Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1569

Are estimated control charts in control?

W. Albers and W.C.M. Kallenberg

February 2001

ISSN 0169-2690

Are estimated control charts in control?

Willem Albers* and Wilbert C.M. Kallenberg

Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract Standard control chart practice assumes normality and uses estimated parameters. Because of the extreme quantiles involved, large relative errors result. Here simple corrections are derived to bring such estimated charts under control. As a criterion, suitable exceedance probabilities are used.

Keywords and phrases: Statistical Process Control, Shewhart, exceedance probability, tolerance region.

2000 Mathematics Subject Classification: 62 F 25, 62 P 30.

1 Introduction

Consider the standard situation where the mean of a production process is controlled through a Shewhart $\overline{X}$-chart. An upper and lower limit are set and as soon as an observation exceeds either limit, an out-of-control signal is given. The common approach is to assume normality of the underlying distribution and to estimate the -typically unknown- parameters involved on the basis of earlier ('Phase I') observations. Plugging in these estimates will then (more or less) allow to proceed as in the case of known parameters. The rationale behind this attitude presumably is that exact results are neither attainable nor desired and that all one needs is reasonable guidance for sensible behavior in practice.

Unfortunately, this appealing no-nonsense approach can lead to grossly wrong results which are highly misleading and on which no sensible behavior can be based. To explain this sad state of affairs, note that the probability p of an in

*Corresponding author: tel: +31534893816; fax: +31534893069;
e-mail: w.albers@math.utwente.nl.

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs

control process producing an out-of-control signal, should typically be very small, e.g. $p = 0.001$. But this means that the estimation will involve rather extreme quantiles, and thus large relative errors will result, unless (very) large sample sizes are applied. However, these are often not available in standard control chart practice. In other words, the situation is highly non-robust: modest estimation effects can largely ruin the supposed validity of the outcomes, a result which may indeed be rather counter-intuitive. A recent reference in which the problems concerning estimation in control charts are explicitly mentioned, is Woodall and Montgomery (1999) (see p. 379; also see references to earlier work in this area).

A study of the effects described above, and subsequently of suitable corrections to remedy these effects, will run along the following lines. A characteristic of interest for the given chart is selected, such as the out-of-control signal probability p or the average run length (ARL). Then it is investigated which sample size is needed to ensure that this characteristic is sufficiently close (e.g. by having a relative error of at most 10%) to its counterpart for the case of known parameters. Next modifications are sought which allow to reduce the resulting, typically high, sample sizes to values suitable for practical use.

Note, however, that this description is not yet complete. In the estimated case the performance characteristics involved are random variables, and hence their being close to the corresponding values for known parameters requires further specification. The most obvious choice seems to be to use expectation. Hence, it can for example be investigated when the stochastic counterpart P of p satisfies $|EP - p|/p \leq 0.1$. Alternatively, $E(1/P)$ can be compared to the fixed ARL$=1/p$. This approach is amply studied in Albers and Kallenberg (2000). It is shown that e.g. for $p = 0.001$ well over 300 observations are needed to achieve a 10% relative error. Moreover, using asymptotic methods some easy to apply correction terms for the control limits are derived which allow to reduce these sample sizes to much more acceptable values of about 40.

However gratifying such results may be, we should bear in mind that using expectation only captures part of the picture. Since the variability of P around its expectation is rather large, other aspects of the distribution of P may also shed important light on the effects of estimation in control charts. As an example to show the limitations of using expectation, we mention the at first sight surprising fact that both EP and $E(1/P)$ turn out to have a positive bias. Although this phenomenon is easily explained (see e.g. Albers and Kallenberg (2000)), its consequence remains that opposite corrections are suggested. The first bias invites widening of the control limits, whereas the second bias calls for making these limits a bit more strict.

Consequently, in the present paper we shall consider an approach based on exceedance probabilities. Rather than worrying about $|EP - p|/p > 0.1$, we try to figure out when e.g. $(P - p)/p > 0.1$ occurs with probability at most α, for some α not too large (e.g. $\alpha = 0.1$ or $\alpha = 0.2$). This approach may be slightly more complicated, but it certainly makes sense from a practical point of view. Once the observations from Phase I have produced their estimates, we are so to say stuck with the corresponding outcome of P, which will govern the subsequent behavior of our chart. If such an outcome happens to be uncharacteristically large, it offers small consolation to state that, since the average is well-behaved, this will be balanced by a small outcome on a future application of the estimation procedure. It seems better to supply information on how likely it is that outcomes beyond a certain level occur. (Note the connection to tolerance intervals!)

To avoid duplication, we shall concentrate on the new aspects in going from expectations to exceedance probabilities. For example, we shall not bother here to first show that without corrections very large sample sizes are required, but instead we derive in section 2 the necessary corrections immediately. As we will see, the results thus obtained easily betray that without corrections things actually go quite wrong. This is even more the case as the present criterion turns out to be more strict than the one based on controlling expectation. Such bias reduction still allows a large variability and that is precisely what the exceedance probability approach focusses on.

In section 3 we present a simple approximation formula, which makes transparent how the final behavior is influenced by the underlying parameters. Some illustrative examples are presented. Finally, in section 4 the out-of-control behavior is investigated. Again, a simple approximation is derived to show the impact of the strong type of correction considered here on the out-of-control performance. In this way it becomes possible to strike a proper balance between protection on the one hand and the price to be paid for it on the other.

2 Exact correction terms

Let $X_1, \ldots, X_n, X_{n+1}$ be independent identically distributed random variables (r.v.'s) with a $N(\mu, \sigma^2)$ distribution. The $X_1, \ldots, X_n$ form our sample from Phase I and enable us to estimate the typically unknown μ and σ. The additional X_{n+1} belongs to the monitoring phase (Phase II). Just as in Albers and Kallenberg (2000), we restrict attention to control charts with an upper limit only. The two-sided case can be treated analogously. Likewise, generalization to the situation of n groups of m (e.g. $m = 5$) observations, rather than n individual observations, is rather straightforward and will not be considered here. Note that we are considering the in-control situation; at the end of the paper we shall also look at what happens when the distribution of X_{n+1} differs from that of $X_1, \ldots, X_n$.

Let p again be the probability of an out-of-control signal for an in control process, then the upper control limit (UCL) is given by

$$\mu + u_p \sigma \tag{1}$$

with $u_p = \Phi^{-1}(1-p) = \overline{\Phi}^{-1}(p)$, where Φ is the standard normal distribution function, $\overline{\Phi} = 1 - \Phi$ and Φ^{-1} and $\overline{\Phi}^{-1}$ are their inverse functions, respectively.

As μ and σ are unknown, the UCL in (1) has to be replaced by an estimated version $\hat{\mu} + u_p\sigma^*$. As concerns the choices for $\hat{\mu}$ and σ^*, let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \overline{X})^2, c_4(n) = \frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}, \quad (2)$$

then we always use $\hat{\mu} = \overline{X}$, together with either $\sigma^* = S$ or $\sigma^* = \hat{\sigma} = S/c_4(n)$. The latter choice is unbiased, as $ES = c_4(n)\sigma$. In the introduction the need for and usefulness of corrections was argued, and hence the next step consists of including such a term: we shall consider estimated UCL's of the form $\hat{\mu} + (u_p + c)\sigma^*$. Hence the standardized UCL u_p is simply increased or decreased somewhat, in order to achieve compliance with the criterion considered for moderate n already.

The stochastic counterpart P of p now is given by

$$P = P(X_{n+1} > \hat{\mu} + (u_p + c)\sigma^*) = \overline{\Phi}(\frac{\hat{\mu} - \mu}{\sigma} + (u_p + c)\frac{\sigma^*}{\sigma}). \quad (3)$$

The performance of the estimated chart can be judged by comparing $g(P)$ to $g(p)$ for some suitable positive function g. Common choices are $g(p) = p, 1/p$ and $1 - (1-p)^k$. The first is evident, the second corresponds to the ARL, while the last represents the probability that the run length is at most some prescribed k (typically k will be a rather small fraction, like 10% or 25%, of the ARL=$1/p$). For increasing g (e.g. the first and third of the examples above), we shall require that

$$P((g(P) - g(p))/g(p) > \varepsilon) \leq \alpha, \quad (4)$$

with $\varepsilon \geq 0$, and $\alpha > 0$ small. Usually we let $\varepsilon = 0.1$, whereas α may vary from 0.05 to 0.25. (For decreasing g (e.g. the second example above), we obviously replace (4) by $P((g(P) - g(p))/g(p) < -\varepsilon) \leq \alpha$). Hence the relative error of $g(P)$ in the relevant direction should be small, except for a small probability.

From (3) and (4) it immediately follows that the left-hand side of (4) can be rewritten into

$$P(\frac{\hat{\mu} - \mu}{\sigma} + (u_p + c)\frac{\sigma^*}{\sigma} < \overline{\Phi}^{-1}(g^{-1}(g(p)(1+\varepsilon)))) \quad (5)$$

(with the result for decreasing g obtained by using $(1-\varepsilon)$ instead of $(1+\varepsilon)$). This expression has the form $P(n^{-1/2}Z + aS/\sigma < b)$, with $Z\, N(0,1)$ and Z and S independent. But this latter probability in its turn equals $P(-n^{-1/2}Z - aS/\sigma > -b) = P(n^{-1/2}Z + b > aS/\sigma) = P((Z + n^{\frac{1}{2}}b)/(S/\sigma) > n^{-\frac{1}{2}}a) = \overline{G}_{n-1,n^{\frac{1}{2}}b}(n^{\frac{1}{2}}a)$, where $G_{n-1,\delta}$ stands for the distribution function of the noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter δ, and $\overline{G}_{n-1,\delta} = 1 - G_{n-1,\delta}$ (cf. Ghosh, Reynolds and Hui (1981) for earlier use of the noncentral t in this connection). Consequently, an outcome α will result for $a =$

$n^{-1/2}\overline{G}^{-1}_{n-1,n^{\frac{1}{2}}b}(\alpha)$. For $\sigma^* = S$, we have $a = u_p + c$ and hence the correction c which precisely produces equality in (4) for increasing g, is given by

$$c = n^{-1/2}\overline{G}^{-1}_{n-1,n^{\frac{1}{2}}b}(\alpha) - u_p, \tag{6}$$

with $b = \overline{\Phi}^{-1}(g^{-1}(g(p)(1+\varepsilon)))$. (For $\sigma^* = \hat{\sigma}$, the factor $n^{-1/2}$ in (6) becomes $n^{-1/2}c_4(n)$, as $a = (u_p + c)/c_4(n)$ in this case; for decreasing g the factor $(1+\varepsilon)$ in b becomes $(1-\varepsilon)$.)

Hence for each of our three g's and any combination of n, p, ε and α, the exact correction c can now readily be evaluated using (6). In Table 1 we provide some examples for the simplest case $g(p) = p$, just to show that the resulting corrections are indeed quite (to very) large. To illustrate the impact of such large corrections, we accompany these values in Table 1 by values of

$$\tilde{\alpha} = \overline{G}_{n-1,n^{\frac{1}{2}}b}(n^{\frac{1}{2}}u_p), \tag{7}$$

which is nothing but the realized value of the exceedance probability from (4) when no correction is used (i.e. $a = u_p$ rather than $a = n^{1/2}\overline{G}^{-1}_{n-1,n^{\frac{1}{2}}b}(\alpha)$).

Table 1 Values of $c(= c_\alpha)$ from (6) and $\tilde{\alpha}$ from (7) for $p = 0.001, \varepsilon = 0.1, \alpha = 0.1$ and $\alpha = 0.2$, using $g(p) = p$

n	$c_{0.1}$	$c_{0.2}$	$\tilde{\alpha}$
25	0.76	0.48	0.51
50	0.48	0.30	0.49
75	0.37	0.23	0.48
100	0.31	0.19	0.47
200	0.21	0.12	0.45
500	0.12	0.065	0.40
1000	0.072	0.037	0.36
2000	0.042	0.017	0.30
5000	0.015	0.0004	0.203

Following simple heuristics, one would suspect $\tilde{\alpha}$ to start well below $\frac{1}{2}$ (as $\frac{1}{2}$ is 'fair' for $\varepsilon = 0$) and to drop off quickly for n as rapidly increasing as in Table 1. Actually, $\tilde{\alpha}$ starts even above $\frac{1}{2}$ (due to the positive bias of P), and decreases very slowly indeed: even for $n = 2000$ and $\alpha = 0.2$ we still have to reach the point where the correction becomes superfluous.

3 Approximation and illustration

The result in (6) is exact, but in itself not very illuminating. Clearly (cf. Table 1), it serves to show that the estimation effects are definitely nonnegligible, but it will be difficult to discover from (6) how matters depend on $n, p, \varepsilon, \alpha$ and g. And precisely such qualitative information would be quite valuable to allow sensible behavior in practice. Hence in the present section we shall derive a simple and transparant approximation which does reveal such relationships.

As S/σ is asymptotically $N(1, 1/(2n))$, it is immediate that $P(n^{-1/2}Z + aS/\sigma < b)$ can be approximated by

$$\overline{\Phi}\Big((a-b)\Big(\frac{a^2+2}{2n}\Big)^{-1/2}\Big). \tag{8}$$

This equals $\overline{\Phi}(u_\alpha) = \alpha$ if $a = u_\alpha\{(a^2+2)/(2n)\}^{1/2} + b$. As $c_4(n)$ from (2) satisfies $c_4(n) \approx 1 - 1/(4n)$, we arrive for both $a = (u_p + c)$ and $a = (u_p + c)/c_4(n)$ at the following approximation to (6):

$$c \approx u_\alpha\Big(\frac{u_p^2+2}{2n}\Big)^{\frac{1}{2}} + (b - u_p). \tag{9}$$

Finally, for ε small (as will usually be the case), $(b - u_p) = \overline{\Phi}^{-1}(g^{-1}(g(p)(1+\varepsilon))) - u_p \approx \overline{\Phi}^{-1}(p + \varepsilon g(p)/g'(p)) - u_p \approx -\varepsilon g(p)/\{\phi(u_p)g'(p)\} \approx -(\varepsilon/u_p)h(p)$ with $\phi = \Phi'$ and $h(p) = g(p)/(pg'(p))$. (Again, for decreasing g, replace $+\varepsilon$ by $-\varepsilon$!) For $g(p) = p$ we get $h(p) = 1$, while for $g(p) = 1/p$ we have $h(p) = -1$. As g is decreasing in the latter case, in either situation we obtain $b - u_p \approx -\varepsilon/u_p$. The third g leads to $h(p) = (1-p)\{(1-p)^{-k} - 1\}/(kp) \approx (e^{kp} - 1)/kp$. Hence a further approximation step produces the simple expression

$$c \approx u_\alpha\Big(\frac{u_p^2+2}{2n}\Big)^{\frac{1}{2}} - \frac{\varepsilon}{u_p}\lambda, \tag{10}$$

with $\lambda = 1$ for the first two choices of g and $\lambda = (e^{kp} - 1)/kp$ for the third. As concerns the uncorrected exceedance probability $\tilde{\alpha}$ from (7), it is immediate from (8) – (10) that it equals

$$\tilde{\alpha} \approx \overline{\Phi}\Big(\Big(\frac{u_p^2+2}{2n}\Big)^{-1/2}(u_p - b)\Big) \approx \overline{\Phi}\Big(\Big(\frac{u_p^2+2}{2n}\Big)^{-1/2}\frac{\varepsilon\lambda}{u_p}\Big). \tag{11}$$

Next, we exploit (10) and (9) to study the behavior of the correction in relation to $n, p, \varepsilon, \alpha$ and g.

1) role of n

In Albers and Kallenberg (2000) it was demonstrated that a term $u_p(u_p^2+2)/(4n)$ is dominant in correcting the bias of P. From (10) we see that here a term $u_\alpha\{(u_p^2+2)/(2n)\}^{1/2}$ occurs. Hence it indeed turns out that controlling the variability calls for stronger intervention: it requires a term of order $n^{-1/2}$ rather than of order n^{-1}. Consequently, much larger sample sizes are needed before this type of correction becomes negligible and thus superfluous. To be more specific, note that 'no correction' means $c = 0$. Hence (9) and (10) immediately supply approximations to the minimal sample size for which (4) is met without correction:

$$n \approx \frac{1}{2}u_\alpha^2(u_p^2+2)/(u_p-b)^2 \approx \frac{1}{2}u_\alpha^2 u_p^2(u_p^2+2)/(\varepsilon^2\lambda^2). \tag{12}$$

For example, for $p = 0.001, \varepsilon = 0.1, \alpha = 0.25$ and $g(p) = p$, we find $n \approx 3250$ and $n \approx 2500$, respectively. Using the exact approach from the previous section, we find for these values of n that the realized value of α equals 0.252 and 0.28, respectively. Hence (9) is quite accurate, but even the more simple (10) suffices to get a clear picture of what is going on in practice. Returning to c itself, we get for the above p, ε and $g(p) = p$ that (9) reduces to

$$c \approx 2.40 u_\alpha n^{-1/2} - 0.028, \tag{13}$$

(for (10) replace the last number in (13) by -0.032), which for $\alpha = 0.1$ and $\alpha = 0.2$ boils down to $c = 3.08n^{-1/2} - 0.028$ and $c = 2.02n^{-1/2} - 0.028$, respectively. It is easily checked that these simple expressions nicely explain the corresponding exact values $c_{0.1}$ and $c_{0.2}$ from Table 1.

2) role of p

As mentioned in the introduction, things go (much) less well than intuitively anticipated because we are dealing with small quantiles. Indeed, if we for example go from $p = 0.001$ to $p = 0.01$, matters improve, as u_p decreases. The n from (12) is reduced by a factor 0.4, leading in the example above to $n \approx 1300$ and $n \approx 900$, respectively (with corresponding realized values of α of 0.253 and 0.29, respectively). Likewise, (13) becomes $c = 1.93u_\alpha n^{-1/2} - 0.036$ (replace the last number by -0.043 for (10) rather than (9)). Hence c is reduced in two respects: the coefficient of the positive term decreases, whereas the magnitude of the negative term increases. Of course, this is in view of (9) hardly surprising, since $(u_p^2+2)^{\frac{1}{2}} = u_p(1+2/u_p^2)^{1/2} \approx u_p + 1/u_p$ behaves like u_p, while $b - u_p$ behaves like $-\varepsilon\lambda/u_p$.

3) role of ε

Clearly, as ε decreases, it becomes harder to satisfy (4) and both c in (9) and (10) and n in (12) will grow. In the limiting case $\varepsilon = 0$ a correction can only be avoided by taking $\alpha = 1/2$, which is a simple consequence of the fact that $g(P)$ is asymptotically normal with mean $g(p)$. The role of ε of reversing to some extent the need for correction (cf. its negative sign in (10)) is, as we saw above, weakened as p decreases. Taking once more $p = 0.001$ and $\alpha = 0.1$, we obtain for small ε through (10) that $c \approx 3.08n^{-1/2} - 0.32\varepsilon\lambda$. This shows that the effect of the second term is rather limited, unless n is really large. Of course, matters change when ε is not small. E.g. for $p = 0.001$ and $g(p) = p$ we then get from (9) that $c \approx 2.40u_\alpha n^{-1/2} - \{3.09 - \overline{\Phi}^{-1}(0.001(1+\varepsilon))\}$. For e.g. $\varepsilon = 9$, the last term equals 0.764, which implies that $c \approx 0$ for $u_\alpha n^{-1/2} \approx 0.318$, i.e. $n \approx 9.90u_\alpha^2$. For $\alpha = 0.05$, this produces $n = 27$. Hence the combination of a small p, a small n and a small α indeed leads to trouble: without correction, P can be off by a huge factor (e.g. 10, as in this example). Incidentally, the approximation again works quite well: from (6) we get for the values under consideration $c = 0.03$, which is indeed rather negligible compared to $u = 3.09$.

4) role of α

This is pretty straightforward: in both (9) and (10), the first term is linear in u_α. Similar remarks can be made and similar examples can be given as in *1) – 3)* above.

5) role of g

For small ε, the role of g is simply given by λ in (10). For both $g(p) = p$ and $g(p) = 1/p$, we get $\lambda = 1$, as $P(\frac{1}{P} < \frac{1}{p}(1-\varepsilon)) = P(P > p(1 + \frac{\varepsilon}{1-\varepsilon}))$ and $\varepsilon/(1-\varepsilon) \approx \varepsilon$ for ε small. Hence it is quite natural that the correction works out the same way for both the probability of an out-of-control signal and for the ARL. In this respect the present criterion is more conforming one's intuition than the expectation, which suggests opposite corrections for these two cases (cf. Albers and Kallenberg (2000)).

The third choice $g(p) = 1-(1-p)^k$ leads to $\lambda = (e^{kp}-1)/kp \approx 1+kp/2$ for kp small. This latter case is usually considered to be of main interest: rather than relying on the ARL, many practitioners prefer to watch the probability that the run length is uncharacteristically small, e.g. less than 10 or 20% of the expected value. Consequently, the moderating effect on the size of the correction may be somewhat larger for the third choice, but the difference will typically not be substantial. All in all we can say that preference for one g or another fortunately has little impact on the actual correction applied. To illustrate this numerically, just reconsider the example from (13). For (10), the last number in (13) now should be replaced by -0.032 λ, with $\lambda = \{\exp(kp) - 1\}/kp$ rather than simply $\lambda = 1$. For $k = 10$ and $k = 200$, this leads to -0.034 and -0.036, respectively.

Summarizing the exposition above, it illustrates that without correction, and with p small and n not large, one should not have great expectations about α and ε! Approximation (9) and (10) make transparant how bad the actual situation is, and what needs to be done to achieve control again.

4 Out-of-control

In this final section we shall briefly consider the out-of-control situation, where X_{n+1} comes from a $N(\mu + d\sigma, \sigma^2)$-distribution, for some $d > 0$. Hence P from (3) transforms into

$$P = \overline{\Phi}\Big(\frac{\hat{\mu} - \mu}{\sigma} + (u_p + c)\frac{\sigma^*}{\sigma} - d\Big). \tag{14}$$

Moreover, arguing as in section 3, we immediately obtain that, for both $\sigma^* = \hat{\sigma}$ and $\sigma^* = S$, the quantity of interest $g(P)$ will behave like

$$g(\overline{\Phi}(u_p - d + c + \tau Z)), \tag{15}$$

with $Z\, N(0,1)$ and $\tau^2 = (u_p^2 + 2)/(2n)$.

The point now is how (15) relates to the corresponding quantity $g(p_1)$, with $p_1 = \overline{\Phi}(u_p - d)$, for the case of known parameters. To put it more bluntly, it may be very nice that the in-control behavior has been corrected, but to what extent do such corrections damage the out-of-control performance? Getting clarity here as well, will allow to find the proper balance in practice. To this end, we begin by making a comparison to the situation considered in Albers and Kallenberg (2000). There it is shown that, for the in-control situation, $g(P)$ has an unacceptably large bias, unless n is very large, a situation which can be remedied by using a c of order n^{-1}. Moreover, it is demonstrated that this type of correction does not disturb the out-of-control behavior. This fortunate state of affairs is, at least partly, based on the phenomenon that the relative error due to estimation effects is considerably smaller around p_1, than around p (cf. the remarks on difficulties concerning estimation of extreme quantiles in the introduction).

As this same effect clearly is present here, we arrive at the following qualitative appraisal of the overall picture: bias removal requires moderate corrections (i.e. of order n^{-1}) with small effects under out-of-control , while bounding exceedance probabilities requires large corrections (i.e. of order $n^{-1/2}$) with moderate out-of-control effects. In other words, the practitioner can buy weak or strong protection at a low or moderate price, respectively.

Obviously, the above is just a general description, trying to convey the flavor of either approach. Specific outcomes may vary, also in view of the ample choice in (values of) parameters. Also, the behavior of $g(P)$ can again be characterized in various ways. For brevity, we just take the simplest one, using expectation:

$$E_1 g(P) \approx g(\overline{\Phi}(u_p - d + c)) \approx g(p_1) - cg'(p_1)\phi(u_p - d), \tag{16}$$

where the last step only makes sense for c still reasonably small. If that is the case, the relative error $(E_1 g(P) - g(p_1))/g(p_1)$ approximately equals

$$-c\frac{g'(p_1)}{g(p_1)}\phi(u_p - d). \tag{17}$$

For $g(p) = p$ this for example gives $-c\phi(u_p - d)/p_1$, which for $0 \leq u_p - d \leq 3$ can be approximated in its turn by

$$-\frac{4}{5}c\{1 + (u_p - d)\}. \tag{18}$$

(Note that for $g(p) = 1/p$ the result from (18) holds with its sign reversed.)

Combination of (16) – (18) with (10) or (9) clearly allows to appraise the impact of c on the out-of-control performance both in a transparant and quantitative manner. Letting once more $p = 0.001$, we obtain e.g. for $d = 1.44$ that $p_1 = 0.05$, while (18) boils down to $-2.12c$ for $g(p) = p$ (and $2.12c$ for $g(p) = 1/p$). Likewise, for $d = 2.25$, we get $p_1 = 0.20$ and $-1.47c$ (or $1.47c$). For e.g. $\varepsilon = 0.1$ and $\alpha = 0.2$, we obtained in section 3 from (13) and (10) that $c \approx 2.02n^{-1/2} - 0.032$, which for $n = 100$ produces $c \approx 0.17$, and thus $2.12c \approx 0.36$ and $1.47c \approx 0.25$. Hence the average out-of-control ARL $1/p_1$ is increased from 20 to 27 and from 5 to 6.3, respectively. This example shows how to find a proper balance between adequate protection against estimation effects during in-control, and the price to be paid for it during out-of-control.

References

Albers, W. and Kallenberg, W.C.M. (2000). Estimation in Shewhart control charts. Technical Report 1559, University of Twente

Ghosh, B.K., Reynolds, M.R.Jr. and Hui, Y.V. (1981). Shewhart $\overline{X}$-charts with estimated process variance. *Comm. Statist. Theory & Methods* **10**, 1797-1822.

Woodall, W.H. and Montgomery, D.C. (1999). Research issues and ideas in statistical process control. *J. Qual. Technol.*, **31**, 376-386.