A Change-Point Model for a Shift in Variance

DOUGLAS M. HAWKINS

School of Statistics, University of Minnesota

K. D. ZAMBA

College of Public Health, Department of Biostatistics, The University of Iowa

A control chart for detecting shifts in the variance of a process is developed for the case where the nominal value of the variance is unknown. As our approach does not require that the in-control variance be known a priori, it avoids the need for a lengthy Phase I data-gathering step before charting can begin. The method is a variance-change-point model, based on the likelihood ratio test for a change in variance with the conventional Bartlett correction, adapted for repeated sequential use. The chart may be used alone in settings where one wishes to monitor one-degree-of-freedom chi-squared variates for departure from control; or it may be used together with a parallel change-point methodology for the mean to monitor process data for shifts in mean and/or variance. In both the solo use and as the scale portion of a combined scheme for monitoring changes in mean and/or variance, the approach has good performance across the range of possible shifts.

KEY WORDS: Average Run Length; Change Point; Control Chart; CUSUM; EWMA; Likelihood Ratio; Phase I; Phase II; Shewhart.

Introduction

THE PURPOSE of control charts is the detection of instants when the process has gone out of control. This detection is the vital precursor to an investigation of the assignable causes of the loss of control. Traditionally, the first step of setting up a charting scheme is to gather a substantial data set in a Phase I study. After data cleaning, this Phase I data set is used to assess the in-control distribution of the process readings and to estimate its parameters—in the simplest case, the mean and standard deviation.

Following the Phase I study, the user then goes on to Phase II monitoring using the calibrated control charts. The most widely used charting framework is to chart means and variances to monitor possible shift in location and scale. A frequent, if often unvoiced, implied supposition is that a model of independent process readings from the normal distribution provides an adequate working model. We will also use this conceptual framework, concentrating on the question of detecting changes in the variance. Historically, emphasis was on detecting increases in variance, as these were a sign of trouble. For example, Shewhart's R chart is incapable of signaling a decrease in variance using small rational groups, and some scale proposals—for example, Crowder and Hamilton (1992)—use this reason to search only for variance increases. Current thinking, by contrast, recognizes variance decreases as the key to quality improvement, and leads to a realization that proper control on variance requires the capability to detect decreases as well as increases.

Shewhart methodology uses the R or S chart for the process variance. This requires rational groups

Dr. Hawkins is Professor, School of Statistics, University of Minnesota, and an ASQ Fellow.

Dr. Zamba is a Visiting Assistant Professor, College of Public Health, Department of Biostatistics, The University of Iowa.

of size $n > 1$. Monitoring the variability of individuals has traditionally been done using a moving range (MR) chart. This, however, has been shown to be only minimally effective (Roes, Does, and Schurink (1993), Rigdon, Cruthis, and Champ (1994)).

Where rational groups of size > 1 are taken, an alternative to the R or S chart is a CUSUM or EWMA control chart. For a more complete exposition of CUSUMs in general and variance CUSUMs in particular, see Hawkins and Olwell (1998). A valuable discussion of both some technical considerations and the choice of rational group size is given by Reynolds and Stoumbos (2004a, 2004b).

The CUSUM has attractive optimality properties for the detection of step changes in parameters. Note, though, that to achieve this optimality requires advance knowledge of both the in-control and out-of-control true variances. This is true for all the standard tools for detecting persistent shifts. These requirements are severe limitations in startup and short-run processes.

Standard practice has been to substitute the Phase I estimates of process parameters for the true parameters and to treat these estimates as if they were the exact true parameters. This, however, is far from the truth. Even large Phase I samples leave considerable random variability in parameter estimates, and these random errors of estimation then carry over to distort the Phase II performance of the charts—see, for example, Quesenberry (1991), Hawkins and Olwell (1998, ch. 7), Jones and Champ (2001), Jones et al. (2004). This distortion is particularly severe for sensitive methods, such as the CUSUM or EWMA. Furthermore, the most obvious remedy—larger Phase I samples—which is unattractive at the best of times, is impossible in short-run or startup settings.

The Change-Point Formulation

The CUSUM is the optimal diagnostic for detecting a step change from one specified in-control level to some other specified out-of-control level. A statistical formulation that encapsulates this for normal data is the "change-point" formulation,

$$X_i \sim \begin{cases} N(\mu_1, \sigma_1) & \text{if } i \leq \tau, \\ N(\mu_2, \sigma_2) & \text{if } i > \tau. \end{cases} \tag{1}$$

The parameter τ is called the change point (if one exists) and is unknown. The remaining parameters μ_1, μ_2, σ_1, σ_2 are the in- and out-of-control means and standard deviations. A location shift occurs if $\mu_1 \neq \mu_2$; a scale shift occurs if $\sigma_1 \neq \sigma_2$. Either or both of these shifts could, in principle, occur.

We will be focussing on the unknown-parameter change-point problem in which none of the parameters is assumed known a priori. Using the change-point model in SPC leads to two statistical tasks: testing for the presence of a change, and estimating the parameters. Whether in SPC, regression analysis, image processing, or discontinuity curves and surfaces, the standard approach to these problems is to split the data into time-ordered segments and test for differences between the segments using some heterogeneity criterion. Most work on change-point problems has focussed on the fixed-sample setting, such as is seen in the analysis of Phase I data. A thorough review of this setting and various techniques addressing it can be found in Bhattacharya (1994), and relevant individual papers include Sen and Srivastava (1975), Hawkins (1977), Worsley (1979), Worsley (1982), and Sullivan and Woodall (1996).

Phase II application of the change-point problem with at least some parameters prespecified is addressed implicitly in Siegmund and Venkatraman (1994), Pollak and Siegmund (1991), and the excellent review paper by Lai (2001). The setting in which none of the parameters is known in advance has received much less attention. Hawkins, Qiu, and Kang (2003) (which we abbreviate to HQK) address the constant-variance mean-change-point formulation in which the mean might change but the variance remains constant.

$$X_i \sim \begin{cases} N(\mu_1, \sigma) & \text{if } i \leq \tau, \\ N(\mu_2, \sigma) & \text{if } i > \tau \end{cases} \tag{2}$$

($\tau, \mu_1, \mu_2, \sigma$ all unknown.) The HQK paper adapts the classic fixed-sample change-point formulation to a Phase II setting by defining a repeated testing framework in which, as each new observation accrues, the change-point test is reapplied to all accumulated data, but this is done in such a way that the probability of a false alarm remains constant.

This paper develops a parallel methodology for detecting changes in variance, with no assumption of the constancy or otherwise of the means.

Testing for a Change-Point in Variance

Let X_i be the sequence of process readings, assumed to follow the change-point model of Equation (1).

Suppose n process readings have accrued and, for $0 \le i < k \le n$, define the summary statistics

$$\bar{X}_{i,k} = \sum_{j=i+1}^{k} X_j/(k-i) \tag{3}$$

$$V_{i,k} = \sum_{j=i+1}^{k} (X_j - \bar{X}_{i,k})^2. \tag{4}$$

Suppose it were known that the change-point was at instant $\tau = k$. Estimates of the remaining parameters would then be

$$\hat{\mu}_1 = \bar{X}_{0,k} \tag{5}$$
$$\hat{\mu}_2 = \bar{X}_{k,n} \tag{6}$$
$$\hat{\sigma}_1^2 = V_{0,k}/(k-1) \tag{7}$$
$$\hat{\sigma}_2^2 = V_{k,n}/(n-k-1), \tag{8}$$

and under the constant-variance assumption, the estimate for the common variance would be

$$\hat{\sigma}^2 = (V_{0,k} + V_{k,n})/(n-2).$$

These variance estimates have the usual degrees-of-freedom divisor to make them unbiased. The likelihood ratio test statistic for equality of variance is (Bartlett and Kendall (1946), Bartlett (1937), Bartlett (1955))

$$G_{k,n} = \left[(k-1)\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_1^2}\right) + (n-k-1)\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_2^2}\right)\right]/C$$

where C is the Bartlett correction factor

$$C = 1 + [(k-1)^{-1} + (n-k-1)^{-1} - (n-2)^{-1}]/3.$$

$G_{k,n}$ has an approximate null chi-squared distribution with 1 degree of freedom. Not surprisingly, $G_{k,n}$ can be written as a function of the familiar F ratio for testing equality of variance: if $F = \hat{\sigma}_2^2/\hat{\sigma}_1^2$, then

$$\begin{aligned} G_{k,n} &= C^{-1}[(k-1)\ln\{k-1+(n-k-1)F\} \\ &\quad + (n-k-1)\ln\{(k-1)F^{-1} + n-k-1\} \\ &\quad - (n-2)\ln(n-2)], \end{aligned}$$

showing that $G_{k,n}$ is a two-sided test, giving large values for F ratios substantially above or below 1. Note, however, that while the distribution of F depends strongly on the values of k and $n-k-1$, the approximate null distribution of $G_{k,n}$ is the same regardless of n and k.

Turning to the situation of a possible change in variance at an unknown change-point τ, the generalized likelihood ratio test is given by maximizing $G_{k,n}$ across all possible k values. As regards the range of k, note that G is undefined if $k = 1$ or $k = n-1$, so that the permissible range of k is $2 \le k \le n-2$.

The GLR test statistic for a change in variance is therefore

$$G_{\max,n} = \max_{2\le k\le n-2} G_{k,n}.$$

Ongoing Monitoring for Variance Change

So far, we have discussed the likelihood ratio test statistic and its asymptotic distribution as a testing problem for a single static sample of size n, as in a Phase I setting. And, indeed, these static change-point formulations provide a potentially valuable tool for verifying that a Phase I sample was homogeneous, as implied by the in-control assumption. It is, however, its potential for ongoing process monitoring that is more exciting.

Paralleling the HQK procedure for monitoring the mean, we will assume that an ongoing stream of process data accrues and that a step change in the variance may occur at any instant. The SPC procedure is then defined by

- After observation n has been added to the total record of the process, compute the statistic $G_{\max,n}$—the GLR for a shift in variance at some unknown previous time.
- If $G_{\max,n} \le h_n$, where h_n is some suitable control limit, then conclude that there is no evidence of a variance shift and leave the process running uninterrupted.
- If, however, $G_{\max,n} > h_n$, then conclude that there is evidence of a variance shift.

Note some important differences between this approach and the more traditional SPC approaches using Shewhart, CUSUM or EWMA charts. One is the point already made—that the traditional charts are commonly calibrated by plugging in parameter estimates from a Phase I data set, but that these estimates are treated as if they were the true values. Our formulation is explicit that the estimates are no more than estimates and aims to control the false-alarm rate without any assumption of knowledge of unknowns.

The other difference is that, unlike the conventional setting, where learning about the true parameter values stops along with the Phase I data gathering, the change-point formulation continues to use the ongoing stream of process data to refine its estimates of process parameters.

The Choice of the Control Limits

There is a vital issue remaining to be resolved, and that is the choice of the sequence of control limits h_n. An intuitively desirable property of the sequence (HQK, Margavio et al. (1995)) would be to give a constant probability of a false alarm for each n. If this probability is to be a constant α, say, then the sequence, which we will relabel $h_{n,\alpha}$, must satisfy

$$P[G_{\max,n} > h_{n,\alpha} \mid G_{\max,j} \leq h_{j,\alpha},\ j < n] = \alpha. \quad (9)$$

It does not seem to be possible to solve for these $h_{n,\alpha}$ values theoretically, and so a simulation was used to estimate them. This involved 5 million random samples of sizes up to 500 and covering α values of 0.05, 0.02, 0.01, 0.005, 0.002, and 0.001 and gave fractiles $h_{n,\alpha}$ with standard errors of approximately 0.02.

Having said that it is possible to start testing right from the fourth observation (a minimum of two observations per segment being required to calculate the test statistic), we think that few practitioners would do so. We think ordinary prudence would lead one, except in short-run settings, to accumulate a few observations before starting testing. Our simulations incorporated this possibility, giving rise to 17 tables of $h_{n,\alpha}$ values, corresponding to initial skips of the first 4, 5, 6, ..., 20 observations. The full table can be found on the Web site `www.stat.umn.edu/hawkins`. Table 1, extracted from this fuller table, is the table of cutoffs for skipping nine initial observations and starting testing with the tenth.

TABLE 1. Control Limits $h_{n,\alpha}$ Starting Testing at Sample 10

	α					
n	0.05	0.02	0.01	0.005	0.002	0.001
10	6.374	8.003	9.229	10.451	12.039	13.238
11	5.651	7.328	8.585	9.840	11.489	12.734
12	5.357	7.077	8.373	9.653	11.357	12.631
13	5.228	6.988	8.312	9.634	11.367	12.672
14	5.173	6.960	8.304	9.658	11.423	12.760
15	5.149	6.960	8.323	9.692	11.469	12.828
16	5.141	6.974	8.357	9.731	11.541	12.885
17	5.145	6.992	8.386	9.776	11.596	12.962
18	5.142	7.010	8.413	9.808	11.651	13.034
19	5.145	7.020	8.434	9.838	11.696	13.070
20	5.150	7.034	8.458	9.875	11.722	13.120
22	5.160	7.064	8.500	9.921	11.788	13.191
24	5.173	7.085	8.529	9.961	11.853	13.297
26	5.184	7.108	8.562	10.000	11.894	13.340
28	5.196	7.125	8.585	10.035	11.947	13.385
30	5.204	7.136	8.610	10.065	11.981	13.408
35	5.224	7.171	8.653	10.133	12.064	13.519
40	5.237	7.187	8.678	10.165	12.114	13.575
45	5.245	7.205	8.698	10.191	12.140	13.604
50	5.243	7.223	8.721	10.210	12.172	13.649
60	5.260	7.235	8.740	10.242	12.210	13.694
70	5.279	7.246	8.757	10.262	12.244	13.715
80	5.291	7.262	8.773	10.278	12.255	13.765
90	5.309	7.261	8.785	10.297	12.288	13.765
100	5.312	7.267	8.789	10.302	12.290	13.806
125		7.277	8.802	10.323	12.323	13.825
150		7.269	8.797	10.352	12.336	13.840
175		7.304	8.831	10.350	12.341	13.854
200		7.334	8.804	10.332	12.356	13.863
250			8.829	10.337	12.356	13.882
300			8.859	10.370	12.370	13.889
350			8.838	10.368	12.395	13.908
400			8.914	10.371	12.379	13.921
500				10.410	12.391	13.907

Using tabled cutoffs is not particularly attractive in a computer implementation, so it is helpful to have some more compact, even if approximate, way of computing suitable cutoff values. Based on this reasoning, we suggest the following approach for getting cutoff values:

- For $n = 10, \ldots 15$, use the cutoffs from Table 1.
- For $n > 15$, use the approximation

$$h_{n,\alpha} = \begin{cases} -1.38 - 2.241\ln(\alpha) + \dfrac{1.61 + 0.691\ln(\alpha)}{\sqrt{(n-9)}} & \text{if } 0.001 \leq \alpha < 0.05, \\ 5 + 0.066\ln(n-9) & \text{if } \alpha = 0.05, \end{cases}$$

which reproduces Table 1 with a maximum absolute error of 0.08. This functional form is not based on any theoretical model but was selected empirically using a combination of graphs and regression fits.

Computational Details and Windows

All the computations necessary for the testing can be found from two tables—one of the running total of the data $W_n = \sum_1^n X_i$, and the other of the running sum of squared deviations from the running mean $V_{0,n}$. These have very simple updates,

$$W_{n+1} = W_n + X_{n+1}$$
$$V_{0,n+1} = V_{0,n} + n(X_{n+1} - W_n/n)^2/(n+1).$$

From these two tables, the quantities needed to perform the test are easy to compute,

$$\overline{X}_{i,k} = (W_k - W_i)/(k-i)$$
$$V_{i,k} = V_{0,k} - V_{0,i} - i(k-i)(\overline{X}_{0,i} - \overline{X}_{i,k})^2/k.$$

As these formulas have some (though limited) potential for loss of precision due to subtractive cancellation, the tables and the arithmetic should be in double precision.

The updates of the W_n and $V_{0,n}$ are fast, but searching for the k maximizing the split statistic involves computing $n-2$ test statistics, and even though the computations themselves are fast, for very large n or in settings where large numbers of process measurements are monitored, this could become a burden. A remedy that comes to mind is to restrict the searching to a window of the most recent observations. An effective way of reducing the computations without discarding information is a windowing approach of retaining only the M most recent values of W and V and restricting the search to these tabled values. When a new observation accrues, the W and V tables are (conceptually rather than literally) moved down one place, the oldest entry dropped and the newly updated entry added. The k search is then confined to the entries in the W and V tables and so does not grow with n, but remains bounded by $M-2$.

Note that this approach is fully statistically efficient in that observations to the left of the window are not lost—they still feature fully in all summary numbers. All that is lost is the ability to declare a split point more than M time intervals in the past.

Performance

The performance of quality-control charts is commonly measured by their average run length (ARL). To explore this, we carried out simulations—each of 10,000 sequences of length up to 10,000. We used the $\alpha = 0.002$ cutoffs so as to get an in-control ARL of 500. There are two further settings that need to be specified—the number of initial familiarization observations gathered before the start of the testing and the instant at which the change in variance occurs.

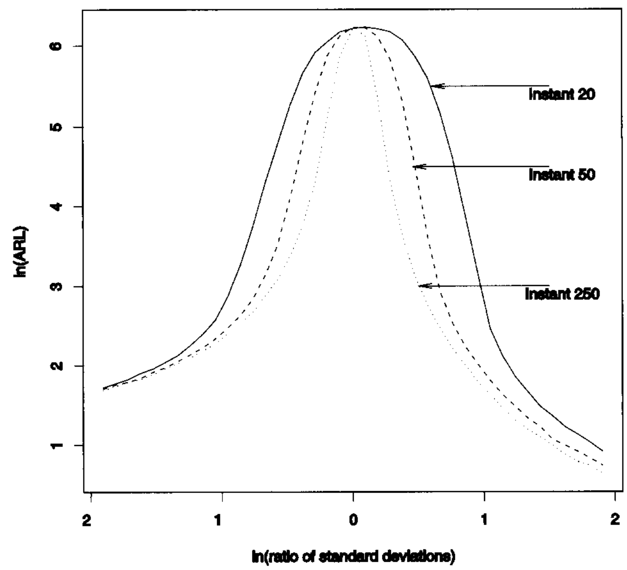

FIGURE 1. ARL Following Changes in Standard Deviation—Instants of Change 20-50-250; x-Axis $\ln(\sigma_2/\sigma_1)$, y-Axis ln(ARL).

We set the first of these at nine, so that formal testing began with the tenth observation of the sequence.

Several instants of change $\tau + 1$ were tested—observations number 20, 50, 80, 150, and 250, but we report only the results for 20, 50, and 250, the two intermediate values adding little insight. The in-control standard deviation was taken as 1, and the out-of-control values ranged from 0.5 to 2. Figure 1 shows the resulting ARLs. Both scales are logarithmic.

The figure shows that large shifts (whether upward or downward) are detected quickly, regardless of how soon after the start of the series they occur. The ability to detect downshifts effectively is a marked contrast with Shewhart R and S charts.

For relatively small shifts, the time of occurrence has a substantial impact. If the standard deviation increases from 1 to 1.6 starting at observation number 20, 50, 80, 150, or 250, the ARL to detection is 354, 90, 38, 26, or 23, respectively. The long ARL to detection of the shift at observation 20 might at first blush suggest that the change-point formulation is not very good. It more accurately reflects the limited precision of the estimate of the prechange variance when only a few observations accrue before the shift. Suppose, for example, that the standard deviation increases by 60% after 19 in-control observations. The

TABLE 2. In- and Out-of-Control ARL of Combination Charts

		Table of ARL(δ, n, λ) λ						
δ	$\tau + 1$	0.512	0.640	0.800	1.000	1.250	1.563	1.953
0.0	20	61.2	180.2	302.4	329.2	274.0	178.6	78.2
	50	22.4	75.9	278.2	325.5	202.4	55.4	13.2
	250	17.6	32.5	208.1	405.7	87.9	19.6	8.6
0.5	20	35.5	98.1	194.1	242.3	216.3	159.0	73.1
	50	19.7	37.5	85.8	129.0	99.5	37.3	11.7
	250	17.6	27.2	39.2	41.8	29.2	14.6	7.7
1.0	20	14.6	20.4	32.6	55.9	79.4	79.8	43.7
	50	12.2	13.7	15.2	16.0	16.0	12.8	7.9
	250	11.1	11.8	11.9	11.5	10.3	8.2	6.0
1.5	20	8.0	8.7	9.7	11.0	13.6	18.9	17.4
	50	6.6	6.7	6.7	6.9	6.7	6.3	5.3
	250	5.9	5.9	5.9	5.7	5.5	5.0	4.3
2.0	20	5.1	5.2	5.5	5.8	6.0	6.5	6.7
	50	4.0	4.1	4.1	4.1	4.1	4.0	3.8
	250	3.6	3.6	3.6	3.6	3.6	3.4	3.2

power of the F test for equality of variance of the first 19 observations against the subsequent observations using a 0.2% significance level increases to an asymptote of 57% as the length of the right segment tends to ∞. Adding observations after the change does not remove the indeterminacy of estimating σ_1 from such a modest sample size.

Control of Both Mean and Variance

There are settings in which this scale control might be used alone, but a more common use will likely be alongside the HQK procedure for a change in mean to provide a pair of charts monitoring data for shifts in mean, variance, or both. This pair of change-point methodologies parallels, for example, the Reynolds and Stoumbos (2001) scheme using EWMA charts. An out-of-control would then be signalled if either of the charts were to exceed its control limit. It is of interest to explore the performance of this combination.

We used simulation (each of 10,000 sequences) to evaluate the performance of this combined scheme when the in-control distribution of $N(0,1)$ changed to one of $N(\delta, \lambda^2)$. The HQK location change-point formulation and this scale proposal were each run using the $\alpha = 0.002$ cutoffs, which would give each of them an in-control ARL of 500.

Table 2 shows the resulting ARLs. As before, we introduced the shifts starting at observations number 20, 50, 80, 150, and 250, but show only the results for 20, 50, and 250. The first row of results shows how the combined scheme reacts to shifts in variance only and the middle ($\lambda = 1$) column shows the response to shift in mean only.

In a fixed-sample setting, the t test for the means and the GLR test for the variances are statistically independent. Thus, one might expect the combined charts to have an in-control ARL near 250. That the actual null ARLs are much higher than this shows that the repeated testing done in the change-point formulation nullifies this independence, creating a positive relationship between the two charts' run lengths.

Because one is likely to interpret a signal on the location chart as indicating a change in mean, and one on the scale chart as indicating a change in variance, it is of interest to know what proportion of the signals

TABLE 3. Percentage of Signal Coming from Location (M) Chart or Scale (V) Chart

		Table of Percentage of Signal(δ,M,V,λ) λ						
δ	mean & var	0.512	0.640	0.800	1.000	1.250	1.563	1.953
0.0	mean	0.82	5.26	31.87	59.87	68.78	56.28	50.00
	var	99.13	94.63	67.97	39.24	27.53	30.88	26.45
	mean & var	0.05	0.11	0.16	0.88	3.69	12.84	23.55
0.5	mean	14.73	36.37	64.21	79.85	82.76	68.87	56.21
	var	84.09	62.66	35.38	19.76	14.70	20.43	21.28
	mean & var	1.18	0.97	0.41	0.39	2.54	10.70	22.51
1.0	mean	66.22	84.63	93.24	97.00	95.43	85.86	68.92
	var	30.60	13.98	6.26	2.59	2.45	5.77	10.65
	mean & var	3.18	1.39	0.50	0.41	2.12	8.37	20.43
1.5	mean	93.77	96.96	98.03	97.91	96.18	90.11	75.49
	var	4.23	2.29	1.22	0.92	0.95	2.05	4.52
	mean & var	2.00	0.75	0.75	1.17	2.87	7.84	19.99
2.0	mean	98.42	98.64	98.13	96.76	94.85	89.05	77.09
	var	0.65	0.46	0.56	0.70	0.70	0.89	1.83
	mean & var	0.93	0.90	1.31	2.54	4.45	10.06	21.08

come from each. Table 3 covers the same ground as Table 2, and shows the percentage of signals coming from the location, from the scale, or from both charts in each setting. The table shows only the results for a change starting at observation 50, as the figures for other epochs are qualitatively similar.

The first row of Table 3 shows that increased variance, even unaccompanied by any shift in mean, is quite likely to trigger the location chart. Though not universally recognized, this is true of all standard location charts. The middle column shows a high specificity for the location chart; signals when the mean shifts but the variance does not are overwhelmingly likely to come just from the location chart with few false alarms from the scale chart. Moving left and right along each row shows that variance decreases tend to reduce the frequency of signals from the location chart and increase that from the scale chart.

Comparison with the CUSUM and EWMA

Comparing our approach with a standard alternative is complicated by the fact that there is no standard alternative, as the standard methods rely on the availability of values for the in-control parameters. We therefore chose a perhaps imperfect comparison—with the cumulative sum and the exponentially weighted moving average following 200 in-control learning samples, the values of which were used to calibrate the charts. We simulated 10,000 sequences, changing the standard deviation starting at the 201st reading, discarding any sequence in which the change-point test signaled before the actual change. The change-point chart was exactly as described, using an in-control ARL of 1,000. One benchmark was the EWMA recommended by Crowder and Hamilton (1992). This used as data the variances of rational groups of size 5, and was tuned for best performance at a 30% increase in standard deviation. Writing S_i^2 for the variance of the ith rational group of size 5, this uses the recursion

$$E_0 = 0; \quad E_i = \max(0, (1-\lambda)E_{i-1} + \lambda \ln(S_i^2))$$

with $\lambda = 0.16$. The chart signals if

$$E_i > K\sqrt{\frac{\lambda}{2-\lambda}\mathrm{Var}(\ln(S_i^2))} = 0.236K.$$

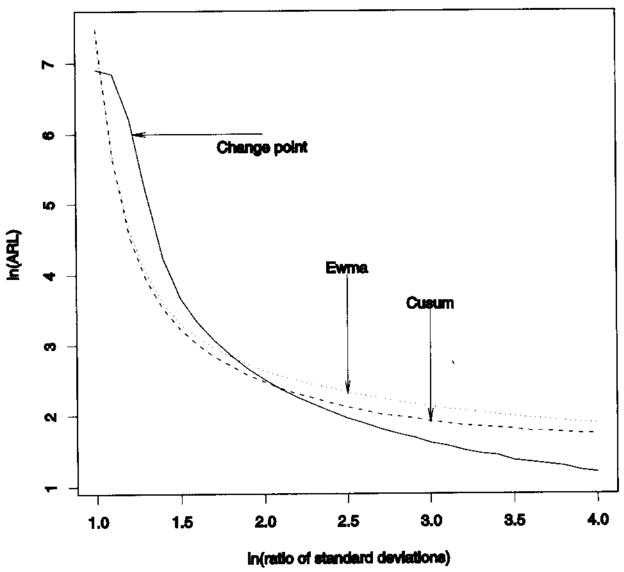

FIGURE 2. Comparison of Change Point with CUSUM and EWMA, x-Axis $\ln(\sigma_2/\sigma_1)$, y-Axis ln(ARL).

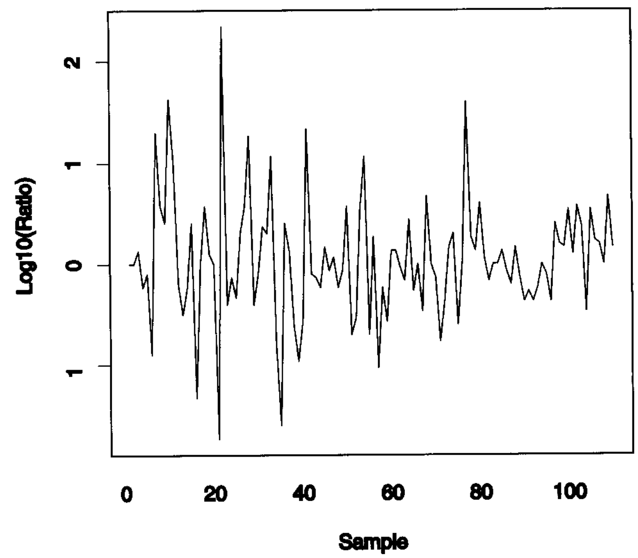

FIGURE 3. Log_{10} of Sampler to Supervisor Gold Content Ratio.

The choice $K = 1.45$ gives an in-control ARL of 200 rational groups, or 1,000 individual readings, matching that of the change point test.

The other benchmark was an upward scale CUSUM using the variances of rational groups of size 5. It was also tuned for optimal response to an upward shift of 30% in standard deviation—details are in Hawkins and Olwell (1998). The CUSUM recursion was

$$C_0 = 0; \quad C_i = \max(0, C_{i-1} + S_i^2 - 1.285),$$

with a signal of an upward shift if $C_i > 3.69$. We note that this comparison is somewhat unfair to the change point test in that the change point test is a two-sided monitoring scheme, but both the schemes with which it is being compared are one-sided. On the other side, Reynolds and Stoumbos (2004b) suggest that the use of rational groups of size greater than 1 in the CUSUM and EWMA is not the best choice for these methodologies either. Neither criticism, however, blunts the broad comparison of the approaches.

Figure 2 shows the ARLs of the three schemes for various post-shift standard deviations. Several points are noteworthy. First, the CUSUM and EWMA have in-control ARLs substantially above the nominal 1,000—that of the CUSUM is 1,750 and the EWMA is 1,600. This shows that calibrating the CUSUM and EWMA using variance estimates having 199 degrees of freedom is not enough to get rid of the substantial effect of estimating parameters. This observation for scale monitoring parallels the observations of Jones and Champ (2001) and Jones et al. (2004) for the location CUSUM and EWMA.

As regards the comparative performance, the figure shows that the CUSUM outperforms the EWMA throughout the range, but increasingly so for large shifts. This is probably due less to the CUSUM's general performance advantage over the EWMA than to the latter's accumulating on the log and not the original scale. The change-point chart lags the other two for standard deviation ratios less than 2, but outperforms them for larger increases in variability.

Example: Gold Mine Sampling Quality Control

Samplers in gold mines cut samples from the face at regular spacings, and submit them for chemical assay for their gold content. As a quality check, supervisors cut out fresh samples at some of the locations already sampled. This gives rise to pairs of samples and of gold content —one by the original sampler and one by the supervisor. The log of the ratio of the gold content of the sampler to that of the supervisor has an approximately normal distribution. The variance of the log ratio is extremely important because of its implication for the precision of the estimates that are used to decide which blocks of ore to extract and which to leave in place. Figure 3, one of a collec-

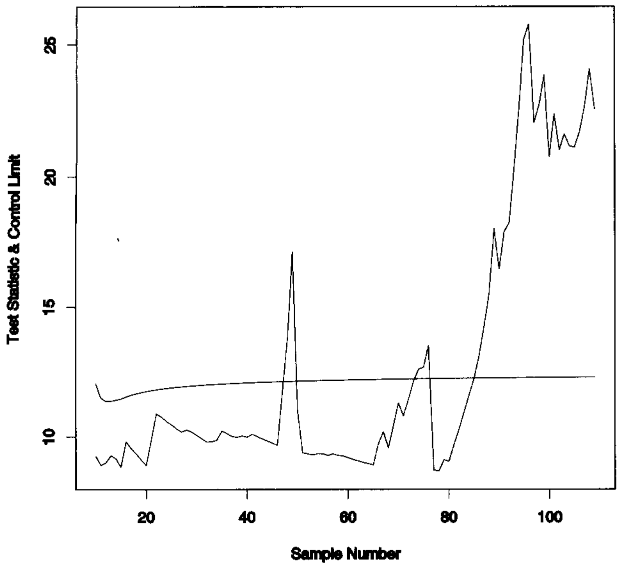

FIGURE 4. Change Point Statistic and Control Limit.

tion given by Rowland and Sichel (1961), shows a sequence of such ratios for a junior sampler. For this analysis, we will assume that the mean of the log ratio is zero throughout the sequence (as Figure 3 suggests), and our main interest will be to see whether there has been change in the variance. Variance reduction will imply quality improvement and knowledge about the sampler as to whether he has gained some learning over time, while increased variance will indicate a deterioration in quality. Did the sampler help improve quality? When did he start getting better? We will address these questions using our variance change-point formulation.

While the log ratio was described as approximately normal, it has two features that make it a less-than-perfect example. One is that, due to rounding of the gold assays, the distribution is grainy, with 10% of the log ratios being exactly zero. The other is that it has a regular sprinkling of outliers. To the extent that our methodology provides good results in the face of these complications, it can be supported for settings of more nearly normal data.

Another potential modeling issue has to do with a possible dependence. Because the entire data reflects the work of a single sampler and a supervisor, there are no cluster effects in the data. Even though serial correlation is possible through the sampler having good or bad days, the sequence did not show any such autocorrelation. Thus, we feel justified in proceeding with the change-point analysis.

Examining the plot in Figure 3, it looks as though the variability in the portion up to around observation 80 is higher than that for the remainder of the sequence. We will apply the change-point technique for change in variance as might be done in routine control in a mine. Figure 4 gives the values of $G_{\max,n}$ along with the control limits $h_{n,0.002}$, corresponding to an in-control ARL of 500 for exactly normal data.

$G_{\max,n}$ exceeds $h_{n,0.002}$ three times, with first exceedances at readings 48, 75, and 87. The details of these signals are

1. Signal at 48. Adding the 48th observation to the process takes $G_{\max,48}$ over the threshold $h_{48,0.002}$ and estimates the change point as having occurred after observation 41. The estimates of the standard deviations of the two segments defined are $\hat{\sigma}_{1,\ldots,41} = 0.839$ and $\hat{\sigma}_{42,\ldots,48} = 0.149$, a substantial decrease. With this clue, we identify the short but unusually constant segment around observation 40 in Figure 3.
2. After spiking at reading 49, $G_{\max,n}$ goes back below $h_{n,0.002}$ following some readings of size more typical of the earlier segment. A second signal then occurs at reading 75. This signal also flags observation 41 as the last of the original-variance segment, but estimates the subsequent standard deviation as $\hat{\sigma}_{40,\ldots,75} = 0.5231$, also lower than the starting variability but not as dramatically so. A visual difference in the variability of these two subsequences is less clear than was that leading to the first signal.
3. Once again, a large discrepancy brings $G_{\max,n}$ below the control limit, where it remains until reading 87, when it punches through the control limit decisively and remains above it for the rest of the sequence (and for several dozen further readings shown in Rowland and Sichel, but not reported here). With this signal, $\hat{\tau}$ moves up to 80, capturing the visual change point in the plot. The before and after standard deviations are $\hat{\sigma}_{1,\ldots,80} = 0.696$; $\hat{\sigma}_{81,\ldots,87} = 0.120$. As the process history grows beyond this point, the estimate of the after standard deviation increases, but to no more than some 0.43. This represents a substantial improvement in the measurement variance than the sampler displayed initially.

This reduction in measurement variance is of major benefit to the mine because it leads to a much-improved precision in the localization of higher grade areas in the ore body.

Reviewing the whole analysis, in hindsight, we can conclude that the two early unconfirmed signals probably did presage some improvement in the variability, but this was initially too small to stay well quantified. It was at around reading 80 that the sampler's training and further experience led to a permanent reduction in variability.

Conclusion

The importance of monitoring process variability has been recognized since Shewhart's earliest writings but has often played second fiddle to monitoring of location. A wider awareness of the importance of variance has come with the quality improvement movement and the increased awareness that variability is a major source of quality problems. In some settings, only the variability needs to be monitored; in others, we require tools for control of both mean and variability.

Conventional charting methods such as the Shewhart, CUSUM, and EWMA charts are calibrated assuming that the in-control process parameters are known exactly. To the extent that this is untrue, the in-control run behavior of the resulting chart will differ from what the user expected. The CUSUM and EWMA have a further failing that realizing their best performance requires tuning to the size of the shift, requiring even more prior knowledge.

There has always been a tension between the objective of starting process monitoring as soon as possible and that of conducting a good process performance study to have confidence in the baseline parameters. While we present the change-point formulation as a method that can start Phase II monitoring after gathering only three Phase I readings, we do not see it being used in any such extreme fashion. In some settings, there may be substantial historical precedent supporting the expectation that the process readings will follow a normal distribution. In this case, the change-point charting could indeed be started with the minimal accumulation of baseline information. At the other end of the spectrum are settings in which there is no advance reason to expect normally distributed process readings and there is no particular obstacle to gathering a reasonably sized Phase I data set. In this setting, we see the change point as a valuable adjunct to the other Phase I analysis tools, leading to the possibility of an earlier transition out of this precursor mode into ongoing production monitoring.

In this paper, we developed a change-point approach for variance shifts. This can be used alone or in conjunction with already-extant change-point control for the process mean to provide a powerful SPC methodology.

Acknowledgments

This work was partially supported by the National Science Foundation under grant DMS 0306304. The authors are grateful to the editor and referees for helpful suggestions.

References

Bartlett, M. S. (1937). "Property of Sufficiency and Statistical Tests". *Proceedings of the Royal Society of London, Series A* 160, pp. 268–282.

Bartlett, M. S. (1955). "Approximate Confidence Intervals: III. A bias correction". *Biometrika* 42, pp. 201–204.

Bartlett, M. S. and Kendall, D. G. (1946). "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation". *Supplement to the Journal of the Royal Society of London* 8, pp. 128–138.

Bhattacharya, P. K. (1994). "Some Aspect of Change Point Analysis". *Change Point Problems, IMS—Lecture Notes Monograph Series* 23, pp. 28–56.

Crowder, S. V. and Hamilton, M. (1992). "An EWMA for Monitoring a Process Standard Deviation". *Journal of Quality Technology* 24, pp. 12–21.

Gombay, E. (2000). "Sequential Change Point Detection with Likelihood Ratios". *Statistics and Probability Letters* 49, pp. 195–204.

Hawkins, D. M. (1977). "Testing a Sequence of Observations for a Shift in Location". *Journal of American Statistical Association* 72, pp. 180–186.

Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement.* Springer Verlag, New York.

Hawkins, D. M.; Qiu, P.; and Kang, C. W. (2003). "The Change Point Model for Statistical Process Control". *Journal of Quality Technology* 35, pp. 355–365.

Jones, L. A. and Champ, C. W. (2001). "The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters". *Technometrics* 43, pp. 156–167.

Jones, L. A.; Champ, C. W.; and Rigdon, S. E., (2004). "The Run Length Distribution of the CUSUM with Estimated Parameters". *Journal of Quality Technology* 36, pp. 95–108.

Lai, T. L. (2001). "Sequential Analysis: Some Classical Problems and New Challenges". *Statistica Sinica* 11, pp. 303–350.

Margavio, T. M.; Conerly, M. D.; Woodall, W. H.; and Drake, L. G. (1995) "Alarm Rates for Quality Control Charts". *Statistics and Probability Letters* 24, pp. 219–224.

POLLAK, M. and SIEGMUND, D. (1991). "Sequential Detection of a Change in Normal Mean When the Initial Value Is Unknown". *Annals of Statistics* 19, pp. 394–416.

QUESENBERRY, C. P. (1991). "SPC Q Charts for Start-Up Processes and Short or Long Runs". *Journal of Quality Technology* 23, pp. 213–224.

REYNOLDS, M. R., JR. and STOUMBOS, Z. G. (2001). "Monitoring the Process Mean and Variance Using Individual Observations and Variable Sampling Intervals". *Journal of Quality Technology* 33, pp. 181–205.

REYNOLDS, M. R., JR. and STOUMBOS, Z. G. (2004a). "Control Charts and the Efficient Allocation of Resources". *Technometrics* 46, pp. 188–199.

REYNOLDS, M. R., JR. and Stoumbos, Z. G. (2004b). "Should Observations Be Grouped for Effective Process Monitoring?", to appear in *Journal of Quality Technology*.

RIGDON, S. E.; CRUTHIS, E. N.; and CHAMP, C. W. (1994). "Design Strategies for Individuals and Moving Range Control Charts". *Journal of Quality Technology* 26, pp. 274–287.

ROES, K. C. B.; DOES, R. J. M. M.; and SCHURINK, Y. (1993). "Shewhart-Type Charts for Individuals". *Journal of Quality Technology* 25, pp. 188–198.

ROWLAND, R. ST. J. and SICHEL, H. S. (1961). "Statistical Quality Control of Routine Underground Sampling". *Journal of the South African Institute for Mining and Metallurgy* 60, pp. 251–284.

SEN, A. and SRIVASTAVA, M. S. (1975). "On Test for Detecting Change in the Mean When Variance Is Unknown". *Annals of the Institute of Statistical Mathematics* 27, pp. 479–486.

SIEGMUND, D. and VENKATRAMAN, E. S. (1994). "Using the Generalized Likelihood Ratio Statistics for Sequential Detection of a Change point". *Annals of Statistics* 23, pp. 255–271.

SULLIVAN, J. H. and WOODALL, W. H. (1996). "A Control Chart for Preliminary Analysis of Individual Observations". *Journal of Quality Technology* 28, pp 265–278.

WORSLEY, K. J. (1982). "An Improved Bonferroni Inequality and Applications". *Biometrika* 69, pp. 297–302.

WORSLEY, K. J. (1979). "On the Likelihood Ratio Test for a Shift in Location of Normal Populations". *Journal of American Statistical Association* 74, pp. 365–367.