98	X			0.05		
44	0.81		IO	IO	0.17		
45	0.85				0.04		
46	0.78				0.07		
47	0.78				0.00	IO	
48	0.64		TC	IO	0.14		
49	0.62				0.02		
50	0.65				0.03		
51	0.74			TC	0.09		
52	0.88		TC	TC	0.14		
53	1.04	X	AO	AO	0.16		
54	0.81				0.23	AO	
55	0.71				0.10		
56	0.53	X	TC	TC	0.18		
57	0.50	X	AO		0.03		
58	0.59				0.09		
59	0.59				0.00	IO	
60	0.73		LS	IO	0.14		
61	0.78				0.05		
62	0.68		LS	LS	0.10		
63	0.72				0.04		

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	0.75				0.03		
65	0.54	X	IO	IO	0.21		
66	0.56				0.02	TC	
67	0.54	X			0.02		
68	0.51	X			0.03		
69	0.47	X			0.04		
70	0.60		IO	LS	0.13		
71	0.72		IO	LS	0.12		
72	0.80				0.08		
73	0.78				0.02		
74	0.73				0.05		
75	0.78				0.05		
76	0.90		AO	AO	0.12		
77	0.76				0.14		
78	0.61		TC	LS	0.15		
79	0.95		AO	AO	0.34	TC	TC
80	0.57				0.38	AO	
81	0.62				0.05	IO	AO
82	0.71	X			0.09		
83	0.50		AO	AO	0.21		
84	0.78				0.28	AO	AO

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	0.84				0.06		
86	0.79	X			0.05		
87	1.11		TC	IO	0.32	IO	AO
88	0.94			LS	0.17		
89	0.82			TC	0.12		
90	1.07	X	AO	AO	0.25	IO	AO
91	0.81				0.26	AO	AO
92	0.82				0.01	TC	
93	0.85				0.03		
94	0.86				0.01		
95	0.48	X	TC	TC	0.38	TC	TC
96	0.60				0.12		
97	0.70		AO	AO	0.10		
98	0.89		TC	TC	0.19		
99	1.03	X	TC	TC	0.14		
100	0.97	X			0.06		
101	0.84		AO	AO	0.13		
102	0.92				0.08		
103	0.83				0.09		
104	0.75				0.08		
105	1.00	X	AO	AO	0.25	AO	AO

Data Type **Product** **Data Set** **Silicon Content**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
106	0.67				0.33	AO	IO
107	0.68				0.01	AO	
108	0.76		AO	AO	0.08		
109	0.69				0.07		
110	0.75				0.06		
111	0.75				0.00		
112	0.84		TC	TC	0.09		
113	0.83				0.01		
114	0.76				0.07		
115	0.64		AO	AO	0.12		
116	0.81				0.17		
117	0.96	X	TC	TC	0.15		
118	1.01	X	TC	TC	0.05		
119	0.92				0.09		
120	0.83				0.09		
121	0.67		LS	LS	0.16		
122	0.36	X	LS	LS	0.31	AO	AO
123	0.24	X	LS	AO	0.12		
124	0.18	X			0.06		
125	0.13	X			0.05		

Data Type **Product** **Data Set** **Ash Percent**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	9.33						
2	9.30				0.03		
3	8.30				1.00		
4	8.17				0.13		
5	8.55				0.38		
6	8.16				0.39		
7	8.13				0.03	`AO	AO
8	9.24		IO	AO	1.11	IO	IO
9	8.51				0.73		
10	8.65				0.14	AO	AO
11	7.84		AO		0.81		
12	8.63				0.79		
13	8.90				0.27	AO	AO
14	8.00		LS		0.90		
15	9.53		IO	IO	1.53	IO	IO
16	8.44				1.09		
17	8.58				0.14	IO	IO
18	10.00		TC	TC	1.42	AO	AO
19	10.34	X	LS	LS	0.34	IO	IO
20	9.67				0.67		
21	9.81				0.14	AO	AO

Data Type **Product** **Data Set** **Ash Percent**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	8.46		AO	AO	1.35	TC	TC
23	9.77				1.31		
24	8.83				0.94		
25	7.82		TC	TC	1.01		
26	7.94				0.12	TC	TC
27	8.24				0.30		
28	8.76				0.52		
29	9.02				0.26		
30	9.94		AO	IO	0.92	AO	AO
31	8.43		AO	AO	1.51	AO	AO
32	8.91				0.48		
33	9.32				0.41		
34	9.78		IO		0.46		
35	9.12				0.66		
36	8.23		AO	AO	0.89		
37	9.56				1.33	AO	AO
38	7.77		TC	TC	1.79	AO	AO
39	8.55				0.78		
40	8.45				0.10	IO	IO
41	8.10		TC	IO	0.35		
42	8.50				0.40		

Data Type **Product** **Data Set** **Ash Percent**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	8.96				0.46		
44	8.14		AO	AO	0.82		
45	8.84				0.70		
46	9.00				0.16	LS	LS
47	9.45				0.45		
48	9.78				0.33		
49	8.00		LS	LS	1.78	AO	AO
50	8.00				0.00		
51	8.22				0.22		
52	7.86				0.36		
53	7.88				0.02		
54	8.15				0.27		
55	9.21		TC	TC	1.06	IO	IO
56	8.81				0.40		
57	8.90				0.09		
58	8.14				0.76	AO	AO
59	9.88		TC	TC	1.74	IO	IO
60	9.55				0.33		
61	8.42		AO		1.13	IO	IO
62	11.85	X	TC	TC	3.43	IO	IO
63	10.94	X			0.91		

Data Type **Product** **Data Set** **Ash Percent**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	9.20		TC	AO	1.74	AO	AO
65	9.10				0.10	IO	IO
66	8.09		TC	AO	1.01		
67	9.18		IO		1.09		
68	8.22				0.96		
69	8.66				0.44	AO	AO
70	9.84				1.18		
71	7.16	X	LS	AO	2.68	AO	AO
72	7.08	X			0.08	AO	UUU

Data Type **Business** **Data Set** **Rejection Rate 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	11.1						
2	12.0				0.9		
3	12.6				0.6		
4	29.5				16.9		
5	17.9				11.6		
6	12.1				5.8		
7	14.8				2.7		
8	15.0				0.2		
9	26.6		IO	IO	11.6		
10	11.0				15.6		
11	9.8				1.2		
12	3.6		AO	AO	6.2		
13	13.8				10.2		
14	12.9				0.9		
15	28.2		AO	AO	15.3		
16	12.4				15.8		
17	8.9				3.5		
18	9.9				1		
19	23.2		AO	AO	13.3		
20	19.1				4.1		
21	14.0				5.1		
22	9.0				5		

Data Type **Business** **Data Set** **Rejection Rate 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	16.0						
2	5.6				10.4		
3	19.4				13.8		
4	16.3				3.1		
5	13.3				3		
6	16.0				2.7		
7	22.9				6.9		
8	21.9				1		
9	6.8			LS	15.1		
10	9.1				2.3		AO
11	15.5				6.4		
12	5.8				9.7		
13	28.6	X	IO	AO	22.8	IO	LS
14	9.2				19.4	IO	
15	10.8				1.6		
16	6.1				4.7		
17	14.3				8.2		
18	12.6				1.7		
19	8.7				3.9		
20	16.7				8		
21	6.8				9.9		
22							

Data Type **Business** **Data Set** **Rejection Rate 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	9.1						
2	15.5				6.4		
3	12.9				2.6		
4	12.0				0.9		
5	11.2				0.8		
6	30.8	X	IO	AO	19.6		
7	14.5				16.3		
8	7.9		TC	TC	6.6		
9	9.0				1.1		
10	15.8		AO	AO	6.8		
11	11.7				4.1		
12	11.0			IO	0.7		
13	25.8	X	AO	AO	14.8	IO	IO
14	10.5				15.3		
15	17.4		TC		6.9		
16	21.4		TC	TC	4		
17	21.4				0		
18	6.0		AO	AO	15.4	TC	AO
19	15.6				9.6		
20	18.3				2.7		
21	15.9				2.4		

Data Type **Business** **Data Set** **Rejection Rate 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	7.4		LS	AO	8.5		
23	10.8				3.4		
24	8.4				2.4		
25	7.5				0.9		

Data Type **<u>Business</u>** **Data Set** **<u>Rejection Rate 4</u>**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	12.7						
2	13.2				0.5		
3	20.3				7.1		
4	7.5				12.8		
5	8.2				0.7		
6	19.5				11.3		
7	8.8		TC		10.7		
8	13.1				4.3		
9	14.3				1.2		TC
10	11.0				3.3		
11	12.4				1.4		
12	9.5				2.9		
13	14.6				5.1		
14	16.8				2.2		
15	8.2				8.6		
16	6.7				1.5		
17	12.6				5.9		
18	13.8				1.2		
19	58.5	X	AO	AO	44.7	TC	
20	26.5		IO	IO	32		TC
21	21.0		IO		5.5	IO	AO

Data Type **Business** Data Set **Rejection Rate 4**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	7.0			LS	14		
23	2.2		LS		4.8		
24	2.8				0.6		
25	0.2				2.6		

Data Type **Business** **Data Set** **Rejection Rate Total**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	11.1						
2	12.0				0.9		
3	12.6				0.6		
4	29.5	X			16.9		
5	17.9				11.6		
6	12.1				5.8		
7	14.8				2.7		TC
8	15.0				0.2		
9	26.6	X	AO	AO	11.6		
10	11.0				15.6	AO	AO
11	9.8				1.2		
12	3.6			AO	6.2		
13	13.8				10.2	AO	
14	12.9		AO		0.9		
15	28.2	X	AO	AO	15.3		
16	12.4				15.8	AO	
17	8.9				3.5		
18	9.9				1.0		
19	23.2			IO	13.3		
20	19.1				4.1	AO	
21	14.0				5.1		

Data Type **Business** **Data Set** **Rejection Rate Total**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	9.0				5.0		
23	16.0				7.0		
24	5.6			AO	10.4		
25	19.4		LS		13.8	AO	
26	16.3				3.1		TC
27	13.3				3.0		
28	16.0				2.7		
29	22.9				6.9		
30	21.9		AO		1.0	AO	
31	6.8			LS	15.1		
32	9.1				2.3	TC	
33	15.5				6.4		
34	5.8		AO		9.7		
35	28.6	X	AO	AO	22.8		IO
36	9.2				19.4	AO	IO
37	10.8				1.6		TC
38	9.1				1.7		
39	14.3				5.2		
40	12.6				1.7		
41	7.8				4.8		
42	16.7				8.9		

Data Type **Business** Data Set **Rejection Rate Total**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	6.8				9.9		
44	9.1				2.3		
45	15.5				6.4		
46	12.9				2.6		
47	12.0				0.9		
48	11.2		AO		0.8		
49	30.8	X	AO	AO	19.6		IO
50	14.50				16.3	AO	IO
51	7.9				6.6		
52	9.0				1.1		
53	15.8				6.8		
54	11.7				4.1		
55	11.0		TC		0.7		
56	25.8			IO	14.8		
57	10.5				15.3	AO	
58	17.4				6.9		
59	21.4			TC	4.0		
60	21.4		AO		0.0		
61	6.0			IO	15.4		AO
62	15.6				9.6		
63	18.3				2.7		

Data Type **Business** **Data Set** **Rejection Rate Total**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	15.9				2.4		
65	7.4				8.5		
66	10.8				3.4	TC	
67	8.4				2.4		
68	7.5				0.9		
69	12.7				5.2		
70	13.2				0.5		
71	20.3		AO	AO	7.1		
72	7.5				12.8		
73	8.2				0.7		
74	19.5			IO	11.3		
75	8.8				10.7		
76	13.1				4.3		
77	14.3				1.2		
78	11.0				3.3		
79	12.4				1.4		
80	9.5				2.9		
81	14.6				5.1		
82	16.8				2.2		
83	8.2				8.6		
84	6.7				1.5		

Data Type **Business** Data Set **Rejection Rate Total**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	12.6				5.9		
86	13.8		TC		1.2		
87	58.5	X		IO	44.7		TC
88	26.5		AO		32.0	AO	
89	21.0				5.5	TC	AO
90	7.0				14.0		
91	2.2				4.8	LS	
92	2.8				0.6		
93	0.2	X			2.6		

Data Type **Business** **Data Set** **Rejected Tons**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	16.9						
2	85.2				68.3		
3	44.3				40.9		
4	44.2				0.1		
5	92.2				48.0		
6	26.1				66.1		
7	51.9				25.8		
8	31.8				20.1		
9	99.6		IO		67.8		
10	18.8				80.8	IO	IO
11	97.1				78.3		
12	48.0				49.1		
13	20.6		LS	LS	27.4		
14	30.4				9.8		
15	22.9				7.5		
16	40.0				17.1		
17	17.7				22.3		
18	33.2				15.5		
19	24.0				9.2		
20	28.2				4.2		
21	20.4				7.8		

Data Type **Business** Data Set **Rejected Tons**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	23.0				2.6		
23	26.0				3.0		
24	19.3				6.7		
25	14.5				4.8		
26	101.6		AO	AO	87.1	TC	TC
27	153.5	X	IO	IO	51.9		
28	13.9				139.6	AO	AO
29	121.9		AO	AO	108.0		
30	17.4				104.5		
31	87.3		IO	IO	69.9		
32	10.9				76.4		
33	77.2				66.3		
34	8.0				69.2		
35	86.7		IO	IO	78.7		
36	21.1				65.6		
37	38.3				17.2		
38	9.4				28.9		
39	49.2				39.8		
40	67.9				18.7		
41	86.3		TC	TC	18.4		
42	87.4				1.1		

Data Type **Business** Data Set **Rejected Tons**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	61.7				25.7		
44	62.9				1.2		
45	151.5	X	IO	IO	88.6	IO	IO
46	10.6				140.9	AO	AO
47	76.6				66.0		
48	25.5				51.1		
49	69.8				44.3		
50	45.8				24		
51	2.9				42.9		
52	104.2		AO	AO	101.3	IO	IO
53	8.2				96.0		
54	45.8				37.6		
55	57.0				11.2		
56	126.1		IO	IO	69.1	LS	LS
57	110.5		AO	IO	15.6	AO	AO
58	6.7				103.8		
59	160.6	X	IO	IO	153.9	AO	AO
60	56.5				104.1		
61	124.0		IO	IO	67.5		
62	14.0				110.0	AO	AO
63	64.0				50.0		

Data Type **Business** **Data Set** **Rejected Tons**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	108.0		AO	AO	44.0		
65	70.8				37.2		
66	18.8				52.0		
67	117.6		AO	AO	98.8	TC	AO
68	22.3				95.3		

Data Type **Business** **Data Set** **Ordered Tons 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	1502						
2	1718				216		
3	2814				1096		
4	2121				693		
5	1814				307		
6	1638				176		
7	3199			TC	1561		
8	609		AO	AO	2590		IO
9	2907				2298		
10	3162				255	TC	
11	1690				1472		
12	3341				1651		
13	2211				1130		
14	220		AO	AO	1991		TC
15	2974				2754		
16	4833			IO	1859		
17	6612		IO	IO	1779		
18	3082				3530		AO
19	2301				781	TC	
20	3134				833		
21	4610			AO	1476		

Data Type **Business** **Data Set** **Ordered Tons 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	3109				1501		
23	1183		AO	IO	1926		
24	5309			IO	4126		AO
25	2641				2668		AO
26	2949				308	TC	
27	3274				325		
28	2987				287		
29	2477				510		
30	4311			TC	1834		
31	3658				653		
32	3279				379		
33	1301		IO	IO	1978		
34	6663		IO	AO	5362	TC	TC
35	3923				2740		
36	7857	X	AO	AO	3934		
37	2960				4897		AO
38	3021				61	TC	TC
39	3040				19		
40	3521				481		
41	3254				267		
42	4860			AO	1606		

Data Type Business **Data Set** Ordered Tons 1

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	2414		TC		2446		AO
44	2170				244	TC	
45	1933				237		
46	3039				1106		
47	2322				717		
48	1589			AO	733		
49	3580				1991		
50	2495				1085		
51	3284				789		
52	2389				895		
53	3296				907		
54	1706				1590		
55	2440				734		
56	6863	X	TC	TC	4423	IO	AO
57	6239				624		
58	2291		TC	TC	3948	IO	AO
59	2911				620		
60	800			IO	2111		
61	6164		IO	IO	5364	AO	AO
62	3177				3987		AO
63	985		TC	TC	2192		

Data Type **Business** Data Set **Ordered Tons 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	548				437		
65	1379				831		
66	3529			IO	2150		
67	4737			IO	1208		
68	3061				1676		
69	2135				926		
70	5664		IO	AO	3529	LS	TC
71	3238				2426		
72	1032			AO	2206		
73	4969			AO	3937		TC
74	9143	X	IO	IO	4174	IO	
75	4010				5133	IO	AO
76	2652				1358		
77	4600				1948		
78	1797			AO	2803		
79	3868				2071		
80	5397			AO	1529		
81	4250				1147		
82	2011			LS	2239		
83	2518				507	TC	
84	3122				604		

Data Type **Business** Data Set **Ordered Tons 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	7681	X	IO	IO	4559	IO	TC
86	2511				5170	IO	AO
87	2834				323		IO
88	2721				113	AO	
89	7072	X	IO	AO	4351	IO	AO
90	2014			AO	5058	IO	AO
91	3772				1758		
92	4189				417		
93	5213			IO	1024		
94	1638			AO	3575	IO	IO
95	3564				1926		
96	4503				939		
97	5051		TC	TC	548		
98	5210				159		
99	6545			IO	1335		
100	5538			AO	1007		
101	3020			AO	2518		YC
102	1287		AO	IO	1733		
103	2998				1711		
104	4369				1371		
105	4814			AO	445		

Data Type **Business** **Data Set** **Ordered Tons 1**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
106	8429	X	AO	AO	3615	AO	AO
107	3148				5281	AO	IO
108	3917				769		
109	846		TC	TC	3071		AO
110	2220				1374		
111	3263				1043		
112	921			IO	2342		
113	2131				1210		
114	6126		AO	AO	3995	TC	TC
115	2319				3807		
116	3694				1375		AO
117	9413	X	AO	AO	5719	TC	AO
118	2899				6514	IO	AO
119	2607				292	AO	AO
120	1164			IO	1443		
121	5420		LS	AO	4256	IO	AO
122	5552				132		

Data Type **Business** **Data Set** **Ordered Tons 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	3860						
2	3323				8955		
3	3769				1299		
4	4629				106		
5	3134				222		
6	4451		AO	AO	415		
7	1511		AO	AO	884		
8	3207				6062		
9	804		AO	AO	626		
10	1407		AO	AO	3015	LS	LS
11	3372				2347		
12	3055				3709		
13	4711		TC	TC	886		
14	4754				6728		
15	732		AO	AO	5161	AO	TC
16	4658		TC	TC	449	AO	
17	4335				1297		TC
18	5488		TC	AO	7349		
19	4617				9301		
20	1508		TC	TC	7584		AO
21	1448				6094		

Data Type **Business** **Data Set** **Ordered Tons 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	6234		AO	AO	533	AO	AO
23	3782		TC	TC	297		
24	3878				835		
25	4933		AO	AO	1083		
26	1405		AO	AO	2579	AO	AO
27	2482				1452		
28	3028				274		
29	2144				557		
30	3459				807		
31	801		AO	AO	5275		AO
32	762		AO	AO	4771		
33	3617				3466		AO
34	4636		AO	AO	65		
35	1570		AO	AO	866		AO
36	3316				2823		
37	7160	X	AO	AO	2680	AO	AO
38	7844	X	AO	AO	15284		
39	1887		LS	LS	7956	AO	AO
40	2836				2863		
41	1698				786		
42	1547		AO		1490		

Data Type **Business** Data Set **Ordered Tons 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	972				1094		
44	1739				5814		
45	2676				4319		
46	310		AO	AO	3135		
47	3169				3986		AO
48	1862				74		
49	1129		TC		4040		
50	1569				2070		
51	4536		AO	AO	2268		TC
52	1409				1559		
53	2187				1742		
54	5239		AO	AO	1235		
55	27		AO	AO	1379	TC	TC
56	2784				532		
57	8208	X	AO	AO	60	AO	AO
58	1995				1577	AO	AO
59	2561				2136		
60	4755		AO	AO	694		
61	1646				868		AO
62	160		AO	AO	1891		
63	5451		AO	AO	2302	IO	AO

Data Type **Business** **Data Set** **Ordered Tons 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	1510				995	AO	AO
65	2115				3157		
66	4331		AO	AO	390		
67	3171				1287		
68	2002				1758		
69	-73		TC	TC	4690		
70	-168				0		
71	4227		AO	AO	4811	AO	AO
72	1705				479		
73	2187				219		
74	2666				1786	LS	
75	1423				3602		
76	658		AO		5788		
77	3334		AO	AO	1885	AO	AO
78	2920				2827		
79	1388				2312		
80	1643				1601		
81	1280				1984		
82	1150				4975		
83	10276	X	AO	AO	2980	IO	AO
84	2736				2736	AO	AO

Data Type **Business** **Data Set** **Ordered Tons 2**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	2720				1848		
86	1131		TC		1568		
87	1744				173		
88	1439				1649		
89	1975				2152		
90	1661				4040		
91	3916		AO	AO	868	TC	TC
92	821		AO		3278		
93	3363		LS	LS	46		
94	4605		AO	AO	626		
95	2165				122		
96	2475				1528		
97	2448				285		
98	3858				5331		
99	3201				644		
100	3050				1926		
101	1610		TC	TC	299		
102	1936				3231		
103	8207	X	AO	AO	219	TC	AO
104	1978				1486		AO
105	3245				1210	TC	

Data Type __**Business**__ **Data Set** __**Ordered Tons 2**__

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
106	3804		AO		3878		
107	7613	X	AO	AO	877	IO	TC
108	5214		AO	AO	4648		
109	10009	X	AO	AO	1117	IO	AO
110	2899				868	AO	AO
111	2543				2004		
112	2282				442		
113	4104		AO	AO	1819		
114	2902				1133		
115	2342				952		
116	1823				2515		
117	5396		AO	AO	277	TC	AO
118	2553				650		

Data Type **Business** **Data Set** **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
1	11731	X					
2	2776				8955		
3	1477				1299		
4	1371				106		
5	1593				222		
6	2008		TC	TC	415		
7	1124				884		
8	7186			IO	6062	AO	AO
9	6560				626		
10	3545				3015		
11	5892				2347		
12	2183		TC	TC	3709		AO
13	3069				886		
14	9797	X		IO	6728	IO	IO
15	4636				5161		
16	5085				449		
17	6382				1297		
18	13731	X	IO	IO	7349	TC	TC
19	4430				9301	TC	TC
20	12014	X		IO	7584		
21	5920				6094		

Data Type **Business** Data Set **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
22	5387				533	TC	TC
23	5090				297		
24	4255		TC		835		
25	3172				1083		
26	5751				2579		
27	4299				1452		
28	4025				274		
29	4582				557		
30	3775	X			807		
31	9050			IO	5275	IO	IO
32	4279				4771		
33	7745			TC	3466		
34	7810				65		
35	6944				866		
36	4121				2823		TC
37	1441			AO	2680		
38	16725	X	IO	IO	15284	IO	IO
39	8769				7956		
40	5906				2863		
41	6692				786		
42	8182		TC		1490		

Data Type **Business** **Data Set** **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
43	9276	X		IO	1094		
44	3462				5814	IO	IO
45	7781				4319		
46	4646				3135		
47	8632			TC	3986	IO	IO
48	8706				74	AO	
49	4666				4040		
50	2596			AO	2070		
51	4864				2268		
52	6423				1559		
53	4681				1742		
54	3446				1235		
55	2067			TC	1379		
56	2599				532		
57	2539				60		
58	4116				1577		
59	1980			LS	2136		
60	3674				694		
61	1806				868		
62	3697				1891		
63	1395				2302		

Data Type **Business** **Data Set** **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
64	400	X			995		
65	3557				3157		IO
66	3167				390		
67	1880				1287		
68	3638				1758		
69	8328		TC	TC	4690	TC	TC
70	8328				0	AO	AO
71	3517				4811		
72	3038				479		
73	3257				219		
74	5043				1786		
75	8645		AO	IO	3602		IO
76	2857				5788	AO	AO
77	4742				1885		
78	1915				2827		
79	4227				2312		
80	2626				1601		
81	642				1984		
82	5617				4975	IO	IO
83	2637				2980		
84	5373				2736		

Data Type **Business** **Data Set** **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
85	3525				1848		
86	1957				1568		
87	1784				173		
88	3433				1649		
89	5585				2152		
90	1545				4040	AO	AO
91	2413				868		
92	5691		LS	TC	3278	AO	TC
93	5645				46	LS	
94	6271			TC	626		
95	6393				122		
96	7921				1528		
97	8206				285		
98	2875				5331	AO	AO
99	3519			AO	644		
100	5445				1926		
101	5146				299		
102	1915				3231	AO	AO
103	1696				219		
104	3182				1486		
105	4392				1210		

Data Type **Business** **Data Set** **Ordered Tons 3**

Item Number	Item Value	2s Outlier Control Chart	2s Outlier Data ARMA(1,1) Type	2s Outlier Data Shortened ARMA(1,1) Type	Range Value	2s Outlier Range ARMA(1,1) Type	2s Outlier Range Shortened ARMA(1,1) Type
106	8270	X	TC	LS	3878	AO	AO
107	9147				877		
108	4499				4648	AO	AO
109	5616				1117		
110	4748				868		
111	2744				2004		
112	3186				442		
113	5005				1819		
114	3872				1133		
115	4824				952		
116	2309				2515		
117	2032				277		
118	1382				650		
119	3215				1833		
120	3044				171		