<table>
<thead>
<tr>
<th>Organization Name and Address</th>
<th>10. Program Element Area &amp; Work Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate Analysis</td>
<td>61102F</td>
</tr>
<tr>
<td>Thackeray Hall</td>
<td></td>
</tr>
<tr>
<td>1970 Pittsburgh, Pittsburgh, PA 15260</td>
<td></td>
</tr>
<tr>
<td>Office of Scientific Research</td>
<td></td>
</tr>
<tr>
<td>O. Air Force</td>
<td></td>
</tr>
<tr>
<td>Force Base, DC 20332</td>
<td>Bldg 41C</td>
</tr>
<tr>
<td>11. Report Date</td>
<td>May 1987</td>
</tr>
</tbody>
</table>

**DTIC ELECTED**

OCT 01 1987

Center for Multivariate Analysis

University of Pittsburgh

**DISTRIBUTION STATEMENT A**

Approved for public release
Distribution: Unlimited
Control charts when the observations are correlated

P. R. Krishnaiah and B. Q. Miao

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh, Pittsburgh, PA 15260

Air Force Office of Scientific Research
Department of the Air Force
Bolling Air Force Base, DC 20332

Approved for public release; distribution unlimited.

Autoregressive model, control ellipse, time series, X-chart.

Traditionally, control charts are based on independently normal samples, but in practice it so happens that the samples are dependent. In this review, dependent samples are considered as ARMA time series. Also multidimensional time series samples are discussed.
CONTROL CHARTS WHEN THE OBSERVATIONS ARE CORRELATED*

P.R. Krishnaiah and B.Q. Miao
Center for Multivariate Analysis
University of Pittsburgh

May 1987
Technical Report No. 87-09

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* Research sponsored by the Air Force Office of Scientific Research (AFOSR) under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notion hereon.
CONTROL CHARTS WHEN THE OBSERVATIONS ARE CORRELATED*

ABSTRACT

Traditionally, control charts are based on independently normal samples, but in practice it so happens that the samples are dependent. In this review, dependent samples are considered as ARMA time series. Also, multidimensional time series samples are discussed.

AMS Subject classification: Primary 62N10.

Key words and phrases: Autoregressive model, control ellipse, time series, $\bar{X}$-chart.

* Research sponsored by the Air Force Office of Scientific Research (AFOSC) under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notion hereon.
1. INTRODUCTION

Traditionally, quality control charts have been designed with respect to statistical criteria only, and the control methodology is based on the independence and normality of serial samples. At first the production process is assumed to be characterized by a single in-control state. For example, if the process has one measurable quality characteristic, then the in-control state will correspond to the mean of this quality characteristic when no assignable cause is present.

Now we consider the model:

\[ X_t = \mu + \xi_t \]  

(1.1)

where \( \mu \) is a constant, \( \xi_t \) is an error and \( X_t \) is t-th observation. It is of interest to select the sample size, statistical characteristic of \( \xi_t \) and control limits so that the power of the test to detect a particular shift in the quality characteristic and the type I error probability are equal to specified values. Usually, \( \xi_t, t = 1, 2, ... \) is considered to be independently normally distributed with zero mean and common variance \( \sigma^2 \), where \( \sigma^2 \) is known or unknown. In this case, consideration of statistical criteria and practical experience have led to general guidelines for the design of control charts resulting in widespread use of samples of size 5, three-sigma control limits, and a sampling frequency of one hour for the \( \bar{X} \)-chart (see Duncan [3]). In the sequel, we set

\[
\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{in+j} \quad \text{and} \quad S_i = \left( \frac{1}{n} \sum_{j=1}^{n} (X_{in+j} - \bar{X}_i)^2 \right)^{1/2},
\]

\[
\bar{X} = \frac{1}{k} \sum_{i=1}^{k} \bar{X}_i \quad \text{and} \quad \bar{S} = \frac{1}{k} \sum_{i=1}^{k} S_i,
\]
Hereafter we use $\bar{x}$ as the number of samples. The general $3\sigma$ quantity control limits on the mean are
\[ X - \frac{3}{\sqrt{n}} A \leq \bar{x} \leq X + \frac{3}{\sqrt{n}} A, \quad A = \frac{3}{\sqrt{n}}. \quad (1.2) \]

The $A_1$ limits are
\[ X - A_1 \sigma \leq \bar{x} \leq X + A_1 \sigma, \quad A_1 = \frac{A}{C_2}. \quad (1.3) \]

When the standard deviation is known or estimated, the $B_1$ and $B_2$ limits on the standard deviation are
\[ B_1 \leq \frac{\sigma}{\bar{x}} \leq B_2. \quad (1.4) \]

The $B_3$ and $B_4$ limits are
\[ B_3 \leq \frac{\sigma}{\bar{x}} \leq B_4. \quad (1.5) \]

Where $\sigma$ is known or estimated, where
\[ B_1 = \frac{C_2}{2}, \quad (1.6) \]
\[ B_2 = \left(1 + \frac{C_2^2}{2}ight)^{1/2}, \quad (1.7) \]
\[ B_3 = 2C_2 + 3C_3, \quad (1.8) \]
\[ B_4 = 1 + \frac{3C_3}{C_2}. \quad (1.9) \]

A, $B_1 - B_4$ are tabulated in the
literature (see Grant [4]). The process will be considered under control if the estimate of the mean and the estimate of the standard deviation of the process remain within prescribed control limits above.

In practice, a number of data sets in economics, business, engineering and the natural science often are present in the form of time series. In other words, the observations are dependent, i.e., $\xi_t$'s of model (1.1) are not white noise; for example, $\{\xi_t, t = 0, \pm 1, \ldots\}$ is an autoregressive moving average (ARMA) with order $(p,q)$. So the problem is how to determine $3\sigma$ control limits. Stamboulis [7] studied AR(1) with parameter $\alpha$. Vasilopoulos [8] extended Stambouli's results to ARMA$(p,q)$ model. Vasilopoulos and Stamboulis [9] together investigated the case AR(2) = ARMA$(2,0)$. It is different from classical control factors. How different it is depends on the stochastic properties of the process. Since the method is similar, we only discuss AR(2).
2. CONTROL CHART ON THE SECOND ORDER AUTOREGRESSIVE MODEL

In model (1.1), assume \( \xi_t \) is an AR(2) model, that is

\[
\xi_t = \alpha_1 \xi_{t-1} + \alpha_2 \xi_{t-2} + \xi_t
\]  

(2.1)

where \( \{\xi_t\} \) is a white noise series with \( E\xi_t = 0 \) and \( V(\xi_t) = \sigma^2_\xi \), \( \alpha_1 \) and \( \alpha_2 \) are constants. For stationarity of AR(2), it is necessary that the roots of the characteristic equation of the AR(2) process

\[
\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 = 0
\]  

(2.2)

must lie outside the unit circle, which is equivalent to the alpha coefficients being in the triangular region:

\[
\alpha_2 + \alpha_1 < 1, \quad \alpha_2 - \alpha_1 < 1, \quad -1 < \alpha_2 < 1
\]  

(see Box and Jenkins [2]). The variance of the AR(2) process is given by

\[
\sigma^2 = \frac{(1 - \alpha_2^2)}{(1 + \alpha_2^2)} \frac{\sigma^2_\xi}{[(1 - \alpha_2^2)^2 - \alpha_1^2]}. 
\]  

(2.3)

Suppose \( \gamma_k, k = 0, \pm 1, \ldots \), are the autocovariance functions of the AR(2) process, then \( \sigma^2 \) and the variance \( \frac{\sigma^2}{\bar{x}} \) of the sample mean \( \bar{x} \) are given in terms of \( \gamma_k \) by

\[
\sigma^2 = \gamma_0, 
\]  

(2.4)

\[
\frac{\sigma^2}{\bar{x}} = \frac{1}{n} \left[ \gamma_0 + 2 \sum_{t=1}^{n-1} \frac{1}{n} (1 - \frac{1}{n}) \gamma_t \right]. 
\]  

(2.5)

In order to evaluate the control limits for \( \bar{x} \), we need to evaluate \( \frac{\sigma^2}{\bar{x}} \). This can be accomplished by (2.3)-(2.5) if \( \alpha_1, \alpha_2 \) and \( \sigma^2_\xi \) are known. Therefore, if the process variance \( \sigma^2 \) is known, the control
limits of the model described by (1.1) and (2.1) is modified to

\[ \bar{x} \pm A(a_1,a_2,n)\sigma \]  

where

\[ A(a_1,a_2,n) = \lambda^{1/2}(a_1,a_2,n) \frac{3}{\sqrt{n}}, \]  

\[ \lambda(a_1,a_2,n) = 1 + 2 \sum_{t=1}^{n-1} (1 - \frac{t}{n})b_t, \]  

\[ b_t = \gamma_t/\gamma_0. \]  

The expression of \( \lambda(a_1,a_2,n) \) is based on the expression of \( \sigma^2_\chi \).

In order to construct the \( \bar{x} \)-chart when the standard deviation is unknown, we must evaluate the auxiliary parameter \( C_2(a_1,a_2,n) \) first, which is also needed to construct control limits of the standard deviation. Since \( S^2/ES^2 \) is distributed as \( \chi^2 \), we get

\[ E(S) = C_2(a_1,a_2,n)\sigma = \sqrt{\frac{2}{n}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 - \frac{2}{n-1} \sum_{t=1}^{n-1} (1 - \frac{t}{n})b_t\right)^{1/2} \sigma. \]  

Hence,

\[ C_2(a_1,a_2,n) = C_2 \left(1 - \frac{2}{n-1} \sum_{t=1}^{n-1} (1 - \frac{t}{n})b_t\right)^{1/2}. \]  

To obtain an approximate expression for \( E(S) \), we use this expression of \( E(S) \):

\[ E(S) = E\sqrt{S^2} = \sqrt{ES^2}(1 - \text{Var} S^2) - \frac{\sqrt{ES^2}}{8(ES^2)^2} \left\{ 2n - 3 - \left( \frac{2\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \right)^2 \right\}. \]  

In general, the last term of (2.12) is smaller than the first term.

For example, let \( n = 5 \). Then the last term of (2.11) is 0.002144 \( \sqrt{ES^2} \).
so we can omit this term and get the approximate formula:

$$E(S) = \sqrt{ES^2} \left(1 - \frac{\text{Var}(S^2)}{8(ES^2)^2}\right). \quad (2.13)$$

The expected value and variance of $S^2$ can be obtained by

$$ES^2 = \gamma_0 \left(1 - \frac{\lambda(\alpha_1, \alpha_2, n)}{n}\right) \quad (2.14)$$

and

$$\text{Var}(S^2) = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{\tau=1}^{n} \gamma_{t-\tau}^2 + \frac{2}{n^2} \left[\frac{1}{n} \sum_{t=1}^{n} \sum_{\tau=1}^{n} \gamma_{t-\tau}\right]^2 - \frac{4}{3} \sum_{t=1}^{n} \sum_{\tau=1}^{n} \sum_{\gamma=1}^{n} \gamma_{t-\gamma} \gamma_{t-\gamma}. \quad (2.15)$$

But, from the expression of $\text{Var}(S^2)$, the complexity of (2.13) is not better than (2.10).

By replacing $A_1(\alpha_1, \alpha_2, n)$, $C_3(\alpha_1, \alpha_2, n)$, $B_i(\alpha_1, \alpha_2, n)$ for $A_1$, $C_3$ and $B_i$, $i = 1, 2, 3, 4$ of (1.3) and (1.6)-(1.8), respectively, the modification of control chart limits in an AR(2) model is obtained.

The substantial ranges in the values of $\lambda(\alpha_1, \alpha_2, n)$ and $C_2(\alpha_1, \alpha_2, n)$ greatly affect the control factors. Vasilopoulos and Stamboulis [9] gave an example to illustrate this result.
3. CONTROL CHARTS IN MULTIVARIATE CASE

Now we consider multivariate case. The model in this case is

\[ X_t = \mu + \xi_t \]  

where \( \mu \) is a \( m \times 1 \) constant vector, \( \{\xi_t, t = 1, 2, \ldots\} \) is a \( m \)-dimensional stationary process with zero mean vector. Set

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})', \]

where the prime means transpose of a matrix (or vector). Let \( \bar{x} \) be the global mean over several subgroups of size \( n \), and \( S \) be the pooled sample covariance. It is well-known that if \( \{\xi_t\}, t = 1, 2, \ldots \) is a series consisting of white noise with distribution \( N_m(\mu, \Lambda) \), the \( \bar{x} \)-control chart has been studied by many authors (for example, Ghare and Torgersen, Jackson and et al). From the facts that \( (\bar{x} - \mu)'(\frac{1}{n}\Lambda)^{-1}(\bar{x} - \mu) \) and \( (\bar{x} - \mu)'(\frac{1}{n}S)^{-1}(\bar{x} - \mu) \) are distributed as \( \chi_m^2 \) and Hotelling \( T^2 \)-statistics respectively, we can construct the quality control region on \( \bar{x} \) based on \( \Lambda \) is known or unknown. When \( \Lambda \) is known, the control region is

\[ D = \{\bar{x}: n(\bar{x} - \bar{x})'\Lambda^{-1}(\bar{x} - \bar{x}) \leq \chi_m^2(a)\}. \]  

(3.2)

This is an elliptical region. When \( \Lambda \) is unknown, the control region is

\[ D = \{\bar{x}: \frac{n-m+1}{m}(\bar{x} - \bar{x})'S^{-1}(\bar{x} - \bar{x}) \leq F_{m,n-m+1}(a)\}. \]  

(3.3)

But in practice, \( \{\xi_t\} \) is not generally a white noise series. When \( \{\xi_t\} \) is serially dependent and described by \( p \)-dimensional ARMA\((p,q)\) model, the control region of mean vector will be modified. For simplicity,
we will discuss $p$-dimensional AR($p$) model. Let

$$x_t = \mu + \xi_t,$$  \hspace{1cm} (3.4)

$$\xi_t = B_1\xi_{t-1} + \cdots + B_p\xi_{t-p} + \varepsilon_t,$$  \hspace{1cm} (3.5)

where $B_i, i = 1, \ldots, p,$ are $m \times m$ matrices and $\xi_t$ a white noise series with distribution $N_m(0, \Sigma), \Sigma > 0$. Furthermore, we can also generalize (3.5) to the following form:

$$\xi_t = B_1\xi_{t-1} + \cdots + B_p\xi_{t-p} + A\varepsilon_t,$$  \hspace{1cm} (3.6)

where $A: m \times r, \varepsilon_t \sim N_r(0, \Sigma)$ such that $A\Sigma A' > 0$. The model described by (3.5) and (3.6) is often met. For example, in the production of synthetic fiber the tensile strength $x_1$ and diameter $x_2$ may be equally important quality characteristics. Their fluctuations mainly result from moisture, then, in proper productive process, $(x_1, x_2)'$ may be described by (3.5) and (3.6). Here we only discuss the model (3.5) because the method treating model (3.6) is the same as (3.5).

It is well-known that the necessary condition that the AR($p$) model (3.5) is stationary is all the roots of determinant of $(\lambda^p I - \lambda^{p-1}B_1 - \cdots - \lambda B_{p-1} - B_p)$ lie within the unit circle. Set

$$\Lambda = E(x_t - \mu)(x_k - x_{t-k}''),$$

then, there also exist "multivariate Yule Walker" equation:

$$\Lambda_0 = B_1\Lambda_1 + \cdots + B_p\Lambda_p + \Sigma,$$  \hspace{1cm} (3.7)

$$\Lambda_k = B_1\Lambda_{k-1} + B_2\Lambda_{k-2} + \cdots + B_p\Lambda_{k-p}, \hspace{1cm} k > 1,$$  \hspace{1cm} (3.8)

$$\Lambda - I = \Lambda_k.$$  \hspace{1cm} (3.9)
Hence, if $B_i, i = 1, \ldots, p$, and $\Sigma$ are identified from data, then all $\Lambda_k, k = 0, 1, \ldots$, can be calculated from (3.7)-(3.9). Furthermore, the covariance matrix of $\overline{x}$ can be obtained:

$$\Lambda_{\overline{x}} = \frac{1}{n} \left( \Lambda_0 + \sum_{t=1}^{n-1} (1 - \frac{1}{n}) (\Lambda_t + \Lambda_t') \right),$$

(3.10)

where $\Lambda_0$ is the covariance matrix of $x_t$.

Since $(\overline{x} - \mu)' \Lambda_{\overline{x}}^{-1} (\overline{x} - \mu)$ is a $\chi^2$ distribution with degrees of freedom $m$, we can get the control region of mean vector within an elliptical region:

$$D = \{ \overline{x}: (\overline{x} - \mu)' \Lambda_{\overline{x}}^{-1} (\overline{x} - \mu) \leq \chi^2_m(\alpha) \}.$$  

(3.11)

Notice that if $\xi_t = \varepsilon_t$ in (3.5), i.e., our process is classical, then the control region described by (3.11) is the same as the one described by (3.2). When $p \neq 0$, these elliptical regions described by (3.2) and (3.11) are different from the lengths and directions of their major axes.
REFERENCES


