Improved control chart performance using cautious parameter learning

Huberts, L.C.E.; Goedhart, R.; Does, R.J.M.M.

DOI
10.1016/j.cie.2022.108185

Publication date
2022

Document Version
Final published version

Published in
Computers & Industrial Engineering

Citation for published version (APA):
Huberts, L. C. E., Goedhart, R., & Does, R. J. M. M. (2022). Improved control chart performance using cautious parameter learning. *Computers & Industrial Engineering*, *169*, Article 108185. https://doi.org/10.1016/j.cie.2022.108185

Computers & Industrial Engineering 169 (2022) 108185

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Improved control chart performance using cautious parameter learning

Leo C.E. Huberts [*], Rob Goedhart, Ronald J.M.M. Does

Department of Business Analytics, University of Amsterdam, Plantage Muidergracht 12, 1018TV Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Estimation Effects
Statistical Process Monitoring
Shewhart
CUSUM
EWMA

ABSTRACT

Parameter estimation is an important topic in Statistical Process Monitoring, as inaccurate estimates may lead to undesirable control chart performance. Updating the control chart limits during the monitoring period reduces estimation uncertainty. However, when out-of-control situations remain undetected, using the corresponding samples to update the parameter estimates can deteriorate the control chart performance in terms of in-control and out-of-control run lengths. For this reason, updating parameter estimates should only occur when there is sufficient evidence of an in-control process state. In this article, we study the performance of a cautious updating scheme for the Shewhart, Cumulative Sum, and Exponentially Weighted Moving Average control charts. We propose simple rules for updating parameters that improve the out-of-control performance of the control charts. We show the added value of using these updating rules in practice through a case study using data from a truck manufacturer.

1. Introduction

Control charts are used to monitor quality indicators in industry and services. A wide range of charts has been developed. The Shewhart, Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts, introduced by Shewhart (1926), Page (1954) and Roberts (1959), respectively, are the most commonly used charts in practice. These three charts were developed to detect changes in the underlying process, often called assignable or special cause variation. The Shewhart chart is simple to interpret and implement and is capable of quickly detecting large shifts in the mean of the process quality indicator. The CUSUM and EWMA charts are more difficult to interpret, as both incorporate previous observations in their plotting statistic. These two charts are generally better at detecting small shifts (see Vera do Carmo et al. (2004) for a comparison).

All three mentioned charts have parameters that need to be estimated in practice. This causes uncertainty in the charts' performance. The effects of this uncertainty have been widely researched in recent years and solutions have been proposed to deal with this uncertainty. Jensen et al. (2006) and Psarakis et al. (2014) conducted literature reviews on this parameter estimation in control charts and identified directions for future research. Recently, several researchers have proposed adjusted control chart designs based on guaranteed in-control performance (see, e.g. Gandy and Kvaløy, 2013, Saleh et al. (2015, 2016) Goedhart et al. (2017a, 2017b), Zwetsloot and Ajadi (2019), Diko et al. (2019)). One of the directions that can further improve conditional control chart performance concerns re-estimating the control limits of a control chart. A few studies have been published on this topic, among which Huberts et al. (2019), investigating the effects of updating the control limits in various scenarios and Capizzi and Masarotto (2020), proposing a delayed updating procedure for the Shewhart, CUSUM, and EWMA charts. This article builds upon this work as Huberts et al. (2019) show that updating can improve performance in certain settings and the proposed delayed updating by Capizzi and Masarotto (2020) is a promising approach.

In this article, we evaluate and extend the approach of Capizzi and Masarotto (2020). Depending on the practitioner's needs important choices have to be made concerning if and when to update. These choices depend on the type of control chart, the sizes of mean deviations deemed important, and the desired false alarm rate. We will therefore investigate the performance of the updating procedure for different control charts, deviations and desired false alarm rates. Based on this, we will extend the approach of Capizzi and Masarotto (2020) to improve performance across a wide range of settings.

The article is structured as follows. In the next section, we provide the designs of the Shewhart, CUSUM, and EWMA charts. This is followed by an explanation of the procedure proposed by Capizzi and Masarotto (2020) in Section 3. In the subsequent section, we analyze the performance of this procedure in various settings. Subsequently, the adjustments to the procedure are motivated in Section 5 and a practical

* Corresponding author.
E-mail address: L.c.e.huberts@uva.nl (L.C.E. Huberts).

https://doi.org/10.1016/j.cie.2022.108185
Received 15 June 2020; Received in revised form 17 January 2022; Accepted 12 April 2022
Available online 16 April 2022

example is given in Section 6. In the last section, we provide some concluding remarks.

2. Control Chart Designs

In this section, we outline the Shewhart, CUSUM, and EWMA control chart designs. Let x_i denote the observation at time i. Assume that in Phase I an in-control sample of m observations is available. Phase I runs from $i = -m+1$ to $i = 0$. Further, we assume that the Phase I samples are independently and identically $N(\mu, \sigma^2)$ distributed and that the observations in Phase II are independently and identically $N(\mu + \delta\sigma, \sigma^2)$ distributed. Note that Phase II starts at $i = 1$.

In practice, for each of the three control charts μ and σ are unknown and need to be estimated. There are a range of estimators that can be used, the choice of which is outside of the scope of this study (see, e.g., Li et al. (2019), Testik et al. (2020), Montgomery (2013) for more details on this). We use the same estimators for each chart in this study. The parameter μ is estimated by

$$\bar{x}_i = \frac{1}{m+i}\sum_{r=-m+1}^{i} x_r, \tag{1}$$

Further, σ is estimated by

$$s_i = \left(\frac{1}{m+i-1}\sum_{r=-m+1}^{i}(x_r - \bar{x}_i)^2\right)^{1/2}. \tag{2}$$

The following subsections outline the Shewhart, CUSUM, and EWMA control chart designs, based on a Phase I in-control sample of size m.

2.1. Shewhart Control Chart

The estimated Shewhart control limits are given by

$$\begin{aligned}\widehat{UCL} &= \bar{x}_0 + L_s s_0,\\ \widehat{LCL} &= \bar{x}_0 - L_s s_0,\end{aligned} \tag{3}$$

where $\bar{x}_0$ is given by (1) and s_0 by (2), while L_s is some positive constant for the Shewhart control chart which we elaborate on in Section 2.4. The Shewhart control chart signals if x_i is larger than $\widehat{UCL}$ or smaller than $\widehat{LCL}$.

2.2. CUSUM Control Chart

The two-sided cumulative sum (CUSUM) control chart uses the cumulative sum of observations to monitor the process. The upper and lower statistics are calculated by

$$C_i^+ = max\left(0, C_{i-1}^+ + \frac{x_i - \bar{x}_0}{s_0} - k\right) \tag{4}$$

$$C_i^- = min\left(0, C_{i-1}^- + \frac{x_i - \bar{x}_0}{s_0} + k\right), \tag{5}$$

with the chart parameter $k \geqslant 0$ and $C_0^+ = C_0^- = 0$. This CUSUM chart signals if either $C_i^- < -L_c$ or $C_i^+ > L_c$, where the critical value L_c is a positive constant for the CUSUM control chart defined in Section 2.4.

2.3. EWMA Control Chart

The EWMA control chart weighs observations over time. The EWMA statistic is defined as

$$Z_i = \lambda x_i + (1-\lambda)Z_{i-1} \tag{6}$$

where $0 < \lambda \leq 1$ and Z_0 equals the mean estimate $\bar{x}_0$. For $\lambda = 1$ the EWMA control chart is equal to the Shewhart control chart. The EWMA control limits for monitoring the process at time $i = m+1, m+2, \ldots$ are

$$\widehat{UCL}_i = \bar{x}_0 + L_e s_0 \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{7}$$

$$\widehat{LCL}_i = \bar{x}_0 - L_e s_0 \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{8}$$

where λ and L_e determine the in-control performance. When Z_i falls above (below) $\widehat{UCL}_i$ ($\widehat{LCL}_i$) the process is considered out-of-control.

2.4. Critical Value L

The performance of the three control charts depends heavily on the choice of the critical value L (L_s for the Shewhart, L_c for the CUSUM, and L_e for the EWMA control chart). Classical control chart design would suggest using a value of L that delivers the desired in-control ARL for known parameters. In recent years the focus has shifted towards using a value for L that guarantees a certain probability that the in-control ARL will be *at least* the desired ARL value, as proposed by Gandy and Kvaløy (2013) and others (see for example, Jones and Steiner (2012), Saleh et al. (2015, 2016), Goedhart et al. (2017a, 2017b)). In that setting, the value L is determined such that $Prob(ARL_{ic} > ARL_0) = 1 - \beta$, where ARL_0 is the desired average run length and β is the accepted (small) probability that the average run length will be shorter than ARL_0. The value of L, given ARL_0 and β, can be determined using analytical or numerical/ Monte Carlo procedures.

3. Updating the Control Chart Limits

The control chart designs as described in the previous section assume a fixed Phase I sample for monitoring. In many cases, in-control Phase II data could be used to re-estimate Eqs. (1) and (2) and update the control limits. Updating the estimates of the mean and standard deviation could occur after every new observation as with self-starting control charts (Hawkins (1987) and Quesenberry (1991)).

Huberts et al. (2019) recently explored a variety of scenarios and concluded that updating is often a good choice but the type of chart and size of shift (δ) are important. Furthermore, the outcome depends on the ability of practitioners to retrospectively identify out-of-control samples. A potential hazard of an updating scheme is that small shifts may not directly be detected, in which case the corresponding out-of-control observations would be used in the updated in-control parameter estimates. An approach to counter this is to use a delay in updating, as is done by Capizzi and Masarotto (2020). The effect of updating the control limits on the control chart performance depends heavily on parameters m, δ, ARL_0, the type of chart, and the choices made by practitioners related to the data and the moment of updating.

The approach of Capizzi and Masarotto (2020) is to update Eqs. (1) and (2) using a delay. Monitoring begins at time $i = 1$. The main concept is that at some time $i > 0$, if it is reasonable to assume the process is in control the newly collected samples together with the initial m Phase I observations can be used to determine updated values of Eqs. (1) and (2). This reduces the parameter estimation uncertainty. The time at which an update occurs could be fixed beforehand or determined using the collected samples. Capizzi and Masarotto (2020) propose a solution for the latter option, using the following inequality

$$\sum_{j=i-d_i+1}^{i}\left(\frac{x_j - \bar{x}_{i-d_i}}{s_{i-d_i}}\right)^2 < Ad_i - B, \tag{9}$$

where d_i counts the number of samples from the last update, $d_1 = 1$, A and B are parameters that need to be set by the practitioner. As long as this inequality does not hold, d_i (the updating delay) is increased by one (i.e. $d_{i+1} = d_i + 1$). Thus, the right-hand side of the inequality increases

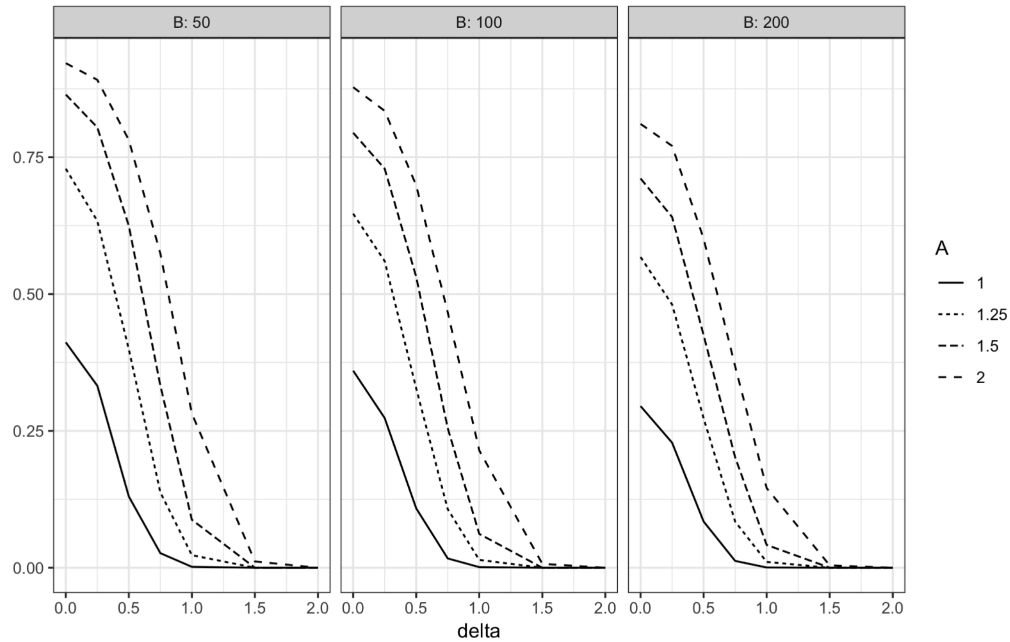

Fig. 1. Percentage of control charts with a first update before a signal for $ARL_0 = 370$.

by A every i as long as there is no update. Capizzi and Masarotto (2020) propose using the values $A = 1.5$ and $B = 50$. As we will show in the following section, this procedure can result in a deterioration of out-of-control chart performance. Improvements can be made to the settings, which we will propose in Section 5.

3.1. Unconditional Expectation

The unconditional expectation of an individual term in the sum on the left-hand side of Inequality (9) can be shown to be (cf. Appendix A)

$$E\left[\left(\frac{x_j - \overline{x}_{i-d_i}}{s_{i-d_i}}\right)^2\right] = \left(\frac{m-1}{m-3}\right)\left(1 + \delta^2 + \frac{1}{m}\right) \tag{10}$$

for $m > 3$. This shows that, in expectation, the left-hand side of Inequality (9) increases faster with larger values of δ thus confirming that the procedure is less likely to update when a mean shift has occurred. For example, given an in-control process with $\delta = 0$ and $m = 50$, the expectation equals 1.0634. Therefore, in expectation, for values $A < 1.0634$ the left-hand side of Inequality (9) increases faster than the right-hand side, preventing updates. For $\delta = 0.5, m = 50$ the expectation in Eq. (10) equals 1.324. Then for $A = 1.5$ as considered by Capizzi and Masarotto (2020), in expectation, the right-hand side of Inequality (9) grows more quickly than the left-hand side. The setting for B does not affect the growth rate but does determine the delay. A larger value of B means that the right-hand side of Inequality (9) starts at a lower negative value, leading to larger updating delays. The settings for A and B are very important for the chart performance, as we demonstrate in the following sections.

3.2. Conditional Expectation

We will now consider the conditional expectation of the sum on the left-hand side of Inequality (9). We only consider the time until the first update, such that $d_i = i$ until the update is done, so that $\overline{x}_{i-d_i} = \overline{x}_0$ and $s_{i-d_i} = s_0$, and such that Inequality (9) becomes

$$\sum_{j=1}^{i}\left(\frac{x_j - \overline{x}_0}{s_0}\right)^2 < Ai - B. \tag{11}$$

In Appendix B we show that the expectation of the left-hand side of Inequality (11), conditional on $\overline{x}_0$ and s_0, is equal to

$$E\left[\sum_{j=1}^{i}\left(\frac{x_j - \overline{x}_0}{s_0}\right)^2 |\overline{x}_0, s_0\right] = i\left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2}. \tag{12}$$

Next, we replace the sum in the left-hand side of Inequality (11) by its expectation, so that we obtain the following inequality

$$i\left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2} < Ai - B \tag{13}$$

We use this inequality to provide an estimate of the expected time to the first update (ETFU). Since B should be a positive number in this method, note that this inequality will never be true if $\left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2} \geqslant A$. If $\left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2} < A$, then we can solve the inequality for i and find that

$$i \geqslant \frac{B}{A - \left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2}} \tag{14}$$

Thus, our estimate of ETFU, conditional on $\overline{x}_0$ and s_0, is equal to

$$ETFU|\overline{x}_0, s_0 = \left\lceil \frac{B}{A - \left(1 + \left(\frac{\mu - \overline{x}_0}{\sigma} + \delta\right)^2\right)\frac{\sigma^2}{s_0^2}} \right\rceil \tag{15}$$

where $\lceil . \rceil$ represents the ceiling function. The *ETFU* shows that B and A are important, as well as the shift size δ and the parameter estimation error. Although the *ETFU* itself is an approximation, it provides some useful insight into the essence of the problem at hand. For example, given $\delta = 0.5, \overline{x_0} = \mu, s_0^2 = \sigma^2, A = 1.5$,and $B = 50$, the expected first update will occur at $ETFU = 200$. For the Shewhart chart with $ARL_0 = 500$ the unconditional ARL for $\delta = 0.5$ equals 202. This means that, in expectation, this Shewhart chart will update the parameter estimates with out-of-control observations before it is able to signal them.

Table 1
Average run lengths (ARL) for 6,000 simulated control charts and $m = 50$ in-control Phase I samples.

	m	50																				
	δ	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2
A	B	Shewhart							EWMA							CUSUM						
0	0	7388	6321	3242	1178	527	103	28	9528	4018	535	78	25	8	5	12572	6878	1341	205	43	9	5
1	50	1255	1197	1160	827	458	98	31	1420	858	276	73	24	8	5	1431	1020	457	139	39	9	5
	100	1417	1358	1197	828	480	103	26	1598	996	295	**69**	24	8	5	1623	1136	478	137	39	9	5
	200	1692	1633	1387	863	451	110	29	2017	1145	308	71	24	8	5	2000	1366	528	136	38	9	5
1.25	50	726	698	654	553	394	101	28	781	583	278	88	26	8	5	802	680	405	148	42	9	5
	100	835	793	736	622	409	100	27	888	633	273	76	24	8	5	884	730	401	139	39	9	5
	200	981	968	856	672	437	105	29	1060	735	271	76	25	8	5	1081	853	425	**135**	39	9	5
1.5	50	686	659	594	469	320	106	28	677	565	296	97	25	8	5	729	651	437	186	50	9	5
	100	726	704	625	491	338	103	27	740	577	293	82	25	8	5	796	645	436	158	43	9	5
	200	798	782	662	536	347	101	27	856	658	**264**	76	24	8	5	887	759	411	150	**38**	9	5
2	50	585	**569**	**548**	466	300	**70**	**20**	561	**487**	269	102	23	**7**	**5**	591	**551**	417	213	68	**8**	**4**
	100	615	611	557	**458**	289	77	21	634	511	273	81	23	8	5	657	588	**396**	180	49	8	4
	200	690	676	589	464	**284**	87	25	713	574	268	71	**22**	8	5	753	650	406	155	40	9	5

3.3. The Updating Parameters

To analyze the impact of A, B, and δ on the time to update and on the ARL performance we use a Monte Carlo simulation because analytical expressions for the three control charts are unfeasible. We do this by determining the probability that a control chart will update the parameter estimates before it produces an out-of-control signal. We perform a simulation for the Shewhart chart, but the same principle applies to the EWMA and CUSUM charts. We apply the following procedure for $ARL_0 = 200, 370$ and 500:

For $j = 1, 2, \ldots, 6{,}000$:

1. Simulate a $N(0,1)$ Phase I sample X_j^I of size $m = 50$ and calculate $x_{0,j}$ and $s_{0,j}$, which are the Phase I estimates according to Eqs. (1) and (2), respectively, for sample X_j^I.
2. Initiate the Shewhart control charts using $x_{0,j}$ and $s_{0,j}$ for all nine combinations of $A \in \{1, 1.25, 1.5\}, B \in \{50, 100, 200\}, m = 50$ and the CautiousLearning R-package of Capizzi and Masarotto (2020)
3. Simulate $N(\delta, 1)$ distributed Phase II samples $X_{\delta j}^{II}$ of size 1,000,000 for a wide range of δ $(0.0, 0.25, 0.5, \ldots, 2.0)$ and calculate the first update $FU_{\delta j}$ and first signal $FS_{\delta j}$ given $\bar{x}_{0,j}$ and $s_{0,j}$.
4. If $j < 6{,}000$ increment j by 1 and go back to step 1
5. Calculate the percentage of charts that have a first update before first signal as $\frac{\sum_{j=1}^{6{,}000} I(FU_{\delta j} < FS_{\delta j})}{6{,}000}$ where $I(FU_{\delta j} < FS_{\delta j}) = 1$ when $FU_{\delta j} < FS_{\delta j}$ and $I(FU_{\delta j} < FS_{\delta j}) = 0$ if $FU_{\delta j} \geqslant FS_{\delta j}$.

The results of the simulation procedure for $ARL_0 = 370$ and various combinations of δ, A and B are shown in Fig. 1. For $ARL_0 = 200$ the percentages of charts that update before signaling are slightly lower and for $ARL_0 = 500$ slightly higher.

Fig. 1 shows that for values of δ smaller than 1.5, the charts often update using out-of-control observations. For example, given $ARL_0 = 370, A = 1.5, B = 50$ and $\delta = 0.5$ the percentage of Shewhart charts that update before signalling is larger than 60%. This means that there is a substantial risk of using out-of-control observations to update in-control parameter estimates, which may negatively affect control chart performance, as we show in the next section.

4. Performance

The previous section has shown that there is a large likelihood of updating control limits using out-of-control samples. The effects on chart performance in terms of in-control and out-of-control average run lengths are studied in this section.

We perform a Monte Carlo simulation to assess the effects of the updating parameters (A, B) and the shift size (δ) on the control chart performance. For the Shewhart, EWMA, and CUSUM charts we set $ARL_0 = 370$, where we consider $k = 1$ (CUSUM) and $\lambda = 0.2$ (EWMA) as in Capizzi and Masarotto (2020) for comparison purposes. We have also analyzed $k = 0.5, \lambda = 0.5, ARL_0 = 200$ and $ARL_0 = 500$ for which the results were very similar. See Hawkins and Wu (2014) for a comparison of the Shewhart, EMWA and CUSUM charts with various design parameters.

We let δ vary from 0 to 2 in steps of 0.25, A from 1 to 2 in steps of 0.5, B from 50 to 200 in steps of 50 and include the reference without-updating ARL values $(A = 0, B = 0)$. For each combination of δ, A, and B we simulate 6,000 Shewhart, EWMA, and CUSUM charts using the CautiousLearning R-package (Capizzi and Masarotto (2020)) and calculate the ARL as the average of the run lengths of these 6,000 charts. Note that the charts are configured to achieve $Prob(ARL_{ic} > ARL_0) = 1 - \beta$ as described in Section 2.4, therefore the realized ARL values with $\delta = 0$ will not be equal to the ARL_0 values, and may differ across control chart types. For $\delta > 0$ the ARL values are a measure of the detection power of the control chart, as also used in Capizzi and Masarotto (2020).

The results for $m = 50$ are reported in Table 1. Since the control charts are designed to provide a guaranteed in-control $(\delta = 0)$ performance when parameters are estimated, we focus on the out-of-control $(\delta > 0)$ performance here. Note that in the out-of-control situation smaller ARL values are preferred. Therefore, for each out-of-control column in Table 1, smaller ARL values are indicated with darker shading. The lowest value is printed in bold.

We can evaluate the performance of the different combinations of A and B in the various scenarios. A first observation is that the values

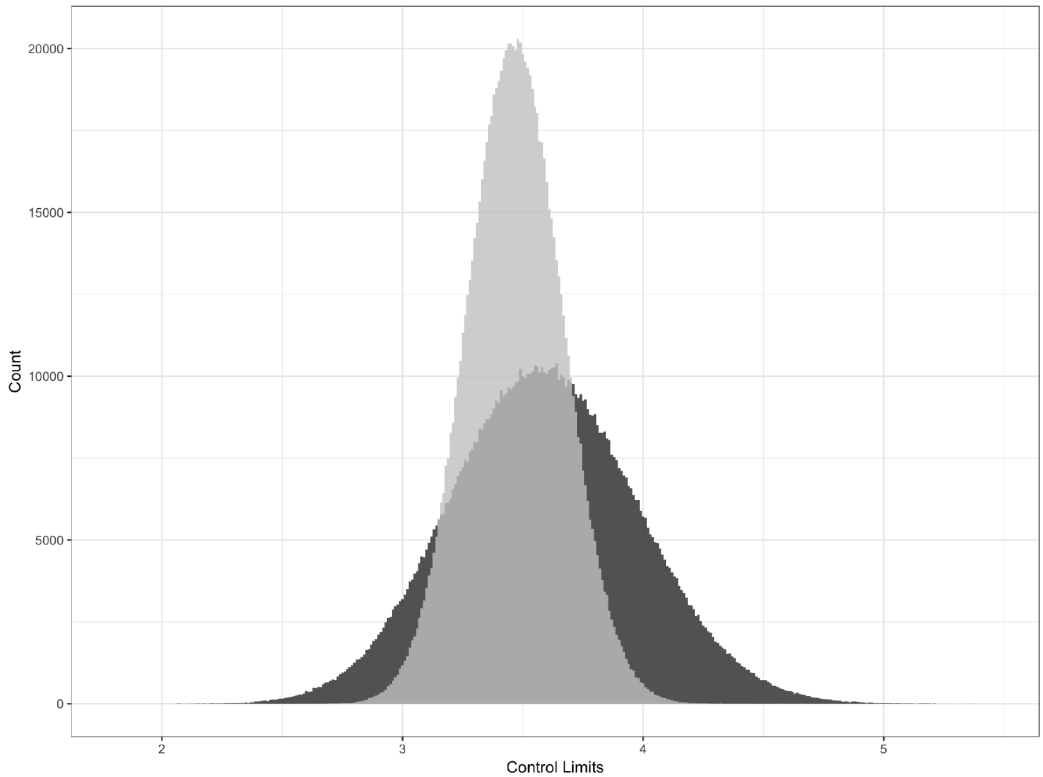

Fig. 2. Histograms of 10 million simulated Shewhart control limits based on $m = 50$ (dark grey) and updated control limits using an additional 115 contaminated observations (light grey).

chosen by Capizzi and Masarotto (2020), $A = 1.5$ and $B = 50$, are sub-optimal for all cases. The Shewhart chart performs best for larger values of A regardless of δ. It appears that updating the parameter estimates is very important for the Shewhart control chart performance in this situation. For the EWMA and CUSUM charts the optimal parameters do not show a clear pattern, but for small values of $\delta = (0.25, 0.5)$ it is best to update quickly using a larger value for A and a smaller value for B. For larger values of δ, better results are achieved for smaller values of A and

Table 2
Average run lengths (ARL) for 6,000 simulated control charts and $m = 250$ in-control Phase I samples.

	m	250																				
	δ	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2
A	B	Shewhart							EWMA							CUSUM						
0	0	895	636	330	163	84	25	9	821	252	58	23	12	6	4	957	428	121	41	18	6	4
1	50	659	541	324	163	81	24	9	624	247	60	22	12	6	4	668	**396**	126	40	18	6	4
	100	715	572	331	162	84	25	9	675	252	**57**	22	12	6	4	740	401	**120**	41	17	6	4
	200	779	603	**323**	**160**	81	25	9	717	**247**	58	22	12	6	4	789	430	123	40	18	6	4
1.25	50	569	516	358	170	84	25	9	538	266	60	22	12	6	4	583	415	142	42	17	6	4
	100	582	526	339	166	81	24	9	569	255	58	22	12	6	4	602	403	132	40	18	6	4
	200	649	547	327	163	84	26	10	619	251	57	22	12	6	4	659	400	122	41	17	6	4
1.5	50	533	508	389	190	82	24	9	507	277	64	21	12	6	4	537	436	176	44	17	6	4
	100	566	519	371	170	**78**	24	9	538	274	58	22	12	6	4	572	420	146	40	**17**	6	4
	200	590	531	338	163	84	24	9	568	250	58	22	12	6	4	599	398	132	40	17	6	4
2	50	528	**492**	453	317	128	**23**	9	488	288	71	**21**	**11**	**6**	4	526	429	208	54	17	6	4
	100	547	503	420	262	98	24	**9**	500	273	62	21	11	6	**4**	541	418	180	46	17	**6**	**4**
	200	554	510	390	205	82	24	9	539	271	59	21	12	6	4	563	418	158	**39**	17	6	4

Table 3
Average run lengths (ARL) for 6,000 simulated control charts and $m = 500$ in-control Phase I samples.

	m	500																				
	δ	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2
A	B	Shewhart							EWMA							CUSUM						
0	0	627	474	247	**117**	65	20	8	572	183	47	19	11	6	4	640	322	97	35	15	6	4
1	50	572	449	254	130	65	21	8	542	180	49	20	11	6	4	578	313	102	35	16	6	**3**
	100	581	483	252	123	66	20	8	543	181	47	20	11	6	4	594	**310**	96	35	16	6	4
	200	604	469	**244**	125	64	21	8	570	**179**	46	20	11	6	4	628	314	**95**	36	15	6	4
1.25	50	523	467	273	124	65	21	8	497	202	47	19	11	6	4	525	346	110	**34**	16	6	4
	100	516	**444**	261	126	66	21	8	518	193	49	20	11	6	4	540	326	101	36	16	6	4
	200	559	459	250	127	64	21	8	528	187	47	19	11	6	4	556	323	99	35	16	6	4
1.5	50	501	459	326	145	65	20	8	489	218	48	19	11	6	4	513	353	121	35	16	6	4
	100	510	468	306	125	**63**	21	8	492	202	47	**19**	11	6	4	531	350	112	35	**15**	6	4
	200	534	446	262	125	66	20	8	512	194	**46**	19	11	6	4	540	323	100	35	16	6	4
2	50	498	461	354	220	84	20	**8**	472	218	51	**19**	**11**	6	**4**	504	359	136	37	16	6	3
	100	502	464	345	186	73	20	8	473	213	48	20	11	**6**	4	507	350	123	36	15	**6**	4
	200	520	452	320	146	65	**20**	8	483	197	47	20	11	6	4	516	345	113	**34**	16	6	4

larger values of B.

We conclude that updating the parameter estimates using contaminated samples can have a positive effect on performance. This surprising finding is due to the large parameter estimation uncertainty when $m = 50$. To illustrate this, we calculate the unconditional expected time to first update using Eq. (15). We then compare the estimated upper control limit values using only the m samples of Phase I to the estimated upper control limit when using $m + ETFU_{\delta,m}$ samples. The latter updates the control limits using contaminated Phase II data.

We simulated control limits based on two scenarios. The first scenario is a Phase I sample consisting of $m = 50$ in-control observations from a standard normal distribution. For the second scenario, we consider Phase II observations with a small shift of $\delta = 0.25$. The expected time to the first update for $\delta = 0.25$ and $m = 50$ equals $ETFU_{\delta=0.25,m=50} = 115$ samples. Because of this, we consider estimated control limits based on 165 observations of which 50 have mean 0 and 115 have mean 0.25 in the second scenario (all with unit variance). The results are displayed in Fig. 2. The control limits for the first scenario are displayed in dark grey, and the control limits for the second scenario are displayed in light grey.

The distribution of the updated control limits in light grey is more narrow due to updating the parameter estimates. A small bias has been introduced, as Phase II samples with mean deviation $\delta = 0.25$ have been included in the parameter estimates. However, the updated limits are on average still more accurate than the original Phase I control limits. The reduction in parameter uncertainty outweighs the small bias that is introduced. This is because the value of L_s (cf. Section 2.4) required to guarantee a minimum in-control performance will be smaller when more observations are available. In particular, for the non-updated limits we have $L_s = 3.61$ for $m = 50$, while for the updated limits we have $L_s = 3.26$ when using estimates based on 165 observations (cf. the CautiousLearning R-package by Capizzi and Masarotto (2020)). As a consequence, even though a positive bias is introduced in the estimate of the mean, the estimated control limits will move closer towards $\bar{x}$ in this situation.

We have repeated the Monte Carlo simulation of Table 1 for larger Phase I sample sizes $m = (250, 500)$. The results for $m = 250$ are reported in Table 2, and for $m = 500$ in Table 3. Consider Table 2 with $m = 250$. Compared to Table 1, the parameter estimation error is smaller. For the smallest $\delta = 0.25$, the Shewhart chart should still update quickly using parameters $A = 2$ and $B = 50$. For values of $\delta > 0.5$ this is not the case, as setting $A = 1$ and $B = 200$ provides better results here. The EWMA and CUSUM charts show a similar pattern for small values of A. The CUSUM does require a lower value of B for small δ. Table 3 shows the results when $m = 500$. In this case the Phase I sample size is larger still and hence parameter estimation is more accurate. Table 3 clearly shows that $A = 1$ or $A = 1.25$ generally performs well. This means updating very slowly or not at all. For the Shewhart chart with $\delta = 0.75$ the best performing chart is the non-updating chart $A = 0, B = 0$. Note that for large shifts ($\delta = 1.5, 2$), for almost all charts and all m, setting $A = 2$ and $B = 50$ achieves the optimal ARL.

5. Improvements

In this section, we discuss the optimal settings when (cautiously) updating the Shewhart, EWMA, and CUSUM charts. As shown in the previous section these settings depend on the number of Phase I samples m, the desired in-control average run length (ARL_0), and the mean shift δ.

The first general result is that the EWMA chart given the chosen parameter settings yields the smallest out-of-control ARL values for all combinations of δ and m. The second general finding is that for large Phase I sample sizes (i.e. $m \geqslant 500$), updating the limits often has negative effects on the control chart performance. This is in line with the recommendations in the literature that at least $m = 300$ samples are needed to sufficiently reduce variability in control chart performance (Quesenberry (1993)). Thus, when a sufficient number of observations ($m \geqslant 500$) are available, we recommend using the EWMA chart for $\delta \leqslant 1$ and not updating the Phase I parameter estimates.

The optimal choice of A and B depends on the value of δ that is important to the practitioner, as well as the number of available in-

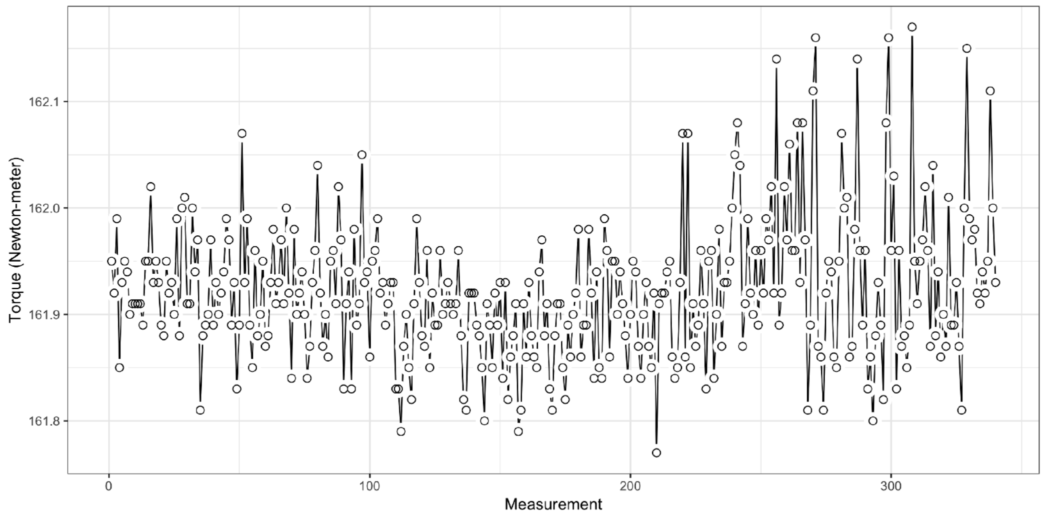

Fig. 3. Torque measurements in Newton-meter used to fasten bolts in truck engines.

control Phase I samples m. Tables 1–3 give guidance on choosing the optimal values of A and B. To be able to use these results in an algorithm or software package, we have translated the findings from these tables into a few very simple rules of thumb. These rules were determined using the set of equations that follow from Tables 1–3 and will result in values for A and B that are close to optimal.

1. For large numbers of Phase I samples ($m \geqslant 500$) consider if updating is still necessary.
2. For detecting moderate to large shifts ($\delta > 1$) set $A = 2, B = 50$.
3. For detecting small shifts ($\delta \leqslant 1$) use the following rules. For the Shewhart chart set A and B as

$$A = max\left(\left\lceil 2 - \frac{1}{2}|\delta| - \frac{m-50}{250}\right\rceil, 0\right) \tag{16}$$

$$B = (m+50)|\delta|. \tag{17}$$

For the EWMA and CUSUM charts set A and B as

$$A = max\left(\left\lceil 2 - \frac{4}{3}|\delta| - \frac{m-50}{250}\right\rceil, 0\right) \tag{18}$$

$$B = 2(m+50)|\delta|. \tag{19}$$

These rules will result in the use of the values of A and B that deliver good out-of-control performance and less unnecessary updating when a large number of Phase I samples are available. Note that these rules apply to the specific settings investigated in this paper and do not (necessarily) generalize to other control chart settings.

5.1. *Signal behavior*

The main motivation for updating control chart limits during monitoring (Phase II) is a lack of sufficient reliable Phase I data when monitoring is required. Thus any updating monitoring scheme should consider signal behavior. Capizzi and Masarotto (2020) advise to re-run Phase I methods on all data collected so far, and re-estimating the parameters with the remaining representative observations. Huberts et al. (2019) presented examples of scenarios where updating and continued use of the chart after a signal is beneficial. If the practitioner can retrospectively identify out-of-control samples and remove them from the data, the chart can safely be updated even after signals. In situations where this is not possible and there is no way to distinguish a false alarm from a correct out-of-control signal, updating is often inadvisable. This does depend on the values of δ, m, and the chart that is used (Huberts et al. (2019)).

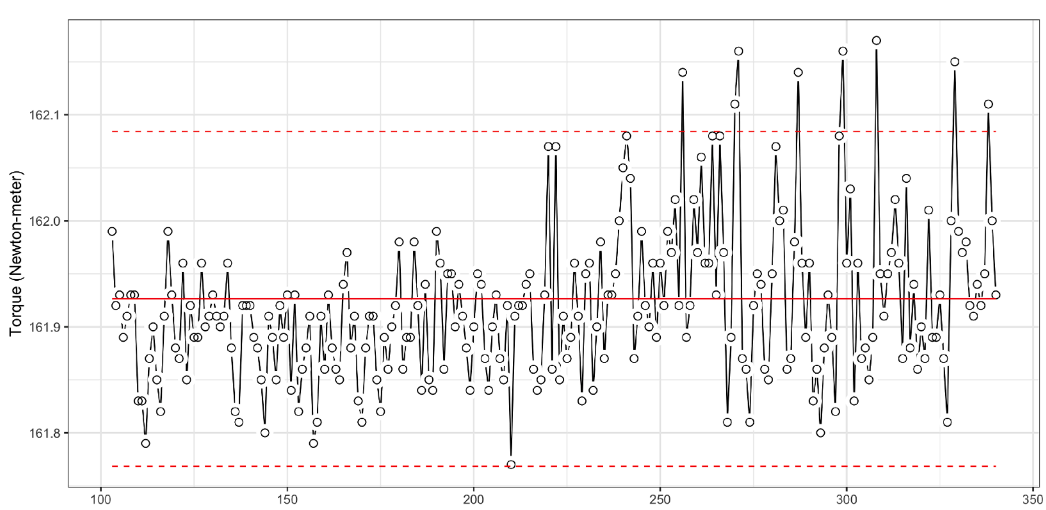

Fig. 4. The Shewhart control chart during monitoring in Phase II, updating the limits ($A = 1.5, B = 50$).

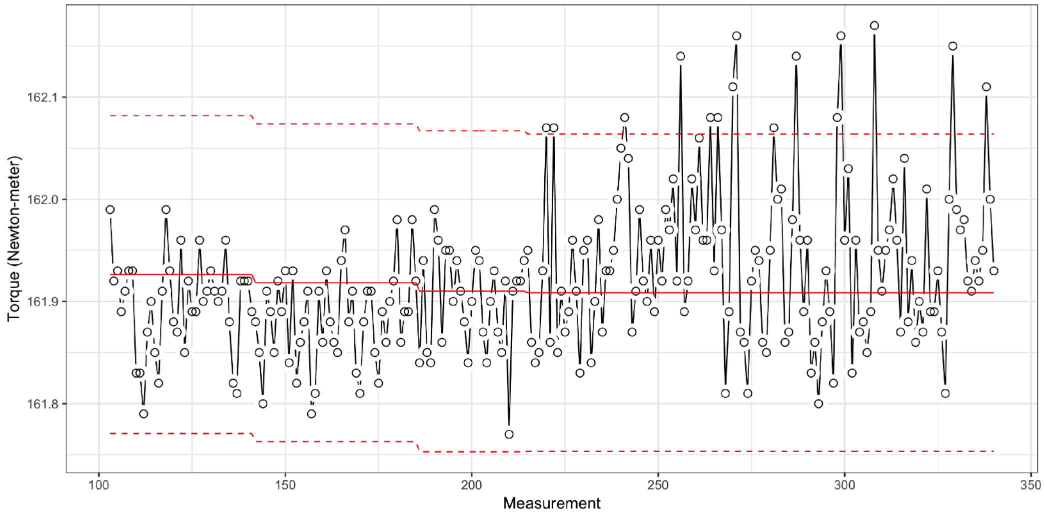

Fig. 5. The Shewhart control chart during monitoring in Phase II, updating the limits using the rules of thumb in Section 5 ($A = 2, B = 30.4$).

6. Case Study

In this section, we demonstrate the (cautious) updating procedure using data of the torque of Torque-to-Yield bolts at PACCAR, a global manufacturer of medium- and heavy-duty trucks (see also Goedhart et al. (2016) and Goedhart et al. (2020)). This example considers the bolts as fasteners at engines, and the applied torque on them. The measurements are performed by a process engineer for process monitoring. The bolts are tightened using a very specific procedure, during which the torque is measured at several moments (in Newton-meters). It is of major importance to monitor these torque values, and substantial changes in these values can indicate that the wrenches need to be recalibrated. For example, the performance of the used wrenches can deteriorate over time, which can result in fasteners being too tight or too loose. It is thus of major importance to detect such out-of-control situations.

To illustrate the use of the updating procedures, we use a data set of 340 observations. The first $m = 102$ observations (in 6 batches of 17 measurements per engine) are used to determine the Phase I estimates. The set of 238 remaining observations (in 14 batches of 17 measurements per engine) are monitored in Phase II. An overview of the 340 observations is given in Fig. 3. Note that a shift appears shortly after measurement 200, as multiple measurements greatly exceed the maximum values detected in the first 200 observations. We set the desired $ARL = 200$ and minimum important deviation $\delta = 0.2$.

The resulting Shewhart control chart using updating parameters $A = 1.5$ and $B = 50$, in line with Capizzi and Masarotto (2020), is depicted in Fig. 4. As described in the previous section, these settings are quite conservative which results in no updates during the 238 observations in Phase II. As can be observed, using these settings leads to an out-of-control signal at the 255th observation.

The Shewhart control chart using the updating rules of Section 5 is presented in Fig. 5. For $m = 102, ARL = 200$ and $\delta = 0.2$ these rules result in updating values $A = 2$ and $B = 30.4$, which trigger multiple updates of the control limits. In this case, this also results in earlier detection of the out-of-control situation (i.e. more than two batches earlier than the 220th observation).

7. Conclusion

In this paper, we investigated the cautious parameter updating approach of Capizzi and Masarotto (2020). Parameter estimation is an issue when determining control limits for the Shewhart, EWMA, and CUSUM control charts, and can have a substantial impact on the control chart performance. One approach to dealing with the estimation error is to update the parameter estimates during Phase II.

We evaluated the cautious updating approach of Capizzi and Masarotto (2020) and propose adjustments to their procedure. An approximation of the expected time to the first parameter update shows that choosing the appropriate updating parameters is important to prevent incorporating contaminated samples in the parameter estimates. We have shown that the average run lengths are a result of the mean deviation δ, the number of Phase I samples m, and the updating parameters A and B. To ensure optimal Phase II performance, formulas were developed for A and B given the available Phase I data and the value of δ that is important to the practitioner. Using these formulas delivers promising control chart performance.

In a case study using data from a multinational truck manufacturer, we demonstrated the added value of updating the control limits for torque measurements. The updating procedure works especially well when using Eqs. (16)–(19) as rules for updating according to Inequality (9).

Updating control chart limits is a logical step towards reducing parameter estimation uncertainty. However, updating using contaminated samples can cause the estimates to spiral out of control. The methods described in this paper greatly reduce the probability of updating using contaminated samples, while still benefitting from the improved estimation accuracy when possible.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the reviewers and the editor for detailed comments and suggestions that have led to various improvements.

Appendix A

A.1. Expectation - Unconditional

In this section, we consider the unconditional expectation of Inequality (9). For the left-hand side of Inequality (9), it is possible to determine the expectation of an individual term in the sum. First, note that

$$x_j \sim N(\mu + \delta\sigma, \sigma^2),$$
$$\overline{x}_{i-d_i} \sim N(\mu, \sigma^2/m),$$
$$\frac{(m-1)s^2_{i-d_i}}{\sigma} \sim \chi^2_{m-1}.$$

Since x_j and $\overline{x}_{i-d_i}$ are independent, we also know that $x_j - \overline{x}_{i-d_i} \sim N(\delta\sigma, \sigma^2(1+1/m))$. Denote $Y = \frac{x_j - \overline{x}_{i-d_i}}{s_{i-d_i}}$. We can then rewrite this into

$$\begin{aligned} Y &= \frac{x_j - \overline{x}_{i-d_i}}{s_{i-d_i}} \\ &= \sqrt{1+1/m}\,\frac{(x_j - \overline{x}_{i-d_i} - \delta\sigma)\Big/\left(\sigma\sqrt{1+1/m}\right) + \delta\sigma\Big/\left(\sigma\sqrt{1+1/m}\right)}{s_{i-d_i}/\sigma} \\ &= \sqrt{1+1/m}\,\frac{Z + \delta\Big/\sqrt{1+1/m}}{\sqrt{V/\nu}} \end{aligned} \tag{20}$$

where $Z = \frac{x_j - \overline{x}_{i-d_i} - \delta\sigma}{\sigma\sqrt{1+1/m}}$ is a standard normal variable, and $V = \frac{(m-1)s^2_{i-d_i}}{\sigma^2}$ is a chi-squared variable with $\nu = m-1$ degrees of freedom. Next, note that

$$T = \frac{Z + \delta\Big/\sqrt{1+1/m}}{\sqrt{V/\nu}}$$

follows a noncentral *t*-distribution with $\nu = m-1$ degrees of freedom and noncentrality parameter $\gamma = \delta/\sqrt{1+1/m}$. Consequently, $F = T^2$ follows a noncentral *F*-distribution with $\nu_1 = 1$ numerator degrees of freedom, $\nu_2 = \nu = m-1$ denominator degrees of freedom, and noncentrality parameter $\lambda = \gamma^2 = \delta^2(1+1/m)$.

To get back to (9), for $m > 3$ the expectation of an individual term in the sum on the left-hand side of the inequality can be calculated to be

$$\begin{aligned} E\left[\left(\frac{x_j - \overline{x}_{i-d_i}}{s_{i-d_i}}\right)^2\right] &= E\left[Y^2\right] = \left(1+\frac{1}{m}\right)E\left[F\right] \\ &= \left(1+\frac{1}{m}\right)\frac{(m-1)\left(1+\delta^2\frac{m}{m+1}\right)}{m-3} \\ &= \left(\frac{m-1}{m-3}\right)\left(1+\delta^2+\frac{1}{m}\right). \end{aligned} \tag{21}$$

A.2. B. Expectation of Sum - Conditional

Consider the conditional expectation of the sum on the left-hand side of Inequality (9). We only consider the time until the first update, such that $d_i = i$ until the update is done, so that $\overline{x}_{i-d_i} = \overline{x}_0$ and $s_{i-d_i} = s_0$, and such that Inequality (9) becomes

$$\sum_{j=1}^{i}\left(\frac{x_j - \overline{x}_0}{s_0}\right)^2 < Ai - B. \tag{22}$$

Consider $Y_j = \frac{x_j - \overline{x}_0}{s_0}$. Conditional on $\overline{x}_0$ and s_0, we know that

$$Y_j|\overline{x}_0, s_0 \sim N\left(\frac{\mu - \overline{x}_0}{s_0} + \delta\frac{\sigma}{s_0}, \frac{\sigma^2}{s_0^2}\right),$$

or equivalently

$$\frac{s_0}{\sigma}Y_j|\overline{x}_0, s_0 \sim N\left(\frac{\mu - \overline{x}_0}{\sigma} + \delta, 1\right).$$

We then rewrite the left-hand side of Inequality (22) into

$$\sum_{j=1}^{i}\left(\frac{x_j-\overline{x}_0}{s_0}\right)^2=\sum_{j=1}^{i}Y_j^2 \\ =\frac{\sigma^2}{s_0^2}C_i \tag{23}$$

where $C_i=\sum_{j=1}^{i}\left(\frac{s_0}{\sigma}Y_j\right)^2$. Note that $C_i|\overline{x}_0,s_0$ follows a noncentral chi-square distribution with i degrees of freedom and noncentrality parameter $i\left(\frac{\mu-\overline{x}_0}{\sigma}+\delta\right)^2$. From this, we calculate the expectation of the left-hand side of Inequality (11), conditional on $\overline{x}_0$ and s_0, to be

$$=E\left[\frac{\sigma^2}{s_0^2}C_i|\overline{x}_0,s_0\right] \\ =i\left(1+\left(\frac{\mu-\overline{x}_0}{\sigma}+\delta\right)^2\right)\frac{\sigma^2}{s_0^2} \tag{24}$$

References

Capizzi, G., & Masarotto, G. (2020). Guaranteed in-control control chart performance with cautious parameter learning. *Journal of Quality Technology*. https://doi.org/10.1080/00224065.2019.1640096

Vera do Carmo, C., Lopes, L. F. D., & Souza, A. M. (2004). Comparative study of the performance of the CuSum and EWMA control charts. *Computers & Industrial Engineering, 46*, 707–724.

Diko, M. D., Chakraborti, S., & Does, R. J. M. M. (2019). An alternative design of the two-sided CUSUM chart for monitoring the mean when parameters are estimated. *Computers & Industrial Engineering, 137*, 1–14.

Gandy, A., & Kvaløy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics, 40*, 647–668.

Goedhart, R., Da Silva, M. M., Schoonhoven, M., Epprecht, E. K., Chakraborti, S., Does, R. J. M. M., & Veiga, A. (2017a). Shewhart control charts for dispersion adjusted for parameter estimation. *IISE Transactions, 49*, 838–848.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2016). Correction factors for Shewhart X and $\overline{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research, 54*, 7464–7479.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2017b). Guaranteed in-control performance for the Shewhart X and $\overline{X}$ control charts. *Journal of Quality Technology, 49*, 155–171.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2020). Nonparametric control of the conditional performance in statistical process monitoring. *Journal of Quality Technology, 52*, 355–369.

Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale. *The Statistician, 36*, 299–316.

Hawkins, D. M., & Wu, Q. (2014). The CUSUM and the EWMA head-to-head. *Quality Engineering, 26*, 215–222.

Huberts, L. C. E., Schoonhoven, M., & Does, R. J. M. M. (2019). The effect of continuously updating control chart limits on control chart performance. *Quality and Reliability Engineering International, 35*, 1117–1128. https://doi.org/10.1002/qre.2448

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology, 38*, 349–364.

Jones, M. A., & Steiner, S. H. (2012). Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *International Journal for Quality in Health Care, 24*, 176–181.

Li, C., Mukherjee, A., & Su, Q. (2019). A distribution-free Phase I monitoring scheme for subgroup location and scale based on the multi-sample Lepage statistic. *Computers & Industrial Engineering, 129*, 259–273. https://doi.org/10.1016/j.cie.2019.01.013

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control* (7th ed.). Hoboken, NJ: John Wiley & Sons.

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

Psarakis, S., Vyniou, A. K., & Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International, 30*, 1113–1129.

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

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

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics, 1*, 239–250.

Saleh, N. A., Mahmoud, M. A., Keefe, M. J., & Woodall, W. H. (2015). The difficulty in designing Shewhart $\overline{X}$ and X control charts with estimated parameters. *Journal of Quality Technology, 47*, 127–138.

Saleh, N. A., Zwetsloot, I. M., Mahmoud, M. A., & Woodall, W. H. (2016). CUSUM charts with controlled conditional performance under estimated parameters. *Quality Engineering, 28*, 402–415.

Shewhart, W. A. (1926). Quality control charts. *Bell System Technical Journal, 5*, 593–603.

Testik, M. C., Kara, O., & Knoth, S. (2020). An algorithmic approach to outlier detection and parameter estimation in Phase I for designing Phase II EWMA control chart. *Computers & Industrial Engineering, 144*, 106440.

Zwetsloot, I. M., & Ajadi, J. O. (2019). A comparison of EWMA control charts for dispersion based on estimated parameters. *Computers & Industrial Engineering, 127*, 436–450.