



## Robust Synthetic Control Charting

Ayu Abdul-Rahman<sup>1\*</sup>, Sharipah Soaad Syed-Yahaya<sup>1</sup>, Abdu Mohammed Ali Atta<sup>2</sup>

<sup>1</sup>*Department of Mathematics and Statistics, School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM, Sintok, Kedah, Malaysia*

<sup>2</sup>*Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, P.O. Box 2713-Doha*

**Abstract.** The proposal of synthetic charting is based on the normality assumption. This assumption, however, is hard to attain in practice. Therefore, it is important to examine how the chart would response under some types of non-normal data. The focus of this article is to monitor location shifts using a synthetic chart and to validate its performance under the  $g$ -and- $h$  distributions. This study shows that the effect of non-normality on the standard synthetic chart is not trivial, especially when the underlying distributions are heavy-tailed. With these types of distribution, Phase II monitoring of location using median-based estimators is advisable. In doing so, the synthetic chart is more robust to departure in the normality assumption with little effect on its out-of-control performance. This paper shows how the synthetic parameters should be attained to reflect the use of the modified one-step  $M$ -estimator ( $MOM$ ) in its Winsorized version, and the median for Phase II. The assessment is based on the average run length and supported by the extra quadratic loss function. Finally, the practical application of the proposed synthetic charts is illustrated using real data.

**Keywords:** Average run length; Extra quadratic loss; Robust estimators; Shewhart chart; Synthetic chart

### 1. Introduction

Competition in the manufacturing sector involves delivering high-quality products or services quickly at a low cost (Baby and Jebadurai, 2018). Undeniably, cost and quality are among key factors in the development of manufacturing strategies (Nurcahyo et al., 2019). To achieve such a feat, lean production is usually advocated. Lean production typically involves the use of quality tools such as control charts, flowcharts and fish bone diagrams, aiming to reduce or eliminate waste that occur during the production process. Today, the lean concept has been extended considerably into various fields beyond manufacturing (Driouach et al., 2019). As such, the use of quality tools, especially the control chart, is vital.

Application of the standard control chart, which typically relies upon the normality assumption, is favorable in many areas of the statistical process monitoring (SPM), since the tool can offer fast detection of special cause(s) in the process. Subsequently, corrective action may be undertaken to preserve process quality. One of the many variable control charts introduced in the literature is the synthetic chart for process mean, by Wu and Spedding (2000). Its use in practice is justified based on its moderate robustness to non-

\*Corresponding author's email: [ayurahman@uum.edu.my](mailto:ayurahman@uum.edu.my), Tel.: +60-12-2515660  
doi: [10.14716/ijtech.v12i2.4216](https://doi.org/10.14716/ijtech.v12i2.4216)

normality for a large sample size  $n$ , i.e.,  $n \geq 6$  (Calzada and Scariano, 2001). Nonetheless, there is a concern for unacceptably high false alarm rates from the synthetic chart based on the sample mean if data distribution is strongly skewed or very heavy-tailed, which brings us to the current discussion.

Under non-normality, it is better to use robust statistics instead of the sample mean to monitor changes in the data (Rocke, 1989). This way, outliers in the subgroups do not cause a signal to occur and, as such, the chart only responds to genuine shifts in the process. Existing work on the design and implementation of synthetic charts, as based on the median estimator, is noted in the literature. See, for example, Hu et al. (2018) and Tran et al. (2019). It is not clear how this robust chart performance is impacted under severe non-normality, but it seems that making use of the robust estimator with the synthetic control structure would lead to better quality tools, as changes in the process can be detected efficiently.

Notably, there is an increasing frequency of coverage on works associated with control charts in the quality technology field, reflecting the relevancy of this quality tool in real application (Jensen et al., 2018). Unfortunately, as reviewed by Bono et al. (2017), real data usually defy the assumption of normality. Thus, it is important to address the robustness (to non-normality) issue and to evaluate the existing and any newly proposed synthetic charts with severe deviation from normality. To fill this gap, we investigate the performance of the synthetic chart that is based on four different estimators, including the aforementioned median estimator, in monitoring the process location of Phase II samples upon violation of the normality assumption. In the next section, we give the details regarding the selected Phase II estimators before explaining the synthetic charting structure.

## 2. Methods

### 2.1. Descriptions of Robust Estimators

Let  $\theta$  be the process location estimator that needs to be monitored through control charting, and  $\hat{\theta}$  be its estimator based on a subgroup of size  $n$ . There are many choices for  $\hat{\theta}$  but we focus on median-based estimators to achieve the highest possible breakdown point ( $BP$ ), i.e., 50%. They are the usual median, modified one-step  $M$ -estimator ( $MOM$ ), and its Winsorized version. For comparison purposes, the classical mean estimator is included in the study. Based on a random sample  $X = \{X_i, X_i, \dots, X_n\}$ , these estimators are defined as follows:

- **Mean** : The sample mean  $\bar{X} = \frac{\sum_i^n X_i}{n}$  has the smallest standard error without comparison under normality and, therefore, is the most efficient estimator when data are normally distributed. Unfortunately, it also has a  $BP$  of 0, suggesting an unreliable estimate under non-normality.
- **Median** : The sample median  $\tilde{X}$  can be computed by taking the average of the two middle order statistics,  $\frac{1}{2} \left[ X_{(\frac{n}{2}-1)} + X_{(\frac{n}{2}+1)} \right]$  when  $n$  is even or, simply, the value of the mid order statistic,  $X_{(\frac{n+1}{2})}$  when  $n$  is odd. With a  $BP$  of 0.5, the estimate is still bounded even when half of the data is “bad” (Rousseeuw and Croux, 1993). Moreover, its efficiency increases and outperforms  $\bar{X}$  as tails of a distribution become heavier or more skewed, suggesting higher power (Figueiredo and Gomes, 2004).
- **MOM** : For better statistical efficiency under normal or light-tailed distribution, an alternative to  $\tilde{X}$ , namely, the  $MOM$ , is preferable (Wilcox and Keselman, 2003a). Under non-normal distribution,  $MOM$  offers better statistical efficiency than does  $\bar{X}$  (Ozdemir,

2010). Whereas  $\bar{X}$  averages all the data, *MOM* only averages the values remaining after outliers are discarded and, as such, produces a reliable estimate regardless of data distribution. More importantly, *MOM* has a *BP* of 0.5, owing to the sturdiness of  $\tilde{X}$  and the median absolute deviation about the median ( $MAD_n$ ), both which are used in ruling out the outliers (Wilcox and Keselman, 2003b). Specifically,  $X_i$  is flagged as an outlier if  $\frac{(X_i - \tilde{X})}{MAD_n} < -K$  or  $\frac{(X_i - \tilde{X})}{MAD_n} > K$  where  $K$  is adjusted to 2.24 for a reasonably small standard error under normality, and  $MAD_n = 1.4826 \text{ med}_i |x_i - \text{med}_j x_j|$ . The use of *MOM* allows for the possibility of asymmetric trimming, which is advantageous when sampling from a skewed distribution.

- **Winsorized *MOM* (*WMOM*)** : The positive features of *MOM* continue in its Winsorized version. The Winsorized *MOM* (*WMOM*) was recommended by Haddad et al. (2012) and it uses the same criteria as in *MOM* to identify outliers. Therefore, *WMOM* also possesses the highest possible *BP* (i.e., 50%). But, unlike the *MOM* estimator, the *WMOM* Winsorizes data before averaging the values in which the Winsorized sample,  $W$ , is obtained, as:

$$W_i = \begin{cases} X_{(i_1+1)}, & \text{if } X_i \leq X_{(i_1+1)} \\ X_i, & \text{if } X_{(i_1+1)} < X_i < X_{(n-i_2)} \\ X_{(n-i_2)}, & \text{if } X_i \geq X_{(n-i_2)} \end{cases} \quad (1)$$

where  $X_i$  is the  $i^{th}$  ordered observations in a sample before replacing the outliers;  $i_1$  is the number of smallest outliers determined in the sample;  $i_2$  is the number of largest outliers determined in the sample.

The monitoring of location as based on the aforementioned estimators with synthetic control structure yields four charts, namely, the synthetic  $\bar{X}$ , synthetic  $\tilde{X}$ , synthetic *MOM* and synthetic *WMOM* charts. In the remainder of this paper, each of the charts will be referred to by its respective location estimator. To equip the reader with knowledge on how these charts work to signal a change when a shift in the mean occurs, we focus on the construction of the charts in the next section.

### 2.2. Design and Derivation of Phase II Limits of the Synthetic Charts

Proposed by Wu and Spedding (2000), a synthetic chart uses two sub-charts: a Shewhart chart and a conforming run length (CRL) chart; both are working in tandem to detect a change in the process mean. Concisely, the Shewhart chart is initiated to mark nonconforming samples, which happen as soon as a sample statistic  $\hat{\theta}_i$  traverses the control limits at  $\theta_0 \pm k_s \sigma_{\hat{\theta}}$ ;  $\theta_0$  is the in-control mean value,  $\sigma_{\hat{\theta}} = \frac{\sigma_X}{\sqrt{n}}$ ,  $\sigma_X$  is the standard deviation of the study variable  $X$ , and  $k_s$  is a specified control limit coefficient. The process will be continuously monitored to detect out-of-control status, which only happens whenever a sample CRL is less than or possibly equal to a lower limit of the CRL chart,  $L_s$ . The CRL sample is attained by calculating the number of samples between two consecutive nonconforming samples, including the ending nonconforming one. Recall that these nonconforming samples are given by the Shewhart chart.

Development and evaluation of the synthetic chart are typically based on the average number of samples (average run length, ARL) required to signal process a mean shift of size  $\delta$  (in standard deviation units). The ARL ( $\delta$ ) corresponds to the out-of-control average run length when the mean changes from  $\theta_0$  to  $\theta_0 + \delta$ . Meanwhile, ARL (0) denotes the in-control value. In designing the chart, i.e., deriving appropriate values of  $k_s$  and  $L_s$ , the two charting constants that form the Phase II limits of the synthetic chart, requires user specification on the anticipated shift, say  $\delta^* > 0$ , and ARL (0). In practice, the pre-specified

ARL (0) is set largely based on the cost of false alarms, while  $\delta^*$  is considered to have substantial impact on the quality of the process.

Considering the intractability of the sampling distribution of the robust Phase II estimators selected in this study, the values of  $k_s$  and  $L_s$  for the robust synthetic charts, i.e.,  $\bar{X}$ ,  $MOM$  and  $WMOM$ , are obtained via simulation. A code developed in SAS/Data Step Version 9.4 is used to identify a set of  $k_s$  and  $L_s$  that minimizes ARL ( $\delta^*$ ) while satisfying ARL (0). For such a purpose, Equation 2 defined by  $ARL(\delta^*) = \frac{1}{p} \times \frac{1}{[1-(1-p)^{L_s}]}$  is first used to retrieve several possible pairs ( $k_s, L_s$ ) that meet the requirement for ARL (0) when  $\delta^* = 0$ . Here,  $p$  is the fraction nonconforming in the process, which depends on the Shewhart limit coefficient,  $k_s$ . Initially,  $L_s$  is set at 1 with some relevant  $k_s$  value for the synthetic design. Then,  $1 \times 10^6$  samples of size  $n$  are simulated for all charts from the standard normal distribution and, for each chart, the sample statistic  $\hat{\theta}_i$  is calculated. It is noted that the respective  $\hat{\theta}_i$  exceeds the control limits of the Shewhart chart, which starts the counter for a non-conforming sample. The fraction nonconforming  $p$  for the process is obtained by averaging over the total  $1 \times 10^6$  samples. The process is repeated by increasing the value of  $L_s$  from 1 to 10 incrementally. The resulting pairs ( $k_s, L_s$ ) will result in different ARL ( $\delta^*$ ) in Equation 2 and for the set of  $k_s$  and  $L_s$  for which ARL ( $\delta^*$ ) reaches minimum, it will be the final design of the synthetic chart. It is worth mentioning that the optimal pair ( $k_s, L_s$ ) for the  $\bar{X}$  obtained through this simulation process is comparable to the numerical results shown in the work of Wu and Spedding (2000), who also noted that the synthetic chart is quite robust with regard to  $L_s$  and, as such, the optimal design can frequently be found among the first ten values of  $L_s$ . However, through this simulation study, we also discover that, when the anticipated shift is quite small, the optimization search may require a much larger set of  $L_s$  (and hence,  $k_s$ ) before ARL ( $\delta^*$ ) can be further reduced. The results in Table 1 for  $n = 5$  and  $\delta^* = 0.5$  corroborate the statement.

**Table 1** Optimizing synthetic charts given ARL (0) = 370

$\delta^*$	$n$	Charts									
		$\bar{X}$		$\tilde{X}$		$MOM$		$WMOM$			
		$L_s$	$k_s$	$L_s$	$k_s$	$L_s$	$k_s$	$L_s$	$k_s$		
0.5	5	14	2.443	15	2.4629	18	2.5558	14	2.4676		
	1	5	4	2.219	4	2.2237	4	2.2599	4	2.233	
	2	5	2	2.085	2	2.0882	2	2.1134	2	2.0943	
0.5	9	8	2.346	8	2.3492	7	2.3617	7	2.3373		
	1	9	2	2.085	2	2.0859	2	2.1043	2	2.0898	
	2	9	1	1.943	1	1.9436	1	1.9536	1	1.9454	

Table 1 presents the optimized synthetic control chart values of  $k_s$  and  $L_s$  when the underlying process data are normally distributed and ARL (0) is set to 370, which is a widely used choice. In this study, the pairs ( $k_s, L_s$ ) are derived for different combinations of the sample size  $n = \{5, 9\}$  and design shift  $\delta^* = \{0.5, 1, 2\}$ .

### 3. Results and Discussion

#### 3.1. Performance of the Proposed Synthetic Control Charts

In assessing control chart performance via ARL, the best charts are those charts that can produce ARL (0) closest to the pre-specified value. Meanwhile the ARL ( $\delta$ ) is expected to be as small as possible, indicating fast detection of a process change. Along with the ARL ( $\delta$ ), this study reports the extra quadratic loss (EQL) value for an out-of-control (OC) process. Both ARL and EQL function differently as efficiency indicators. The ARL, in general,

measures the performance of the control chart for a specific shift size,  $\delta$ , and the EQL measures the overall performance over the range of shifts under consideration. The EQL is defined as  $EQL(\delta) = \frac{1}{\delta_{max}} \sum_0^{\delta_{max}} \delta^2 \times ARL(\delta)$  where  $\delta_{max}$  is the upper boundary of the range of shifts under consideration (Malela-Majika et al., 2019). The EQL formula gives penalty to the ARL values for larger shifts using a weight  $\delta^2$ . Thus, if the chart had a slowly declining ARL ( $\delta$ ), it yields a larger EQL. Conversely, if the chart had a quickly declining ARL ( $\delta$ ), which is an indication of a good chart, it yields a smaller EQL (i.e., minimum EQL value) and is considered to be the winner (in terms of fast detection) among the charts being compared.

3.1.1. In-control (IC) performance

To study robustness of the proposed synthetic charts, we applied the parameters of the optimized synthetic charts from Table 1 to process data from the  $g$ -and- $h$  distributions (Hoaglin, 1985). For each parameter combinations, all ARL values reported for the synthetic charts studied here were computed using 10,000 simulations per distribution.

The  $g$ -and- $h$  distributions are based on transformation of the standard normal, which allows for symmetric and heavier tails. Specifically, the standard normal variates,  $Z_i$ , are converted to random variables  $X_i$  as follows (Martinez and Iglewicz, 1984):

$$X_i = \begin{cases} ((e^{gZ_i} - 1)/g) \times (e^{hZ_i^2/2}), & g \neq 0; \\ Z_i (e^{hZ_i^2/2}), & g = 0. \end{cases} \tag{2}$$

The parameters  $g$  and  $h$  in Equation 2 control the amount of skewness and kurtosis, respectively. To illustrate, Figure 1 shows plots of several  $g$ -and- $h$  distributions which cover various degrees of non-normality of the underlying data. As shown in the Figure, tails of the distribution become heavier as  $h$  increased and are further skewed as  $g$  increased.

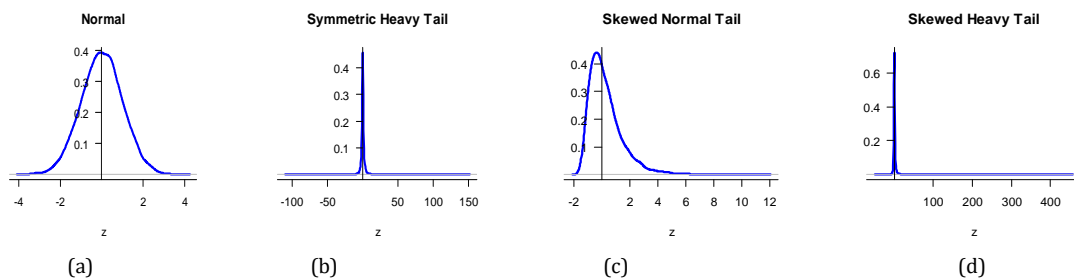


Figure 1 Distributional shapes: (a)  $g = 0, h = 0$ ; (b)  $g = 0, h = 0.5$ ; (c)  $g = 0.5, h = 0$ ; (d)  $g = 0.5, h = 0.5$

By using process data from the  $g$ -and- $h$  distributions with  $(g, h) = (0, 0), (0, 0.5), (0.5, 0)$ , and  $(0.5, 0.5)$  (Figure 1), the effect of non-normality on the performance of the four investigated synthetic charts, i.e.,  $\bar{X}, \tilde{X}, MOM$ , and  $WMOM$ , can be studied thoroughly. Some properties of these four  $g$ -and- $h$  distributions can be referred from the paper by Wilcox and Keselman (2003b). By including the case  $g = h = 0.5$ , the ability of a synthetic chart to perform under extreme deviation from the normality assumption can be judged exhaustively. If the chart can perform well under this extreme non-normality, it is safe to conclude the same satisfactory performance under distributions of lesser skewness and/or heavy-tailedness.

Table 2 displays the ARL (0) for the proposed synthetic charts of different sample sizes when the underlying distribution actually follows the  $g$ -and- $h$  distributions discussed above. In an attempt to judge robustness, these synthetic charts were designed using values of  $k_s$  and  $L_s$  in Table 1, which are technically only appropriate for normally distributed process data. The results in Table 2 show variations in the control chart performance in



relation to the types of distribution. First, consider the situation where the underlying process data are normally distributed. All four charts have a similar performance, i.e., they have a comparable ARL (0) of approximately 370, as they were designed to have.

Next, consider the situation where data are non-normally distributed. Combining the two factors—the distributional shape and design shift—we notice that the ARL (0) produced by synthetic charts based on  $\bar{X}$ ,  $\tilde{X}$ ,  $MOM$ , and  $WMOM$  are well controlled in any non-normal data, if they were designed for a medium ( $\delta^* = 1$ ) and a large ( $\delta^* = 2$ ) shift. It is also interesting to note here that several ARL (0) for these charts actually exceed 370 when data distribution is skewed normal-tailed, with  $MOM$  consistently emerging as the best chart. Nonetheless, if we specify a too-small shift in designing the chart, a few of them may not perform robustly under heavy-tailed distributions. However, their IC performance can certainly be enhanced with larger sample sizes  $n$ , resulting in  $MOM$  producing ARL (0) that are much closer to the pre-specified value when compared to the rest of the charts under consideration. Once again, the application of  $MOM$  is advantageous if we design the chart for small shift and apply it on the skewed normal-tailed process data. The  $MOM$  generates ARL (0) = 246.41 for  $n = 5$  and ARL (0) = 382.05 for  $n = 9$ , beating other supposedly robust charts for the case  $g = 0.5$  and  $h = 0$ .

**Table 2** ARL(0) for sample size  $n = 5$  and  $n = 9$

$\delta^*$	$g$	$h$	Charts							
			$n = 5$				$n = 9$			
			$\bar{X}$	$\tilde{X}$	$MOM$	$WMOM$	$\bar{X}$	$\tilde{X}$	$MOM$	$WMOM$
0.5	0.0	0.0	371.78	370.34	372.99	372.80	372.03	366.32	374.92	375.63
		0.5	1660.85	119.43	118.30	127.84	2836.71	176.98	214.89	188.51
	0.5	0.0	202.89	188.76	246.41	248.70	300.36	286.36	382.05	332.24
		0.5	6706.56	133.94	137.34	139.61	6986.98	184.40	219.32	198.19
1	0.0	0.0	366.94	369.97	372.78	372.21	372.56	365.77	368.52	374.51
		0.5	3657.79	203.72	221.67	205.64	5010.58	270.34	289.21	267.65
	0.5	0.0	326.80	310.63	389.22	349.99	400.20	397.51	426.20	385.98
		0.5	7114.94	253.71	267.53	251.83	7355.55	310.12	321.58	310.92
2	0.0	0.0	371.88	373.58	372.92	371.48	372.07	370.77	370.97	376.80
		0.5	4739.19	268.63	283.57	271.91	5783.29	325.47	331.96	320.41
	0.5	0.0	399.47	394.75	450.34	396.83	430.82	433.06	438.67	405.31
		0.5	7438.38	351.37	360.92	345.25	7370.50	380.08	386.75	389.56

Despite being able to control false alarms under skewed normal-tailed cases, the  $\bar{X}$  fails to demonstrate the same virtue when applied to heavy-tailed data, as its ARL (0) are significantly higher than the expected 370. This is true for both sample sizes, alluding to a potential power problem (to be discussed next).

**3.1.2. Out-of-control (OC) performance**

Sample sizes of  $n = 5$  and  $9$  are common in SPM because real data based on these numbers are easily obtained. However, in synthetic charts, it is advisable to use  $n \geq 6$  to achieve good detection (Calzada and Scariano, 2001). Thus, only cases for  $n = 9$  are discussed in this section.

Table 3 displays results for the ARL ( $\delta$ ) which are obtained using ARL (0) = 370,  $n = 9$  and  $\delta^* = \{0.5, 1, 2\}$  for the shift size  $\delta = \{0.25, 0.5, 0.75, 1, 1.5, 2, 3\}$ . The EQL values are listed as well.

For the reader’s convenience, the minimum EQL value for different combinations of  $n$  and  $\delta^*$  is in bold, suggesting the best overall performance. We first start with the case of  $g = h = 0$  (the upper part of Table 3). Noticeably, all four charts perform consistently across all large  $\delta$  values, when the underlying process data are normally distributed.

**Table 3** ARL ( $\delta$ ) using process data from various  $g$ -and- $h$  distributions when  $n = 9$

$g$	$h$	$\delta^*$	Charts	ARL ( $\delta$ )							EQL	
				0.25	0.5	0.75	1	1.5	2	3		
0	0	0.5	$\bar{X}$	48.59	5.94	2.21	1.35	1.02	1.00	1.00	<b>7.47</b>	
			$\tilde{X}$	48.86	6.02	2.22	1.34	1.01	1.00	1.00	<b>7.47</b>	
			MOM	55.85	6.77	2.27	1.35	1.02	1.00	1.00	7.70	
			WMOM	51.99	6.22	2.17	1.33	1.02	1.00	1.00	7.55	
	0	1	0.5	$\bar{X}$	61.20	7.42	2.19	1.26	1.01	1.00	1.00	<b>7.81</b>
				$\tilde{X}$	61.71	7.41	2.15	1.26	1.01	1.00	1.00	<b>7.81</b>
				MOM	67.40	7.94	2.24	1.27	1.01	1.00	1.00	8.00
				WMOM	62.29	7.69	2.19	1.26	1.01	1.00	1.00	7.86
	0	2	0.5	$\bar{X}$	70.43	9.32	2.58	1.36	1.01	1.00	1.00	<b>8.27</b>
				$\tilde{X}$	71.35	9.30	2.61	1.36	1.01	1.00	1.00	8.30
				MOM	75.49	9.50	2.63	1.36	1.02	1.00	1.00	8.41
				WMOM	71.21	9.18	2.61	1.37	1.01	1.00	1.00	8.29
0	0.5	0.5	$\bar{X}$	1615.09	222.83	2.94	1.04	1.00	1.00	1.00	58.20	
			$\tilde{X}$	51.20	7.33	2.23	1.31	1.02	1.00	1.00	<b>7.63</b>	
			MOM	57.88	8.05	2.26	1.31	1.02	1.00	1.00	7.83	
			WMOM	55.38	7.52	2.20	1.29	1.02	1.00	1.00	7.73	
	0	1	0.5	$\bar{X}$	3217.60	216.24	1.44	1.02	1.00	1.00	1.00	90.75
				$\tilde{X}$	69.75	8.68	2.12	1.23	1.01	1.00	1.00	<b>8.08</b>
				MOM	74.01	9.17	2.17	1.23	1.01	1.00	1.00	8.22
				WMOM	72.51	8.87	2.12	1.23	1.01	1.00	1.00	8.15
	0	2	0.5	$\bar{X}$	4010.05	163.89	1.32	1.03	1.00	1.00	1.00	102.88
				$\tilde{X}$	86.77	10.60	2.46	1.30	1.02	1.00	1.00	<b>8.69</b>
				MOM	88.59	10.57	2.52	1.32	1.02	1.00	1.00	8.74
				WMOM	88.05	10.70	2.48	1.30	1.02	1.00	1.00	8.72
0.5	0	0.5	$\bar{X}$	35.37	6.58	2.39	1.38	1.00	1.00	1.00	<b>7.28</b>	
			$\tilde{X}$	34.96	6.80	2.42	1.39	1.00	1.00	1.00	7.30	
			MOM	43.02	6.98	2.44	1.37	1.01	1.00	1.00	7.48	
			WMOM	39.85	6.71	2.36	1.37	1.01	1.00	1.00	7.37	
	0.5	1	0	$\bar{X}$	56.59	8.68	2.40	1.26	1.00	1.00	1.00	7.86
				$\tilde{X}$	55.51	8.88	2.45	1.25	1.00	1.00	1.00	<b>7.85</b>
				MOM	59.49	8.97	2.47	1.27	1.00	1.00	1.00	7.96
				WMOM	56.58	8.68	2.39	1.26	1.00	1.00	1.00	<b>7.85</b>
	0.5	2	0	$\bar{X}$	70.94	11.11	2.86	1.33	1.00	1.00	1.00	8.47
				$\tilde{X}$	70.79	11.16	2.91	1.33	1.00	1.00	1.00	8.48
				MOM	72.62	10.98	2.86	1.36	1.00	1.00	1.00	8.50
				WMOM	70.46	10.74	2.88	1.35	1.00	1.00	1.00	<b>8.44</b>
0.5	0.5	0.5	$\bar{X}$	5623.46	1516.86	11.67	1.00	1.00	1.00	1.00	251.16	
			$\tilde{X}$	42.45	8.10	2.50	1.33	1.01	1.00	1.00	<b>7.56</b>	
			MOM	47.99	8.82	2.53	1.34	1.01	1.00	1.00	7.75	
			WMOM	44.57	8.35	2.51	1.32	1.01	1.00	1.00	7.62	
	0.5	1	0.5	$\bar{X}$	6568.34	1949.92	1.05	1.00	1.00	1.00	1.00	304.95
				$\tilde{X}$	69.86	10.74	2.44	1.21	1.01	1.00	1.00	<b>8.30</b>
				MOM	73.36	11.30	2.48	1.22	1.01	1.00	1.00	8.43
				WMOM	70.74	11.32	2.47	1.20	1.00	1.00	1.00	8.37
	0.5	2	0.5	$\bar{X}$	6835.23	1782.46	1.01	1.00	1.00	1.00	1.00	296.55
				$\tilde{X}$	91.83	13.54	2.79	1.28	1.01	1.00	1.00	<b>9.08</b>
				MOM	92.65	13.42	2.82	1.29	1.01	1.00	1.00	9.10
				WMOM	92.42	13.85	2.87	1.26	1.01	1.00	1.00	9.12

For  $\delta \geq 1$ , all three designs offer equivalent ARL performance, suggesting the charts will not lose performance even if we misdesigned them. For  $\delta \leq 0.75$ , we perceive a

variation in the control chart performance, noticeably, as soon as we introduced potential shifts in the process. For instance, the ARL ( $\delta = 0.25$ ) of the  $\bar{X}$  is the smallest, suggesting quickest detection among all charts under consideration. This chart also yields the best overall performance (in terms of EQL) across all design shifts, as to be expected under normality.

Next, the case  $g = 0$  and  $h = 0.5$  is considered. The general observation is that  $\bar{X}$  yields the worst performance with a slowly declining ARL ( $\delta$ ). The effect of heavy-tailed distribution on the chart is trivial for  $\delta \geq 1$  but, below the threshold, it loses performance. It has the worst performance when  $\delta^* = 2$ . Evidently, synthetic charting using median-based estimators is more beneficial when employed with this type of distribution. By having significantly smaller standard errors than  $\bar{X}$  under non-normality, the robust charts produce narrower control limits, which lead to better detection of the OC status. To illustrate, in Table 3, the ARL ( $\delta = 0.25$ ) of the  $\bar{X}$  designed for  $\delta^* = 2$  is 4010.05, which is significantly (about 45 times) higher than the next largest ARL ( $\delta = 0.25$ ) value produced by the *MOM* (88.59). According to the EQL determinant,  $\tilde{X}$  is the most effective in detecting shifts.

Arriving at the case  $g = 0.5$  and  $h = 0$ , the general conclusion is that the effect of change in design shift follows the finding under normality, and that the effect can be generalized across all estimators. However, the ranking changes, unlike that seen in the normal distribution. Indeed, there is no variation in the control chart performance if a large shift ( $\delta > 1$ ) occurs, and that minimal difference exists if the magnitude of the shift is small ( $\delta < 0.75$ ). However, when we introduced the smallest shift ( $\delta = 0.25$ ) in the process, the ARL ( $\delta$ ) for the synthetic chart based on *MOM* and *WMOM* are slightly higher than the rest of the charts under consideration. While it seems counterintuitive that the *WMOM* gives the best overall performance if designed for  $\delta^* = 2$ , this can be explained by the weightage assigned to the ARL based on  $\delta^2$  in the EQL formula. The ARL ( $\delta$ ) for *WMOM* is much smaller for larger values of  $\delta$  compared to the rest of the charts under consideration. Therefore, the chart is more dominant than others. For  $\delta^* = 0.5$  and 1, both  $\bar{X}$  and  $\tilde{X}$  are at par according to the EQL determinant, and they both outweigh the *MOM* and *WMOM* in detecting OC status.

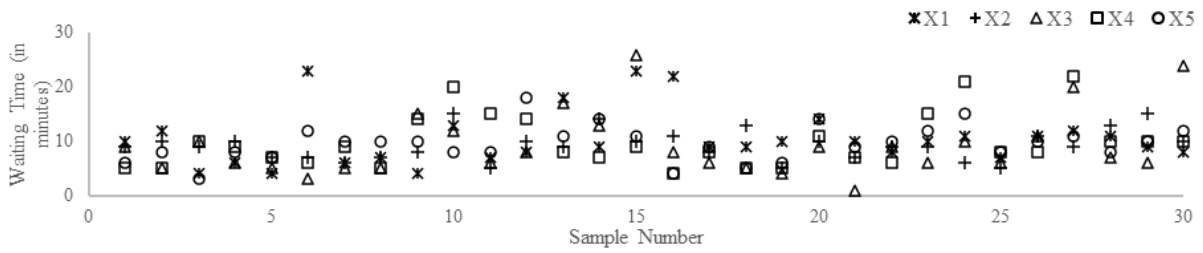
Under extreme non-normality, i.e., when  $g = h = 0.5$ ,  $\tilde{X}$  is regarded as most effective in signalling OC status in accordance with the ARL and EQL. This is true for all design shifts. Conversely,  $\bar{X}$  yields the worst performance with substantially large ARL values for small shifts. The results support our conjecture in Subsection 3.1.1. regarding detection delay of a process change by  $\bar{X}$ .

Evaluation of synthetic charts under severe deviation from normality as illustrated in this section confirms that the standard synthetic chart ( $\bar{X}$ ) is highly susceptible to the effect of non-normality. The situation can be easily managed by using robust statistics in the place of sample mean.

### 3.2. A Numerical Example

The purpose of this example is to illustrate the performance of the synthetic charts when (i) data is real, and (ii) process parameters are estimated. The data concern the wait times (in minutes) of patients' who underwent a colonoscopy procedure in a regional medical center. Thirty samples, each of size five patients, had been taken and were shown to be positively skewed by [Jones-Farmer et al. \(2009\)](#). The data are presented in Figure 2.





**Figure 2** Patients’ waiting time (in minutes)

From these 30 samples, the in-control mean,  $\theta_0$ , is estimated by the average  $\hat{\theta}$ , computed as  $\hat{\theta}_0 = \sum_{j=1}^{30} \hat{\theta}_j / 30$ . With four different choices for  $\hat{\theta}$ , as discussed in Subsection 2.1, we then have four different estimators for  $\hat{\theta}_0$  to be considered in this example, namely  $\bar{X}$ ,  $\tilde{X}$ ,  $MOM$ , and  $WMOM$ . For a fair comparison of these location estimators, the standard deviation estimator  $\hat{\sigma}$  for each chart is the same. In this paper,  $\sigma$  is estimated by the average range  $R$  defined by  $\hat{\sigma} = \left(\frac{1}{d_2}\right) \sum_{j=1}^{30} R_j / 30$ , where the value of  $d_2$  for  $n = 5$  is taken to be equal to 2.326 (Montgomery, 2013). For this example, the ARL (0) is set at 370 and the design shift  $\delta^*$  at 0.5. Then, the control limits coefficient  $k_S$  of the Shewhart sub-chart and the lower control limit  $L_S$  of the CRL sub-chart for the different synthetic charts can be obtained from Table 1, because the pairs  $(k_S, L_S)$  only depend on  $n$ , ARL (0) and  $\delta^*$  (Wu and Spedding, 2000). For example, if  $\hat{\theta} = \bar{X}$ , then  $k_S = 2.443$  and  $L_S = 14$ . Substituting the values of  $k_S$  and  $L_S$  into  $\hat{\theta}_0 \pm k_S \frac{\hat{\sigma}}{\sqrt{n}}$  yield 5.59 and 13.1373 for the lower (LCL) and upper (UCL) limits of the Shewhart sub-chart, respectively. In the same manner, we obtain the control limits of the Shewhart sub-chart for the three robust synthetic charts: (i)  $\tilde{X}$ , (ii)  $MOM$ , and (iii)  $WMOM$ . Accordingly, their control limits are (i) (5.3, 13.10), (ii) [4.9, 13.10], and (iii) [4.89, 13.11], (in brackets, the first value is LCL followed by UCL). Applying all four charts on the skewed data, no OC status is detected by the robust charts but the  $\bar{X}$  signals at subgroup 27. In practice, further investigation needs to be done before taking corrective action, but if the process is actually stable, the false signal by  $\bar{X}$  simply leads to unneeded distraction, not to mention wasted effort. This concern can easily be managed via the application of robust synthetic charts, as supported by the findings in Section 3.1.

#### 4. Conclusions

Robustification in this study is intended to lead to a synthetic chart, the IC ARL of which is not sensitive to the non-normal data or the particular values of design parameters,  $k_S$  and  $L_S$ , but which will detect location shifts reliably. We identify  $MOM$  as best fit for the synthetic chart. This way, researchers in many application domains will not be constrained by the normality assumption, but may instead work with the original data without having to worry about the shape of the distributions. While this paper concentrated on the performance of the synthetic chart based on the assumption of known process parameters, we believe that the findings can be easily extended to the estimated parameters case, which will be considered in our future work.

The decision of what shift for which to design rarely gets attention in the literature, and even more rarely under non-normal cases. The three design shifts employed in this study provide valuable information for further implementation and for the designing of synthetic charts in practice. Generally, the synthetic charts designed for a small shift ( $\delta = 0.5$ ) exhibit lack of IC robustness, and therefore can be misleading in the detection of OC status. However, for a large sample size ( $n = 9$ ), the synthetic  $MOM$  and  $WMOM$  charts are better at

controlling the  $ARL(0)$ , which leads to the best charts (in terms of in-control robustness) across the distributional shapes. Moreover, the results confirm that these synthetic median-based charts (including the use of the usual median estimator) perform reliably and quickly if designed for a medium-sized ( $\delta = 1$ ) or a large shift ( $\delta = 2$ ). For OC processes, the data of which follow heavy-tailed distributions, these charts offer quickly declining ARLs, suggesting not much loss in responsiveness to genuine shifts in the process.

### Acknowledgements

The authors would like to acknowledge the work that has led to this paper, which is supported by Universiti Utara Malaysia, Fundamental Research Grant Scheme (S/0 Code 13578) of the Ministry of Higher Education, Malaysia.

### References

- Baby, B.N.P., Jebadurai, D.S., 2018. Implementation of Lean Principles to Improve the Operations of a Sales Warehouse in the Manufacturing Industry. *International Journal of Technology*. Volume 9(1), pp.46–54
- Bono, R., Blanca, M.J., Arnau, J., Gomez-Benito, J., 2017. Non-normal Distributions Commonly used in Health, Education, and Social Sciences: A Systematic Review. *Frontiers in Psychology*, Volume 8, pp. 1–6
- Calzada, M.E., Scariano, S.M., 2001. The Robustness of the Synthetic Control Chart to Non-normality. *Communications in Statistics - Simulation and Computation*, Volume 30(2), pp. 311–326
- Driouach, L., Zarbane, K., Beidouri, Z., 2019. Literature Review of Lean Manufacturing in Small and Medium-sized Enterprises. *International Journal of Technology*. Volume 10(5), pp. 930–941
- Figueiredo, F., Gomes, M.I., 2004. The Total Median in Statistical Quality Control. *Applied Stochastic Models in Business and Industry*, Volume 20(4), pp. 339–353
- Haddad, F.S., Syed-Yahaya, S.S., Alfaro, J.L., 2012. Alternative Hotelling's T<sup>2</sup> Charts using Winsorized Modified One-Step M-estimator. *Quality and Reliability Engineering International*, Volume 29, pp. 583–593
- Hoaglin, D.C., 1985. Summarizing shape numerically: The *g*-and-*h* Distributions. In D. Hoaglin, F. Mosteller, J. Tukey (Eds.), Wiley, Hoboken, NJ. *Exploring Data Tables, Trends, and Shapes*, pp. 461–513
- Hu, X-L., Wu, S., Zhong, J-L., 2018. Optimal Design of the Synthetic Median Chart based on Average Run Length and Expected Average Run Length. In: Proceedings of the 24<sup>th</sup> ISSAT International Conference on Reliability and Quality in Design 2018, Toronto, 2 - 4 August, Canada
- Jensen, W.A., Montgomery, D.C., Tsung, F., Vining, G.G., 2018. 50 Years of the Journal of Quality Technology. *Journal of Quality Technology*, Volume 50(1), pp. 2–16
- Jones-Farmer L.A., Jordan, V., Champ, C.W., 2009. Distribution-Free Phase I Control Charts for Subgroup Location. *Journal of Quality Technology*, Volume 41(3), pp. 304–316
- Malela-Majika, J.C. Motsepa, C.M., Graham, M.A., 2019. A New Double Sampling  $\bar{X}$  Control Chart for Monitoring an Abrupt Change in the Process Location. *Communication in Statistics- Simulation and Computation*, Volume 50(3), pp. 917–935
- Martinez, J., Iglewicz, B., 1984. Some Properties of the Tukey *g* and *h* Family of Distributions. *Communications in Statistics - Theory and Methods*, Volume 13(3), pp. 353–369
- Montgomery, D.C., 2013. *Introduction to Statistical Quality Control* (7th Ed.), John Wiley & Sons, Hoboken, NJ

- Nurchahyo, R., Wibowo, A.D., Robasa, R., Cahyati, I., 2019. Development of a Strategic Manufacturing Plan from a Resource-Based Perspective . *International Journal of Technology*. Volume 10(1), pp. 178–188
- Ozdemir, A.F., 2010. Comparing Measures of Location When the Underlying Distribution has Heavier Tails than Normal. *Istatistikçiler Dergisi*, Volume 3, pp. 8–16
- Rocke, D.M., 1989. Robust Control Charts. *Technometrics*, Volume 31(2), pp. 173–184
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, Volume 88(424), pp. 1273–1283
- Tran , P.H., Tran, K.P., Rakitzis, A., 2019. A Synthetic Median Control Chart for Monitoring the Process Mean with Measurement Errors. *Quality and Reliability Engineering International*, Volume 35, pp. 1100–1116
- Wu, Z., Spedding, T.A., 2000. A Synthetic Control Chart for Detecting Small Shifts in the Process Mean. *Journal of Quality Technology*, Volume 32(1), pp. 32–38
- Wilcox, R.R., Keselman, H.J., 2003a. Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods*, Volume 8(3), pp. 254–274
- Wilcox, R.R., Keselman, H.J., 2003b. Repeated Measures One-Way ANOVA based on a Modified One-Step M-Estimator. *British Journal of Mathematical and Statistical Psychology*, Volume 56(1), pp. 15–25