QUALITY ENGINEERING®
Vol. 16, No. 3, pp. 407–421, 2004

# Robust Individuals Control Chart for Exploratory Analysis

**Jeroen de Mast[1,*] and Kit C. B. Roes[2]**

[1]Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA), Amsterdam, The Netherlands
[2]NV Organon, Oss, The Netherlands

## ABSTRACT

Apart from their uses in the context of statistical process control, control charts are also used as an exploratory tool in the context of exploratory data analysis. These two application situations are not similar and, as a consequence, control charting methodology should be adjusted to the exploratory context. In this article we make an inventory of the requirements for control charts that are used for exploratory analysis and we propose a procedure that meets these requirements. Robustness against assignable causes of variation appears to be important. The proposed methodology is illustrated from two real-life examples.

*Key Words:* A-estimator; Change-point estimation; Robust estimation; Preliminary analysis; Shewhart chart.

## INTRODUCTION

Control charts are applied for a number of purposes and in a range of contexts. Their original application is as part of a control loop, for instance, in the form of Shewhart's continuing and self-correcting operation to bring a process into a state of statistical control (Shewhart, 1939, p. 25). In this application, measurements from the running process are added one by one to the chart. The chart's purpose is to instigate and guide the search for assignable causes of variation. A slightly different situation arises once the process is more or less in statistical control and the control chart is used to monitor the process. Literature on control charts (Woodall and Montgomery, 1999) focuses predominantly on the control chart in this context. Consequently, the current charting methodology has developed charts that function well in the context of monitoring, but this does not necessarily make them perform well as a technique in other applications. A different application of control charts is their

*Correspondence: Dr. Jeroen de Mast, Senior Consultant, Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA), Plantage Muidergracht 24, 1018 TV, Amsterdam, The Netherlands; Fax: +31 20 5255101; E-mail: jdemast@science.uva.nl.



DOI: 10.1081/QEN-120027943
0898-2112 (Print); 1532-4222 (Online)
www.dekker.com

use to judge from a given data set whether the process that produced the data is in statistical control. This type of application is called *retrospective analysis* (Woodall and Montgomery, 1999).

In this article we study the use of control charts for exploratory analysis. This type of analysis is performed on a given data set (such as in retrospective analysis but unlike monitoring) and has as its purpose the detection of assignable causes of variation, i.e., assist an inquirer who seeks to discover sources of variation in a process by revealing evidence of the variation sources. This purpose is identical to the function of control charting in the monitoring context but deviates from its application in retrospective analysis, where the (primary) purpose is not the detection of assignable causes (which is an analysis of individual observations) but the inference whether or not the process is in statistical control (which is an inference about the complete data sample). Control charts are used in this manner in exploratory studies (Hoaglin et al., 1983) as well as an after-the-trial check in the analysis of experimental data to check for evidence of assignable causes. It is the purpose of this article to develop a control-charting procedure that meets the needs of exploratory analysis.

We limit ourselves to control charts for data that are collected as individual measurements, leaving situations in which data are collected in subgroups for further research. Furthermore, we consider situations in which the data can be described reasonably adequately by means of the normal distribution, although the methodology is robustified against deviations from the normal distribution, especially in the tails. Finally, we assume that the data are independent. The control chart is designed with moderate sample sizes (say, 20 to 100 data points) in mind.

The article is organized as follows. First we investigate the differences between the exploratory context and the monitoring context. Two motivating examples are discussed, highlighting the problems that are encountered if an inquirer applies control charts for exploratory analysis. The fourth section enumerates the requirements for a control chart in the exploratory context and incorporates these in a control-charting procedure. The statistical technicalities are developed in the fifth section. Two real-life examples demonstrate the proposed procedure's effectiveness in comparison with alternative control-charting methods. A discussion concludes the article.

## THE EXPLORATORY VS. THE MONITORING CONTEXT

To illustrate the importance of studying control-charting procedures for exploratory analysis apart from the monitoring context, we consider differences between the two situations. These differences bear upon:

1. *Continuing time series vs. finite sample*. In the monitoring situation, samples are collected one by one. Each newly collected data point is compared to the control limits to verify whether the process is still in statistical control. The probability $\alpha$ of a false signal is associated with a tail area under the distribution of the individual measurements. If the normal distribution is taken as an approximation of this distribution, the control limits are usually set at a distance of $3\sigma$ of the central line, which corresponds with an approximate false alarm probability of 0.0027 per observation.

In the exploratory case, the inquirer deals with a finite sample. Consequently, he could consider basing the control limits on an overall false alarm probability. The approximate false alarm probability $\alpha$ is then based on the joint distribution of all the data in the sample. Woodall and Montgomery (1999) advocate this procedure in the context of retrospective analysis, and it seems appropriate when the chart ought to provide answers relating to the complete data set (such as, Was the process controlled?). We shall argue that it is inappropriate in the exploratory context.

2. *State of the process*. Monitoring is especially useful if the process has first been brought in a more or less controlled state. In the exploratory situation, however, there is no reason to assume that the process is in control, hence it is very possible to encounter a chaotic sample of measurements. It is, therefore, important to protect oneself by making the control chart robust against deviations from the assumptions that underlie the method. In line with exploratory data analysis theory in general, robust methods ought to play a dominant role (Hoaglin et al., 1983).

3. *Dependence between test statistics and control limits*. The control limits for a chart that is used for monitoring are computed from an initial sample or from historical data. Hence, the measurements that are collected from the production process are stochastically independent of the control limits. In the exploratory situation this is not the case: control limits are computed from the measurements that are plotted in the chart. As a consequence, disturbances


*Figure 1.* Control chart with control limits based on the average moving range.

in the observations can inflate the control limits, making the chart less sensitive in detecting the remaining assignable causes. This is a second motivation to use robust statistical procedures.

4. *Dependence between successive tests*. If the control limits are computed from the measurements that are indicated in the chart, the successive comparisons of observations to the control limits are no longer independent, as was observed by Quesenberry (1993). Consequently, the fraction of points that are mistakenly identified as outliers is no longer $\alpha$. Quesenberry shows that the dependence between the successive tests is related to the variance of the control limits. This is a motivation to base the control limits on efficient estimators.

5. *Difference in functionality*. In monitoring, the emphasis is on the signal function of the control chart, whereas the emphasis in exploratory analysis is on clue generation. For this reason, the requirement that the chart provides an easily interpreted display of the data is even more important in the exploratory context than in the monitoring context.

## TWO MOTIVATING EXAMPLES

Two examples demonstrate some of the problems that could arise when control charts are applied in exploratory analysis.

The data in Fig. 1 were drawn from the $\mathcal{N}(0,1)$ distribution. Observations 11 and 20 were replaced with the value 6.0, observation 14 with the value −6.0. Given the way in which the data were obtained, one would like a control chart to give three signals: at observations 11, 14, and 20. The chart that is often used for exploratory analysis is the individuals control chart (Roes et al., 1993). In this control chart—referred to as the average moving range (AMR) chart—the control limits are computed from:

$$\begin{aligned} \mathrm{UCL} &= \hat{\mu} + 2.66\,\mathrm{AMR} \\ \mathrm{LCL} &= \hat{\mu} - 2.66\,\mathrm{AMR} \end{aligned} \tag{1}$$

Here, $\hat{\mu}$ is the overall mean of the data and:

$$\mathrm{AMR} = \frac{1}{n-1}\sum_{i=1}^{n-1} |y_i - y_{i+1}| \tag{2}$$

The outliers in the data result in an inflated error estimate; consequently, the control limits are too wide: the chart in Fig. 1 gives one signal (observation 14)—and even this point is only just below the lower control limit.

The data in Fig. 2 were obtained from a simulation of the model:

$$y_i = \mu_i + \epsilon_i \tag{3}$$

with $\mu_i = 10.0$ for $i = 1, \ldots, 20$; $\mu_i = 12.0$ for $i = 21, \ldots, 30$; and $\mu_i = 8.0$ for $i = 31, \ldots, 50$. The $\epsilon_i$ are i.i.d. $\mathcal{N}(0,1)$. Observation 47 was set to the value 12.5. We would like a control chart to make the following suggestions:

- Shift in the mean at $i = 21$ and $i = 31$.
- Outlier at $i = 47$.
- The remaining variation is white noise.

The AMR chart gives one signal. It does not give any information about the moment when the shifts occurred, and the shifts make the detection of further

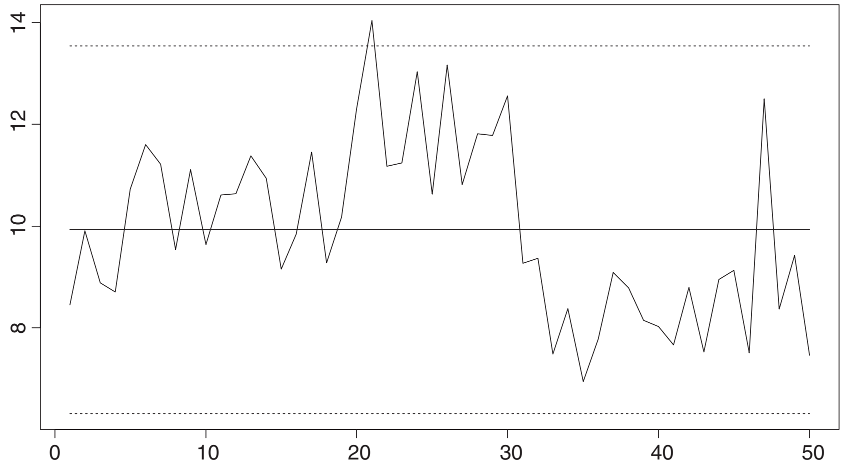


*Figure 2.* Control chart with control limits based on the average moving range.

assignable causes—such as the outlier at $i = 47$—difficult.

## PROPOSED CONTROL CHART

### Requirements

In order to identify requirements for a control chart for exploratory analysis, we study the requirements listed by Shewhart (Shewhart, 1939, p. 30) for control charts:

1. They should indicate the presence of assignable causes of variation.
2. They should not only indicate the presence of assignable causes but also should do this in a way to facilitate the discovery of these causes.
3. They should be as simple as possible and adaptable in a continuing and self-corrective operation of control.
4. They should be such that the chance of looking for assignable causes when they are not present does not exceed some prescribed value.

We discuss these requirements for the control chart in the exploratory context.

#### Ad 1. Indicate the Presence of Assignable Causes

Assignable causes are detected by anything that indicates nonrandomness. An obvious deviation from randomness is an outlier. Other criteria should be based on the order of data, in which generic patterns could be observed by which assignable causes manifest themselves. We propose to discern, apart from outliers, one generic pattern that the control chart should detect, namely shifts in the mean. The importance of this pattern is acknowledged by the extensive literature on CUSUM and EWMA charts. The discussion about the inclusion of more generic patterns is presented in the last section of this article.

As demonstrated in the examples in the previous section, the presence of assignable causes of either kind, outliers or shifts in the mean, seriously affects the chart's ability to detect the remaining assignable causes. Therefore, and in view of points 2 and 3 in section 2, the procedure should be robust against both outliers and shifts in the mean. The former is achieved by employing robust estimation methods, the latter by incorporating detected shifts in the analysis. Finally, a clear visualization is vital in the exploratory context to facilitate the revelation of more nonrandom patterns, as noted in point 5 in section 2.

#### Ad 2. Facilitate the Discovery of Detected Assignable Causes

In exploratory analysis, this is an important point and it is the reason why the chart should not use an overall false alarm probability $\alpha$. The inquirer is interested in questions such as, *when* were assignable causes present? and *how many* assignable causes were present?, which are questions related to individual measurements. Hence, the chart's performance should be based on the distribution of the individual data points and individual $\alpha$s, such as in the monitoring context, are appropriate.

Ad 3. Simple and Adaptable

The computations required for the set-up of a control chart are nowadays always performed by a computer, and as a consequence, the restrictions on computational efforts are much looser than in Shewhart's days. We do not, therefore, consider simplicity or adaptability of computation an important requirement. Simplicity of the visual presentation, on the contrary, is important.

Ad 4. Low False-Alarm Rate

Shewhart warns against confusing criteria for the identification of assignable causes and hypothesis tests. This relates to the issue that as long as the process is not in statistical control, the assumptions that are required for a hypothesis test (e.g., concerning the assumed distribution) cannot be made rigorously. The mathematical theory of hypothesis testing serves as a background for the development of control-chart procedures, yet it is fundamentally different (Shewhart, 1939, pp. 39–40). In particular, the chance of looking for an assignable cause when there is none present is associated with the tail probability of an assumed distribution function, but this association should be seen as a rough approximation.

Various authors have argued against plotting control limits in a control chart when it is used in the context of exploratory data analysis. In reply to Woodall (2000), Hoerl says that "the plot of the data over time is much more valuable than the control limits." However, Shewhart's requirement 4 suggests that the purpose of the control chart is not only to specify when to search for assignable causes but also when not to do this. In our experience, engineers find it difficult to assess which patterns typically arise in random noise and which are indicative for assignable causes. The control limits give the guidance that these engineers need. This does not mean that the comparison of observations to control limits should be seen as a hypothesis test (Woodall, 2000). It only means that our chart should indicate which points are suspicious and which are not and that our operational definition of 'suspicious' is based on tail areas of distribution functions.

## Proposed Procedure

For the analysis of individual measurements in the exploratory context, we propose the subsequent control-charting procedure.

1. Estimate the locations of possible shifts and test significance of these shifts. Upon completion of this step, the original data set is divided into intervals on which the mean of the measurements is presumed constant.
2. Estimate (using robust estimators) the means of the intervals between successive shifts. Also, estimate (using robust estimators) the variance of the in-control measurements.
3. Based on these estimates, determine a pair of control limits for each interval. Measurements are identified as outliers if they fall beyond these control limits.

A preview of the effectiveness of this procedure in dealing with the problems illustrated in section 3 is shown in Fig. 3.

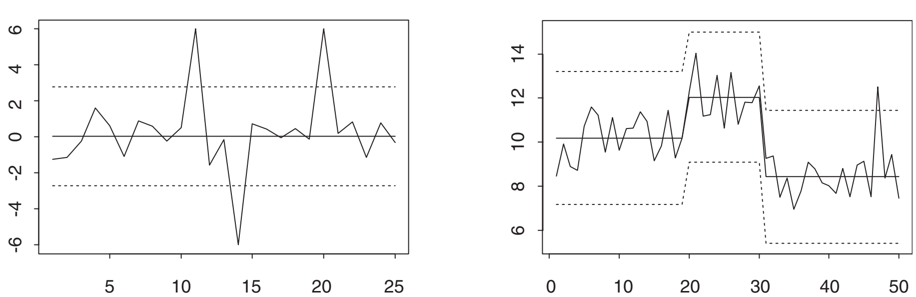


*Figure 3.* Proposed control chart.



Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

### Relationship with Propositions in Literature

Hawkins (1993) distinguishes isolated assignable causes from persistent assignable causes and links these to the Shewhart-type control chart for the first type and charts that accumulate information from successive measurements (CUSUM, EWMA) for the second. The proposed control chart is a combination of both types of techniques: it employs change-point analysis techniques to detect shifts and Shewhart-type control charts for the intervals between shifts. This combination has been proposed in the literature (Lucas, 1982; Sullivan and Woodall, 1996) but not in a form that has the two analyses incorporated in a single graph. There is a lot of literature on the application of change-point analysis in control-charting methodology, such as literature on CUSUM techniques (Page, 1954) and Cuscore charts (Box and Ramirez, 1992). Especially in the context of retrospective analysis, a powerful method is proposed by Koning and Does (2000). However, these articles focus on detection of shifts and do not address the issue of estimating their location. Sullivan and Woodall (1996) do consider estimation of the location of a shift. Their procedure is less powerful in detecting shifts than the procedure of Koning and Does, but the gain in power of the latter is appearance, since the extra detected shifts are so small that they cannot be estimated accurately.

The role that robust estimation ought to play in control-charting methodology was well stated by Rocke (1989), who observed that:

- statistics that are indicated in the control chart should be sensitive to outliers, whereas
- statistics that are used to calculate the control limits should be robust against outliers.

In past decades, a number of articles has appeared that deal with robust control charts. A robust mean and range chart was proposed by Langenberg and Iglewicz (1986), who base their control limits on the trimmed mean of the subgroup means and the trimmed mean of ranges. Rocke (1989) studied the use of the inter quartile range (IQR) as a robust estimator of error and found it to perform well in a number of out-of-control scenarios. More advanced estimators for the mean and range chart where studied by Tatum (1997). He recommended the use of a variant of the biweight A-estimator (Lax, 1985). This estimator is very robust against deviations from the model assumptions but has nonetheless an efficiency that is only slightly less than that of the sample standard deviation.

Robust methods for the individuals chart were studied by Bryce et al. (1997) and Boyles (1997). The median absolute deviation (MAD) from the median seems to outperform other estimators in situations in which outliers are present. However, neither of the two articles draws the biweight A-estimator or other advanced robust estimators into the comparison.

In the following section we derive the necessary estimation and testing procedures. In order to satisfy the robustness requirement, the theory of M-estimators plays an important role. The reader is referred to Hoaglin et al. (1983) for a concise introduction.

## ESTIMATION AND TESTING PROCEDURES

### Estimation of the Location of a Shift

Suppose that an inquirer has collected a sample of $n$ measurements that are to be analyzed using a control chart. The null model that is assumed for the in-control process has all measurements $y_i$, $i = 1, \ldots, n$, share the same normal distribution:

$$y_i = \mu + \epsilon_i, \quad i = 1, \ldots, n, \tag{4}$$

with $\epsilon_i$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

To derive the required estimation and testing methods, we start studying how to estimate the location $\tau$ of a shift, given that a single shift has occurred. The situation is described by the following model:

$$\begin{cases} y_i = \mu_1 + \epsilon_i & \text{for} \quad i = 1, \ldots, \tau \\ y_i = \mu_2 + \epsilon_i & \text{for} \quad i = \tau + 1, \ldots, n \end{cases} \tag{5}$$

$\epsilon_i$ i.i.d. $\mathcal{N}(0, \sigma^2)$; $\mu_1$, $\mu_2$, $\sigma$, and $\tau$ are unknown. Sullivan and Woodall (1996) derive the maximum likelihood (ML) estimator for $\tau$. It is found, together with the ML estimators for $\mu_1$, $\mu_2$, and $\sigma$, from:

$$(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}, \tilde{\tau}) = \arg \min_{\mu_1, \mu_2, \sigma, \tau} \ell(\mu_1, \mu_2, \sigma, \tau) \tag{6}$$

with

$$\ell(\mu_1, \mu_2, \sigma, \tau) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \left( \sum_{i=1}^{\tau} (y_i - \mu_1)^2 + \sum_{i=\tau+1}^{n} (y_i - \mu_2)^2 \right) \tag{7}$$

Arg min $f(x)$ denotes the value of $x$ for which $f(x)$ is minimized. Taking derivatives for $\mu_1$, $\mu_2$, and $\sigma$ and equating to zero, we find for fixed $\tau$:

$$c\tilde{\mu}_1[\tau]=\frac{1}{\tau}\sum_{i=1}^{\tau}y_i$$
$$\tilde{\mu}_2[\tau]=\frac{1}{n-\tau}\sum_{i=\tau+1}^{n}y_i$$
$$\tilde{\sigma}^2[\tau]=\frac{1}{n}\left(\sum_{i=1}^{\tau}(y_i-\tilde{\mu}_1[\tau])^2+\sum_{i=\tau+1}^{n}(y_i-\tilde{\mu}_2[\tau])^2\right) \quad (8)$$

$\tau=2,3,\ldots,n-2$. Next,

$$\tilde{\tau}=\arg\min_{2\le\tau\le n-2}\ell(\tilde{\mu}_1[\tau],\tilde{\mu}_2[\tau],\tilde{\sigma}[\tau],\tau)$$
$$=\arg\min\{\log(\tilde{\sigma}[\tau])\} \quad (9)$$

Finally,

$$\tilde{\mu}_1=\tilde{\mu}_1[\tilde{\tau}];\ \tilde{\mu}_2=\tilde{\mu}_2[\tilde{\tau}];\ \tilde{\sigma}=\tilde{\sigma}[\tilde{\tau}] \quad (10)$$

However, we are well aware that not all measurements are necessarily produced by the in-control model Eq. (5). We anticipate that some observations—outliers—might stem from a more heavily tailed distribution. In order to make the procedure robust against the possible presence of outliers, it is modified in the following way. For $\tau=2,\ldots,n-2$, let $\hat{\mu}_1[\tau]$ be the solution of $\sum_{i=1}^{\tau}\psi((y_i-\mu_1)/cs_0[\tau])=0$ and let $\hat{\mu}_2[\tau]$ be the solution of $\sum_{i=\tau+1}^{n}\psi((y_i-\mu_2)/cs_0[\tau])=0$, where:

$$s_0[\tau]=\text{median}\{|y_i-m|\}_{i=1,\ldots,n} \quad (11)$$

with $m=m_1[\tau]$ if $1\le i\le\tau$, and $m=m_2[\tau]$ if $\tau+1\le i\le n$, where $m_1[\tau]$ and $m_2[\tau]$ are the medians of $y_1,\ldots,y_\tau$ and $y_{\tau+1},\ldots,y_n$, respectively. Furthermore, $\psi$ is an odd function, whereas $c$ is a tuning constant. Thus, $\hat{\mu}_1[\tau]$ and $\hat{\mu}_2[\tau]$ are M-estimators for location based on a preliminary estimate $s_0[\tau]$ for scale. Our final scale estimate is given by:

$$\hat{\sigma}[\tau]=\frac{\sqrt{n}cs_0[\tau]\left(\sum_{i=1}^{\tau}\psi^2((y_i-\hat{\mu}_1[\tau])/cs_0[\tau])+\sum_{i=\tau+1}^{n}(\psi^2((y_i-\hat{\mu}_2[\tau])/cs_0[\tau])\right)^{1/2}}{\left|\sum_{i=1}^{\tau}\psi'((y_i-\hat{\mu}_1[\tau])/cs_0[\tau])+\sum_{i=\tau+1}^{n}\psi'((y_i-\hat{\mu}_2[\tau])/cs_0[\tau])\right|} \quad (12)$$

$\psi'$ denoting the derivative of $\psi$. The given procedure to obtain a scale estimate uses the asymptotic variance of the M-estimators for location in order to estimate the standard deviation of the error. This type of scale estimator is named A-estimators by Lax (1985). Lax finds A-estimators to perform well when compared to other robust scale estimators, including M-estimators. See also Hoaglin et al. (1983), Gross (1976), and Shoemaker (1984).

We propose to take for $\psi$ the derivative of the bisquare function:

$$\psi(u)=\begin{cases}u(1-u^2)^2, & |u|\le 1,\\ 0, & |u|>1.\end{cases} \quad (13)$$

Although this choice can be well motivated, we mention Huber's, Hampel's and Andrews' $\psi$ functions as alternative options (Hoaglin et al., 1983). Observations $y_1,\ldots,y_\tau$ in the neighborhood of the estimated mean $\hat{\mu}_1[\tau]$ have standardized values $u=(y_i-\hat{\mu}_1[\tau])/cs_0[\tau]$ around 0, as do observations $y_{\tau+1},\ldots,y_n$ in the neighborhood of $\hat{\mu}_2[\tau]$. For these observations, $\psi(u)$ behaves as $ku$ (for a nonzero constant $k$), which is the $\psi$-function associated with the ML estimator for normally distributed measurements. In view of the assumed normality of the center of the distribution, this is an important property (Hoaglin et al., 1983, p. 363). Observations further removed from the mean are more and more downweighted, and observations further removed than $cs_0[\tau]$ are completely rejected. Simulations show that the value $c=9$ for the tuning constant results in a good performance in a range of situations.

Analogous to Eq. (9), $\hat{\tau}$ could be determined from:

$$\hat{\tau}=\arg\min\{\log(\hat{\sigma}[\tau])\} \quad (14)$$

but because of the chosen $\psi$ function, evidence of (larger) shifts is mistakenly downweighted (or rejected) as stemming from outliers. Therefore, a better estimate for $\tau$ is found from:

$$\hat{\tau}=\arg\min\{\log(\hat{\sigma}^{\#}[\tau])\} \quad (15)$$

where $\hat{\sigma}^{\#}[\tau]$ is found by substituting $\psi^{\#}$ for $\psi$ in Eq. (12). Function $\psi^{\#}$ is a stretched version of $\psi$ so as to reflect the fact that we have data from two normal distributions (with mean $\mu_1$ and mean $\mu_2$). Figure 4 illustrates the idea. Letting $\Delta=|\hat{\mu}_1[\tau]-\hat{\mu}_2[\tau]|/cs_0[\tau]$,

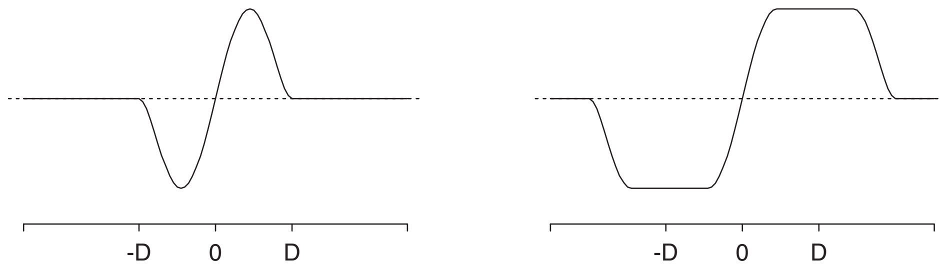


***Figure 4.*** $\psi$ (left) and $\psi^{\#}$ (right). The locations of $u=0$ and $u=\pm\Delta$ are marked with 0 and $D$ or $-D$.

we define:

$$\psi^{\#}(u) = \begin{cases} u(1-u^2)^2, & |u| \le \frac{1}{\sqrt{5}}, \\ \text{sgn}(u)\frac{16}{25\sqrt{5}}, & \frac{1}{\sqrt{5}} < |u| \le \Delta + \frac{1}{\sqrt{5}}, \\ \text{sgn}(u)(|u|-\Delta) \\ \quad \times(1-(|u|-\Delta)^2)^2, & \Delta + \frac{1}{\sqrt{5}} < |u| \le \Delta + 1, \\ 0, & |u| > \Delta + 1. \end{cases} \tag{16}$$

We use $\psi^{\#}$ only in Eq. (15); all other estimators are based on $\psi$. As in Eq. (10), $\hat{\mu}_1 = \hat{\mu}_1[\hat{\tau}]$; $\hat{\mu}_2 = \hat{\mu}_2[\hat{\tau}]$; $\hat{\sigma} = \hat{\sigma}[\hat{\tau}]$.

***Table 1.*** Suitable values for $n_1$ and $n_2$ for various sample sizes $n$.

| $n$ | $n_1$ | $n_2$ |
|---|---|---|
| 5 | 2.09 | 1.15 |
| 8 | 2.57 | 1.95 |
| 10 | 2.98 | 3.00 |
| 15 | 3.26 | 5.70 |
| 20 | 3.50 | 10.90 |
| 30 | 3.76 | 29.60 |
| 40 | 3.97 | 55.30 |
| 50 | 4.13 | 90.60 |
| 60 | 4.23 | $\infty$ |
| 70 | 4.33 | $\infty$ |
| 100 | 4.42 | $\infty$ |
| 150 | 4.56 | $\infty$ |

## Testing Significance of a Shift

We test the null-model Eq. (4) against the alternative model Eq. (5). The likelihood ratio statistic for this testing problem is $(\tilde{\sigma}_0^2/\tilde{\sigma}^2)^{n/2}$, with $n\tilde{\sigma}_0^2 = \sum_{i=1}^n (y_i - \tilde{\mu}_0)^2$ and $n\tilde{\mu}_0 = \sum_{i=1}^n y_i$. This statistic can be shown to be equivalent with:

$$T = \sqrt{\frac{\tilde{\tau}(n-\tilde{\tau})}{n}}\left(\frac{\tilde{\mu}_2 - \tilde{\mu}_1}{\bar{\sigma}}\right) \tag{17}$$

with $\bar{\sigma}^2 = n\tilde{\sigma}^2/(n-2)$. This statistic has the same appearance as the test statistic in the two sample $t$-test, but because of the fact that $\tilde{\tau}$ is a random variable, it does not follow a $t$-distribution under the null model.

The test statistic defined in Eq. (17) can be robustified by plugging in the robust estimators proposed in the preceding section. Thus, we arrive at:

$$RT = \sqrt{\frac{\hat{\tau}(n-\hat{\tau})}{n}}\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{\hat{\sigma}}\right) \tag{18}$$

Simulations show that the distribution of $RT^2/n_1$ under the null model can be well approximated by an $F$-distribution having $n_1$ degrees of freedom in the numerator and $n_2$ degrees of freedom in the denominator. For various $n$, suitable choices for $n_1$ and $n_2$ are given in Table 1. The appendix provides details about the determination of these values.

The proposed approximation is accurate up to the 0.99 quantile of the distribution of $RT^2/n_1$. An accurate approximation to the values for $n_1$ and $n_2$ can be found from the following formulas:

$$\begin{aligned} n_1 &\approx 4.58 - 22.4/n + 52.2/n^2 \\ n_2 &\approx 2.41 - 0.424n + 0.0438n^2 \\ &\quad (\text{For } n > 50, \text{ take } n_2 = \infty.) \end{aligned} \tag{19}$$

Altogether, the test signals if $(1/n_1)RT^2 > F^{-1}_{n_1,n_2} \times (1-\alpha)$, with $\alpha$ the probability of falsely detecting a shift. Table 2 illustrates the performance of the procedure. For shifts of various magnitudes (in multiples of $\sigma$) and for various sample sizes $n$, the table gives the percentage of detected shifts for shifts that occur on one fourth of the data series and halfway the data series.



Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

***Table 2.*** Percentage of detected shifts depending on magnitude of the shift, sample size, and location of the shift.

| Shift (multiples of $\sigma$) | Sample size $n$ and location of shift $\tau$: 10 | | 20 | | 40 | | 80 | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 5 | 5 | 10 | 10 | 20 | 20 | 40 |
| 1 | 6.0% | 6.3% | 21.9% | 29.3% | 47.0% | 61.8% | 83.4% | 93.4% |
| 1.5 | 10.5% | 10.2% | 49.3% | 62.4% | 84.8% | 95.5% | 99.7% | 100.0% |
| 2 | 14.4% | 19.6% | 74.5% | 88.7% | 98.3% | 99.8% | 100.0% | 100.0% |
| 2.5 | 23.4% | 28.7% | 90.9% | 98.2% | 100.0% | 100.0% | 100.0% | 100.0% |
| 3 | 30.6% | 38.9% | 98.2% | 99.8% | 100.0% | 100.0% | 100.0% | 100.0% |

## Multiple Shifts

Having detected a shift, the data are split into two groups: $y_1, \ldots, y_{\hat{\tau}}$ and $y_{\hat{\tau}+1}, \ldots, y_n$. The same procedure is applied to both groups, verifying whether more shifts can be detected. This procedure is continued recursively until no more shifts are detected or until the size of the groups becomes smaller than 4 or any other chosen minimum value. In the situation of a larger number of alternating shifts, this procedure suffers from a masking problem. Since the control chart is designed with moderate sample sizes in mind, the number of shifts will not often be that large and, as a consequence, the adequacy of the proposed control chart is not seriously affected. However, the construction of a procedure that deals more effectively with multiple shifts is an interesting topic for further research.

## Detection of Outliers

Denoting by $\hat{\mu}_1, \ldots, \hat{\mu}_k$ the estimated means of the groups in between the detected shifts $\hat{\tau}_2, \ldots, \hat{\tau}_k$, we re-estimate the error from the formula:

$$\hat{\sigma} = \frac{n(cs_0)\left(\sum_{i=1}^{\hat{\tau}_1} \psi^2((y_i - \hat{\mu}_1)/cs_0) + \cdots + \sum_{i=\hat{\tau}_{k-1}+1}^{n} \psi^2((y_i - \hat{\mu}_k)/cs_0)\right)^{1/2}}{\sqrt{n-k}\left|\sum_{i=1}^{\hat{\tau}_1} \psi'((y_i - \hat{\mu}_1)/cs_0) + \cdots + \sum_{i=\hat{\tau}_{k-1}+1}^{n} \psi'((y_i - \hat{\mu}_k)/cs_0)\right|} \tag{20}$$

As before, $s_0$ is the MAD:

$$s_0 = \text{median}\{y_i - m\} \tag{21}$$

with $m$ the median of the relevant subgroup. The factor $\sqrt{n}$ in the numerator of Eq. (12) is replaced with $\sqrt{n^2/(n-k)}$ in order to account for the loss of degrees of freedom for the estimation of $\mu_1, \ldots, \mu_k$. Note that $\hat{\sigma}$ is calculated from $\psi$ instead of $\psi^{\#}$; this ensures that our final scale estimate is robust against nuisance stemming from slight misestimation of $\tau$.

For each group $j = 1, \ldots, k$, the chart has control limits at;

$$\text{UCL}_j = \hat{\mu}_j + h\sqrt{\frac{\hat{\tau}_{j+1} - \hat{\tau}_j - 1}{\hat{\tau}_{j+1} - \hat{\tau}_j}}\hat{\sigma}$$
$$\text{LCL}_j = \hat{\mu}_j - h\sqrt{\frac{\hat{\tau}_{j+1} - \hat{\tau}_j - 1}{\hat{\tau}_{j+1} - \hat{\tau}_j}}\hat{\sigma} \tag{22}$$

where we define $\hat{\tau}_1 = 0$ and $\hat{\tau}_{k+1} = n$ for notational ease. The scalar $h$ determines how much evidence we require before we are prepared to identify an observation as an outlier. We work with the traditional value $h = 3$. The factor $\sqrt{(\hat{\tau}_{j+1} - \hat{\tau}_j - 1)/(\hat{\tau}_{j+1} - \hat{\tau}_j)}$ stems from the dependency between the $y_i$ and the control limits.

## TWO EXAMPLES

In two examples stemming from recent projects, we compare the method with other control charts. The first example stems from the food industry. The principal stage in the production of decaffeinated coffee is an extraction process during which most of the caffeine is removed from the coffee beans. During the extraction, an auxiliary substance DCM is added to the beans and removed after the process. The percentage of DCM that is absorbed by the beans and as a result cannot be removed is an important quality characteristic.

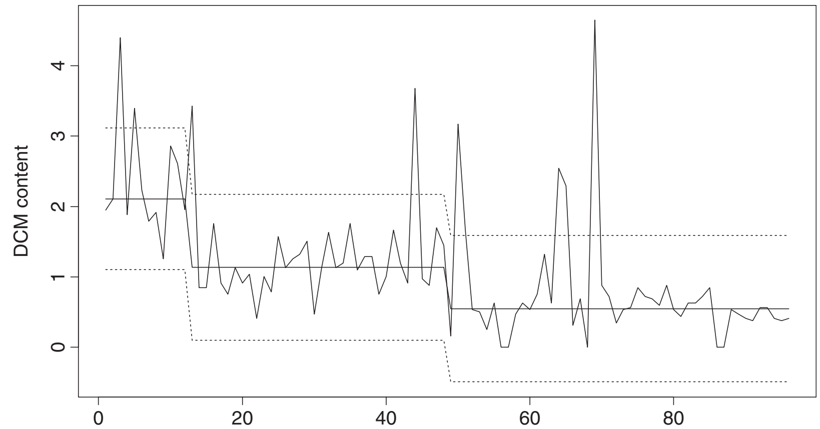


***Figure 5.*** Proposed control chart for DCM percentage measurements.

A quality improvement project was started in order to bring the DCM percentages safely below a certain requirement. In an early stage in the project, DCM percentage measurements were collected from 96 batches of coffee beans. The robust exploratory control chart of these data, which are multiplied by a scalar for reasons of confidentiality, is presented in Fig. 5.

The control chart reveals the presence of several assignable causes: it provides evidence that in the time period in which the 96 measurements were collected eight isolated disturbances occurred as well as two shifts in the mean. Removing the outliers and correcting the remaining measurements for the shifts, the inquirer is left with measurements that can be described reasonably well by a normal distribution with standard deviation 0.35. Note that the identification of outliers is possible because the risk of a false signal is specified per observation (cf. section 4.1, ad. 2). Had we chosen to work with an overall false alarm risk, it would have been unclear what guidance the control limits provide for the identification of assignable causes.

The unstable state of this process is a serious complication for experimentation. In the first instance, improvement efforts should focus on preventive actions against the disturbances. The control chart in the figure gives important indications of their nature and the time instants on which they occurred. A good strategy would be to discuss the control chart with the operators and process engineers who work with the process and, if possible, to consult log books in order to find more information about the process conditions during those time when disturbances occurred.

Figure 6 shows the AMR chart for the DCM percentage measurements. This chart detects the larger outliers. It does not, however, provide indications about the presence of the shifts. Moreover, two of the smaller outliers are not detected. The standard deviation of the in-control process is estimated to be 0.58, which seems too large. This large estimate can be explained by the fact that it includes the additional variation that is caused by the shifts and by the fact that the average moving range is not as robust an estimator as the A-estimator that was used in the proposed procedure.

The AMR chart can be augmented with additional runs rules (Nelson, 1984). These extra tests make the chart more powerful in detecting shifts and drifts. We apply two popular rules to the DCM percentage data:

1. Signal when nine consecutive measurements are on the same side of the center line.
2. Signal when four out of five measurements are on the same side of and more than a distance of 1 sigma from the center line.

The first rule signals at observations 9, 10, 11, 12, 13, 60, 61, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96. The second rule signals at observations 4, 5, 6, 7, 8, 10, 11, 12, 13, 56, 57, 58, 60, 89, 90, 91. Clearly, the chart finds evidence for the shifts. However, it does not provide clear information about their number or the time instants on which they occur. Moreover, the detected shifts are not incorporated in the analysis, and, as a consequence, the chart is less sensitive in detecting the remaining assignable causes. Also CUSUM and EWMA charts



Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

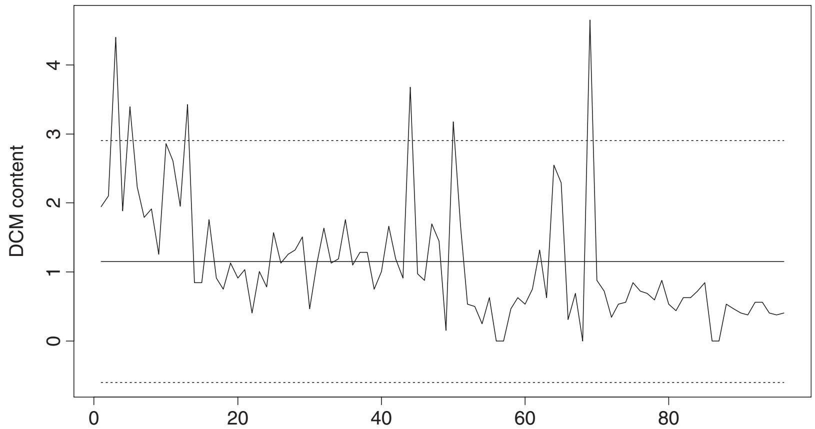


*Figure 6.* AMR chart for DCM percentage measurements.

(with sensibly chosen parameters) detect the shifts but fail to show to what extent the remaining variation is white noise.

The second example demonstrates the performance of our procedure in the case of a smaller data set (30 observations). The example concerns the lengths of biscuits when they leave the oven. In order to identify assignable causes of variation in their length, three biscuits were measured every 2 minutes, for 60 minutes on a row. The resulting data set does not consist of individual measurements. However, the proposed control chart can be applied to the averages of the groups of three measurements. The chart for these averages is presented in Fig. 7.

The chart suggests that the average length has shifted between minute 22 and minute 24. Inquiry with the operators on duty revealed that the speed of a conveyor belt was modified at (precisely) that moment in order to adjust the length a bit. In addition to this adjustment, the average length appears in statistical control, indicating that no assignable causes should be sought on the basis of these measurements.

The traditional chart for subgrouped measurements is the $(\bar{X}, R)$ control chart. The $\bar{X}$-chart is shown in Fig. 8. The control limits are computed from an estimate of the within-groups spread based on the average range (see formula 6–5 in Montgomery (1991), p. 204). The chart gives four signals. The $\bar{X}$-chart signals when there is (a significant amount of) between-groups variation that cannot be accounted for by the within-groups variation alone. In view of this behavior, the four signals would suggest that the length of the biscuits is not perfectly stable in time, which is in itself a valuable suggestion but not specific enough to pinpoint the assignable cause.

If the inquirer is less interested in establishing whether additional between-groups variation is present, but rather wants to know whether this additional variation is random, he could plot an AMR-chart with the average lengths considered as individual observations (this procedure is advocated for the monitoring context in Does et al. (1999)). The chart is shown in Fig. 9. The chart fails to detect the shift in the mean (even when it is augmented with the previously mentioned runs rules). It is concluded that the proposed control chart provides the most revealing presentation of this data set.

## DISCUSSION

The term *control* has little bearing on the application of control charts in exploratory analysis. Consequently, the terms *control chart* and *control limits* are somewhat misleading, and *process behavior chart* and *natural process limits* might be better terms (this terminology is used by Wheeler and Poling, 1998).

It has been argued that charts for monitoring in which the control limits are computed from small samples are misleading (Quesenberry, 1993). Simulations show that the probability that the proposed exploratory chart gives a misleading image is modest, even for smaller sample sizes. Judging a chart to be misleading either if it incorrectly detects a shift or if it gives two or more false alarms for outliers, we find that for a sample size of 40, the probability that the chart is misleading

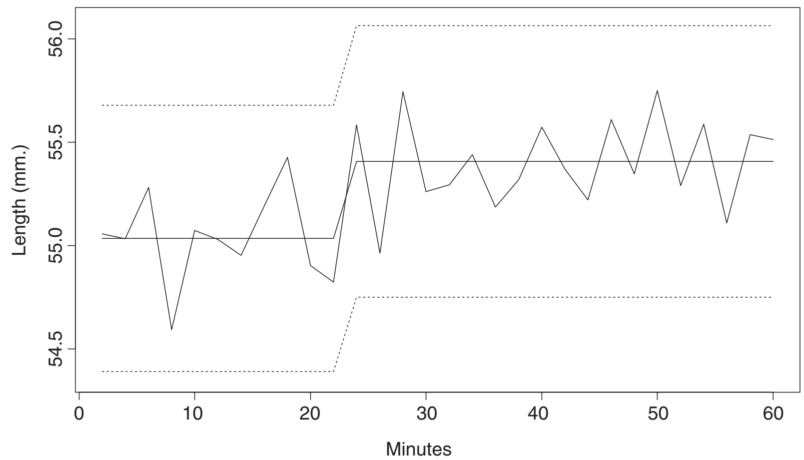


***Figure 7.*** Proposed control chart for the averages of lengths.

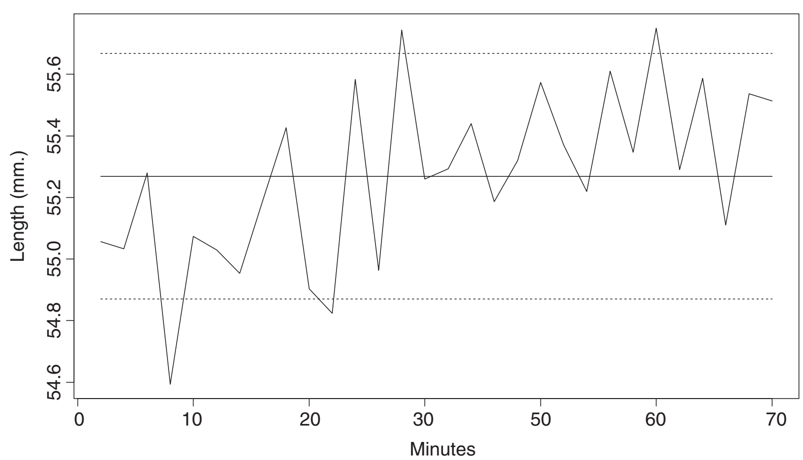


***Figure 8.*** $\bar{X}$-chart for the lengths of biscuits.

is 5.72%. Simulations also confirm the high efficiency of the A-estimator: 84% for a sample size of 40 (compared to the standard deviation). The efficiency of a scale estimator based on the AMR is 61%.

An alternative for using robust estimators for the mean and in-control standard deviation is the following iterative procedure:

1. Plot a provisional control chart of the data.
2. Identify outliers.
3. Recalculate the control limits, ignoring the identified outliers.

Steps 2 and 3 could be done once or iteratively until no more outliers are detected. It should be understood that this procedure is nothing but a robust estimation procedure. The statistical properties of procedures like this, however, are only mediocre to bad (Hampel et al., 1986, pp. 56–71). There is, moreover, an inherent problem in the given procedure: outliers might inflate the error estimate so much that the chart becomes too insensitive to detect even a single outlier (the chart in Fig. 1 comes quite close to this situation). The given iterative procedure is inferior to the use of the A-estimator in the proposed control chart.

The proposed procedure could, and perhaps should, be generalized to include more generic patterns. Obvious candidates include linear (or polynomial) trends, autoregressive or moving average terms, and shifts in the variance. The selected generic patterns should span most of the space of patterns in which assignable causes manifest themselves. The procedure should suggest one or a number of possible

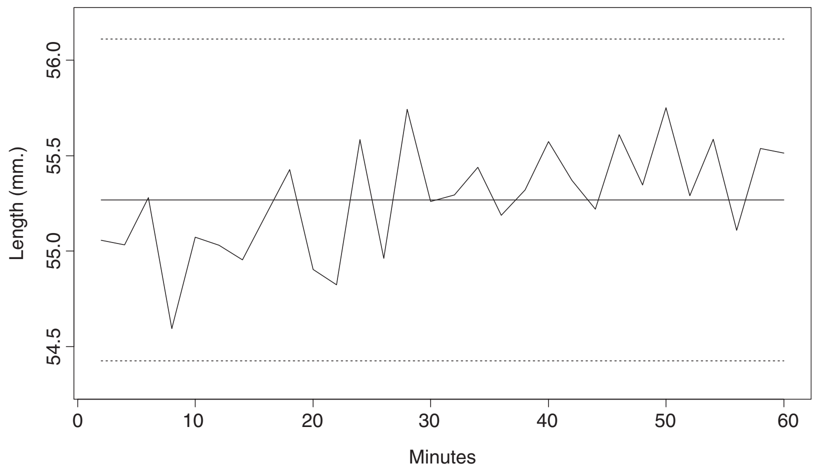


***Figure 9.*** AMR-chart for the averages of lengths.

interpretations of the data and it should indicate to what extent the given interpretations succeed in reducing the measurements to white noise. This procedure appears to be a valuable aid for exploratory data analysis. The inquirer must, however, always look for clues that the automated procedure does not detect. Moreover, the suggested interpretations should be interpreted as hypotheses and further testing should be required before the inquirer can arrive at a final interpretation of a set of data.

## APPENDIX

### Approximating the Distribution of RT$^2$

For suitable choices of $n_1$ and $n_2$, we approximate the distribution of $RT^2/n_1$ [see Eq. (18)] under model Eq. (4) by the $F_{n_1,n_2}$ distribution. This appendix shows how suitable values for $n_1$ and $n_2$ were obtained for various sample sizes $n$.

For fixed $n$, we draw 10,000 samples of size $n$ from a standard normal distribution and we calculate the corresponding 10,000 realizations of $RT^2$. The first two moments of the empirical cumulative distribution function (ECDF) that is induced by these 10,000 values are given by:

$$M_1 = \frac{1}{n}\sum_{i=1}^{10,000} RT_i^2 \qquad M_2 = \frac{1}{n}\sum_{i=1}^{10,000} RT_i^4 \tag{23}$$

The first two moments of the ECDF approximating the distribution of $RT^2/n_1$ can be determined analogously, and equating these with the first two moments of the $F_{n_1,n_2}$ distribution (Johnson and Kotz, 1970), we find:

$$\frac{M_1}{n_1} = \frac{n_2}{n_2 - 2} \qquad \frac{M_2}{n_1^2} = \frac{n_2^2(n_1+2)}{n_1(n_2-2)(n_2-4)} \tag{24}$$

Solving $n_1$ and $n_2$ from these equations gives:

$$n_2 = \frac{4M_2 - 2M_1^2}{M_2 - M_1^2 - 2M_1} \tag{25}$$

$$n_1 = M_1\frac{n_2 - 2}{n_2} \tag{26}$$

### Further Specifics

For $n > 50$, we find $n_2$ to increase rapidly. Since $F_{n_1,n_2}$ converges to $\mathcal{X}_{n_1}^2/n_1$ if $n_2 \to \infty$ (for fixed $n_1$ and $\mathcal{X}_{n_1}^2/n_1$ being the distribution of a statistic having a chi-square distribution with $n_1$ degrees of freedom, divided by $n_1$), we approximate the distribution of $RT^2$ by the $\mathcal{X}_{n_1}^2$ distribution for sample sizes larger than 50. Following the same idea as above, $n_1$ is determined from:

$$n_1 = M_1 \tag{27}$$

For $n \le 20$, we run into the problem that the proposed approximation is inadequate in the remote

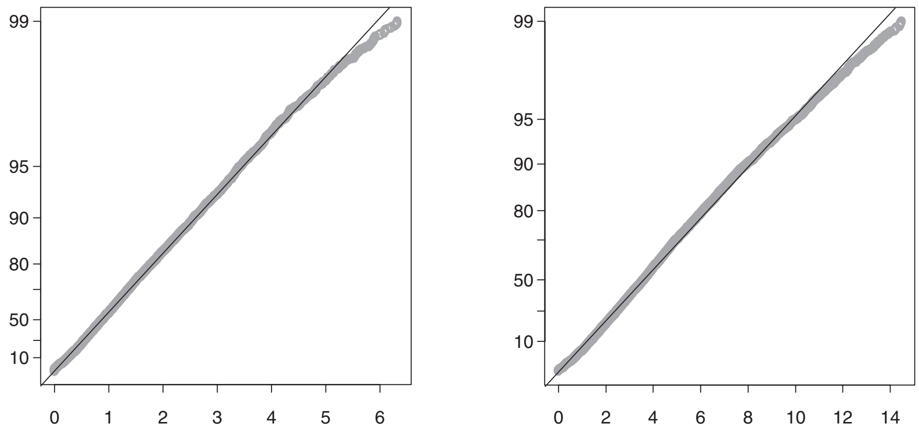


***Figure 10.*** Probability plot for $n=20$ (left) and $n=60$ (right).

tails (beyond the 0.99 quantile); the remote tails of the distribution of $RT^2/n_1$ are heavier than those of the approximating $F$-distribution. These heavy tails have a large impact on the moments of the ECDF; consequently, the found values for $n_1$ and $n_2$ do not give the best approximation on the relevant domain, namely up to the 0.99 quantile. Ideally, we would approximate the ECDF truncated at the 0.99 quantile by an $F$-distribution truncated at the 0.99 quantile. We would then equate moments of these truncated distributions. Unfortunately, it seems not feasible to obtain expressions for the moments of the truncated $F$-distribution.

Another approach to circumvent the problem is inspired by the following idea. Writing $RT^2_{(i)}$ for the ordered $RT^2_i$, we note that the points $(RT^2_{(i)}/n_1, F^{-1}_{n_1,n_2}(\frac{i-1/2}{10{,}000}))$, $i=1,\ldots,9900$ would lie on the line $y=x$ if the approximation were perfect (up to the 0.99 quantile). Using this idea;

$$n_1,n_2=\arg\min_{\nu_1,\nu_2}\sum_{i=1}^{9900}\left(RT^2_{(i)}/\nu_1-F^{-1}_{\nu_1,\nu_2}\left(\frac{i-1/2}{10{,}000}\right)\right)^2 \quad (28)$$

i.e., $n_1$ and $n_2$ are chosen so that they minimize the $L^2$ distance from the line $y=x$. In practice, $n_1$ and $n_2$ were found taking the values of Eq. (25) and Eq. (26) as starting points for a numerical routine minimizing Eq. (28).

The reported values in Table 1 were determined from Eq. (28) (for $n=5$, 6, 8, 10, 15, and 20), Eq. (25), and Eq. (26) (for $n=30$, 40, and 50), or Eq. (27) (for $n=60$, 70, 100, and 150). The adequacy of the proposed approximation for $n=20$ and $n=60$ is demonstrated in Fig 10, which shows the probability plots (up to the 0.99 quantile) for the ECDF. It displays the points $(RT^2_{(i)}/n_1, F^{-1}_{n_1,n_2}(\frac{i-1/2}{10{,}000}))$. The indices on the $y$-axis are $F_{n_1,n_2}(y)$ instead of $y$. The plots are representative of the plots for other values of $n$ and they show that the approximation is accurate up to the 0.97 quantile and fairly accurate up to the 0.99 quantile.

## ABOUT THE AUTHORS

Jeroen de Mast works as a statistical consultant and researcher for the Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA). He teaches courses in six sigma and supervises improvement projects in Dutch industry. He wrote a Ph.D. thesis titled "Quality Improvement from the Viewpoint of Statistical Method." He is a member of the executive committee of the European Network for Business and Industrial Statistics (ENBIS).

Kit C. B. Roes has worked for various Dutch companies as a statistical consultant. He has been head of the Centre for Biometry Wageningen in the Netherlands. Currently, he is director of Biometrics of NV Organon.

## REFERENCES

Box, G. E. P., Ramirez, J. (1992). Cumulative score charts. *Quality and Reliability Engineering International* 8(1):17–27.

Boyles, R. A. (1997). Estimating common-cause sigma in the presence of special causes. *Journal of Quality Technology* 29(4):381–395.

Bryce, G. R., Gaudard, M. A., Joiner, B. L. (1997). Estimating the standard deviation for individuals control charts. *Quality Engineering* 10(2):331–341.

Does, R. J. M. M., Roes, K. C. B., Trip, A. (1999). *Statistical Process Control in Industry*. Dordrecht: Kluwer.

Gross, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *Journal of the American Statistical Association* 71(354):409–416.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986). *Robust Statistics—The Approach Based on Influence Functions*. New York: Wiley.

Hawkins, D. M. (1993). Robustification of cumulative sum charts by winsorization. *Journal of Quality Technology* 25(4):248–261.

Hoaglin, D. C., Mosteller, F., Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. New York: Wiley.

Johnson, N. I., Kotz, S. (1970). *Continuous Univariate Distributions—2*. Boston: Houghton Mifflin.

Koning, A. J., Does, R. J. M. M. (2000). CUSUM charts for preliminary analysis of individual observations. *Journal of Quality Technology* 32(2):122–132.

Langenberg, P., Iglewicz, B. (1986). Trimmed mean $\bar{X}$ and $R$ charts. *Journal of Quality Technology* 18(2):152–161.

Lax, D. A. (1985). Robust estimators of scale: finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association* 80(391):736–741.

Lucas, J. M. (1982). Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology* 14(2):51–59.

Montgomery, D. C. (1991). *Statistical Quality Control*. 2nd ed. New York: Wiley.

Nelson, L. S. (1984). The Shewhart control Chart—tests for special causes. *Journal of Quality Technology* 16(4):237–239.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* 41:100–115.

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

Rocke, D. M. (1989). Robust control charts. *Technometrics* 31(2):173–184.

Roes, K. C. B., Does, R. J. M. M., Schurink, Y. (1993). Shewhart-type control charts for individual observations. *Journal of Quality Technology* 25(3):188–198.

Shewhart, W. A. (1939). *Statistical Method from the Viewpoint of Quality Control*. Washington, D.C.: The Graduate School of the Department of Agriculture. [Reprinted by Dover Publications: New York, 1986].

Shoemaker, L. H. (1984). Robustness properties for a class of scale estimators. *Communications in Statistics A: Theory and Methods* 13(1):15–28.

Sullivan, J. H., Woodall, W. H. (1996). A control chart for preliminary analysis of individual observations. *Journal of Quality Technology* 28(3):265–278.

Tatum, L. G. (1997). Robust estimation of the process standard deviation for control charts. *Technometrics* 39(2):127–141.

Wheeler, D. J., Poling, S. R. (1998). *Building Continual Improvement—A Guide for Business*. Knoxville: SPC Press.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control, with discussion. *Journal of Quality Technology* 32(4):341–378.

Woodall, W. H., Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology* 31(4):377–386.