SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Modified Levey-Jennings Chart with Robust Estimator: A Case of Semiconductor Manufacturing Process

Sufinah Dahari[1,*]

[1] Department of Research and Development, Dominant Opto Technologies Sdn. Bhd., 75350 Melaka, Malaysia

ARTICLE INFO

Article history:
Received 1 September 2023
Received in revised form 8 November 2023
Accepted 1 March 2024
Available online 17 April 2024

Keywords:
Robust statistics; Automated control chart; Levey-Jennings; Non-stationary

ABSTRACT

In the era of Industrial Revolution 4.0 and smart manufacturing, the development and deployment of control charts used in the semiconductor industry need to be automated. Consequently, artificial intelligence-based automation methods typically encompass the deployment of statistical software such as JMP. Automation involves frequent dataset updates; the control limits are recalculated as the parameters change (non-stationary behaviour). This requires the user to define the control chart type before its deployment on the production floor. An initially normally distributed dataset may be skewed during the process owing to the influence of outliers. If a user selects a chart based on normality assumptions, detecting a process-mean shift may be impossible if the recalculated limit fluctuates. In the semiconductor industry, a process-mean shift occurs owing to special cause variations in the process. This signals process deterioration, which may affect the quality of the product. It is unknown when outliers will affect the equilibrium of normality assumptions; therefore, it is important to develop an automated, robust control chart that can detect special cause variations under non-stationary conditions. This study proposes the use of Huber's M-estimators in the Levey-Jennings chart to detect special cause variations in a semiconductor manufacturing process. This study computes the robust M-estimates of all available samples to calculate the new limits in the Levey-Jennings chart. This new chart is referred to as the modified Levey-Jennings with a robust Huber M-estimator (MLVHM). Using production data from Dominant Opto Technologies Sdn. Bhd., Malaysia, a statistical comparison of the MLVHM and Levey-Jennings charts was performed. While the MLVHM is stable, the absolute difference in dispersion between the two charts ranges between 25.26% and 47.91% owing to standard deviation variation in the Levey-Jennings chart in non-stationary situations with outliers. The study concludes that the MLVHM chart is robust and suitable for industrial automatic flow applications.

1. Introduction

Statistical process control (SPC) charts were developed by Walter Shewhart in 1924 to monitor central location and dispersion towards preventing the manufacturing of defective products [1].

* *Corresponding author.*
E-mail address: sufinah.dahari@dominant-semi.com

https://doi.org/10.37934/araset.43.2.189202

In the SPC chart, the output may indicate special cause variation or signal process deterioration as an "out-of-control" event, or OOC. There are a few types of Shewhart control charts, such as the X-bar chart to monitor the mean and R or S charts to monitor the dispersion within batch observations [2]. For a large number of samples within the batch (e.g., n > 10), it is recommended to use the S chart to monitor the dispersion because the R chart is sensitive to changes in the process standard deviation [2]. The Levey–Jennings (LV) control chart was introduced for clinical laboratories as an alternative to the Shewhart control chart to detect special cause variations [3]. Generally, the LV method uses all the samples in a dataset to calculate the statistics required for the control limit. This differs from the Shewhart version of the control chart, which calculates limits from batch observations in terms of the central location (estimated mean) and dispersion (estimated standard deviation) [4]. Table 1 shows the differences between LV and classic Shewhart (X-bar and S chart) control limit computation [5,6]. The LV chart is commonly used to monitor individual measurements against central locations whereas Shewhart charts (X-bar and S chart) are used to monitor the mean or standard deviation within subgroups against the weighted average of subgroup means and standard deviations. The Shewhart chart requires the user to input several subgroups of data for the initial control limit computation, typically 20–25 subgroups [2]. However, in a fast-moving semiconductor industry, an automated control chart based on the first available subgroup data is preferred.

Table 1
Comparison between Levey-Jennings and Shewhart Control limit formula

Control Chart	Centre Line, CL (Estimated location)	Dispersion (Estimated sigma)	Upper Control Limit (UCL)	Lower Control Limit (LCL)
Shewhart (S-chart)	$\bar{s} = \frac{1}{m} * (\sum_{j=1}^{m} sj)$	$\frac{\bar{s}}{c_4} * (\sqrt{1 - c_4^2})$	$\bar{s} + k * \frac{\bar{s}}{c_4} * (\sqrt{1 - c_4^2})$	$\bar{s} - k * \frac{\bar{s}}{c_4} * (\sqrt{1 - c_4^2})$
Shewhart (X-bar chart)	$\bar{\bar{X}} = \frac{1}{m} * (\sum_{j=1}^{m} \frac{\sum_{i=1}^{n} x_i}{n})$	$\frac{1}{m} * (\sum_{j=1}^{m} sj)$	$\bar{\bar{X}} + k * \frac{\bar{s}}{c_4\sqrt{n}}$	$\bar{\bar{X}} - k * \frac{\bar{s}}{c_4\sqrt{n}}$
Levey-Jennings	$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}$	$\sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}}$	$\bar{x} + k * \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}}$	$\bar{x} - k * \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}}$

where s denotes the standard deviation, m represents the number of subgroups, n denotes the number of samples in the subgroup (sample size), N denotes the number of total samples (n x m), k denotes k-sigma (typically three), $c4$ is a constant that depends on the sample size n if the underlying distribution is normal, and $\bar{\bar{X}}$ denotes the weighted average of subgroup means [5,6].

2. Literature Review

Numerous researchers have extensively discussed issues with the classic Shewhart or LV control charts related to outliers, normal distribution assumptions, and non-stationary data [7-19]. Abu-Shawiesh and Abdullah proposed a univariate robust SPC chart for location based on the Hodges–Lehmann (HL) estimator using simulated data. They remarked that the advantages of the robust control chart outweighed its drawbacks in terms of statistical effectiveness evaluated using the average run length (ARL). As an alternative to Shewhart's control chart, Abu-Shawiesh developed a robust control chart based on the median absolute deviation (MAD) of the process dispersion. He used a Monte Carlo simulation study to illustrate the ARL performance of the proposed robust control

chart. Raji *et al.,* applied Shewhart–Tukey- and MAD-based models to detect and screen outliers before computing the control limits on the photolithography process in the semiconductor industry using real univariate variable data [20]. Their method also used the ARL to evaluate the chart performance. However, Lazariv and Schmid demonstrate that ARLs do not exist in certain processes. Atalay *et al.,* provided strategies for practitioners to deploy an automated control chart regarding the performance of the control chart. However, their proposed strategies were based on stationary conditions [21].

Most studies on process control charts with time-dependent data assume that the target process is stationary and use ARLs to evaluate the chart's performance [19]. Woodall argues that, in theory, the ARL can be controlled under strong assumptions, but this is not practical [22]. Thus, in this study, ARL is not used to evaluate chart performance. In summary, although the research on managing outliers and normal distribution assumptions using robust control charts is ample, non-stationary issues remain unresolved. Another promising technique for resolving non-stationary issues is artificial intelligence (AI) and machine learning in control charts. Yet, this is too intricate for the semiconductor manufacturing process. Another control chart that may be appropriate for non-stationary conditions is the cumulative sum (CUSUM) chart [23]; however, it is difficult to develop and interpret [24,25].

This research aims to develop an automated robust control chart for variable data that can detect unexpected special cause variations in non-stationary conditions and is robust to outliers and normality departures. Although the LV chart is appropriate for stationary conditions, this method employs all samples in the dataset and is thus suitable for automated control chart deployment. However, the LV chart (which uses mean and standard deviation as measure of location and dispersion, respectively) is sensitive to outliers. For this reason, a different estimator that is robust to outliers is needed. Huber's M-estimator is less affected (robust) by outliers and is thus potentially able to resist non-stationary conditions. As there is no universal best control chart [26], this study attempts to apply the Huber M-estimator to a modified LV control chart to monitor the non-stationary process quality of an automated system. Based on the literature review, it is hypothesized that initially normally distributed data may be tampered with by the influence of unexpected outliers when the process is non-stationary. This may cause the automated LV control limits to deviate and become unstable during deployment on the production floor. The scope of this research is limited to a case study conducted on the semiconductor manufacturing process at Dominant Opto Technologies Sdn. Bhd., Malaysia, from February 3 to 27, 2023.

The remainder of this paper is organized as follows: Section 3 reviews the theory of automation and AI in the SPC chart, LV chart, and Huber M-estimator. Sections 4 and 5 present the methodology, results, and discussion. Section 6 presents the conclusion, which includes research implications for the semiconductor manufacturing industry, the limitations of this study, and future research avenues.

3. Theory

3.1 Automation and Artificial Intelligence in the Statistical Process Control Chart

Major SPC tasks involve monitoring the process, identifying the deviated process, and taking corrective actions in sequence [27]. These tasks can be automated following the development of computer-based applications in a computer-integrated manufacturing environment to support Industrial Revolution 4.0 and smart manufacturing [28,29]. Guh proposed the application of hybrid AI techniques to build a real-time SPC system. In the research, Guh integrates an artificial-neural network-based control chart monitoring subsystem and an expert system-based control chart alarm

interpretation subsystem to implement an SPC task automatically. This concept has been adopted in many statistical monitoring software packages used in the semiconductor industry, including JMP by SAS and Minitab. Machine learning, a subset and application of AI, uses mathematical data models to enable computer systems to continue learning without direct instruction [30]. Tran *et al.,* comprehensively discuss how a machine learning approach can be used to resolve issues in classic control charts. They remark that automation and machine learning strategies for SPC chart development must be implemented effectively so that they can adapt to unexpected variations and abnormal patterns in production systems, especially in non-stationary processes. Zhao *et al.,* observe that complex industrial processes exhibit non-stationary process characteristics (e.g., time-varying mean and/or variance) that require time-varying control chart parameters [31]. Non-stationary behaviour can also be statistically tested using the augmented Dickey–Fuller (ADF) test, where the *p*-value is compared against the significance level (typically 0.01) [32]. If the *p*-value is greater than the significance level, the data are considered non-stationary [32]. Lazariv and Schmid propose control charts to be developed using the likelihood ratio method, sequential probability ratio test, and Shiryaev–Roberts procedure to detect a change from the mean structure in state-space processes using simulation studies.

Nonetheless, the challenges of applying machine-learning-based control charts in the industry still exist because most studies thereon use simulated data [13]. It is important to have an automated control chart with robust parameters that can accurately detect changes in a non-stationary process using industrial data [33]. Pittino *et al.,* investigated various machine-learning models for detecting OOC events in an industrial semiconductor manufacturing process. Yet, the proposed machine-learning control chart built in Python is highly dependent on dataset availability for training and better suited to circumstances where outliers are few or absent [34]. Outliers, on the other hand, can be unpredictable or cannot yet be estimated in actual settings [35].

3.2 Levey–Jennings Control Chart

Levey and Jennings suggest that a dataset should remain valid for a certain period before the system can be reassessed. However, there is no clear specification for the number of samples that should be used to derive the population metrics [4]. Previous research shows that the estimated mean and standard deviation over an entire dataset change as the number of observations increases [4]. When assessing the mean and control limits for the entire dataset, using fewer observations tends to underestimate the scale of dispersion [4]. This could result in more OOC events, where the observations fall below the lower control limit (LCL) or above the upper control limit (UCL). Bramwell comments that there is an evident trade-off between generating stable and meaningful limits and the number of samples required to produce them. This method is suitable for automation because it uses all samples in the dataset to calculate the control limits. However, the LV chart is not suitable for deployment to the production floor under non-stationary conditions where unexpected outliers may affect the chart's stability (i.e., time-varying control limits), subsequently causing special cause variations left undetected.

3.3 The Huber M-Estimator

Huber introduced a robust regression estimator to limit the effect of outliers, thereby providing stable results in the presence of outliers [36]. Huber's M-estimation is the most widely used method for robust regressions [37]. Huber's M-estimator $\hat{\theta}_m$ of θ minimizes the sum of the less rapidly increasing functions of the residuals, instead of minimizing the sum of squares of the residuals [38].

M-estimator T_N is a minimization problem ($min!$) explained by Huber (1981, 1996) as follows [39,40]:

$$\sum_{i=1}^{N} \rho(x_i\,;\, T_N) = min! \tag{1}$$

or by an implicit equation

$$\sum_{i=1}^{N} \psi(x_i\,; T_N) = 0, \tag{2}$$

where ρ is an arbitrary function (Huber loss) and ψ is a derivative of ρ

To determine the Huber M-estimates of location, $\widehat{\theta_m}$

$$\sum_{i=1}^{N} \rho(x_i - T_N) = min! \tag{3}$$

or

$$\sum_{i=1}^{N} \psi(x_i - T_N) = 0 \tag{4}$$

if

$$\sum_{i=1}^{N} w_i(x_i - T_N) = 0 \tag{5}$$

with

$$w_i = \frac{\psi(x_i - T_N)}{x_i - T_N} \tag{6}$$

where w_i is the weight depending on the sample.

T_N can be represented as a weighted mean

$$T_N = \frac{\sum_{i=1}^{N}(w_i \times x_i)}{\sum_{i=1}^{N} w_i} \tag{7}$$

by weight depending on the sample used. In the final form, the objective function can be represented as,

$$\rho\,(x) = \begin{cases} 1, & if\ |x_i - T_N| \le c \\ \frac{c}{|x_i - T_N|}, & otherwise \end{cases} \tag{8}$$

where *c* denotes a tuning constant that regulates robustness. Huber recommends using *c* = 1.345 to retain the asymptotic efficiency of an estimator for normally distributed data [41].

In practice, the M-estimates of the location must be supplemented by a simultaneous estimate of the scale [38]. Therefore, robust regression is required. A robust regression procedure requires an iterative solution referred to as iteratively reweighted least squares (IRLS), in which a weighted least-squares fit is performed inside an iteration loop [37]. For each iteration, the least-squares fit uses a set of weights for the observations. The weights are constructed by applying a weight function to the

current residuals. The initial weights are based on the residuals from an initial fit, and the iteration terminates when the convergence criterion is satisfied [37].

To determine the Huber M-estimator of scale if σ is known,

$$\sum_{i=1}^{N} \psi\left(\frac{r_i}{\sigma}\right) x_{ij} \; = \; 0, \; j = 1, \ldots, \rho \tag{9}$$

where r_i is the residual of *i*th given by

$$r_i = \; r_i(\theta) \; = y_i - \sum_{j=1}^{\rho} (x_{ij} \; \times \; \theta_j) \tag{10}$$

However, if σ is unknown, both θ and σ are estimated by minimizing the function below with $a > 0$ over θ and σ by alternately improving $\hat{\theta}$ in the location step and $\hat{\sigma}$ in the scale step [37]. a is the location step parameter.

$$\mathrm{Q}(\theta, \sigma) \; = \sum_{i=1}^{N} [\rho(\tfrac{r_i}{\sigma}) \; + \; a] \; \sigma, \; a > 0 \tag{11}$$

$\hat{\sigma}$ is obtained by the iteration of the below function

$$(\hat{\sigma}^{(m+1)})^2 = \; \frac{1}{N.h} \sum_{i=1}^{N} \chi_d(\tfrac{r_i}{\hat{\sigma}^{(m)}}) \; (\hat{\sigma}^{(m)})^2, \tag{12}$$

with the Huber function given below, where d is the scale parameter

$$\chi_d(x) \; = \begin{cases} x^2/2 & if \; |x| \; < d \\ d^2/2 & otherwise \end{cases} \tag{13}$$

where Huber constant [36],

$$h \; = \; \frac{N - p}{N} (d^2 + \; (1 - d^2)\Phi(d) - 0.5 \; - d\sqrt{2\pi e^{-\frac{1}{2}d^2}} \tag{14}$$

Because Huber's M-estimator is less affected by outliers, it may be able to resist varying parameter changes under non-stationary conditions, which is the focus of this study.

4. Methodology

4.1 Research Conceptual Framework and Procedure

This section outlines the framework and procedures used in the study. The LV chart is compared to a newly proposed chart, the modified Levey-Jennings with a robust Huber M-estimator (MLVHM), under non-stationary conditions. The comparison is evaluated in terms of unexpected special cause variation detection, parameter robustness to outliers, normality departures, and chart stability. First, a programming code is written to load the collected data into the statistical software environment. The code is also programmed to perform statistical analysis, which includes summary statistics, normality tests, and control chart development. The statistical analysis results are saved automatically. This step is repeated daily to collect data during the study. Second, summary statistics are compiled in Microsoft Excel 365 and the absolute dispersion difference between the standard deviation (used in the LV chart) and the robust standard deviation (used in the MLVHM chart) is

computed as the amount of data increases. This is essential for verifying whether the statistics fluctuate over a non-stationary period. The data is also tested using the ADF test in R-studio software at 0.01 significance level to confirm the non-stationary behaviour. Subsequently, the control limit deviation movement of both charts is tracked to verify their stability.

This case study was conducted at Dominant Opto Technologies, Sdn. Bhd., Malaysia. A semiconductor company manufactures light-emitting-diodes and performs periodic quality checks on the plating process using an automated inspection machine that employs image processing to convert the visual representation of the plated surface colour into digital form. The digital form of the image processing is referred to as the measurement of the surface plating tone which must be monitored to detect contamination on the plated surface and can result in product quality concerns. The inspection machine evaluates the surface plating tone of 64 samples daily for each plating machine for Product Z. The 64 samples indicate the number of inspection points on the plating surface of the product. Therefore, the subgroup size in this study is $n = 64$. The number of subgroups collected is $m = 31$. Hence, the total number of samples inspected (n x m) during the study over a period of 23 days was *N* = 1984 (the raw data are available in the supplementary data).

This is the primary data point where the data are collected automatically, as measurements are made using an automated inspection machine. Each subgroup's data, which are saved in a .csv file, will be digitalized and used as source data for automated SPC development in the JMP Version 17.0 software. An algorithm (code) was written in the JSL programming language to automate the development of summary statistics with a normality test chart and SPC charts that monitor the increase in subgroup data for both the LV and MLVHM charts. A common chart type in JMP is the LV chart; however, to create an MLVHM chart, the LV control limit table must be replaced using robust statistics and limits. JMP computes a robust mean and standard deviation based on the Huber M-estimation method using all measurements available at a given point in time [38]. In JMP, the Huber M-estimator of location and scale are computed using an iteratively robust regression procedure called "ROBUSTREG" [38]. JMP hardcodes the default scale parameter *d* for the Huber M-estimate as 2.5 and the tuning constant for the weight function *c* is 1.345 [38]. Regarding the MLVHM chart, the control limits are computed using the following formula:

Centre line, CL : $\widehat{\theta_m}$ (15)

Upper control limit (UCL) : $\widehat{\theta_m} + k * \hat{\sigma}$ (16)

Lower control limit (LCL) : $\widehat{\theta_m} - k * \hat{\sigma}$ (17)

where $\hat{\theta}_m$ represents the Huber M-estimate for location, $\hat{\sigma}$ represents the Huber M-estimate for scale, and *k*-sigma = 3.

Microsoft Task Scheduler Ver. 1.0 is used to automatically schedule the running of the JSL code daily. Figure 1 presents a flowchart of the steps used to automate the SPC charts in this study.

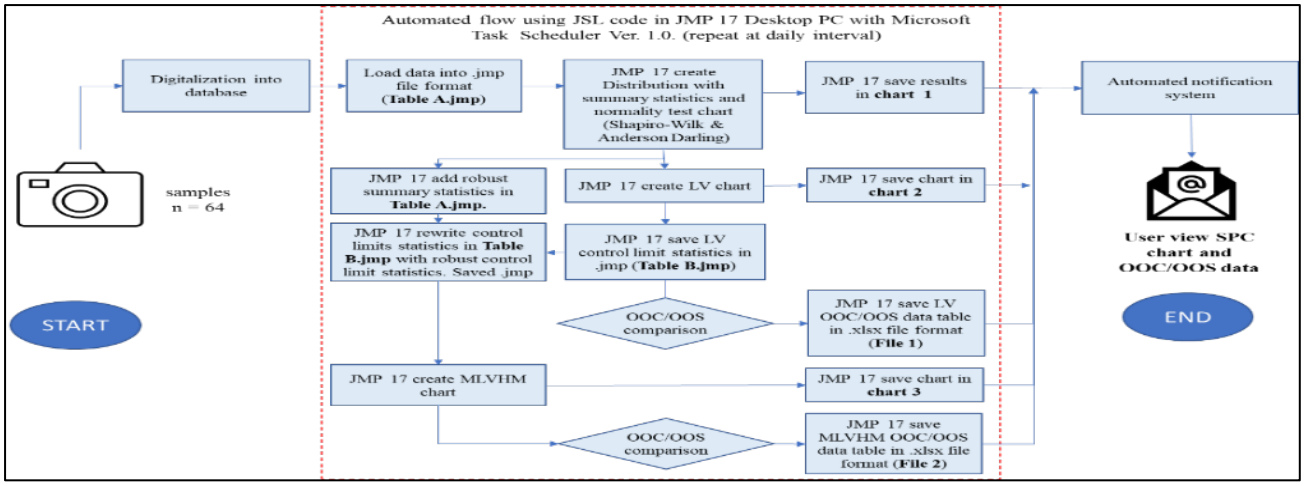


Fig. 1. Procedure flow chart of the automated SPC chart. The red box indicates the scope of this research

The summary statistics and absolute dispersion difference between the standard deviation and robust standard deviation as the number of subgroups increases are computed. The absolute dispersion difference is given by

$$Abs.DispersionDiff_i = \frac{|\hat{\sigma}_{std.i} - \hat{\sigma}_i|}{\hat{\sigma}_{std.i}} \times 100\% \, , i = 1,2,3 \ldots . m \tag{18}$$

The control limit deviation, VCD, is further calculated using the formula below to assess the stability of the computed limit towards the initial control limits (UCL_1 and LCL_1) over non-stationary data. This erratic VCD movement implies that the chart is unreliable for deployment on the production floor.

$$VCD_i = \frac{\sum_i^m |UCL_i - UCL_1| + \sum_i^m |LCL_i - LCL_1|}{2} \, , i = 2,3 \ldots . m \tag{19}$$

where $m = 31$ in this case study, and the denominator is 2 because there are two sides from the centerline (upper and lower).

The control limits of both charts are compared using the Levene's test, a statistical test for unequal variances with a significance level of 0.01, which is less sensitive towards normality departures [42]. The statistical test is used to verify whether the variances of the LV and MLVHM control limits are equal, and if they are not equal, which chart is more reliable and stable.

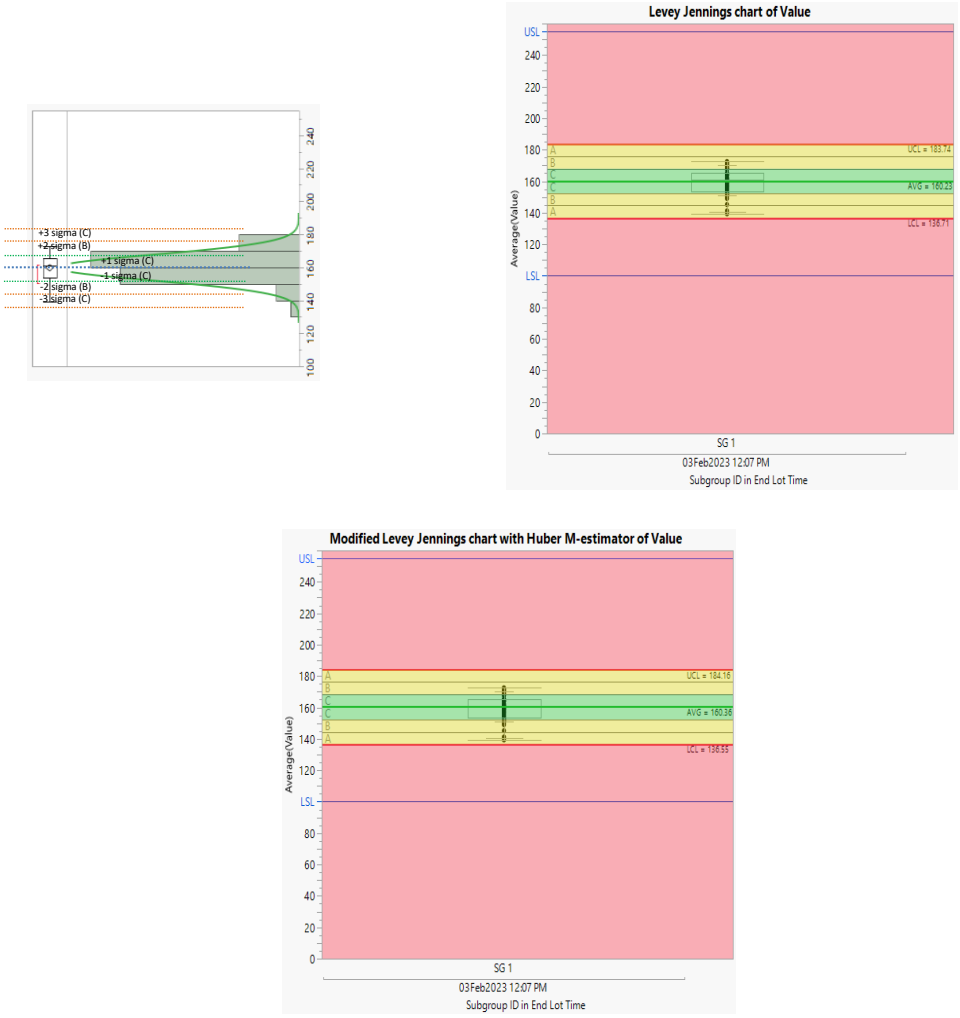
5. Results and Discussion

This section presents automated SPC charts, summary statistics, and normality test results of non-stationary data computed using JMP.

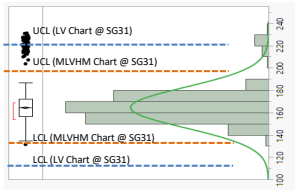
Figure 2 shows the SPC charts of the non-stationary dataset at the beginning of the study (subgroup ID 1) and at the end of the study (subgroups ID 1–31) with respect to the Gaussian curve for normality visualization. "SG" is the prefix of the subgroup ID. For example, "SG1" refers to subgroup ID 1. Each subgroup's data points are arranged in order of time and in a vertical boxplot for easy visualization. The average (mean) of each subgroup is indicated visually by the connected grey line between the subgroups in the LV chart and the purple line in the MLVHM chart. The green horizontal line represents the centre of the SPC chart. The green (C) and yellow (B and A) areas in the SPC chart indicate the 1-sigma and 2-to-3-sigma region, respectively. The red area indicates more

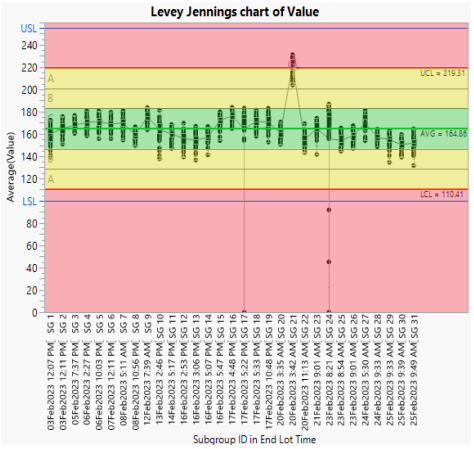
than 3-sigma away from the centre green line or "OOC" region. Notably, USL and LSL are two arbitrary reference limits added to the charts that remain constant over the period; however, the computed control limits vary as the dataset is updated. Furthermore, OOS were defined as observations outside the reference limit, while OOC was defined as a subgroup of mean observations outside the control limit. (Note: The reference lines used in this case study are for illustration only and do not imply the actual reference limits used on the actual production floor.)

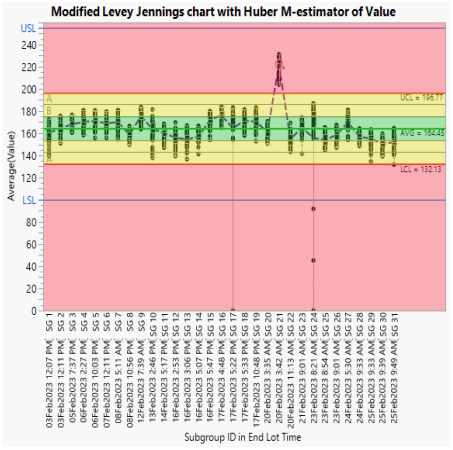
Figure 2(a) shows the SPC limits at the beginning of the study (SG1) for both charts. The Gaussian curve for SG1 showed that the distribution was normal at the beginning of the study. The control limits at SG1 are similar in both charts. Figure 2(b) shows the SPC control limits at the end of the study (SG31). The Gaussian curve at SG31 indicates that the distribution deviated from normality. Figure 2(b) shows that unexpected outliers have occurred at SG17 (1 outlier in 64 samples) and SG24 (3 outliers in 64 samples). Upon observing the outliers in the SPC chart at the respective subgroup intervals, the engineer can perform an investigation for corrective action. A process-mean shift was detected at SG21 in both charts. However, the LV chart marginally detected the process-mean shift (almost half of the 64 samples fell above the UCL_{LV}), as indicated by the Gaussian curve. In the MLVHM chart, all 64 samples fell outside the UCL_{MLVHM} and were isolated. The process of mean-shift detection helps engineers take corrective action immediately.



(a) Gaussian curve with respect to the automated LV and MLVHM control chart at subgroup ID1







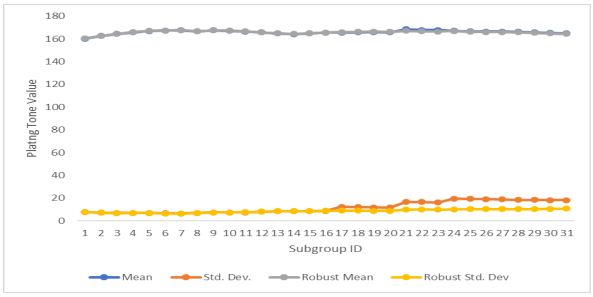
(b) Gaussian curve with respect to the automated LV and MLVHM control chart at subgroup ID31

Fig. 2. SPC chart monitoring of subgroup ID 1 to 31

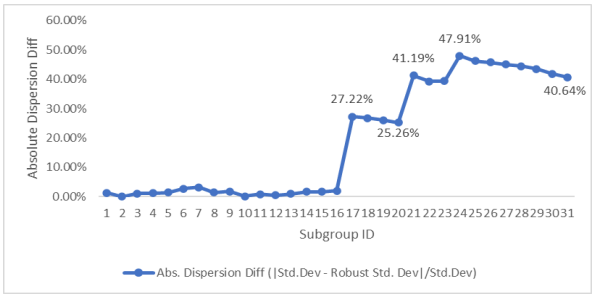
Generally, the LV control limits are wider than the MLVHM control limit at the end of the study (SG31), although the limits are almost the same at the starting point (SG1) (refer to the yellow area in Figure 2). Outliers must be shown in the SPC chart because they are real data from automated systems. Excluding outliers from SPC charts will mislead practitioners regarding potential loopholes in the system for continuous quality improvement. Figure 3 depicts the movement of the statistical parameters, control limit, and VCD regarding time or subgroup intervals.

As shown in Figure 3(a), there is no evident difference in location between the mean and robust Huber M-estimates. However, the statistical parameters for dispersion showed a time-varying standard deviation starting at SG17, indicating non-stationary behaviour in the process. There are clear differences between the standard deviation and robust Huber M-estimates of the scale for non-stationary data. The ADF test also reveals that the *p*-values of the statistical parameters for location and dispersion are more than 0.01 significance level, confirming that the process is non-stationary (the statistical analysis related to non-stationary behaviour is available in the supplementary data). The standard deviation used in the LV chart increases with the presence of outliers (SG17 and SG24) and process-mean shift (SG21). This finding is similar to that in the previous literature (Bramwell).

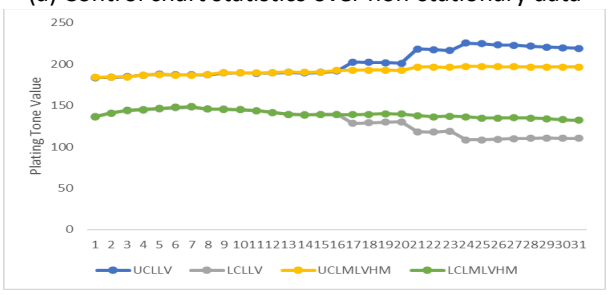
Figure 3(b) shows that the absolute difference in dispersion between the standard deviation and robust Huber M-estimates of scale fluctuates every time outliers (SG17 and SG24) and process-mean shifts (SG21) are present. The fluctuations range from 25.26% to 47.91%. This figure shows that in contrast to the Huber M-estimates of scale, the standard deviation is vulnerable to outliers.



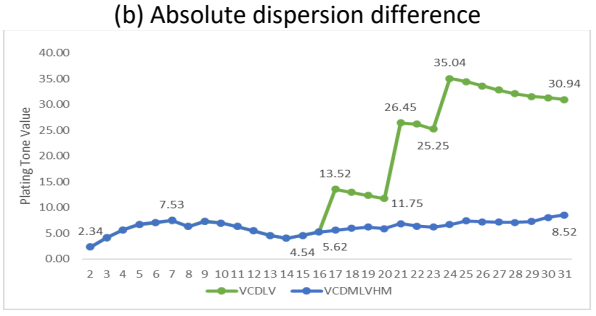
(a) Control chart statistics over non-stationary data



(b) Absolute dispersion difference



(c) Movement of control limits over non-stationary data between LV and MLVHM



(d) Movement of control limit deviation, VCD over non-stationary data between LV and MLVHM

Fig. 3. Results of statistics and control limit monitoring over non-stationary data

Figure 3(c) and (d) show the movement of the control limits every time new subgroup data are added. In Figure 3(c), the LV control limits widen with the unexpected presence of outliers and a mean shift. However, the MLVHM control limits are mostly stable over non-stationary data, even in the presence of 3% of outliers in the control chart. Figure 3(d) confirms the extent of the control limit changes for both charts in Figure 3(c). It shows an increasing trend of the control limit deviation (VCD) from SG2 to SG7 because both charts automatically adapt to common variations inherent to the plating process (such as machine and material variation).

As the adaptation to common variations continues, the trend decreases from SG7 to SG14. The VCD of the LV and MLVHM charts is similar and stabilizes at SG14 (N = 896), approximately 4.01 to 5.62 before outliers appear in the non-stationary process. Interestingly, the VCD of the MLVHM chart stabilizes between 5.62 and 8.52, even in the presence of outliers in the system. In contrast, the VCD of the LV chart continues to fluctuate erratically between 11.75 and 35.04. With both common and special cause variations, the non-stationary LV chart control limit widens to 35.04. However, the MLVHM chart widened to only 8.52.

The results presented in Figure 3 are important for quality control because they show that the LV chart is not reliable for detecting special cause variations over a non-stationary process with outliers. Levene's test shows that the p-value comparing the LV and MLVHM control limits is less than the significance level of 0.01. Thus, the variances between these two methods are not equal, and MLVHM control limits show lower standard deviation around 4.3 to 4.6 (versus LV control limits with a standard deviation around 14.7 to 15.6). Therefore, if users select the LV chart to be automated for the production floor, an unexpected process-mean shift (which consists of several outliers and is categorically defined as special cause variation) may remain undetected in the future. In comparison, the automated MLVHM chart has been demonstrated as consistent and stable in detecting special cause variations in non-stationary processes with outliers (statistical analysis to support this claim is available in the supplementary data).

6. Conclusions

This case study research shows that the proposed MLVHM chart is more stable and suitable for detecting special cause variations than the LV chart in the presence of unexpected outliers and departures from normality throughout the non-stationary period. By selecting the MLVHM for this process, the automatic system can run continuously for control chart deployment with minimal supervision. This case study is important in the semiconductor manufacturing industry because it allows an organization to set up a robust automated control chart for variable data that can detect special cause variations in a non-stationary process. The inability to detect unexpected special cause variations for immediate corrective action will incur higher losses to the organization owing to rework or scrapped products. Consistent with the latest trends in the Industrial Revolution 4.0 era and smart manufacturing in the semiconductor industry, the proposed method works better in an automated environment. Therefore, the MLVHM chart is superior to removing outliers for control-limit computation, particularly when the control chart has been deployed to the production floor. Exclusion of outliers in dataset can lead to loss of information [35]. The interruption of manufacturing processes after control chart deployment is not productive. This case study, however, has few limitations. First, 64 samples were used in the initial subgroup for the computation of the initial control limit. The effect of a sample size of < 64 samples within the subgroup was not evaluated. Second, to utilize the proposed method, it is necessary to use statistical software that can support the computation of robust Huber's M-estimates. In this case, JMP Version 17.0 software from SAS was used. This study does not assess different statistical software to estimate the robust Huber M-statistics. The JMP software in SAS hardcodes the value of the Huber constant for the M-estimates. Therefore, the impact on the MLVHM chart if the Huber constant varied is not known. Another well-known mathematical software such as MATLAB R2020b [43], could potentially be used to replicate this study for real-time application. Future research should investigate the effect of a smaller sample size within the subgroup, and the effect of different Huber constant values for M-estimates on the MLVHM chart using industrial data. This is important to ensure that the MLVHM chart can be applied across different quality-monitoring scenarios in the semiconductor industry. As the LV chart was developed for clinical laboratory studies, future research on automated MLVHM applications in clinical studies would be beneficial. Robust M-estimates have immense potential for industrial SPC applications but are rarely used in practice. Lack of creativity and innovation in AI has been quoted as one of main challenges faced by manufacturing firms to adopt Industrial Revolution 4.0 [44]. Therefore, this case study can be used as a practical reference by industrial practitioners to set up automatic control charts for non-stationary processes.

Acknowledgement
This research was not funded by any grant. The author would like to thank Dominant Opto Technologies, Sdn. Bhd. for providing the data and the statistical software JMP Version 17.0 for this case study. The author also thanks Dr. Adilah Abdul Ghapor and Dr. Muzalwana Abdul Talib at the University of Malaya for discussions and suggestions on the research concept. The author further acknowledges the contributions of practitioners in JMP user communities for coding development using JSL language programming. The author thanks Judy for proofreading the manuscript. Special thanks to the anonymous reviewers, referees and editors for their insightful comments and suggestions that helped to improve the structure and content of this research paper.

References

[1] Karkalousos, Petros, and Angelos Evangelopoulos. "The history of statistical quality control in clinical chemistry and haematology (1950–2010)." *International Journal of Biomedical Laboratory Science (IJBLS)* 4, no. 1 (2015): 1-11.
[2] Montgomery, Douglas C. *Introduction to statistical quality control*. John wiley & sons, 2019.
[3] Levey, Stanley, and E. R. Jennings. "The use of control charts in the clinical laboratory." *American Journal of Clinical Pathology* 20, no. 11_ts (1950): 1059-1066. https://doi.org/10.1093/ajcp/20.11_ts.1059
[4] Bramwell, David. "An introduction to statistical process control in research proteomics." *Journal of proteomics* 95 (2013): 3-21. https://doi.org/10.1016/j.jprot.2013.06.010
[5] JMP. "Statistical Details for Levey-Jennings Charts. https://www.jmp.com/support/help/en/17.0/index.shtml#page/jmp/statistical-details-for-leveyjennings-charts.shtml
[6] US Department of Commerce. https://www.itl.nist.gov/div898/handbook/pmc/section3/pmc321.htm
[7] Huat, Ng Kooi, and Habshah Midi. "Robust individuals control chart for shifts in process mean and variance." In *Proceedings of the 9th WSEAS international conference on Applications of computer engineering*, pp. 270-275. 2010.
[8] Huat, Ng Kooi, and Habshah Midi. "The performance of robust control chart for change in variance." *Trends in Applied Sciences Research* 6, no. 10 (2011): 1172. https://doi.org/10.3923/tasr.2011.1172.1184
[9] Abu-Shawiesh, Moustafa O., and Mokhtar B. Abdullah. "New robust statistical process control chart for location." *Quality Engineering* 12, no. 2 (1999): 149-159. https://doi.org/10.1080/08982119908962572
[10] Abu-Shawiesh, Moustafa OA. "A simple robust control chart based on MAD." *Journal of Mathematics and Statistics* 4, no. 2 (2008): 102. https://doi.org/10.3844/jmssp.2008.102.107
[11] Schoonhoven, Marit, Muhammad Riaz, and Ronald JMM Does. "Design and analysis of control charts for standard deviation with estimated parameters." *Journal of Quality Technology* 43, no. 4 (2011): 307-333. https://doi.org/10.1080/00224065.2011.11917867
[12] Mathieu, Timothée. "M-estimation and Median of Means applied to statistical learning." PhD diss., Université Paris-Saclay, 2021.
[13] Tran, Phuong Hanh, Adel Ahmadi Nadi, Thi Hien Nguyen, Kim Duc Tran, and Kim Phuc Tran. "Application of machine learning in statistical process control charts: A survey and perspective." In *Control charts and machine learning for anomaly detection in manufacturing*, pp. 7-42. Springer, Cham, 2022. https://doi.org/10.1007/978-3-030-83819-5_2
[14] Schoonhoven, Marit, and Ronald JMM Does. "A robust control chart." *Quality and Reliability Engineering International* 29, no. 7 (2013): 951-970. https://doi.org/10.1002/qre.1447
[15] Stoumbos, Zachary G., Marion R. Reynolds Jr, Thomas P. Ryan, and William H. Woodall. "The state of statistical process control as we proceed into the 21st century." *Journal of the American Statistical Association* 95, no. 451 (2000): 992-998. https://doi.org/10.1080/01621459.2000.10474292
[16] Nazir, Hafiz Zafar. *Robust control charts in statistical process control*. Universiteit van Amsterdam [Host], 2014.
[17] Chen, Qian, Uwe Kruger, and Andrew YT Leung. "Cointegration testing method for monitoring nonstationary processes." *Industrial & Engineering Chemistry Research* 48, no. 7 (2009): 3533-3543. https://doi.org/10.1021/ie801611s
[18] Liu, Jialin, and Ding-Sou Chen. "Nonstationary fault detection and diagnosis for multimode processes." *AIChE journal* 56, no. 1 (2010): 207-219. https://doi.org/10.1002/aic.11999
[19] Lazariv, Taras, and Wolfgang Schmid. "Surveillance of non-stationary processes." *AStA Advances in Statistical Analysis* 103 (2019): 305-331. https://doi.org/10.1007/s10182-018-00330-4
[20] Raji, Ishaq Adeyanju, Muhammad Hisyam Lee, Muhammad Riaz, Mu'azu Ramat Abujiya, and Nasir Abbas. "Outliers detection models in shewhart control charts; An application in photolithography: A semiconductor manufacturing industry." *Mathematics* 8, no. 5 (2020): 857. https://doi.org/10.3390/math8050857
[21] Atalay, Murat, Murat Caner Testik, Serhan Duran, and Christian H. Weiß. "Guidelines for automating Phase I of control charts by considering effects on Phase-II performance of individuals control chart." *Quality Engineering* 32, no. 2 (2020): 223-243. https://doi.org/10.1080/08982112.2019.1641208
[22] Woodall, William H. "Bridging the gap between theory and practice in basic statistical process monitoring." *Quality Engineering* 29, no. 1 (2017): 2-15.
[23] Espadinha-Cruz, Pedro, Radu Godina, and Eduardo MG Rodrigues. "A review of data mining applications in semiconductor manufacturing." *Processes* 9, no. 2 (2021): 305. https://doi.org/10.3390/pr9020305
[24] Woodall, William H., George Rakovich, and Stefan H. Steiner. "An overview and critique of the use of cumulative sum methods with surgical learning curve data." *Statistics in Medicine* 40, no. 6 (2021): 1400-1413. https://doi.org/10.1002/sim.8847

[25] Prajapati, D. R. "Effectiveness of conventional CUSUM control chart for correlated observations." *International Journal of Modeling and Optimization* 5, no. 2 (2015): 135. https://doi.org/10.7763/IJMO.2015.V5.449
[26] Clifford, Paul C. "Control charts without calculations." *Industrial Quality Control* 15, no. 11 (1959): 40-44.
[27] Guh, Ruey-Shiang. "Integrating artificial intelligence into on-line statistical process control." *Quality and reliability engineering international* 19, no. 1 (2003): 1-20. https://doi.org/10.1002/qre.510
[28] Nasir, Norazlin, Ahmad Yusairi Bani Hashim, M. H. F. M. Fauadi, and Teruaki Ito. "Statistical process control as a traceability tools for industry 4.0." *Proceedings of Mechanical Engineering Research Day 2018* 2018 (2018): 89-90.
[29] Hon, Lim Chong, and Chin Jeng Feng. "Cloud-based statistical process control mobile application development for smart manufacturing."
[30] Microsoft. "Artificial intelligence (AI) vs. machine learning (ML)". https://azure.microsoft.com/en-us/solutions/ai/artificial-intelligence-vs-machine-learning/#introduction
[31] Zhao, Chunhui, He Sun, and Feng Tian. "Total variable decomposition based on sparse cointegration analysis for distributed monitoring of nonstationary industrial processes." *IEEE Transactions on Control Systems Technology* 28, no. 4 (2019): 1542-1549. https://doi.org/10.1109/TCST.2019.2908339
[32] Lin, Yuanling, Uwe Kruger, Fengshou Gu, Andrew Ball, and Qian Chen. "Monitoring nonstationary and dynamic trends for practical process fault diagnosis." *Control Engineering Practice* 84 (2019): 139-158. https://doi.org/10.1016/j.conengprac.2018.11.020
[33] Lazariv, Taras, and Wolfgang Schmid. "Challenges in monitoring non-stationary time series." In *Frontiers in Statistical Quality Control 12*, pp. 257-275. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-75295-2_14
[34] Pittino, Federico, Michael Puggl, Thomas Moldaschl, and Christina Hirschl. "Automatic anomaly detection on in-production manufacturing machines using statistical learning methods." *Sensors* 20, no. 8 (2020): 2344. https://doi.org/10.3390/s20082344
[35] Hochkamp, Florian, and Markus Rabe. "Outlier detection in data mining: Exclusion of errors or loss of information?." In *Changing Tides: The New Role of Resilience and Sustainability in Logistics and Supply Chain Management–Innovative Approaches for the Shift to a New Era. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 33*, pp. 91-117. Berlin: epubli GmbH, 2022.
[36] Huber, Peter J. "Robust regression: asymptotics, conjectures and Monte Carlo." *The annals of statistics* (1973): 799-821. https://doi.org/10.1214/aos/1176342503
[37] Chen, Colin. "Paper 265-27 Robust regression and outlier detection with the ROBUSTREG procedure." In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*. 2002.
[38] Inc, S. I. "SAS/STAT 14.2 User's Guide." *Cary, NC, USA: SAS Institute Inc* (2016).
[39] Huber, P. J. "Robust statistics John Wiley & Sons New York." (1981). https://doi.org/10.1002/0471725250
[40] Huber, Peter J. *Robust statistical procedures*. Society for Industrial and Applied Mathematics, 1996. https://doi.org/10.1137/1.9781611970036
[41] Wang, Lili, Chao Zheng, Wen Zhou, and Wen-Xin Zhou. "A new principle for tuning-free Huber regression." *Statistica Sinica* 31, no. 4 (2021): 2153-2177. https://doi.org/10.5705/ss.202019.0045
[42] Levene, H. "Levene test for equality of variances." *Contributions to probability and statistics* (1960): 278-292.
[43] Ab Wahab, Nurul Ain, and Mohd Agos Salim Nasir. "Graphical User Interface for Solving Non-Linear Equations for Undergraduate Students." *International Journal of Advanced Research in Future Ready Learning and Education* 30, no. 1 (2023): 25-34.
[44] Ahmad, Md Fauzi, Noor Athirah Adila Husin, Ahmad Nur Aizat Ahmad, Huda Abdullah, Chan Shiau Wei, and Mohd Nasrun Mohd Nawi. "Digital Transformation: An Exploring Barriers and Challenges Practice of Artificial Intelligence in Manufacturing Firms in Malaysia." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 29, no. 1 (2022): 110-117. https://doi.org/10.37934/araset.29.1.110117
[45] Dahari, Sufinah (2023). G003_SupplementaryData1.xlsx. figshare. Dataset.