Model based control charts in stage 1 quality control

Alex J. Koning
Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738
NL-3000 DR Rotterdam
The Netherlands
koning@few.eur.nl

Econometric Institute Report EI-9958/A

Abstract

In this paper a general method of constructing control charts for preliminary analysis of individual observations is presented, which is based on recursive score residuals. A simulation study shows that certain implementations of these charts are highly effective in detecting assignable causes.

Key Words: *Cusum chart, control chart, statistical process control.*

Introduction

When a process is new or just has been modified, one is faced with two seperate problems. Firstly, it is unknown whether the process is in-control; secondly, the parameters governing the in-control statistical behavior of the process are unknown. Stage 1 quality control aims at solving both problems simultaneously on the basis of historical data. If stage 1 quality control yields the conclusion that the process is indeed in-control, then stage 1 is followed by stage 2 quality control. Stage 2 quality control aims at detecting departures from the in-control state on the basis of the in-control parameter estimates obtained in stage 1 and prospective data.

Additional information concerning stage 1 and stage 2 quality control may be found in the introduction in Sullivan and Woodall (1996) and the introduction in Koning and Does (1999). In Sullivan and Woodall (1996) the importance of the detection of assignable causes is emphasized. In

Koning and Does (1999) the relevance of stage 1 quality control for current manufacturing processes is underlined.

Stage 1 quality control relies heavily on control charts, which often are constructed following either the likelihood ratio approach or the recursive residual approach.

In the likelihood ratio approach [cf. Quandt (1960), Hinkley (1971), Worsley (1986), Sullivan and Woodall (1996)] the full sample of historical data $X_1, \ldots, X_n$ is divided into two subsamples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$, and these two subsamples are compared by computing a two-sample likelihood ratio test statistic Λ_k under the given model. The procedure is repeated for every $1 \leq k < n$, and Λ_k is plotted is versus k. Finally, a horizontal decision line is added to the chart and used to asses whether the process may be classified as being in-control.

In the recursive residuals approach [Brown, Durbin and Evans (1975), Hawkins (1987), Quesenberry (1991, 1995), Del Castillo and Montgomery (1994), Koning and Does (1999), Koning (1999)] the unknown in-control parameters are eliminated by cleverly transforming the historical data so as to obtain a number of independent random variables which have distributions [virtually] not depending on the unknown in-control parameters but reacting to out-of-control conditions nevertheless. The transformed data are used to asses whether the original data are in-control. In Koning and Does (1999) and Koning (1999) the recursive residual approach is combined with the theory of uniformly most powerful tests [cf. Lehmann (1994)] so as to obtain control charts which are optimal for detecting a particular form of trend in normally distributed random variables. The chart proposed in Koning and Does (1999), which was developed for detecting linear trend in normally distributed random variables, showed the best properties to detect linear trends and shifts in the data in a comparison with various other charts [including the LRT-chart of Sullivan and Woodall (1996) and the Cusum chart of Brown, Durbin and Evans (1975)].

The "recursive residuals" literature mentioned above concentrates on the situation where the observations follow a normal distribution. In contrast to the likelihood ratio approach there is no obvious extension of the recursive residual approach to the nonnormal situations. However, in this paper a possible way of extending the recursive residual approach to nonnormal situations is presented.

The structure of this paper is as follows. First, we introduce the concept of recursive score residuals, which form the basis of the extension of the recursive residual approach to nonnormal situations. After describing how recursive score residual arise naturally in detecting assignable causes in exponential families, we generalize to distributions which do not necessarily belong to an exponential family, and discuss several ways of implementing the charts. Finally, we focus on the normal distribution to select an

implementation which is highly effective according to simulation results, and apply this implementation to the data sets in Sullivan and Woodall (1996).

Recursive score residuals

In this section we introduce the concept of recursive score residuals, which will allow us to extend the recursive residual approach to nonnormal situations.

Let the random variable X have density function $f(x;\theta)$, where the parameter θ belongs to some parameter space $\Theta \subset I\!R^k$. Define the classical score function by

$$\rho(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta).$$

Note that ρ is a vector-valued function, of the same dimension k as the parameter θ.

The individual score $\rho(X;\theta)$ is obtained by transforming the random variable X via the classical score function. It is well-known that the mathematical expectation $\mathcal{E}_\theta \rho(X;\theta)$ is equal to zero; moreover, if the $k \times k$ Fisher information matrix Σ exists, then

$$\mathcal{E}_\theta \rho(X;\theta)\rho(X;\theta)^T = \Sigma$$

[cf. Section 5.1.2 in Lindsey (1996)]. Recall that the Fisher information matrix depends on θ.

Now, consider the situation in which we have observed n independent copies $X_1, \ldots, X_n$ of X. Define

$$Y_i = \sqrt{\frac{i-1}{i}} \left\{ \rho(X_i;\theta) - \frac{1}{i-1} \sum_{j=1}^{i-1} \rho(X_j;\theta) \right\}.$$

One may view Y_i as a recursive residual formed from the individual scores. We shall refer to Y_i as the i^{th} recursive score residual. Note that Y_1 is degenerate in zero, and hence does not convey any information whatsoever. Thus, it suffices to only consider $Y_2, \ldots, Y_n$.

It is easily seen that $\mathcal{E}_\theta Y_i = 0$. Moreover, one may show that

$$\mathcal{E}_\theta Y_i Y_j^T = \begin{cases} \Sigma & \text{if } i = j > 1; \\ 0 & \text{if } i \neq j; \end{cases}$$

Thus, the recursive score residuals $Y_2, \ldots, Y_n$ have expectation zero, covariancematrix Σ and are uncorrelated. Note that being uncorrelated does not imply independence, except for normal distributions.

Recursive score residuals in exponential families

A random variable X is said to belong to an exponential family if its density function admits the representation

$$f(x;\theta) = c(\theta)h(x)\exp\left\{\tau(\theta)^T t(x)\right\} \tag{1}$$

[cf. Lehmann (1994), Section 2.7]. If in addition the dimension of the vector $\tau(\theta)$ coincides with the dimension of θ, then the random variable X is said to belong to a full exponential family. In Section 6.2 of Hawkins and Olwell (1998) the relevance of exponential families for statistical quality control is discussed, and examples are given.

It is often convenient to reparametrize a full exponential family by taking $\tau = \tau(\theta)$ as parameter instead of θ. In this way we obtain the natural or canonical reparametrization in which X has density function

$$f^*(x;\tau) = c^*(\tau)h(x)\exp\left\{\tau^T t(x)\right\},$$

where $c^*(\tau)$ satisfies $c^*(\tau(\theta)) = c(\theta)$. Let $\rho^*(x;\tau)$ denote the classical score function derived under the natural parametrization. One may show that

$$\rho^*(x;\theta) = \frac{\partial}{\partial\tau}\log f^*(x;\tau) = t(x) - \mathcal{E}_\tau t(X)$$

[see Equation (3.106) in Lindsey (1996), p. 124]. Hence, it follows that the i^{th} "natural" recursive score residual is given by

$$Y_i^* = \sqrt{\frac{i-1}{i}}\left\{t(X_i) - \frac{1}{i-1}\sum_{j=1}^{i-1} t(X_j)\right\}.$$

Note that Y_i^* does not depend on θ.

The reparametrization above was introduced because of its technical convenience. However, in stage 1 quality control the original parametrization may be carefully chosen so as to optimize the detection of assignable causes, and reparametrization should be avoided. Fortunately, the i^{th} "original" recursive score residual Y_i are easily derived from the i^{th} "natural" recursive score residual Y_i^*. Since differentiating via the chain rule yields

$$\rho(x;\theta) = \rho^*(x;\tau(\theta))\cdot\frac{\partial\tau(\theta)}{\partial\theta},$$

it immediately follows that

$$Y_i = Y_i^*\cdot\frac{\partial\tau(\theta)}{\partial\theta}.$$

Note that $\partial\tau(\theta)/\partial\theta$ is a $k \times k$ matrix, possibly depending on θ. Also note that the Fisher information matrix Σ in the original parametrization takes the form

$$\Sigma = \frac{\partial\tau(\theta)}{\partial\theta}\Sigma^* \left(\frac{\partial\tau(\theta)}{\partial\theta}\right)^T,$$

where Σ^* is the Fisher information matrix in the canonical parametrization.

Detecting assignable causes: exponential families

In this section we consider independent random variables $X_1, \ldots, X_n$ with X_i having density $f(x;\theta_i)$ which can be written in the form (1).

Note that the situation considered in the previous section is obtained by setting all θ_i's equal to θ, and corresponds to a process which is in-control. The situation in this section also allows for θ_i's not sharing the same value, corresponding to a process which is not in-control.

Let us suppose that the θ_i's satisfy

$$\tau(\theta_i) = \tau_* + \delta a_i d, \tag{2}$$

where δ is a scalar representing the magnitude of the deviation between $\tau(\theta_i)$ and τ_*, a_i is a scalar representing the type of deviation [sudden shift, linear trend, etcetera], and d is a k-dimensional vector representing the direction of the deviation. Observe that the process is in-control if and only if the $\tau(\theta_i)$'s admit a representation (2) with δ equal to zero.

The joint density of $X_1, \ldots, X_n$ may now be written in the form

$$c(\delta, a_1, \ldots, a_n, d)h(x_1, \ldots, x_n)\exp\left\{\tau_*^T \sum_{j=1}^n t(x_j) + \delta \sum_{j=1}^n a_j d^T t(x_j)\right\}.$$

It follows from Corollary 2 in Chapter 3 of Lehmann (1994) that the test statistic

$$\sum_{j=1}^n \{(\tau_* - \tau_0) + a_i d\}^T t(X_j) \tag{3}$$

is most powerful for testing the simple null hypothesis $H_0 : \tau(\theta_1) = \cdots = \tau(\theta_n) = \tau_0$ versus the alternative described by (2). Moreover, in the in-control situation the statistic $\sum_{j=1}^n t(X_j)$ is sufficient for τ_*, the common value of the $\tau(\theta_i)$'s [Lehmann (1994), p.57]. Now note that the null hypothesis covariance between the most powerful test statistic (3) and the

in-control sufficient statistic $\sum_{j=1}^{n} t(X_j)$ is equal to

$$\sum_{j=1}^{n} \{(\tau_* - \tau_0) + a_j d\}^T \mathcal{E} t(X_1)^T t(X_1),$$

which becomes zero when τ_0 equals $\tau_* + \bar{a}_n d$, where $\bar{a}_n$ denotes $n^{-1} \sum_{j=1}^{n} a_j$. One may interpret this particular value of τ_0 as some sort of a "least favorable" null parameter[1] in the sense of Hájek and Šidák (1967), p. 30. Note that substituting $\tau_* + \bar{a}_n d$ for τ_0 in (3) yields the test statistic

$$\sum_{j=1}^{n} (a_j - \bar{a}_n) d^T t(X_j).$$

Hence, this test statistic [which does not depend on δ] should be highly efficient in testing the composite null hypothesis that all θ_i's are equal versus the alternative described by (2).

By letting the direction d vary, it follows that plots of the components of the k-dimensional vector U_i defined by

$$U_i = \sum_{j=1}^{i} (a_j - \bar{a}_i) t(X_j)$$

may be highly effective in detecting out-of-control behavior of the type described by $a_1, \ldots, a_n$. Note that U_1 is degenerate in zero. Moreover, since we have

$$\begin{aligned}
U_i - U_{i-1} &= \sum_{j=1}^{i} (a_j - \bar{a}_i) t(X_j) - \sum_{j=1}^{i-1} (a_j - \bar{a}_{i-1}) t(X_j) \\
&= (a_i - \bar{a}_i) t(X_i) - (\bar{a}_i - \bar{a}_{i-1}) \sum_{j=1}^{i-1} t(X_j) \\
&= (a_i - \bar{a}_i) \left\{ t(X_i) - \frac{1}{i-1} \sum_{j=1}^{i-1} t(X_j) \right\} \\
&= (a_i - \bar{a}_i) \sqrt{\frac{i}{i-1}} Y_i^* \\
&= (a_i - \bar{a}_{i-1}) \sqrt{\frac{i-1}{i}} Y_i^*,
\end{aligned}$$

we may alternatively express U_i in terms of the natural recursive score residuals $Y_1^*, \ldots, Y_i^*$ as

$$U_i = \sum_{j=2}^{i} c_i Y_j^*, \tag{4}$$

[1] According to Hájek and Šidák (1967) the most powerful test statistic and the sufficient statistic should be independent for the least favorable null parameter. Recall that independence implies zero covariance, but not vice-versa.

where

$$c_i = (a_i - \bar{a}_{i-1})\sqrt{\frac{i-1}{i}}. \tag{5}$$

We defer the precise description of the chart based on the U_i's to the end of the next section.

Detecting assignable causes: the general case

In this section we consider independent random variables $X_1, \ldots, X_n$ with X_i having density $f(x;\theta_i)$, not necessarily belonging to an exponential family.

In exponential families we were able to exploit the special structure of the natural classical score function, which may be viewed as the difference between a term depending only on x and a term only depending on the parameter θ. Such a structure is in the general case not available.

However, if we are willing to assume that all θ_i's are relatively close to the in-control parameter θ, then locally around θ the classical score function approximately has the structure encountered in exponential families, provided that the approximation

$$\rho(x;\vartheta) \approx \rho(x;\theta) - \Sigma(\vartheta - \theta) \tag{6}$$

holds for every ϑ in the vicinity of θ. As before, Σ denotes the in-control Fisher information. In the stage 2 quality control context, a similar approximation may be found in Box and Ramirez (1992).

In exponential families we were also able to exploit the theory of most powerful tests in order to construct test statistics. In the general case we have to resort to asymptotic efficiency concepts and intuitions gained in the previous section.

We start with observing that (6) implies the approximation

$$Y_i \approx \sqrt{\frac{i-1}{i}}\Sigma\left(\theta_i - \frac{1}{i-1}\sum_{j=1}^{i-1}\theta_j\right) + W_i, \tag{7}$$

where

$$W_i = \sqrt{\frac{i-1}{i}}\left\{\rho(x_i;\theta_i) - \frac{1}{i-1}\sum_{j=1}^{i-1}\rho(x_j;\theta_j)\right\}.$$

Since $\mathcal{E}_{\theta_i}\rho(X_i;\theta_i) = 0$ for every $i = 1, 1, \ldots, n$, the random variable W_i has mathematical expectation equal to zero, which coincides with the in-control expectation of Y_i. Moreover, for θ_i close to θ we have

$$\mathcal{E}_{\theta_i}\rho(X_i;\theta_i)\rho(X_i;\theta_i)^T \approx \Sigma;$$

hence, if $\theta_1, \ldots, \theta_i$ are all close to θ, then the random variable W_i approximately has covariance matrix Σ, the in-control covariance matrix of Y_i. In a similar way we may show that W_i and W_j are uncorrelated if $i \neq j$. Thus, the behavior of the sequence $W_2, \ldots, W_n$ approximately is the same as the in-control behavior of the sequence $Y_2, \ldots, Y_n$, as far as first and second moments are concerned. In combination with (7) this leads us to the conclusion that the main effect of the transition from the in-control situation to the present situation is a shift in the distribution of the Y_i's. We shall refer to this shift as the slope of the test statistic Y_i.

Suppose that the θ_i's in fact satisfy

$$\theta_i = \theta_* + \delta a_i d,$$

where as before δ is a scalar representing the magnitude of the deviation between θ_i and θ_*, a_i is a scalar representing the type of deviation, and d is a k-dimensional vector representing the direction of the deviation. It follows from (7) that the slope of Y_i is approximately

$$\sqrt{\frac{i-1}{i}}\Sigma \left(\theta_i - \frac{1}{i-1} \sum_{j=1}^{i-1} \theta_j \right) = \delta \sqrt{\frac{i-1}{i}} \left(a_i - \bar{a}_{i-1} \right) \Sigma d = \delta c_i \Sigma d,$$

and c_i is given by (5).

In the previous section we found that linear combinations of the Y_i's were most powerful. This leads us now to concentrate on the performance of linear combinations $\sum_{i=2}^{n} w_i^T Y_i$ as test statistics when testing the null hypothesis $H_0 \ : \ \delta = 0$ versus a one-sided alternative. Here the w_i's are given k-dimensional weight vectors, for which we shall derive an optimal choice shortly.

Various efficiency concepts suggest that efficiency of a test statistic is indicated by the ratio between the its squared slope and its null-hypothesis variance [see Kallenberg and Koning (1995)]. Straightforward calculations show that the variance of $\sum_{i=2}^{n} w_i^T Y_i$ is approximately equal to the in-control variance $\sum_{i=2}^{n} w_i^T \Sigma w_i$. Moreover, the slope of $\sum_{i=2}^{n} w_i^T Y_i$ is approximately equal to

$$\delta \sum_{i=2}^{n} c_i w_i^T \Sigma d.$$

Thus, the eficiency of linear combinations $\sum_{i=2}^{n} w_i^T Y_i$ is indicated by the ratio

$$\frac{\left(\sum_{i=2}^{n} c_i w_i^T \Sigma d \right)^2}{\sum_{i=2}^{n} w_i^T \Sigma w_i}.$$

An application of the Inequality of Cauchy-Schwarz yields that this ratio is maximized by choosing w_i proportional to $c_i d$. We thus obtain a statistic

$d^T U_i$, where

$$U_i = \sum_{j=2}^{i} c_i Y_j$$

[compare with (4)].

Observe that U_i approximately has mathematical expectation $\delta \sum_{j=2}^{i} c_j^2$ $d^T \Sigma d$ and variance $\sum_{j=2}^{i} c_j^2 d^T \Sigma d$. Thus, both the expectation and the variance of U_i are approximately linear in $\sum_{j=2}^{i} c_j^2$, which indicates that U_i should be plotted versus $\sum_{j=2}^{i} c_j^2$ rather than versus i itself. A disturbing consequence is that [in contrast to traditional cumulative sums] the time instance $\sum_{j=2}^{n} c_j^2$ at which the "last" observation in the sample is observed, does not depend linearly on the sample size anymore. This can be repaired by introducing

$$b_n = \left(n^{-1} \sum_{j=2}^{n} c_j^2 \right)^{-1/2}, \tag{8}$$

and plotting the cumulative sum $b_n U_i$ versus $b_n^2 \sum_{j=2}^{i} c_j^2$. Moreover, since a nonzero value of δ approximately corresponds to a linear trend in the plot of $b_n U_i$ versus $b_n^2 \sum_{j=2}^{i} c_j^2$, applying a V-mask procedure to this plot seems reasonable.

It was shown in Lucas (1982) [cf. Montgomery (1996)] that a V-mask cumulative sum chart may be represented by means of a pair of so-called tabular cumulative sums. Following the same line of reasoning, one may show that the V-mask procedure applied to the plot of $b_n U_i$ versus $b_n^2 \sum_{j=2}^{i} c_j^2$ is equivalent to imposing a control limit h on the pair of one-sided cumulative sums $C_{H,i}$ and $C_{L,i}$ defined by

$$\begin{aligned} C_{H,i} &= \max\left(0, C_{H,i-1} + b_n c_i \left(Y_i - f b_n c_i\right)\right), \\ C_{L,i} &= \max\left(0, C_{L,i-1} + b_n c_i \left(-Y_i - f b_n c_i\right)\right), \end{aligned} \tag{9}$$

where f is the so-called reference value. Observe that both $C_{H,i}$ and $C_{L,i}$ are k-dimensional random vectors; the control limit h should be imposed on all components of these random variables simultaneously. That is, an out-of-control signal is given when at least one of these components exceeds h.

Implementing the charts

In the previous sections we presented the general ideas behind the proposed charts, ideas that give the charts the ability to detect the assignable cause efficiently. However, until now we have avoided technical issues that

emerge when implementing the charts. In this section we discuss these issues and offer some solutions.

The first technical issue concerns the dependence between the components of U_i. Earlier, we suggested to plot components of U_i versus an appropriate time-scale $\sum_{j=2}^{i} c_j^2$. Due to the fact that the covariance matrix of U_i is approximately proportional to the in-control Fisher information Σ, it may be that the components of U_i are highly correlated, which makes it difficult to identify the precise nature of the assignable cause when detected.

To avoid this, we should look at the components of $\Sigma^{-1/2}U_i$ rather than U_i itself; here $\Sigma^{-1/2}$ denotes a matrix such that $\Sigma^{-1/2}\Sigma^{1/2}$ is the identity matrix for some matrix $\Sigma^{1/2}$ which satisfies $\Sigma^{1/2}(\Sigma^{1/2})^T = \Sigma$. In general, several choices of $\Sigma^{-1/2}$ are available. For instance, one may obtain a matrix $\Sigma^{-1/2}$ as a result of inverting the LU-root of Σ; this would be appropriate if the components of θ could be ranked according to importance. Alternatively, one may set $\Sigma^{-1/2}$ equal to $RD^{-1/2}R^T$, where D is a diagonal matrix and R is a rotation matrix satisfying $RDR^T = \Sigma$; this may be appropriate if the components of θ do not differ in importance.

The second technical issue concerns the computation of the in-control Fisher information matrix Σ. It may well be that the computation of Σ becomes too intricate. An alternative is to estimate Σ from the data. One may estimate Σ by

$$\frac{1}{n}\sum_{j=1}^{n}\rho(X_j;\theta)\rho(X_j;\theta)^T. \tag{10}$$

The third technical issue concerns the fact that Y_i, Σ and estimators of Σ may depend on the unknown in-control parameter θ. One may resolve this issue by replacing occurrences of θ by occurrences of $\hat{\theta}_n$, where $\hat{\theta}$ is a full sample estimator of θ. If $\hat{\theta}_n$ takes values close to θ, it should be noted that by virtue of (6) we have

$$\begin{aligned}\sqrt{\frac{i-1}{i}}&\left\{\rho(X_i;\hat{\theta}_n) - \frac{1}{i-1}\sum_{j=1}^{i-1}\rho(X_j;\hat{\theta}_n)\right\}\\ &\approx Y_i + \sqrt{\frac{i-1}{i}}\Sigma\left((\hat{\theta}_n-\theta) - \frac{1}{i-1}\sum_{j=1}^{i-1}(\hat{\theta}_n-\theta)\right\} = Y_i.\end{aligned}$$

As remarked earlier, in exponential families the natural recursive score residual Y_i^* does not depend on θ.

The fourth technical issue concerns the choice between full-sample and "running" estimators. Above, we have proposed using full sample estimators to estimate θ and Σ, as they are the most obvious estimators under in-control conditions. Unfortunately, there is a practical drawback: the use of full sample estimators introduces a masking effect, since out-of-control conditions may inflate full sample estimators.

To avoid the drawback of full sample estimation, one may instead use "running" estimation. Rewrite $\Sigma^{-1/2}U_i$ in the form

$$\sum_{j=2}^{i} c_j \Sigma^{-1/2} Y_j,$$

and replace every occurence of θ in $\Sigma^{-1/2}Y_j$ by $\hat{\theta}_{j-1}$, an estimator solely based on the first $j-1$ observations $X_1, \ldots, X_{j-1}$. If desired, one may first estimate Σ by (10), with n replaced by $i-1$.

Although running estimation does not suffer from the masking effect, it has a drawback of its own. Especially early in the sequence, the variability of the running estimators may lead to "estimated" recursive score residuals with extremely heavy tails. To remedy this, one may consider the use of transformations along the lines of Hawkins (1987) and Quesenberry (1991, 1995). However, such transformations may in turn depend on the unknown in-control parameter θ; estimating this parameter may produce estimated recursive score residuals which have correlated components.

The fifth technical issue concerns the method of estimating θ. Since the classical score function already belongs to the realm of likelihood, it seems natural to use maximum likelihood estimator. The full sample maximum likelihood estimator $\hat{\theta}_n^{\mathrm{ML}}$ is found by solving the likelihood equations

$$\sum_{j=1}^{n} \rho(X_i; \hat{\theta}_n^{\mathrm{ML}}) = 0. \tag{11}$$

When replacing n by $i-1$ in (11) one obtains the maximum likehood equations for the running maximum likelihood estimator $\hat{\theta}_{i-1}^{\mathrm{ML}}$. Observe that these "running" likelihood equations directly imply that

$$\sqrt{\frac{i-1}{i}} \left\{ \rho(X_i; \hat{\theta}_{i-1}^{\mathrm{ML}}) - \frac{1}{i-1} \sum_{j=1}^{i-1} \rho(X_j; \hat{\theta}_{i-1}^{\mathrm{ML}}) \right\} = \sqrt{\frac{i-1}{i}} \rho(X_i; \hat{\theta}_{i-1}^{\mathrm{ML}}),$$

which underlines the naturalness of maximum likelihood estimation in the presence of classical score functions.

The sixth and final technical issue concerns time reversal. Although the proposed charts were developed to detect a particular assignable cause, they are typically higly effective in detecting other assignable causes as well. For instance, the chart for detecting linear trend is also sensitive to sudden shifts occurring not too early in the sample. Reversing time yields other charts, with "secundary" properties that may be favourable: the "reversed time" chart for detecting linear trend is sensitive to sudden shifts occurring not too late in the sample.

Detecting linear trend in normal quality characteristics

In this section we exemplify the methods discussed previously by applying them to the problem of detecting linear trend in quality characteristics which follow a normal distribution. The in-control parameters μ and σ^2 are unknown.

The in-control density of these quality characteristics is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\},$$

which can be written in exponential form as

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \times \exp\left\{\tau(\mu,\sigma^2)^T t(x)\right\}$$

with

$$\tau(\mu,\sigma^2) = \begin{pmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{pmatrix}, \quad t(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}.$$

Let θ denote the vector $(\mu,\sigma^2)^T$, and let X a random variable with density $f(x;\mu,\sigma^2)$. Since

$$\mathcal{E}_\theta t(X) = \begin{pmatrix} \mu \\ \mu^2+\sigma^2 \end{pmatrix},$$

it follows that the natural classical score function is given by

$$\rho^*(x;\mu,\sigma^2) = \begin{pmatrix} x-\mu \\ x^2-(\mu+\sigma^2) \end{pmatrix}.$$

Moreover, since

$$\frac{\partial\tau(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ -\frac{\mu}{\sigma^4} & \frac{1}{2\sigma^4} \end{pmatrix},$$

it follows that the original classical score function is given by

$$\rho(x;\mu,\sigma^2) = \begin{pmatrix} x-\mu \\ x^2-(\mu+\sigma^2) \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ -\frac{\mu}{\sigma^4} & \frac{1}{2\sigma^4} \end{pmatrix} = \frac{1}{\sigma^2}\begin{pmatrix} x-\mu \\ \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}-1 \end{pmatrix}.$$

Alternatively, we could have obtained the original classical score function by directly differentiating

$$\log f(x;\mu,\sigma^2) = -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log\sigma^2 - \tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

with respect to μ and σ^2.

It is now easily derived that

$$Y_i = \sqrt{\frac{i-1}{i}} \frac{1}{\sigma^2} \begin{pmatrix} X_i - \frac{1}{i-1}\sum_{j=1}^{i-1} X_j \\ \frac{1}{2}\left\{ \frac{(X_i-\mu)^2}{\sigma^2} - \frac{1}{i-1}\sum_{j=1}^{i-1} \frac{(X_j-\mu)^2}{\sigma^2} \right\} \end{pmatrix}.$$

The in-control Fisher information Σ, the in-control covariance matrix of Y_i, is given by

$$\Sigma = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Since Σ happens to be a diagonal matrix, the issue of choosing $\Sigma^{-1/2}$ does not emerge here. The only available choice is

$$\Sigma^{-1/2} = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^2\sqrt{2} \end{pmatrix}.$$

Thus, we obtain

$$\Sigma^{-1/2} Y_i = \sqrt{\frac{i-1}{i}} \begin{pmatrix} \frac{1}{\sigma}\left\{ X_i - \frac{1}{i-1}\sum_{j=1}^{i-1} X_j \right\} \\ \frac{1}{\sigma^2\sqrt{2}}\left\{ (X_i-\mu)^2 - \frac{1}{i-1}\sum_{j=1}^{i-1}(X_j-\mu)^2 \right\} \end{pmatrix}. \qquad (12)$$

Note that $\Sigma^{-1/2} Y_i$ depends on the unknown in-control parameters μ and σ^2.

Replacing these unknown values by their full sample estimators

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$$

yields

$$Z_{n,i} = \sqrt{\frac{i-1}{i}} \begin{pmatrix} \frac{1}{S_n}\left\{ X_i - \frac{1}{i-1}\sum_{j=1}^{i-1} X_j \right\} \\ \frac{1}{S_n^2\sqrt{2}}\left\{ (X_i - \bar{X}_n)^2 - \frac{1}{i-1}\sum_{j=1}^{i-1}(X_j - \bar{X}_n)^2 \right\} \end{pmatrix}$$

as an approximation to $\Sigma^{-1/2} Y_i$. Observe that each of the $Z_{n,i}$'s has the same distribution.

To derive charts for detecting linear trend, we should set a_i equal to i, which yields that the tabular Cusum scheme (9) should be applied with

$$c_i = \sqrt{\frac{i-1}{i}}(a_i - \bar{a}_{i-1}) = \sqrt{\frac{i-1}{i}}(i - i/2) = \sqrt{(i-1)i/2}$$

and Y_i replaced by $Z_{n,i}$. Note that considering only the first components of $C_{H,i}$ and $C_{L,i}$ leads to the Cusum chart proposed in Koning and Does (1999).

Replacing the unknown in-control parameters μ and σ^2 their running estimators $\bar{X}_{i-1}$ and S_{i-1}^2 yields

$$Z_{i-1,i} = \sqrt{\frac{i-1}{i}} \begin{pmatrix} \left(X_i - \bar{X}_{i-1}\right)/S_{i-1} \\ \frac{1}{\sqrt{2}}\left\{ \left(\left(X_i - \bar{X}_{i-1}\right)/S_{i-1}\right)^2 - 1 \right\} \end{pmatrix}$$

as an approximation to $\Sigma^{-1/2}Y_i$. Observe that $Z_{2,3}, Z_{3,4}, \ldots, Z_{n-1,n}$ remain uncorrelated, but have widely varying in-control behavior: one may consider $Z_{i-1,i}$ to be a function of $\sqrt{(i-1)/i}(X_i - \bar{X}_{i-1})/S_{i-1}$, which has a t-distribution with $i-1$ degrees of freedom. It is well-known that the heavy-tailed Cauchy distribution is a special case of the t-distribution, and the light-tailed normal distribution is a limiting case. To remedy the varying behavior we could consider replacing $\sqrt{(i-1)/i}(X_i - \bar{X}_{i-1})/S_{i-1}$ by

$$Q_i = \Phi^{-1}\left(G_{i-1}\left(\sqrt{\frac{i-1}{i}}\left(X_i - \frac{1}{i-1}\sum_{j=1}^{i-1} X_j\right)/S_{i-1}\right)\right),$$

where Φ^{-1} denotes the standard normal inverse cumulative distribution function, and G_{i-1} the cumulative distribution function belonging to the t-distribution with $i-1$ degrees of freedom. In Quesenberry (1991) it is shown that $Q_3, \ldots, Q_n$ are independent random variables under in-control conditions. Using the Q_i's as replacements leads to

$$Z^*_{i-1,i} = \begin{pmatrix} Q_i \\ \frac{1}{\sqrt{2}}\left(\sqrt{\frac{i}{i-1}}Q_i^2 - 1\right) \end{pmatrix}$$

as an approximation to $\Sigma^{-1/2}Y_i$. To derive charts for detecting linear trend, the tabular Cusum scheme (9) should be applied with Y_i replaced by $Z^*_{i-1,i}$, and c_i set equal to $((i-1)i/2)^{1/2}$.

The in-control variance of $Z^*_{i-1,i}$ depends on i, which is undesirable. Alternatively, one may consider using

$$Z^{**}_{i-1,i} = \begin{pmatrix} Q_i \\ \frac{1}{\sqrt{2}}\left(Q_i^2 - 1\right) \end{pmatrix} \tag{13}$$

as an approximation to $\Sigma^{-1/2}Y_i$.

A comparison of the methods

In this section the implementations described in the previous section are compared to each other, and to the likelihood ratio test chart.

Under three different out-of-control "directions" and six different out-of-control "types" 10,000 samples of size $n = 6, 12, 18, 24, 30$ were simulated according to the model

$$X_i = (\delta a_i d_1 + \epsilon_i)\exp\{\delta a_i d_2\}, \quad i = 1, \ldots, n,$$

where the ϵ_i's are independent standard normal random variables, the a_i's depend on the condition and satisfy $\sum_{i=1}^n (a_i - \bar{a}_n)^2 = 1$, and δ is a quantity indicating the magnitude of the deviance from the in-control condition. Under the in-control condition δ is equal to zero.

Although we do not recommend performing preliminary control on only six observations, we have included $n = 6$ in our simulations in order to be able to clearly distinguish unwanted behavior of the charts in small samples.

The vector $(d_1, d_2)^T$ determines the out-of-control "direction"; we used the choices $(1, 0)^T$ [location], $(0, 1)^T$ [scale] and $(.8, .6)^T$ [combined location-scale].

Below the structure of the a_i's in each of the six out-of control types is described.

Type SS2 The a_i's exhibit a sudden shift at relative position $1/2$.

Type SS3 The a_i's exhibit a sudden shift at relative position $2/3$.

Type SS6 The a_i's exhibit a sudden shift at relative position $5/6$.

Type LT0 The a_i's are linearly dependent on i.

Type LT1 The a_i's are constant up to relative position $1/3$ within the sample, and linearly dependent on i from this position onwards.

Type LT3 The a_i's are constant up to relative position $2/3$, and linearly dependent on i from this position onwards.

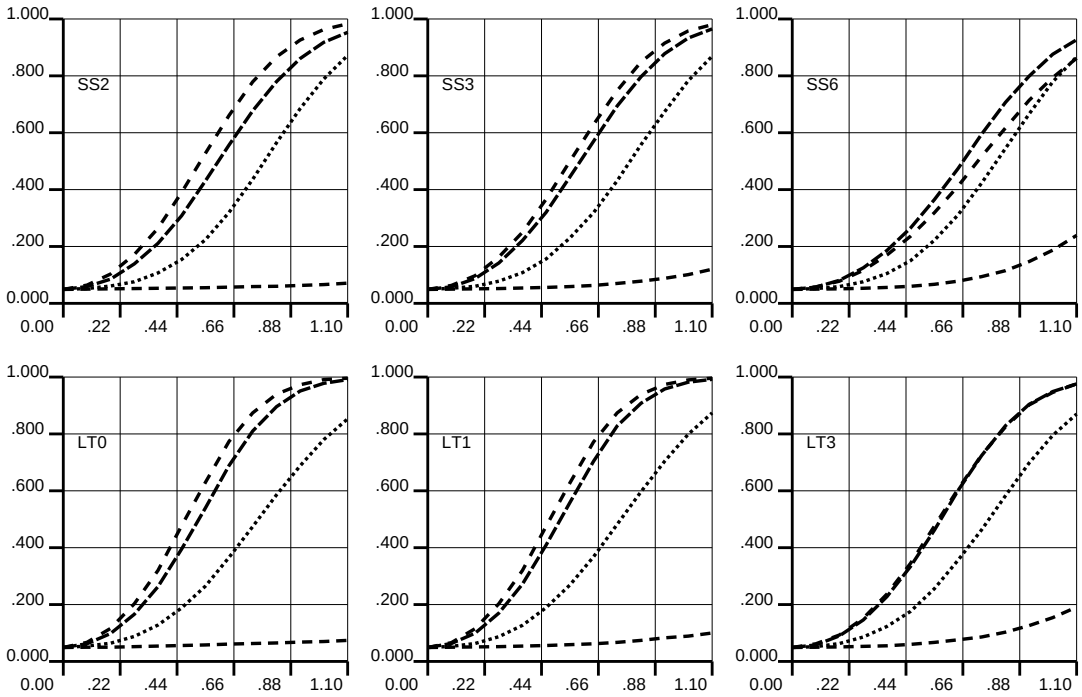
These out-of-control types also feature in Koning and Does (1999).

We did not consider out-of-control types which reflect a sudden shift occurring relative early in the sample. If one [without looking at the data] has reason to suspect the presence of an early sudden shift, then one should reverse time before constructing the charts. The performance of the time-reversed charts for early sudden shifts can be immediately deduced from the performance of the "ordinary" chart for sudden shifts occurring relatively late.

Under each of the six out-of-control types we estimated the signalling probabilities of all the charts. All charts were designed to have an overall in-control signalling probability equal to 0.05.

The extensiveness of the simulation results does not permit a detailed discussion. However, in broad outline the following conclusions may be drawn.

- Charts based on $Z^{**}_{i-1,i}$ perform marginally better than charts based on $Z^{*}_{i-1,i}$.
- Variants of tabular Cusum schemes (9) with $c_i = ((i-1)i/2)^{1/2}$ perform clearly better than their counterparts with $c_i = 1$.



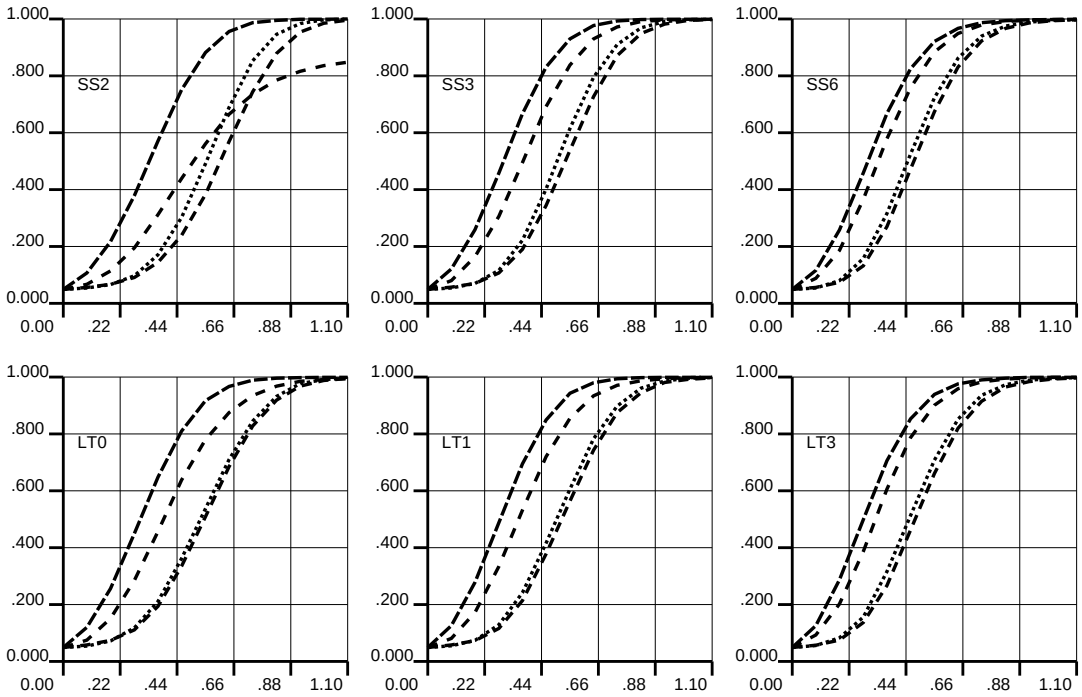
Legend

- - · Variant based on $Z_{i,n}$
- - - Variant based on $Z_{i,i-1}$
- - -· Variant based on $\tilde{Z}_{i,i-1}$
- ····· Likelihood ratio chart

Figure 1: Simulation results for sample size 24: signalling probability versus δ. Out-of-control conditions are in the location direction. On the upper row out-of-control types SS2, SS3, SS6, on the lower row out-of-control types LT0, LT1, LT3.

- If a tabular Cusum scheme (9) with $f = 0$ performs well, setting f equal to a positive value does not lead to a better performance [but may lead to a worse performance instead].
- Running estimation should not be used without subsequent transformation.
- If the out-of-control conditions are purely of the location type, then full sample estimation is to be preferred over running estimation [with or without subsequent transformation]. If out-of-control conditions influence scale, then full estimation should not be used, and running estimation with subsequent transformation is to be preferred.

To illustrate these findings we present in Figures 1–3 the simulation results for $n = 24$ for four different charts. Three of these charts are variants of the tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by either $Z_{i,n}$ [full sample estimation], $Z_{i,i-1}$ [running estimation] or $Z^{**}_{i,i-1}$



Legend
- – – · Variant based on $Z_{i,n}$
- – – – Variant based on $Z_{i,i-1}$
- – – – · Variant based on $\tilde{Z}_{i,i-1}$
- ····· Likelihood ratio chart

Figure 2: Simulation results for sample size 24: signalling probability versus δ. Out-of-control conditions are in the scale direction. On the upper row out-of-control types SS2, SS3, SS6, on the lower row out-of-control types LT0, LT1, LT3.

[running estimation with subsequent transformation]. The fourth chart is the likelihood ratio chart of Sullivan and Woodall (1996).

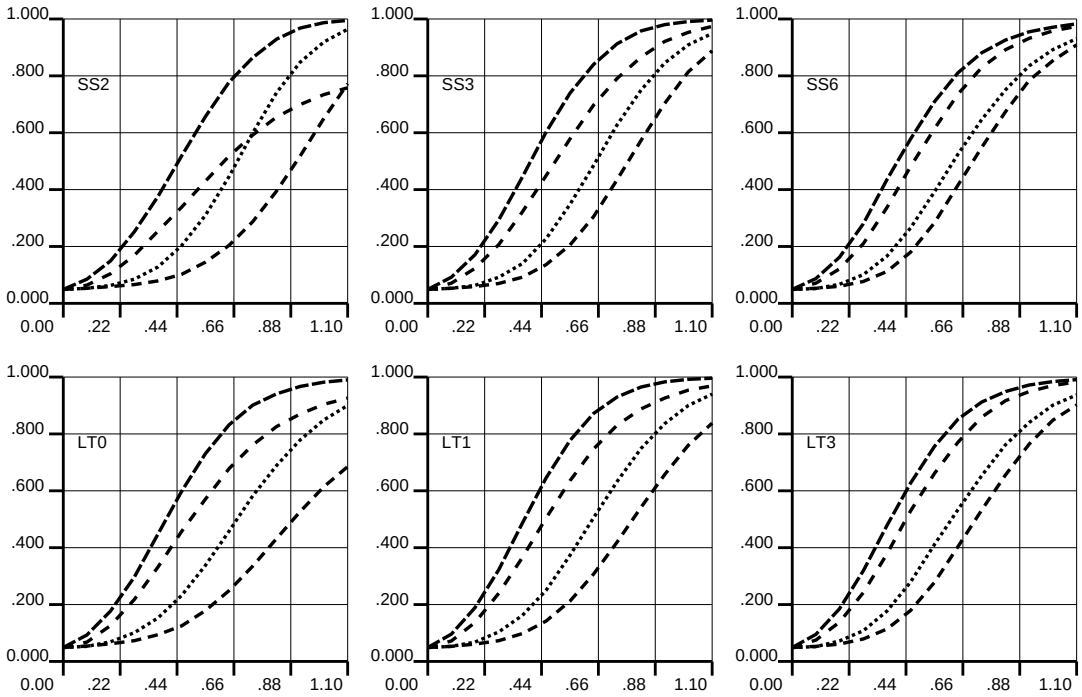
In particular, the simulation results suggest that the third variant considered in Figures 1–3 [the tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$] is highly effective in detecting out-of-control conditions which may affect location as well as scale. In the remainder of this section we investigate the behavior of this tabular scheme under in-control conditions.

For n sufficiently large, theoretical considerations based on formula (11.12) in Billingsley (1968) yield that $h_{0.001}$, $h_{0.005}$, $h_{0.01}$ and $h_{0.05}$ may be approximated by $\sqrt{14.71n}$, $\sqrt{11.70n}$, $\sqrt{10.41n}$ and $\sqrt{7.45n}$, respectively. Plots of the values in Table 1 versus n suggest the following refinements:

$$h_{0.001} \approx \sqrt{14.71n + 26.07\sqrt{n}},$$

$$h_{0.005} \approx \sqrt{11.70n + 7.28\sqrt{n}},$$

$$h_{0.01} \approx \sqrt{10.41n + 1.61\sqrt{n}},$$



Legend
- – – · Variant based on $Z_{i,n}$
- – – – Variant based on $Z_{i,i-1}$
- – – – Variant based on $\tilde{Z}_{i,i-1}$
- ····· Likelihood ratio chart

Figure 3: Simulation results for sample size 24: signalling probability versus δ. Out-of-control conditions are in the combined location-scale direction. On the upper row out-of-control types SS2, SS3, SS6, on the lower row out-of-control types LT0, LT1, LT3.

$$h_{0.05} \approx \sqrt{7.45n - 5.91\sqrt{n}},$$

which may be used for $n \geq 50$.

An application

In this section the variant of the tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$ [see (13)] is applied to the data sets in Examples 1–3 in Sullivan and Woodall (1996). These data sets were generated from a normal distribution.

The data set in Example 1 contains a shift in mean after 15 of 30 observations [in our terminology: sample size 30, type SS2, location direction]. The left plot of Figure 4 shows the "location" components of the tabular

n	Control limits			
	$h_{0.001}$	$h_{0.005}$	$h_{0.01}$	$h_{0.05}$
5	12.29	9.04	7.53	4.62
10	15.71	11.95	10.41	7.45
15	17.67	14.25	13.09	9.49
20	21.10	16.75	15.05	11.22
25	22.91	19.19	16.98	12.58
30	23.79	19.51	17.84	13.99
35	25.96	20.93	19.40	15.24
40	26.79	22.54	20.61	16.20
45	29.51	23.66	21.36	17.37
50	31.51	25.63	23.32	18.13
60	32.84	27.63	25.24	20.03
70	34.94	30.25	27.10	21.45
80	38.17	31.94	29.22	23.25
90	38.39	32.45	30.72	24.83

Table 1: Simulated control limits h_α for tabular Cusums $S_{L,i}$ and $S_{H,i}$ with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$, resulting in an overall in-control signalling probability $\alpha = 0.001, 0.005, 0.01, 0.05$. Control limits are based on 10,000 simulations.

Cusums $S_{L,i}$ and $S_{H,i}$, where

$$b_n = \left(n^{-1} \sum_{j=1}^{n} c_j^2 \right)^{-1/2} = \sqrt{\frac{3}{n(n+1)}}$$

[cf. (8)]. Note that the upper tabular Cusum exceeds $h_{0.05} = 13.99$, and hence at the 5 percent level an out-of-control signal should be given. The right plot shows the "scale" components of $S_{L,i}$ and $S_{H,i}$; they do not move beyond $h_{0.05}$.

The data set in Example 2 contains a shift in variance after 5 of 30 observations. As Figure 5 shows, both the location and the scale components of the tabular Cusums $S_{L,i}$ and $S_{H,i}$ do not exceed $h_{0.05}$, and hence at the 5 percent level no out-of-control signal should be given.

If it is known that a possible shift is likely to emerge relatively early in the sample, then one may reverse time in order to profit of the sensitivity of the tabular Cusums to shifts occurring relatively late in the sample. As Figure 6 shows, reversing time also does not lead to an out-of-control signal at the 5 percent level.

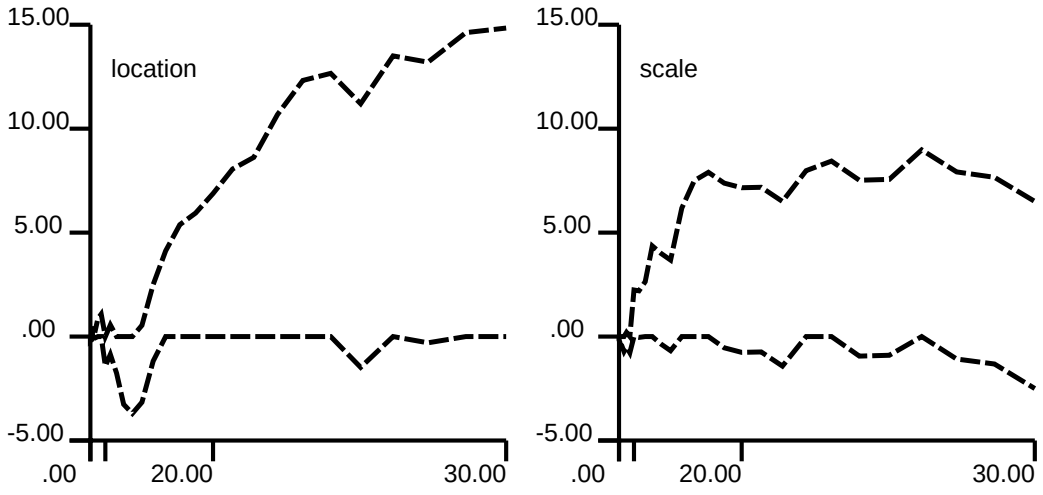


Figure 4: Application to Example 1 in Sullivan and Woodall (1996) of tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$ [running estimation with subsequent transformation].

In contrast, the likelihood ratio test chart does give an out-of-control signal. This is contrary to what is expected: in our terminology we are dealing here with the SS6 type in reversed time, and our simulations show that in this case the likelihood ratio test chart is inferior to the chart depicted in Figure 6 [cf. the upper-right plot of Figure 2].

Finally, the data set in Example 3 exhibits a shift in both mean and variance after 15 of 30 observations. Figure 7 shows that both the location and the scale component of the upper tabular Cusum exceed $h_{0.05}$, and hence at the 5 percent level an out-of-control signal should be given.

References

[1] Box, G., Ramirez, J. (1992). "Cumulative score charts". *Quality and Reliability Engineering International* **8**, 17–27.

[2] Brown, R.L., Durbin, J., Evans, J.M. (1975). "Techniques for testing the constancy of regression relationships over time". *Journal of the Royal Statistical Society* **B37**, 149-163.

[3] Del Castillo, E., Montgomery, D.C. (1994). "Short-run statistical process control: Q-chart enhancements and alternative methods". *Quality and Reliability Engineering International* **10**, 87–97.

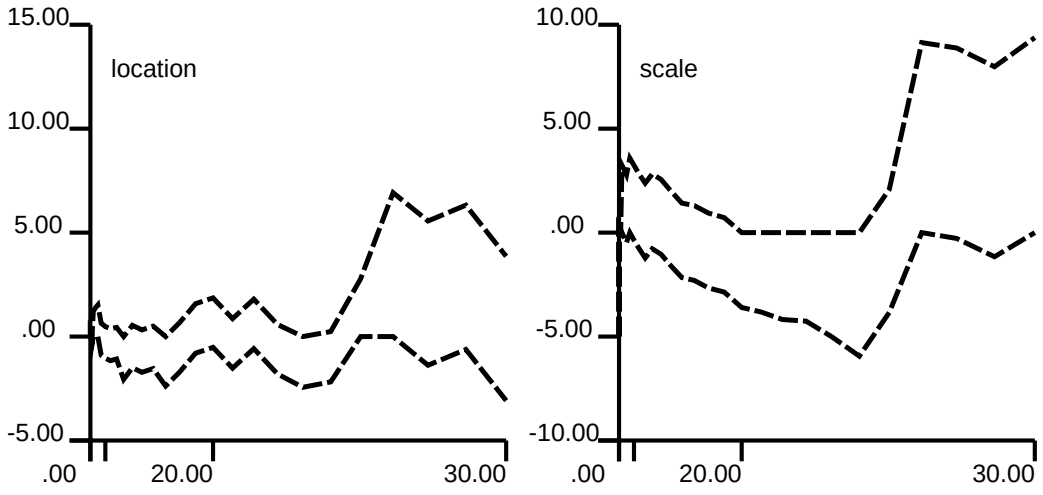


Figure 5: Application to Example 2 in Sullivan and Woodall (1996) of tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$ [running estimation with subsequent transformation].

[4] Hájek, J., Šidák, Z. (1967). *Theory of rank tests*. Academic Press, New York.

[5] Hawkins, D.M. (1991). "Diagnostics for use with regression recursive residuals". *Technometrics* **33**, 221–234.

[6] Hawkins, D.M., Olwell, D.H. (1998). *Cumulative sum charts and charting for quality improvement*. Springer-Verlag, New York.

[7] Hinkley, D.V. (1971). "Inference about the change-point from cumulative sum tests". *Biometrika* **58**, 509–523.

[8] Kallenberg, W.C.M., Koning, A.J. (1995). "On Wieand's theorem". *Statistics & Probability Letters* **25**, 121–132.

[9] Koning, A.J. (1999). "A general Cusum chart for preliminary analysis of individual observations". *Bulletin of the International Statistical Institute.*

[10] Koning, A.J. and Does, R.J.M.M. (1999). "Cusum charts for preliminary analysis of individual observations". *Journal of Quality Technology*, accepted for publication.

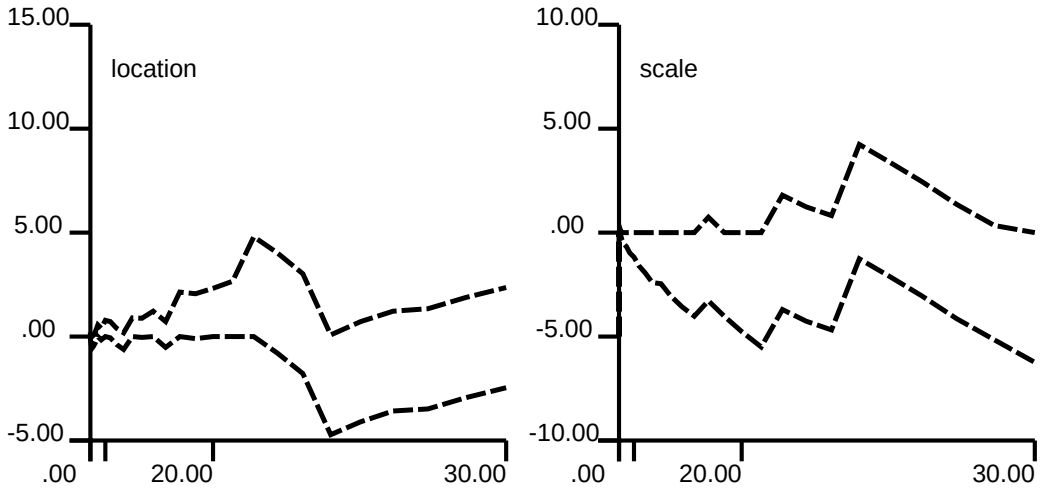


Figure 6: Application to Example 2 in Sullivan and Woodall (1996) of tabular scheme (9) with reversed time, $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$ [running estimation with subsequent transformation].

[11] Lehmann, E.L. (1994). *Testing Statistical Hypotheses. Second edition*. Chapman and Hall, London.

[12] Lindsey, J.K. (1996). *Parametric Statistical Inference*. Oxford University Press.

[13] Lucas (1982). "Combined Shewhart-CUSUM quality control schemes". *Journal of Quality Technology* **14**, 51–59.

[14] Montgomery, D.C. (1996). *Introduction to Statistical Quality Control, third edition*. Wiley, New York.

[15] Quandt, R.E. (1960). "Tests of the hypothesis that a linear regression system obeys two seperate regimes". *Journal of the American Statistical Association* **55**, 324–330.

[16] Quesenberry, C.P. (1991). "SPC Q charts for start-up processes and short or long runs". *Journal of Quality Technology* **23**, 213–224.

[17] Quesenberry, C.P. (1995). "On properties of Q charts for variables". *Journal of Quality Technology* **27**, 184–203.

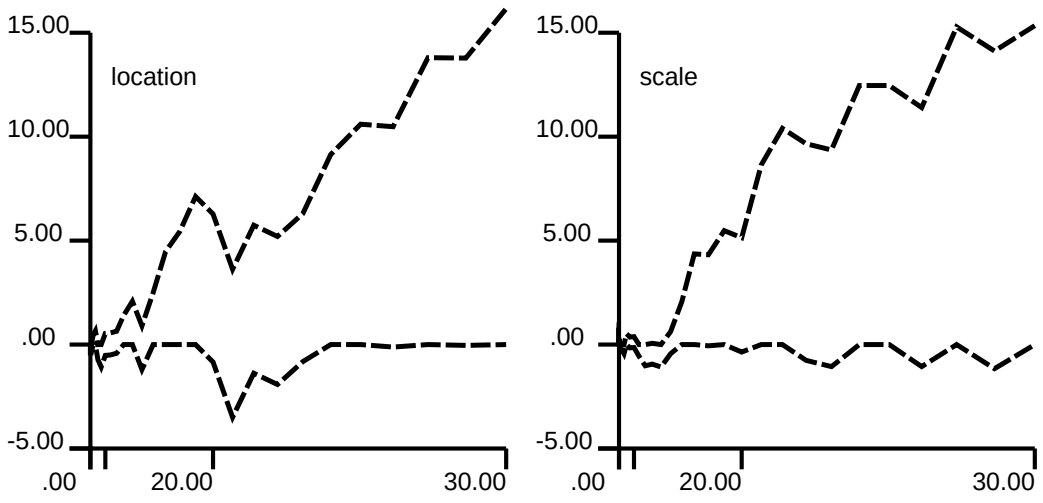


Figure 7: Application to Example 3 in Sullivan and Woodall (1996) of tabular scheme (9) with $c_i = ((i-1)i/2)^{1/2}$, $f = 0$ and Y_i replaced by $Z^{**}_{i,i-1}$ [running estimation with subsequent transformation].

[18] Sullivan, J.H., Woodall, W.H. (1996). "A control chart for preliminary analysis of individual observations". *Journal of Quality Technology* **28**, 265–278.

[19] Worsley, K.J. (1986). "Confidence regions and tests for a change-point in a sequence of exponential family random variables". *Biometrika* **73**, 91–104.