Effect of Estimation of the Process Parameters on the Control Limits of the Univariate Control Charts for Process Dispersion

Petros Maravelakis and John Panaretos and Stelios Psarakis

2002

COMMUN. STATIST.—SIMULA., 31(3), 443–461 (2002)

EFFECT OF ESTIMATION OF THE PROCESS PARAMETERS ON THE CONTROL LIMITS OF THE UNIVARIATE CONTROL CHARTS FOR PROCESS DISPERSION

P. E. Maravelakis, J. Panaretos,* and S. Psarakis

Department of Statistics,
Athens University of Economics and Business,
76 Patission St., 10434, Athens, Greece

1. INTRODUCTION

Control charts are used for controlling and monitoring variables in any product or process. They have found considerable applications in industry for improving the quality of the products. The most known of them are the Shewhart type control charts for monitoring process mean and dispersion.

Quesenberry (1) examined the effect of estimation of the process mean and standard deviation on the control limits of the Shewhart chart for both rational subgroups and individual observations. Chen (2) extended this work by using three different estimators of the standard deviation in the $\bar{X}$ chart case. Nedumaran and Pigniatiello (3) investigated the estimation effect on the T^2 control charts. Woodall and Montgomery (4) emphasized the need for much more research in this area since it is proved that more data than usually recommended is needed for the control charts to behave as expected from theory. In the same paper, Woodall and Montgomery state that much work has been done concerning the control of the process mean but not that much for the process dispersion. In an earlier paper Lowry et al. (5) examined the effect of run rules on the performance of Shewhart control charts for detecting shifts in process standard deviation. Recently, Klein (6) proposed modified Shewhart S-charts for keeping stable the process variability. Chen (7) deals with the run length properties of the R, s and s^2 control charts in the case that σ is estimated. In this paper, we examine the effect of estimation of the process parameters on the control limits of charts for process dispersion by extending the results of Chen (7) for both rational subgroups and individual observations.

The paper is organized as follows. In Section 2, we present the classical S chart with three sigma limits and extensive numerical calculations of the effect of estimating the process standard deviation on the values of average run length (ARL) and standard deviation of the run length ($SDRL$). Section 3 outlines the S chart using probability limits and results of estimating the process standard deviation on the ARL and $SDRL$ values again. The X chart for individual observations is presented on Section 4 and its use for process dispersion when we have estimated limits. Finally, in Section 5 conclusions are listed.

2. THE S (THREE SIGMA) CONTROL CHART

Assume that X_{ij}, $i=1,\ldots,m$ and $j=1,\ldots,n$ are observations from a stable $N(\mu,\sigma^2)$ process comprising m samples of size n each. In this process we want to keep its variability in control. In order to develop control limits we need to know the value of the true standard deviation σ. If this value is known the control limits are

$$UCL = \left(c_4 + 3\sqrt{1 - c_4^2}\right)\sigma \tag{2.1}$$

$$LCL = \left(c_4 - 3\sqrt{1 - c_4^2}\right)\sigma \tag{2.2}$$

Usually, we do not know the value of σ and therefore we have to estimate it from past data. The estimate used is

$$\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$$

where m is the number of past samples used, $S_i^2 = (1/(n-1)) \times \sum_{j=1}^{n} \left(X_{ij} - \bar{X}_i\right)^2$ is the unbiased estimator of σ^2 and n is the sample size. However, we know that S is not an unbiased estimator of σ. It has been proved (see e.g., Ryan (8)) that an unbiased estimate of σ is $\bar{S}/c_4$ where $c_4 = (2/(n-1))^{1/2}\Gamma(n/2)/\Gamma((n-1)/2)$ and that the standard deviation of S equals $\sigma\sqrt{1 - c_4^2}$. The upper and lower control limits of the chart known as the S chart are:

$$\widehat{UCL} = \left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)\bar{S} \tag{2.3}$$

$$\widehat{LCL} = \left(1 - \frac{3}{c_4}\sqrt{1 - c_4^2}\right)\bar{S} \tag{2.4}$$

Approaches making use of these limits are known as the three sigma approaches based on the normal approximation proposed by Shewhart in the early thirties. However, it is easy to prove that this approximation is not satisfactory since as is known

$$\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2 \tag{2.5}$$

Although this approximation is not accurate, it is usually used as a first check (see e.g., Ryan (8), Klein (6), Lowry et al. (5)).

Let A_i denote the event that the ith sample standard deviation S_i exceeds UCL or is exceeded by LCL. Then, since S_i and S_j are independent for $i \neq j$, the sequence of trials A_i and A_j are independent meaning that they constitute a sequence of Bernoulli trails. The mean and standard deviation of the run length distribution, ARL and $SDRL$ respectively, of this process is that of a geometric distribution given by the following formulas

$$ARL = \frac{1}{1-\beta} \tag{2.6}$$

$$SDRL = \frac{\sqrt{\beta}}{1-\beta} \tag{2.7}$$

where $\beta = 1 - \Pr(A_i) = \Pr(LCL \leq S_i \leq UCL)$.

Assume now that we are in the case when the true value of the standard deviation is not known, which is the most usual case. Let B_i denote the event that the ith sample standard deviation S_i exceeds $\widehat{UCL}$ or is exceeded by $\widehat{LCL}$. The formulas (2.6) and (2.7) for ARL and $SDRL$ are not valid any more because the events B_i and B_j are not independent for $i \neq j$. We can prove that $E(\widehat{UCL}) = UCL$ and $\text{Var}(\widehat{UCL}) = (1 + (3/c_4)\sqrt{(1-c_4^2)})^2 \sigma^2((1-c_4^2)/m)$ and using these relations we prove after some calculations that

$$\text{Cov}(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \text{Var}(\widehat{UCL}) = \left(1 + \frac{3}{c_4}\sqrt{1-c_4^2}\right)^2 \sigma^2 \frac{(1-c_4^2)}{m}$$

and

$$\text{Var}(S_i - \widehat{UCL}) = \left[1 + \frac{\left(1 + \frac{3}{c_4}\sqrt{1-c_4^2}\right)^2}{m}\right] \sigma^2(1-c_4^2)$$

Therefore, the correlation between the random variables $S_i - \widehat{UCL}$ and $S_j - \widehat{LCL}$ is

$$\text{Corr}(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \frac{\text{Var}(\widehat{UCL})}{\text{Var}(S_i - \widehat{UCL})} = \frac{\left(1 + \frac{3}{c_4}\sqrt{1-c_4^2}\right)^2}{m + \left(1 + \frac{3}{c_4}\sqrt{1-c_4^2}\right)^2}$$

It is obvious that the correlation is a function of m and n only. In Table 1 we present values of the correlation for combinations of m and n. From the table we see that as the sample size and the number of samples increase the correlation decreases. For small or moderate sample size ($n \leq 20$) we need 200 samples for the correlation to be negligible. However, for larger sample size we can afford $m = 50$.

In order to examine the values of the first two moments of the run length distribution, we performed a simulation study based on various numbers of samples and various sample sizes. In particular the number of

Table 1. Correlation for Several Values of m and n

	n			
m	5	10	20	50
5	0.46581	0.37055	0.30735	0.25370
10	0.30362	0.22741	0.18158	0.14528
20	0.17898	0.12829	0.09986	0.07833
30	0.12689	0.08935	0.06886	0.05362
50	0.08020	0.05560	0.04249	0.03288
100	0.04178	0.02859	0.02171	0.01671
200	0.02133	0.01450	0.01097	0.00843
500	0.00864	0.00585	0.00442	0.00339
1000	0.00434	0.00293	0.00221	0.00170

samples and samples sizes considered were $m = 5$, 10, 20, 30, 50, 100, 200, 500, 1000 and $n = 5$, 10, 20, 50. For every combination of m and n we simulated m samples of size n from a $N(\mu, \sigma_0^2)$ distribution and computed $\widehat{UCL}$ and $\widehat{LCL}$. Then, we simulated samples from a $N(\mu, \sigma_1^2)$ distribution until we obtained a value above $\widehat{UCL}$ or below $\widehat{LCL}$. The number of samples simulated up to the one that lead to a value outside the control limits constitutes one observation of the run length distribution. This procedure was repeated 10,000 times in order to get an estimate of the values of ARL and $SDRL$. The results are presented in Tables 2–5.

From Tables 2 through 5 certain conclusions are drawn. We see that we have results for both upward and downward shifts when $n > 5$ but only for upward when $n = 5$. This happens because for $n \leq 5$ the lower control limit is set to zero. Therefore, it can never be crossed by a simulation study, or in reality. For upward shifts as m increases the ARL and $SDRL$ values decrease and approach their theoretical values. For downward shifts as m increases the same thing happens for $n = 50$. For $n = 10$, 20 the ARL and $SDRL$ values do not follow a specific trend. In the in-control state we also do not have a clear pattern for either ARL or $SDRL$ values. What we can say in every case is that ARL and $SDRL$ values behave in the same way.

As m increases the ARL is getting closer to the theoretical value faster than the $SDRL$. Moreover, as n increases the theoretical values, in the in-control state, approach the ones from a normal distribution which are $ARL = 370.4$ and $SDRL = 369.9$. The same of course happens for the out of control states.

If we use this type of chart for identifying shifts in process dispersion we have to use samples of size n at least 20, for minimizing the effect of estimating S. If n is less than this value the practitioner will face an

Table 2. *ARL* and *SDRL* Values for the *S* (Three Sigma) Chart When $n=5$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	$4\cdot10^5$	$1\cdot10^5$	2223.5	$7\cdot10^4$	594.02	$2\cdot10^4$	105.38	1353.30	37.41	143.16
10	2200.10	$3\cdot10^4$	310.65	2288.70	86.99	330.39	39.29	104.72	21.06	45.34
20	551.16	1699.70	139.42	297.08	54.88	95.14	27.47	41.75	16.48	21.70
30	415.06	840.14	112.55	182.48	48.74	74.75	25.79	34.47	15.52	18.54
50	346.68	545.72	101.62	134.96	43.32	55.43	23.36	27.19	14.73	16.19
100	298.59	407.05	91.09	106.99	40.75	44.42	22.35	23.56	14.20	14.55
200	276.08	318.09	85.28	93.97	39.28	41.12	21.55	22.18	13.93	13.92
500	262.29	275.14	85.20	88.07	38.55	39.24	21.75	21.79	13.94	13.52
1000	253.76	258.84	84.37	87.67	37.32	37.06	20.97	20.66	13.59	13.14
∞	249.31	248.81	82.44	81.94	37.72	37.21	21.22	20.71	13.69	13.18

Table 3. *ARL* and *SDRL* Values for the *S* (Three Sigma) Chart When $n=10$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	606.61	1064.81	236.14	634.06	78.31	263.87	29.43	112.67	14.39	41.18
10	538.65	919.10	145.57	329.89	45.83	99.37	19.21	31.64	10.17	13.87
20	461.44	725.80	106.92	175.82	33.86	48.22	15.85	19.75	9.04	10.34
30	430.50	626.79	95.59	137.72	32.34	40.07	15.02	17.08	8.59	9.14
50	389.91	510.09	88.05	106.54	30.29	33.98	14.38	15.65	8.37	8.58
100	359.35	411.69	80.89	88.11	28.79	30.81	13.85	14.24	8.26	8.16
200	344.38	367.08	78.19	82.25	28.46	28.96	13.43	13.31	7.97	7.58
500	334.53	340.97	76.10	76.60	27.45	27.27	13.52	13.14	8.06	7.65
1000	334.56	337.96	75.88	75.93	27.31	27.01	13.50	13.04	7.98	7.48
∞	331.17	330.67	75.66	75.16	27.52	27.01	13.47	12.96	8.00	7.48

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	24.01	37.43	306.28	520.37	1019.6	1274.5	1136.2	1433.6
10	21.04	25.88	254.28	377.34	1071.7	1253.2	1316.9	1514.7
20	19.77	22.62	230.60	275.74	1079.4	1207.5	1472.6	1603.4
30	19.15	20.62	223.33	249.71	1056.5	1155.1	1569.2	1656.9
50	18.47	19.15	218.20	229.50	1047.3	1106.2	1644.7	1686.9
100	18.21	18.10	210.57	215.40	1037.5	1061.3	1696.2	1729.9
200	17.95	17.93	205.32	205.49	1023.1	1027.8	1744.5	1746.7
500	18.20	17.79	205.59	203.14	1009.1	1022.0	1773.9	1785.0
1000	17.53	17.28	206.96	204.95	1006.7	1007.9	1768.3	1773.9
∞	17.90	17.39	206.06	205.56	1011.7	1011.2	1777.2	1776.7

increased number of false alarms. The effect of estimation is also severe for $m \leq 20$, especially in the in-control state and for small shifts. For values $30 \leq m \leq 50$ the effect is moderate and for values of 100 or larger the effect is small enough. A last point we have to make is that when we have small downward shifts for $n \leq 20$ the *ARL* and *SDRL* values are larger than the corresponding in-control values. This result is also confirmed by Klein (6). Consequently, in such cases special care must be given and it is better to use control charts for small shifts like CUSUM and EWMA.

In Figure 1, we present the empirical run length distribution functions (ERL) for $n=5$, 10, 20, 50. In each figure we plot six different lines

Table 4. *ARL* and *SDRL* Values for the *S* (Three Sigma) Chart When $n = 20$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	332.72	444.02	121.96	244.63	32.29	78.86	11.46	27.03	5.54	10.33
10	362.96	457.01	92.99	166.56	23.32	45.79	8.71	11.78	4.62	5.39
20	371.24	439.25	75.00	115.39	19.63	25.36	7.87	8.77	4.32	4.20
30	372.32	430.13	68.53	86.90	18.23	21.84	7.67	8.17	4.29	4.12
50	362.66	403.51	63.80	76.74	17.52	18.73	7.49	7.60	4.24	3.97
100	364.01	393.80	60.17	65.57	17.01	17.34	7.36	7.15	4.11	3.58
200	359.00	374.30	59.56	60.31	16.61	16.45	7.15	6.82	4.11	3.59
500	355.18	358.14	59.11	59.61	16.36	16.15	7.13	6.70	4.09	3.56
1000	353.23	353.28	57.59	57.23	16.26	15.79	7.15	6.66	4.08	3.59
∞	356.50	356.00	57.37	56.87	16.39	15.88	7.15	6.63	4.07	3.53

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	1.32	0.75	11.92	18.83	111.01	190.51	383.43	457.13
10	1.28	0.64	10.03	12.23	90.20	127.11	423.04	463.05
20	1.26	0.60	9.21	9.96	80.28	94.68	442.70	473.45
30	1.26	0.58	8.97	9.20	78.03	88.22	444.96	471.63
50	1.24	0.54	8.90	8.59	75.70	80.02	451.20	469.84
100	1.25	0.57	8.68	8.20	73.42	75.19	450.90	455.92
200	1.23	0.54	8.70	8.27	73.69	74.61	447.50	446.43
500	1.24	0.55	8.55	8.05	73.62	73.39	441.19	441.96
1000	1.24	0.56	8.54	8.14	72.08	71.77	445.81	448.29
∞	1.24	0.54	8.56	8.04	72.91	72.41	449.79	449.29

representing the ERL functions for $m = 5, 20, 50, 100, 1000$ and the theoretical run length distribution (inf). It is obvious that as m increases the ERL approaches the theoretical run length distribution. Moreover, as n increases the ERLs for the m values approach the theoretical run length distribution faster.

3. THE *S* (PROBABILITY LIMITS) CONTROL CHART

A modification of the control limits (2.1), (2.2) and (2.3), (2.4) based on property (2.5) uses probability limits in place of the three sigma limits

Table 5. *ARL* and *SDRL* Values for the *S* (Three Sigma) Chart When $n=50$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	263.03	325.32	59.79	125.84	9.49	18.56	3.23	3.96	1.86	1.57
10	304.52	359.74	44.11	76.18	7.69	9.98	2.89	2.87	1.73	1.23
20	328.25	365.28	36.59	49.56	6.91	7.54	2.83	2.52	1.69	1.15
30	340.23	369.51	33.55	39.88	6.65	6.77	2.76	2.37	1.68	1.11
50	345.02	369.81	32.36	35.89	6.64	6.61	2.72	2.24	1.67	1.09
100	355.17	366.97	30.64	31.98	6.37	6.11	2.70	2.20	1.67	1.08
200	357.85	364.35	30.75	30.97	6.39	6.06	2.67	2.09	1.67	1.06
500	362.32	358.59	30.32	30.28	6.38	5.87	2.65	2.10	1.67	1.06
1000	356.30	352.76	30.62	29.97	6.29	5.89	2.67	2.08	1.67	1.05
∞	365.96	365.46	30.23	29.72	6.35	5.83	2.67	2.11	1.66	1.04

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	1	0	1.25	0.66	8.68	13.45	124.56	199.43
10	1	0	1.23	0.56	7.20	8.36	110.20	171.69
20	1	0	1.21	0.51	6.80	7.10	97.84	128.55
30	1	0	1.20	0.50	6.55	6.53	93.47	110.48
50	1	0	1.20	0.48	6.51	6.38	89.64	98.31
100	1	0	1.20	0.48	6.44	6.04	85.98	91.30
200	1	0	1.19	0.47	6.37	5.91	85.92	88.10
500	1	0	1.18	0.47	6.34	5.92	85.47	85.74
1000	1	0	1.18	0.47	6.26	5.82	85.82	85.86
∞	1	0	1.19	0.48	6.28	5.76	84.25	83.75

(see e.g., Ryan (8)). If the value of the standard deviation σ is known the control limits (in place of (2.1) and (2.2)) are:

$$UCL = \sigma\sqrt{\frac{\chi^2_{0.999}}{n-1}}$$

$$LCL = \sigma\sqrt{\frac{\chi^2_{0.001}}{n-1}}$$

In these limits, if the process variability operates in control, the probability that the standard deviation of future subgroups will fall between them

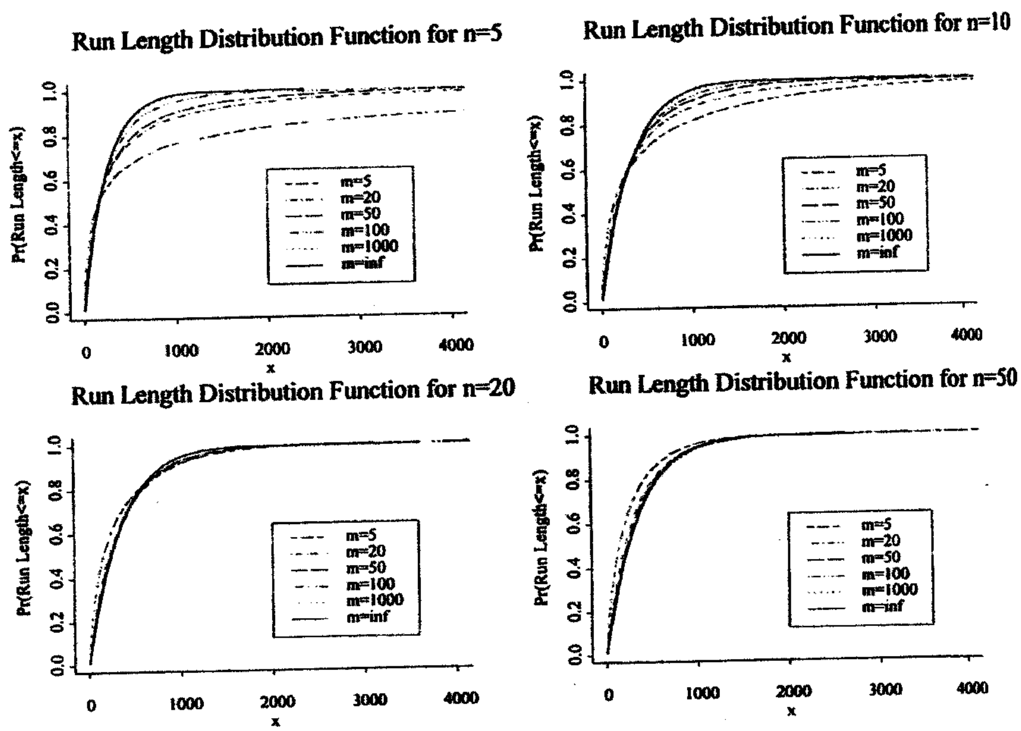

Figure 1. Empirical run length distribution functions for the 3 sigma chart.

is 0.998, which is approximately equal to the 0.9973, the probability assumed when using the 3 sigma ones. If the true standard deviation is not known we use its unbiased estimate $\bar{S}/c_4$. The limits then become (in place of (2.3) and (2.4)):

$$\widehat{UCL} = \frac{\bar{S}}{c_4}\sqrt{\frac{\chi^2_{0.999}}{n-1}}$$

$$\widehat{LCL} = \frac{\bar{S}}{c_4}\sqrt{\frac{\chi^2_{0.001}}{n-1}}.$$

It is obvious that these limits are based on property (2.5). In the same way of thinking as in the case of three sigma limits we can prove that $\mathrm{Var}(\widehat{UCL}) = [\sigma^2(1-c_4^2)\chi^2_{0.999}]/[(n-1)c_4^2 m]$ and consequently

$$\mathrm{Cov}(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \mathrm{Var}(\widehat{UCL}) = \frac{\sigma^2(1-c_4^2)\chi^2_{0.999}}{(n-1)c_4^2 m}$$

Moreover,

$$\mathrm{Var}(S_i - \widehat{UCL}) = \sigma^2(1 - c_4^2)\left[1 + \frac{\chi^2_{0.999}}{(n-1)c_4^2 m}\right]$$

and finally

$$\mathrm{Corr}(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \frac{\mathrm{Var}(\widehat{UCL})}{\mathrm{Var}(S_i - \widehat{UCL})} = \frac{\chi^2_{0.999}}{\chi^2_{0.999} + (n-1)c_4^2 m}.$$

As in the case of three sigma limits this correlation depends only on m and n. In Table 6 we calculated the correlation for various combinations of m and n. From this table we conclude again that as the sample size and the number of samples increase the correlation decreases. The recommendation for sample sizes and number of samples is the same as in the previous section.

We computed the ARL and $SDRL$ values for several values of m and n via simulation along the same lines as in the three sigma limits. The number of samples and samples sizes considered were $m = 5$, 10, 20, 30, 50, 100, 200, 500, 1000 and $n = 5$, 10, 20, 50. The results are presented in Tables 7–10. From Tables 7 through 10 we deduce the following points. For upward shifts as m increases the ARL and $SDRL$ values generally decrease and approach their theoretical values. For downward shifts as m increases the same thing happens for $n = 20$, 50. For $n = 5$, 10 the ARL and $SDRL$ values do not follow a specific pattern. In the in-control state the ARL and $SDRL$ values increase until they get close to their theoretical values, which is in

Table 6. Correlation for Several Values of m and n

	n			
m	5	10	20	50
5	0.51095	0.39568	0.32137	0.26032
10	0.34314	0.24663	0.19144	0.14964
20	0.20710	0.14066	0.10585	0.08087
30	0.14831	0.09839	0.07315	0.05541
50	0.09460	0.06145	0.04521	0.03400
100	0.04965	0.03170	0.02313	0.01729
200	0.02545	0.01611	0.01170	0.00872
500	0.01034	0.00650	0.00471	0.00351
1000	0.00520	0.00326	0.00236	0.00176

Table 7. *ARL* and *SDRL* Values for the *S* (Probability Limits) Chart When $n=5$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	359.97	463.12	267.35	405.54	173.40	312.06	111.11	231.57	71.17	173.00
10	401.46	491.51	268.52	395.19	154.68	263.77	83.88	161.01	47.93	102.77
20	441.09	495.15	254.39	350.22	127.40	199.04	64.92	106.92	36.69	58.40
30	462.04	509.78	247.68	320.05	115.34	164.84	58.02	84.90	33.35	49.45
50	472.24	504.56	239.29	295.97	108.19	137.80	52.48	65.23	30.29	35.47
100	489.90	512.64	229.28	262.50	99.08	115.37	49.79	54.68	28.81	31.03
200	498.35	505.20	221.61	240.21	94.66	102.45	48.20	50.97	27.67	29.10
500	500.93	505.59	216.74	223.58	93.45	95.24	46.06	46.00	28.00	27.60
1000	497.73	503.09	213.01	217.36	92.29	94.50	47.12	47.08	27.31	26.70
∞	500.02	499.52	214.74	214.24	91.78	91.28	46.51	46.01	27.33	26.82

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	59.28	92.90	207.50	279.98	367.15	426.84	423.91	494.16
10	51.33	60.99	188.80	229.67	383.62	419.59	478.62	506.71
20	49.25	53.77	178.80	195.31	381.22	406.88	535.06	558.84
30	47.36	50.56	174.26	182.63	378.10	395.01	551.71	561.82
50	47.06	47.90	172.37	175.45	374.69	387.19	572.90	579.90
100	46.45	46.53	170.21	172.41	369.25	373.72	588.21	585.37
200	44.99	44.60	169.92	169.76	369.89	371.74	595.42	594.84
500	45.64	45.22	168.09	168.67	364.29	363.97	604.03	601.67
1000	45.64	44.65	165.86	166.04	364.05	369.30	598.00	601.84
∞	45.09	44.59	167.40	166.90	366.87	366.37	597.91	597.41

accordance with the results of Chen (7). As an overall conclusion we can say that the *ARL* and *SDRL* values behave in the same way except that as *m* increases the *ARL* is getting closer to the theoretical value faster than the *SDRL*.

When we are in-control we need at least $m=200$, otherwise the practitioner will face many false alarms whereas the value of n is not equally important. In the out-of-control situations the value of n is important for minimizing the effect of estimating *S*. Specifically when $\sigma_1^2/\sigma_0^2 = 1.2$ the ARL values for $n=5$, 10, 20, 50 are 239.29, 178.40, 117.98, and 50.77, respectively. Therefore, we observe a dramatic reduction as n becomes larger. A similar situation occurs for downward shifts. Consequently,

Table 8. *ARL* and *SDRL* Values for the *S* (Probability Limits) Chart When $n = 10$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	341.44	422.99	217.14	329.46	110.54	218.04	52.43	130.38	25.60	62.05
10	391.03	456.05	208.08	307.21	86.61	155.83	36.61	67.22	17.72	28.51
20	428.95	469.37	194.55	257.65	70.14	106.07	28.50	39.48	14.96	18.23
30	448.41	480.41	187.90	234.75	65.03	88.79	27.33	33.58	14.20	16.15
50	464.28	481.37	178.40	209.85	60.27	72.62	25.81	28.64	13.61	14.78
100	479.05	488.20	169.77	184.28	56.35	61.10	24.52	25.62	13.16	13.69
200	484.86	493.03	166.70	176.09	54.73	56.70	24.26	24.50	12.81	12.66
500	490.54	489.97	161.32	164.74	52.91	53.41	24.02	24.08	13.11	12.82
1000	492.16	480.65	161.60	161.87	53.81	53.11	23.60	23.23	12.70	12.38
∞	500.05	499.55	161.99	161.48	53.44	52.94	23.46	22.95	12.74	12.23

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	5.34	7.41	46.09	81.74	182.01	262.85	339.52	406.60
10	4.63	4.92	38.28	48.93	162.80	206.09	378.42	422.41
20	4.40	4.27	35.18	40.33	154.72	181.67	396.79	417.98
30	4.36	4.10	34.06	37.31	147.37	159.98	400.78	424.06
50	4.20	3.76	33.36	35.08	144.77	156.54	401.66	421.77
100	4.27	3.81	32.66	34.24	139.43	140.95	402.89	413.00
200	4.26	3.73	32.78	33.25	137.34	137.24	400.48	403.36
500	4.17	3.64	32.44	31.76	136.91	133.59	405.11	405.09
1000	4.21	3.63	31.95	31.10	133.69	132.26	398.98	400.52
∞	4.23	3.70	32.13	31.62	136.47	135.97	400.85	400.35

large values of n, larger than 20, are recommended. The effect of estimation is severe for $m \leq 20$, especially for small shifts. For values $30 \leq m \leq 50$ the effect is moderate and for values of 100 or larger the effect is small enough. When we have small downward shifts for $n = 5$, and $n = 10$ when $m \leq 10$, the *ARL* and *SDRL* values are larger than the corresponding in-control values. In such a situation it is better to use control charts for detecting small shifts like CUSUM and EWMA.

In Figure 2 we present the empirical run length distribution functions (ERL) for $n = 5$, 10, 20, 50. In each figure we plot six different lines representing the ERL functions for $m = 5$, 20, 50, 100, 1000 and the theoretical run length distribution (inf). We seen that as m increases the ERL

Table 9. *ARL* and *SDRL* Values for the *S* (Probability Limits) Chart When $n=20$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	327.65	381.36	170.43	279.59	56.40	125.43	18.47	47.23	8.07	17.04
10	379.94	415.88	154.94	241.56	40.09	71.48	13.24	19.53	6.45	8.01
20	421.10	434.89	135.60	194.87	33.08	46.55	11.89	14.66	5.93	6.32
30	442.13	451.34	126.95	170.03	30.45	37.64	11.46	12.67	5.81	6.00
50	461.32	467.99	117.98	139.49	29.17	32.83	11.09	11.54	5.69	5.50
100	476.40	478.77	113.42	126.66	27.69	28.82	10.97	10.94	5.57	5.26
200	486.38	486.97	109.86	115.12	27.35	27.82	10.50	10.20	5.50	5.17
500	485.13	488.22	108.31	108.90	26.81	26.19	10.43	10.12	5.41	4.89
1000	494.29	488.10	106.57	108.41	26.63	25.79	10.30	9.78	5.48	4.91
∞	500.01	499.51	106.64	106.14	26.67	26.17	10.42	9.91	5.46	4.93

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	1.17	0.51	7.00	10.62	56.41	103.25	245.11	324.91
10	1.14	0.43	6.11	7.00	45.91	69.83	247.16	304.65
20	1.13	0.40	5.71	5.85	40.82	47.93	233.76	273.54
30	1.13	0.39	5.44	5.32	40.07	44.78	228.37	256.46
50	1.12	0.36	5.49	5.21	38.59	41.06	223.43	242.44
100	1.12	0.37	5.36	4.95	37.85	39.05	219.40	225.87
200	1.11	0.36	5.40	4.92	37.99	38.81	217.81	217.89
500	1.11	0.36	5.36	4.87	37.41	36.61	213.84	209.40
1000	1.12	0.37	5.29	4.80	37.69	37.00	213.70	213.52
∞	1.12	0.36	5.29	4.77	37.44	36.94	215.93	215.43

approaches the theoretical run length distribution. Also, an increasing n value causes the ERLs for the m values to approach the theoretical run length distribution faster.

4. THE X CHART FOR MONITORING PROCESS DISPERSION

Let X_i, $i=1,\ldots,n$ denote independent and identically distributed observations from a $N(\mu,\sigma^2)$ process. If the parameters μ and σ^2 are

Table 10. *ARL* and *SDRL* Values for the *S* (Probability Limits) Chart When $n = 50$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	320.32	380.78	93.28	184.90	13.83	29.14	4.02	5.49	2.10	1.90
10	369.19	410.82	70.91	122.34	10.73	15.32	3.61	4.02	1.93	1.49
20	411.62	433.18	58.04	85.60	9.50	10.86	3.37	3.19	1.93	1.42
30	431.22	447.76	53.56	68.63	8.96	9.54	3.42	3.10	1.90	1.36
50	452.27	459.10	50.77	58.78	8.96	8.95	3.28	2.85	1.89	1.30
100	472.90	472.99	48.14	50.64	8.62	8.59	3.25	2.79	1.88	1.32
200	482.50	481.24	47.71	48.24	8.51	8.24	3.25	2.75	1.86	1.29
500	493.58	498.61	47.47	48.19	8.60	8.11	3.24	2.69	1.85	1.23
1000	490.32	499.04	47.59	47.66	8.56	8.10	3.23	2.66	1.86	1.27
∞	500.01	499.51	47.23	46.73	8.52	8.01	3.22	2.67	1.86	1.27

	0.2		0.4		0.6		0.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	1	0	1.21	0.62	7.47	11.58	107.80	188.75
10	1	0	1.19	0.50	6.19	6.96	92.22	141.22
20	1	0	1.18	0.47	5.86	5.90	79.17	102.14
30	1	0	1.16	0.45	5.69	5.50	74.86	87.60
50	1	0	1.16	0.45	5.65	5.46	71.97	78.71
100	1	0	1.15	0.42	5.64	5.22	69.87	73.40
200	1	0	1.16	0.43	5.62	5.18	69.61	69.69
500	1	0	1.15	0.42	5.49	4.95	68.90	70.23
1000	1	0	1.15	0.42	5.51	5.06	69.27	70.27
∞	1	0	1.16	0.43	5.48	4.96	68.04	67.54

known, the control limits are

$$UCL = \mu + 3\sigma$$
$$CL = \mu$$
$$LCL = \mu - 3\sigma$$

Usually, these parameters are not known and they have to be estimated. In this case, the variability is usually controlled using moving ranges. Nevertheless, Nelson (9), Roes et al. (10) and Rigdon et al. (11) have recommended either against the use of the moving range chart or its use together with the classical *X* chart. Moreover, Sullivan and Woodall (12) showed that

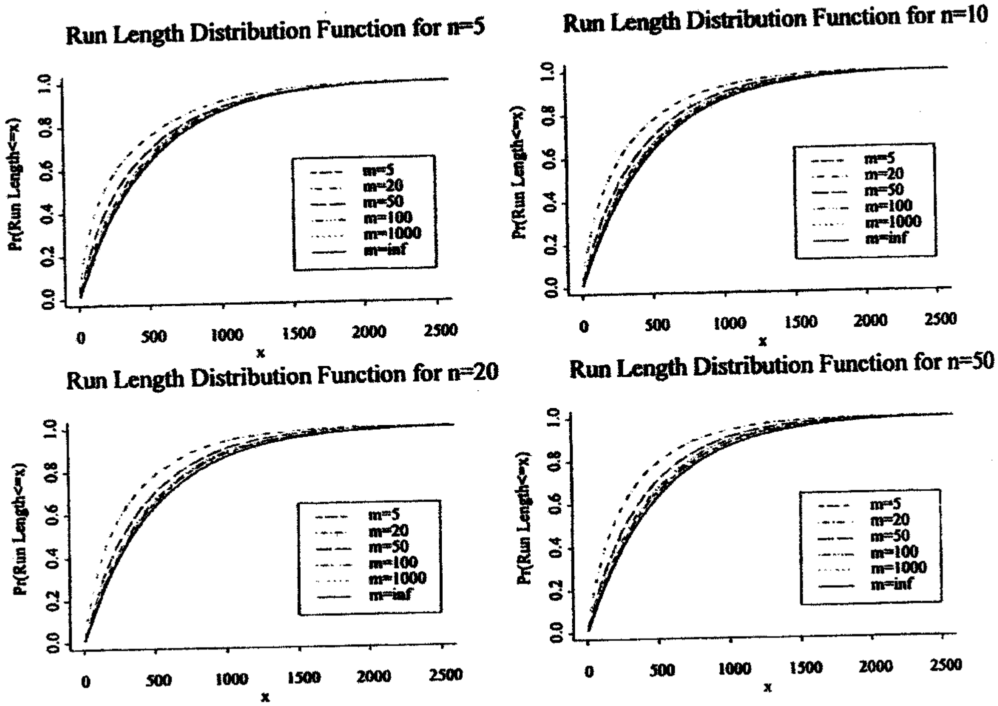

Figure 2. Empirical run length distribution functions for the probability limits chart.

a moving range control chart does not contribute significantly to the identification of out of control situations. Therefore, the use of the X control chart for monitoring the process standard deviation is recommended. The control limits of the X control chart are

$$\widehat{UCL} = \bar{X} + 3\hat{\sigma}$$
$$\widehat{CL} = \bar{X}$$
$$\widehat{LCL} = \bar{X} - 3\hat{\sigma}$$

where $\bar{X}$ is an unbiased estimate of the mean of the process and $\hat{\sigma}$ is an estimate of the standard deviation σ of the process. Usually, the estimate of the standard deviation used is $\overline{MR}/d_2$ where $\overline{MR}$ denotes the average of the moving ranges and d_2 is a constant used to make the estimator unbiased. However, Cryer and Ryan (13) showed that a preferable estimate of σ is s/c_4 where c_4 is defined the same way as in the case of rational subgroups and s is the standard deviation of the observations.

In order to assess the effect of the number of observations on the control limits of the X chart we performed a simulation study. The results

Table 11. *ARL* and *SDRL* Values for the *X* Control Chart

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
N	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
30	986.31	5024.83	315.36	1058.44	147.93	439.79	84.36	187.50	53.74	98.54
50	614.94	1565.00	229.95	476.60	116.69	200.50	69.61	107.23	47.23	66.81
75	503.75	948.78	202.02	318.54	105.18	150.77	64.51	84.15	43.99	54.87
100	467.07	770.60	190.53	274.54	100.73	131.39	61.98	75.26	42.78	50.48
200	413.88	518.65	173.68	205.96	93.86	105.77	58.63	63.56	40.67	42.81
300	398.94	476.34	167.79	187.69	92.76	100.47	57.93	61.37	41.26	42.29
500	387.38	429.45	167.90	179.39	90.34	93.58	56.80	58.96	39.69	40.54
1000	379.32	401.55	162.96	168.50	89.12	91.10	57.03	57.78	39.90	39.85
2000	372.64	383.71	162.70	166.87	89.45	89.41	56.35	55.82	39.62	39.17
∞	370.40	369.90	162.08	161.58	89.05	88.55	56.48	55.98	39.45	38.95

are presented in Table 11. For each value in the table we simulated *N* values from a $N(\mu, \sigma_0^2)$ distribution, we computed the $\widehat{UCL}$ and $\widehat{LCL}$ and subsequently we generated values from a $N(\mu, \sigma_1^2)$ distribution until we obtained a value above $\widehat{UCL}$ or below $\widehat{LCL}$. The number of samples simulated up to the one that was outside the control limits constitutes one observation of the run length. This procedure was repeated 32 000 times in order to get an estimate of the values of *ARL* and *SDRL*.

From Table 11 we see that we do not have results for downward shifts. This happens because a decreasing standard deviation will never cause a value below the lower control limit. The simulation reveals that the *ARL* and *SDRL* values decrease until they approach their theoretical values. We need at least 300 observations to minimize the effect of estimation in the control limits of the *X* chart.

In Figure 3 we present the empirical run length distribution functions (ERL) for $n = 30$, 50, 100, 200, 500, 1000, 2000 and the theoretical run length distribution (inf). The result is that as *n* increases the ERL approaches the theoretical run length distribution.

5. CONCLUSIONS

In this paper, we examined the effect of estimation on the control limits for process dispersion on charts using rational subgroups and individual observations. Extensive numerical studies for several combinations of numbers of

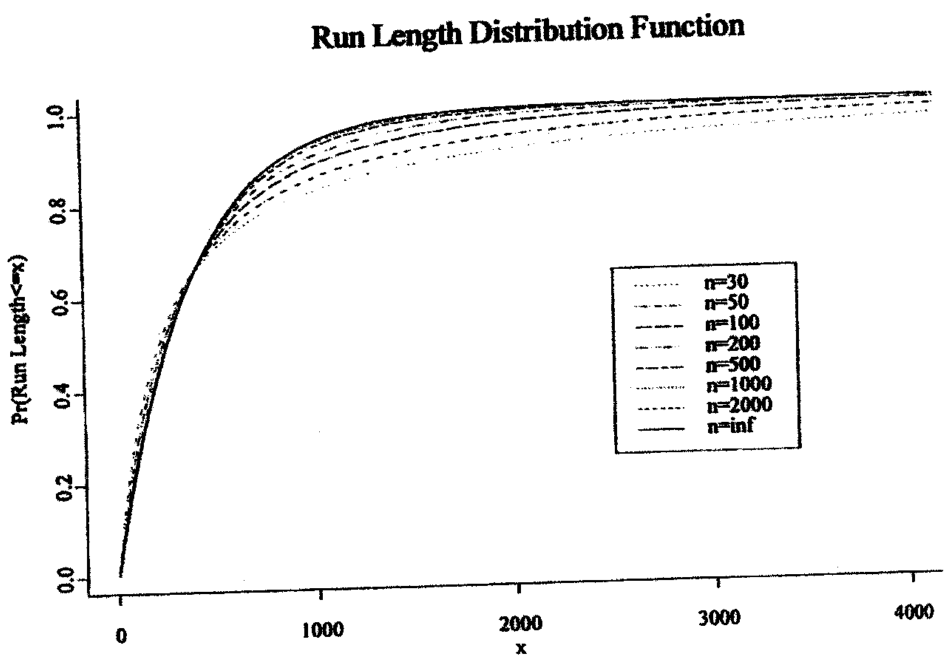

Figure 3. Empirical run length distribution functions for the X chart.

samples and of sample sizes in the case of rational subgroups and of numbers of observations in the case of individual control charts were presented. These values are used for proposing the m and n values that a practitioner should use in order to reduce the estimation effect on the univariate dispersion control charts.

In the rational subgroups case we propose larger n values than usual and someone may report that this is a problem. However, Woodall and Montgomery (4) remarked that in industry now there are large data sets available in contrast to the past. Therefore, such values for the sample size should not be a problem, generally. On the other hand, if for some special applications this still remains a problem, the practitioner should keep in mind the great influence on the estimated control chart performance displayed in the tables of this work.

REFERENCES

1. Quesenberry, C.P. The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts. Journal of Quality Technology **1993**, *25*(4), 237–247.

2. Chen, G. The Mean and Standard Deviation of the Run Length Distribution of $\bar{X}$ Charts When Control Limits Are Estimated. Statistica Sinica **1997**, *7*, 789–798.
3. Nedumaran, G.; Pigniatiello, J.J. On Constructing T^2 Control Charts for On-Line Process Monitoring. IIE Transactions **1999**, *31*, 529–536.
4. Woodall, W.H.; Montgomery, D.C. Research Issues and Ideas in Statistical Process Control. Journal of Quality Technology **1999**, *31*, 376–385.
5. Lowry, C.A.; Champ, C.W.; Woodall, W.H. The Performance of Control Charts for Monitoring Process Variation. Communications in Statistics—Simulation and Computation **1995**, *24*(2), 409–437.
6. Klein, M. Modified *S*-Charts for Controlling Process Variability. Communication is Statistics—Simulation and Computation **2000**, *29*(3), 919–940.
7. Chen, G. The Run Length Distribution of the *R*, *s* and s^2 Control Charts When σ Is Estimated. The Canadian Journal of Statistics **1998**, *26*(2), 311–322.
8. Ryan, T.P. *Statistical Methods for Quality Improvements.* John Wiley & Sons: New York, 2000.
9. Nelson, L.S. Control Charts for Individual Measurements. Journal of Quality Technology **1982**, *14*(3), 172–173.
10. Roes, K.C.B.; Does, R.J.M.M.; Schuring, Y. Shewhart-Type Control Charts for Individual Observations. Journal of Quality Technology **1993**, *25*(3), 188–198.
11. Rigdon, S.E.; Cruthis, E.M.; Champ, C.W. Design Strategies for Individuals and Moving Range Control Charts. Journal of Quality Technology **1994**, *26*, 274–287.
12. Sullivan, J.H.; Woodall, W.H. A Control Chart for Preliminary Analysis of Individual Observations. Journal of Quality Technology **1996**, *28*, 265–278.
13. Cryer, J.D.; Ryan, T.P. The Estimation of Sigma for an *X* Chart: $\overline{MR}/d_2$ or s/c_4? Journal of Quality Technology **1990**, *22*, 187–192.