Quality Control

Online Control Charts for Process Averages Based on Repeated Median Filters

ABHIJIT GUPTA AND SUKALYAN SENGUPTA

SQC & OR Unit, Kolkata, Indian Statistical Institute, Kolkata, India

Two types of estimates of process level, namely repeated median estimates (Siegel, 1982) and full online estimates (Gather et al., 2006) based on repeated median filters, are used to develop control charts. The distributional properties of the estimates are studied using simulation and these are found to closely follow normal distribution. The repeated median being robust against outliers with asymptotically 50% breakdown value and having small standard deviation is found to be useful as a basis for monitoring process averages. The control charts using repeated median estimates have been recommended for general use.

Keywords Control chart; Empirical distribution; Full online estimate; Normal approximation; Repeated median estimate.

Mathematics Subject Classification 62P30.

1. Introduction

The repeated median (RM) being a robust statistic has been used successfully for online signal extraction (Davies et al., 2004). In this case, the data points are processed by a moving window of length n $(=2k+1)$ using repeated median (Siegel, 1982) as a filter to approximate the signal (μ_t) underlying an observation (x_t) by $\tilde{\mu}_t$ (repeated median of the window). This estimated signal has a time lag of length k. The repeated median was the first regression estimator to achieve a breakdown point of 50% asymptotically.

In order to overcome the time lag, Gather et al. (2006) later proposed a full online estimate of the latest observation of a window based on the repeated median filters. Thus, starting from a set of n points forming a window, a sequence of update steps is performed to move the window. At each step the top-most point of the window is deleted and one new point is inserted at the bottom of the existing window to get the successive windows.

Address correspondence to Abhijit Gupta, SQC & OR Unit, Kolkata, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India; E-mail: agupta@isical.ac.in

The repeated median gives a good protection against outliers and shows moderate variability and therefore is a very good estimator of the signal.

In the situations where control charts are used to monitor a process average, the most important objectives are: (i) to identify any trend in the process; and (ii) to detect process shifts early so that the process can be quickly reset. It is also desirable that the outliers should not unduly affect the control charts. Different types of control charts (Montgomery, 2005) are used in the shop floors for this purpose. In the case of Shewhart Control Chart, the monitoring is done on the basis of observed averages of subgroups taken at different time intervals. The windows considered here are similar to such subgroups. In this article, we propose two types of control charts based on repeated median filters: (i) control chart using repeated medians (to be called RM Control Chart); and (ii) control chart using full online estimate based on repeated median filters (to be called Full Online Control Chart). The first one uses the repeated median of a window as the control variable, whereas the second one uses the estimate of the latest observed value of a window for control.

In this article, we propose the repeated median filter-based control charts because of the following:

(i) the charts will not be unduly affected by outliers;
(ii) the estimates of the process level (signal) have small variances as we have seen by studying their distribution by simulation; and
(iii) these can be installed on-line as fast algorithms are available (Bernholt and Fried, 2003; Fried et al., 2006).

Following Fried et al. (2006), we define the repeated medians of a window $(x_{t-k}, \ldots, x_{t+k})$, where $-k, \ldots, k$ are the time periods, as

$$\tilde{\mu}_t = med\left(x_{t-k} + k\tilde{\beta}_t, \ldots, x_{t+k} - k\tilde{\beta}_t\right)$$
$$\tilde{\beta}_t = \underset{i=-k,\ldots,+k}{med}\left(\underset{j=-k,\ldots,+k,\, j\neq i}{med} \frac{x_{t+i} - x_{t+j}}{i-j}\right). \tag{1}$$

The Full Online estimate of the latest value of a window, i.e., x_{t+k} is as follows;

$$\tilde{\mu}_{t+k} = \tilde{\mu}_t + k\tilde{\beta}_t. \tag{2}$$

We will do the following in this article:

(a) Study the distributions of both the estimates using simulations, assuming each x_i to be a standard normal variate. We will also study the effect of non normality of x_i for some chosen non normal distributions. The effect of contamination by outliers on these distributions will be studied as well.
(b) Based on these the distributions, obtain the control limits at different levels of significance.
(c) Study the performance of the control charts by obtaining the corresponding average run lengths (ARL) required to detect process shifts of different magnitudes.

2. Empirical Distribution of the RMs

We have studied the distribution of the RMs for different values of k from 2–20 on the basis of 50,000 simulated observations from $N(0, 1)$ for each k. The simulation and analysis were carried out using Matlab 7.0.1 (2004).

It may be pointed out that we are interested in small values of k since larger values would lead to longer delay in initiating control on the process.

The results of the analysis are given in Table 1a. In this table, the lower and upper ordinates covering 95, 99, and 99.73% area (equal tail) of the obtained distributions are denoted as 95% ordinates, 99% ordinates, and 99.73% ordinates, respectively. Lilliefors test (Matlab 7.0.1, 2004) was used to check the normality of the distributions.

As an illustration, we also give in Table 1b the transformed values of the ordinates by standardizing the obtained normal variables for each k to communicate the degree of agreement with the standard normal distribution. The 95, 99, and 99.73% ordinates of $N(0, 1)$ are ± 1.96, ± 2.57, and ± 3.00, respectively.

It is thus concluded from the above analysis that for all the values of k considered, the RMs closely follow normal distributions with standard deviations (SD) much smaller than those of the underlying distribution. It is also observed that the SD decreases with an increase in k.

We have tried to fit a relationship between the SD and k for values of k lying between 2 and 20. Two relationships, namely, (a) logarithmic relationship and (b) root inverse of k were considered. The root inverse relationship was found to be a better fit and was chosen as the model. The relationship obtained is as follows:

$$\mathrm{SD}(k) = 0.0409 + 0.7313\frac{1}{\sqrt{k}} \tag{3}$$

where $\mathrm{SD}(k)$ denotes the value of the standard deviation for a particular value of k.

The values of r^2, adjusted r^2, and standard error for this model are 0.9956, 0.9951, and 0.0076, respectively. The curves showing the fitted values and observed values of SD against k are given in Fig. 1.

We can, therefore, conveniently approximate the distributions of RM estimates by $N(0, \mathrm{SD}(k))$ for the range of the values of k considered. Siegel (1982) also mentioned in his article that under suitable conditions, RM estimates are unbiased and follow normal distribution.

2.1. *Effect of Non Normality*

We have looked into the effect of non normality of x_i on the distribution of RM to judge the robustness of RM estimates. The non normal distributions chosen were:

- Weibull with scale parameter 1 and shape parameter 2, as a moderately skewed distribution, and
- Lognormal with shape parameter 0.75 as a long tailed distribution.

These distributions are often encountered in the area of quality control. The pdfs of these distributions are shown in Fig. 2.

The normal probability plots of the RMs for $k = 2$ to 8, based on 200 simulations, are given in Fig. 3. These plots, in both the cases, exhibit fair degree of normality, although a slight departure in the tails is observed for smaller values of k.

Table 1a
Distribution of RM based on simulation for different values of k

			95% Ordinates		99% Ordinates		99.73% Ordinates		Lilliefors statistic	
k	Mean	SD	Lower	Upper	Lower	Upper	Lower	Upper	Calculated	Critical value
2	0.0051	0.5461	−1.0667	1.0837	−1.4185	1.4188	−1.6377	1.6794	0.0035	
3	−0.0030	0.4665	−0.9175	0.9085	−1.2091	1.2141	−1.4244	1.4114	0.0020	
4	−0.0021	0.4124	−0.8111	0.8101	−1.0692	1.0754	−1.2520	1.2497	0.0035	
5	−0.0066	0.3724	−0.7413	0.7193	−0.9790	0.9382	−1.1347	1.0926	0.0037	
6	0.0059	0.3416	−0.6693	0.6784	−0.8722	0.8909	−1.0095	1.0536	0.0039	
7	0.0016	0.3251	−0.6412	0.6390	−0.8515	0.8411	−0.9860	0.9807	0.0036	0.0040
8	0.0089	0.3058	−0.5858	0.6113	−0.7786	0.8093	−0.9116	0.9382	0.0034	
10	−0.0018	0.2714	−0.5403	0.5265	−0.7182	0.7015	−0.8616	0.8038	0.0036	
15	0.0054	0.2197	−0.4305	0.4280	−0.5628	0.5403	−0.6690	0.6104	0.0065	
20	0.0008	0.1974	−0.3919	0.3860	−0.5171	0.5135	−0.5795	0.5899	0.0041	

Table 1b
Standardized ordinates

	95% Ordinates		99% Ordinates		99.73% Ordinates	
k	Lower	Upper	Lower	Upper	Lower	Upper
2	−1.9626	1.9751	−2.6068	2.5887	−3.0082	3.0659
3	−1.9603	1.9539	−2.5854	2.6090	−3.0469	3.0319
4	−1.9617	1.9694	−2.5875	2.6128	−3.0308	3.0354
5	−1.9729	1.9492	−2.6114	2.5371	−3.0293	2.9517
6	−1.9766	1.9687	−2.5706	2.5907	−2.9725	3.0670
7	−1.9772	1.9606	−2.6241	2.5823	−3.0378	3.0117
8	−1.9747	1.9699	−2.5752	2.6174	−3.0101	3.0389
10	−1.9842	1.9466	−2.6396	2.5914	−3.1680	2.9683
15	−1.9841	1.9235	−2.5863	2.4347	−3.0696	2.7538
20	−1.9894	1.9514	−2.6236	2.5973	−2.9397	2.9843

2.2. *The Proposed Control Charts*

The control limits obtained here are for a $N(0, 1)$ variate. In practice, these limits will have to be suitably modified for a specific $N(\mu, \sigma)$. The values of μ and σ are to be established using data collected from the process. One must be careful to base the estimation using a sufficiently large data set after proper cleaning of the data and ensuring the constancy of mean and SD during the period of data collection. We will ignore the amount of uncertainty due to use of these estimates while developing the control charts.

2.2.1. Based on Empirical Distribution. The central line of the control charts will be taken as zero and the percentile points given in Table 1 can be used as the control

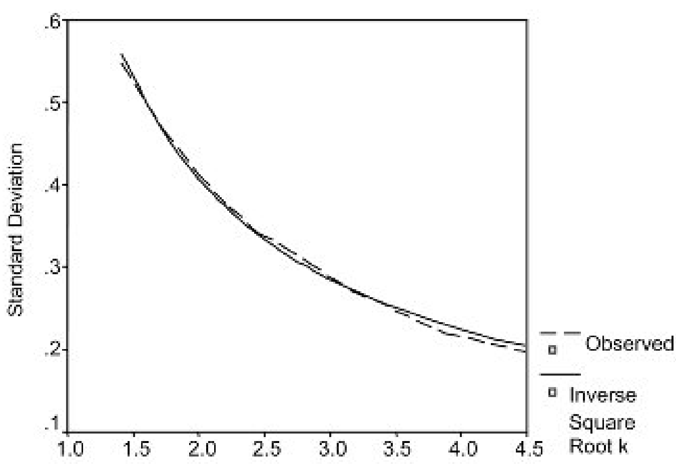


Figure 1. Curve showing fitted $\left[\frac{1}{\sqrt{k}}\right]$ vs. observed values of standard deviation.

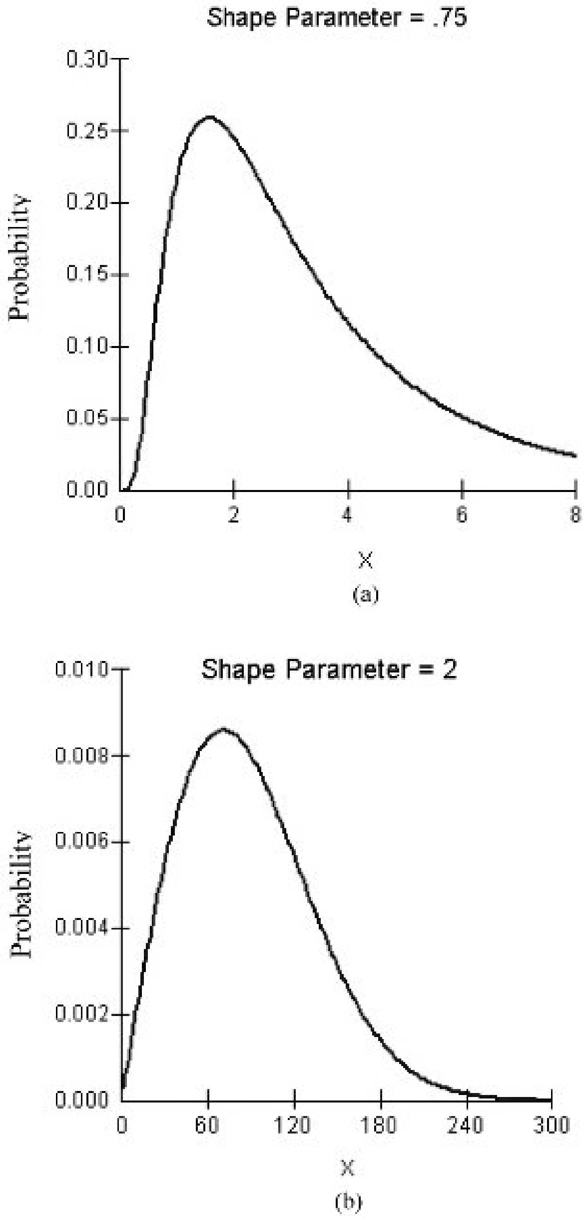
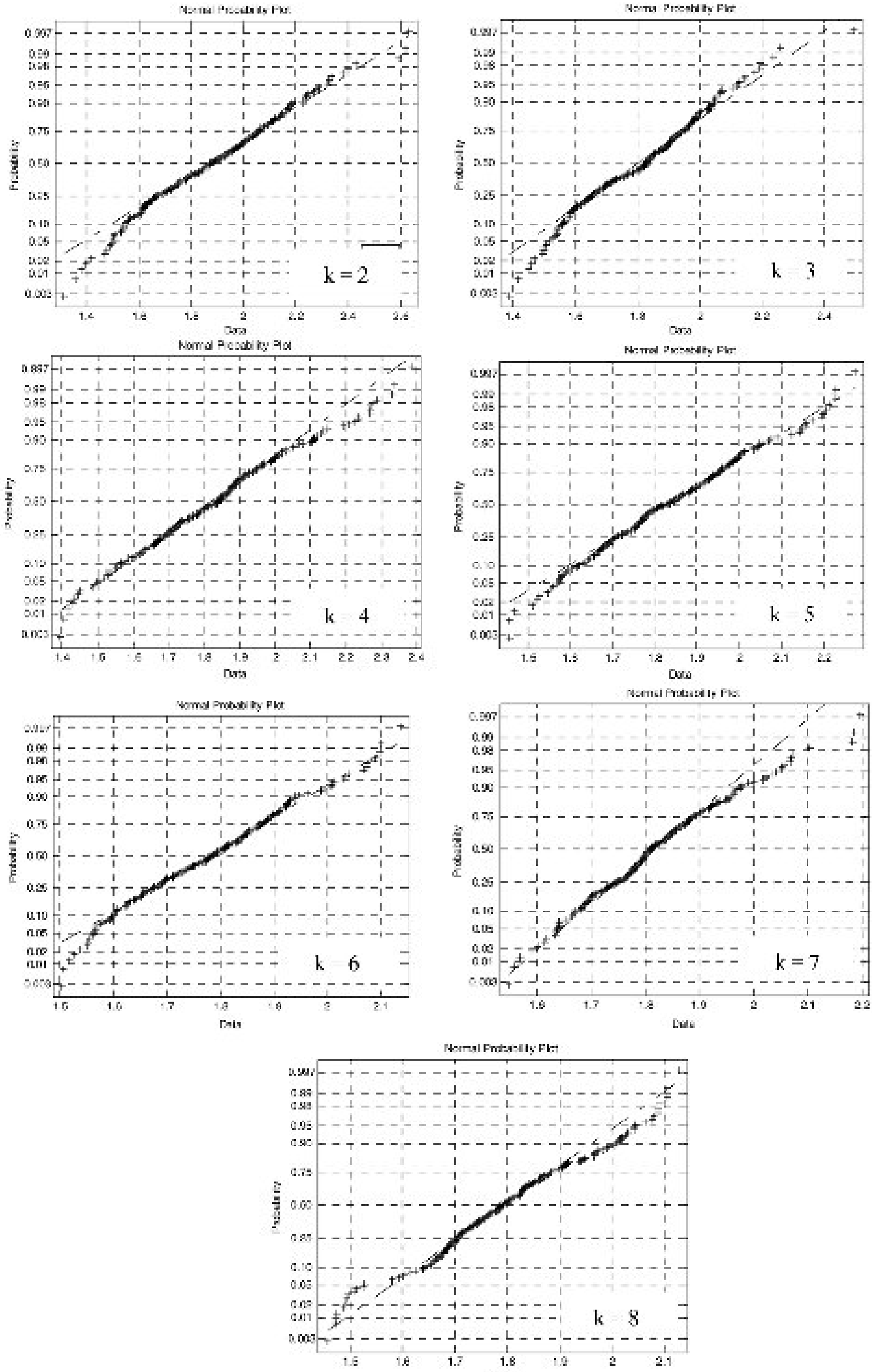
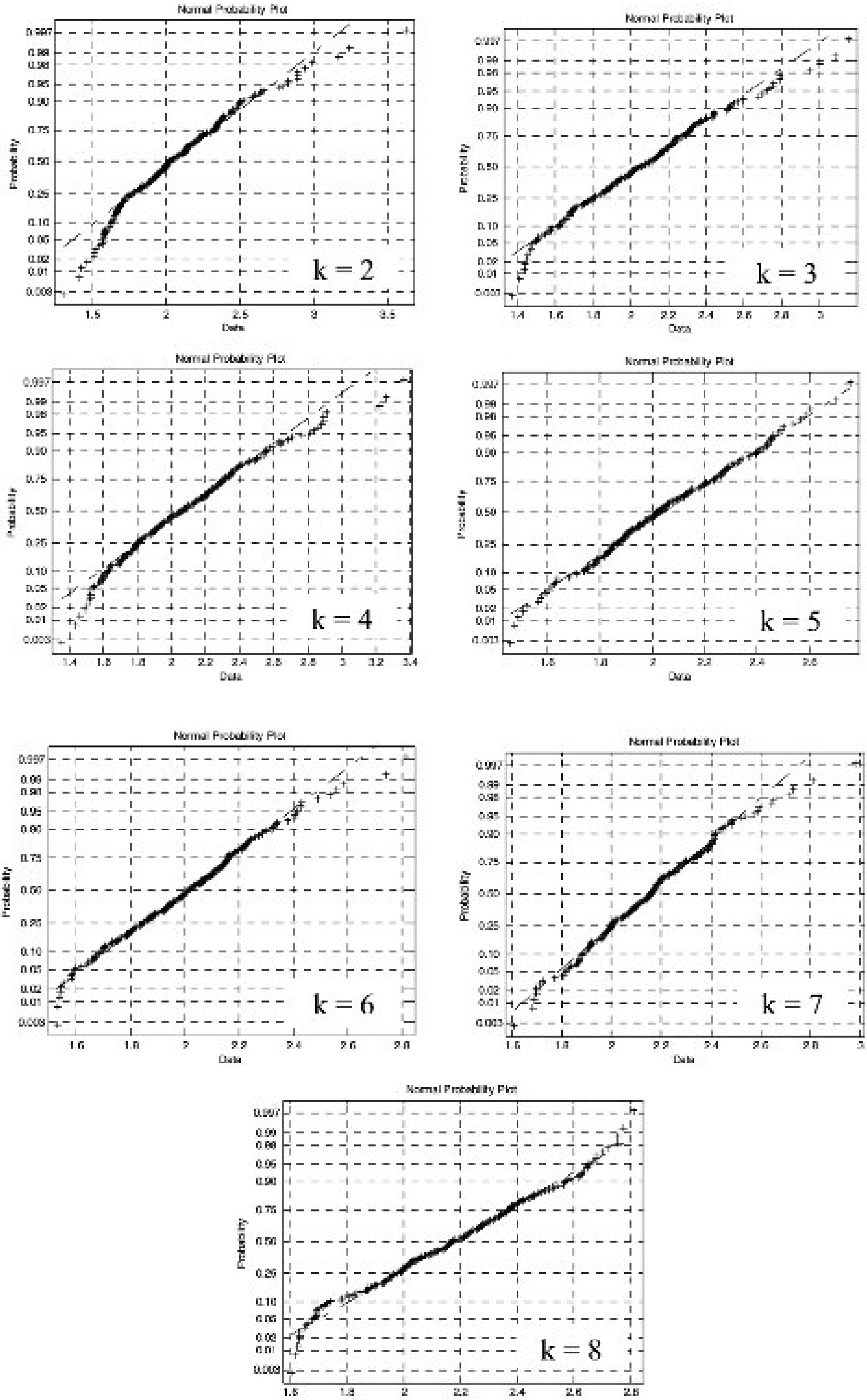


Figure 2. Probability density functions. (a) Lognormal distribution; (b) Weibull distribution.



(a)

Figure 3a. Normal probability plots under Weibull distribution.



(b)

Figure 3b. Normal probability plots under Lognormal Distribution for RM estimates.

Table 2
Control limits and corresponding Type I error (in %) (based on RM estimate)

	95%		99%		99.73%	
k	Control limits	Error	Control limits	Error	Control limits	Error
2	−1.067, 1.084	5.046	−1.418, 1.419	1.134	−1.638, 1.679	0.328
3	−0.918, 0.908	4.940	−1.209, 1.214	0.980	−1.424, 1.411	0.310
4	−0.811, 0.810	5.266	−1.069, 1.075	1.032	−1.252, 1.250	0.270
5	−0.741, 0.719	5.134	−0.979, 0.938	1.094	−1.135, 1.093	0.282
6	−0.669, 0.678	5.280	−0.872, 0.891	1.232	−1.010, 1.054	0.370
7	−0.641, 0.639	4.678	−0.852, 0.841	0.854	−0.986, 0.981	0.202
8	−0.586, 0.611	4.796	−0.779, 0.809	0.786	−0.912, 0.938	0.200

limits. The Shewhart Control Chart is based on 99.73% limits. The control limits along with the first type of error are given in Table 2.

We have studied the performance of these control charts by calculating the first type of error and the ARLs to detect shifts of specific magnitudes for different values of k up to 8.

In order to calculate ARL we have assumed that there was no shift in the starting window and that the shift has occurred while moving to the subsequent windows. The ARLs are estimated based on 1,000 runs. The estimated ARLs are given in Table 3 for different values of k corresponding to shifts of different magnitudes. The magnitudes of process shifts considered are $\delta \times \sigma$, where $\delta = 0, 0.5, 1.0, 1.5$ and $\sigma = 1$.

Table 3
ARL for the suggested control charts

			Confidence interval		
k	Shift	ARL	95%	99%	99.73%
		I. 95% Control Limit			
2	0	26.76	0–99	0–140	0–160
	0.5	12.12	0–41	0–61	0–78
	1.0	5.76	0–18	0–23	0–34
	1.5	3.74	0–9	0–11	0–13
3	0	32.72	0–126	0–177	0–232
	0.5	13.76	0–48	0–67	0–80
	1.0	6.42	0–19	0–27	0–28
	1.5	4.07	0–8	0–12	0–15
4	0	37.90	0–137	0–185	0–225
	0.5	14.75	0–47	0–76	0–88
	1.0	6.70	0–16	0–22	0–24
	1.5	4.79	0–9	0–11	0–13

(continued)

Table 3
Continued

			Confidence interval		
k	Shift	ARL	95%	99%	99.73%
5	0	43.34	0–165	0–264	0–357
	0.5	15.76	0–51	0–69	0–88
	1.0	7.22	0–17	0–22	0–24
	1.5	5.08	0–10	0–11	0–14
6	0	50.39	0–181	0–249	0–326
	0.5	16.61	0–56	0–77	0–89
	1.0	7.84	0–18	0–23	0–24
	1.5	5.64	0–11	0–13	0–19
7	0	61.72	0–229	0–334	0–445
	0.5	17.74	0–52	0–68	0–82
	1.0	8.48	0–18	0–26	0–29
	1.5	6.30	0–12	0–14	0–19
8	0	67.84	0–262	0–358	0–535
	0.5	18.73	0–58	0–79	0–100
	1.0	9.19	0–19	0–24	0–29
	1.5	6.57	0–13	0–15	0–16
			II. 99% Control limit		
2	0	121.91	2–397	0–551	0–679
	0.5	33.91	2–118	0–117	0–221
	1.0	10.53	1–33	0–46	0–49
	1.5	5.33	1–13	0–18	0–21
3	0	149.63	4–520	0–704	0–877
	0.5	35.45	2–118	0–163	0–220
	1.0	10.68	1–29	0–43	0–55
	1.5	5.96	2–12	0–18	0–22
4	0	181.14	5–664	0–833	0–922
	0.5	35.90	3–122	0–167	0–175
	1.0	10.71	2–28	0–38	0–42
	1.5	6.54	2–13	0–17	0–24
5	0	184.05	3–699	0–900	0–982
	0.5	31.46	2–98	0–144	0–162
	1.0	10.57	1–28	0–39	0–42
	1.5	6.62	1–12	0–14	0–15
6	0	197.28	3–650	0–877	0–969
	0.5	32.86	3–100	0–126	0–153
	1.0	11.70	1–29	0–41	0–44
	1.5	7.52	2–13	0–15	0–20

(continued)

Table 3
Continued

k	Shift	ARL	Confidence interval 95%	99%	99.73%
7	0	240.03	8–818	1–939	0–969
	0.5	36.52	4–124	0–171	0–199
	1.0	12.70	2–28	0–42	0–54
	1.5	8.38	2–14	0–16	0–19
8	0	253.47	4–890	0–966	0–995
	0.5	39.94	5–133	0–172	0–190
	1.0	13.63	3–29	0–40	0–45
	1.5	8.90	2–15	1–17	0–19
		III. 99.73% Control limit			
2	0	286.20	0–891	0–959	0–973
	0.5	88.05	5–309	1–467	0–503
	1.0	18.96	2–62	1–86	1–96
	1.5	7.48	2–22	1–27	1–31
3	0	343.18	2–934	0–985	0–994
	0.5	73.57	6–263	2–400	0–462
	1.0	15.29	4–46	1–64	0–75
	1.5	7.50	3–17	2–23	0–25
4	0	353.44	12–945	3–981	0–994
	0.5	69.51	5–238	1–330	0–408
	1.0	15.79	4–45	1–69	0–76
	1.5	7.71	3–16	1–20	0–26
5	0	340.11	9–926	0–994	0–998
	0.5	56.96	5–206	0–268	0–350
	1.0	14.40	4–41	1–53	0–57
	1.5	7.92	3–15	2–19	1–25
6	0	358.44	5–920	0–994	0–997
	0.5	66.01	7–231	2–293	0–314
	1.0	15.59	5–40	2–51	1–75
	1.5	9.23	4–16	1–22	0–28
7	0	391.13	10–956	1–988	0–997
	0.5	62.95	7–196	2–286	1–394
	1.0	16.41	5–41	2–56	0–80
	1.5	9.74	4–16	2–19	0–21
8	0	413.42	12–958	2–992	0–997
	0.5	65.84	9–212	2–291	0–374
	1.0	17.28	7–42	1–55	0–62
	1.5	10.49	4–17	2–19	0–22

Table 4
Type I error (in %) under normality assumption

	Type I error (%) corresponding to		
k	95% Control limit	99% Control limit	99.73% Control limit
2	6.292	1.540	0.514
3	5.482	1.232	0.380
4	4.812	0.994	0.286
5	4.506	0.866	0.218
6	4.194	0.744	0.166
7	4.674	0.864	0.258
8	4.604	0.856	0.210

Looking into the performance of the control charts, we recommend the values of $k = 2$ to 6 for shop floor use. In the case of $\overline{X} - R$ chart, subgroup size is generally chosen within 4 to 6, whereas for $\overline{X} - s$ chart, a subgroup size of at least 10 is considered to be appropriate. A delay of 2 to 6 observations, in case of RM chart, will not be unacceptable for the shop-floor use. However, the user may choose a suitable combination of k and control limits on the basis of ARL values acceptable to him.

2.2.2. *Based on Normality Assumption.* The control limits based on the normal distribution $N(0, \mathrm{SD}(k))$ will be:

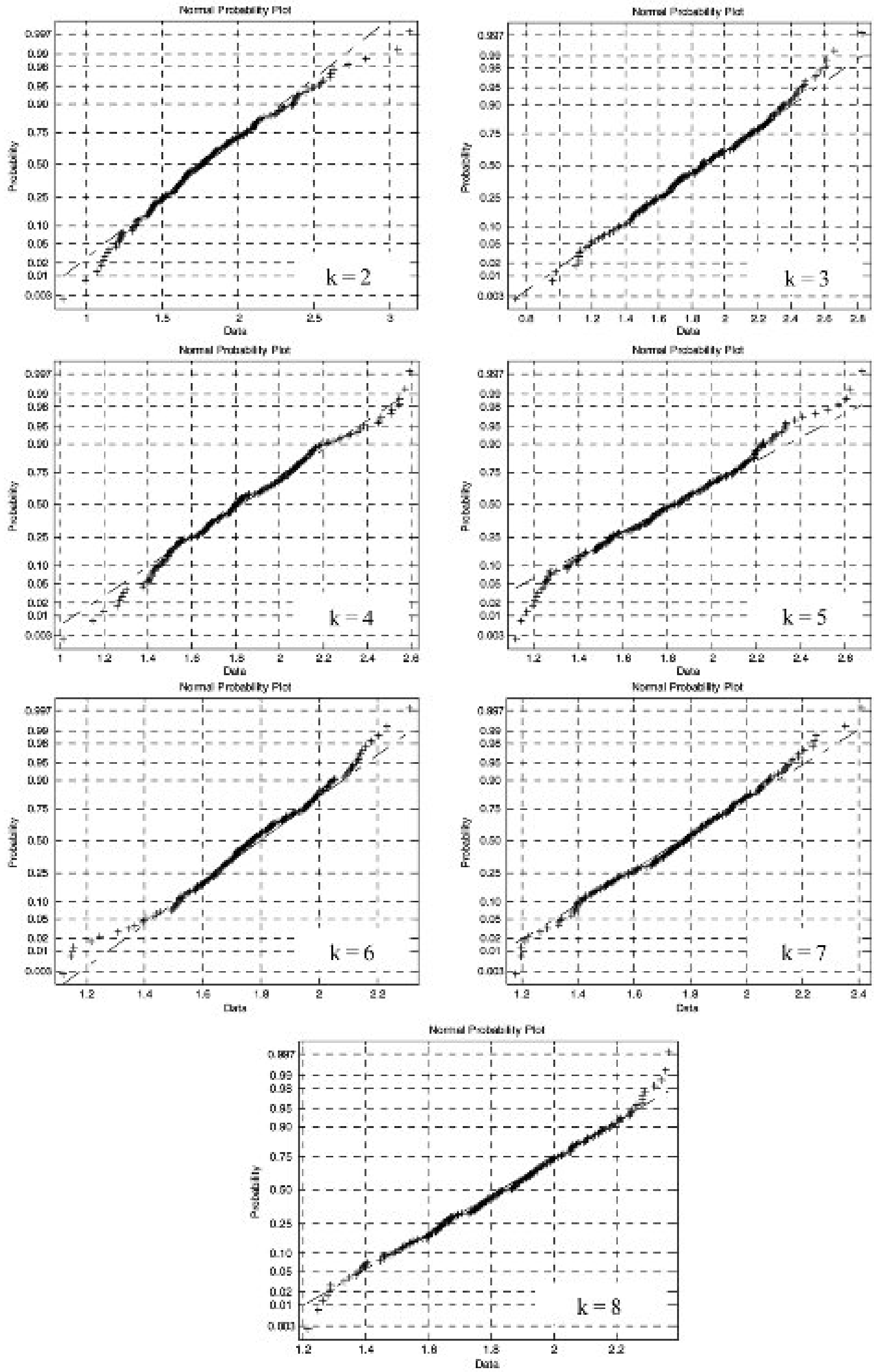
$$\begin{aligned} 95\% \text{ control limits } &= 0 \pm 1.96 \times \mathrm{SD}(k) \\ 99\% \text{ control limits } &= 0 \pm 2.575 \times \mathrm{SD}(k) \\ 99.73\% \text{ control limits } &= 0 \pm 3.0 \times \mathrm{SD}(k). \end{aligned}$$

Table 5
ARL for the proposed control charts under normality assumption

	k				
Shift	2	4	6	8	10
			95% Control limit		
0	28.44	37.83	48.66	57.80	76.44
0.5	12.69	14.65	15.72	17.70	18.48
1.0	5.92	6.83	8.01	8.80	9.99
1.5	3.81	4.67	5.65	6.53	7.04
			99% Control limit		
0	130.46	151.99	196.06	237.48	305.80
0.5	36.06	32.38	35.03	36.84	36.04
1.0	10.99	10.59	11.42	12.54	14.02
1.5	5.46	6.32	7.36	8.71	9.59
			99.73% Control limit		
0	444.20	463.18	610.26	770.63	960.63
0.5	83.22	60.65	58.97	55.69	55.84
1.0	18.89	14.92	14.89	15.98	16.81
1.5	7.72	7.64	8.71	10.04	11.30

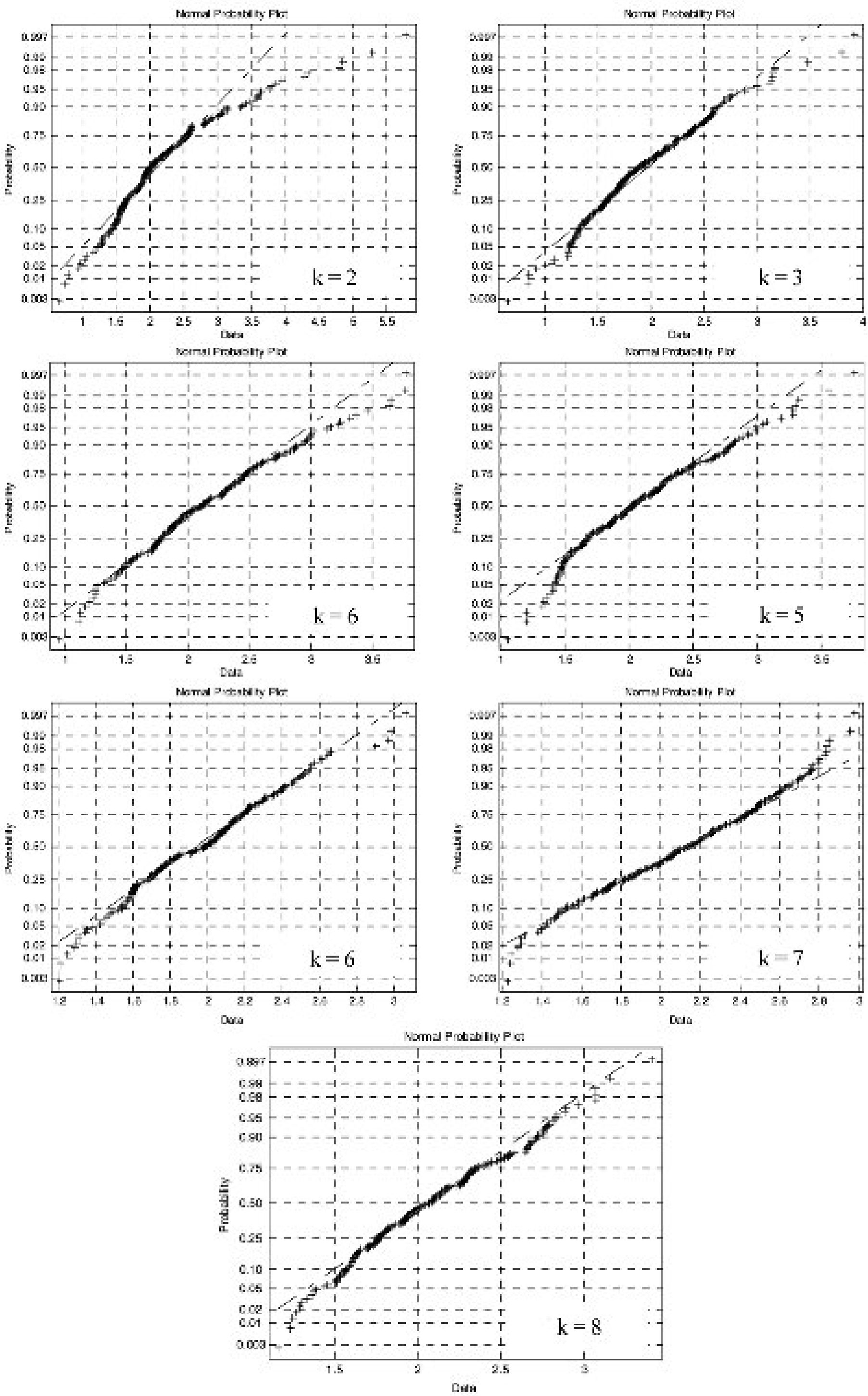
Table 6
Distribution of full online estimates based on simulation for different values of *k*

			95% Ordinates		99% Ordinates		99.73% Ordinates		Lilliefors statistic	
k	Mean	SD	Lower	Upper	Lower	Upper	Lower	Upper	Calculated	Critical value
2	0.0055	0.9213	–1.8198	1.8259	–2.4502	2.4622	–2.8802	2.9713	0.0055	
3	–0.0013	0.8276	–1.6306	1.6311	–2.1688	2.2051	–2.5810	2.6425	0.0057	
4	–0.0080	0.7395	–1.4583	1.4363	–1.9949	1.9111	–2.3489	2.2177	0.0034	
5	0.0023	0.6839	–1.3476	1.3382	–1.8070	1.8022	–2.0872	2.1472	0.0040	
6	–0.0028	0.6263	–1.2293	1.2231	–1.6546	1.6144	–1.9407	1.8885	0.0019	
7	–0.0026	0.5849	–1.1537	1.1407	–1.5467	1.4925	–1.8544	1.7758	0.0035	0.0040
8	0.0017	0.5670	–1.1092	1.1277	–1.4768	1.4954	–1.7231	1.7926	0.0041	
10	0.0078	0.5103	–1.0011	1.0010	–1.3713	1.2898	–1.5420	1.5028	0.0033	
15	0.0099	0.4089	–0.7980	0.8033	–1.0470	1.0488	–1.1882	1.2147	0.0037	
20	0.0053	0.3711	–0.7243	0.7203	–0.9482	0.9707	–1.1021	1.1208	0.0049	



(a)

Figure 4a. Normal probability plots under Weibull distribution.



(b)

Figure 4b. Normal probability plots under Lognormal distribution for full online estimates.

The performance of these control charts has also been studied as in Sec. 2.2.1. Table 4 gives the Type I errors for the control limits for different values of k.

ARL values as obtained in this case are given in Table 5.

It is seen that ARLs obtained for the empirically based control charts and those under normality assumption are of comparable magnitudes.

3. Investigations on Full Online Estimate

Investigations on full online estimates have been carried out in the same manner as for the RMs. The findings are as follows:

(a) It can be seen from Table 6 that the full online estimates also follow normal distributions for different values of k, however, with standard deviations somewhat larger than those of RM estimates.

(b) the standard deviation can be expressed as

$$\mathrm{SD}(k) = 0.1351 + 1.1727 \times \frac{1}{\sqrt{k}} \tag{4}$$

for k lying in the range of values considered with values of r^2, adjusted r^2, and standard error as 0.9816, 0.9793, and 0.0251, respectively. This standard error is much larger than the corresponding standard error of 0.0076 as in the case of RM estimate.

(c) The normal probability plots, based on simulation, to study the effect of non normality are given in Fig. 4.
When the parent distribution is Weibull, the full online estimates are observed to follow normality for all k. However, these estimates do not show adequate normality for $k < 4$ in the case of Lognormal Distribution.

(d) The control limits based on empirical distributions along with Type I error have been given in Table 7.

(e) The ARL values for the control charts based on the empirical distribution of full online estimates have been given in Table 8. It can be seen that these ARL values are larger than those for RM estimates for small shifts, particularly for 99.73% control chart.

Table 7
Control limits and corresponding Type I error (in %) (based on full online estimate)

	95%		99%		99.73%	
k	Control limits	Error	Control limits	Error	Control limits	Error
2	−1.820, 1.826	4.960	−2.450, 2.462	1.004	−2.880, 2.972	0.310
3	−1.631, 1.631	5.188	−2.169, 2.205	1.036	−2.581, 2.642	0.282
4	−1.458, 1.436	5.012	−1.995, 1.911	1.018	−2.349, 2.218	0.332
5	−1.348, 1.338	4.984	−1.807, 1.802	0.976	−2.087, 2.147	0.238
6	−1.229, 1.223	4.940	−1.655, 1.614	0.928	−1.941, 1.888	0.220
7	−1.154, 1.141	5.004	−1.547, 1.492	0.972	−1.854, 1.776	0.248
8	−1.109, 1.128	4.604	−1.477, 1.495	0.894	−1.723, 1.793	0.258

Table 8
ARL for the suggested control charts based on full online estimates

			Confidence interval		
k	Shift	ARL	95%	99%	99.73%
			I. 95% Control limit		
2	0	29.28	0–107	0–145	0–160
	0.5	18.41	0–69	0–99	0–123
	1.0	7.71	0–29	0–36	0–47
	1.5	3.85	0–14	0–18	0–23
3	0	27.70	0–103	0–156	0–200
	0.5	16.48	0–64	0–98	0–114
	1.0	6.93	0–26	0–33	0–39
	1.5	3.61	0–11	0–18	0–21
4	0	29.72	0–111	0–141	0–171
	0.5	15.83	0–59	0–72	0–119
	1.0	6.14	0–21	0–28	0–38
	1.5	3.67	0–10	0–14	0–17
5	0	33.63	0–128	0–225	0–234
	0.5	17.07	0–69	0–85	0–105
	1.0	6.41	0–21	0–28	0–46
	1.5	3.72	0–9	0–13	0–15
6	0	37.47	0–155	0–206	0–232
	0.5	16.38	0–61	0–86	0–110
	1.0	6.64	0–23	0–30	0–35
	1.5	3.95	0–9	0–12	0–14
7	0	39.95	0–156	0–233	0–294
	0.5	16.54	0–62	0–88	0–97
	1.0	6.54	0–20	0–32	0–37
	1.5	4.21	0–9	0–12	0–15
8	0	46.06	0–182	0–233	0–325
	0.5	18.48	0–71	0–95	0–168
	1.0	6.65	0–19	0–31	0–49
	1.5	4.48	0–9	0–13	0–20
			II. 99% Control limit		
2	0	129.80	2–477	0–695	0–729
	0.5	66.44	2–253	0–358	0–406
	1.0	22.83	2–80	0–124	0–129
	1.5	9.17	1–33	0–47	0–64
3	0	114.53	1–406	0–607	0–772
	0.5	53.41	1–206	0–271	0–372
	1.0	17.33	1–69	0–88	0–94
	1.5	6.73	1–24	0–33	0–42

(continued)

Table 8
Continued

k	Shift	ARL	Confidence interval 95%	99%	99.73%
4	0	123.04	3–458	0–641	0–814
	0.5	50.36	2–181	0–262	0–357
	1.0	15.75	1–57	0–91	0–107
	1.5	5.75	1–20	0–30	0–36
5	0	131.56	2–459	0–681	0–953
	0.5	52.21	2–215	0–270	0–333
	1.0	13.80	2–53	0–74	0–98
	1.5	5.69	1–17	0–24	0–30
6	0	163.31	2–646	0–866	0–991
	0.5	50.13	1–187	0–246	0–336
	1.0	13.34	2–48	0–69	0–82
	1.5	5.59	1–14	0–21	0–29
7	0	184.74	2–663	0–873	0–1363
	0.5	50.15	2–174	0–266	0–328
	1.0	12.63	2–43	0–67	0–88
	1.5	6.02	1–14	0–23	0–26
8	0	196.92	3–712	0–986	0–1418
	0.5	50.08	2–176	0–259	0–332
	1.0	12.33	1- 45	0–68	0–83
	1.5	6.28	1–16	0–21	0–26
			III. 99.73% Control limit		
2	0	378.08	7–1510	1–1927	0–2846
	0.5	194.66	5–733	1-1053	1–1192
	1.0	57.83	2–214	1–340	1–370
	1.5	20.32	2–76	1–103	1—136
3	0	331.10	6–1271	0–1902	0–2131
	0.5	136.40	3–517	0–678	0–844
	1.0	37.29	3–137	1–180	0–226
	1.5	12.05	2–49	0–65	0–83
4	0	313.43	8–1107	0–1479	0–2271
	0.5	123.26	3–460	1–576	0–719
	1.0	30.28	2–113	0–161	0–214
	1.5	9.97	2–38	0–51	0–77
5	0	395.82	9–1502	2–1981	1–2208
	0.5	123.51	4–460	1–672	0–782
	1.0	25.69	3–101	1–156	0–193
	1.5	8.13	2–29	0–45	0–50
6	0	476.02	13–1695	0–2284	0–3423
	0.5	116.41	5–472	2–693	0–811
	1.0	24.14	3–88	0–136	0–153
	1.5	8.04	2–25	1–44	0–53

(continued)

Table 8
Continued

k	Shift	ARL	Confidence interval 95%	99%	99.73%
7	0	518.76	9–2060	0–2603	0–3107
	0.5	115.24	3–453	0–595	0–772
	1.0	24.75	3–99	0–124	0–161
	1.5	7.74	2–23	1–37	0–47
8	0	596.19	16–2006	1–3225	0–4003
	0.5	128.01	6–445	3–613	0–701
	1.0	21.71	3–81	1–120	0–133
	1.5	7.58	3–19	1–25	0–33

(f) The control limits using normal approximations are:

$$95\% \text{ control limits } = 0 \pm 1.96 \times \mathrm{SD}(k)$$

$$99\% \text{ control limits } = 0 \pm 2.575 \times \mathrm{SD}(k)$$

$$99.73\% \text{ control limits } = 0 \pm 3.0 \times \mathrm{SD}(k).$$

4. Effect of Outliers on the Normality of RM and Full Online Estimates

Repeated median estimate having high breakdown point, its normality will also remain unaffected from contaminations by outliers. We have studied this aspect by simulating sets of 1,000 observations from $N(0, 1)$. The basis of the simulation is as under:

(a) each window was contaminated by a specific number of outliers;
(b) position of contamination inside a window was chosen at random and the corresponding value of the data set was replaced by outliers, selected at random, falling in the range of 3 to 4. The same was repeated for outliers in the range of 4.5 to 6;
(c) these replaced values were assigned signs (i.e., $+/-$) at random.

Table 9
Effect of outliers on normality of RM and full online estimates—number of outliers up to which normality is maintained

k	Number of outliers: RM estimate	Full online estimates
2	1	0
4	2	1
6	5	3
8	6	4

The normal probability plots for different combinations of the chosen values have shown that the normality is not greatly affected. The result of the analysis is given in Table 9. To illustrate, the plots for $k = 4$ with outliers falling in the range 4.5 to 6 for both the RM and full online estimates are given in Fig. 5. The effect of the contamination on the control charts has also been studied for the example considered in the next section.

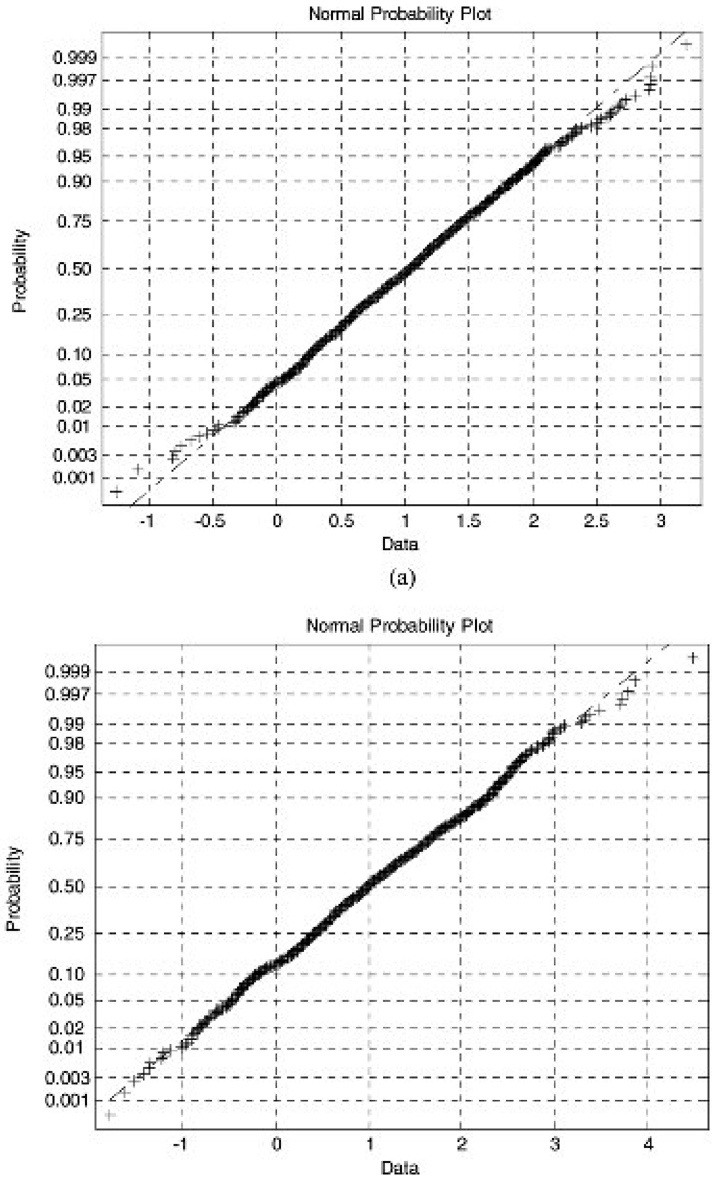


Figure 5. Normal probability plots for $k = 4$ with outliers falling in the range 4.5 to 6. (a) For repeated medians (2 outliers) and (b) for full online estimates (1 outlier).

5. An Example

We use a live data from a pharmaceutical industry to illustrate the RM based control chart. The data relates to weight of 65 uncoated tablets observed sequentially on a particular day between 9:45 AM and 1:12 PM. Plot of the observed values and the corresponding 3σ RM control chart for $k = 4$ are shown in Fig. 6.

An early upward trend is clearly discernable from the RM chart, which would help the process owner to initiate proactive actions. However, the individual observations have failed to bring out such a trend.

The same data was used to construct the 99.73% control chart using the full online estimate. The chart is shown in Fig. 7. It is evident that in this case the trend, which was clearly visible in case RM control chart, is not prominent. The trend

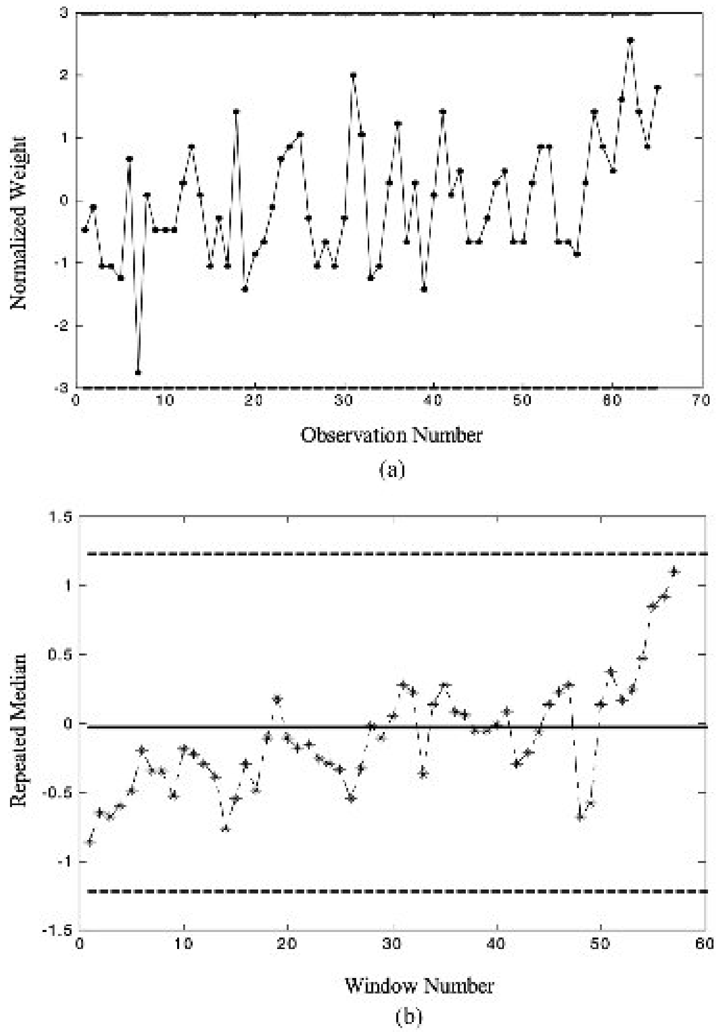


Figure 6. (a) Transformed [$N(0, 1)$] weight of uncoated tablets along with 3σ-control limits and (b) RM control chart for the transformed weight of uncoated tablets.

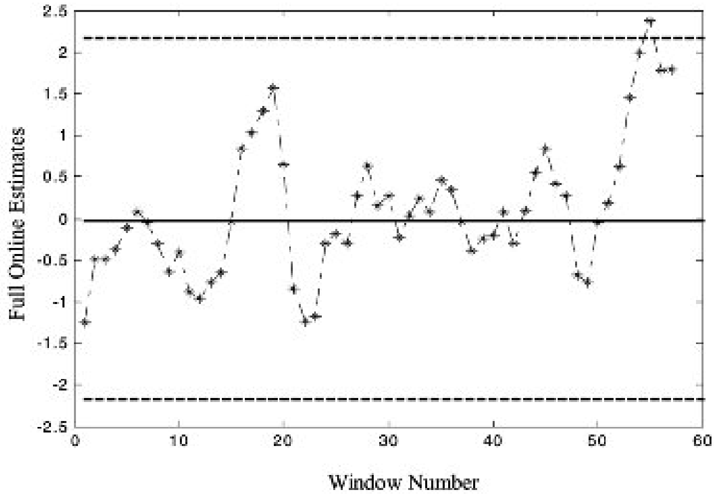


Figure 7. Control chart based on full online estimates for the transformed weight of uncoated tablets.

is seen only towards the end as was also indicated by the graph of the individual observations. It has also shown one out-of-control situation.

To study the effect of outliers on the control charts the observed values were contaminated. Each window with $k = 4$ was contaminated with 1 outlier and then with 2 outliers, falling in the ranges 3 to 4 and 4.5 to 6, respectively, and control charts were drawn. To illustrate the control charts with outliers in the range 4.5 to 6, are shown in Fig. 8 for both the RM and full online estimates. It is seen that the effect of outliers is more pronounced in the case of control chart based on full online estimates when number of outliers is 2.

6. Choice of Control Chart

It is observed from Tables 3 and 8 that for detecting large shifts, full online control chart performs with a somewhat lower values of ARLs for large k in the case of 95% and 99.73% control limits. The in-control ARLs for this chart are also larger than those of RM control chart in the case of 99.73% control limits. However, for detecting smaller shifts, RM control charts perform better having lower ARL values. Thus, the choice will depend on the nature of the shift to be detected.

It is also be noted from the earlier analysis that the RM control charts perform better in detecting trends and the effect of outliers are less pronounced than in the case of full online control chart. Without loosing much in performance we therefore recommend the RM control charts for routine use in the shop floor to monitor the process averages.

7. Conclusion

It is seen that for an underlying normal distribution, the RMs closely follow a normal distribution and are not much affected either by outliers or by non

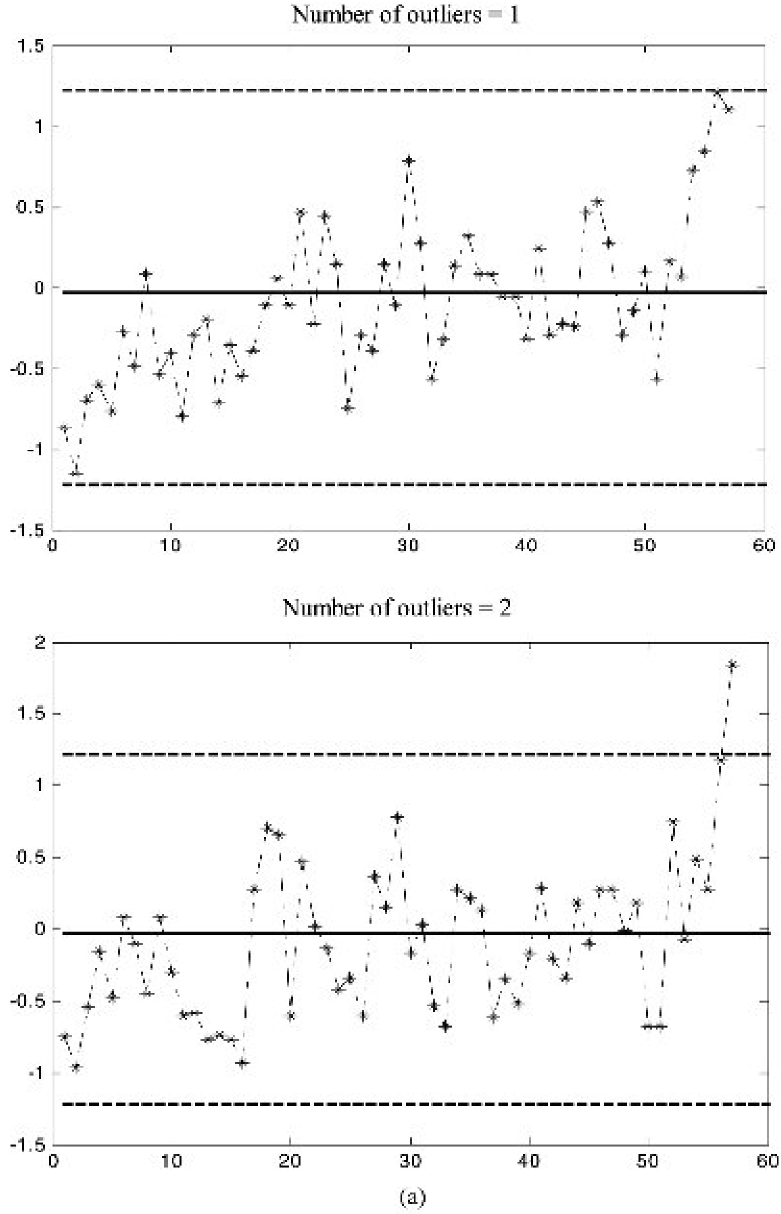


Figure 8a. RM control charts for the data on transformed weight of uncoated tablets with contamination.

normality. Moreover, RMs are

(i) highly resistant to outliers,
(ii) able to make an early detection a trend, and
(iii) found to have small variability.

The SDs steadily decrease with increase in the window size. The control charts based on the RMs have good performance in detection of trends and shifts and

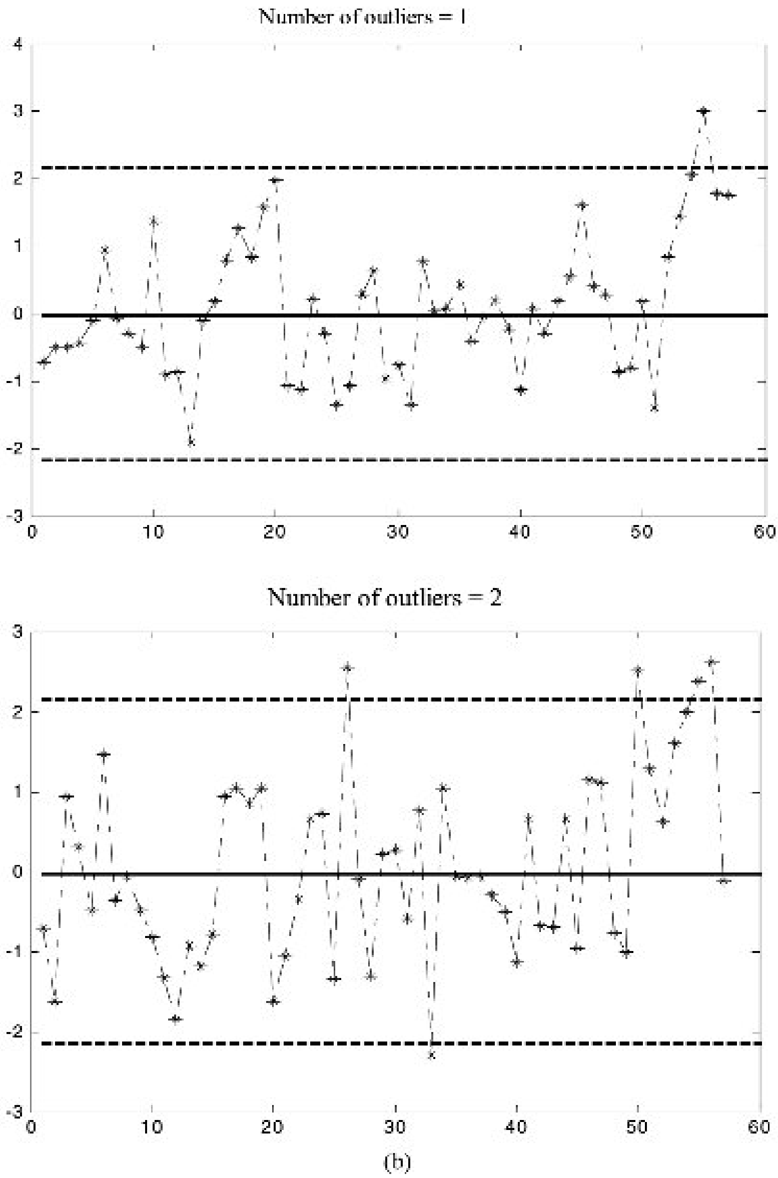


Figure 8b. Full online control charts for the data on transformed weight of uncoated tablets with contamination.

therefore the control chart based on normal approximation as developed in this article can be used beneficially in practice to monitor the process average.

The full online estimates, although, have similar robust properties, having been found to be lacking in early detection of trend which means that it would lead to delay in taking proactive action on the process. The effect of outliers is also seen to be more pronounced on the control charts based on full online estimates.

We would therefore prefer to recommend the control charts based on the RMs. However, if the user can fix the desirable value of in-control ARL and the extent of shift to be detected, he can choose the appropriate control chart using Tables 3 and 8.

Acknowledgment

The authors are indebted to the referee for his suggestions in revising and improving the article.

References

Bernholt, T., Fried, R. (2003). Computing the update of the repeated median regression line in linear time. *Information Processing Letters* 88:111–117.

Davies, P. L., Fried, R., Gather, U. (2004). Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference* 122:65–78.

Fried, R., Bernholt, T., Gather, U. (2006). Repeated median and hybrid filters. *Computational Statistics & Data Analysis* 50:2313–2338.

Gather, U., Schettlinger, K., Fried, R. (2006). Online signal extraction by robust linear regression. *Computational Statistics* 21:33–51.

Matlab 7.0.1. (2004). *The MathWorks Inc*. Massachusetts: Natick.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*. 5th ed. New York: Wiley.

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika* 69:242–244.